Handbook of Exponential and Related Distributions for Engineers and Scientists
Handbook of Exponential and Related Distributions for Engineers and Scientists

Nabendu Pal
University of Louisiana at Lafayette
Lafayette, LA

Chun Jin
Central Connecticut State University
New Britain, CT

Wooi K. Lim
William Paterson University
Wayne, NJ
To our parents and all our teachers for the wonderful things of life that we have today.

Authors
Preface

The motivation to write this book came a few years back when a colleague from one of the engineering departments (University of Louisiana at Lafayette) approached us with a problem that required modeling a real life data set which looked highly positively skewed.

There are many good statistics books available on areas like survival analysis, reliability, and life testing, where one often encounters positively skewed data sets. There are a few recent books that devote entirely on one or a few specific positively skewed probability distributions. These books, which are quite thorough and rigorous, tend to be highly mathematical, and as a result may discourage many applied scientists who may lack a sound background in mathematical statistics. On the other hand, introductory level statistics books, which may appeal to a wider audience, place over-emphasis on the normal distribution (the Bell Curve) due to its popularity, but may not be suitable enough for modeling positively skewed data sets for obvious reasons. Hence we felt the need for a handbook which can explain and discuss the applications of some of the well known positively skewed probability distributions in a simple step by step manner.

We wanted to cover the Exponential Distribution due to its wide applications in modeling lifetime data and its famous Lack of Memory property. Immediately, we felt the need for generalizing this to a few other natural extensions, namely — the Gamma, the Weibull, and the Extreme Value Distributions, which are also popular in engineering fields. A quick glance at the CIS (Current Index in Statistics) shows more than 300 research papers on these few distributions, and naturally this put us in a bind to determine what to cover and what not to cover. As a result, we had to limit ourselves to the aforementioned distributions only, leaving out a few other common positively skewed distributions like the Logistic or the Log – Normal. Even for the above four distributions we wanted to cover as much material as possible, yet didn’t want to sacrifice the conciseness and simplicity of the presentation which, we hope, would be one of the high points of this book. Since statistical computations with the help of numerous mathematical/statistical packages have become very convenient and less time consuming, we decided to trade some of the complex theoretical results with the tools based on simulations.
Statistical simulations give an applied researcher plenty of room to analyze and study the data without getting lost in the mish-mash of theory. Of course, no one (including us) can deny the importance of the theory, and a sound theoretical knowledge is very helpful in statistical analyses. But we had to walk a fine line between covering sufficient rigorous theoretical results (which may not evoke much appeal to a nonstatistician), and simulation techniques based on simple ideas by which anyone with basic statistical knowledge can model a data set and visualize its implications. One effect of this approach is that we ended up with a large number of graphs.

“In attempts to understand ... it is a good idea”, — as Socrates once said — “to begin at the beginning”. So we have started the book from scratch. If the reader needs a little brush-up of the elementary or the standard concepts of mathematics/statistics, he/she still finds those covered here briefly. The book has been divided in two parts. The first part has five chapters covering most of the basic concepts and general ideas needed (or used) in statistics — all the way up to inference. One chapter (Chapter 5) is devoted solely to statistical computations including the increasingly popular bootstrap method. The second part of the book, which also has another five chapters, covers the four positively skewed distributions mentioned above. For each distribution we have followed a clear-cut path (divided in four sections):

(i) Discuss the general properties and show the applicability with a few real data sets

(ii) Characterize the distribution (some theoretical results without much derivations or proofs)

(iii) Estimation of parameters and related inferences

(iv) Goodness of fit test (to see how well the specific distribution fits the given data)

The last chapter (Chapter 10) deals with system reliability, mainly for series and parallel systems, which may be of interest to the engineers. One can find a heavy usage of simulations (with a lot of graphs) since the theoretical results are hopelessly complicated. All the simulations and computations have been carried out by the package ‘Mathematica’ (with
the addition package ‘mathStatica.m’); and the graphs are drawn using
GNUPLOT.

We sincerely hope that the book will be useful as a handbook (or
as a reference book) for the applied researchers who are more interested
in data analysis rather than mathematical developments of the meth-
ods. Apart from engineers and applied scientists, seniors and beginning
graduate students of statistics can benefit greatly from this book. The
references cited here, especially those with real life data sets, can be a
treasure trove for any one who is interested in statistics.

We would like to thank many of our colleagues for their help. A special
thanks to Dr. Robert Crouse (Central Connecticut State University) for
his assistance with some of the computations and editorial help.

We greatly appreciate the encouragement that we have received from
our spouses (Pallavi, Xiaoyan, and Guo-Zhen) and children (Nabapallav
and Devpallav (Pal), Daniel (Jin), and Alice Wu and Gary Wu (Lim))
who understood the meaning of the lost family time. Without their
continuous support this book could not have been possible.

Last but not least, we are thankful to Mr. Robert Stern and Ms.
Theresa Delforn (of CRC Press) for their infinite patience. While we
worked on this book at a very slow rate of convergence and missed the
deadline repeatedly to submit the manuscript, they were generous enough
to extend the deadline without any question every time we requested it,
irrespective of the excuse we gave the previous time. Truly a memoryless
property exhibited by the exponential distribution only.

Nabendu Pal, University of Louisiana at Lafayette

Chun Jin, Central Connecticut State University

Wooi K. Lim, William Paterson University
About the Authors

Nabendu Pal is Professor of Statistics in the Department of Mathematics, University of Louisiana at Lafayette (formerly University of Southwestern Louisiana), where he has been a faculty member since August 1989. Earlier (in Spring 1989), he was a visiting faculty in the Department of Mathematics and Statistics, University of Arkansas, Fayetteville. He received his master’s in statistics (M. Stat) degree from the Indian Statistical Institute (ISI), Calcutta, in 1986, and Ph.D. in statistics from the University of Maryland Baltimore County in 1989. His areas of interest include Decision Theory, Biostatistics, Reliability, and Life Testing. He has served as the department’s graduate coordinator from August 2001 to July 2005. He has also co-authored the book *Statistics: Concepts and Applications* (by Prentice-Hall of India, 2005).

Chun Jin is Professor of Statistics in the Department of Mathematical Sciences at Central Connecticut State University (CCSU). He received his MS degree in mathematics from the University of Southern Mississippi in 1988 and his Ph.D. in statistics from the University of Louisiana at Lafayette (formerly University of Southwestern Louisiana) in 1992. Before joining the faculty of CCSU in 1994, he taught in the Department of Mathematics at the University of New Orleans for two years. His main area of interest is Multivariate Analysis. He has also co-authored the book *Ready, Set, Run! A Student Guide to SAS Software for Microsoft Windows* (by Mayfield Publishing Company, 1998).

Wooi K. Lim received his Ph.D. in statistics from the University of Louisiana at Lafayette (formerly University of Southwestern Louisiana) in December 1996. Since August 1999, he is an Assistant Professor in the Department of Mathematics, William Paterson University. Earlier (1995 to 1999), he taught in the Department of Mathematics, University of New Orleans. His areas of interest include Decision Theory, Statistical Simulations, and Numerical Computations.
Acronyms

AM : Arithmatic mean
BEL : Bayes expected loss
cdf : Cumulative distribution function (of a random variable)
cf : Characteristic function
CI : Confidence interval
CLT : Central Limit Theorem
corr : Correlation coefficient
df : Degrees of freedom
EV-I (II, III) : Extreme value distribution of Type-I (or Type-II, or Type-III)
GM : Geometric mean
HM : Harmonic mean
iid : Independent and identically distributed (for a set of random variables)
ln c : Natural logarithm of the value c
LRT : Likelihood ratio test
mgf : Moment generating function
MLE : Maximum likelihood estimator
MME : Method of moment(s) estimator
MMLE : Modified maximum likelihood estimator
MSE : Mean squared error
pdf : Probability density function (of a continuous random variable)
pmf : Probability mass function (of a discrete random variable)
PB-Bias (MSE) : Parametric bootstrap bias (mean squared error)
r.v. : Random variable (or vector)

SB-Bias (MSE) : Semiparametric bootstrap bias (mean squared error)

SEL : Squared error loss function

UMVUE : Unique minimum variance unbiased estimator

Var : Variance
Notations

A ∪ B : Union of events A and B.

A ∩ B : Intersection of events A and B.

A^c : Complement of event A.

A ⊆ B : Event A is a subset of event B.

B(n, θ) : Binomial distribution where n is the number of trials, and θ is the probability of success in a single trial.

B(α, β) : Beta distribution with first parameter α and second parameter β.

corr(X, Y) : Correlation coefficient between the random variables X and Y.

CU(θ₁, θ₂) : Continuous uniform distribution over the interval (θ₁, θ₂).

DU(θ₁, θ₂) : Discrete uniform distribution over the countable set {θ₁, θ₁+1, ..., θ₂}.

Example a.b.c. : Example number c of Section b of Chapter a. [Same for Theorem, Remark, Lemma, Corollary].

E(X) : Expectation or expected value of the random variable X.

E(X|Y) : Conditional expectation of the random variable X for the fixed value of another random variable Y.

EV − I(β, μ) : Type-I extreme value distribution with location parameter μ and scale parameter β.

EV − II(α, β, μ) : Type-II extreme value distribution with location parameter μ and scale parameter β and shape parameter α.

EV − III(α, β, μ) : Type-III extreme value distribution with location parameter μ and scale parameter β and shape parameter α.

F_{k₁, k₂} : (Snedecor’s) F distribution with first (or numerator) df k₁ and second (or denominator) df k₂.
$\mathcal{F}$ (or $\mathcal{F}_X$) : Usually indicates a family of probability distributions (identified by pmfs or pdfs indexed by a parameter $\theta$ (possibly vector valued)).

$F_X(x)$ (or $F(x)$) : Usually denotes a cdf at $x$.

$\bar{F}(t)$ : For a random variable $X$ with cdf $F(\cdot)$, $\bar{F}(t) = 1 - F(t) = P(X > t)$, reliability at time $t$ if $X$ indicates the lifetime.

$\bar{F}_S(t)$ : Reliability of a series system at time $t$.

$\bar{F}_P(t)$ : Reliability of a parallel system at time $t$.

$G(\alpha, \beta, \mu)$ : (Three parameter) Gamma distribution with location parameter $\mu$, shape parameter $\alpha$ and scale parameter $\beta$.

$G(\alpha, \beta)$ : (Two parameter) Gamma distribution with shape parameter $\alpha$ and scale parameter $\beta$. Also same as $G(\alpha, \beta, 0)$.

$G(\theta)$ : Geometric distribution where $\theta$ is the probability of success in a single trial.

$LN(\mu, \sigma^2)$ : Lognormal distribution with parameters $\mu$ and $\sigma^2$.

$L(\mu, \sigma)$ : Logistic distribution with location parameter $\mu$ and scale parameter $\sigma$.

$\mu_X$ (or $\mu$) : Same as $E(X)$.

$\mu_{(k)}^\prime$ : $k^{th}$ raw moment, i.e., $E(X^k)$, expected value of the $k^{th}$ power of the random variable $X$.

$\mu_{(k)}$ : $k^{th}$ central moment, i.e., $E((X - E(X))^k)$, expected value of the $k^{th}$ power of $(X - E(X))$.

$NB(r, \theta)$ : Negative binomial distribution where $r$ is the fixed number of successes to be obtained, and $\theta$ is the probability of success in a single trial.

$O(n)$ : A sequence $\{c_n\}$ such that $(c_n/n)$ is bounded as $n \to \infty$.

$o(n)$ : A sequence $\{c_n\}$ such that $(c_n/n) \to 0$ as $n \to \infty$.

$P(A|B)$ : Conditional probability of the event $A$ given the event $B$. 

© 2006 by Taylor & Francis Group, LLC
\( P(\alpha, \beta) \) : Pareto distribution with parameters \( \alpha \) and \( \beta \).

\( \text{Poi}(\theta) \) : Poisson distribution with (intensity) parameter \( \theta \).

\( N(\mu, \sigma^2) \) : Normal distribution with location parameter (also mean) \( \mu \) and scale parameter (also variance) \( \sigma^2 \).

\( \sigma^2_X \) (or \( \sigma^2 \)) : Variance of the random variable \( X \), i.e., \( E(X - E(X))^2 \).

\( \sigma_X \) (or \( \sigma \)) : \( \sqrt{\sigma^2_X} \), standard deviation of the random variable \( X \).

\( \rho_{12} \) (or \( \rho_{X,Y} \) or \( \rho \)) : Usually indicates the correlation coefficient between two random variables.

\( (a_{ij}) \) : A matrix where \( a_{ij} \) is the element of \( i^{th} \) row and \( j^{th} \) column.

\( \text{diag}(d_1, \ldots, d_n) \) : A square matrix where \( (i, i) \)-element is \( d_i \), \( 1 \leq i \leq n \); and all other elements are 0.

\( \chi^2_k \) : Chi-square distribution with \( k \) degrees of freedom (df). Same as \( G(k/2, 1/2) \) distribution.

\( t_k(\theta) \) : (Student's) \( t \)-distribution with \( k \) degrees of freedom and location parameter \( \theta \).

\( t_k \) : Same as \( t_k(0) \).

\( W(\alpha, \beta) \) : (Two parameter) Weibull distribution with shape parameter \( \alpha \) and scale parameter \( \beta \).

\( W(\alpha, \beta, \mu) \) : (Three parameter) Weibull distribution with location parameter \( \mu \), shape parameter \( \alpha \) and scale parameter \( \beta \).

\( \Gamma(\alpha) \) : \( \int_0^\infty \exp(-x)x^{\alpha-1} \, dx \), the gamma function evaluated at \( \alpha \), \( \alpha > 0 \).

\( X_k \overset{d}{\rightarrow} X \) : The sequence of random variables \( \{X_k\} \) converges in distribution to the random variable \( X \) as \( k \to \infty \).

\( X_k \overset{p}{\rightarrow} X \) : The sequence of random variables \( \{X_k\} \) converges in probability to the random variable \( X \) as \( k \to \infty \).

\( X_k \overset{a.s.}{\rightarrow} X \) : The sequence of random variables \( \{X_k\} \) converges almost surely to the random variable \( X \) as \( k \to \infty \).

\( X \sim f(x) \) : Indicates that the random variable \( X \) has pmf or pdf \( f(x) \).
\[ \mathbb{R} : \text{Real line, the set of all real numbers.} \]

\[ F_{k_1, k_2, \alpha} : \text{Right (or upper) tail } \alpha \text{-probability cut-off point of the } F_{k_1, k_2} \text{ distribution.} \]

\[ z_\alpha : \text{Right (or upper) tail } \alpha \text{-probability cut-off point of the standard normal distribution.} \]

\[ t_{k, \alpha} : \text{Right (or upper) tail } \alpha \text{-probability cut-off point of the } t_k \text{ distribution.} \]

\[ \chi^2_{k, \alpha} : \text{Right (or upper) tail } \alpha \text{-probability cut-off point of the } \chi^2_k \text{ distribution.} \]

\[ \Theta : \text{Usually indicates the unrestricted parameter space.} \]

\[ \text{Exp}(\beta, \mu) : \text{(Two-parameter) exponential distribution with location parameter } \mu \text{ and scale parameter } \beta. \]

\[ \text{Exp}(\beta) : \text{Same as } \text{Exp}(\beta, 0). \]

\[ \Gamma_c(\alpha) : \int_0^c \exp(-x)x^{\alpha-1} \, dx. \]

\[ L(\delta(x), \theta) : \text{Loss function at } \theta \text{ for the decision } \delta(x) \text{ (with data } X = x). \]

\[ R(\delta, \theta) : \text{Risk function of the decision } \delta \text{ at } \theta. \]

\[ r(\delta, \theta) : \text{Bayes risk of the decision } \delta \text{ with prior distribution } \pi \text{ of } \theta \text{ on } \Theta. \]

\[ \hat{\theta}_U : \text{Unbiased estimator of } \theta. \]

\[ \hat{\theta}_{\text{ML}} : \text{MLE of } \theta. \]

\[ \hat{\theta}_{\text{MMLE}} : \text{MMLE of } \theta. \]

\[ \hat{\theta}_{\text{MME}} : \text{MME of } \theta. \]
Contents

Preface ................................................................. vii

About the Authors ................................................... xi

Acronyms ............................................................ xiii

Notations ............................................................. xv

List of Figures ....................................................... xxiii

List of Tables ....................................................... xxix

Part-I

General Statistical Theory 1

1 Basic Concepts 3
  1.1 The System of Real Numbers ............................ 3
  1.2 Some Useful Algebraic Results ......................... 4
    Some Algebraic Inequalities ............................. 7
    Matrices .................................................. 7
    Calculus .................................................. 10
    Order ‘big oh’ and ‘small oh’ ......................... 12
  1.3 Set Theory ............................................... 13
  1.4 Introduction to Probability Theory ..................... 14
    Meaning of Probability .................................. 17
    Classical Definition of Probability ................. 18
    Probability as Long Term Relative Frequency .......... 20
Conditional Probability ................. 21
Bayes’ Principle ........................ 23
1.5 Random Variables ..................... 25
Expectations of Random Variables ........ 33
1.6 Joint Probability Distributions ............. 35
1.7 Moment Generating Function ................. 44
Some Moment Inequalities .................. 46
1.8 Order Statistics ........................ 47
1.9 Characteristic Function .................. 49

2 Some Common Probability Distributions 51
2.1 Discrete Distributions .................... 52
Discrete Uniform Distribution ................ 52
Binomial Distribution .................... 53
Geometric Distribution .................... 54
Negative Binomial Distribution ............... 54
Poisson Distribution ...................... 54
Hypergeometric Distribution ................. 55
2.2 Continuous Distributions .................. 56
Continuous Uniform Distribution .............. 56
Normal Distribution ..................... 56
Gamma Distribution ..................... 58
Lognormal Distribution ................... 60
Logistic Distribution ..................... 62
Weibull Distribution ..................... 62
Pareto Distribution ...................... 63
Student’s $t$-Distribution .................. 64
Snedecor’s $F$-Distribution ................. 66
Beta Distribution ...................... 67
2.3 Some Limit Theorems .................... 68

3 Concepts of Statistical Inference 71
3.1 Introduction .......................... 71
3.2 Sufficiency and Completeness ................. 74
3.3 Methods of Estimation .................. 82
Method of Moments Estimators (MMEs) ........ 82
Maximum Likelihood Estimators (MLEs) ........ 82
3.4 Methods of Evaluating Estimators ............ 85
3.5 Elements of Hypothesis Testing .......................... 92
   Likelihood Ratio Tests ..................................... 95
   Large Sample Tests ..................................... 100
   The Notion of P-Value ..................................... 102
3.6 Set (or Interval) Estimation ................................. 104
   Methods of Finding Set (or Interval) Estimators ...... 105

4 Elements of Decision Theory .............................. 107
   4.1 Introduction ............................................. 107
   4.2 Optimality Criteria ..................................... 109
   4.3 Loss Functions .......................................... 112
   4.4 Admissible, Minimax, and Bayes Rules .............. 115

5 Computational Aspects ..................................... 127
   5.1 Preliminaries ............................................. 127
   5.2 Numerical Integration ................................... 132
   5.3 Monte-Carlo Simulation ................................. 135
   5.4 Bootstrap Method of Resampling ..................... 140
   5.5 Testing and Interval Estimation Based on Computations . 145

Part-II
Exponential and Other Positively Skewed Distributions with Applications 149

6 Exponential Distribution .................................. 151
   6.1 Preliminaries ............................................. 151
   6.2 Characterization of Exponential Distribution ....... 156
   6.3 Estimation of Parameter(s) ............................ 161
   6.4 Goodness of Fit Tests for Exponential Distribution .... 167

7 Gamma Distribution ....................................... 171
   7.1 Preliminaries ............................................. 171
   7.2 Characterization of Gamma Distribution ............ 179
   7.3 Estimation of Parameters ................................ 181
List of Figures

1.4.1 Venn Diagram for the Events \((A \cap B)\) and \((A \cap B^c)\)........ 23
1.5.1 Bar Graph of the pmf (1.5.1).............................. 28
1.5.2 Probability Histogram of the pmf (1.5.1).............. 29
1.5.3 Graph of the discrete \(cdf\) (1.5.3)..................... 30
1.6.1 The Event \((X_1 > X_2)\) is the shaded area.......... 38

2.2.1 The pdfs of \(N(\mu, 1)\) Distribution, \(\mu = 0, 5\)......... 57
2.2.2 The pdfs of \(G(\alpha, 1)\) Distribution, \(\alpha = 1, 5\)....... 60
2.2.3 The pdfs of \(LN (\mu, 1), \mu = 0, 3\). .................. 61
2.2.4 The pdfs of \(L(\mu, 1), \mu = 0, 5\). .................... 63
2.2.5 The pdfs of \(W(\alpha, 1), \alpha = 1, 5\). ................. 64
2.2.6 The pdfs of \(P(\alpha, 1), \alpha = 1, 5\). ................ 65
2.2.7 The pdfs of \(t_\nu(\theta)\) with \(\theta = 0, \nu = 1, 5, 30\).... 66

3.5.1 The Power Function \(\beta(\theta)\)......................... 97

4.2.1 Risk Curves of \(\delta_1, \delta_2, \delta_3, \text{ and } \delta_4\). ....... 110
4.2.2 Risk Curves of \(\delta_1, \delta_2, \text{ and } \delta_3\). ........ 111
4.4.1 Risk Curves of \(\delta_0 \text{ and } \delta^*_\pi\).................. 124

5.2.1 Plots of \(B_5(\hat{\sigma}_0^2| 1, \lambda)\) and \(B_{10}(\hat{\sigma}_0^2| 1, \lambda)\)........ 134
5.4.1 A Histogram of the Glucose Data Set.................. 144

6.1.1 The Exponential pdfs................................. 152
6.1.2 The relative frequency histogram and a matching exponential pdf of female ramet size with \(\beta = 3.0\)........... 156
6.1.3 The relative frequency histogram of male ramet size and a matching exponential pdf with \(\beta = 2.5\)........... 157
6.1.4 The pdfs of \(\text{Exp}(\beta, \mu)\) with \((\beta, \mu) = (1, 1), (2, 3)\).... 158
6.3.1 The graph of $\hat{\xi}_c$ against $c \in (0, 1)$.......................... 162
6.3.2 Plots of $\text{Var}(\hat{\xi}_c (ML))$ for $n = 8$: (a) against $\beta$ (with $c = 0.90$); (b) against $c$ (with $\beta = 1$). .................. 163
6.3.3 MSE curves plotted against $\beta$ for $n = 8$: (a) for $\beta$ estimation; (b) for $\mu$ estimation. .......................... 166

7.1.1 Two-parameter Gamma pdfs. .......................... 173
7.1.2 Histogram of the vessel lifetime data and a matching gamma pdf with $\alpha = 1.45$ and $\beta = 300$. .......................... 175
7.1.3 Histogram of the A/C equipment operating hours and a matching pdf with $\alpha = 2.0$ and $\beta = 49.0$. .......................... 177
7.1.4 Histogram of Transit Time of Africanized Honey Bee and a matching pdf with $\alpha = 1.15$ and $\beta = 4.0$. .......................... 178

7.3.1 Graphs of bias of various estimators plotted against $\alpha$ when $n = 10$ and data come from $G(\alpha, 1, 0)$ distribution: (a) for $\alpha$ estimation, (b) for $\beta$ estimation, (c) for $\mu$ estimation. .......................... 187
7.3.2 Graphs of MSE of various estimators plotted against $\alpha$ when $n = 10$ and data come from $G(\alpha, 1, 0)$ distribution: (a) for $\alpha$ estimation, (b) for $\beta$ estimation, (c) for $\mu$ estimation. .......................... 189
7.3.3 Graphs of bias of various estimators plotted against $\alpha$ when $n = 20$ and data come from $G(\alpha, 1, 0)$ distribution: (a) for $\alpha$ estimation, (b) for $\beta$ estimation, (c) for $\mu$ estimation. .......................... 191
7.3.4 Graphs of MSE of various estimators plotted against $\alpha$ when $n = 20$ and data come from $G(\alpha, 1, 0)$ distribution: (a) for $\alpha$ estimation, (b) for $\beta$ estimation, (c) for $\mu$ estimation. .......................... 193
7.4.1 Plots of $\hat{\alpha}_L$ and $\hat{\alpha}_U$ against $\alpha_0$ for the data set in Example 7.1.2. .......................... 198

8.1.1 Two-parameter Weibull pdf’s. .......................... 200
8.1.2 Histogram of the Test of Endurance Data and a matching $W(\alpha, \beta, \mu)$ pdf with $\alpha = 1.7, \beta = 70.0, \mu = 16.0$. .......................... 203
8.1.3 Histogram of Springs’ Failure Data and a matching $W(\alpha, \beta, \mu)$ pdf with $\alpha = 2.4, \beta = 78.0, \mu = 99.0$. .......................... 204
8.4.1 Plots of $\hat{\alpha}_L$ and $\hat{\alpha}_U$ against $\alpha_0$ for the data set in Example 8.1.1 ........................................ 223

9.1.1 Type-I EV distribution pdfs for various $\beta$ and $\mu$ ........ 226
9.1.2 Type-II EV distribution pdfs for various $\alpha$, $\beta$, and $\mu$ .... 227
9.1.3 Type-III EV distribution pdfs for various $\alpha$, $\beta$, and $\mu$ . . 228
9.1.4 (a) Histogram for altitude 40 km and a matching EV-I pdf. 232
9.1.4 (b) Histogram for altitude 48 km and a matching EV-III pdf. .................. 233
9.1.5 Relative frequency histogram of gust-velocity data and a matching EV-I pdf. .............................. 235

10.1.1 Two possible ways of handling a system reliability. ........ 247
10.2.1 Bias of $\hat{F}_{ML}$ as a function of $t_*/\beta$ with $n = 8$ .......... 251
10.2.2 MSEs of $\hat{F}_{ML}(t)$ and $\hat{F}_{U}(t)$ as a function of $t_*/\beta$ with $n = 8$ . ................................. 252
10.2.3 Bias of $\hat{F}_{ML}(t)$ as a function of $t_*/(t - \mu)/\beta$ $(n = 8)$ . 255
10.2.4 MSEs of $\hat{F}_{ML}(t)$ and $\hat{F}_{U}(t)$ as functions of $t_*/(t - \mu)/\beta$ $(n = 8)$. ........................................ 256
10.2.5 The $\hat{F}_{ML}$ curve for the data set in Example 10.1.2 .......... 258
10.2.6 (a) Simulated bias of $\hat{F}_{ML}$ plotted against $\alpha$ $(\beta = 1, t = 1)$ 259
10.2.6 (b) Simulated MSE of $\hat{F}_{ML}$ plotted against $\alpha$ $(\beta = 1, t = 1)$ 260
10.2.7 (a) Simulated bias of $\hat{F}_{ML}$ plotted against $t > 0$ ($\alpha = 5, \beta = 1$) ............................. 260
10.2.7 (b) Simulated MSE of $\hat{F}_{ML}$ plotted against $t > 0$ ($\alpha = 5, \beta = 1$). ........................................ 261
10.2.8 (a) SB-Bias of $\hat{F}_{ML}$ (Example 10.2.2) .................. 262
10.2.8 (b) SB-MSE of $\hat{F}_{ML}$ (Example 10.2.2) ....................... 263
10.2.9 (a) PB-Bias of $\hat{F}_{ML}$ (Example 10.2.2) ..................... 265
10.2.9 (b) PB-MSE of $\hat{F}_{ML}$ (Example 10.2.2) ..................... 265
10.2.10 (a) Simulated Biases of $\hat{F}_{MML}$ and $\hat{F}_{MM}$ plotted against $\alpha$ $(\beta = 1, t = 1)$ ........................................ 266
10.2.10 (b) Simulated $MSE$s of $\hat{F}_{MML}$ and $\hat{F}_{MM}$ plotted against $\alpha$ $(\beta = 1, t = 1)$ .............................. 267
10.2.11 (a) Simulated Biases of $\hat{F}_{MML}$ and $\hat{F}_{MM}$ plotted against $t > 0$ $(\alpha = 5.0, \beta = 1.0)$ .............................. 267
10.2.11 (b) Simulated $MSE$s of $\hat{F}_{MML}$ and $\hat{F}_{MM}$ plotted against $t > 0$ $(\alpha = 5.0, \beta = 1)$ ......... 268

© 2006 by Taylor & Francis Group, LLC


10.2.12 (a) SB-Biases of F̄
MML and F̄ MM (Example 10.2.2). . . .


10.2.12 (b) SB-M SEs of F̄
. .
MML and F̄ MM (Example 10.2.2).



10.2.13 The F̄ ML , F̄ MML , and F̄ MM curves for the data set in
Example 10.2.2 with the G(α, β, μ) model. . . . . . . . . .
 , F̄


10.2.14 (a) Simulated Biases of F̄
ML
MML , and F̄ MM plotted
against t > 0 ( α = 5.0, β = 1.0, μ = 0.0) for n = 10 with
a G(α, β, μ) model. . . . . . . . . . . . . . . . . . . . . . .
 , F̄


10.2.14(b) Simulated M SEs of F̄
ML
MML , and F̄ MM plotted
against t > 0 ( α = 5.0, β = 1.0, μ = 0.0) for n = 10 with
a G(α, β, μ) model. . . . . . . . . . . . . . . . . . . . . . .
 , F̄


10.2.15 (a) Simulated Biases of F̄
ML
MML , and F̄ MM plotted
against t > 0 ( α = 5.0, β = 1.0, μ = 0.0) for n = 20 with
a G(α, β, μ) model. . . . . . . . . . . . . . . . . . . . . . .


 , F̄
10.2.15(b) Simulated M SEs of F̄
ML
MML , and F̄ MM plotted
against t > 0 ( α = 5.0, β = 1.0, μ = 0.0) for n = 20 with
a G(α, β, μ) model. . . . . . . . . . . . . . . . . . . . . . .
 , F̄


10.2.16 (a) SB-Biases of F̄
ML
MML , and F̄ MM using a G(α, β, μ)
model for the data in Example 10.2.2. . . . . . . . . . . .
 , F̄


10.2.16 (b) SB-M SEs of F̄
ML
MML , and F̄ MM using a G(α, β, μ)
model for the data in Example 10.2.2. . . . . . . . . . . .

10.2.17 The F̄
. .
ML curve for the data set in Example 10.2.3.

10.2.18 The F̄ MM curve for the data set in Example 10.2.3. . .
 and F̄

10.2.19 (a) Simulated Biases of F̄
ML
MM plotted against α
with n = 10 and 20. . . . . . . . . . . . . . . . . . . . . .
 and F̄

10.2.19 (b) Simulated Biases of F̄
ML
MM plotted against t
with n = 10 and 20. . . . . . . . . . . . . . . . . . . . . .
 and F̄

10.2.20 (a) Simulated MSEs of F̄
ML
MM plotted against α
(β = 1, t = 1.0) with n = 10 and 20. . . . . . . . . . . . .


10.2.20 (b) Simulated MSEs of F̄
ML and F̄ MM plotted against
t > 0 (α = 5.0, β = 1.0) with n = 10 and 20. . . . . . . . .


10.2.21 (a) SB-Biases of F̄
. . .
ML and F̄ MM (Example 10.2.3).


10.2.21 (b) SB-MSEs of F̄ ML and F̄ MM (Example 10.2.3). . . .

10.2.22 (a) PB-Bias of F̄
. . . . . . . . .
ML (Example 10.2.3).

10.2.22 (b) PB-MSE of F̄ ML (Example 10.2.3). . . . . . . . . .


10.2.23 F̄
. .
ML and F̄ MM plotted against t (Example 10.2.3).


F̄
10.2.24 (a) Simulated Biases of F̄
and
plotted
against
ML
MM
t > μ (α = 5.0, β = 1.0, μ = 0.0). . . . . . . . . . . . . .
xxvi
© 2006 by Taylor & Francis Group, LLC

269
269
271

271

272

272

273
273
274
275
276
278
279
279
280
281
282
283
283
284
285


10.2.24 (b) Simulated Biases of \( \hat{F}_{ML} \) and \( \hat{F}_{MM} \) plotted against 
\( \alpha \ (t = 1.0, \beta = 1.0, \mu = 0.0) \). ........................................ 285

10.2.25 (a) Simulated MSEs of \( \hat{F}_{ML} \) and \( \hat{F}_{MM} \) plotted against 
\( t > \mu \ (\alpha = 5.0, \beta = 1.0, \mu = 0.0) \). ................................. 286

10.2.25 (b) Simulated MSEs of \( \hat{F}_{ML} \) and \( \hat{F}_{MM} \) plotted against 
\( \alpha \ (t = 100, \beta = 1.0, \mu = 0.0) \). ........................................ 286

10.3.1 A Series System consisting of \( k \) components .................................. 287

10.3.2 Graphs of \( \hat{F}_{s(ML)} \) and its 95\% lower confidence bound
\( \hat{F}_{s} \) plotted against \( t \) (Example 10.3.1). .................................. 293

10.3.3 Plot of \( \hat{F}_{s(ML)} \) against \( t > 0 \) (Example 10.3.2). ................. 296

10.3.4 Plot of \( \hat{F}_{s} \) against \( \alpha \) (Example 10.3.2). .......................... 297

10.4.1 A parallel system consisting of \( k \) components .................................. 298

10.4.2 (a) Bias of \( \hat{F}_{p(ML)} \) for \( k = 2 \) and \( (n_1, n_2) = (5, 10) \). .... 301

10.4.2 (b) Bias of \( \hat{F}_{p(ML)} \) for \( k = 2 \) and \( (n_1, n_2) = (10, 10) \). .... 301

10.4.3 (a) MSE of \( \hat{F}_{p(ML)} \) for \( k = 2 \) and \( (n_1, n_2) = (5, 10) \). .... 302

10.4.3 (b) MSE of \( \hat{F}_{p(ML)} \) for \( k = 2 \) and \( (n_1, n_2) = (10, 10) \). .... 302

10.4.4 Plot of \( \hat{F}_{p(ML)} \) as a function of \( t \). .................................. 304

10.4.5 Plot of the lower confidence bound \( \hat{F}_{p} \) against \( \alpha \). ........ 305
List of Tables

1.4.1 Student Smoking Habit vs. Gender. .......................... 21
1.6.1 Joint Probability Distribution of $(X_1, X_2)$. ............. 36

3.5.1 Type-I and Type-II Errors. ................................. 94

5.2.1 Computed Values of $B_n(\hat{\sigma}^2, 1, \lambda)$. .......... 134
5.3.1 Simulated Bias of $\hat{\sigma}^2$. ........................... 140
5.4.1 Glucose per 100 Milliliters of Blood (in mg). ............ 144
5.4.2 Bootstrap Estimates of Bias and SE. ........................ 145

6.1.1 Ramet Size Distributions. .................................. 155
6.4.1 Computational Results for Calculating $T^*$. ............... 169
6.4.2 Computational Results for Calculating $T^{**}$. ............ 170

7.1.1 Frequency Table for Pressure Vessel Lifetime Data. ....... 174
7.1.2 Frequency Table for Operating Hours of A/C Equipment. 176
7.1.3 Frequency Table for Transit Time of Africanized Honey Bee. 177

8.1.1 Frequency Table of the Test of Endurance Data. ............ 202
8.1.2 Frequency Table for Springs’ Failure Data. ............... 203
8.4.1 Simulated Critical Values for $\Delta^*$ with 10,000 Replications. 217
8.4.2 Simulated Critical Values for $\Delta^{**}$ with 10,000 Replications. 218
8.4.3 Critical Values for the Test Statistics $\Delta_1, \Delta_2, \Delta_3$. .... 220

9.1.1 Ozone Level Values (in percentage deviation from midnight = 0 hour) at Two Altitudes. .................. 231
9.1.2 Frequency Tables for the Ozone Level Data. ............. 232
9.1.3 Frequency Distribution of 485 Maximum Gust-Velocity Measurements. ............................ 234
9.1.4 Relative Frequency Distribution of Gust-Velocity Measurements. ........................................... 234
9.1.5 Percentiles of the Standard Type-I Extreme Value Distribution. ............................................. 236
9.3.1 Values of $l, k_{r,n},$ and $d_{r,n}$ for Selected $n$ and $r$ ($0.8 \leq r/n < 1.0$). ................................. 241
9.3.2 Values of $k_n$ for $2 \leq n < \infty$ ................................................................. 242
9.3.3 Estimated Parameters with Various Methods. ................................................................. 243
9.3.4 Estimated Mean, Bias, and MSE of the Parameter Estimates Using Bootstrap. ...................... 244
9.4.1 Values of $\alpha, \alpha_0, \alpha_1,$ and $\alpha_2.$ .................................................. 245
9.4.2 Values of $W_{n,\alpha}$ for Selected Values of $n$ and $\alpha.$ ...................................... 245
9.4.3 The Coefficients $W_i$'s for $n = 24.$ ............................................. 246
10.3.1 A Comparison of First Order Bias and MSE of $\hat{F}_{s(ML)}$ with the Exact Ones ($k = 2$). .................................................................................................................. 290
Part-I
General Statistical Theory
Chapter 1

Basic Concepts

1.1 The System of Real Numbers

A real number or a real value is one which can be experienced in real life. In fact, any value that can occur in our real life is indeed a real value. For example, it is possible to experience a temperature of (−22.4) degrees Fahrenheit. Hence the value (−22.4) is a real value. Take the example of a radio active element which emits, say, 623000000000 particles per unit of time. Hence the value 623000000000 is a real value.

To put it more rigorously, real numbers or real values are those which satisfy the following axioms (A.1 to A.5):

(A.1) : \( x + y = y + x, \quad xy = yx. \)

(A.2) : \( x + (y + z) = (x + y) + z, \quad x(yz) = (xy)z. \)

(A.3) : \( x(y + z) = xy + xz. \)

(A.4) : Given two real values \( x \) and \( y \), there exists a real value \( z \) such that \( x + z = y \). This \( z \) is denoted by \( y - x \). For any real value \( x \), \( x - x \) is denoted by 0 which is independent of \( x \). Also, for any real value \( x \), there exists another real value \( -x \) (called negative \( x \)) which means \( 0 - x \).

(A.5) : Given two real values \( x \) and \( y \) where \( x \) is not 0, there exists a real value \( z \) such that \( xz = y \). This \( z \) is denoted by \( y/x \). For any nonzero real value \( x \), \( x/x \) is denoted by 1 which is independent of \( x \).
Also, for any nonzero real value \( x \), there exists another real value \( x^{-1} \) (called \( x \)-inverse) which means \( 1/x \).

From the above axioms all the usual algebraic rules can be derived. For instance, \((-(-x)) = x\), \((x^{-1})^{-1} = x\), \((-(-x)) = x\), etc.

The collection of all real numbers (or values) is called a real set and is denoted by \( \mathbb{R} \). The following concepts and notations are widely used:

1. The whole numbers \( \ldots, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, \ldots \) are called integers. \( \mathbb{Z} \) denotes the collection of all integers; and \( \mathbb{Z}^+ (\mathbb{Z}^-) \) denotes the subcollection of all positive (negative) integers.

2. A rational number is a real number which can be expressed as a quotient of two integers. \( \mathbb{Q} \) denotes the collection of all rational numbers. A real number, which is not a rational number, is called an irrational number. The two popular irrational numbers are \( \pi \) and \( e \) (discussed later).

### 1.2 Some Useful Algebraic Results

If we have \( n \) real values, say, \( x_1, x_2, \ldots, x_n \), their sum is denoted by \( \sum_{i=1}^{n} x_i \) or by \( \sum_{1}^{n} x_i \) or by \( \sum_{i} x_i \) when the range of values of \( i \) is evident from the context. \( \sum_{i=1}^{n} x_i \) is read as “the summation of \( x_i \) where \( i \) goes from 1 to \( n \).” The following simple results can be easily verified:

\[
\begin{align*}
\text{(i)} \quad \sum_{i} (x_i \pm y_i) &= \sum_{i} x_i \pm \sum_{i} y_i \\
\text{(ii)} \quad \sum_{i} cx_i &= c \sum_{i} x_i \\
\text{(iii)} \quad \sum_{i} (c + x_i) &= nc + \sum_{i} x_i
\end{align*}
\]

where \( c \) is a real value.

Suppose we have \( m \times n \) real values arranged in \( m \) rows and \( n \) columns as follows:

\[
\begin{array}{cccc}
x_{11} & x_{12} & \cdots & x_{1n} \\
x_{21} & x_{22} & \cdots & x_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
x_{m1} & x_{m2} & \cdots & x_{mn}
\end{array}
\]

where \( x_{ij} \) is the real value occurring in both the \( i \)th row and the \( j \)th column. The sum of all the \( m \times n \) real values can be obtained by first
adding the values in each row and then adding each row totals:

\[ (x_{11} + x_{12} + \ldots + x_{1n}) + (x_{21} + x_{22} + \ldots + x_{2n}) + \ldots + (x_{m1} + x_{m2} + \ldots + x_{mn}) \]

\[ = \sum_{j} x_{1j} + \sum_{j} x_{2j} + \ldots + \sum_{j} x_{mj} \]

\[ = \sum_{i} \sum_{j} x_{ij} \quad (1.2.5) \]

This grand total can also be obtained by first adding the values in each column and then adding the column totals:

\[ (x_{11} + x_{21} + \ldots + x_{m1}) + (x_{12} + x_{22} + \ldots + x_{m2}) + \ldots + (x_{1n} + x_{2n} + \ldots + x_{mn}) \]

\[ = \sum_{i} x_{i1} + \sum_{i} x_{i2} + \ldots + \sum_{i} x_{in} \]

\[ = \sum_{j} \sum_{i} x_{ij} \quad (1.2.6) \]

The product of \( n \) real values \( a_1, a_2, \ldots, a_n \) is denoted by \( \prod_{i=1}^{n} a_i \) or by \( \prod a_i \) or simply by \( \prod a_i \) when the range of \( i \) is clear from the context. Like \( \sum \), \( \prod \) also obeys some simple rules as given below:

(i) \[ \prod_{i} (a_i b_i) = \left( \prod_{i} a_i \right) \left( \prod_{i} b_i \right) \quad (1.2.7) \]

(ii) \[ \prod_{i} (a_i/b_i) = \left( \prod_{i} a_i \right) / \left( \prod_{i} b_i \right) \quad \text{(for } b_i \neq 0) \quad (1.2.8) \]

(iii) \[ \prod_{i} (ca_i) = c^n \left( \prod_{i} a_i \right) \quad (1.2.9) \]

where \( c \) is a real value.

A **sequence** is a collection of quantities \( c_1, c_2, \ldots \), called elements, which can be arranged in an order, so that when an integer \( k \) is given, the \( k \)th element of the sequence is completely specified. A sequence is denoted by \( \{c_k\} \) with elements \( c_1, c_2, \ldots \), etc.

A sequence \( \{c_k\} \) is said to be bounded if there exists an arbitrary positive number \( M \) such that \( |c_k| \leq M \) for all \( k \) (where \( |c| \), called ‘modulus’ \( c \), denotes the **absolute value** of \( c \), i.e., \( |c| = c \) if \( c \geq 0 ; = -c \) if \( c < 0 \)).

For a sequence \( \{c_k\} \) if there exists a value \( c_* \) such that for every choice of \( \varepsilon > 0 \), there exists an integer \( N \) such that \( |c_k - c_*| < \varepsilon \) for all \( k \geq N \),
then the sequence \( \{c_k\} \) is said to be **convergent** and has the limit \( c_\ast \). It can be written as \( \lim_{k \to \infty} c_k = c_\ast \). A sequence which is not convergent is called **divergent**.

A **series** is an expression of the form \( c_1 + c_2 + \ldots + c_k + \ldots \), which may have a finite or an infinite number of terms. Its partial sums, \( s_1 = c_1, s_2 = c_1 + c_2, \ldots, s_k = \sum_{i=1}^{k} c_i \), constitute a sequence \( \{s_k\} \). A series is said to be convergent if the sequence of these partial sums is convergent. If \( \lim_{k \to \infty} s_k = s_\ast \) for some value \( s_\ast \), then \( s_\ast \) is the sum of the convergent series and \( s_\ast = \sum_{k=1}^{\infty} c_k \). If \( \lim_{k \to \infty} s_k = \pm \infty \), then the series is said to oscillate if the sequence of the partial sums does not tend to a definite limit, but oscillates between two values. A simple example of oscillating series is

\[
\sum_{i=1}^{\infty} (-1)^i
\]

is denoted by the letter ‘e’. Also, for any positive value \( a \),

\[
a^x = 1 + \frac{x \ln a}{1!} + \frac{x^2(\ln a)^2}{2!} + \ldots + \frac{x^k(\ln a)^k}{k!} + \ldots
\]  

where \( (\frac{k}{r}) = \frac{k(k-1)\ldots(k-r+1)}{r!} \) and \( r! = (r-1)\ldots2\cdot1 \). The right-hand side (RHS) expansion of the above expression \( (a+b)^k \) is called a ‘**Binomial series**’ and holds for any positive integer \( k \).

Two important results are

\[
(1 - a)^{-1} = 1 - a + a^2 - a^3 + a^4 - \ldots + (-a)^r + \ldots
\]  

and

\[
(1 + a)^{-1} = 1 + a + a^2 + a^3 + a^4 + \ldots + a^r + \ldots
\]  

which are valid for \(|a| < 1\).

The logarithmic series which gives the expansion of \( \ln(1 + a) \) is

\[
\ln(1 + a) = a - \frac{a^2}{2!} + \frac{a^3}{3!} - \frac{a^4}{4!} + \ldots
\]
Also,

$$\ln(1 - a) = -a - \frac{a^2}{2!} - \frac{a^3}{3!} - \frac{a^4}{4!} - \ldots$$

Both the above logarithmic expansions hold for \(|a| < 1\).

### Some Algebraic Inequalities

1. If \(a\) and \(b\) are any two real values, then \(|a + b| \leq |a| + |b|\).

2. Let \(a_1, a_2, \ldots, a_n\) be a set of positive real values. The arithmetic mean (AM), geometric mean (GM), and harmonic mean (HM) of \(a_1, a_2, \ldots, a_n\) are defined respectively as follows:

$$\text{AM} = \sum_{i} a_i / n, \quad \text{GM} = \left( \prod_{i} a_i \right)^{1/n}, \quad \text{HM} = n / \sum_{i} \left(1/a_i\right)$$

(1.2.18)

It can be shown that \(\text{HM} \leq \text{GM} \leq \text{AM}\).

3. If \(a_1, a_2, \ldots, a_n\) and \(b_1, b_2, \ldots, b_n\) are two sets of real values, then

$$\left( \sum_{i} a_i b_i \right)^2 \leq \left( \sum_{i} a_i^2 \right) \left( \sum_{i} b_i^2 \right)$$

or

$$\left| \sum_{i} a_i b_i \right| \leq \sqrt{\left( \sum_{i} a_i^2 \right) \left( \sum_{i} b_i^2 \right)}$$

(1.2.19)

This is known as Cauchy-Schwartz inequality. The equality holds when \(a_i = cb_i\) for all \(i\) and a constant \(c\).

### Matrices

A matrix \(A_{m \times n}\) (or simply \(A\)) of order \(m \times n\) is a rectangular array of elements arranged in \(m\) rows and \(n\) columns as

$$A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix} = \begin{bmatrix}
    (a_{ij})
\end{bmatrix}$$

(1.2.20)

where \(a_{ij}\) represents the element in the \(i\)th row and \(j\)th column.\(^1\) Also, a matrix of order \(m \times n\) is a set of \(m \times n\) elements arranged in a convenient

\(^1\)A is called a row vector when \(m = 1\); and a column vector when \(n = 1\).
Chapter 1 Basic Concepts

way. For instance, closing stock prices of \( m \) publicly held companies at the end of \( n \) different quarters may be represented by a matrix \( A = \left[ \begin{array}{c} a_{ij} \end{array} \right] \) of order \( m \times n \) where \( a_{ij} \) is the closing stock price of \( i \)th company at the end of \( j \)th quarter.

Two matrices \( A = \left[ \begin{array}{c} a_{ij} \end{array} \right] \) and \( B = \left[ \begin{array}{c} b_{ij} \end{array} \right] \) are called equal (written as \( A = B \)) if and only if they are of the same order and \( a_{ij} = b_{ij} \) for all \( i \) and \( j \).

A matrix with the same number of rows as of columns is called a square matrix. The elements \( a_{ii} \) of a square matrix are called the main diagonal elements. A square matrix is symmetric if \( a_{ij} = a_{ji} \) for all \( i \) and \( j \).

In the following, some standard matrix operations (properties) are defined (discussed):

1. The product of a matrix \( A = \left[ \begin{array}{c} a_{ij} \end{array} \right] \) and a real value \( c \) is defined as:

\[
cA = c \left[ \begin{array}{c} a_{ij} \end{array} \right] = \left[ \begin{array}{c} ca_{ij} \end{array} \right]
\]

Hence \( cA = Ac \), and \( (-1)A \) is written as \( (-A) \).

2. The sum of \( A_{m \times n} \) and \( B_{k \times l} \) is defined only when \( m = k \) and \( n = l \), and \( A + B = \left[ \begin{array}{c} a_{ij} + b_{ij} \end{array} \right] \). Similarly, \( A - B = \left[ \begin{array}{c} a_{ij} - b_{ij} \end{array} \right] \).

3. The product of \( A_{m \times n} \) and \( B_{n \times l} \) is defined only when \( n = k \), and \( A_{m \times n}B_{n \times l} = C_{m \times l} = \left[ \begin{array}{c} c_{ij} \end{array} \right] \) where \( c_{ij} = \sum_{r=1}^{n} a_{ir}b_{rj} \). Note that \( A_{m \times n}B_{n \times l} \) is undefined unless \( n = k \). Also, for square matrices of same order, \( AB \) need not be same as \( BA \).

4. The transpose \( A' \) of a matrix \( A \) is obtained by interchanging the rows and columns of \( A \). Therefore,

\[
(A')' = A, \quad (A \pm B)' = A' \pm B', \quad (AB)' = B'A'
\]

Thus a square matrix \( A \) is symmetric if \( A = A' \).

5. A diagonal matrix is a square matrix where all off-diagonal elements are zero, and hence it is symmetric too. A square matrix is called an identity matrix if it is diagonal and all diagonal elements are equal to \( 1 \). An identity matrix of order \( n \) (or \( n \times n \)) is denoted by \( I_n \) (or simply \( I \)). It can be verified that

\[
I_mA_{m \times n} = A_{m \times n} = A_{m \times n}I_n
\]
6. A square matrix $A$ is called **idempotent** if $A^2 = A \cdot A = A$. Note that if $A$ is idempotent, then for any positive integer $k$,

$$A^k = A \cdot A \ldots A = A \quad (1.2.24)$$

7. A matrix is called a **null matrix** (denoted by $O_{m \times n}$ or simply $O$) when all elements are equal to zero. It can be seen that for any matrix $A_{m \times n}$,

$$A_{m \times n}O_{n \times l} = O_{m \times l}, \quad O_{k \times m}A_{m \times n} = O_{k \times n}$$

and $A_{m \times n}O_{m \times n} = A_{m \times n} \quad (1.2.25)$

8. A square matrix $Q$ is called an **orthogonal matrix** if $QQ' = I = Q'Q$.

9. For every square matrix $A_{n \times n}$, the determinant of $A$ (denoted by $|A|$ or $\det(A)$) is defined as $|A| = \sum_{\pi} (\pm a_{i_1 i_2 \ldots a_{i_n}})$, where $\pi$ is the collection of $n!$ different permutations $\{i_1, i_2, \ldots, i_n\}$ of $\{1, 2, \ldots, n\}$ and the sign of each term is plus ($+$) if the permutation consists of even number of cycles, whereas it is minus ($-$) if the number of cycles is odd.

The **minor** of an element $a_{ij}$ in $A$ is defined to be the determinant of a matrix of order $(n - 1)$ obtained from $A$ by deleting the $i$th row and the $j$th column. The **cofactor** of $a_{ij}$, denoted by $A_{ij}$ is $A_{ij} = (-1)^{i+j} \left(\text{the minor of } a_{ij}\right)$.

The determinant of a matrix $A_{n \times n}$ can also be obtained as

$$|A| = \sum_{j=1}^{n} a_{ij}A_{ij} \quad \text{for any } i = 1, 2, \ldots, n \quad (1.2.26)$$

In the simplest case, when $A$ is of order $2 \times 2$, $|A| = a_{11}a_{22} - a_{12}a_{21}$.

Also, it can be verified that $|A| = |A'|$ for any square matrix $A_{n \times n}$.

10. A square matrix $A_{n \times n}$ is called **nonsingular** if $|A| \neq 0$. If $A_{n \times n}$ is nonsingular then said to have a rank $n$. $A_{n \times n}$ is called a **singular** matrix if $|A| = 0$. For a nonsingular matrix $A$, there exists another

---

*A cycle of length $k$ means that, given integers $\{i_1, i_2, \ldots, i_k\}$, we have a permutation $\{i_2, i_3, \ldots, i_k, i_1\}$. 

© 2006 by Taylor & Francis Group, LLC
matrix (of order $n \times n$), denoted by $A^{-1}$ such that $AA^{-1} = I_n = A^{-1}A$. Such matrix $A^{-1}$ is called the inverse of $A$, and it can be obtained as

$$A^{-1} = (a_{ij}) \quad \text{where} \quad a_{ij} = \frac{1}{|A|} A_{ji} \quad (1.2.27)$$

It can be shown that:

(a) $|A^{-1}| = 1/|A|$
(b) If $A$ is symmetric, then so is $A^{-1}$
(c) For an orthogonal matrix $Q$, $Q^{-1} = Q'$

11. For a symmetric matrix $A$ of order $n \times n$, and variables $x_1, x_2, \ldots, x_n$, the expression

$$Q_A(x) = x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_i x_j \quad (1.2.28)$$

where $x = (x_1, x_2, \ldots, x_n)'$, is called a quadratic form in $x$. If $Q_A(x) \geq 0$ for all $x \in \mathbb{R}^n$, then $A$ is called positive semidefinite; and positive definite if $Q_A(x) > 0$ for all $x \neq 0$. Likewise negative semidefinite and negative definite. For two symmetric matrices $A$ and $B$ of same order, if $Q_A(x) = Q_B(x)$ for all $x \in \mathbb{R}^n$ (the $n$-dimensional real space), then $A = B$.

**Calculus**

If $x$ and $y$ are two variables such that the value of $y$ depends on the value of $x$, then $y$ is said to be a function of $x$. The variable $x$, whose value $y$ depends on, is called the independent variable; and $y$ is called the dependent variable. To denote the dependence of $y$ on $x$, we usually write $y = f(x)$, and $f(x_0) = y_0$ indicates the value of $y$ when $x = x_0$.

Suppose the variable $y$ is a function of the variable $x$. The collection of all possible values of $x$, for which the function is defined, is called the domain of the function, and the collection of all possible values of $y$ is called the range of the function. A function $y = f(x)$ is called bounded if there exists a real value $M > 0$ such that $|y| < M$ for all $x$ in the domain.
1.2 Some Useful Algebraic Results

If there is a finite value \( y_0 \) to which \( f(x) \) tends as \( x \) tends to the value \( x_0 \), then \( y_0 \) is called the limit of \( f(x) \) as \( x \) approaches to \( x_0 \). This is denoted as:

\[
\lim_{x \to x_0} f(x) = y_0
\]

More rigorously, it means, given any \( \varepsilon > 0 \), there exists a \( \delta > 0 \) such that \( |f(x) - y_0| < \varepsilon \) whenever \( |x - x_0| < \delta \).

For two functions \( f_1(x) \) and \( f_2(x) \) of \( x \), common limit operations are:

1. \( \lim_{x \to x_0} \{f_1(x) + f_2(x)\} = \lim_{x \to x_0} f_1(x) + \lim_{x \to x_0} f_2(x) \)
2. \( \lim_{x \to x_0} \{f_1(x) \cdot f_2(x)\} = \lim_{x \to x_0} f_1(x) \cdot \lim_{x \to x_0} f_2(x) \)
3. \( \lim_{x \to x_0} \{f_1(x)/f_2(x)\} = \left\{ \lim_{x \to x_0} f_1(x) \right\}/\left\{ \lim_{x \to x_0} f_2(x) \right\} \)

provided \( \lim_{x \to x_0} f_2(x) \neq 0 \)

Some common limit results are:

1. \( \lim_{x \to \infty} (1 + 1/x)^x = e \)
2. \( \lim_{x \to 0} (1 + x)^{1/x} = e \)
3. \( \lim_{x \to a} \frac{x^k - a^k}{x - a} = ka^{k-1} \)
4. \( \lim_{x \to 0} \frac{\sin x}{x} = 1 \)

A function \( f(x) \) of \( x \) is said to be continuous at a point \( x = x_0 \) if \( \lim_{x \to x_0} f(x) = f(x_0) \). Otherwise, \( f(x) \) is said to be discontinuous at \( x = x_0 \). If \( f(x) \) is continuous at \( x = x_0 \), then this point \( x_0 \) is called a continuity point of \( f(x) \).

Given a function \( y = f(x) \), the limiting value

\[
\lim_{h \to 0} \frac{(f(x_0 + h) - f(x_0))}{h}
\]

when it exists, is called the derivative of \( f(x) \) with respect to \( x \) at \( x = x_0 \). In this case, we say that \( f(x) \) is differentiable at \( x = x_0 \). If the limit does not exist, then \( f(x) \) is not differentiable at \( x = x_0 \). The derivative of \( f(x) \) at \( x = x_0 \) is denoted as \( f'(x_0) \).
Chapter 1 Basic Concepts

If \( y = f(x_1, x_2, \ldots, x_k) \) is a function of several variables \( x_1, x_2, \ldots, x_k \), then the limit,

\[
\lim_{h \to 0} \left( f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_k) - f(x_1, x_2, \ldots, x_k) \right) / h
\]

if it exists when \( x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k \) are treated as constants, is called the (first) partial derivative of \( f \) with respect to \( x_i \) and is denoted by \( f^{(i)}(x_1, x_2, \ldots, x_k) \) or \( \partial f / \partial x_i \).

Let \( y = f(x) \) be a bounded function on the interval \( [a, b] \). Let the interval \( [a, b] \) be partitioned into \( n \) subintervals with a width \( h = (b-a)/n \), i.e.,

\[
[a, b] = [a, a+h] \cup [a+h, a+2h] \cup \ldots \cup [a+(n-1)h, b]
\]

such that \( h \to 0 \) as \( n \to \infty \). Let \( n \) arbitrary points, \( x_1, x_2, \ldots, x_n \), be chosen from the above subintervals, respectively. If for any selection of points \( \{x_1, x_2, \ldots, x_n\} \), the sum \( h \sum_{i=1}^n f(x_i) \) tends to a fixed limit, then that limit is called the definite integral of \( f(x) \) over the interval \( [a, b] \) and is denoted by \( \int_a^b f(x) \, dx \). Geometrically, the value \( \int_a^b f(x) \, dx \) represents the area bounded by the curve \( y = f(x) \), the \( x \)-axis, and two ordinates at \( a \) and \( b \).

The procedure of finding the integral of a function is called integration, and it is the inverse of differentiation. Therefore, if \( y = f(x) \) is a function with derivative \( f'(x) \), then the integral of \( f'(x) \) with respect to \( x \) between \( a \) and \( b \) is \( f(b) - f(a) \). Hence, for integrating a function \( f(x) \), we look for an expression \( F(x) \) such that \( F'(x) = f(x) \). The function \( F(x) \) is called the indefinite integral of \( f(x) \) and is denoted by \( \int f(x) \, dx \).

Since \( \frac{d}{dx} \left( F(x) + c \right) = F'(x) \) for any constant \( c \), the integral of \( f(x) \) is \( (F(x) + c) \). Hence \( F(x) + c \) is also an indefinite integral and therefore the constant \( c \) must appear in any indefinite integral, i.e., \( \int f(x) \, dx = F(x) + c \).

Order ‘big oh’ and ‘small oh’:

A function \( f(x) \) is of order:

(a) \( O(x) \) (called, ‘big oh’ of \( x \)) if \( (f(x)/x) \) is finite as \( x \to \infty \)

(b) \( o(x) \) (called, ‘small oh’ of \( x \)) if \( (f(x)/x) \to 0 \) as \( x \to \infty \)
1.3 Set Theory

A set is a collection of well-defined distinct objects. Usually, a set is denoted by a capital letter (with or without a suitable subscript), like $A$, $B$, or $C$, etc., or $A_1, A_2, \ldots$, etc.

For example, let $A$ be the collection of all positive integers less than 6, i.e., $A$ contains the integers 1, 2, 3, 4, and 5. We can write this set $A$ as $A = \{1, 2, 3, 4, 5\}$. Each member of the set $A$ is called a set element. Note that this collection is well defined and all set elements are distinct.

Another example could be the collection of stocks listed on NYSE (New York Stock Exchange) at the end of December 31, 2000. We can denote this collection as $B = \{IBM, IP, XRX, NU, T, \ldots\}$. Again, this collection is well defined and all elements are distinct.

If we consider the collection of students enrolled at a particular college, then it is not well defined because the student body changes from semester to semester, and hence this collection, in true sense, cannot be a set. However, if we specify the collection of students enrolled in Spring 2005 at a particular college, then it makes sense; we can identify all qualifying students in the collection, and this collection is a set.

A few more of examples of a set are listed below:

(a) The collection of U.S. cities with reported homicides not less than 1 per thousand people in a particular year

(b) The collection of ‘Dot-Com’ start-ups world wide in 2004

(c) The collection of rational numbers between 0 and 1

A set is called an empty set or a null set if it does not have any element in it. Such a set is denoted by $\phi$. For example, let $A$ denote the set of integers between $1/3$ and $2/3$. It can be seen that $A = \phi$ because there is no integer between $1/3$ and $2/3$.

Given two sets, say $A$ and $B$, $B$ is said to be a subset of $A$ if each element of $B$ is also an element of $A$ (i.e., $B$ is a sub-collection of $A$). This is denoted as ‘$B \subseteq A$’. For instances, (i) let $A = \{1, 2, 3, 4, 5\}$ and $B = \{2, 4\}$, here $B \subseteq A$. (ii) Let $A_1$ be all dollar bills issued by the U.S. Federal Reserve in 2000, and let $A_2$ be all dollar bills issued by the U.S. Federal Reserve in 2000 with even serial numbers. Here, $A_2 \subseteq A_1$.

For any two sets $A$ and $B$: 

© 2006 by Taylor & Francis Group, LLC
Chapter 1 Basic Concepts

- \( A \cup B \) (called ‘A union B’) is the collection of all elements which are members of A or B or both

- \( A \cap B \) (called ‘A intersection B’) is the collection of those elements which are common to both A and B

Example 1.3.1: (i) Let \( A = \{1, 2, 3, 4, 5\} \) and \( B = \{4, 5, 6, 7\} \), then \( A \cup B = \{1, 2, 3, 4, 5, 6, 7\} \) and \( A \cap B = \{4, 5\} \). (ii) Let \( A_1 = \{1, 2, 3, 4\} \) and \( A_2 = \{2, 3\} \), then \( A_1 \cup A_2 = \{1, 2, 3, 4\} \) and \( A_1 \cap A_2 = \{2, 3\} \).

For any set \( A \), \( A^c \) (called ‘A-complement ’) is the collection of all elements which are not members of A. The following set theoretic rules are easy to follow:

(a) For any set \( A \), \( A \cup \emptyset = A \) and \( A \cap \emptyset = \emptyset \)

(b) For any two sets \( A \) and \( B \),

\[
(A \cup B)^c = A^c \cap B^c \quad \text{and} \quad (A \cap B)^c = A^c \cup B^c
\]

(c) For any three sets \( A \), \( B \), and \( C \),

\[
A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{and} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)
\]

(d) For any set \( A \), \( (A^c)^c = A \)

1.4 Introduction to Probability Theory

In statistics, we are concerned with collection, presentation, and interpretation of ‘chance outcomes’ which occur in a scientific study. Chance outcomes are those which cannot be predicted with certainty. For example, (i) we may record the number of burglaries that take place per month in a neighborhood to justify more resources for the law enforcement agency; or (ii) we may be interested in investigating voters’ mood for an upcoming election where two candidates, say Candidate-D and Candidate-R, are in the fray. Note that the number of burglaries in a particular month can be any nonnegative value and cannot be predicted beforehand. Similarly, whether a particular voter votes for D or R is unknown unless or until he/she casts his/her ballot. But once the particular month is elapsed,
the number of burglaries are known; and once the election is over, a voter’s choice is known. [Here we are dealing with quantitative data representing counts or measurements, or perhaps with qualitative data which can be categorized according to some criterion.]

From now on we shall refer to any recording of information, quantitative or qualitative, as an observation. Thus the quantitative information of 3, 5, 7, 1, representing the number of burglaries per month in a neighborhood for four consecutive months, constitutes a set of observations or a quantitative data set. In a similar way, the qualitative information of R, R, D, R, D, R, D, D, R, representing the preferences of ten randomly selected registered voters in a town, constitutes a set of observations or a qualitative data set.

In statistics, we often use the word experiment to describe a process or a mechanism that generates a set of observations or a data set. Some simple examples of statistical experiments are: (i) Tossing a coin only once. In this experiment, there are only two possible outcomes, Head (H) or Tail (T). (ii) Rolling a die only once and recording its face value. Here the possible outcomes are 1, 2, 3, 4, 5, or 6. (iii) Selecting a student at random from the local community college and recording his or her ethnic background. The possible outcomes of this experiment are — African American, Caucasian, Asian, Native Indian, etc. Note that the outcome of an experiment can not be predicted with certainty and depends on the chance or probability of each outcome.

Before we discuss ‘probability’ or ‘chance’ of outcomes, let us introduce some commonly used notations and terminology.

**Definition 1.4.1**

(a) A statistical experiment is a mechanism whose outcome is uncertain.

(b) The set of all possible outcomes of a statistical experiment is called the sample space, and it is denoted by \( S \).

Each outcome in a sample space \( S \) is called an element or a member of the sample space, or simply a sample point. A sample space may have a finite number of sample points or can have infinitely many sample points. For example, if a geologist measures the magnitude of an earthquake at the foothills of the Himalayan Mountain range, then he/she might record any possible magnitude between 0 and 10 on the Richter scale. Hence, the sample space of the geologist’s experiment...
contains infinitely many sample points. On the other hand, if an experiment consists of drawing a card at random from a standard deck of 52 cards, then the sample space is finite and has only 52 possible sample points.

If the sample space has a finite number of sample points, then we can list all elements separated by commas and enclosed in brackets. Thus the sample space $S$ of possible outcomes when a die is rolled, is given as

$$S = \{1, 2, 3, 4, 5, 6\}$$

Suppose a coin is tossed three times consecutively. The sample space $S$ can be written as

$$S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$$

where H and T denote ‘head’ and ‘tail’ respectively.

It is convenient to describe a sample space with a large or infinite number of sample points by a statement or rule. For instance, if an experiment consists of drawing a card at random from a standard deck of 52 playing cards, then the sample space can be written as

$$S = \{\text{all 52 playing cards in a standard deck}\}$$

Similarly, consider the experiment where a geologist records the magnitude of an earthquake, the sample space may be described as

$$S = \{x \mid 0 < x < 10\}$$

where the vertical bar ‘$\mid$’ means ‘such that’; or in short, $S = (0, 10)$ (the open interval from 0 to 10).

In any given experiment, we may be interested in the occurrence of certain types of outcomes rather than a single outcome. For instance, when a die is rolled, we might be interested in those outcomes which are divisible by 3. This will happen if the outcome is an element of the subset, say, $A = \{3, 6\}$, of the sample space $S = \{1, 2, 3, 4, 5, 6\}$. Such a subset is called an event. 3

**Definition 1.4.2** An event is a subset of the sample space.

---

3Events are usually denoted by capital letters (with or without subscripts).
1.4 Introduction to Probability Theory

Let us consider an experiment of observing the lifespan of a 60-watt (120 volts) electric bulb. The sample space \( \mathcal{S} = \{ t \mid t \geq 0 \} \), where \( t \) is the life-time in hours. The manufacturer claims that the average life of such bulbs is 750 hours. Hence, we are interested in the event that a bulb fails before completing 750 hours. Therefore, the event of interest can be described by the subset \( B = \{ t \mid 0 \leq t < 750 \} \).

It is possible that an event can be the entire sample space that includes all sample points of \( \mathcal{S} \). The empty set \( \phi \) is called the empty event and contains no sample points at all.

Consider an experiment in which the drinking habit of airline pilots are being recorded. The sample space of possible outcomes might be classified as a collection of nondrinkers, light drinkers, moderate drinkers, and heavy drinkers. Let all the nondrinkers correspond to a subset of \( \mathcal{S} \) which is called the complement of the subset of all drinkers.

**Definition 1.4.3** The complement of an event \( A \), with respect to the sample space \( \mathcal{S} \), is the event of all elements of \( \mathcal{S} \) that are not in \( A \). The complement of \( A \) is denoted by \( A^c \).

Basically \( A^c \) indicates ‘nonoccurrence’ of \( A \). For example, let \( R \) be the event in which a red card is drawn from a standard deck of 52 playing cards. Then \( R^c \) is the event that a black card is drawn from the deck.

Since events are subsets of the sample space, we can invoke the set theoretic operations discussed earlier (in Section 1.3) for events also.

**Meaning of Probability**

The word **probability** may be used in two different contexts. First, it may be used regarding some proposition. Consider the following statement: “It is very probable that the digital divide between the Hispanics and non-Hispanics in the United States will decrease in the next twenty years.” Probability here means the degree of belief in the above proposition of the individual making the statement. This is usually called **Subjective Probability**.

Alternatively, the word **probability** may be used regarding an event given rise by an experiment that can conceivably be repeated infinitely many times under similar conditions. The probability of an event here
refers to the proportion of cases in which the event occurs in such repetitions of the experiment. This type of probability is called **Objective Probability**.

Objective probability, which reflects the real world, is the sense that we should be concerned with most of the time in this handbook. Subjective probability will be discussed later when we deal with Bayesian analysis.

**Classical Definition of Probability**

To develop the concepts of probability, we first consider only those experiments for which the sample space $S$ has a finite number of sample points. This is done for the sake of simplicity and will not be applicable to the situations where $S$ consists of infinitely many elements (either countable or noncountable).

Secondly, it is assumed that the experiment is such that all the possible outcomes in $S$ are equally likely, i.e., when all relevant evidence is taken into account, no single outcome can be expected to occur in preference to the others.

The probability of any event $A$, denoted by $P(A)$, is defined as

$$P(A) = \frac{\# \text{ elements in } A}{\# \text{ elements in } S} \quad (1.4.1)$$

The above ratio is called the **classical definition** of probability. Thus probability can be treated as a function $P$ defined on the class of all events which assigns a unique value between 0 and 1 to an event $A$. Note that:

(a) $0 \leq P(A) \leq 1$

(b) If $A = \phi$, then $P(A) = 0$

(c) If $A = S$, then $P(A) = 1$

(d) If $A \subseteq B \subseteq S$, then $P(A) \leq P(B)$

(e) $P(A) = 1 - P(A^c)$

(f) For any two events $A$ and $B$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

---

\footnote{This is usually called the additive rule.}
Example 1.4.1: An unbiased coin is tossed twice. What is the probability that at least one head occurs?

The sample space for this experiment is $\mathcal{S} = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$. Since this is a fair coin, each of these outcomes would be equally likely to occur. If $A$ represents the event that at least one head is observed, then $A = \{\text{HH}, \text{HT}, \text{TH}\}$, and $P(A) = \frac{3}{4}$.

Limitations of Classical Definition

The classical definition of probability has some obvious drawbacks. Surely, direct application of this definition is not possible if the total number of elements in the sample space is infinite. Even if the number of elements in $\mathcal{S}$ is finite, they must have equally likely outcomes in order to apply the classical definition of probability.

In Example 1.4.1, it was crucial to know that the coin was balanced or unbiased. Because of the unbiasedness of the coin, all possible outcomes were equally likely and hence we could compute the probability of the event $A$ of at least one head occurring. But what about an experiment where outcomes are not equally likely? Can we still find the probability of an event? Well, we can, provided we know what kind of bias is involved in the experiment, and this is explained through the following example.

Example 1.4.2: A die is loaded in such a way that an even number is three times as likely to occur as an odd number. If the die is rolled only once, then find the probability of the event that the face value is less than or equal to 3.

The sample space of the above example is $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. We assign a probability, say $p$, to each odd number and a probability of $3p$ to each even number. The total probability (of the whole sample space) must add up to 1, i.e.,

$$p + p + p + 3p + 3p + 3p = 12p = 1$$

or $p = \frac{1}{12}$. Let $B = \{1, 2, 3\}$ be the event of interest, then

$$P(B) = \frac{1}{12} + \frac{3}{12} + \frac{1}{12} = \frac{5}{12}$$
Probability as Long Term Relative Frequency

In Example 1.4.2, we knew the type of bias involved in the experiment. We could use the known bias to assign a probability to every sample point in the sample space. If we have a reason to believe that a certain sample point is more likely to occur when the experiment is conducted, the probability assigned should be close to 1. On the other hand, a probability closer to 0 is assigned to a sample point that is less likely to occur. Still, if we do not know the nature of the experiment (whether or not the outcomes are equally likely), then it is impossible to assign probabilities to the elements of $\mathcal{S}$. Therefore, we conceptualize the notion of probability as a relative frequency in the long run. By doing this, we avoid the difficulties of applying the classical definition of probability when the sample space is not finite.

Suppose we perform a sequence of $n$ repetitions (or replications) of an experiment. Let $f_n$ be the number of times an event $A$ is occurring among these $n$ repetitions. The value $f_n$ is called the absolute frequency (or simply, frequency) of $A$ in $n$ repetitions, and the ratio, $f_n/n$, is called the relative frequency of $A$.

In many cases, it is observed that the successive values of the sequence $\{f_n/n\}$ differ from one another by small amounts when $n$ gets large. In such cases, there is a tendency in $\{f_n/n\}$ to stabilize around a fixed value. This limiting value of $\{f_n/n\}$ as $n \to \infty$ is regarded as the probability of $A$ in the experiment.

As we adopt the above notion of probability, it must be kept in mind that, practically speaking, it is impossible to determine the exact probability of an event (without any further information or assumption as we had in the last two examples). We cannot even prove the existence of a limit to the relative frequency in any given case. On the other hand, we can treat any observed relative frequency as an estimate of the supposed limit. In this sense, we can speak of the probability that it would rain in Washington, D.C., on 4th of July, or of the stock market acting positively in an election year. If we observe the weather on 4th of July at Washington, D.C. for many years, then the proportion of times it rained would be an approximation of the actual probability under consideration. Similarly, one can approximate the actual probability of the stock market acting positively in an election year.
1.4 Introduction to Probability Theory

With probability being interpreted in this way, the following relations are taken as axioms:

(A1) for any event \( A \), \( P(A) \geq 0 \)

(A2) \( P(S) = 1 \)

(A3) if \( A_1, A_2, \ldots \) are pairwise disjoint events, or \( A_i \cap A_j = \phi \) for \( i \neq j \),

then \( P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \)

Thus, any function \( P \), defined on the events of a sample space \( S \), that obeys axioms (A1)-(A3) is called a probability function.

Conditional Probability

The probability of an event \( B \) occurring when it is known that some other event \( A \) has occurred is called a conditional probability and is denoted by \( P(B \mid A) \). The notation \( P(B \mid A) \) is usually read as the probability of \( B \), given \( A \).

Definition 1.4.4 The conditional probability of \( B \), given \( A \), denoted by \( P(B \mid A) \) is defined as

\[
P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad \text{provided } P(A) > 0 \quad (1.4.2)
\]

As an illustration, suppose that our sample space consists of all students enrolled at a local community college in Fall 2004. We categorize the students according to their smoking habit (at least one cigarette per day or not) and gender in the following table.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Smoker</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Male</td>
<td>312</td>
<td>918</td>
</tr>
<tr>
<td>Female</td>
<td>117</td>
<td>773</td>
</tr>
<tr>
<td>Total</td>
<td>429</td>
<td>1691</td>
</tr>
</tbody>
</table>

Table 1.4.1: Student Smoking Habit vs. Gender.

One of these students is selected at random for a campaign to publicize the ill effects of smoking on health. We are concerned with the following
two events:

\[ F : \text{the selected student is a female} \]
\[ N : \text{the selected student is a nonsmoker} \]

Note that the sample space \( S \) consists of all 2120 students. We want to find the conditional probability \( P(F|N) \), i.e., the student we selected at random is a female given that the student is a nonsmoker.

Using the above definition,

\[
P(F|N) = \frac{P(F \cap N)}{P(N)} = \frac{\text{(# elements in } F \cap N) / \text{(# elements in } S)}{\text{(# elements in } N) / \text{(# elements in } S)}
\]

\[
= \frac{\text{(# elements in } F \cap N)}{\text{(# elements in } N)} = \frac{773}{1691} = 0.4571
\]

**Definition 1.4.5** Two events \( A \) and \( B \) are independent if and only if

\[ P(B|A) = P(B) \quad \text{or} \quad P(A|B) = P(A) \]

Otherwise, \( A \) and \( B \) are called dependent events.

**Theorem 1.4.1** If in an experiment the events \( A \) and \( B \) can both occur, then

\[ P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \]

**Theorem 1.4.2** Two events \( A \) and \( B \) are independent if and only if

\[ P(A \cap B) = P(A) \cdot P(B) \]

The following theorem, called the multiplication law, is concerned with the probability of the joint occurrence of several events.

**Theorem 1.4.3** Let \( A_1, A_2, \ldots, A_m \) be events of an experiment, then

\[ P(A_1 \cap A_2 \ldots \cap A_m) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots \]
\[ P(A_m|A_1 \cap A_2 \ldots \cap A_{m-1}) \]

Independence of several events can be defined as follows.

**Definition 1.4.6** A collection of events \( A_1, A_2, \ldots \) are called independent provided for any finite subcollection \( A_{i_1}, A_{i_2}, \ldots, A_{i_k}, \ k \geq 2 \),

\[ P(A_{i_1} \cap A_{i_2} \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}), \quad \text{for } 1 \leq i_1 < \ldots < i_k \]
Bayes’ Principle

Note that for a given event $A$, we can write

$$A = (A \cap B) \cup (A \cap B^c)$$

where $B$ is any other suitable event. This is illustrated through the Figure 1.4.1.

Figure 1.4.1: Venn Diagram for the Events $(A \cap B)$ and $(A \cap B^c)$.

Since $A \cap B$ and $A \cap B^c$ are mutually exclusive, using the addition and multiplication laws of probability, one can obtain the following equation:\(^5\)

$$P(A) = P(A \cap B) + P(A \cap B^c)$$
$$= P(B)P(A|B) + P(B^c)P(A|B^c)$$

The above probability equation is useful in many real life problems. The following example is one of such applications.

\(^5\)This equation is called the total probability formula.
Example 1.4.3: A medical test given to an individual is known to be 90% reliable when the person is HIV positive and 99% reliable when the person is HIV negative. In other words, 10% of the HIV positive individuals are diagnosed as negative, and 1% of the HIV negative individuals are diagnosed as positive by the medical test. If a person is selected from a population where 2% of the individuals are known to be HIV positive and the test indicates he or she is positive, what is the probability that he or she is actually HIV negative?

Consider the experiment of drawing an individual at random from a population consisting of 2% HIV positive people. The sample space \( S \) is the collection of all individuals in the population. Let \( B \) be the event of HIV positive individuals. It is known that \( P(B) = 0.02 \). Let \( A \) be the event that a person is diagnosed as HIV positive. From the reliability information about the test we have

\[
P(A | B) = 0.90 \quad \text{and} \quad P(A^c | B^c) = 0.99
\]

Now we need to find \( P(B^c | A) \).

From the conditional probability formula,

\[
P(B^c | A) = \frac{P(B^c \cap A)}{P(A)} = \frac{P(B^c \cap A)}{P(B \cap A) + P(B^c \cap A)} = \frac{P(B^c | A)}{P(B)}P(A | B) + P(B^c | A)P(A | B^c) = \frac{(1 - 0.02)(0.90) + (1 - 0.02)(1 - 0.99)}{0.8448}
\]

The above concept is generalized in the Bayes’ theorem as given below.

**Theorem 1.4.4 (Bayes’ Theorem)** Let the collection of events \( \{B_1, B_2, \ldots, B_m\} \) constitute a partition of the sample space \( S \) (i.e., \( B_i \cap B_j = \phi \) for \( i \neq j \); and \( S = B_1 \cup B_2 \ldots \cup B_m \)). Then given any event \( A \),

\[
P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^{m} P(B_j)P(A | B_j)}, \quad \text{for any } B_i, \quad 1 \leq i \leq m \quad (1.4.3)
\]
Remark 1.4.1 In many situations, $B_1, B_2, \ldots, B_m$ may be looked upon as the possible causes of the effect $A$. Bayes’ theorem then gives the probability of the cause $B_i$ when the effect $A$ has been observed, or the probability of the ‘hypothesis’ $B_i$ in the light of the ‘data’ $A$. In Bayesian analysis, Theorem 1.4.4 provides the ‘posterior probability’ of $B_i$ in terms of the ‘prior probabilities’ $P(B_i)$, $i = 1, 2, \ldots, m$, and the conditional probabilities of $A$.

1.5 Random Variables

A sample space $S$ could be difficult or inconvenient to describe if the sample points are not numbers. Often we are not interested in the details associated with each sample point but only in some numerical description of the outcome. We shall now discuss how we can use a rule by which an element $s$ of $S$ may be associated with a real value $x$.

Consider the example of tossing a coin three times in a row. The sample space $S$ is given as

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

If one is concerned only with the number of heads observed, then a numerical value of 0, 1, 2, or 3 will be assigned to each sample point. Notice that the numerical values 0, 1, 2, and 3 are random quantities determined by the outcome of the experiment. They can be thought of as values assumed by some random variable $X$, which in this case represents the number of heads when a coin is tossed three times.

Generally speaking, given an experiment with a sample space $S$, a function $X$ that assigns to each element $s$ in $S$ one and only one real number $X(s) = x$, is called a random variable. The range (or space) of $X$ is the set of real numbers $\{x \mid x = X(s), s \in S\} = R_X$ (say).

Definition 1.5.1 A random variable (r.v.) is a function that associates a real number with each element in the sample space.

We shall use capital letters with or without subscripts, like $X, Y, Z$ or $X_1, X_2, \ldots$ etc. to denote random variables.

Example 1.5.1: Let an experiment be rolling a die only once and observing the face value of the die. The sample space is known as
$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. For each $s \in \mathcal{S}$, define a random variable $Y$ as $Y(s) = s$. The range (or space) of the random variable $Y$ is then $R_Y = \{1, 2, 3, 4, 5, 6\}$.

**Definition 1.5.2** If the range of a random variable contains a finite number of values or an infinite sequence of values indexed by natural numbers (i.e., countable infinite), then the random variable is called a discrete random variable.

**Definition 1.5.3** If the range of a random variable contains infinitely many values equal to the number of points on a line segment or a collection of line segments (i.e., uncountably many values), then the random variable is called a continuous random variable.

For notational convenience, given a random variable $X$, we denote the event $\{s \mid s \in \mathcal{S} \text{ and } X(s) = a\}$ by $\{X = a\}$. That is, the event $\{X = a\}$ is the set of elements in the sample space that is mapped onto the real value, $a$, by the function $X$. Similarly, the event $\{s \mid s \in \mathcal{S} \text{ and } a < X(s) < b\}$ will be denoted by $\{a < X < b\}$. Now if we want to find the probabilities associated with the events described in terms of $X$, such as $\{X = a\}$ or $\{a < X < b\}$, we use the probabilities of those corresponding events in the original sample space $\mathcal{S}$. In other words, the probabilities for a random variable $X$ is defined as

$$P(\{X = a\}) = P(\{s \mid s \in \mathcal{S} \text{ and } X(s) = a\}) \quad \text{or}$$

$$P(\{a < X < b\}) = P(\{s \mid s \in \mathcal{S} \text{ and } a < X(s) < b\})$$

and these are called **induced probabilities**.

For a discrete random variable $X$, the induced probability $P(X = x)$, denoted by $f(x)$, is called the probability mass function (pmf). A pmf $f(x)$ of $X$ with range $R_X$ satisfies the following properties:

1. $f(x) > 0, \forall x \in R_X$
2. $\sum_{x \in R_X} f(x) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x), \quad \text{where } A \subseteq R_X$
1.5 Random Variables

Usually the pmf \( f(x) = 0 \) when \( x \notin R_X \). Since the probability \( f(x) = P(X = x) > 0 \) for \( x \in R_X \) where \( R_X \) contains all values of \( X \) with positive probabilities. \( R_X \) is also referred to as the \textbf{support} of \( X \).

**Example 1.5.2:** In an introductory statistics class of 40 students, there are 6 math majors, 11 nursing majors, 16 engineering majors, 3 criminal justice majors, and 4 undecided. One student is selected at random and the random variable \( X \) is assigned a score 0 for a student with undecided major, 1 for math, 2 for nursing, 3 for engineering, and 4 for criminal justice.

The range of \( X \) is thus \( R_X = \{0, 1, 2, 3, 4\} \). The pmf of \( X \) is given by

\[
\begin{align*}
    f(0) &= P(X = 0) = \frac{4}{40} \\
    f(1) &= P(X = 1) = \frac{6}{40} \\
    f(2) &= P(X = 2) = \frac{11}{40} \\
    f(3) &= P(X = 3) = \frac{16}{40} \\
    f(4) &= P(X = 4) = \frac{3}{40}
\end{align*}
\]

(1.5.1)

It is understood that \( f(x) = 0 \) when \( x \notin R_X \).

**Definition 1.5.4** The set of ordered pairs \( (x, f(x)) \) is called the probability distribution of the discrete random variable \( X \).

It is often helpful to look at a probability distribution in terms of a graph that depicts the pmf of \( X \). Two types of graphs can be used to give a visual representation of the pmf, namely, a bar graph and a probability histogram.

A bar graph of the pmf \( f(x) \) of a discrete random variable \( X \) is a graph of \( f(x) \) having a vertical line segment joining the points \( (x, 0) \) and \( (x, f(x)) \) for each \( x \in R_X \).

Instead of plotting the points \( (x, f(x)) \), we can also draw a rectangle of height \( f(x) \) and a base of length 1 centered at \( x \) for each \( x \in R_X \). Such a graphical representation is called a probability histogram. Even if the bases are not of unit width, we can still adjust the heights of rectangles to obtain areas that would be equal to \( f(x) \) at \( x \). This concept of using areas to represent probabilities is necessary for our consideration of the probability distribution of a continuous random variable.
Figure 1.5.1 and Figure 1.5.2 display a bar graph and a probability histogram, respectively, for the *pmf* \( f(x) \) (1.5.1).

\[ \text{Figure 1.5.1: Bar Graph of the pmf (1.5.1).} \]

In many problems, instead of considering events such as \( \{ X = x \} \) and the corresponding induced probabilities \( f(x) = P(X = x) \), we may wish to consider events of the form \( \{ X \leq x \} \) and the corresponding probabilities \( P(X \leq x) \). Writing \( F(x) = P(X \leq x) \) for every real number \( x \), we define \( F(x) \) to be the cumulative distribution function (cdf) of the random variable \( X \).

**Definition 1.5.5** *The cumulative distribution function (cdf) \( F(x) \) of a discrete random variable \( X \) with the probability mass function \( f(x) \) is given by*

\[
F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \quad \text{for } -\infty < x < \infty \quad (1.5.2)
\]
In Example 1.5.2, the random variable $X$ has the range $R_X = \{0, 1, 2, 3, 4\}$ with the pmf $f(x)$ in (1.5.1). Therefore,

\[
\begin{align*}
F(0) &= f(0) = \frac{4}{40} \\
F(1) &= f(0) + f(1) = \frac{10}{40} \\
F(2) &= f(0) + f(1) + f(2) = \frac{21}{40} \\
F(3) &= f(0) + f(1) + f(2) + f(3) = \frac{37}{40} \\
F(4) &= f(0) + f(1) + f(2) + f(3) + f(4) = 1
\end{align*}
\]
Hence

\[
F(x) = \begin{cases}
0 & \text{if } x < 0 \\
\frac{4}{40} & \text{if } 0 \leq x < 1 \\
\frac{10}{40} & \text{if } 1 \leq x < 2 \\
\frac{21}{40} & \text{if } 2 \leq x < 3 \\
\frac{37}{40} & \text{if } 3 \leq x < 4 \\
1 & \text{if } x \geq 4 
\end{cases}
\] (1.5.3)

The graph of the cdf in (1.5.3) is obtained by plotting points \((x, F(x))\) in Figure (1.5.3) which looks like a step function.

![Graph of the discrete cdf (1.5.3).](image)

Figure 1.5.3: Graph of the discrete cdf (1.5.3).

For a continuous random variable \(X\), the induced probability \(P(X = x), x \in R_X\), is zero. This may seem a bit puzzling, but it
1.5 Random Variables

becomes more plausible when we consider the following example.

Consider a random variable \( X \) representing the percentage change in the NASDAQ stock index at the end of a business day. Between any two fixed values, say 1.5 and 2.9%, there are infinitely many other possible values. What is the probability of the NASDAQ index going up at the end of a business day by exactly 2%? Note that the value 2% is only one of those infinitely many possible values, and hence a probability of zero is assigned to this event. This is not the case if we try to find the probability of the NASDAQ stock index going up by an amount that is strictly between 1.75 and 2.25% of the value on the previous day; i.e., we are talking about the probability of the event \( \{1.75 < X < 2.25\} \). Also, by the above argument,

\[
P(1.75 < X < 2.25) = P(1.75 < X < 2.25) + P(X = 2.25)
= P(1.75 < X \leq 2.25)
\]
because it does not matter whether we include an end point of the interval or not. Probability of an interval \((a, b]\) can be expressed as

\[
P(a < X \leq b) = P(X \leq b) - P(X \leq a)
\]
and similar to the discrete case here too for a continuous random variable \( X \), we have the cdf of \( X \) at \( a \) and \( b \) respectively as \( F(a) = P(X \leq a) \) and \( F(b) = P(X \leq b) \). Hence,

\[
P(a < X \leq b) = P(X \leq b) - P(X \leq a)
= F(b) - F(a)
\]

In the case of a continuous random variable \( X \), quite often it is possible to express the above probability \( F(b) - F(a) \) in terms of the integration of another function \( f(x) \) as

\[
P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) \, dx \quad (1.5.4)
\]
Such a function \( f(x) \), if it exists, is called the probability density function (pdf) of \( X \).

**Definition 1.5.6** A function \( f(x) \) is called a probability density function (pdf) for a continuous random variable \( X \) defined over \( \mathbb{R} \), the set of real numbers, if
Chapter 1 Basic Concepts

(a) $f(x) \geq 0$ for all $x \in \mathbb{R}$

(b) $\int_{-\infty}^{\infty} f(x) \, dx = 1$

(c) $P(a < X \leq b) = \int_{a}^{b} f(x) \, dx$ for all $-\infty < a < b < \infty$

Note that for a continuous random variable, we cannot define the range exactly the way it was done for a discrete random variable since, as we discussed earlier, a continuous random variable can assume a single value with probability zero. However, the concept of support can be extended to the continuous case as the set of all values for which the pdf is strictly positive, i.e., for a continuous random variable $X$ with pdf $f(x)$,

$$R_X = \text{support of } X = \{x \mid f(x) > 0\}$$

**Example 1.5.3:** A continuous random variable $X$ is defined as $2 \leq X \leq 5$, with the pdf $f(x) = \frac{2(1 + x)}{27}$. Find that the probability of the event $(X \leq 3)$.

It is known that the pdf $f(x)$ is zero where $x \notin [2, 5]$. Therefore, using Definition 1.5.6, we have

$$P(X \leq 3) = \int_{-\infty}^{3} f(x) \, dx = \int_{2}^{3} \frac{2}{27}(1 + x) \, dx = \frac{7}{27}$$

Similar to the discrete case, we define the cdf of a continuous random variable as follows.

**Definition 1.5.7** The cumulative distribution function (cdf) $F(x)$ of a continuous random variable $X$ with the pdf $f(x)$ is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t) \, dt \quad \text{for } -\infty < x < \infty$$

An immediate consequence of the above definition is that $f(x) = F'(x) = dF(x)/dx$ provided the derivative exists. It is noticed that for any random variable $X$, the cdf always exists, but the existence of the pdf is not guaranteed.

Any cumulative distribution function $F(x)$ possesses the following four properties:

(a) $0 \leq F(x) \leq 1$ because $F(x)$ is a probability

(b) $F(x)$ is a nondecreasing function of $x$
1.5 Random Variables

(c) \( \lim_{x \to -\infty} F(x) = 0 \) and \( \lim_{x \to +\infty} F(x) = 1 \)

(d) \( F(x) \) is right-continuous, i.e., \( \lim_{x \to x_0^+} F(x) = F(x_0) \) for any \( x_0 \)

[“\( x \to x_0^+ \)” means that \( x \) approaches to \( x_0 \) only from the right side of \( x_0 \).]

Definition 1.5.8 Let \( X \) be a random variable with pdf or pmf \( f(x) \). A value (a) \( M \) is called the median of \( X \) provided \( P(X \leq M) = 0.5 = P(X > M) \); (b) \( M_0 \) is called a mode of \( X \) provided \( f(M_0) \geq f(x) \), for all \( x \) in the range (or support) of \( X \).

Expectations of Random Variables

An important concept in summarizing basic characteristics of the probability distribution of a random variable is that of mathematical expectations. We shall introduce this concept through an example.

Example 1.5.4: Suppose we device a game where a participant pays a certain fee to toss a fair coin three times. If the person gets no heads, he/she receives nothing; if exactly one head occurs, he/she receives $1; if exactly two heads occur, he/she receives $2; and if all the three heads occur, he/she receives $3. Therefore, the events of interest are:

\[
A_1 = \{0 \text{ heads}\} = \{TTT\} \\
A_2 = \{1\text{ head}\} = \{HTT, THT, TTH\} \\
A_3 = \{2\text{ heads}\} = \{HHT, HTH, THH\} \\
A_4 = \{3\text{ heads}\} = \{HHH\}
\]

So, let the random variable \( X \) be the $ amount the participant receives. \( X \) has range \( R_X = \{0, 1, 2, 3\} \). The pmf of \( X \) can be obtained as

\[
P(X = 0) = \frac{1}{8}; \quad P(X = 1) = \frac{3}{8}; \quad P(X = 2) = \frac{3}{8}; \quad \text{and} \quad P(X = 3) = \frac{1}{8}
\]

If the game is played a large number of times, then about 1/8 of the times need a payment of $0, about 3/8 of the times need a payment of $1, about 3/8 of the times need a payment of $2, and 1/8 of the times need a payment of $3. Thus, the approximate average payment (in $) is

\[
(0) \left( \frac{1}{8} \right) + (1) \left( \frac{3}{8} \right) + (2) \left( \frac{3}{8} \right) + (3) \left( \frac{1}{8} \right) = 1.5
\]
Note that a participant never receives exactly $1.5; the rewards are either $0, $1, $2, or $3. However, the weighted average $1.5 is what we expect to pay per participant if the game is played for a large number of times. This weighted average is called the mathematical expectation of the participant’s reward. Thus, if we decide to charge a fee of $2 to each participant per play, then we can make, on average, a profit of 50 cents per play.

**Definition 1.5.9** If $X$ is a random variable (either discrete or continuous) with pmf or pdf $f(x)$ over a suitable range (or support) $R_X$, then the expectation (or the expected value) of $X$, denoted by $E(X)$ or $\mu_X$ or simply $\mu$ (when the subscript is clear from the context), is defined as

$$\mu_X = E(X) = \begin{cases} \sum_{x \in R_X} xf(x) & \text{if } X \text{ is discrete} \\ \int_{R_X} xf(x) \, dx & \text{if } X \text{ is continuous} \end{cases} \tag{1.5.5}$$

More generally, if $Y = g(X)$ is a real valued function of $X$ with pmf or pdf $f(x)$, then

$$\mu_Y = E(Y) = E[g(X)] = \begin{cases} \sum_{x \in R_X} g(x) f(x) & \text{if } X \text{ is discrete} \\ \int_{R_X} g(x) f(x) \, dx & \text{if } X \text{ is continuous} \end{cases} \tag{1.5.6}$$

The expectation of a random variable is also called its mean (or mean value).

**Definition 1.5.10** Let $X$ be a random variable with pmf or pdf $f(x)$ and mean $\mu_X$. The variance of $X$, denoted by $\sigma^2_X$ or simply $\sigma^2$, is defined as

$$\sigma^2_X = E[(X - \mu_X)^2] = \begin{cases} \sum_{x \in R_X} (x - \mu_X)^2 f(x) & \text{if } X \text{ is discrete} \\ \int_{R_X} (x - \mu_X)^2 f(x) \, dx & \text{if } X \text{ is continuous} \end{cases} \tag{1.5.7}$$

The positive square root of the variance, $\sigma_X$, is called the standard deviation of $X$.

Unless mentioned otherwise, we shall use $\mu$ and $\sigma^2$ to denote the mean and variance of a random variable $X$. An alternative expression of the
1.6 Joint Probability Distributions

variance of $X$ is given as

$$\sigma^2 = E(X^2) - \mu^2$$

(1.5.8)

The first term in the above expression, $E(X^2)$, is called the second raw moment of the random variable $X$. Thus, for any integer $k$, the $k$th raw moment of $X$, denoted by $\mu'_(k)$, is

$$\mu'_(k) = E(X^k)$$

(1.5.9)

provided the summation or integration involved is meaningful. Similarly, $k$th central moment of $X$, denoted by $\mu_(k)$, is

$$\mu_(k) = E[(X - \mu)^k]$$

(1.5.10)

Hence, we have $\mu = \mu'_(1)$ and $\sigma^2 = \mu_(2)$.

Some simple yet useful properties of expectation are shown in the following result.

**Theorem 1.5.1** Given a random variable $X$ with either pmf or pdf $f(x)$, the expectation operation ‘$E$’ satisfies the following properties whenever it exists.

(i) If $c$ is a constant, then $E(c) = c$.

(ii) If $c$ is a constant and $u(X)$ is a real valued function of $X$, then

$$E[cu(X)] = cE[u(X)]$$

(iii) If $c_1$ and $c_2$ are constants, and $u_1(X)$ and $u_2(X)$ are real valued functions of $X$, then

$$E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)]$$

1.6 Joint Probability Distributions

So far, we have restricted ourselves to one-dimensional sample spaces, and recorded outcomes of an experiment in terms of numerical values assumed by a single random variable. But in many real life situations, we may find it useful to record simultaneous outcomes of several random variables. For instance, one can measure the height (H) and weight (W) of a patient selected at random in a clinical trial, thus giving rise to a
two-dimensional range, say $R_{HW}$, consisting of all possible outcomes of $(H, W) = (h, w)$.

If $X_1$ and $X_2$ are two discrete random variables, the probability distribution for their simultaneous occurrence can be represented by a function $f(x_1, x_2)$ for any pair of values $(x_1, x_2)$ within the range (say, $R_{X_1X_2}$ or simply $R_{12}$) of the random variables $X_1$ and $X_2$. It is customary to refer this function as the joint probability distribution of $X_1$ and $X_2$. Therefore, in the discrete case,

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

Suppose $X_1$ and $X_2$ represent, respectively, the number of children and the number of pets in a household. Then $f(3, 2)$ is the probability that a randomly selected household has 3 children and 2 pets.

**Definition 1.6.1** The function $f(x_1, x_2)$ is a joint probability distribution of the discrete random variables $X_1$ and $X_2$ with range $R_{12}$ if

(i) $f(x_1, x_2) > 0$ for all $(x_1, x_2) \in R_{12}$;

(ii) $\sum_{(x_1, x_2) \in R_{12}} f(x_1, x_2) = 1$; and

(iii) $P((X_1, X_2) \in A) = \sum_A f(x_1, x_2)$ for any subregion $A \subseteq R_{12}$.

**Example 1.6.1:** Consider a small town where the joint probability distribution of $X_1 = \#$ children per household and $X_2 = \#$ pets per household is given through the Table 1.6.1. If a household is selected at random, what is the probability that the total number of kids and pets in the household is less than or equal to 2?

<table>
<thead>
<tr>
<th>$f(x_1, x_2)$</th>
<th>$X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>$X_1$ 0</td>
<td>6/30</td>
</tr>
<tr>
<td>1</td>
<td>6/30</td>
</tr>
<tr>
<td>2</td>
<td>4/30</td>
</tr>
<tr>
<td>3</td>
<td>1/30</td>
</tr>
</tbody>
</table>

Table 1.6.1: Joint Probability Distribution of $(X_1, X_2)$.

Note that in terms of $X_1$ and $X_2$, we are interested in finding the probability $P(0 \leq X_1 + X_2 \leq 2)$. The subregion $A = \{(x_1, x_2) \mid 0 \leq x_1 + x_2 \leq 2\}$
$x_1 + x_2 \leq 2$) includes the pairs of (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), and (2, 0). Therefore,

$$P(0 \leq X_1 + X_2 \leq 2) = f(0, 0) + f(0, 1) + f(0, 2) + f(1, 0) + f(1, 1) + f(2, 0) = \frac{25}{30}$$

When $X_1$ and $X_2$ are continuous random variables, the joint probability density function $f(x_1, x_2)$ is a surface lying above the $x_1$-$x_2$ plane, and $P((X_1, X_2) \in A)$, where $A$ is any subregion in the $x_1$-$x_2$ plane, is equal to the volume of the right cylinder bounded by the base $A$ and the surface $f(x_1, x_2)$.

**Definition 1.6.2** The function $f(x_1, x_2)$ is a joint probability distribution of the continuous random variables $X_1$ and $X_2$ with range $R_{12}$ if

(i) $f(x_1, x_2) > 0$ for all $(x_1, x_2) \in R_{12}$;

(ii) $\int_{R_{12}} f(x_1, x_2) \, dx_1 \, dx_2 = 1$; and

(iii) $P((X_1, X_2) \in A) = \int_A f(x_1, x_2) \, dx_1 \, dx_2$ for any subregion $A \subseteq R_{12}$.

**Example 1.6.2:** Two continuous random variables, $X_1$ and $X_2$, have the joint density given by

$$f(x_1, x_2) = \begin{cases} 
  k(x_1^2 + x_2^2), & 0 < x_1 < 2, \, 1 < x_2 < 4 \\
  0, & \text{otherwise}
\end{cases}$$

Find $P(X_1 > X_2)$.

First, we need to find the value of $k$. Using (ii) of Definition 1.6.2, we have

$$\int_0^2 \int_1^4 k(x_1^2 + x_2^2) \, dx_2 \, dx_1 = 1$$

After a straightforward integration, we obtain that $k = 1/50$. Hence, using (iii) of Definition 1.6.2 (also see Figure 1.6.1),

$$P(X_1 > X_2) = \int_0^2 \int_1^{x_1} \frac{1}{50} (x_1^2 + x_2^2) \, dx_2 \, dx_1$$

A further simplification gives that $P(X_1 > X_2) = 7/150$. 

© 2006 by Taylor & Francis Group, LLC
Given the joint probability distribution $f(x_1, x_2)$ of the discrete random variables $X_1$ and $X_2$, the probability distribution $f_1(x_1)$ of $X_1$ alone is obtained by summing $f(x_1, x_2)$ up over the values of $X_2$. Similarly, the probability distribution $f_2(x_2)$ of $X_2$ alone is obtained by summing $f(x_1, x_2)$ up over the values of $X_1$. The functions $f_1(x_1)$ and $f_2(x_2)$ are called the marginal distributions of $X_1$ and $X_2$, respectively. For continuous random variables, summations are replaced by integrals.

The marginal distributions $f_1(x_1)$ and $f_2(x_2)$ are indeed the probability distributions of the individual random variables $X_1$ and $X_2$, respectively.
Definition 1.6.3 The marginal distributions of individual random variables \( X_1 \) and \( X_2 \) are given by

\[
f_1(x_1) = \sum_{x_2 \in R_{12}^x} f(x_1, x_2) \quad \text{and} \quad f_2(x_2) = \sum_{x_1 \in R_{12}^x} f(x_1, x_2)
\]

for the discrete case, and by

\[
f_1(x_1) = \int_{x_2 \in R_{12}^x} f(x_1, x_2) \, dx_2 \quad \text{and} \quad f_2(x_2) = \int_{x_1 \in R_{12}^x} f(x_1, x_2) \, dx_1
\]

for the continuous case; where \( R_{12}^x \) is the set of possible values that \( X_2 \) can assume when \( X_1 = x_1 \), and \( R_{12}^x \) is the set of possible values that \( X_1 \) can assume when \( X_2 = x_2 \).

The concept of conditional probability can now be invoked in the context of joint probability distributions.

Define the two events \( A_1 \) and \( A_2 \) as \( A_1 = \{ X_1 = x_1 \} \) and \( A_2 = \{ X_2 = x_2 \} \) where \( X_1 \) and \( X_2 \) are two discrete random variables. Using Definition 1.4.4 we can obtain

\[
P(X_2 = x_2 \mid X_1 = x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \frac{f(x_1, x_2)}{f_1(x_1)} = f_{2|1}(x_2 \mid x_1) \quad \text{(say)}
\]

provided \( f_1(x_1) > 0 \).

It can be shown that \( f_{2|1}(x_2 \mid x_1) \), which is strictly a function of \( x_2 \) when \( x_1 \) is fixed, is also a probability distribution, and this is true even if the random variables are continuous.

Definition 1.6.4 Let \( X_1 \) and \( X_2 \) be two random variables (discrete or continuous). The conditional distribution of the random variable \( X_2 \), given that \( X_1 = x_1 \), is given by

\[
f_{2|1}(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}, \quad \text{provided} \ f_1(x_1) > 0
\]

Similarly, the conditional distribution of the random variable \( X_1 \), given that \( X_2 = x_2 \), is given by

\[
f_{1|2}(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}, \quad \text{provided} \ f_2(x_2) > 0
\]
Example 1.6.3: Consider the Table 1.6.1 representing the joint probability distribution of $X_1$ and $X_2$, the number of children and the number of pets per household, respectively. The marginal distributions are

$$f_1(0) = \frac{13}{30}, \quad f_1(1) = \frac{10}{30}, \quad f_1(2) = \frac{5}{30}, \quad f_1(3) = \frac{2}{30}$$

and

$$f_2(0) = \frac{17}{30}, \quad f_2(1) = \frac{10}{30}, \quad f_2(2) = \frac{2}{30}, \quad f_2(3) = \frac{1}{30}$$

Given that a household has two children, what is the probability that the household has no pets? In terms of $X_1$ and $X_2$, it is seen that

$$P(X_2 = 0 | X_1 = 2) = \frac{f_2(0)}{f_1(2)} = \frac{4}{5}$$

As it can be seen in the above example,

$$f(2, 0) = \frac{4}{30} \neq f_1(2) \cdot f_2(0) = \frac{5}{30} \times \frac{17}{30}$$

But there are cases where $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ for all $(x_1, x_2)$ in the range $R_{12}$.

Definition 1.6.5 Given two random variables $X_1$ and $X_2$ (discrete or continuous) with joint probability distribution $f(x_1, x_2)$ and two marginal distributions $f_1(x_1)$ and $f_2(x_2)$, $X_1$ and $X_2$ are said to be independent if and only if $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ for all $(x_1, x_2)$ within the range $R_{12}$ of $(X_1, X_2)$.

Example 1.6.4: Two continuous random variables have the joint density given by

$$f(x_1, x_2) = \begin{cases} 
4x_1x_2, & 0 < x_1 < 1, \ 0 < x_2 < 1 \\
0, & \text{otherwise}
\end{cases}$$

Are $X_1$ and $X_2$ independent?

The marginal densities of $X_1$ and $X_2$ are respectively

$$f_1(x_1) = \int_0^1 4x_1x_2 \, dx_2 = 2x_1$$

$$f_2(x_2) = \int_0^1 4x_1x_2 \, dx_1 = 2x_2$$

Hence, $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ for all $(x_1, x_2)$ such that $0 < x_1 < 1$ and $0 < x_2 < 1$. Therefore, $X_1$ and $X_2$ are independent.
It is also worth mentioning that two random variables are independent if and only if
\[ f_{1|2}(x_1|x_2) = f_1(x_1) \quad \text{or} \quad f_{2|1}(x_2|x_1) = f_2(x_2) \]
for all \((x_1, x_2)\) in the range \(R_{12}\).

**Definition 1.6.6** Let \(X_1\) and \(X_2\) be random variables with joint probability distribution \(f(x_1, x_2)\). The covariance of \(X_1\) and \(X_2\), denoted by \(\sigma_{12}\), is defined as
\[
Cov(X_1, X_2) = \sigma_{12} = E \left[ (X_1 - E(X_1))(X_2 - E(X_2)) \right]
\]
If we denote \(E(X_1)\) and \(E(X_2)\) by \(\mu_1\) and \(\mu_2\), respectively, then \(\sigma_{12}\) can be rewritten as
\[
\sigma_{12} = \begin{cases} 
\sum_{(x_1, x_2) \in R_{12}} (x_1 - \mu_1)(x_2 - \mu_2)f(x_1, x_2) & \text{discrete case} \\
\int_{R_{12}} (x_1 - \mu_1)(x_2 - \mu_2)f(x_1, x_2) \, dx_1 \, dx_2 & \text{continuous case}
\end{cases}
\]
\[
= E(X_1X_2) - \mu_1\mu_2
\]
It can be shown that for constants \(a_1, a_2, b_1, \) and \(b_2, \)
\[
Cov(a_1X_1 + b_1, a_2X_2 + b_2) = a_1a_2 \, Cov(X_1, X_2)
\]
Therefore, covariance of two random variables can be affected by the choice of scales for recording \(X_1\) and \(X_2\). Covariance can also be viewed as a measure of association between the random variables \(X_1\) and \(X_2\). But since the covariance depends on the units being used for \(X_1\) and \(X_2\), a unit free measure of the association between \(X_1\) and \(X_2\) is found through correlation coefficient.

**Definition 1.6.7** The correlation coefficient between two random variables \(X_1\) and \(X_2\) is defined as
\[
\rho_{12} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1) \cdot Var(X_2)}}.
\]
Properties of the correlation coefficient:

(a) Obviously, the correlation coefficient \(\rho_{12}\) is a pure number, that is, \(\rho_{12}\) is independent of units of measurement of both \(X_1\) and \(X_2\).
(b) Let \( X^*_1 = (X_1 - a_1)/b_1 \) and \( X^*_2 = (X_2 - a_2)/b_2 \) where \( a_1, b_1, a_2, \) and \( b_2 \) are four arbitrarily chosen constants. Let \( \rho^*_1 = \rho^*_2 = \rho^*_{12} \) be the correlation coefficient between \( X^*_1 \) and \( X^*_2 \). It can be shown that \( \rho^*_{12} = \rho_{12} = \) correlation coefficient between \( X_1 \) and \( X_2 \).

(c) For any two random variables, \( X_1 \) and \( X_2 \), it is true that \(-1 \leq \rho_{12} \leq +1\). The correlation coefficient \( \rho_{12} \) takes the lowest value \(-1\) when \( X_2 = \mu_2 - \frac{a_2}{\sigma_2}(X_1 - \mu_1) \), and the highest value \(+1\) when \( X_2 = \mu_2 + \frac{a_2}{\sigma_2}(X_1 - \mu_1) \). Here, \( \mu_i \) and \( \sigma_i \) are mean and standard deviation of \( X_i \), respectively, \( i = 1, 2 \).

(d) Positive correlation (i.e., \( 0 < \rho_{12} \leq 1 \)) implies a positive association between \( X_1 \) and \( X_2 \). In other words, large values of \( X_1 \) tend to be associated with large values of \( X_2 \). On the other hand, negative correlation (i.e., \( -1 \leq \rho_{12} < 0 \)) implies a negative association between \( X_1 \) and \( X_2 \), i.e., large values of \( X_1 \) tend to be associated with small values of \( X_2 \) and vice-versa. If \( X_1 \) and \( X_2 \) are independent, then \( \rho_{12} = 0 \), but the converse is not always true.

All the preceding concepts concerning two random variables can be generalized to the case of \( n \) random variables.

Let \( f(x_1, x_2, \ldots, x_n) \) be the joint probability distribution (pmf or pdf) of the random variables \( X_1, X_2, \ldots, X_n \). The marginal distribution of \( X_i \) is given by

\[
 f_i(x_i) = \sum_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} f(x_1, x_2, \ldots, x_n)
\]

for the discrete case, and by

\[
 f_i(x_i) = \int \cdots \int f(x_1, x_2, \ldots, x_n) \, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n
\]

for the continuous case where the summation or integration is taken over all possible values of \( (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \) where the \( i^{th} \) component is fixed at \( x_i \). Similarly, we can obtain the joint marginal distribution of \( (X_1, X_2, \ldots, X_k), \ 1 \leq k < n, \) as

\[
 f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k) = \begin{cases} 
 \sum_{x_{k+1}, \ldots, x_n} f(x_1, x_2, \ldots, x_n), & \text{discrete case} \\
 \int \cdots \int f(x_1, x_2, \ldots, x_n) \, dx_{k+1} \cdots dx_n, & \text{continuous case}
\end{cases}
\]
As a result, the joint conditional distribution of \( (X_{k+1}, \ldots, X_n) \) given \( (X_1 = x_1, \ldots, X_k = x_k) \) is

\[
f_{k+1,\ldots,n|1,\ldots,k}(x_{k+1}, \ldots, x_n \mid x_1, \ldots, x_k) = \frac{f(x_1, x_2, \ldots, x_n)}{f_{1,\ldots,k}(x_1, x_2, \ldots, x_k)}
\]

provided \( f_{1,\ldots,k}(x_1, x_2, \ldots, x_k) > 0 \). The random variables, \( X_1, X_2, \ldots, X_n \) with joint distribution \( f(x_1, x_2, \ldots, x_n) \) and marginal distributions \( f_1(x_1), f_2(x_2), \ldots, f_n(x_n) \), respectively, are said to be mutually independent if and only if

\[
f(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = f_1(x_{i_1}) f_2(x_{i_2}) \cdots f_k(x_{i_k})
\]

for all \( 1 < k \leq n \) and \((x_{i_1}, x_{i_2}, \ldots, x_{i_k})\) within the underlying range.

Note that with the above setup of \( n \) random variables \( X_1, X_2, \ldots, X_n \) with joint probability distribution \( f(x_1, x_2, \ldots, x_n) \), the mean of \( X_i \) is

\[
\mu_i = \mu_{X_i} = E(X_i) = \left\{ \begin{array}{ll}
\sum x_i f_i(x_i), & \text{discrete case} \\
\int x_i f_i(x_i) \, dx_i, & \text{continuous case}
\end{array} \right.
\]

where the summation or integration takes over the appropriate range (unconditional) of the random variable \( X_i \). The vector \( \mu = (\mu_1, \mu_2, \ldots, \mu_n)' \) is called the mean vector of the random vector \( X = (X_1, X_2, \ldots, X_n)' \).

The matrix \( \Sigma_{n \times n} \) defined as

\[
\Sigma = E[(X - \mu)(X - \mu)']
\]

\[
= \begin{pmatrix}
\text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\
\text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n)
\end{pmatrix}
\]

is called the dispersion matrix of the random vector \( X \). Using the notations \( \sigma_{ij} = \text{Cov}(X_i, X_j), i \neq j \), and \( \sigma_{ii} = \text{Var}(X_i) = \sigma_i^2 \) (say), \( 1 \leq i \leq n \), the matrix \( \Sigma \) can be denoted as \( \Sigma = \begin{pmatrix} (\sigma_{ij}) \end{pmatrix} \). It is easy to decompose \( \Sigma \) as

\[
\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n) \cdot \rho \cdot \text{diag}(\sigma_1, \ldots, \sigma_n)
\]

\( ^6\text{diag}(\sigma_1, \ldots, \sigma_n) \) is a \( n \times n \) diagonal matrix with diagonal elements of \( \sigma_1, \ldots, \sigma_n \).
where the matrix $\mathbf{\rho}$ is given as

$$
\mathbf{\rho} = \begin{pmatrix}
1 & \rho_{12} & \cdots & \rho_{1n} \\
\rho_{21} & 1 & \cdots & \rho_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{n1} & \rho_{n2} & \cdots & 1
\end{pmatrix}
$$

and is called the correlation matrix. For $i \neq j$, $\rho_{ij}$ is the correlation coefficient between $X_i$ and $X_j$.

### 1.7 Moment Generating Function

Earlier, we discussed the $k$th moment (raw) $\mu'_k$ of a random variable $X$ given as $\mu'_k = E(X^k)$, provided it exists. A unified approach to find all moments (if they exist) of $X$ can be obtained through the moment generating function (mgf) of $X$.

The moment generating function of $X$, denoted by $\phi_X(t)$, is defined as

$$
\phi_X(t) = E(e^{tX})
$$

(1.7.1)

provided the above expectation exists for all $t \in (-\varepsilon, \varepsilon)$, $\varepsilon > 0$, a small neighborhood of 0.

The mgf $\phi_X(t)$ of $X$ acts as a tool to derive all moments of $X$. It can be shown that

$$
\phi^{(k)}_X(0) = \phi^{(k)}_X(t) \bigg|_{t=0} = \left( \frac{d^k}{dt^k} \phi_X(t) \right) \bigg|_{t=0} = \mu'_k
$$

(1.7.2)

As a consequence, we have

$$
\mu_X = E(X) = \mu'_1 = \frac{d}{dt} \phi_X(t) \bigg|_{t=0} = \phi^{(1)}_X(t) \bigg|_{t=0}
$$

$$
E(X^2) = \mu'_2 = \frac{d^2}{dt^2} \phi_X(t) \bigg|_{t=0} = \phi^{(2)}_X(t) \bigg|_{t=0}, \ldots, \text{etc.}
$$

It should be pointed out that the existence of the mgf of a random variable $X$ is a stronger condition than the existence of its moments. If the mgf $\phi_X(t)$ exists, then all moments $\mu'_k$ of $X$ exist and can be obtained by (1.7.2). However, all or a few moments of $X$ may exist, yet $\phi_X(t)$ may not exist. A few such examples will be discussed later in Chapter 2.
Apart from its ability to generate moments, a major usefulness of the moment generating function is that it characterizes a probability distribution uniquely. Note that if the mgf exists, it characterizes an infinite collection of moments. Therefore, a natural question is whether characterizing the infinite set of moments uniquely determines a probability distribution. The answer to this question is ‘no’. Characterizing the set of moments is not sufficient to determine a probability distribution uniquely since there can be two distinct probability distributions having the same set of moments. This can be illustrated through the following example.

Example 1.7.1: Consider two continuous random variables, \( X_1 \) and \( X_2 \), having pdfs \( f_1(x_1) \) and \( f_2(x_2) \), respectively, given by

\[
f_1(x_1) = \frac{1}{\sqrt{2\pi x_1}} \exp\left(-\left(\ln x_1\right)^2/2\right), \quad 0 < x_1 < \infty
\]

and

\[
f_2(x_2) = \left(1 + \sin(2\pi \ln x_2)\right)f_1(x_2), \quad 0 < x_2 < \infty
\]

It can be shown that for any positive integer \( k \), \( E\left(X_1^k\right) = E\left(X_2^k\right) = \exp(k^2/2) \). Thus, even though \( X_1 \) and \( X_2 \) have distinct pdfs, they have the same moments.

If the mgf exists then it actually says more than the sequence of moments it can generate.

The above problem of nonuniqueness of the distributions does not arise if the probability distributions have bounded support. In that case, the infinite sequence of moments does uniquely determine the probability distribution. Also, if the mgf \( \phi_X(t) \) of a random variable \( X \) exists in a neighborhood of 0, then the probability distribution of \( X \) is uniquely determined regardless of the support. The next theorem summarizes the above discussion.

**Theorem 1.7.1** Let \( X_1 \) and \( X_2 \) be two random variables with probability distributions, \( f_1(x_1) \) and \( f_2(x_2) \), respectively. Let the corresponding mgfs be \( \phi_1(t) \) and \( \phi_2(t) \), respectively (assuming that they both exist).

(i) If \( f_1 \) and \( f_2 \) have bounded support, then \( f_1(x) = f_2(x) \) for all \( x \) if, and only if, \( E\left(X_1^k\right) = E\left(X_2^k\right) \) for all integers \( k = 0, 1, 2, \ldots \).

(ii) If \( \phi_1(t) = \phi_2(t) \) for all \( t \) in some neighborhood of 0, then \( f_1(x) = f_2(x) \) for all \( x \).
Chapter 1  Basic Concepts

The *mgf* of a random vector \( \mathbf{X} = (X_1, X_2, \ldots, X_n)' \) is defined as

\[
\phi_\mathbf{X}(t) = E\left(e^{t'\mathbf{X}}\right) = E\left(e^{\sum_{i=1}^{n} t_i X_i}\right) \tag{1.7.3}
\]

provided the above expectation exists for all \( t = (t_1, t_2, \ldots, t_n)' \) in a neighborhood of \( \mathbf{0} = (0, \ldots, 0)' \). In the special case, when the random variables \( X_1, X_2, \ldots, X_n \) are mutually independent,

\[
\phi_\mathbf{X}(t) = \prod_{i=1}^{n} E(e^{t_i X_i}) = \prod_{i=1}^{n} \phi_i(t_i)
\]

where \( \phi_i(t_i) \) is the *mgf* of \( X_i, 1 \leq i \leq n \).

Some Moment Inequalities

The following moment inequalities are widely used in mathematical statistics, and they can help to explain the characteristics of many statistical experiments.

(a) **Cauchy-Schwarz Inequality**: For any two random variables \( X \) and \( Y \),

\[
E|XY| \leq \sqrt{E(X^2)E(Y^2)} \tag{1.7.4}
\]

(b) **Jensen’s Inequality**: For any random variable \( X \) and a convex function \( h(x) \),

\[
E[h(X)] \geq h(E(X)) \tag{1.7.5}
\]

(c) **Hölder’s Inequality**: For any two random variables \( X \) and \( Y \), and two positive real values \( a \) and \( b \) such that \( a + b = 1 \),

\[
E|XY| \leq \left(E(|X|^{1/a})\right)^a \left(E(|Y|^{1/b})\right)^b \tag{1.7.6}
\]

It is seen that (1.7.4) is a special case of (1.7.6) with \( a = b = 1/2 \).

(d) **Liapounov’s Inequality**: If one takes \( Y \equiv 1 \) in (1.7.6), then the Hölder’s Inequality reduces to:

\[
E|X| \leq \left(E(|X|^{1/a})\right)^a, \quad \text{for } 0 < a \leq 1
\]

A function \( h(x) \) is called convex if \( h(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda h(x_1) + (1 - \lambda)h(x_2) \) for all \( x_1 \) and \( x_2 \), and \( \lambda \in (0, 1) \). Also, if \( h''(x) \geq 0 \), then \( h(x) \) is convex. On the other hand, \( h(x) \) is said to be concave if \( ( - h(x) ) \) is convex, or \( h''(x) \leq 0 \).
(c) Minkowski’s Inequality: For any two random variables $X$ and $Y$, and any real value $a$, $0 < a \leq 1,$

$$\left( E(|X+Y|^{1/a})\right)^a \leq \left( E(|X|^{1/a})\right)^a + \left( E(|Y|^{1/a})\right)^a \quad (1.7.7)$$

### 1.8 Order Statistics

If $X_1, X_2, \ldots, X_n$ are random variables, and $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are the same variables arranged in ascending order of magnitude such as

- $X_{(1)}$ = the smallest observation of $\{X_1, X_2, \ldots, X_n\}$
- $X_{(2)}$ = the second smallest observation of $\{X_1, X_2, \ldots, X_n\}$
- $\vdots$
- $X_{(n-1)}$ = the second largest observation of $\{X_1, X_2, \ldots, X_n\}$
- $X_{(n)}$ = the largest observation of $\{X_1, X_2, \ldots, X_n\}$

then $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are called the order statistics of the random variables $X_1, X_2, \ldots, X_n$.

Ordered values of the random observations have many real life applications. For instance, to study the strength of steel cable used in heavy-duty cranes, a random sample of ten specimens has been obtained from a manufacturer. Each specimen is then tested in the lab where the stress level is measured at which a cable breaks (and this is called the ‘breaking point’ of a cable). For the sample of ten cable specimens, we thus obtain $X_1, X_2, \ldots, X_{10}$ where $X_i$ represents the $i^{th}$ breaking point. For field applications of this particular type of cable, one may be interested in the probability distribution of either $X_{(1)}$ or $X_{(10)}$, or $(X_{(5)} + X_{(6)})/2$, the sample median of these ten ‘breaking point’ values.

If the random variables $X_1, X_2, \ldots, X_n$ are iid with cdf $F(x)$ and pdf $f(x)$, then the pdf of $X_{(1)}$ is

$$f_{(1)}(x) = nf(x)(1-F(x))^{n-1} \quad (1.8.1)$$

and that of $X_{(n)}$ is

$$f_{(n)}(x) = nf(x)(F(x))^{n-1} \quad (1.8.2)$$
More generally, the pdf of $X_{(j)}$, $1 \leq j \leq n$, is

$$f_{(j)}(x) = \frac{n!}{(j-1)! (n-j)!} f(x) (F(x))^{j-1} (1 - F(x))^{n-j} \quad (1.8.3)$$

Joint distribution of $(X_{(i_1)}, \ldots, X_{(i_j)})$, where $1 \leq i_1 \leq i_2 \leq \ldots \leq i_j \leq n$, can be found as

$$f_{(i_1),\ldots,(i_j)}(x_{i_1}, \ldots, x_{i_j}) = \frac{n!}{(i_j - i_1)! (i_j - i_2)! (n - i_j)!} f(x_{i_1}) \cdots f(x_{i_j}) \times 
\prod_{k=1}^{i_1} \left\{ F(x_{i_k}) - F(x_{i_{k-1}}) \right\}^{i_k - i_{k-1}} \times \prod_{k=1}^{j} f(x_{i_k})$$

$$\quad (1.8.4)$$

with $i_0 = 0, i_{j+1} = n, F(x_{i_0}) = 0, F(x_{i_{j+1}}) = 1$ and $x_{i_1} \leq \ldots \leq x_{i_j}$.

As a special case of the above formula (1.8.4), one can find the joint pdf of $X_{(i_1)}$ and $X_{(i_2)}$, $1 \leq i_1 < i_2 \leq n$, as

$$f_{(i_1),(i_2)}(x_{i_1}, x_{i_2}) = \frac{n!}{(i_2 - i_1 - 1)! (n - i_2)!} f(x_{i_1}) f(x_{i_2}) \times 
\left( F(x_{i_2}) - F(x_{i_1}) \right)^{i_2 - i_1 - 1} \times \left[ 1 - F(x_{i_2}) \right]^{n-i_2}$$

$$\quad (1.8.5)$$

for $-\infty < x_{i_1} < x_{i_2} < \infty$. Further, by taking $i_1 = 1$ and $i_2 = n$ we get the joint pdf of $X_{(1)}$ and $X_{(n)}$ (the smallest and the largest of the observations, respectively) as

$$f_{(1),(n)}(x_1, x_n) = n(n-1) f(x_1) f(x_n) \left[ F(x_n) - F(x_1) \right]^{n-2} \quad (1.8.6)$$

The joint pdf (1.8.6) can be used to find the probability distribution of, say $R = X_{(n)} - X_{(1)}$, the sample range (a measure of dispersion), or $V = (X_{(1)} + X_{(n)})/2$, the mid-range (a measure of location).

Order statistics play a major role in experiments where one can obtain censored observations. If an experiment is stopped after the first $k$ ordered observations are obtained out of $n$ test units, then it is called Type-II censoring on the right. Thus one obtains only $X_{(1)} \leq \ldots \leq X_{(k)}$. Similarly, Type-II censoring on the left would happen if the smallest $k$ observations are ignored and one ends up with $X_{(k+1)} \leq \ldots \leq X_{(n)}$.

Analogously, if an experiment is discontinued after a fixed length of time $t_*$ then this is called Type-I censoring on the right. For this kind
of censoring the length of the experiment is fixed, but the number of observations (say, \( N \)) before time \( t_* \) is a discrete random variable, and one observes \( X_{(1)} \leq \ldots \leq X_{(N)} \) where \( N = \# \) observations before the fixed time \( t_* \).

Suppose we are interested in the lifespan of a particular brand of electric bulbs and \( X_i \) denote the lifespan of the \( i \)th randomly selected bulb \((1 \leq i \leq n)\). If we decide to stop the experiment as soon as the \( k \)th ordered observation is obtained (i.e., Type-II right censoring is employed), then the expected length of the experiment is \( E(X_{(k)}) \). On the other hand, if Type-I censoring with fixed time \( t_* \) is used, then a bulb fails before time \( t_* \) with probability \( P(X_i \leq t_*) = F(t_*) \) (\( F \) being the common cdf of \( X_i \)'s). Further, \( N \) follows a binomial probability distribution (discussed in Chapter 2) \( P(N = k) = \binom{n}{k} F(t_*)^k (1 - F(t_*))^{n-k} \). This can be used to find the joint distribution of \((X_{(1)}, \ldots, X_{(N)}, N)\) as

\[
g(x_{(1)}, \ldots, x_{(N)}, N = k) = g(x_{(1)}, \ldots, x_{(N)}|N = k)P(N = k) \\
= k! \prod_{i=1}^{k} \left\{ \frac{f(x_{(i)})}{F(t_*)} \right\} \binom{n}{k} F(t_*)^k \\
\times (1 - F(t_*))^{n-k} \\
= \frac{n!}{(n-k)!} (1 - F(t_*))^{n-k} \prod_{i=1}^{k} f(x_{(i)})
\]

where \( x_{(1)} \leq \ldots \leq x_{(k)} < t_* \), \( k = 1, 2, \ldots, n \) and \( f(\cdot) \) is the pdf of \( X_i \)'s. The quantity \( E(N) \) gives an idea about the number of failures one expects to see within the fixed time \( t_* \).

### 1.9 Characteristic Function

In Section 1.7 we have discussed the mgf of a r.v. \( X \) which may or may not exist. But the most important function which characterizes the probability distribution of a random variable is the characteristic function \((cf)\) which is defined at a real value \( t \) as

\[
\Psi_X(t) = E(e^{itX}) \tag{1.9.1}
\]

where \( i = \sqrt{-1} \), the complex number. The above expectation exists always and requires complex (or contour) integration. The \( cf \) uniquely determines the probability distribution associated with the r.v. \( X \) (i.e.,
given a cf expression, it is possible to extract the cdf of \( X \). Also, if the moments of a r.v. \( X \) exist, then they can be generated from \( \Psi_X(t) \) just the way it can be done for the mgf, assuming that the mgf exists, (i.e., by taking derivative(s) of \( \Psi_X(t) \) w.r.t. \( t \) and then setting \( t = 0 \)).
Chapter 2

Some Common
Probability Distributions

In this chapter, we discuss some commonly used probability distributions. We will see that, while discussing a particular probability distribution, we essentially deal with a family of probability distributions indexed by one or more parameters. This family of distributions gives us some liberty to vary the parameter value(s) while staying with a particular functional form. As a result, when it comes to model the outcomes of an experiment (i.e., approximating the actual probability distribution of a random variable), we can choose a suitable parameter value that can give the resultant, commonly used probability distribution a good fit to the actual one. Hence, to emphasize the role of the parameter(s) and keep track of the parameter(s), a typical pdf or pmf will be denoted by, for example, $f(x|\theta)$ instead of $f(x)$, where $\theta$ is a relevant parameter involved in the probability distribution. If there is a single parameter involved, then $\theta$ is taken as scalar valued, otherwise $\theta$ is taken as vector valued. The collection of all possible values of $\theta$, denoted by $\Theta$, which makes $f(x|\theta)$ a pdf or pmf, is called the (natural) parameter space of the probability distribution $f(x|\theta)$. For the sake of simplicity, we take $\theta$ as scalar valued only, but if a situation demands, then $\theta$ can be vector valued too.

**Notation:** If a random variable has probability distribution (pmf or pdf) $f(x|\theta)$, then it is written as $X \sim f(x|\theta)$. 
Chapter 2  Some Common Probability Distributions

2.1  Discrete Distributions

Discrete Uniform Distribution

A random variable $X$ has a Discrete Uniform $(1, \theta)$ distribution (or DU(1, $\theta$) distribution) if

$$f(x|\theta) = P(X = x|\theta) = \frac{1}{\theta}, \quad x = 1, 2, \ldots, \theta \quad (2.1.1)$$

where $\theta$ is a specified integer. Note that the parameter space $\Theta = \{1, 2, 3, \ldots\}$, the set of all positive integers.

Using the identities

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2} \quad \text{and} \quad \sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}$$

we can find the mean and variance of $X$ as

$$\mu_X = E(X) = \sum_{x=1}^{\theta} x \frac{1}{\theta} = \frac{(\theta + 1)}{2}$$

and

$$E(X^2) = \sum_{x=1}^{\theta} x^2 \frac{1}{\theta} = \frac{(\theta + 1)(2\theta + 1)}{6} \quad (2.1.2)$$

which gives

$$\sigma_X^2 = Var(X) = \frac{(\theta + 1)(\theta - 1)}{12} \quad (2.1.3)$$

The mgf of DU(1, $\theta$) is

$$\phi_X(t) = \left( \sum_{x=1}^{\theta} e^{xt} \right) / \theta \quad (2.1.4)$$

The above distribution can be generalized to Discrete Uniform $(\theta_1, \theta_2)$ distribution (or DU($\theta_1, \theta_2$) distribution) with pmf

$$f(x|\theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1 + 1)}, \quad x = \theta_1, \theta_1 + 1, \ldots, \theta_2$$

and

$$\Theta = \{(\theta_1, \theta_2) | \theta_1 \leq \theta_2, \text{ and both } \theta_1 \text{ and } \theta_2 \text{ are integers}\}$$

Moments of DU($\theta_1, \theta_2$) can be derived easily and hence omitted.
2.1 Discrete Distributions

Binomial Distribution

The binomial distribution is based on the idea of a Bernoulli trial. A Bernoulli trial is an experiment with only two possible outcomes, say $S$ (for ‘success’) and $F$ (for ‘failure’). Let $\theta = P(S)$ be the probability of ‘success’ in a Bernoulli trial. Define a random variable $X$ as $X = \#$ successes in a Bernoulli trial. The range of $X$ is $R_X = \{0, 1\}$ and

$$X = \begin{cases} 1 & \text{with probability } \theta \\ 0 & \text{with probability } (1 - \theta) \end{cases} \quad (2.1.5)$$

where $0 \leq \theta \leq 1$. The probability distribution of $X$ given by the pmf

$$f(x|\theta) = \theta^x(1 - \theta)^{1-x}, \quad x = 0, 1 \quad (2.1.6)$$

where $\theta \in \Theta = [0, 1]$, is said to be a Bernoulli($\theta$) distribution.

Now consider a larger experiment which consists of a fixed number of independent and identical Bernoulli trials with $\theta$ being probability of success in a single trial. Let $n = \text{the total number (fixed) of Bernoulli trials}$. Define the random variable $X$ as $X = \#$ successes in those $n$ trials. The range of $X$ is then $R_X = \{0, 1, 2, \ldots, n\}$, and the probability distribution of $X$ is

$$f(x|\theta) = P(X = x|\theta) = \binom{n}{x} \theta^x(1 - \theta)^{n-x}, \quad x = 0, 1, \ldots, n \quad (2.1.7)$$

where $\theta \in \Theta = [0, 1]$. A random variable having the pmf (2.1.7) is said to have a Binomial($n, \theta$) or B($n, \theta$) distribution. Thus, a Bernoulli($\theta$) distribution is a B(1, $\theta$) distribution.

Note that $\theta = 0$ implies $P(X = 0) = 1$ and $\theta = 1$ implies $P(X = 1) = 1$, and both happen rarely or only in extreme cases. Therefore, we ignore those cases by restricting $\Theta = (0, 1)$ (open interval).

It can be shown that for a B($n, \theta$) distribution,

$$E(X) = n\theta, \quad Var(X) = n\theta(1 - \theta) \quad (2.1.8)$$

and the mgf of $X$ is

$$\phi_X(t) = (\theta e^t + (1 - \theta))^n \quad (2.1.9)$$

A B($n, \theta$) random variable $X$ (i.e., the random variable $X$ having a B($n, \theta$) distribution) can be characterized as a sum of $n$ independent and
identically distributed (iid) random variables each having a Bernoulli(θ) distribution. In other words, let $X_1, X_2, \ldots, X_n$ be iid Bernoulli(θ) (or $B(1, \theta)$) random variables, then $X = X_1 + X_2 + \ldots + X_n$ has a $B(n, \theta)$ distribution.

**Geometric Distribution**

Suppose we have a sequence of independent and identical Bernoulli(θ) trials. We define a random variable $X$ which counts the number of trials to get the very first success. The range of $X$ is then $R_X = \{1, 2, 3, \ldots\}$. The probability distribution of $X$ given by the pmf

$$f(x|\theta) = P(X = x|\theta) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, 3, \ldots \quad (2.1.10)$$

where $\theta \in \Theta = (0, 1)$, is called a Geometric (G(θ)) distribution with parameter θ. A generalization of G(θ) distribution is presented below.

**Negative Binomial Distribution**

Suppose we count the number of Bernoulli trials required to obtain a fixed number of successes (say, $r$ successes). Let the random variable $X$ denote the trial number at which the $r$th success occurs ($r$ is a fixed positive integer). The range of $X$ is $R_X = \{r, r+1, \ldots\}$ with pmf

$$f(x|r,\theta) = P(X = x|r,\theta) = \binom{x - 1}{r - 1} \theta^r (1 - \theta)^{x-r} \quad (2.1.11)$$

where $\theta \in \Theta = (0, 1)$, and $X$ is said to have a Negative Binomial($r, \theta$) (or NB($r, \theta$)) distribution. For NB($r, \theta$) distribution, we have

$$E(X) = \frac{r(1 - \theta)}{\theta}, \quad Var(X) = \frac{r(1 - \theta)}{\theta^2} \quad (2.1.12)$$

$$\phi_X(t) = \left( \frac{\theta}{1 - (1 - \theta)e^t} \right)^r, \quad t < -\ln(1-\theta) \quad (2.1.13)$$

It can be seen that NB(1, θ) = G(θ). Hence, one can find the $E(X)$, $Var(X)$, or the mgf of G(θ) distribution from (2.1.13) by taking $r = 1$.

**Poisson Distribution**

A random variable $X$ with range $R_X = \{0, 1, 2, \ldots\}$ is said to have a Poisson distribution with parameter $\lambda$ (Poi($\lambda$)) if the pmf of $X$ is

$$f(x|\lambda) = P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \quad (2.1.14)$$
where $\lambda \in \Lambda = (0, \infty)$. The parameter $\lambda$ of Poi($\lambda$) distribution is often called the intensity parameter. Mean, variance, and the mgf of Poi($\lambda$) are given by

\[ E(X) = \text{Var}(X) = \lambda \quad \text{and} \quad \phi_X(t) = e^{\lambda(e^t-1)} \quad (2.1.15) \]

Poisson distribution may be looked upon as a limiting form of a binomial distribution. Suppose that in a binomial distribution, the number of trials ($n$) tends to infinity and the probability of success ($\theta$) in a single trial tends to 0 in such a way that

\[ n\theta = (\# \text{ trials}) \times (\text{probability of success}) \rightarrow \lambda \quad \text{(finite)} \]

then

\[ \lim_{n \to \infty} \binom{n}{x} \theta^x (1-\theta)^{n-x} = \frac{e^{-\lambda \theta^x}}{x!}, \quad x = 0, 1, 2, \ldots \]

**Hypergeometric Distribution**

The probability distribution of the hypergeometric random variable $X$, the number of successes in a random sample of size $n$ selected from $N$ items of which $K$ are labeled ‘success’ and $N-K$ labeled ‘failure’, is

\[ f(x|N,K,n) = \binom{K}{x} \binom{N-K}{n-x} / \binom{N}{n}, \quad x = 0, 1, 2, \ldots, n \quad (2.1.16) \]

Note that there is, implicit in (2.1.16), an additional assumption on the range of $X$. Binomial coefficients of the form $\binom{a}{b}$ are meaningful only if $0 \leq b \leq a$, and so the range of $X$ is additionally restricted by the pair of inequalities

\[ 0 \leq x \leq K \quad \text{and} \quad 0 \leq n-x \leq N-K \]

which can be combined as

\[ n-(N-K) \leq x \leq K \quad (2.1.17) \]

In many cases, $n$ is a small value compared to $N$ and $K$, and hence, the range $0 \leq x \leq n$ will be contained in the above range (2.1.17). We, therefore, state the formal range of the hypergeometric random variable $X$ as $R_X = \{0, 1, 2, \ldots, n|n \leq N-K \text{ and } n \leq K\}$.

Deriving moments of a hypergeometric distribution is a bit tedious. However, its mean and variance are given as

\[ E(X) = \frac{nK}{N} \quad \text{and} \quad \text{Var}(X) = \frac{nK(N-n)(N-K)}{N^2(N-1)} \quad (2.1.18) \]
Chapter 2  Some Common Probability Distributions

Unfortunately, no closed form expression of the moment generating function of a hypergeometric distribution is available.

\section{Continuous Distributions}

\subsection{Continuous Uniform Distribution}

The continuous uniform distribution over the range \([\theta_1, \theta_2]\) is defined by spreading mass (or weight) uniformly over the interval \([\theta_1, \theta_2]\) and denoted by \(CU[\theta_1, \theta_2]\). A random variable \(X\) with the pdf given by

\[ f(x|\theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \text{if } x \in [\theta_1, \theta_2] \\ 0 & \text{otherwise} \end{cases} \]  

(2.2.1)

is said to have \(CU[\theta_1, \theta_2]\) distribution. It is easy to verify that

\[ E(X) = \frac{\theta_1 + \theta_2}{2}, \quad Var(X) = \frac{(\theta_2 - \theta_1)^2}{12} \]

\[ \text{and } \phi_X(t) = \frac{(e^{\theta_2 t} - e^{\theta_1 t})}{t(\theta_2 - \theta_1)} \]

(2.2.2)

\subsection{Normal Distribution}

One of the most important continuous probability distributions in the entire field of statistics is the normal distribution. The graph of a normal pdf, called the \textbf{normal curve}, is a bell shaped symmetric curve, and it describes (probability-wise) many real life data sets. The normal distribution is often referred to as the Gaussian distribution in honor of Karl Friedrich Gauss who derived the equation of a normal pdf.

The normal distribution has two parameters, usually denoted by \(\mu\) and \(\sigma^2\), which are its mean and variance. A random variable \(X\) is said to have a normal distribution with mean \(\mu\) and variance \(\sigma^2\) which is denoted by \(N(\mu, \sigma^2)\), if its pdf is given by

\[ f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \]  

(2.2.3)

If \(X\) has \(N(\mu, \sigma^2)\) distribution, then the random variable \(Z = (X - \mu)/\sigma\) has a \(N(0, 1)\) distribution, known as the \textbf{standard normal distribution}. The relationship between the standard normal \(cdf\)
and any other \( N(\mu, \sigma^2) \) cdf can be established as
\[
P(X \leq x) = P\left(\frac{(X - \mu)}{\sigma} \leq \frac{(x - \mu)}{\sigma}\right) = P(Z \leq z), \; z = \frac{(x - \mu)}{\sigma}
\]
\[
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} \, du \quad (2.2.4)
\]

Therefore, all normal probabilities can be computed in terms of the standard normal distribution. Throughout this book we will denote the pdf and cdf of the standard normal distribution by \( \varphi(\cdot) \) and \( \Phi(\cdot) \) respectively, i.e.,
\[
\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{and} \quad \Phi(z) = \int_{-\infty}^{z} \varphi(u) \, du \quad (2.2.5)
\]

Figure 2.2.1: The pdfs of \( N(\mu, 1) \) Distribution, \( \mu = 0, 5 \).
The following simple properties of any $N(\mu, \sigma^2)$ distribution are observed:

(a) The two parameters $\mu$ and $\sigma^2$ which uniquely determine a $N(\mu, \sigma^2)$
distribution are indeed mean and variance of the distribution, i.e.,
$E(X) = \mu$ and $Var(X) = \sigma^2$.

(b) The mode, which is the point on the horizontal axis where the
normal curve (pdf) is a maximum, occurs at $x = \mu$.

(c) The normal curve is symmetric about a vertical axis through the
mean $\mu$.

(d) The normal curve has its points of inflexion at $x = \mu \pm \sigma$. In other
words, the normal curve is concave if $\mu - \sigma < x < \mu + \sigma$, and is
convex otherwise.

(e) The normal curve approaches the horizontal axis asymptotically as
we proceed in either direction away from the mean $\mu$.

(f) The mgf of $N(\mu, \sigma^2)$ distribution is $\phi_X(t) = e^{\mu t + \sigma^2 t^2 / 2}$ for
$-\infty < t < \infty$.

(g) Probability coverage under the standard normal curve is given in
Table A.1 (Appendix-A). As a result, one can easily find the prob-
ability content within 1, 2, or 3 standard deviations of mean for
$N(\mu, \sigma^2)$ as

\[
P(|X - \mu| \leq \sigma) = 0.6826 \\
P(|X - \mu| \leq 2\sigma) = 0.9544 \\
P(|X - \mu| \leq 3\sigma) = 0.9974
\] (2.2.6)

Gamma Distribution

The gamma family of distributions is a flexible family of distributions on
$[0, \infty)$. If $\alpha$ is a positive real value, then the integral

\[
\Gamma(\alpha) = \int_{0}^{\infty} t^{\alpha-1} e^{-t} dt
\] (2.2.7)

is called the gamma function evaluated at $\alpha$. The value of $\Gamma(\alpha)$ is finite
for $0 < \alpha < \infty$. It can be shown through integration by parts that

\[
\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha > 0
\] (2.2.8)
and for any positive integer $k$,

$$\Gamma(k+1) = k! \quad (2.2.9)$$

A random variable $X$ is said to have a (two parameter) gamma distribution with parameters $\alpha$ and $\beta$, denoted by $G(\alpha, \beta)$, if the pdf of $X$ is given by

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty \quad (2.2.10)$$

where $\alpha > 0$ and $\beta > 0$. The parameter $\alpha$, which influences the peakedness of the distribution, is known as the shape parameter, and $\beta$ is called the scale parameter.

The mean, variance, and the mgf of a $G(\alpha, \beta)$ are respectively

$$E(X) = \alpha \beta, \quad Var(X) = \alpha \beta^2, \text{ and } \phi_X(t) = (1 - \beta t)^{-\alpha} \text{ for } t < 1/\beta \quad (2.2.11)$$

Two important special cases of the gamma distribution are worth mentioning.

If we set $\alpha = k/2$ where $k$ is an integer, and $\beta = 2$, then the pdf (2.2.10) becomes

$$f(x|k) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2}, \quad 0 < x < \infty \quad (2.2.12)$$

and this is known as **chi-square distribution** with $k$ degrees of freedom ($df$), denoted by $\chi^2_k$.

The chi-square distribution plays an important role in statistical inferences, especially when a random sample is taken from a normal population. For example, if $Z$ is a $N(0,1)$ random variable, then $W = Z^2$ follows a $\chi^2_1$ distribution. Also, if $W_1, W_2, \ldots, W_k$ are independent and identically distributed as $\chi^2_1$, then $W = \sum_{i=1}^k W_i$ follows $\chi^2_k$ distribution.

The other important special case of the gamma distribution is obtained if we set $\alpha = 1$. This gives the special gamma pdf

$$f(x|\beta) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0 \quad (2.2.13)$$

known as an **exponential distribution** with scale parameter $\beta$. Exponential distribution has tremendous applications in modeling lifetimes.
Figure 2.2.2: The pdfs of $G(\alpha, 1)$ Distribution, $\alpha = 1, 5$. 

due to some interesting properties. Both Chi-square and Exponential distributions will be discussed later (Chapter 6 and Chapter 7) in greater detail in this book.

Lognormal Distribution

A random variable $X$ is said to have a lognormal distribution if $(lnX)$ is normally distributed. Using the fact that $(lnX)$ is a $N(\mu, \sigma^2)$ distribution, one can derive the pdf of $X$ as (denoted by $LN(\mu, \sigma^2)$), and

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}x^{-1}e^{-((lnx-\mu)^2/(2\sigma^2))}, \quad 0 < x < \infty \quad (2.2.14)$$
where $-\infty < \mu < \infty$ and $\sigma > 0$. Using the normal mgf, we can find the mean and variance of the LN($\mu, \sigma^2$) distribution as

$$E(X) = E(e^{\ln X}) = e^{\mu + \sigma^2/2} \text{ and } Var(X) = e^{2\mu + \sigma^2} \left(e^{\sigma^2} - 1\right)$$

Interestingly, LN($\mu, \sigma^2$) is one such distribution for which the mgf does not exist, yet all the moments exist. In fact,

$$E(X^k) = e^{k\mu + k^2\sigma^2/2} \text{ for any } k \geq 1$$

(2.2.16)

The lognormal distribution is very similar in appearance to the gamma distribution as shown in Figure 2.2.2 and Figure 2.2.3.

Figure 2.2.3: The pdfs of LN($\mu, 1$), $\mu = 0, 3$. 

© 2006 by Taylor & Francis Group, LLC
Logistic Distribution

A random variable $X$ is said to have a logistic distribution with parameters $\mu$ and $\sigma$ (denoted by $L(\mu, \sigma)$), provided the pdf of $X$ is

$$f(x|\mu, \sigma) = \frac{1}{\sigma} \frac{e^{-(x-\mu)/\sigma}}{(1 + e^{-(x-\mu)/\sigma})^2}, \quad -\infty < x < \infty$$  \hspace{1cm} (2.2.17)

where $-\infty < \mu < \infty$ and $\sigma > 0$.

The logistic distribution is appealing because of its simple looking cdf given as

$$F(x|\mu, \sigma) = P(X \leq x) = \left(1 + e^{-(x-\mu)/\sigma}\right)^{-1}$$  \hspace{1cm} (2.2.18)

and it has numerous applications in clinical trials. Mean, variance, and mgf of $L(\mu, \sigma)$ are respectively

$$E(X) = \mu, \quad Var(X) = \frac{\pi^2}{3} \sigma^2$$

and

$$\phi_X(t) = \Gamma(1 - \sigma t) \Gamma(1 + \sigma t) e^{\mu t}, \quad |t| < \frac{1}{\sigma}$$  \hspace{1cm} (2.2.19)

Weibull Distribution

A random variable $X$ has a Weibull distribution with parameters $\alpha$ and $\beta$, denoted $W(\alpha, \beta)$, if its pdf is given by

$$f(x|\alpha, \beta) = \left(\frac{\alpha}{\beta}\right) x^{\alpha-1} e^{-(x/\beta)^\alpha}, \quad 0 < x < \infty$$  \hspace{1cm} (2.2.20)

where $\alpha > 0$ and $\beta > 0$. Note that if we set $\alpha = 1$, then the pdf in (2.2.20) becomes an exponential pdf in (2.2.13).

One can easily derive the moments of $W(\alpha, \beta)$ distribution as

$$E(X^k) = \beta^k \Gamma\left(1 + \frac{k}{\alpha}\right)$$  \hspace{1cm} (2.2.21)

from which mean and variance can be obtained as

$$E(X) = \beta \Gamma\left(1 + \frac{1}{\alpha}\right)$$

and

$$Var(X) = \beta^2 \left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right)$$  \hspace{1cm} (2.2.22)
2.2 Continuous Distributions

The mgf of $W(\alpha, \beta)$ exists only for $\alpha \geq 1$ and has a very complicated form (not very useful).

The Weibull distribution is often applied to reliability and life testing problems such as the time to failure or life span of a component, measured from some specified time until it fails. For more on this see Chapter 8.

Pareto Distribution

A random variable $X$ is said to have a Pareto distribution with parameters $\alpha$ and $\beta$, denoted by $P(\alpha, \beta)$, if its pdf is given by

$$f(x|\alpha, \beta) = \beta \alpha^\beta x^{-(\beta+1)}, \quad \alpha < x < \infty \quad (2.2.23)$$

where $\alpha > 0$ and $\beta > 0$. 

Figure 2.2.4: The pdfs of $L(\mu, 1), \mu = 0, 5$. 

© 2006 by Taylor & Francis Group, LLC
Chapter 2  Some Common Probability Distributions

The mean and variance of P(\( \alpha, \beta \)) are

\[
E(X) = \frac{\alpha \beta}{(\beta - 1)}, \quad \beta > 1
\]  \hspace{1cm} (2.2.24)

and

\[
Var(X) = \frac{\alpha^2 \beta}{(\beta - 1)^2(\beta - 2)}, \quad \beta > 2
\]  \hspace{1cm} (2.2.25)

If \( \beta \leq 2 \), then \( Var(X) \) does not exist, and \( E(X) \) does not exist for \( \beta \leq 1 \). Any \( k \)th moment of \( X \) will exist subject to the restriction on \( \beta > k \). The mgf of P(\( \alpha, \beta \)) distribution does not exist.

**Student’s \( t \)-Distribution**

A distribution which is very similar in shape to the normal distribution is the Student’s \( t \)-distribution with \( \nu \) degrees of freedom, in short called
The special case, when $\nu = 1$ is called the Cauchy distribution and
Chapter 2  Some Common Probability Distributions

the corresponding pdf is

\[ f(x|\theta) = \frac{1}{\pi(1 + (x - \theta)^2)}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty \] (2.2.28)

The parameter \( \theta \) in (2.2.26) measures the center of the distribution and is the median of \( t_\nu(\theta) \) distribution. Note that the mean of the Cauchy distribution (2.2.28) does not exist. Also, from the pdf (2.2.26) of \( t_\nu(\theta) \), as \( \nu \to \infty \), \( t_\nu(\theta) \) tends to the \( N(\theta, 1) \) distribution.

Snedecor’s F-Distribution

A random variable \( X \) is said to have a Snedecor’s F-distribution (or ‘F-distribution’ in short) with ordered degrees of freedom (dfs) \( \nu_1 \) and
2.2 Continuous Distributions

\[ \frac{\nu_2}{\nu_2 - 2}, \text{ provided its pdf is given as} \]

\[ f(x|\nu_1, \nu_2) = \frac{\Gamma((\nu_1 + \nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1/\nu_2}{x}\right)^{\nu_1/2} x^{(\nu_2 - 2)/2} \left(1 + \left(\frac{\nu_1/\nu_2}{x}\right)^{-(\nu_1 + \nu_2)/2}\right) \]

\[ x > 0 \quad (2.2.29) \]

where the dfs \( \nu_1 \) and \( \nu_2 \) usually take integer values. The above distribution is denoted by \( F_{\nu_1, \nu_2} \). Mean and variance of the \( F_{\nu_1, \nu_2} \) distribution exist subject to restrictions on \( \nu_1 \) and \( \nu_2 \), and they are

\[ E(X) = \frac{\nu_2}{\nu_2 - 2}, \quad \nu_2 > 2 \]

and

\[ \text{Var}(X) = 2 \left(\frac{\nu_2}{\nu_2 - 2}\right)^2 \frac{(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)}, \quad \nu_2 > 4 \quad (2.2.30) \]

The mgf of \( F_{\nu_1, \nu_2} \) does not exist, however, the \( k \)th moment is given by

\[ E(X^k) = \frac{\Gamma((\nu_1 + 2k)/2)\Gamma((\nu_2 - 2k)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_2}{\nu_1}\right)^k, \quad \nu_2 > 2k \quad (2.2.31) \]

Suppose \( X_1 \) and \( X_2 \) are two independent random variables having \( \chi^2_{\nu_1} \) and \( \chi^2_{\nu_2} \) distributions, respectively. Then the distribution of \( X = (X_1/\nu_1)/(X_2/\nu_2) \) is \( F_{\nu_1, \nu_2} \). Also, if a random variable \( Y \) follows \( t_{\nu}(0) \) distribution, then \( X = Y^2 \) follows \( F_{1, \nu} \) distribution.

**Beta Distribution**

A random variable \( X \) with range (or support) \( R_X = [0, 1] \) is said to have a Beta(\( \alpha, \beta \)) distribution, provided the pdf is given by

\[ f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1}, \quad 0 \leq x \leq 1 \quad (2.2.32) \]

and \( \alpha > 0, \beta > 0 \). The constant \( B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta) \). Mean, variance, and mgf of Beta(\( \alpha, \beta \)) distribution are

\[ E(X) = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (2.2.33) \]

and

\[ \phi_X(t) = 1 + \sum_{i=1}^{\infty} \left(\prod_{j=0}^{i-1} \frac{\alpha + j}{\alpha + \beta + j}\right) \frac{t^i}{i!} \quad (2.2.34) \]

For the special case \( \alpha = \beta = 1 \), the Beta(\( \alpha, \beta \)) distribution reduces to a C[0,1] distribution in (2.2.1). Also, if \( X_1 \) and \( X_2 \) are independent random variables with \( G(\alpha_1, \beta) \) and \( G(\alpha_2, \beta) \) distributions respectively, then the distribution of \( Y = X_1/(X_1 + X_2) \) is \( B(\alpha_1, \alpha_2) \).
2.3 Some Limit Theorems

Definition 2.3.1 (Convergence in Distribution) A sequence \( \{F_k(x)\} \) of distribution functions (or cdfs) converges in distribution (or weakly) to a distribution function \( F(x) \), if
\[
\lim_{k \to \infty} F_k(x) = F(x)
\]
for all continuity points \( x \) of \( F(x) \).

Let \( F_k(x) \) and \( F(x) \) be the distribution functions of \( X_k \) \((k \geq 1)\) and \( X \), respectively. Then, the sequence \( \{X_k\} \) of random variables converges in distribution (or weakly) to \( X \) if \( \{F_k(x)\} \) converges in distribution to \( F(x) \). This is denoted as
\[
X_k \overset{d}{\to} X
\]

Theorem 2.3.1 Let \( X_k \overset{d}{\to} X \) and \( g \) be a continuous function, then \( g(X_k) \overset{d}{\to} g(X) \).

Definition 2.3.2 (Convergence in Probability) Let \( \{X_k\} \) be a sequence of random variables, then \( \{X_k\} \) converges in probability to a random variable \( X \) if for every \( \epsilon > 0 \),
\[
P(|X_k - X| > \epsilon) \to 0 \quad \text{as} \ k \to \infty
\]
It is denoted as
\[
X_k \overset{p}{\to} X
\]

Definition 2.3.3 (Almost Sure Convergence) A sequence of random variables \( \{X_k\} \) is said to converge almost surely to a random variable \( X \) if
\[
P \left( \lim_{k \to \infty} X_k = X \right) = 1
\]
and this is denoted as
\[
X_k \overset{a.s.}{\to} X
\]
or \( X_k \to X \) with probability 1.

The next theorem establishes an order among the above three types of convergences.
Theorem 2.3.2

\[ X_k \xrightarrow{a.s.} X \implies X_k \xrightarrow{p} X \implies X_k \xrightarrow{d} X \]

The following theorem provides a result in which convergence in distribution can lead to convergence in probability.

Theorem 2.3.3 If \( X_k \xrightarrow{d} c \) (constant), then \( X_k \xrightarrow{p} c \).

In the following we will see applications of the above three types of convergences.

Theorem 2.3.4 (Central Limit Theorem or CLT) Let \( X_1, X_2, \ldots \) be a sequence of iid random variables with a finite mean \( \mu \) and a finite variance \( \sigma^2 > 0 \). Define

\[ Y_k = \frac{1}{k} \sum_{i=1}^{k} X_i \]  
\[ Z_k = \frac{\sqrt{k}(Y_k - \mu)}{\sigma} \]

Then

\[ Z_k \xrightarrow{d} Z \]

where \( Z \) is a standard normal random variable (i.e., \( Z \) follows \( N(0,1) \) distribution).

Remark 2.3.1 Since any real value is a continuity point of the standard normal cdf, the central limit theorem (CLT) says that for any \( c \in \mathbb{R} \),

\[ P(Z_k \leq c) \longrightarrow \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du, \quad \text{as } k \to \infty \]

Generally speaking, \( Z_k \) is approximately a \( N(0,1) \) random variable for a sufficiently large \( k \); i.e., the average of sufficiently large number of iid random variables is approximately normally distributed.

Theorem 2.3.5 (Weak Law of Large Numbers or WLLN) Let \( X_1, X_2, \ldots \) be a sequence of iid random variables with a finite mean \( \mu \). Define

\[ Y_k = \frac{1}{k} \sum_{i=1}^{k} X_i \]

then

\[ Y_k \xrightarrow{p} \mu \]
Theorem 2.3.6 (Strong Law of Large Numbers or SLLN) Let \(X_1, X_2, \ldots\) be a sequence of iid random variables with a finite mean \(\mu\). Define
\[
Y_k = \frac{1}{k} \sum_{i=1}^{k} X_i
\]
then
\[
Y_k \xrightarrow{a.s.} \mu
\]

The above limit theorems help us study the behavior of certain sample quantities when the sample size approaches infinity. Although the concept of an infinitely large sample size is a hypothetical one, it can often provide us with some useful approximations for the finite sample case. Consider the following applications.

Suppose \(X_1, X_2, \ldots, X_n\) be iid with a common cdf \(F(x)\) which is assumed to be strictly monotonic. For \(0 < c < 1\), let \(\xi_c\) denotes the 100\(c\)th population percentile and let \(k\) be an integer such that \((k/n) \to c\) as \(n \to \infty\). Then the asymptotic distribution of \(X_{(k)}\) is \(N(\xi_c, \sigma_k^2)\) (i.e., as \(n \to \infty\), \(X_{(k)} \xrightarrow{d} N(\xi_c, \sigma_k^2)\)) where \(\sigma_k^2 = c(1-c)/\{n(F'(\xi_c))^2}\).

The above result says that in a large sample, the \(k\)th order statistic can be used to estimate the population percentile \(\xi_c\) provided \((k/n) \approx c\). Further, the variance \(\sigma_k^2\) can tell us how fast the convergence is as \(n\) grows.

Theorem 2.3.7 (Slutsky’s Theorem) If \(X_k \xrightarrow{d} X\) and \(Y_k \xrightarrow{p} c\) (where \(c\) is a constant), then:

(i) \(X_k Y_k \xrightarrow{d} cX\)

(ii) \(X_k + Y_k \xrightarrow{d} (c + X)\)
Chapter 3

Concepts of Statistical Inference

3.1 Introduction

In any statistical investigation, we collect data which, in many cases, can be described by observable random variables $X_1, X_2, \ldots, X_n$ (all or some of which may as well be multi-dimensional). We then assume a suitable probability distribution, called a probability model or simply a model, for the data $X = (X_1, X_2, \ldots, X_n)$. The choice of the model is dictated partly by the nature of the experiment which gives rise to the data and partly by our past experience with a similar data. Often, mathematical simplicity plays a significant role in choosing the model. In general, the model takes the form of a specification of the joint probability distribution of the observable random variables $X_1, X_2, \ldots, X_n$. According to the model, the joint distribution function $F_X$ is supposed to be some unspecified member of a suitable class $\mathcal{F}_X$ of distributions. For the sake of simplicity, we assume that $X_i$'s are single dimensional if not mentioned otherwise.

In many situations, we assume that the random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) random variables having a common but unspecified cumulative distribution function (cdf) $F$. Thus, our model states that $F_X$ is a member of some class $\mathcal{F}_X$ of the
where $F$ belongs to some class $\mathcal{F}$ of distribution functions.

As an example, imagine the experiment where $n$ individuals are selected at random (without replacement) from a very large population, and let the $i^{th}$ random variable $X_i$ represent the height of the $i^{th}$ individual, $i = 1, 2, \ldots, n$. Since the individuals are selected without replacement, the probability distribution of $X_i$, the height of the $i^{th}$ individual, depends on the population composition after excluding the previous $(i-1)$ individuals. But the overall population size (say, $N$) is so large compared to the overall sample size $n$ that inclusion or exclusion of a few individuals does not make much difference, and as a result, the probability distribution of $X_i$ is virtually the same as those of $X_1, X_2, \ldots, X_{i-1}$, $i = 2, 3, \ldots, n$, and they all are independent. $F$ indicates the cdf of a randomly selected individual’s height. Because the randomly selected individual can represent any individual in the population, $F$ therefore is the cdf over the range $R_X = \text{collection of ‘height’ values of all individuals in the population, which is a subset of the positive side of } \mathbb{R}$.

Besides making the assumption that $X_1, X_2, \ldots, X_n$ are iid random variables with a common distribution $F$, we may further assume that the structure of $F$ is a $\text{N}(\mu, \sigma^2)$ distribution with $\mu$ and $\sigma^2$ being left unspecified. Thus, $F \in \mathcal{F}$, where

$$\mathcal{F} = \{ \text{N}(\mu, \sigma^2) \mid \mu \in \mathbb{R} \text{ and } \sigma^2 > 0 \}$$  \hspace{1cm} (3.1.2)

With the assumption (3.1.2), the joint distribution function of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ given in (3.1.1) becomes

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \int_{-\infty}^{x_i} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2\sigma^2}(u_i - \mu)^2\right) \, du_i$$  \hspace{1cm} (3.1.3)

where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

In carrying out the statistical investigation, we then take as our goal the task of specifying $F$ more completely than is done by the model (say,
3.1 Introduction

This task is achieved through the observable random variables \( X_1, X_2, \ldots, X_n \) with realizations \( X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \). The observations \( (x_1, x_2, \ldots, x_n) \) are used to make an educated guess about the distribution \( F_X \) (the model) which is partly unknown since \( \mu \) and \( \sigma^2 \) are unknown in (3.1.2).

The process of making an educated guess about the probability distribution of the outcome of an experiment is called statistical inference. The inference is based on the data \( (X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \) obtained from one or more experiment(s), and the theory of probability enters into the process in three ways. First, the model used to explain the occurrence of the data is probabilistic. Second, certain probabilistic principles are followed in making the inference. Third, the precision or the dependability of our inference is also evaluated in probabilistic terms.

The problem of statistical inference generally takes one of the two following forms:

(a) estimation

(b) hypothesis testing

(a) Sometimes we are interested in a particular characteristic \( \theta \) of the probability model \( F_X \) of \( X = (X_1, X_2, \ldots, X_n) \) where \( X_1, X_2, \ldots, X_n \) are iid with the common distribution \( F \) and \( \theta \) may be the mean or variance of the distribution \( F \). In order to make a conjecture about this \( \theta \), we use some function of the data \( X = (X_1, X_2, \ldots, X_n) \), say \( \hat{\theta} = \hat{\theta}(X) \). More precisely, if we observe \( X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \), then we put forward the corresponding value of \( \hat{\theta}(X) \), i.e., \( \hat{\theta} = \hat{\theta}(x) = \hat{\theta}(x_1, x_2, \ldots, x_n) \) as a ‘likely’ value of the characteristic \( \theta \). This \( \hat{\theta} \) is called a ‘point estimate’ of \( \theta \), and the functional form \( \hat{\theta}(X) \) is called a point estimator of \( \theta \). Therefore, in point estimation, the unknown value \( \theta \) is estimated by a point \( \hat{\theta} \) or a single value \( \hat{\theta} \).

Instead of a single value \( \hat{\theta} \) for \( \theta \), one might prefer a collection of values in the form of an interval, say \( C(X) = (\tilde{\theta}_L(X), \tilde{\theta}_U(X)) \), such that the actual value \( \theta \) is likely to belong to this set \( C(X) \). Such an estimation is called set or interval estimation of \( \theta \).

(b) On some occasions, we start with some tentative notion about the characteristic \( \theta \) of the distribution which we are interested in. This idea about \( \theta \), called a hypothesis, may come from our past experience in
dealing with similar problems, or may be suggested by some authority (e.g., a leading anthropologist making a guess about the average height of male members of a lost tribe, or a manufacturer making a claim about the reliability of an improved product released in the market). We may then like to know how tenable or valid the idea about \( \theta \) is in the light of the observations \( X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \). The inference problem is thus regarded as testing a hypothesis about the unknown characteristic \( \theta \) of the model. Note that the model used and the hypothesis being tested are based upon the probability distribution of \( X = (X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \). However, while the hypothesis is an assumption the validity of which is questionable, the validity of the model is not questioned, but is taken for granted.

In the development of statistical methods, the techniques of inference that were the first to appear were those which involved a number of assumptions about the distribution of \( X = (X_1, X_2, \ldots, X_n) \). In most cases, it is assumed that \( X_1, X_2, \ldots, X_n \) are iid random variables, and in any case it would be assumed that the joint distribution has a particular parametric form with some or all of the model parameters being unknown. Statistical inference in those cases would relate solely to the value(s) of the unknown parameter(s). This is called parametric inference. On the other hand, statistical inference when no parametric form of the distribution is assumed is called nonparametric inference.

### 3.2 Sufficiency and Completeness

As mentioned earlier, in parametric statistics, the model, i.e., the probability distribution \( F_X \) of the data has a known structural form in which one or more parameters are unknown. For convenience, let us assume that the parametric form \( F_X \) depends on the parameter \( \theta \), and for simplicity, we also assume that \( \theta \) is single dimensional even though multidimensional \( \theta \) poses no ambiguity. The parameter \( \theta \) is indeed a characteristic of the probability distribution \( F_X \). To show the dependence of \( F_X \) on \( \theta \), we modify our notation slightly and the distribution of the data \( X = (X_1, X_2, \ldots, X_n) \) will henceforth be noted by \( F_{X, \theta} \) or by \( F_{\theta} \).

Since values of \( \theta \) over a suitable range \( \Theta \) constitute the family of possible probability distribution for our data \( X \), \( \theta \) is thus an index indicating a suitable class of distributions \( \mathcal{F}_X = \{ F_{X, \theta} \mid \theta \in \Theta \} \). In a statistical in-
vestigation, the aim of the investigator is to know which member of the above class $F_X$ would be most appropriate to his/her study, i.e., to make a decision about the value of the unknown parameter $\theta$.

The only information that helps the investigator in making a decision is supplied by the observed values $x_1, x_2, \ldots, x_n$ of the random variables $X_1, X_2, \ldots, X_n$. However, in most of cases, the observations would be too numerous and too complicated a set of numbers to be directly dealt with, and so a condensation or summarization of data would be most desirable. It is this motivation that leads one to the principle of sufficiency or the notion of sufficient statistics.

A sufficient statistic for a parameter $\theta$ is a statistic (or a function of the data) that retains all the essential information about $\theta$. Any additional information in the data, besides the value of the sufficient statistic, is not relevant for $\theta$.

**Definition 3.2.1** A $k$-dimensional statistic $T(X) = (T_1(X), T_2(X), \ldots, T_k(X))$ is called a sufficient statistic for $\theta$ (or rather for the family of distributions $F_X = \{F_{X,\theta} | \theta \in \Theta\}$) if the conditional probability distribution of $X = (X_1, X_2, \ldots, X_n)$, given the value of $T$, does not depend on $\theta$.

It should be noted that in the above definition, the components $T_1, T_2, \ldots, T_k$ of $T$ need to be functionally unrelated, otherwise, one or more components can be dropped from $T$ without compromising any relevant information. Also, in most of the subsequent examples, the parameter $\theta$ is single dimensional and so will be a sufficient statistic. In such a case, we use the notation, $T(X) = T$, to denote a sufficient statistic.

**Example 3.2.1:** Suppose our observations $X_1, X_2, \ldots, X_n$ are independent, and $X_i$ follows a $B(m_i, \theta)$ distribution, $i = 1, 2, \ldots, n; 0 < \theta < 1$.

Consider the statistic $T = \sum_{i=1}^{n} X_i$ which has a $B(m, \theta)$ distribution with $m = m_1 + m_2 + \ldots + m_n$. Now, the pmf of $T$ is

$$g(t|\theta) = \begin{cases} \binom{m}{t} \theta^t(1 - \theta)^{m-t}, & t = 0, 1, 2, \ldots, m \\ 0, & \text{otherwise} \end{cases}$$

For $t \in \{0, 1, 2, \ldots, m\}$ = the range (or support) of $T$,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n|T = t),$$

---

1Any function of the data is called a statistic.
$x_i \in \{1, 2, \ldots, m_i\}, \ i = 1, 2, \ldots, n$

$$P(X_1 = x_1, \ldots, X_n = x_n, T = t)/P(T = t) = \begin{cases} P(X_1 = x_1, \ldots, X_n = x_n)/P(T = t) & \text{if } \sum_{i=1}^{n} x_i = t \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \left(\prod_{i=1}^{n} \binom{n}{x_i}\right) \theta^t (1 - \theta)^{m-t} / \left(\binom{m}{t}\theta^t (1 - \theta)^{m-t}\right) & \text{if } \sum_{i=1}^{n} x_i = t \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \prod_{i=1}^{n} \binom{n}{x_i} / \binom{m}{t} & \text{if } \sum_{i=1}^{n} x_i = t \\ 0 & \text{otherwise} \end{cases}$$

Thus, when $T$ is given, the conditional distribution of $X$ is independent of $\theta$. Hence, $T$ is a sufficient statistic for $\theta$.

When it comes to the continuous case, direct application of the definition becomes difficult since $P(T = t) = 0$ for any $t \in \mathbb{R}$, the range of $T$. The following result can be used to determine whether a statistic $T$ is sufficient for $\theta$ or not.

**Theorem 3.2.1** If $f_X(x \mid \theta)$ is the joint pdf or pmf of $X$, and $g_T(t \mid \theta)$ is the pdf or pmf of $T(X)$, then $T(X)$ is a sufficient statistic for $\theta$ if, and only if, for every $x \in \mathbb{R}^n$, the range of $X$, the ratio $f_X(x \mid \theta) / g_T(T(x) \mid \theta)$ is free from $\theta$.

**Example 3.2.2:** Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ where $\sigma^2$ is known. We would like to verify that $T(X) = \bar{X} = \sum_{i=1}^{n} x_i / n$ is a sufficient statistic for $\mu$.

The joint pdf of $X$ is

$$f_X(x \mid \mu) = \prod_{i=1}^{n} (2\pi \sigma^2)^{-1/2} \exp \left( - \frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

$$= (2\pi \sigma^2)^{-n/2} \exp \left( - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right)$$

Also, note that $\bar{X}$ follows a $N(\mu, \sigma^2/n)$ distribution (which can be obtained from the mgf of $\bar{X}$), i.e., the pdf of $\bar{X}$ given by $g_{\bar{X}}(\bar{x} \mid \mu)$ is

$$g_{\bar{X}}(\bar{x} \mid \mu) = (2\pi \sigma^2/n)^{-1/2} \exp \left( - \frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right)$$
3.2 Sufficiency and Completeness

Thus,

\[ f_X(x|\mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2\right) \]

which is free from \( \mu \), and hence, \( \bar{X} \) is a sufficient statistic for \( \mu \).

While dealing with sufficiency, we encounter two questions: first, how to verify whether a statistic is sufficient or not; and second, how to find a sufficient statistic. Theorem 3.2.1 answers the first question, whereas the following theorem settles the second one.

**Theorem 3.2.2 (Fisher-Neyman Factorization Theorem)** Let

\[ f_X(x|\theta) \]

be the joint pdf or pmf of the data \( X \). A statistic \( T(X) \) is a sufficient statistic for \( \theta \) if, and only if, there exists functions \( g_T(t|\theta) \) and \( h(x) \) such that for all \( x \) and all \( \theta \in \Theta \),

\[ f_X(x|\theta) = g_T(T(x)|\theta) \cdot h(x) \]  \hspace{1cm} (3.2.1)

Basically, we factor the joint pdf or pmf of the data into two components. The component that is free from \( \theta \) constitutes the function \( h(x) \). The other component that depends on \( \theta \) also depends on the observed data \( x \) only through some function \( T(x) \), and this function \( T \) is a sufficient statistic for \( \theta \).

**Example 3.2.3:** Let \( X_1, X_2, \ldots, X_n \) be iid random variables from the continuous uniform \( \text{CU}(\theta_1, \theta_2) \) with common pdf

\[ f_X(x|\theta) = (\theta_2 - \theta_1)^{-1}I_{(\theta_1, \theta_2)}(x) \]

where \( \theta = (\theta_1, \theta_2) \) and both \( \theta_1 \) and \( \theta_2 \) are unknown. The parameter space is \( \Theta = \{(\theta_1, \theta_2) | -\infty < \theta_1 < \theta_2 < \infty \} \). The joint pdf of \( (X_1, X_2, \ldots, X_n) \) is

\[
\begin{align*}
 f_X(x|\theta) &= \begin{cases} 
 (\theta_2 - \theta_1)^{-n} & \text{if } \theta_1 < x_i < \theta_2 \ \forall i \\
 0 & \text{otherwise}
\end{cases} \\
&= \begin{cases} 
 (\theta_2 - \theta_1)^{-n} & \text{if } \theta_1 < x^{(1)} \leq x^{(n)} < \theta_2 \\
 0 & \text{otherwise}
\end{cases}
\end{align*}
\]

where \( x^{(1)} \) and \( x^{(n)} \) are the smallest and the largest of \( x_1, x_2, \ldots, x_n \), respectively. We can further write the joint pdf as

\[ f_X(x|\theta) = (\theta_2 - \theta_1)^{-n}I_{(\theta_1, \infty)}(x^{(1)}) \cdot I_{(-\infty, \theta_2)}(x^{(n)}) \]

\[ I_{(\theta_1, \theta_2)}(x) = 1, \text{ when } \theta_1 < x < \theta_2; = 0, \text{ otherwise.} \]
Define $t_1 = T_1(x) = x(1)$, $t_2 = T_2(x) = x(n)$, $h(x) = 1$, and $g_\theta(t_1, t_2) = (\theta_2 - \theta_1)^{-n} I(\theta_1, \infty)(x(1)) \cdot I(-\infty, \theta_2)(x(n))$. Thus, by applying Theorem 3.2.2, $T = (T_1, T_2) = (X(1), X(n))$ is a sufficient statistic for $\theta = (\theta_1, \theta_2)$.

**Example 3.2.4:** Assume that $X_1, X_2, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ where both the parameters $\mu$ and $\sigma^2$ are unknown, and can be written as a two-dimensional parameter $\theta = (\mu, \sigma^2)$. It is easy to see that the joint pdf of $X = (X_1, X_2, \ldots, X_n)$ depends on the sample $X$ only through the following two values:

$$T_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad T_2 = \frac{1}{n-1} S = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

It is seen that the joint pdf $f_X(x | \theta)$ of $X$ can be written as

$$f_X(x | \theta) = g_\theta(T_1(x), T_2(x) | \theta) \cdot h(x)$$

where

$$g_\theta(t_1, t_2 | \theta) = (2\pi \sigma^2)^{-n/2} \exp \left(-\frac{(n(t_1 - \mu)^2 + (n-1)t_2)}{2\sigma^2}\right)$$

and $h(x) = 1$. Therefore, by the Factorization Theorem, $T(X) = (T_1(X), T_2(X)) = (\bar{X}, S/(n-1))$ is a sufficient statistic for $(\mu, \sigma^2)$.

The sufficient statistic $(\bar{X}, S/(n-1))$, in Example 3.2.4, contains all the relevant information about $(\mu, \sigma^2)$ for a normal model when both $\mu$ and $\sigma^2$ are unknown. But this may not be the case if the model is other than normal. For example, if our $iid$ observations have a Cauchy distribution with location parameter $\mu$ and scale parameter $\sigma$, then the common pdf is given by

$$f(x | \mu, \sigma) = \frac{1}{\pi \sigma} \left(1 + \left(\frac{x - \mu}{\sigma}\right)^2\right)^{-1}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \text{ and } \sigma > 0$$

This pdf, when plotted, looks very similar to the pdf of $N(\mu, \sigma^2)$, but the statistic $(\bar{X}, S/(n-1))$ is **not** sufficient for $(\mu, \sigma^2)$. One can verify that with $(T_1, T_2) = (\bar{X}, S/(n-1))$, the decomposition (3.2.1) of the joint pdf of $X$ under a Cauchy model (3.2.2) does not hold. Therefore, the choice of a sufficient statistic is very much model dependent.
So far we have discussed the sufficiency property that a statistic $T$ may have in relation to a parameter $\theta$ which essentially is an index identifying a family of probability distributions. Now we shall consider another property, called completeness.

Consider a statistic $T$ (may be multidimensional) based on the random variables $X_1, X_2, \ldots, X_n$ having a joint distribution $f_{X}(x|\theta)$, $\theta \in \Theta$. The distribution of $T$, in general, depends on $\theta$. Let the probability distribution of $T$ be $g_T(t|\theta)$, $\theta \in \Theta$.

**Definition 3.2.2** A statistic $T$ or, more precisely, the family of distributions $\{g_T(t|\theta), \theta \in \Theta\}$ of $T$, is called complete if for any function $\psi(T)$ of $T$, $E(\psi(T)) = 0$ for all $\theta \in \Theta$ implies $\psi(T) = 0$ with probability 1 for all $\theta \in \Theta$.

Notice that completeness is a property of a family of probability distributions, not of a particular distribution.

**Example 3.2.5:** We have already seen that if $X_1, X_2, \ldots, X_n$ are iid Bernoulli($\theta$), then the statistic $T = \sum_{i=1}^{n} X_i$ is sufficient for $\theta$ (see Example 3.2.1). $T$ has a $B(n, \theta)$ distribution. Let $\psi$ be a function such that $E(\psi(T)) = 0$ for all $\theta \in (0, 1)$. Then,

$$0 = E(\psi(T)) = \sum_{t=0}^{\infty} \psi(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = (1-\theta)^n \sum_{t=0}^{\infty} \psi(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t \hspace{1cm} (3.2.3)$$

for all $\theta \in (0, 1)$. Since $(1-\theta) \neq 0$, (3.2.3) implies

$$\sum_{t=0}^{\infty} \psi(t) \binom{n}{t} \eta^t = 0 \hspace{1cm} (3.2.4)$$

for all $\eta = \theta/(1-\theta) \in (0, \infty)$. By taking limit $\eta$ tending to 0 (i.e., $\eta \to 0$), we can get $\psi(0) = 0$. Next, by differentiating both sides of (3.2.4) with respect to $\eta$, and then taking $\eta \to 0$, we can obtain $\psi(1) = 0$. Repeat the same procedure by differentiating further and taking $\eta \to 0$ and we eventually have $\psi(2) = 0, \psi(3) = 0, \ldots, \psi(n) = 0$. Thus, $\psi(T) \equiv 0$. Hence, $T$ is a complete statistic.

**Example 3.2.6:** Let $X_1, X_2, \ldots, X_n$ be a random sample from Poisson($\theta$) (Poi($\theta$)) distribution with common pmf $f(x|\theta) = \exp(-\theta) \theta^x/x!$, $x =$
0, 1, 2, \ldots, \text{ where } \theta \in \Theta = (0, \infty). \text{ By Factorization Theorem, it is seen that } T = \sum_{i=1}^{n} X_i \text{ is a sufficient statistic for } \theta. \text{ Also, } T \text{ has a Poisson}(n\theta) \text{ distribution (easily seen from the mgf of } T). \text{ Let } \psi(T) \text{ be a function such that } E(\psi(T)) = 0 \text{ for all } \theta > 0, \text{ i.e.,}
\begin{align*}
0 &= E(\psi(T)) = \exp(-n\theta) \sum_{t=0}^{\infty} k(t) \theta^t \quad (3.2.5)
\end{align*}
where \(k(t) = \psi(t)n^t/t!\) and \(\theta > 0\). The equation (3.2.5) implies that
\[\sum_{t=0}^{\infty} k(t) \theta^t = 0 \quad \text{for all } \theta \in (0, \infty) \quad (3.2.6)\]
Differentiating both sides of (3.2.6) \(j\) times with respect to \(\theta\) \((j = 1, 2, 3, \ldots)\), and then taking \(\theta \to 0\), we can get \(k(j) = 0, \ j = 1, 2, 3, \ldots\), and it then provides \(k(0) = 0\). Therefore, \(k(T) = 0\) with probability 1; i.e., \(\psi(T) = 0\) with probability 1 implying that \(T\) is complete.

**Example 3.2.7:** Let \(X_1, X_2, \ldots, X_n\) be iid \(N(\mu, \sigma^2)\) where \(\mu\) is unknown but \(\sigma^2\) is known. Earlier we have seen that \(T = \overline{X}\) is sufficient for \(\theta\) and has \(N(\mu, \sigma^2/n)\) distribution. Let \(\psi(T)\) be a function such that \(E(\psi(T)) = 0\) for all \(\mu \in \mathbb{R}\) and it implies
\[\int_{-\infty}^{\infty} \psi(t) \exp\left(-nt^2/(2\sigma^2) + n\mu t/\sigma^2\right) dt = 0 \quad (3.2.7)\]
for \(\mu \in \mathbb{R}\). The left hand side of (3.2.7) is the bilateral Laplace transform of the function \(\psi(t) \exp\left(-nt^2/(2\sigma^2)\right)\). From the unicity theorem of Laplace transform, it follows that
\[\psi(t) \exp\left(-nt^2/(2\sigma^2)\right) = 0 \quad \text{for all } \mu \in \mathbb{R}\]
i.e., \(\psi(t) = 0\) for all \(\mu \in \mathbb{R}\). Hence, \(T = \overline{X}\) is a complete statistic.

As we have seen earlier, a sufficient statistic retains only the information relevant for the unknown parameter \(\theta\). The remaining information in the original sample (i.e., total information in the sample minus the information in a sufficient statistic) is redundant for \(\theta\), and this gives rise to the concept of an ancillary statistic.

**Definition 3.2.3** A statistic \(S(\mathbf{X})\) whose distribution does not depend on the parameter \(\theta\) is called an ancillary statistic for \(\theta\).
Notice that the distribution of an ancillary statistic is free from $\theta$, and hence, carries no information about $\theta$. It is therefore natural to expect that an ancillary statistic should have nothing to do with a sufficient statistic, i.e., an ancillary statistic should be independent of a sufficient statistic. The following result tells exactly when this is going to happen.

**Theorem 3.2.3 (Basu’s Theorem)** If $T(X)$ is sufficient for $\theta$ and is complete, then $T(X)$ is independent of any ancillary statistic $S(X)$.

Basu’s Theorem is useful in the sense that it allows us to prove the independence of two statistics without finding the joint probability distribution of the two statistics.

**Example 3.2.8:** Let $X_1, X_2, \ldots, X_n$ be iid with pdf
\[ f(x | \theta) = e^{-(x-\theta)}, \quad -\infty < \theta < x < \infty \] (3.2.8)
The joint pdf of $X = (X_1, X_2, \ldots, X_n)$ is
\[
 f_X(x | \theta) = \prod_{i=1}^{n} f(x_i | \theta) \\
 = e^{-\sum(x_i-\theta)}I(\theta, \infty)(x_1) \cdots I(\theta, \infty)(x_n) \\
 = e^{-\sum(x_i(1)-\theta)}I(\theta, \infty)(x_1(1)) \tag{3.2.9}
\]
where $x(1) \leq x(2) \leq \ldots \leq x(n)$. Using Factorization Theorem, it is easy to see that a sufficient statistic for $\theta$ is $T(X) = X(1)$. The pdf of $X(1)$ is given by
\[
 g_{X(1)}(x(1) | \theta) = n e^{-n(x(1)-\theta)}I(\theta, \infty)(x(1)) \tag{3.2.10}
\]
Also, if for some function $\psi(T) = \psi(X(1))$ we have $E(\psi(T)) = 0$ for all $\theta \in \mathbb{R}$, then we can get
\[
 \int_{\theta}^{\infty} \psi(t)e^{-nt} \, dt = 0 \quad \text{for all} \quad \theta \in \mathbb{R} \tag{3.2.11}
\]
Differentiating both sides of (3.2.11) with respect to $\theta$ yields
\[
 -\psi(\theta)e^{-n\theta} = 0 \quad \text{for all} \quad \theta \in \mathbb{R}
\]
i.e., $\psi(T) = 0$ with probability 1. Hence $T = X(1)$ is a complete sufficient statistic for $\theta$. Is $T = X(1)$ independent of $S = \sum_{i=1}^{n}(X_i - \bar{X})^2$? Note that $S$ can be also written as $S = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$ where $Y_i = X_i - \theta$ and $Y_i$’s are iid having pdf $f(y) = e^{-y}$, $y > 0$. Clearly, $Y_i$’s are ancillary and so are $\bar{Y}$ and $S$. By Basu’s Theorem, $X(1)$ is independent of $S$.\(^3\)

\(^3\)X(1) \leq X(2) \leq \ldots \leq X(n) are called order statistics of $X = (X_1, X_2, \ldots, X_n)$.
3.3 Methods of Estimation

In this section, we will consider two popular methods of estimating a parameter, namely — (i) method of moments estimation, and (ii) maximum likelihood estimation.

Method of Moments Estimators (MMEs)

Let \(X_1, X_2, \ldots, X_n\) be iid with pdf or pmf \(f(x | \theta)\), where \(\theta = (\theta_1, \theta_2, \ldots, \theta_k)\). Method of moments estimators (MMEs) of \(\theta_1, \theta_2, \ldots, \theta_k\) are found by equating the first \(k\) sample moments to the corresponding \(k\) population moments, and then solving the resultant system of \(k\) equations for \(k\) unknown parameters. In other words, let \(\mu'_1(\theta), \mu'_2(\theta), \ldots, \mu'_k(\theta)\) be the first \(k\) population raw moments, and \(m'_{(j)}(X) = \sum_{i=1}^{n} X_i^j / n, j = 1, 2, \ldots, k\), be the first \(k\)-sample moments. We can obtain the MMEs of \(\theta_1, \theta_2, \ldots, \theta_k\), say \(\hat{\theta}_1^{(MM)}, \hat{\theta}_2^{(MM)}, \ldots, \hat{\theta}_k^{(MM)}\), by solving

\[
m'_{(j)}(X) = \mu'_{(j)}(\theta), \quad i = 1, 2, \ldots, k
\]

Maximum Likelihood Estimators (MLEs)

Suppose we have a random vector \(X = (X_1, X_2, \ldots, X_n)\) with joint pdf or pmf \(f_X(x | \theta)\) where \(\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta\) (\(k\)-dimensional). The likelihood function is defined by

\[
L(\theta | X) = f_X(X | \theta)
\]

For each observed \(X = x\), let \(\hat{\theta}(x)\) be a parameter value at which \(L(\theta | X = x)\) attains its supremum (or maximum) as a function of \(\theta\). Then the maximum likelihood estimator (MLE) of \(\theta\) based on \(X\) is \(\hat{\theta}(X)\).

If the likelihood function \(L(\theta | X)\) is differentiable in each \(\theta_i\), then the MLE of \(\theta_i\)'s are found by solving the following equations

\[
\frac{\partial}{\partial \theta_i} L(\theta | X) = 0, \quad i = 1, 2, \ldots, k
\]  

and the solution \(\hat{\theta}_{ML} = (\hat{\theta}_{1(ML)}, \hat{\theta}_{2(ML)}, \ldots, \hat{\theta}_{k(ML)})\) satisfies

\[
\left. \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\theta | X) \right) \right|_{\theta = \hat{\theta}_{ML}} \leq 0
\]
The left hand side of \((3.3.4)\) is a \(k \times k\) matrix and ‘\(\leq 0\)’ implies non-positive definite.

Note that quite often \(L(\theta|X)\) is too complicated to differentiate. Hence, it is more convenient to use \(lnL(\theta|X)\) since ‘logarithm’ preserves monotonicity of \(L(\theta|X)\).

**Example 3.3.1:** Let \(X_1, X_2, \ldots, X_n\) be \(iid\) Bernoulli\((\theta)\). Find the MME and MLE of \(\theta\).

Since there is only one unknown parameter involved, we find the MME of \(\theta\) by equating the first sample moment with the population analogue as \(\bar{X} = E(X_1) = \theta\). Hence, \(\hat{\theta}_{\text{MM}} = \bar{X}\).

To find the MLE of \(\theta\), first note that the likelihood function is

\[
L(\theta|X) = \prod_{i=1}^{n} \theta^{X_i} (1 - \theta)^{1-X_i} = \theta^{\Sigma X_i} (1 - \theta)^{n - \Sigma X_i}
\]

Let \(X = \sum_{i=1}^{n} X_i\), then the log likelihood function can be written as

\[
lnL(\theta|X) = Xln\theta + (n - X)ln(1 - \theta)
\]

It is seen that \(\hat{\theta}_{\text{ML}} = X/n = \bar{X}\) which maximizes \(lnL(\theta|X)\).

**Example 3.3.2:** Let \(X_1, X_2, \ldots, X_n\) be \(iid\) CU\((0, \theta)\) (continuous uniform over the interval \((0, \theta)\)). Find the MME and MLE of \(\theta\).

The \(pdf\) of \(X_i\) is

\[
f(x_i|\theta) = \frac{1}{\theta} I(0, \theta)(x_i), \quad i = 1, 2, \ldots, n
\]

Hence, the likelihood function is

\[
L(\theta | X_1, \ldots, X_n) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0, \theta)(X_i) = \frac{1}{\theta^n} I(X_{(n)}, \infty)(\theta)
\]

where \(X_{(n)} = \max_i X_i\). Note that the equation, \(\frac{\partial}{\partial \theta} lnL(\theta|X) = 0\), has no solution. But we observe that the likelihood function \(L(\theta|X)\) is monotonically decreasing for \(\theta \geq X_{(n)}\). Therefore, \(L(\theta|X)\) is maximized at \(\theta = \hat{\theta}_{\text{ML}} = X_{(n)}\).

On the other hand, if we equate the first sample moment with the first population moment, we get \(\bar{X} = \theta/2\), i.e., \(\hat{\theta}_{\text{MM}} = 2\bar{X}\). It can be seen that the MME and MLE of \(\theta\) above are vastly different.
Both the estimation methods described above have advantages and disadvantages. For example, one may not obtain a closed form expression of the MLE always, whereas MMEs may be easier to derive and have simpler closed form expressions. A specific example of this situation is seen when iid observations are available from a Gamma distribution with both scale and shape parameters unknown (see Chapter 7 for details).

On the other hand, MLE is always a function of a sufficient statistic and invariant under transformations, as mentioned in the next two theorems.

**Theorem 3.3.1** If $T$ is a sufficient statistic for $\theta$, then the MLE of $\theta$ is a function of $T$.

**Theorem 3.3.2** If $\hat{\theta}_{ML}$ is the MLE of $\theta$, and $\eta = \psi(\theta)$ is a function of $\theta$, then $\hat{\eta}_{ML} = \psi(\hat{\theta}_{ML})$.

The above two theorems are useful tools in finding estimators of a parameter $\theta$ (scalar or vector valued). Theorem 3.3.1 says that it is enough to focus on a sufficient statistic as far as maximum likelihood estimation is concerned. Reducing the original data set by a sufficient statistic makes the problem less cumbersome and more tractable. Also, quite often we are not interested in estimating the original model parameter, rather our focus may be on some function of the original model parameter. Theorem 3.3.2 helps in deriving MLE’s of such functions of the model parameter.

**Example 3.3.3:** As an application of Theorem 3.3.2, we consider a random variable $X$ representing the number of daily auto accidents in a small town. It is assumed that $X$ follows a Poisson($\theta$) distribution where $\theta > 0$ is the average number of auto accidents per day. To estimate $\eta = P(X = 0)$, the probability of not having any auto accident in a day, we take a random sample of $n$ distinct days and record the number of auto accidents on those days as $X_1, X_2, \ldots, X_n$. Obviously, $X_1, X_2, \ldots, X_n$ are iid Poisson($\theta$), and a sufficient statistic is $T = \sum_{i=1}^{n} X_i$ following a Poisson($n\theta$) distribution. Also, it can be seen that the MLE of $\theta$ is $\hat{\theta}_{ML} = \bar{X}$. But we are only interested in estimating the parameter, $\eta = P(X = 0) = e^{-\theta}$. By Theorem 3.3.2, the MLE of $\eta$ is $\hat{\eta}_{ML} = \exp(-\hat{\theta}_{ML}) = e^{-\bar{X}}$. 

© 2006 by Taylor & Francis Group, LLC
3.4 Methods of Evaluating Estimators

Based on our data $X = (X_1, X_2, \ldots, X_n)$ following a joint pdf or pmf $f(x|\theta)$, we wish to estimate a real valued function $\eta = \eta(\theta)$ of $\theta$. Let $\hat{\eta} = \hat{\eta}(X)$ be an estimator of $\eta$. Note that $\hat{\eta}$ is a random variable. When $X = x$ is observed, $\hat{\eta}(X) = \hat{\eta}(x)$ is used to estimate the value of $\eta$, or just an ‘estimate’ of $\eta$. But how do we judge whether $\hat{\eta}$ is a ‘good’ or a ‘bad’ estimator of $\eta$?

Intuitively, $\hat{\eta} = \hat{\eta}(X)$ may be looked upon as a ‘good’ estimator if and only if it gives values $\hat{\eta}(x)$ that deviate from $\eta = \eta(\theta)$ only by a ‘small’ margin, i.e., if the probability distribution of $\hat{\eta}$ has a high degree of concentration around $\eta = \eta(\theta)$, whatever the true value of $\theta$ ($\in \Theta$) may be. More precisely, we may say that $\hat{\eta}$ is the best estimator if, for every $d > 0$,

$$P[|\hat{\eta} - \eta| < d] \geq P[|\hat{\eta}^* - \eta| < d]$$

for all $\theta \in \Theta$ (3.4.1)

whatever the rival estimator $\hat{\eta}^* = \hat{\eta}^*(X)$ may be.

Even though the condition (3.4.1) is a desirable one in finding the ‘best’ estimator of $\eta$, it is difficult to apply due to the mathematical complexities involved. Given two estimators $\hat{\eta}$ and $\hat{\eta}^*$ of $\eta$, it may not be easy to derive the probability terms on two sides of the inequality in (3.4.1).

A simpler and mathematically convenient criterion, known as mean squared error, is thus adopted to judge the performance of an estimator.

**Definition 3.4.1** The mean squared error (MSE) of an estimator $\hat{\eta}$ of a parameter $\eta = \eta(\theta)$ is defined by $\text{MSE}(\hat{\eta}) = E[(\hat{\eta} - \eta)^2]$ (which is a function of $\theta$ and will be assumed to exist for each $\theta \in \Theta$).

Note that the MSE has two components, of which one is due to bias and the other is due to variance. For $\text{MSE}(\hat{\eta})$,

$$E[(\hat{\eta} - \eta)^2] = E\left[ (\hat{\eta} - E(\hat{\eta})) + (E(\hat{\eta}) - \eta) \right]^2$$

$$= (E(\hat{\eta}) - \eta)^2 + E\left[ (\hat{\eta} - E(\hat{\eta}))^2 \right]$$

$$= (B_\theta(\hat{\eta}))^2 + \text{Var}_\theta(\hat{\eta})$$ (3.4.2)

We use the notation $f(x|\theta)$ to denote the joint pdf or pmf of $X = (X_1, X_2, \ldots, X_n)$ instead of $f_X(x|\theta)$ since the subscript is obvious from the context.
where $B_\theta(\hat{\eta}) = E(\hat{\eta}) - \eta$ is called the bias of $\hat{\eta}$. Note that both the bias and variance of $\hat{\eta}$ depend, in general, on $\theta$; but we will drop this subscript often when it is clear from the context.

**Definition 3.4.2** An estimator $\hat{\eta}$ of $\eta = \eta(\theta)$ is called unbiased if it has zero bias for all $\theta$, or $E(\hat{\eta}) = \eta$ for all $\theta \in \Theta$.

From (3.4.2), it is obvious that for an unbiased estimator, its MSE is same as its variance. Therefore, a good amount of attention has been paid, while developing the theory of statistical inference, to find estimators which are not only unbiased but also have minimum variance.

**Definition 3.4.3** An estimator is called a (uniformly) minimum variance unbiased estimator (MVUE) of $\eta(\theta)$ if it is unbiased and has the smallest variance (for each $\theta \in \Theta$) among all unbiased estimators of $\eta(\theta)$. Thus, $\hat{\eta}$ is an MVUE of $\eta(\theta)$ if

$$E(\hat{\eta}) = \eta(\theta) \quad \text{for all } \theta \in \Theta$$

and

$$\text{Var}(\hat{\eta}) \leq \text{Var}(\hat{\eta}^*) \quad \text{for all } \theta \in \Theta$$

for any other unbiased estimator $\hat{\eta}^*$ of $\eta(\theta)$.

In the following, we now present some useful theorems on MVUEs and ordinary unbiased estimators.

**Theorem 3.4.1** An MVUE, if it exists, is unique, in the sense that if both $\hat{\eta}_1$ and $\hat{\eta}_2$ are MVUEs, then $\hat{\eta}_1 = \hat{\eta}_2$ with probability 1.

According to Theorem 3.4.1, since an MVUE is unique, henceforth we shall refer to it as UMVUE (unique minimum variance unbiased estimator).

While one can check directly whether an estimator is unbiased or not, it is not immediately apparent how to satisfy oneself that an estimator has the smallest variance among all unbiased estimators. Two approaches are available to achieve this task. One method is based on the use of the Cramér-Rao inequality, and other one is based on the use of Rao-Blackwell inequality.

**Theorem 3.4.2 (Cramér-Rao Inequality)** Let $X = (X_1, X_2, \ldots, X_n)$ be a sample with pdf or pmf $f(x | \theta)$, and let $\tilde{\eta}(X)$ be any estimator where
3.4 Methods of Evaluating Estimators

$E[\hat{\eta}(X)]$ is differentiable with respect to $\theta$. Suppose the joint pdf or pmf $f(x|\theta)$ satisfies

(a) \[
\frac{\partial}{\partial \theta} \int \ldots \int h(x)f(x|\theta) \, dx_1 \ldots dx_n = \int \ldots \int h(x) \frac{\partial}{\partial \theta} f(x|\theta) \, dx_1 \ldots dx_n
\]
if $X$ is continuous; or

(b) \[
\frac{\partial}{\partial \theta} \sum_{x_1} \ldots \sum_{x_n} h(x)f(x|\theta) = \sum_{x_1} \ldots \sum_{x_n} h(x) \frac{\partial}{\partial \theta} f(x|\theta)
\]
if $X$ is discrete;

for any function $h(x)$ such that $E|h(x)| < \infty$. Then,

\[
\text{Var}(\hat{\eta}(X)) \geq \left( \frac{d}{d\theta} E[\hat{\eta}(X)] \right)^2
\]

The quantity in the denominator on the right hand side of (3.4.3) is called **Fisher information**, or the **information number**, and is denoted by

\[
I(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X|\theta) \right)^2 \right] \quad (3.4.4)
\]

If we have iid observations $X_1, X_2, \ldots, X_n$ with common pdf or pmf $f(x|\theta)$, then,

\[
I(\theta) = nE \left[ \left( \frac{\partial}{\partial \theta} \ln f(X|\theta) \right)^2 \right] = n \cdot i(\theta) \quad (\text{say}) \quad (3.4.5)
\]

where $i(\theta)$ is called Fisher information per observation.

**Example 3.4.1:** Suppose $X_1, X_2, \ldots, X_n$ are iid from a $N(\theta, \sigma^2)$ population where $\sigma^2$ is known. Consider the problem of estimation of $\eta(\theta) = \theta$, $\theta \in \Theta = \mathbb{R} = \text{the set of real numbers}$.

Here,

\[
f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x-\theta)^2}{2\sigma^2} \right)
\]
and so $\ln f(x|\theta) = (\text{constant}) - (x-\theta)^2/(2\sigma^2)$ where the ‘constant’ is free from $\theta$. It is seen that $i(\theta) = 1/\sigma^2$. Using Theorem 3.4.2,

\[
\text{Var}(\hat{\theta}(X)) \geq \left( \frac{d}{d\theta} E[\hat{\theta}(X)] \right)^2 \quad \left( \frac{n}{\sigma^2} \right)
\]

© 2006 by Taylor & Francis Group, LLC
Chapter 3  Concepts of Statistical Inference

Let \( E[\hat{\theta}(X)] = \theta + B(\hat{\theta}) \), where \( B(\hat{\theta}) \) is the bias of \( \hat{\theta} \), and then (3.4.6) can be written as

\[
Var(\hat{\theta}(X)) \geq \frac{\sigma^2}{n} (1 + B'(\hat{\theta}))^2 \tag{3.4.7}
\]

If \( \hat{\theta}(X) \) is an unbiased estimator of \( \theta \), or \( B(\hat{\theta}) = 0 \) for all \( \theta \), then (3.4.7) reduces further to

\[
Var(\hat{\theta}(X)) \geq \frac{\sigma^2}{n} \tag{3.4.8}
\]

Note that the estimator \( \hat{\theta} = \bar{X} \) has variance of \( \sigma^2/n \). Hence, \( \bar{X} \) is the UMVUE of \( \theta \). Also, the regularity condition of Theorem 3.4.2 is satisfied in this example.

**Example 3.4.2:** Suppose we have iid observations \( X_1, X_2, \ldots, X_n \) following \( N(\mu, \sigma^2) \) where \( \sigma^2 \) is unknown, but \( \mu \) is known. Call \( \theta = \sigma^2 \), the parameter space being \( \Theta = (0, \infty) \). We want to estimate (i) \( \eta(\theta) = \sigma^2 = \theta \), and (ii) \( \eta(\theta) = \sqrt{\theta} = \sigma \).

(i) Estimation of \( \sigma^2 \) (Normal Variance)

The pdf of each observation is

\[
f(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(x-\mu)^2}{2\theta}\right)
\]

and then,

\[
\ln f(x|\theta) = c - \frac{1}{2} \ln \theta - \frac{(x-\mu)^2}{2\theta}
\]

where ‘\( c \)’ is a constant free from \( \theta \). It is now easy to see that

\[
i(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X|\theta) \right)^2 \right] = \frac{1}{2\theta^2} \tag{3.4.9}
\]

Therefore, by Theorem 3.4.2, for any estimator \( \hat{\theta} = \hat{\sigma}^2 \) of \( \eta(\theta) = \sigma^2 = \theta \) with bias function \( B(\hat{\theta}) \),

\[
Var(\hat{\theta}) \geq \frac{2\theta^2}{n} (1 + B'(\hat{\theta}))^2 \tag{3.4.10}
\]

where \( B'(\hat{\theta}) = \frac{d}{d\theta} B(\hat{\theta}) \).

Now, we consider the MLE of \( \theta \) given by

\[
\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \tag{3.4.11}
\]
Since $\sum_{i=1}^{n} (X_i - \mu)^2 / \theta$ follows a $\chi^2_n$ distribution, it can be verified that $\hat{\theta}_{ML}$ is unbiased for $\theta$ and $\text{Var}(\hat{\theta}_{ML}) = 2\theta^2 / n$. Therefore, the variance of $\hat{\theta}_{ML}$ attains the Cramér-Rao lower bound, hence, $\hat{\theta}_{ML}$ is the UMVUE for $\theta = \sigma^2$.

(ii) Estimation of $\sigma$ (Normal Standard Deviation)
The parametric function to be estimated now is $\eta(\theta) = \sqrt{\theta}$. If $\hat{\eta}(X)$ is an estimator of $\eta(\theta) = \sqrt{\theta}$, then by Theorem 3.4.2,

$$\text{Var}(\hat{\eta}(X)) \geq \frac{2\theta^2}{n} \left( \frac{1}{2\sqrt{\theta}} + B'(\hat{\eta}) \right)^2$$

(3.4.12)

If $\hat{\eta}(X)$ is an unbiased estimator of $\eta(\theta) = \sqrt{\theta}$, then

$$\text{Var}(\hat{\eta}(X)) \geq \frac{\theta}{2n}$$

(3.4.13)

Now, we consider the MLE of $\eta(\theta)$ given by

$$\hat{\eta}_{ML} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 / n}$$

(3.4.14)

which is not an unbiased estimator, since

$$E(\hat{\eta}_{ML}) = \frac{\sqrt{2} \Gamma((n+1)/2)}{\sqrt{n} \Gamma(n/2)} \sqrt{\theta}$$

(3.4.15)

An unbiased estimator of $\eta(\theta) = \sqrt{\theta}$ is, thus,

$$\hat{\eta}_U(X) = \sqrt{n} \frac{\Gamma(n/2)}{2 \Gamma((n+1)/2)} \hat{\eta}_{ML}$$

(3.4.16)

A direct derivation of the variance of $\hat{\eta}_U(X)$ shows

$$\text{Var}(\hat{\eta}_U(X)) = \theta \left[ \frac{n}{2} \left( \frac{\Gamma(n/2)}{\Gamma(n+1)/2} \right)^2 - 1 \right]$$

(3.4.17)

which does not match the right hand side of (3.4.13). As a result, one may wonder whether it is possible to obtain another unbiased estimator with variance smaller than that of $\hat{\eta}_U$ in (3.4.17). We will address this question in Theorem 3.4.4.

Before we go to the next theorem, it should be pointed out that the regularity condition in Theorem 3.4.2 prohibits certain distributions to
take advantage of the Cramér-Rao inequality, as illustrated by the following example.

**Example 3.4.3:** Let $X_1, X_2, \ldots, X_n$ be iid $\text{CU}(0, \theta)$ (continuous uniform), $\theta \in \Theta = (0, \infty)$. Our goal is to estimate $\eta(\theta) = \theta$.

Note that the pdf of a single observation is

$$f(x|\theta) = \frac{1}{\theta}I_{(0,\theta)}(x)$$

As a result,

$$\frac{d}{d\theta} \int_0^\theta h(x) f(x|\theta) \, dx \neq \int_0^\theta h(x) \frac{\partial}{\partial \theta} f(x|\theta) \, dx$$

The above inequality is true in general if the support of the distribution depends on the parameter we intend to estimate. The regularity condition of Theorem 3.4.2 is not satisfied, and hence, we cannot apply the Cramér-Rao inequality to evaluate the performance of an estimator of $\theta$.

A generalization of Cramér-Rao inequality, known as Chapman-Robbins inequality, helps us handle the cases like the one in Example 3.4.3.

**Theorem 3.4.3 (Chapman-Robbins Inequality)** Let $X$ have pmf or pdf $f(x|\theta)$. $\hat{\eta}_0(X)$ is an unbiased estimator of $\eta(\theta)$. If $\theta_0$ is a fixed point in $\Theta$, and $\epsilon \neq 0$ is such that $\theta_0 + \epsilon \in \Theta$ and $f(x|\theta_0) = 0 \Rightarrow f(x|\theta_0 + \epsilon) = 0$, then

$$\text{Var}_{\theta_0}(\hat{\eta}_0(X)) \geq \sup_{-\theta < \epsilon < 0} \frac{(\eta(\theta_0 + \epsilon) - \eta(\theta_0))^2}{E_{\theta_0}\left[ \frac{f(X|\theta_0 + \epsilon)}{f(X|\theta_0)} - 1 \right]^2}$$

(3.4.18)

**Example 3.4.4:** (Continuation of Example 3.4.3). Note that (from Example 3.4.3) $f(X|\theta) = 0 \Rightarrow f(x|\theta + \epsilon) = 0$ when $-\infty < \epsilon < 0$. Hence, Chapman-Robbins inequality can be applied and the lower bound to the variance of an unbiased estimator of $\eta(\theta) = \theta$ is

$$\sup_{-\theta < \epsilon < 0} \frac{\epsilon^2}{E_{\theta}\left[ (f(X|\theta + \epsilon) - f(X|\theta))^2 \right]} = \sup_{-\theta < \epsilon < 0} \left( \frac{\epsilon^2(\theta + \epsilon)^n}{\theta^n - (\theta + \epsilon)^n} \right)$$

If $n = 1$, the above lower bound is $\theta^2/4$.

The Cramér-Rao inequality, or more generally, the Chapman-Robbins inequality can help us in obtaining UMVUE estimators provided the lower
3.4 Methods of Evaluating Estimators

bound to the variance is attained. But if an unbiased estimator does not attain the variance lower bound, then still it does not mean that the estimator is not UMVUE. The following theorem, coupled with the concept of completeness discussed earlier, helps us in settling the question whether an unbiased estimator is UMVUE or not even if the estimator does not attain the Cramér-Rao lower bound.

**Theorem 3.4.4 (Rao-Blackwell Theorem)** Let \( \hat{\eta}^*_U(X) \) be any unbiased estimator of \( \eta(\theta) \), and let \( T \) be a sufficient statistic for \( \theta \). Define \( \hat{\eta}_U(T) = E[\hat{\eta}^*_U | T] \). Then \( \hat{\eta}_U(T) \) is unbiased for \( \eta(\theta) \) and \( \text{Var}_\theta(\hat{\eta}_U(T)) \leq \text{Var}_\theta(\hat{\eta}^*_U(X)) \) for all \( \theta \in \Theta \).

**Corollary 3.4.1** If the sufficient statistic \( T \) is complete, then the unbiased estimator \( \hat{\eta}_U(T) \), as defined in Theorem 3.4.4, is the UMVUE.

**Example 3.4.5:** (Continuation of Example 3.4.2, part (ii)).
The statistic \( T = \sum_{i=1}^n (X_i - \mu)^2 \) is sufficient for \( \theta = \sigma^2 \). Also, \( T \) has a complete family of distributions. Therefore, by Corollary 3.4.1 and Theorem 3.4.4,

\[
\hat{\eta}_U(T) = \sqrt{\frac{n}{2}} \frac{\Gamma(n/2)}{\Gamma((n+1)/2)} \sqrt{\frac{T}{n}}
\]

is the UMVUE of \( \eta(\theta) = \sqrt{\theta} = \sigma \).

The real utility of Theorem 3.4.4 and Corollary 3.4.1 lies in problems where \( \eta(\theta) \) is complicated.

**Example 3.4.6:** Let \( X_1, X_2, \ldots, X_n \) be iid Poisson(\( \theta \)). We want to estimate

\[
\eta(\theta) = P(X = k) = \frac{\theta^k e^{-\theta}}{k!}, \quad \theta \in \Theta = (0, \infty)
\]

Define a random variable \( Y \) as \( Y = I(X_1 = k) \).

Then, \( E_\theta(Y) = 1 \cdot P(X_1 = k) + 0 \cdot P(X_1 \neq k) = \eta(\theta) \), \( \forall \theta \in \Theta \).

So, \( Y \) is an unbiased estimator of \( \eta(\theta) \). Now, recall that \( T = \sum_{i=1}^n X_i \) is sufficient for \( \theta \), and \( T \), following a Poisson(n\( \theta \)), has a complete family of distributions. Let \( \hat{\eta}_U(T) = E(Y | T) \). It can be seen that for \( T = t \),

\[
\hat{\eta}_U(t) = P(X_1 = k | T = t) = \binom{t}{k} \frac{(n-1)^{t-k}}{n^t}.
\]

Thus,

\[
\hat{\eta}_U(T) = \binom{T}{k} \frac{(n-1)^{T-k}}{n^T}
\]

is the UMVUE of \( \eta(\theta) = \theta^k e^{-\theta} / k! \).

\[I(X_1 = k) = 1, \text{ when } X_1 = k; = 0, \text{ otherwise.}\]
3.5 Elements of Hypothesis Testing

Consider the random vector \( X = (X_1, X_2, \ldots, X_n)' \) having a joint probability distribution \((pmf \ or \ pdf) \ f(x|\theta)\) where \( \theta \in \Theta \), the parameter space. It is assumed that \( f \) is a known function except for the parameter \( \theta \). Our inference problem is to have more specific knowledge about the true value of \( \theta \) within \( \Theta \). This problem is to test a statement (called a hypothesis) when we have before us a suggestion in the form

\[
H_0 : \theta \in \Theta_0
\]  

(3.5.1)

where \( \Theta_0 \) is a specified proper subset of \( \Theta \). A statement such as (3.5.1), whose validity needs to be verified on the basis of a set of observations on \( X \), is called a statistical hypothesis and is generally denoted, as in (3.5.1), by \( H_0 \).

In this section, we will discuss the procedure that one needs to follow in testing (or verifying) a hypothesis.

The formulation of criteria that are used in constructing a test for any given hypothesis is the subject-matter of the theory of hypothesis testing. For this purpose, it will be useful to make a distinction between a simple hypothesis and a composite hypothesis.

**Definition 3.5.1** If a hypothesis \( H_0 \) specifies the joint distribution of \( X = (X_1, X_2, \ldots, X_n)' \) completely, then \( H_0 \) is called a simple hypothesis. Otherwise, \( H_0 \) is said to be a composite hypothesis.

Consider the case where we have observations \( X_1, X_2, \ldots, X_n \) iid \( N(\mu, 1) \), \( \mu \in \mathbb{R} \), and we want to test the validity of the statement (hypothesis) “\( \mu = 0 \)”. The statement “\( \mu = 0 \)” makes the above probability distribution completely known, and hence it is a simple hypothesis. On the other hand, if \( X_1, X_2, \ldots, X_n \) are independent with \( X_i \) following \( N(\mu_i, 1) \), \( \mu_i \in \mathbb{R} \), \( i = 1, 2, \ldots, n \), and one wants to verify the statement “\( \mu_1 = \mu_2 = \ldots = \mu_n \)”, then under this statement \( X_i \)'s are supposed to be iid \( N(\mu, 1) \) for some unknown \( \mu \in \mathbb{R} \) (common for all \( X_i \)'s). As a result, the distribution of \( X_i \)'s still remains unspecified (nonunique), and hence the statement “\( \mu_1 = \mu_2 = \ldots = \mu_n \)” is termed as a composite hypothesis.

The validity of a hypothesis \( H_0 \) regarding the distribution of \( X = (X_1, X_2, \ldots, X_n)' \) is judged on the basis of the realization (observation) \( x = (x_1, x_2, \ldots, x_n)' \) of \( X \). The value of \( x \) will dictate whether
we should consider $H_0$ to be valid or not. Suppose $R_X$ denotes the range of $X$ (i.e., the set of all possible values that $X$ can assume). The range $R_X$ is partitioned into two regions, say, $A$ and $R$, called acceptance region and rejection region, respectively. If the observation $x$ falls in the region $A$, then $H_0$ is considered to be valid, and $H_0$ is said to be accepted. On the other hand, if $x$ belongs to the region $R$ ($R = A^c$ since $A \cap R = \emptyset$), then $H_0$ is not considered to be valid and $H_0$ is said to be rejected. Often the rejection region $R$ is referred to as critical region.

It should be kept in mind that the acceptance (rejection) of $H_0$ does not mean that $H_0$ has been proved to be true (untrue); rather it only means that as far as the evidence $x$ is concerned, $H_0$ seems to be a plausible (nonplausible) hypothesis.

Testing a hypothesis $H_0$ is thus a procedure to partition $R_X$ into $A$ and $R(= A^c)$ which will determine whether to accept $H_0$ or not based on the evidence $x$. To each test procedure there corresponds a unique partition of $R_X$; and conversely, to each partition there corresponds a unique test procedure. Furthermore, each partition can be characterized by the corresponding critical region (rejection region) only.

Any hypothesis which differs from the null hypothesis is called an alternative hypothesis, and is denoted by $H_1$. An alternative hypothesis is considered in such a way that it is the one to be accepted when the null hypothesis is rejected.

For a parameter $\theta$, the general format of the null and alternative hypotheses is $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_0^c$ where $\Theta_0$ is some subset of the parameter space $\Theta$ and $\Theta_0^c$ is its complement. But practicality and common sense can make our alternative hypothesis confined to a smaller subclass (smaller than $\Theta_0^c$).

For example, an automobile manufacturer claims that its new compact models get, on an average, 30 miles per gallon (mpg) on highways. Let $\theta$ be the mean of the mileage distribution for these cars. Common sense tells that the manufacturer is not going to underrate the car, but we suspect that the mileage claim might be overstated. We want to see if the manufacturer’s claim of $\theta = 30$ mpg can be rejected. Therefore, we set out null hypothesis as $H_0 : \theta = 30$. If $H_0$ is not accepted, we are sure that $\theta$ is less than 30 mpg since we believe that the manufacturer’s claim is too high. Hence, our alternative hypothesis is $H_1 : \theta < 30$.

If we reject the null hypothesis when it is in fact true, we have made
an error that is called a **type-I error**. On the other hand, if we accept the null hypothesis when it is in fact *false*, we have made another error that is called a **type-II error**. Table 3.5 indicates how these errors occur.

<table>
<thead>
<tr>
<th>Truth about ( H_0 )</th>
<th>Decision Made</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0 is true</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>H0 is false</td>
<td>Reject ( H_0 )</td>
</tr>
</tbody>
</table>

![Table 3.5.1: Type-I and Type-II Errors.](image)

Note that we make a decision about \( H_0 \) (acceptance or rejection) based on \( X \), and hence, commission of an error of either type is a random event. Therefore, we can only talk about the probability of making each error. Usually for a given sample size, an attempt to reduce the probability of one type of error results in an increase in the probability of the other type of error. In practical applications, one type of error may be more serious than the other. In such a case, careful attention is given to the probability of more serious error. It is possible to reduce both types of errors simultaneously by increasing the sample size, but this may not always be possible due to limited resources.

The probability with which we are willing to risk a type-I error is called the **level of significance** of a test and it is often denoted by \( \alpha \). The probability of making a type-II error is denoted by \( \beta \). The quantity \((1 - \beta)\) represents the probability of rejecting \( H_0 \) when it is in fact false, and this \((1 - \beta)\) is called the **power** of the test.

According to the partition of \( A \) (acceptance region) and \( R \) (rejection or critical region) discussed earlier, the probability of type-I error associated with a given \( \theta \) is \( P_\theta(X \in R) = P_\theta(R) \), provided \( \theta \in \Theta_0 \) (in (3.5.1)). On the other hand, the probability of type-II error associated with a given \( \theta \) is \( P_\theta(X \in A) = P_\theta(A) \), provided \( \theta \notin \Theta_0 \). Since \( R = A^c \), we can express the probability of type-II error as \((1 - P_\theta(R))\) for \( \theta \notin \Theta_0 \).

The probability \( P_\theta(R) \), regarded as a function of \( \theta, \theta \in \Theta \), is called the **power function** (or the **operating characteristic function**) of the test, i.e.,

\[
P_\theta(R) = \begin{cases} 
\text{Probability of type-I error} & \text{if } \theta \in \Theta_0 \\
1 - \text{Probability of type-II error} & \text{if } \theta \notin \Theta_0 
\end{cases} \quad (3.5.2)
\]
3.5 Elements of Hypothesis Testing

For fixed level of significance $\alpha$ ($0 < \alpha < 1$) we choose a partition of $R_X = A \bigcup R$ such that $P_\theta(R) \leq \alpha$ for $\theta \in \Theta_0$. The value $\sup_{\theta \in \Theta_0} P_\theta(R)$ is called the size of the test and it may not necessarily be same as $\alpha$.

Likelihood Ratio Tests

The likelihood ratio method of hypothesis testing is based on maximum likelihood estimation. If $X = (X_1, X_2, \ldots, X_n)'$ is a random vector with pdf or pmf $f(x|\theta)$, then the likelihood function is defined as

$$L(\theta|X) = f(X|\theta)$$

Definition 3.5.2 The likelihood ratio test statistic for testing $H_0: \theta \in \Theta_0$ against $H_1: \theta \in \Theta^c_0$ is

$$\lambda(X) = \frac{\sup_{\theta \in \Theta_0} L(\theta|X)}{\sup_{\theta \in \Theta} L(\theta|X)}$$ (3.5.3)

A likelihood ratio test (LRT) is a test procedure that has a rejection region of the form $R = \{x|\lambda(x) \leq c\}$, where $c$ is a suitable value between 0 and 1.

The rationale behind LRT principle is that if $H_0$ is really true then it would be supported by the data, and hence the restricted maximization of $L(\theta|X)$ in the numerator of (3.5.3) would come close to the unrestricted maximization in the denominator of (3.5.3) making $\lambda(X)$ come close to 1. Otherwise, if $H_0$ is really false, $\lambda(X)$ would be close to 0. The value of $c$ which determines acceptance or rejection of $H_0$ is determined by the level of significance, i.e., $\alpha$. Also, if $T = T(X)$ is a sufficient statistic for $\theta$, then for convenience, one can use the likelihood function based on $T$ instead of $X$.

Example 3.5.1: Let $X_1, X_2, \ldots, X_n$ be iid $N(\theta, 1)$ random variables. Consider testing $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$, where $\theta_0$ is a prefixed value. It can be shown that the LRT statistic is

$$\lambda(X) = \exp \left\{-n(X - \theta_0)^2/2\right\}$$ (3.5.4)

where $X = \sum^n_1 X_i/n$. The rejection region $R$ is $R = \{x|\lambda(x) \leq c\}$ (with $c$ strictly between 0 and 1) which can be rewritten as

$$R = \{x|X - \theta_0| \geq \sqrt{-2/n}\log c\}$$
Call $\sqrt{-(2/n)\ln c} = c_*$, which ranges between 0 and $\infty$ as $c$ ranges between 0 and 1. How do we choose $c$, and hence $c_*$? Note that there is a one-to-one correspondence between $c$ and $c_*$. If we set our probability of type-I error at level $\alpha$, then

$$\alpha = P_{\theta_0}(|X - \theta_0| \geq c_*)$$

$$= P_{\theta_0}(|\sqrt{n}(X - \theta_0)| \geq \sqrt{n}c_*)$$

Since $\sqrt{n}(X - \theta_0)$ follows $N(0,1)$ distribution, $\sqrt{n}c_*$ must be equal to $z(\alpha/2)$, the upper $(\alpha/2)$-cutoff point of a standard normal curve; i.e.,

$$c_* = \frac{1}{\sqrt{n}}z(\alpha/2).$$

The rejection region thus becomes

$$R = \{x \mid x \leq \theta_0 - \frac{1}{\sqrt{n}}z(\alpha/2) \text{ or } x \geq \theta_0 + \frac{1}{\sqrt{n}}z(\alpha/2)\}$$

The power function $\beta(\theta)$ is obtained as

$$\beta(\theta) = P_\theta(|X - \theta_0| \geq \frac{1}{\sqrt{n}}z(\alpha/2))$$

$$= 1 - P_\theta(\theta_0 - \frac{1}{\sqrt{n}}z(\alpha/2) \leq X \leq \theta_0 + \frac{1}{\sqrt{n}}z(\alpha/2))$$

$$= 1 - \left[\Phi\left(\sqrt{n}(\theta_0 - \theta) + z(\alpha/2)\right) - \Phi\left(\sqrt{n}(\theta_0 - \theta) - z(\alpha/2)\right)\right]$$

(3.5.5)

where $\Phi(\cdot)$ is the standard normal cdf. Figure 3.5.1 gives the plot of $\beta(\theta)$ in (3.5.5) for $\theta_0 = 0$, $\alpha = 0.05$, and $n = 5$.

**Definition 3.5.3** A test with power function $\beta(\theta)$ is unbiased if $\beta(\theta^*) \geq \beta(\theta_0)$ for any $\theta_0^* \in \Theta_0$ and $\theta^* \in \Theta_0^c$.

Note that in Example 3.5.1, the power function $\beta(\theta)$ is always greater than $\alpha$ for $\theta \neq \theta_0$ (which can be proved analytically by using the symmetry and unimodality of a normal curve). Therefore, the LRT test for the normal model does have an unbiased power function. Also, unbiasedness of a test procedure is a reasonable requirement which essentially says that the probability of making a correct decision is greater than or equal to the probability of making a wrong decision.

**Definition 3.5.4** Let $\mathcal{C}$ be a class of tests for testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_0^c$. A test in class $\mathcal{C}$ with power function $\beta_0(\theta)$, is a uniformly
Elements of Hypothesis Testing

Figure 3.5.1: The Power Function $\beta(\theta)$.

A most powerful (UMP) test in class $C$ provided $\beta_0(\theta) \geq \beta(\theta)$ for all $\theta \in \Theta_0^c$, where $\beta(\theta)$ is the power function of any other test in class $C$.

Usually, the class $C$, mentioned in the above definition, is taken as the collection of all tests with significance level $\alpha$. For obvious notational simplicity, we can thus denote such a class by $C_\alpha$ instead of $C$.

In many problems with standard probability distributions, the UMP tests do not exist since $C_\alpha$ is too large, and consists of tests with power function very good ($\beta(\theta)$ value is closed to 1) in a subregion of $\Theta_0^c$, but very poor ($\beta(\theta)$ value is closed to 0) in another subregion of $\Theta_0^c$. In a situation like this, we look for an optimal test in a smaller class of tests like $C_\alpha^U$, the class of all unbiased tests with significance level $\alpha$. An UMP test in $C_\alpha^U$ is called a uniformly most powerful unbiased (UMPU) test with significance level $\alpha$. 
One of the major goals in hypothesis testing is to identify UMP tests within suitable classes. The following results, which clearly describes UMP level $\alpha$ tests for testing a simple null against a simple alternative, is useful for more general situations.

**Theorem 3.5.1 (Neyman-Pearson Lemma)** For testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, consider a rejection region of the form

$$R_{NP} = \left\{ x \mid f(x \mid \theta_1) > k f(x \mid \theta_0) \right\}$$

where $k \geq 0$ is such that $\alpha = P_{\theta_0}(X \in R)$. A test is UMP (in $C_\alpha$) with significance level $\alpha$ if and only if it has the above rejection region, $R_{NP}$.

**Example 3.5.2:** For iid observations $X_1, X_2, \ldots, X_n$ with $N(\theta, 1)$ distribution, the rejection region for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 (\theta_1 > \theta_0)$, is

$$R_{NP} = \left\{ x \mid f(x \mid \theta_1) > k f(x \mid \theta_0) \right\} = \left\{ x \mid \sqrt{n}(\overline{X} - \theta_0) > k_s(\theta_0, \theta_1, n) \right\}$$

where $k_s = k_s(\theta_0, \theta_1, n) = ((\theta_1 - \theta_0)/2) + \ln k/(n(\theta_1 - \theta_0))$. Since $\theta_0, \theta_1,$ and $n$ are all known, finding $k$ is equivalent to finding $k_s$, and this is done subject to the level condition, i.e.,

$$\alpha = P_{\theta_0}(\overline{X} - \theta_0 > k_s) = P_{\theta_0}(\sqrt{n}(\overline{X} - \theta_0) > \sqrt{n}k_s)$$

Since $\sqrt{n}(\overline{X} - \theta_0)$ follows a standard normal distribution, from (3.5.7) we can obtain

$$k_s = \frac{1}{\sqrt{n}} z_\alpha$$

Therefore, the rejection region for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 (\theta_1 > \theta_0)$ is

$$R_{NP} = \left\{ x \mid \overline{X} - \theta_0 > z_\alpha/\sqrt{n} \right\}$$

which is free from $\theta_1$. The test with this rejection region has maximum power at $\theta_1$. But since the rejection region is free from $\theta_1$, it does not really matter which $\theta_1 (> \theta_0)$ is being used. As long as $\theta_1$ is greater than $\theta_0$, the test with rejection region (3.5.8) is the most powerful in the class of all tests with the level $\alpha$ for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$. 

© 2006 by Taylor & Francis Group, LLC
Using the monotone likelihood ratio (MLR) property of the distribution of $\sqrt{n}(\bar{X} - \theta_0)$ (which is $N(0,1)$), it can be shown that the test with the critical region $(3.5.8)$ is indeed UMP for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$. The MLR property and its usage are discussed in the following.

**Definition 3.5.5** A family of pdfs or pmfs $\{h(s|\theta), \theta \in \Theta\}$ of a univariate random variable $S$ with real valued parameter $\theta$ has a monotone likelihood ratio (MLR) if, for every $\theta_2 > \theta_1$, $r(s) = (h(s|\theta_2)/h(s|\theta_1))$ is a nondecreasing function of $s$ on the set $\{s | h(s|\theta_1) > 0 \text{ or } h(s|\theta_2) > 0\}$, where $r(s)$ takes the value $\infty$ if $h(s|\theta_1) = 0$.

For a nonnegative random variable $S$ we have the following useful results which essentially use the MLR property. For notational simplicity we may use $h_\theta(s)$ instead of $h(s|\theta)$ as the pdf or pmf of $S$.

**Lemma 3.5.1** Let $S$ be a nonnegative random variable with pdf or pmf $h_\theta(s)$, $\theta \in \Theta$. If

$$\pi(s) = \frac{h_{\theta_1}(s)}{h_{\theta_2}(s)}$$

is increasing in $s$ for $\theta_1, \theta_2 \in \Theta$, then $P_{\theta_1}(S > c) \geq P_{\theta_2}(S > c)$ for all $c > 0$.

**Lemma 3.5.2** $P_{\theta_1}(S > c) \geq P_{\theta_2}(S > c)$ for all $c > 0$ implies that

$$E_{\theta_1}[S] \geq E_{\theta_2}[S]$$

**Theorem 3.5.2** Let $S$ be a nonnegative random variable with pdf or pmf $h_\theta(s)$, $\theta \in \Theta$. If

$$\pi(s) = \frac{h_{\theta_1}(s)}{h_{\theta_2}(s)}$$

is increasing in $s$ for $\theta_1, \theta_2 \in \Theta$, then

$$E_{\theta_1}[\phi(S)] \leq E_{\theta_2}[\phi(S)]$$

for any nonincreasing function $\phi$. 

© 2006 by Taylor & Francis Group, LLC
Large Sample Tests

Recall Example 3.5.1 where we had to test \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta \neq \theta_0 \) for iid observations \( X_1, X_2, \ldots, X_n \) from a \( N(\theta, 1) \) distribution. The LRT test procedure is characterized by the rejection region

\[
R = \left\{ x \mid |x - \theta_0| \geq \frac{1}{\sqrt{n}} z_{(\alpha/2)} \right\}
\]

which is actually determined by \((x - \theta_0)\) or just by \(x\). In fact, the boundary of the rejection region, i.e., \((\theta_0 \pm z_{(\alpha/2)}/\sqrt{n})\) is determined by the probability distribution of the random variable \( \bar{X} \) under the null hypothesis (or \( \theta = \theta_0 \)). The random variable, in this case \( \bar{X} \), which determines the boundary of the rejection region, is called the test statistic of the test procedure.

In many problems, it is not easy to derive the boundary of a rejection region due to the complicated probability distribution of the test statistic under the null hypothesis. In such cases, it may be possible to derive approximate rejection regions by using large sample (asymptotic) properties of useful yet common statistics (functions of the sample \( X \)).

Suppose we wish to test a hypothesis about a real valued parameter \( \theta \), and \( T(X) \) is a point estimator of \( \theta \) based on the random sample \( X = (X_1, X_2, \ldots, X_n)' \) of size \( n \). For example, \( T(X) \) could be the MLE of \( \theta \). Suppose \( \sigma_T^2(\theta) \) is the variance of \( T(X) \), then \( Z_* = (T(X) - \theta)/\sigma_T \) follows \( N(0, 1) \) distribution asymptotically,\(^6\) and we can use \( T(X) \) as an approximate test statistic for testing a hypothesis on \( \theta \). If \( \sigma_T \) is unknown, then replace it by a suitable estimate \( \hat{\sigma}_T \) so that the asymptotic normality still holds.

The likelihood ratio test procedure discussed earlier gives an explicit definition of the test statistic \( \lambda(X) \) (see 3.5.3) and an explicit form of the rejection region as \( R = \{ x \mid \lambda(x) \leq c \} \). The constant \( c \) is chosen subject to the level condition,

\[
P_{\theta}(\lambda(X) \leq c) \leq \alpha \quad \text{for all } \theta \in \Theta_0
\]

Hence, we need to derive the probability distribution of \( \lambda(X) \) under the null hypothesis. If \( \lambda(X) \) has a complicated probability distribution which might frustrate one from deriving the value of \( c \), then we can look at the

\(^6\)For large sample size \( n \), the distribution of \( Z_* \) can be approximated by the \( N(0, 1) \) distribution.
asymptotic distribution of $\lambda(X)$ (or a suitable monotone function of it) and derive an approximate value of $c$ for large sample sizes. This is given in the following theorem.

**Theorem 3.5.3** Let $X_1, X_2, \ldots, X_n$ be iid following a probability distribution with pdf or pmf $f(x|\theta), \theta \in \Theta$. For testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, the distribution of $(-2 \ln \lambda(X))$ is asymptotically $\chi^2_\nu$ (chi-square distribution with $\nu$ degrees of freedom) provided some regularity conditions are satisfied by $f(x|\theta)$. Also, $\nu$, the degrees of freedom, is determined by the difference between the number of free parameters specified by $\theta \in \Theta$ and the number of free parameters specified by $\theta \in \Theta_0$.

Theorem 3.5.3 says that for sufficiently large sample size (and subject to some regularity conditions) the distribution of $(-2 \ln \lambda(X))$ is approximated by $\chi^2_\nu$. Since rejection of $H_0$ for ‘small’ values of $(-2 \ln \lambda(X))$ is equivalent to rejection of $H_0$ for ‘large’ values of $(-2 \ln \lambda(X))$, asymptotically $H_0$ is rejected if and only if $(-2 \ln \lambda(X)) \geq \chi^2_{\nu, \alpha}$, where $\chi^2_{\nu, \alpha}$ is the right (or upper) tail $\alpha$-probability cutoff point of the $\chi^2_\nu$ distribution.

**Example 3.5.3:** Consider the random vector $X = (X_1, X_2, \ldots, X_k)'$ having a joint probability mass function (pmf):

$$f(x|\theta) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k|\theta) = \frac{n!}{x_1!x_2! \ldots x_k!} \theta_1^{x_1} \theta_2^{x_2} \ldots \theta_k^{x_k},$$  \hspace{1cm} (3.5.9)

where $n$ is a fixed positive integer, $\theta = (\theta_1, \theta_2, \ldots, \theta_k), 0 \leq \theta_i \leq 1, 1 \leq i \leq k; \sum_{i=1}^{k} \theta_i = 1$, and $x_1 + x_2 + \ldots + x_k = n$. The above probability distribution is called a multinomial distribution (a generalization of the binomial distribution), and it is denoted by $M(n, \theta_1, \ldots, \theta_k)$. Assume that $\theta_i$’s are all unknown and we want to test

$$H_0 : (\theta_1, \theta_2, \ldots, \theta_k) = (\theta_1^0, \theta_2^0, \ldots, \theta_k^0)$$

against

$$H_1 : \theta_i \neq \theta_i^0 \text{ for some } i$$

Such a hypothesis testing problem arises when we want to test a population composition. For example, five years ago, an artificial lake was stocked with fish in the following proportions: 30% bluegill, 25% bass, 25% catfish, and 20% pike. Five years later, a sample of size 400 fish is drawn from the lake, and the number of each type fish was recorded as below:
Do these data indicate that the composition of fish has changed over the last five years?

Note that for testing $H_0: \theta = (\theta_1, \theta_2, \ldots, \theta_k) = (\theta_1^0, \theta_2^0, \ldots, \theta_k^0)$ (say), the MLEs of $\theta_i$ under $H_0$ are simply $\hat{\theta}_i^0$. The unrestricted MLEs are $\hat{\theta}_i = X_i/n, i = 1, 2, \ldots, k$, where $X_i = \#$ fish of type-$i$. Therefore, the likelihood ratio is

$$
\lambda(x) = \frac{(\theta_1^0)^{x_1} \cdots (\theta_k^0)^{x_k}}{(\hat{\theta}_1)^{x_1} \cdots (\hat{\theta}_k)^{x_k}} = \prod_{i=1}^{k} \left( \frac{n\theta_i^0}{x_i} \right)^{x_i}
$$

or

$$
-2 \ln \lambda(x) = -2 \sum_{i=1}^{k} x_i \ln \left( \frac{n\theta_i^0}{x_i} \right)
$$

$$
= 2 \sum_{i=1}^{k} x_i \ln \left( \frac{x_i}{e_i} \right)
$$

(3.5.10)

where $e_i = n\theta_i^0$ is the expected number of items of $i^{th}$ type when $H_0$ is true. By Theorem 3.5.3, the distribution of $(-2 \ln \lambda(x))$ is asymptotically (as $n \to \infty$) $\chi^2_k$. When $H_0$ is true, all $\theta_i$ values are specified, and hence the number of free parameters under $H_0$ is 0. In the full parameter space $\Theta$, the $\theta_i$ values are subject to the restriction: $\theta_1 + \theta_2 + \ldots + \theta_k = 1$.

Therefore, the number of free parameters in $\Theta$ are $(k - 1)$. Hence, for large $n$, we reject $H_0$ if $(-2 \ln \lambda(x))$ exceeds $\chi^2_{k-1, \alpha}$.

For the above mentioned fish example, $k = 4$, and $(\theta_1^0, \theta_2^0, \theta_3^0, \theta_4^0) = (0.30, 0.25, 0.25, 0.20)$. Using (3.5.10), it is found that

$$
-2 \ln \lambda(x) = 2.4543
$$

Taking $\alpha = 0.05$, from the chi-square distribution table in Appendix A, we can obtain $\chi^2_{k-1, \alpha} = \chi^2_{3, 0.05} = 7.8147$. Since $(-2 \ln \lambda(x)) = 2.4543 < 7.8147$, we accept the $H_0$ statement, and conclude that five years later, the proportions of each fish type remain the same as 30% bluegill, 25% bass, 25% catfish, and 20% pike. Moreover, since $n = 400$ is sufficiently large, we can use this asymptotic result easily.

The Notion of P-Value

The level of significance $\alpha$ in a hypothesis testing is the probability of rejecting $H_0$ when it is true. The value of $\alpha$ is set before the test is done.
However, two different values of $\alpha$ may lead to two different conclusions for the same sample data as seen in the following example.

**Example 3.5.4:** A lending agency has lowered their auto finance rate by 0.5%. Past records of accounts before the interest rate change showed the mean amount borrowed by new car buyers to be $11,520. The management believes that the lower rate will encourage new car buyers to borrow more money. A random sample of 81 accounts were examined one year after the interest rate reduction. The average amount borrowed was $\bar{x} = 12,300$ with standard deviation $s = 3,000$. Test the management’s belief at 0.05 and 0.01 levels of significance. Assume that amount borrowed by a new car buyer follows a normal distribution.

In this problem, we would like to test $H_0 : \mu = 11,520$ against $H_1 : \mu > 11,520$, where $\mu$ represents the mean amount borrowed by new car buyers after the rate change. The test statistic $T(X) = \sqrt{n}(\bar{X} - 11520)/s$ follows $t_{n-1}$, i.e., $t_{80}$, distribution under $H_0$. The sample of size $n = 81$ gives $\bar{x} = 12,300$ with $s = 3,000$ resulting the value of $T$ as $t = \sqrt{80}(12300 - 11520)/3000 = 2.34$. If $\alpha$ is taken as 0.05, then the critical value is $t_{n-1,\alpha} = t_{80,0.05} = 1.6641 < t = 2.34$, and hence $H_0$ is rejected at 5% level. On the other hand, if $\alpha$ is used as 0.01, then $t_{80,0.01} = 2.374 > t = 2.34$ and $H_0$ is accepted at 1% level.

In general, if the data $x$ falls in the rejection region of a test procedure with level $\alpha$, then it will also be in the rejection region of the same test with level $\alpha' > \alpha$. A natural question is: “What is the smallest level of significance at which the sample data will tell us to reject $H_0$?” The answer is the ‘$p$-value’ associated with the test statistic value.

The $p$-value for the data $x$ (test statistic $T(x)$) is the smallest value of $\alpha$ for which the data $x$ (test statistic $T(x)$) will lead to rejection of $H_0$.

In Example 3.5.4, the test statistic value $T(x)$ is $t = T(x) = 2.34$. The test statistic value $T(X) = \sqrt{n}(X - \mu_0)/s$ follows $t_{n-1}$-distribution under $H_0 : \mu = \mu_0$. In this case, the $p$-value is computed as

$$p\text{-value} = P_{\mu_0}(T(X) \geq t) = P(t_{n-1} \geq 2.34) \approx 0.014$$

More generally, if $H_0 : \theta \in \Theta_0$ is a composite hypothesis, then the
Chapter 3 Concepts of Statistical Inference

\[ p \text{-value is described as } \]
\[ p \text{-value} = \sup_{\theta \in \Theta} P_{\theta} \left( \text{test statistic is as or more extreme than the observed test statistic value} \right) \]
\[ (3.5.11) \]

Instead of a fixed \( \alpha \) value, \( p \)-value has more appeal because it presents the evidence against accepting \( H_0 \) and in favor of \( H_1 \). The smaller the \( p \)-value that one can obtain, the stronger the sample evidence that shows \( H_1 \) is true.

3.6 Set (or Interval) Estimation

In Section 3.3, we discussed point estimation of a parameter \( \theta \) based on \( X = (X_1, X_2, \ldots, X_n)' \) where we had to guess a single value \( \hat{\theta}(X) \) as the value of \( \theta \). Since \( X \) is random (having a probability distribution \( f(x|\theta), \theta \in \Theta \) it is unlikely that \( \hat{\theta}(x) \), when \( X = x \) is observed, would be the exact value of \( \theta \). All that we try by taking \( \hat{\theta}(x) \) to be the estimate of \( \theta \) is that \( \hat{\theta}(x) \) is likely to differ from \( \theta \) by a ‘small amount’. That is why it is customary to provide standard error (standard deviation of \( \hat{\theta}(X) \)) or an estimate of the standard error as a measure of variation of \( \hat{\theta}(X) \). Armed with the standard error (SE) or its estimate, one expects \( \theta \) to lie between \( \hat{\theta}(X) \pm \text{SE} \), or more likely to lie between \( \hat{\theta}(X) \pm 2 \text{SE} \). This idea leads us to the concept of set estimation based on the observed data \( X = x \). We propose a set \( C = C(x) \) such that \( C \subseteq \Theta \) and

\[ P(\theta \in C(X)) \geq (1 - \alpha) \]
\[ (3.6.1) \]

for some preassigned probability \( (1 - \alpha) \), regardless of the true value of \( \theta \). Such a set \( C \) is called a set estimate of \( \theta \) with confidence level \( (1 - \alpha) \), or in short, a ‘(1 - \alpha) level confidence set’ of \( \theta \).

If \( \theta \) is real valued (which we will assume throughout this section for simplicity), then we usually prefer the set estimate \( C \) to be an interval of the form \( C = [L, U] \), where \( L = L(X) \) and \( U = U(X) \) are two statistics calculated from \( X = (X_1, X_2, \ldots, X_n)' \) such that

\[ P(L \leq \theta \leq U) \geq (1 - \alpha) \]
\[ (3.6.2) \]

for a preassigned probability \( (1 - \alpha) \) and any \( \theta \in \Theta \). Note that being functions of \( X \), \( L \) and \( U \) are random variables and hence, \([L,U]\) is a
3.6 Set (or Interval) Estimation

random interval. From (3.6.2), it is said that the random interval \([L, U]\) will cover the unknown parameter \(\theta\) with at least probability \((1 - \alpha)\), whatever the true value of \(\theta\) may be.

For instance, let \(\alpha = 0.05\) (or \(1 - \alpha = 0.95\)), we can interpret the meaning of (3.6.2) as the following: let repeated observations \(x\) be taken on the random vector \(X = (X_1, X_2, \ldots, X_n)'\) (representing the outcome of an experiment), and in each case, \(l = L(x)\) and \(u = U(x)\) are determined. Then, in about 95% or more of the cases the interval \((l, u)\) will include \(\theta\), and it will fail to do so in about 5% or less of the cases.

Also, for any interval estimator \(C = [L, U] = [L(X), U(X)]\) of a parameter \(\theta\), the coverage probability of \(C\) is the probability that the random interval covers the true parameter \(\theta\). Usually, the coverage probability of \(C\) is a function of \(\theta\) since the probability distribution of \(X\) depends on \(\theta\). The infimum of the coverage probabilities, i.e.,

\[
\inf_{\theta \in \Theta} P(L(X) \leq \theta \leq U(X))
\]

is called the confidence coefficient of the interval \(C = [L, U]\).

Methods of Finding Set (or Interval) Estimators

Two popular methods of obtaining set or interval estimators are: (a) inverting a test statistic, and (b) making use of pivots.

(a) Inverting a test statistic

There is a strong relationship between set (or interval) estimation and hypothesis testing. Consider testing \(H_0 : \theta = \theta_0\) against \(H_1 : \theta \neq \theta_0\) for any \(\theta_0 \in \Theta\). Let \(A\) be the acceptance region (which depends on \(\theta_0\) apart from \(X\)) of a test with level \(\alpha\). For each \(x \in R_X\), define a set \(C(x) \subseteq \Theta\) such that

\[
C(x) = \{\theta_0 \mid x \in A\}
\]

Then, the random set \(C(X)\) is a \((1 - \alpha)\) level confidence set for \(\theta\).

**Example 3.6.1:** Let \(X_1, X_2, \ldots, X_n\) be iid \(N(\theta, 1)\) and we would like to derive a \((1 - \alpha)\) level confidence interval (CI) for \(\theta\).

First, note that for testing \(H_0 : \theta = \theta_0\) against \(H_1 : \theta \neq \theta_0\), the UMPU test after observing \(X = x\) gives the acceptance region

\[
A = \left\{x \mid |\bar{x} - \theta_0| \leq \frac{1}{\sqrt{n}} z_{(\alpha/2)}\right\}
\]

© 2006 by Taylor & Francis Group, LLC
Chapter 3  Concepts of Statistical Inference

Rewrite \( A \) for \( \theta_0 \) as
\[
A = \left\{ x \mid \bar{x} - \frac{1}{\sqrt{n}} z_{(\alpha/2)} \leq \theta_0 \leq \bar{x} + \frac{1}{\sqrt{n}} z_{(\alpha/2)} \right\}
\]
which gives
\[
C(x) = \left\{ \theta_0 \mid x \in A \right\} = \left[ \bar{x} - \frac{1}{\sqrt{n}} z_{(\alpha/2)}, \bar{x} + \frac{1}{\sqrt{n}} z_{(\alpha/2)} \right]
\]
Thus, \( C(X) = \left[ \bar{X} - \frac{1}{\sqrt{n}} Z_{(\alpha/2)}, \bar{X} + \frac{1}{\sqrt{n}} Z_{(\alpha/2)} \right] \) is a \((1 - \alpha)\) level CI for \( \theta \).

(b) Making use of pivots

A functional \( Q(X, \theta) \) (comprising both data and parameters) is called a pivot if the probability distribution of \( Q(X, \theta) \) is completely known (or free of all parameters).

Since the distribution of \( Q(X, \theta) \) is completely known, it is possible to find a set (known) \( B \) such that \((1 - \alpha) = P(Q(X, \theta) \in B) \). We then find a set estimate \( C(X) \) of \( \theta \) defined as
\[
C(x) = \left\{ \theta \mid Q(x, \theta) \in B \right\} \quad (3.6.4)
\]

Example 3.6.2: Let \( X_1, X_2, \ldots, X_n \) be iid \( N(0, \sigma^2) \). Find a \((1 - \alpha)\) level confidence interval (CI) for \( \sigma^2 \).

The sufficient statistic for \( \sigma^2 \) is \( S = \sum_{i=1}^{n} X_i^2 \), and \( Q(S, \sigma^2) = S/\sigma^2 \)
follows a \( \chi^2_n \)-distribution which is completely known since the sample size \( n \) is known. By denoting the lower and upper \((\alpha/2)\)-cutoff points of a \( \chi^2_n \) distribution by \( \chi^2_{n,(1-\alpha/2)} \) and \( \chi^2_{n,\alpha/2} \), respectively, we have
\[
(1 - \alpha) = P\left(\chi^2_{n,(1-\alpha/2)} \leq \frac{S}{\sigma^2} \leq \chi^2_{n,\alpha/2}\right)
= P\left(\frac{S}{\chi^2_{n,\alpha/2}} \leq \sigma^2 \leq \frac{S}{\chi^2_{n,(1-\alpha/2)}}\right)
\]
Hence, \( \left[ S/\chi^2_{n,\alpha/2}, S/\chi^2_{n,(1-\alpha/2)} \right] \) is a \((1 - \alpha)\) level CI for \( \sigma^2 \).

Remark 3.6.1 There are problems where pivots do not exist, and in such cases, one must invert the acceptance region for testing a hypothesis to derive a set estimate of \( \theta \). For instance, if \( X \) follows Binomial\((n, \theta)\) or Poisson\((\theta)\), it is impossible to find a pivot which can allow us to follow \((3.6.4)\).
Chapter 4

Elements of Decision Theory

4.1 Introduction

In Chapter 3, we have discussed how to estimate a parameter $\theta$ given a set of observations $X = (X_1, X_2, \ldots, X_n)'$ following a joint probability distribution, $f(x|\theta), \theta \in \Theta$. (For the sake of simplicity we assume that the distribution of $X$ depends only on a single, scalar valued parameter $\theta$. Multiparameter cases, i.e., vector valued parameter $\theta$ can be considered suitably if the situation demands.) A point estimate of $\theta$ was denoted by $\hat{\theta}$ or $\hat{\theta}(X)$ (with suitable subscripts to indicate the type of estimator or the method used to derive it). On the other hand, we used the notation $C$ or $C(X) = [L(X), U(X)]$ to denote an interval estimate of $\theta$. But whether we consider point or interval estimation of $\theta$, it is always a tricky issue to evaluate the performance of an estimator. Traditionally, if $\hat{\theta} = \hat{\theta}(X)$ is a point estimate of $\theta$, then we study the bias of $\hat{\theta}$ given as $B(\hat{\theta}) = (E(\hat{\theta}) - \theta)$, and the mean squared error (MSE) of $\hat{\theta}$ given as $\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$. Both $B(\hat{\theta})$ and $\text{MSE}(\hat{\theta})$ are functions of $\theta$. Similarly, if $C = [L(X), U(X)]$ is an interval estimate of $\theta$, then we study Probability Coverage (PC) of $C$ given as $P(C) = P(L(X) \leq \theta \leq U(X))$, and Expected Length (EL) of $C$ given as $\text{EL}(C) = E[U(X) - L(X)]$. (More generally, when we consider a set estimate $C = C(X)$ for a parameter vector $\theta$, EL is replaced by Expected Volume (EV) of the set $C(X)$ given as $\text{EV}(C) =$.
$E[\text{volume of } C].$ Again, both $P(C)$ and $\text{EL}(C)$ are functions of $\theta, \theta \in \Theta.$

In decision theory, not only do we unify both the estimation types (point and interval estimations) under a single framework, but also we are allowed to consider other ways to evaluate estimators and identify the optimal ones. The general framework of decision theory is described below in simple terms.

Based on the data $X = (X_1, X_2, \ldots, X_n)'$ following a probability distribution $f(x|\theta), \theta \in \Theta,$ we take a decision $\delta(X)$ about the parameter $\theta.$ The decision $\delta(X)$ could be a point estimate, or an interval estimate of $\theta,$ or a procedure to test a hypothesis on $\theta.$ Roughly speaking, the decision $\delta(X)$ indicates our ‘understanding’ about the parameter $\theta$ based on the data $X.$

Since the data $X$ gives only a partial information about $\theta,$ our decision $\delta(X)$ is never perfect (in most of the real life problems), and as a result, we incur a loss (due to our imperfect decision) ‘$L$’ which depends on $\delta(X)$ as well as $\theta$ and can be expressed as $L = L(\delta(X), \theta).$ Note that the loss depends on three things: (i) the parameter value $\theta$ (unknown to us), (ii) the random data $X,$ and (iii) the structural form of $\delta.$ It is assumed that $L$ is bounded below by a constant $M,$ and without loss of generality we can take $M = 0,$ i.e., $L(\delta(X), \theta) \geq 0, \forall \delta(X), \theta.$

Since our decision $\delta(X)$ is data dependent, so is the loss $L(\delta(X), \theta).$ Therefore, to evaluate the performance of our decision making process $\delta$ (the functional form), we look at the risk of $\delta$ defined as

$$R(\delta, \theta) = E[L(\delta(X), \theta)] = \int L(\delta(x), \theta)f(x|\theta)dx \quad (4.1.1)$$

The expression $R(\delta, \theta)$ is called the risk function of the decision making process $\delta,$ provided the expectation in (4.1.1) exists. Note that $R(\delta, \theta)$ is nonnegative because the loss is assumed to be so. Suppose that $\mathcal{D}$ denotes the class of all possible decisions about $\theta,$ then it is seen that $R: \mathcal{D} \times \Theta \to \mathbb{R}^+, \text{ where } \mathbb{R}^+ \text{ is the positive part of the real line. The class } \mathcal{D} \text{ has uncountably many elements since it contains all possible real functions of } X.$

\footnote{If $M \neq 0,$ then we can look at a new loss $L^* \text{ defined as } L^* = L - M \geq 0.$}
4.2 Optimality Criteria

Using the risk $R(\delta, \theta)$, we can partially order the elements of $D$. An element $\delta$ of $D$ is called a ‘decision rule’ (or just a ‘rule’).

**Definition 4.2.1** Given two decision rules $\delta_1$ and $\delta_2$,

(a) $\delta_1$ is said to be as good as $\delta_2$ if

$$R(\delta_1, \theta) \leq R(\delta_2, \theta) \quad \forall \theta \in \Theta \quad (4.2.1)$$

(b) $\delta_1$ is said to be better (or uniformly better) than $\delta_2$ if (4.2.1) holds and

$$R(\delta_1, \theta) < R(\delta_2, \theta) \quad \text{for some } \theta \in \Theta$$

**Definition 4.2.2** A decision rule $\delta_0$ is called admissible if there does not exist any other rule $\delta$ which is better than $\delta_0$. A decision rule is called inadmissible if it is not admissible.

Even though admissibility is an optimality criterion, it is rather weak. In order to be admissible all a decision rule $\delta$ needs to possess is ‘very small’ risk at some $\theta \in \Theta$. The Figure 4.2.1 gives a hypothetical example with $D = \{\delta_1, \delta_2, \delta_3, \delta_4\}$ where $\delta_3$ is admissible yet it has ‘high’ risk (poor performance) in most of the parameter space. On the other hand, inadmissibility of a decision rule $\delta_0$ is a strong result since it implies that there exists another decision rule (say, $\delta^*$) which is better than $\delta_0$. In the Figure 4.2.1, $\delta_2$ is inadmissible since $\delta_3$ (also $\delta_1$) is better than $\delta_2$.

Another optimality criterion is derived from **minimax principle** where a rule $\delta$ is judged by its supremum risk (or maximum risk), i.e.,

$$\sup_{\theta \in \Theta} R(\delta, \theta)$$

Note that we can order $D$ fully in terms of supremum risk of the rules.

**Definition 4.2.3** A decision rule $\delta_0$ is said to be minimax if it has the smallest supremum risk in the class of all rules, i.e.,

$$\sup_{\theta \in \Theta} R(\delta_0, \theta) = \inf_{\delta \in D} \sup_{\theta \in \Theta} R(\delta, \theta) \quad (4.2.2)$$

The minimax approach is like guarding against the worst possibility, but again it can result into an ‘overkill’ as the Figure 4.2.2 illustrates. Here $\delta_3$ is minimax even though it may not be too appealing.

The other popular criterion which again helps us order $D$ fully is based on the **Bayes’ principle**. Traditionally the parameter $\theta$ is believed to be
completely unknown. The way it is understood is that the nature selects \( \theta \) (unknown to us) from \( \Theta \) following a probability distribution \( \pi(\theta) \) (on \( \Theta \)), and the probability distribution of \( X \), i.e., \( f(x|\theta) \), depends on \( \theta \) thus selected. The probability distribution \( \pi(\theta) \) of \( \theta \) is called a prior distribution.

As \( \theta \) is looked upon as a random variable, the expected value of \( R(\delta, \theta) \), for variation of \( \theta \), will be a guiding principle in evaluating a rule \( \delta \). For a rule \( \delta \), its Bayes risk with respect to (w.r.t.) prior \( \pi(\theta) \),
Figure 4.2.2: Risk Curves of $\delta_1$, $\delta_2$, and $\delta_3$.

$\theta \in \Theta$, is defined as

$$r(\delta, \pi) = E\left[R(\delta, \theta)\right] = \begin{cases} 
\int_{\Theta} R(\delta, \theta)\pi(\theta) \, d\theta, & \text{if } \pi \text{ is continuous} \\
\sum_{\theta \in \Theta} R(\delta, \theta)\pi(\theta), & \text{if } \pi \text{ is discrete} 
\end{cases}$$

(4.2.3)

**Definition 4.2.4** A decision rule $\delta_\pi$ is called a Bayes rule w.r.t. prior $\pi(\theta)$ if

$$r(\delta_\pi, \pi) \leq r(\delta, \pi) \text{ for all } \delta \in D$$

(4.2.4)
The concept of a prior distribution of an unknown parameter has always been debatable as whether this could be a practical approach in reality. Example 4.2.1 illustrates the usefulness of the Bayes principle.

**Example 4.2.1:** Consider the case of lot sampling where an inspector has to decide whether to accept or reject a lot of known size $N$ of items (may be manufactured or agricultural products) by observing the number of defective items in a random sample of size $n$ from the lot. The actual number of defective items in the whole lot, say $K$, is unknown to the inspector.

The observable random variable $X$, the number of defective items in a sample of size $n$, has a hypergeometric distribution with pmf

$$f(x|K) = \binom{K}{x} \binom{N-K}{n-x} \binom{N}{n}, \quad x = 0, 1, 2, \ldots, n \quad (4.2.5)$$

It is assumed that $n$ is a small integer compared to $N$ and $K$. But if the original lot is formed by taking $N$ items at random from a production line then $K$, the number of defective items, would itself be a random variable. Assuming that the production process yields a proportion $p$ of defective items, it is reasonable to take a binomial prior distribution of $K$ as

$$\pi(k) = P(K = k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad k = 0, 1, 2, \ldots, N \quad (4.2.6)$$

where $p$ may be known from past experience.

The prior distribution may in turn depend on other parameters, known as super or hyper parameters, which may or may not be known from past experience. In case these super parameters are unknown (i.e., we just know the structural form of the prior distribution), then they can be estimated in the Empirical Bayes setup (not discussed here).

### 4.3 Loss Functions

In a point estimation problem, the loss function $L(\delta(X), \theta)$ should be ‘small’ when $\delta(X)$ is close to $\theta$, and the loss generally increases as $\delta(X)$ deviates more from $\theta$. For a real valued parameter $\theta$, let

$$\Delta = \delta(X) - \theta \quad (4.3.1)$$
denote the error in estimating $\theta$ by $\delta(X)$. Usually, $L(\delta(X), \theta)$ is taken as a convex function, say $L(\Delta)$, of $\Delta$. Two commonly used loss functions are

**Absolute Error Loss (AEL):** $L(\delta(X), \theta) = |\Delta|$ \hspace{1cm} (4.3.2)

and

**Squared Error Loss (SEL):** $L(\delta(X), \theta) = \Delta^2$ \hspace{1cm} (4.3.3)

For point estimation of a parameter vector $\theta = [\theta_1, \theta_2, \ldots, \theta_k]'$ by a decision rule $\delta(X) = [\delta_1(X), \delta_2(X), \ldots, \delta_k(X)]'$, the loss functions AEL and SEL can be generalized as

$$
\sum_{i=1}^{k} |\delta_i(X) - \theta_i| \quad \text{and} \quad \sum_{i=1}^{k} (\delta_i(X) - \theta_i)^2,
$$

respectively.

A variation of SEL, one that penalizes over and under estimations differently, is reflected through the loss function

$$
L(\delta(X), \theta) = \begin{cases} K_1 \Delta^2 & \text{if } \delta(X) \geq \theta \\ K_2 \Delta^2 & \text{if } \delta(X) < \theta \end{cases} \hspace{1cm} (4.3.4)
$$

where $K_1$ and $K_2$ ($K_1 \neq K_2$) are known constants. Another loss function which penalizes over and under estimations differently and has applications in real estate market is the LINEX$^2$ (Varian (1975)) loss function given as

$$
L(\delta(X), \theta) = b \exp(a \Delta) - c \Delta - b \hspace{1cm} (4.3.5)
$$

where $a, c \neq 0, b > 0$ are known suitable constants. Note that in (4.3.5), $L(\delta(X), \theta) = L(\Delta)$ and $L(0) = 0$. For a minimum to exist at $\Delta = 0$, we must have $ab = c$. Thus, (4.3.5) can be rewritten as

$$
L(\delta(X), \theta) = L(\Delta) = b \{ \exp(a \Delta) - a \Delta - 1 \} \hspace{1cm} (4.3.6)
$$

For $a > 0$, the loss (4.3.6) is asymmetric with overestimation being more costly than underestimation and vice-versa for $a < 0$.

For a nonnegative scalar valued parameter $\theta$ (for example, when $\theta$ is a scale parameter), two popular loss functions are

**Squared Log Error (SLE):** $L(\delta(X), \theta) = (\ln \delta(X) - \ln \theta)^2$ \hspace{1cm} (4.3.7)

and

**Entropy Loss (EL):** $L(\delta(X), \theta) = \ln(\delta(X)/\theta) - (\delta(X)/\theta) - 1$ \hspace{1cm} (4.3.8)

\footnote{It is abbreviated from ‘Linear-Exponential’}
where \( \delta(X), \theta > 0 \).

In testing a hypothesis \( H_0 : \theta \in \Theta_0 \) against \( H_1 : \theta \in \Theta_0^c \), any decision rule \( \delta(X) \) can take two possible values, \( a_0 = \) ‘accept \( H_0 \)’ and \( a_1 = \) ‘accept \( H_1 \)’. The sets \( A = \{ x | \delta(x) = a_0 \} \) and \( R = \{ x | \delta(x) = a_1 \} \) are respectively the acceptance and rejection regions of a rule \( \delta(X) \).

A loss function in a hypothesis testing problem should reflect the errors if \( \theta \in \Theta_0 \) and \( \delta(x) = a_1 \), or if \( \theta \in \Theta_0^c \) and \( \delta(x) = a_0 \). The simple loss, known as ‘0-1 loss’, which takes into account such errors (of wrong decisions) is given as

\[
L(\delta(x), \theta) = \begin{cases} 
0 & \text{if } \theta \in \Theta_0 \text{ and } \delta(x) = a_0 \text{ or } \theta \in \Theta_0^c \text{ and } \delta(x) = a_1 \\
1 & \text{if } \theta \in \Theta_0 \text{ and } \delta(x) = a_1 \text{ or } \theta \in \Theta_0^c \text{ and } \delta(x) = a_0
\end{cases}
\] (4.3.9)

The risk of any decision rule \( \delta(X) \) under the loss (4.3.9) is given as

\[
R(\delta, \theta) = \begin{cases} 
0 \cdot P(\delta(X) = a_0) + 1 \cdot P(\delta(X) = a_1) & \text{if } \theta \in \Theta_0 \\
0 \cdot P(\delta(X) = a_1) + 1 \cdot P(\delta(X) = a_0) & \text{if } \theta \in \Theta_0^c
\end{cases}
\]

which is closely related to the concept of power function described in (3.5.2) and this is due to the special nature of the 0-1 loss function which judges whether a decision is right or wrong.

A more general loss function for testing \( H_0 \) against \( H_1 \) can be given as

\[
L(\delta(x), a_0) = \begin{cases} 
0 & \text{if } \theta \in \Theta_0 \\
L(a_0, \theta) & \text{if } \theta \in \Theta_0^c
\end{cases}
\]

(4.3.11)

\[
L(\delta(x), a_1) = \begin{cases} 
L(a_1, \theta) & \text{if } \theta \in \Theta_0 \\
0 & \text{if } \theta \in \Theta_0^c
\end{cases}
\]

where both \( L(a_0, \theta) \) and \( L(a_1, \theta) \) are suitable nonnegative scalar values.

The risk of a rule \( \delta(X) \) under (4.3.11) is

\[
R(\delta, \theta) = L(a_0, \theta)P(\delta(X) = a_0) + L(a_1, \theta)P(\delta(X) = a_1)
\]

\[
= L(a_0, \theta)(1 - \beta(\theta)) + L(a_1, \theta)\beta(\theta)
\]

(4.3.12)

where \( \beta(\theta) = P(\delta(X) = a_1) \) is the power function based on the decision rule \( \delta \).
The loss function in an interval estimation problem usually has two components: (i) a measure of whether the interval estimate correctly includes the true value $\theta$, and (ii) a measure of the size of the index estimate. Let $C = [L(X), U(X)]$ be an interval estimate of a scalar valued parameter $\theta$. Then the length of $C$, say $l(C)$, is called a measure of size, and the indicator function,

$$I_C(\theta) = \begin{cases} 
1 & \text{if } \theta \in C \\
0 & \text{if } \theta \notin C 
\end{cases}$$

expresses the correctness measure of $C$ in containing $\theta$.

The loss function should reflect the fact that a good interval estimate $C = [L(X), U(X)]$ should have smaller $l(C)$ and large correctness $I_C(\theta)$. One such loss function for interval estimation of $\theta$ by $C$ is given as

$$L(C, \theta) = a l(C) - b I_C(\theta) \quad (4.3.13)$$

where $a$ and $b$ are suitable positive constants reflecting the relative weights of the two components of the loss function. Using the loss (4.3.13) one can get the associated risk function as

$$R(C, \theta) = a E[l(C)] - b P(\theta \in C) \quad (4.3.14)$$

which also has two components: (i) the expected length of the interval, and (ii) the coverage probability of the interval.

In the multiparameter case, when $\theta$ is vector valued and $C$ is a set estimate of $\theta$, we replace $l(C)$ by $v(C)$, the volume of the set $C = C(X)$.

## 4.4 Admissible, Minimax, and Bayes Rules

We start this section by studying the structure of a Bayes rule first. The risk function of a decision rule $\delta$ is

$$R(\delta, \theta) = \int_{R_x} L(\delta(x), \theta) f(x | \theta) \, dx$$

and the Bayes risk of $\delta$ w.r.t. a prior distribution $\pi(\theta)$ of $\theta$ over $\Theta$ is

$$r(\delta, \pi) = \int_\Theta \int_{R_x} L(\delta(x), \theta) f(x | \theta) \pi(\theta) \, dx \, d\theta \quad (4.4.1)$$
We assume that the functions $L(\delta(x), \theta)$ and $f(x \mid \theta)$ are such that $r(\delta, \pi)$ admits a change in the order of integration in (4.4.1), so that we have

$$r(\delta, \pi) = \int_{\mathbb{R}_X} \int_\Theta L(\delta(x), \theta)f(x \mid \theta)\pi(\theta)\,d\theta \,dx \quad (4.4.2)$$

In a Bayesian setup where $\theta$ is treated as a variable with a (marginal) distribution $\pi(\theta)$, the probability distribution $f(x \mid \theta)$, in true sense, is the conditional distribution of $X \mid \theta$. Thus the joint distribution of $(X, \theta)$ is

$$m(x, \theta) = f(x \mid \theta)\pi(\theta) \quad (4.4.3)$$

and the conditional distribution of $\theta$ given $X = x$, called the posterior distribution of $\theta$ given $X = x$, is

$$\pi(\theta \mid x) = \frac{m(x, \theta)}{m(x)} \quad (4.4.4)$$

where $m(x) = \int_\Theta m(x, \theta)\,d\theta$. Using (4.4.3) and (4.4.4), $r(\delta, \pi)$ can be rewritten as

$$r(\delta, \pi) = \int_{\mathbb{R}_X} \int_\Theta L(\delta(x), \theta)\pi(\theta \mid x)m(x)\,d\theta \,dx$$

$$= \int_{\mathbb{R}_X} \left\{ \int_\Theta L(\delta(x), \theta)\pi(\theta \mid x)\,d\theta \right\}m(x)\,dx$$

$$= \int_{\mathbb{R}_X} \left\{ \mathbb{E}_\theta \left[ L(\delta(x), \theta) \mid \theta \sim \pi(\theta \mid x) \right] \right\}m(x)\,dx \quad (4.4.5)$$

The inner expectation in (4.4.5), or

$$\mathbb{E}_\theta \left[ L(\delta(x), \theta) \mid \theta \sim \pi(\theta \mid x) \right] \quad (4.4.6)$$

is called the **Bayes Expected Loss** (BEL) when the posterior distribution of $\theta$ is $\pi(\theta \mid x)$ for given $X = x$.

Suppose for each $X = x$ in $\mathbb{R}_X$ there exists an element $\delta_x(x)$ such that

$$\mathbb{E}_\theta \left[ L(\delta(x), \theta) \mid \theta \sim \pi(\theta \mid x) \right] = \inf_{\delta(x)} \mathbb{E}_\theta \left[ L(\delta(x), \theta) \mid \theta \sim \pi(\theta \mid x) \right] \quad (4.4.7)$$

Then the function $\delta_x(x), x \in \mathbb{R}_X$, minimizes (4.4.5), and hence $\delta_x$ is the Bayes rule w.r.t. $\pi$. In other words, for fixed $X = x$, a Bayes rule is a value $\delta_x(x)$ of $\delta(x)$ that minimizes the BEL.
4.4 Admissible, Minimax, and Bayes Rules

Theorem 4.4.1 For point estimation of a real valued parameter \( \theta \), the Bayes rule \( \delta_\pi(X) \) is:

(a) \( \delta_\pi(X) = E[\theta \mid \theta \sim \pi(\theta \mid X)] \) if the loss is SEL (4.3.3)

(b) \( \delta_\pi(X) = \text{Median}(\theta \mid \theta \sim \pi(\theta \mid X)) \) if the loss is AEL (4.3.2)

To prove the above theorem, it is easy to note that the BEL under SEL is the MSE of \( \theta \) where \( \theta \) follows \( \pi(\theta \mid X) \). The MSE of \( \theta \) is minimized when \( \delta(x) = E[\theta \mid \theta \sim \pi(\theta \mid x)] \). Similarly, the BEL under AEL is nothing but the \( E[\theta - \delta(X)] \) where \( \theta \) follows \( \pi(\theta \mid x) \), and is minimized when \( \delta(x) = \text{median of } \pi(\theta \mid x) \).

In many real life problems where the loss functions may not be as simple as AEL or SEL, and may reflect complex monetary or other resource losses, the minimization in (4.4.7) may not be possible analytically. But the integration can be evaluated and minimization can be carried out numerically. It is seen that after observing \( X = x \), we need to solve (4.4.7) only for the value \( X = x \), and need not be concerned with what the value of \( \delta_\pi(x^*) \) would be for \( X = x^* \) where \( x^* \) is different from \( x \).

Example 4.4.1: Let \( X_1, X_2, \ldots, X_n \) be iid \( N(\theta, \sigma^2) \) and let the prior distribution of \( \theta \), \( \pi(\theta) \), be \( N(\mu, \tau^2) \), where \( \mu \), \( \tau^2 \) and \( \sigma^2 \) are known. Our goal is to estimate \( \theta \).

The problem can be reduced first by sufficiency. It is known that all the relevant information concerning \( \theta \) is contained in \( \bar{X} \) which follows \( N(\theta, \sigma^2/n) \).

Following the steps in (4.4.3) and (4.4.4) it can be shown that for \( \bar{X} = \bar{x} \) the posterior distribution of \( \theta \) is \( \pi(\theta \mid \bar{x}) = N(\eta, \delta^2) \) where

\[
\eta = E[\theta \mid \bar{x}] = \frac{\tau^2}{\tau^2 + (\sigma^2/n)} \bar{x} + \frac{\sigma^2/n}{\tau^2 + (\sigma^2/n)} \mu
\]

and

\[
\delta^2 = \text{Var}(\theta \mid \bar{x}) = \frac{\tau^2 \sigma^2/n}{\tau^2 + (\sigma^2/n)}
\]

Since the posterior distribution is normal which is symmetric about its mean (= median), the Bayes rule under SEL or AEL is \( \delta_\pi(\bar{X}) = \eta \) (as given above).

At this point we need to add a few more words about Bayes rules. Usually the prior distribution \( \pi(\theta) \) is a probability distribution with \( \int_{\Theta} \pi(\theta) d\theta = 1 \). This guarantees that the posterior \( \pi(\theta \mid x) \) is also a probability distribution since it is a conditional distribution of \( \theta \) given \( x \). But
the existence of the posterior distribution is not conditional on \( \pi(\theta) \) being a probability distribution. In other words, it is possible to have \( \pi(\theta) \) with \( \int_{\Theta} \pi(\theta) \, d\theta = \infty \), yet the posterior \( \pi(\theta | x) \) is a probability distribution. Such a prior \( \pi(\theta) \) is called an improper or generalized prior, and the resultant Bayes rule is called an improper or generalized Bayes rule. A prior \( \pi(\theta) \) with \( \int_{\Theta} \pi(\theta) \, d\theta < \infty \) is called a proper prior and the corresponding Bayes rule is called a Proper Bayes or just a Bayes rule.

We now turn our attention to the optimality criterion of admissibility. We shall explore the connection between admissibility of a rule and its being Bayes (w.r.t. some prior distribution) while answering the following two questions:

(Q.1) For a given problem, how to find an admissible rule? and
(Q.2) Given a rule, how to check its admissibility?

The following theorem which answers above (Q.1) states that Bayes rules are generally admissible.

**Theorem 4.4.2** A proper Bayes rule is admissible if it is unique except for risk equivalence. (Two rules are called equivalent if they have identical risk functions.)

Theorem 4.4.2 can be used as a tool to generate admissible decision rules. As long as the proper Bayes rule \( \delta_{\pi} \) (w.r.t. prior \( \pi(\theta) \)) is unique, it is admissible irrespective of the fact that the prior distribution \( \pi(\theta) \) may or may not reflect the true nature of the parameter \( \theta \).

**Example 4.4.2:** Let \( X_1, X_2, \ldots, X_n \) be iid Bernoulli(\( \theta \)), \( \theta \in (0, 1) \). \( X = \sum_{i=1}^{n} X_i \) is the minimal sufficient statistic for \( \theta \), and \( X \) follows \( \text{B}(n, \theta) \), \( R_X = \{0, 1, \ldots, n\} \). Suppose that we would like to find an admissible rule for \( \theta \). Take a prior \( \pi(\theta) \) of \( \theta \) as a Beta(\( \alpha, \beta \)) \( (\alpha > 0, \beta > 0) \) distribution. It can be shown that the posterior distribution \( \pi(\theta | X) \) is Beta(\( X + \alpha, n - X + \beta \)) distribution. Therefore, the unique Bayes rule under SEL is \( \delta_{\pi}^{\alpha,\beta}(X) = E(\theta | X) = (X + \alpha) / (\alpha + \beta + n) \). By Theorem 4.4.2, \( \delta_{\pi}^{\alpha,\beta}(X) \) is admissible for \( \theta \) under SEL. In fact, we have achieved more; we have just produced a class of admissible rules,

\[
\mathcal{D}_{\alpha,\beta} = \left\{ \delta_{\pi}^{\alpha,\beta}(X) = (X + \alpha) / (\alpha + \beta + n) \mid \alpha > 0, \beta > 0 \right\}
\]
4.4 Admissible, Minimax, and Bayes Rules

since \( \delta_\alpha^\alpha(X) \) is admissible for every fixed pair of \( \alpha > 0, \beta > 0 \).

Interestingly, in Example 4.4.2, the class of admissible rules, \( D_{\alpha,\beta} \), does not contain the traditional estimator \( \delta_0 = X/n \) which is UMVUE as well as MLE. The only way \( \delta_\pi^{\alpha,\beta}(X) = \delta_0 \) if and only if \( \alpha = \beta = 0 \). But \( \text{Beta}(\alpha, \beta) \) is not a probability distribution for \( \alpha = \beta = 0 \). So the question is: “Does there exist a prior probability distribution for which \( \delta_0(X) \) is a Bayes rule?”

More generally, this brings out our question (Q.2) that is “Given a rule, how do we check its admissibility?” We first shall look at a more special case that deals with unbiased estimators and SEL.

**Theorem 4.4.3** Under SEL, no unbiased estimator can be a proper Bayes rule.

Theorem 4.4.3 states that there does not exist any prior (proper) \( \pi(\theta) \) for which an unbiased estimator can be made as a proper Bayes rule under SEL. Therefore, in Example 4.4.2, no proper prior can make \( \delta_0(X) = X/n \) a proper Bayes rule under SEL, and hence the admissibility of \( \delta_0(X) \) under SEL can not be answered by Theorem 4.4.2.

However, note that, if we consider an improper prior \( \pi_\star(\theta) \propto \frac{\theta(\theta - 1)}{-1} \), then \( \delta_0(X) = X/n \) is improper (or generalized Bayes) w.r.t. \( \pi_\star(\theta) \). But admissibility of improper or generalized rules is not guaranteed always. Also, by choosing \( \alpha \) and \( \beta \) very close to 0 (yet \( > 0 \)), we can make the proper prior \( \text{Beta}(\alpha, \beta) \) arbitrarily close to above \( \pi_\star(\theta) \propto \frac{\theta(\theta - 1)}{-1} \). At the same time, the proper Bayes rule \( \delta_\pi^{\alpha,\beta}(X) \) can be made arbitrarily close to \( \delta_0(X) \).

**Definition 4.4.1** Let \( \{\pi_k(\theta)\}_{k \geq 1} \) be a sequence of proper priors and \( \{\delta^k_\pi(X)\} \) be the corresponding proper Bayes rules (under some loss function).

(a) A rule \( \delta(X) \) is said to be a limit of Bayes rules \( \{\delta^k_\pi(X)\} \), hence \( \delta(X) \) is called a ‘limiting Bayes’ rule, if for almost all \( x \), \( \delta^k_\pi(x) \to \delta(x) \) as \( k \to \infty \).

(b) A rule \( \delta(X) \) is said to be Bayes ‘in the wide sense’ provided

\[
\lim_{k \to \infty} \int \frac{r(\delta^k_\pi, \pi_k) - r(\delta, \pi_k)}{\int_0^1 \frac{\theta(\theta - 1)}{-1} d\theta} = 0
\]

\( \int_0^1 \frac{\theta(\theta - 1)}{-1} d\theta = \infty. \)
(Note that a proper Bayes rule is necessarily a Bayes rule in the wide sense.)

**Theorem 4.4.4** Under some mild regularity conditions, a decision rule \( \delta \) is admissible if and only if there exists a sequence \( \{ \pi_k \} \) of prior distributions (possibly improper) such that:

(i) Each \( \pi_k \) gives mass only to a closed and bounded subset of \( \Theta \) (possibly different for each \( k \)), and hence has finite total mass

(ii) There is a closed and bounded set \( \Theta_0 \subset \Theta \) to which each \( \pi_k \) gives mass 1

(iii) \( \lim_{k \to \infty} \left[ r(\delta^k_{\pi}, \pi_k) - r(\delta, \pi_k) \right] = 0 \), where \( \delta^k_{\pi} \) is the Bayes rule w.r.t \( \pi_k \)

The regularity conditions under which Theorem 4.4.4 holds include the continuity of the risk function and those needed for Cramér-Rao inequality. Also, the above condition (ii) says that an admissible rule is essentially Bayes ‘in the wide sense’. The next theorem is a weaker version of Theorem 4.4.4 but easy to implement for checking the admissibility of a given rule.

**Theorem 4.4.5** Consider a decision problem where \( \Theta \) is a nondegenerate convex subset. Assuming that risk is continuous in \( \theta \), a decision rule \( \delta \) is admissible if there exists a sequence \( \{ \pi_k \} \) of improper priors such that:

(i) The Bayes risks \( r(\delta, \pi_k) \) and \( r(\delta^k_{\pi}, \pi_k) \) are finite for all \( k \), where \( \delta^k_{\pi} \) is the Bayes rule w.r.t \( \pi_k \)

(ii) For any nondegenerate convex subset \( \Theta_0 \subset \Theta \), there exists a \( K > 0 \) and an integer \( k_0 \) such that for \( k \geq k_0 \), \( P_{\pi_k}(\theta \in \Theta_0) \geq K \)

(iii) \( \lim_{k \to \infty} \left[ r(\delta^k_{\pi}, \pi_k) - r(\delta, \pi_k) \right] = 0 \)

While one can use the last two general theorems to prove admissibility of a given rule for a specific problem, there are other techniques available for special cases. One such technique is the use of Cramér-Rao inequality to prove admissibility only under squared error loss (SEL) as shown in the next example.
4.4 Admissible, Minimax, and Bayes Rules

Example 4.4.3: Without loss of generality, assume that $X$ follows $N(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$. Then, under SEL, $\delta_0(X) = X$ is an admissible rule for $\theta$.

We prove the admissibility of $\delta_0(X) = X$ by the method of contradiction using Cramér-Rao inequality.

Suppose $\delta_0$ is not admissible (i.e., inadmissible), then there exists another rule $\delta_1(X) \neq \delta_0(X)$ such that

$$1 = E[(\delta_0(X) - \theta)^2] \geq E[(\delta_1(X) - \theta)^2], \quad \forall \theta \in \mathbb{R} \quad (4.4.8)$$

with strict inequality for some $\theta$. But by Cramér-Rao inequality,

$$E[(\delta_1(X) - \theta)^2] \geq b^2(\theta) + (1 + b'(\theta))^2 \quad (4.4.9)$$

Since Fisher information $I(\theta) = 1$, and $b(\theta) = E[\delta_1(X) - \theta] = \text{the bias of } \delta_1(X)$, combining (4.4.8) and (4.4.9), we get

$$b^2(\theta) + (1 + b'(\theta))^2 \leq 1, \quad \forall \theta \in \mathbb{R} \quad (4.4.10)$$

with strict inequality for some $\theta$. It is seen from (4.4.10) that $b'(\theta) \leq 0 \quad \forall \theta$. In other words, the bias function $b(\theta)$ of $\delta_1(X)$ is decreasing in $\theta$. Note that $b(\theta)$ is not identically equal to 0. (Otherwise, $\delta_1(X)$ would be an unbiased estimator of $\theta$, and by completeness of $N(\theta, 1)$ distribution, $\delta_1(X)$ would be identical to $\delta_0(X)$ which would lead to a contradiction.) So, there must exist $\theta_0 \in \mathbb{R}$ such that $b(\theta_0) \neq 0$.

Case-(i): Let $b(\theta_0) > 0$. So $b(\theta) > 0, \forall \theta < \theta_0$, since $b(\theta)$ is a decreasing function of $\theta$. From (4.4.10), we have that

$$b^2(\theta) + 2b'(\theta) \leq 0, \quad \forall \theta \in \mathbb{R}$$

then

$$1 \leq -2 \frac{b'(\theta)}{b^2(\theta)}, \quad \forall \theta < \theta_0$$

$$\int_\theta^{\theta_0} d\theta \leq -2 \int_\theta^{\theta_0} \frac{b'(\theta)}{b^2(\theta)} d\theta \quad \text{for any } \theta < \theta_0$$

or

$$\left(\theta_0 - \theta\right) \leq 2\left(\frac{1}{b(\theta_0)} - \frac{1}{b(\theta)}\right) \quad \text{for any } \theta < \theta_0 \quad (4.4.11)$$

By letting $\theta \to -\infty$, we have the left hand side of (4.4.11) approaching to $\infty$ whereas the right hand side is finite which leads to a contradiction. Hence, we must obtain that $b(\theta_0) \leq 0$. 

© 2006 by Taylor & Francis Group, LLC
Case-(ii): Let \( b(\theta_0) < 0 \). Then \( b(\theta) < 0, \forall \theta > \theta_0 \), because \( b(\theta) \) is a decreasing function of \( \theta \). Similar to the argument of Case-(i), it can be shown that \( b(\theta_0) \geq 0 \).

Combining the both cases (i) and (ii), it is true that \( b(\theta) \equiv 0 \), or \( E[\delta_1(X) - \delta_0(X)] = 0 \), leading to a contradiction since \( N(\theta, 1), \theta \in \mathbb{R} \) is a complete family of distributions. Hence, \( \delta_0(X) = X \) is admissible under SEL.

**Remark 4.4.1** Using the technique of Cramér-Rao lower bound, one can similarly prove admissibility of \( \delta_0 \) under SEL for the following distributions: (a) \( X \) follows Poisson(\( \lambda \)) and \( \delta_0 = X \), (b) \( X \) follows \( B(n, \theta) \) and \( \delta_0(X) = X/n \).

Now we turn our attention to minimaxity. While we explore the connection between minimaxity and Bayes principle, we address the following two questions:

\((Q.3)\) For a given problem, how to find a minimax rule?

\((Q.4)\) Given a decision rule, how to check its minimaxity?

The following theorem is widely used to construct minimax rules and it partially answers \((Q.3)\).

**Theorem 4.4.6** An admissible equalizer rule is minimax.

Note that a unique proper Bayes rule is admissible, and hence a unique proper Bayes equalizer rule is minimax.

Recall that an equalizer rule is the one with constant risk. To construct an admissible rule, we first take a suitable (preferably conjugate\(^4\)) family of proper prior distributions, and derive the Bayes rule and its risk. We then choose super parameters (parameters involved in the prior distributions) such that the risk of the Bayes rule turns out to be a constant. The next example illustrates the application of Theorem 4.4.6.

**Example 4.4.4:** Let \( X \) follow \( B(n, \theta) \). Find a minimax rule for \( \theta \) under SEL.

Consider a Beta prior family of distributions,

\[
\pi(\theta) = \theta^{a-1}(1 - \theta)^{b-1}/B(a, b), \quad a > 0, b > 0
\]

\(^4\)A family of prior distributions is called a **conjugate family** for which the posterior distribution has the same structure as the prior.
The posterior distribution of \( \theta \), given \( X = x \), is
\[
\pi(\theta|x) \propto \theta^{a+x-1}(1-\theta)^{n-x+b-1}
\]
and hence the Bayes rule under SEL is
\[
\delta_\pi(X) = \frac{a + X}{a + b + X}
\]
We want to choose \( a \) and \( b \) (both are positive) such that \( \delta_\pi \) has a constant risk. First we derive the regular risk of \( \delta_\pi \) under SEL.
\[
R(\delta_\pi, \theta) = E \left[ (\delta_\pi - \theta)^2 \right] = E \left[ \left( \frac{a + X}{a + b + n} \right)^2 \right] - 2\theta E \left[ \frac{a + X}{a + b + n} \right] + \theta^2
\]
\[
= \frac{a^2 + 2na\theta + n\theta(1-\theta) + n^2\theta^2}{(a + b + n)^2} - \frac{2\theta(a + n\theta)}{a + b + n} + \theta^2
\]
\[
= K_0 + K_1\theta + K_2\theta^2
\]
where \( K_0, K_1, \) and \( K_2 \) are suitable constants depending only on \( a, b, \) and \( n \). So it is seen that the risk of \( \delta_\pi \) is a constant when \( K_1 = K_2 = 0 \).
Equating the explicit expressions of \( K_1 \) and \( K_2 \) with 0 and then solving for \( a \) and \( b \), we get \( a = b = \sqrt{n}/2 \). Therefore, a minimax estimator of \( \theta \) under SEL is \( \delta^*_\pi \) (say), where
\[
\delta^*_\pi(X) = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}
\]
A natural question thus arises is that ‘Is the usual estimator, \( \delta_0(X) = X/n \), which is UMVUE, MLE, and admissible under SEL, minimax under SEL?’

By plugging in \( a = b = \sqrt{n}/2 \) in the expression (4.4.12), we get the risk of \( \delta^*_\pi \) as
\[
R(\delta^*_\pi, \theta) = \frac{1}{4(1 + \sqrt{n})^2}
\]
On the other hand,
\[
R(\delta_0, \theta) = \frac{\theta(1-\theta)}{n}
\]
and it is seen that (Figure 4.4.1)
\[
\frac{1}{4n} = \sup_\theta R(\delta_0, \theta) > \sup_\theta R(\delta^*_\pi, \theta) = \frac{1}{4(1 + \sqrt{n})^2}
\]
Figure 4.4.1: Risk Curves of $\delta_0$ and $\delta^*_\pi$.

So, $\delta_0$ is not minimax for $\theta$ under SEL. Even though the difference between the supremum risks in (4.4.14) and (4.4.15) is negligible for large $n$, it still can be substantial for small $n$.

**Remark 4.4.2** For the above binomial example, if one uses the weighted squared error loss $L_*(\delta, \theta) = (\delta - \theta)^2 / (\theta(1 - \theta))$, then $\delta_0 = X/n$ is still admissible under $L_*$. Under $L_*$, the risk of $\delta_0$ is a constant $= 1/n$, hence, $\delta_0$ is an admissible equalizer rule under $L_*$. By Theorem 4.4.6, this rule, $\delta_0$, is minimax under $L_*$.

Even though in the Binomial case, the usual estimator $\delta_0 = X/n$ (admissible, MLE, UMVUE) is not minimax under SEL, many times the
4.4 Admissible, Minimax, and Bayes Rules

standard estimators are indeed minimax under common loss functions.
The following theorem is useful to prove minimaxity of a given rule and partially answers the above question (Q.4).

**Theorem 4.4.7** Let \( \{\pi_k(\theta)\}_{k \geq 1} \) be a sequence of proper priors, and \( \{\delta_k^X(X)\} \) be the corresponding proper Bayes rules (under some loss function). Suppose the Bayes risk \( r(\delta_k^X, \pi_k) \) of \( \delta_k^X \) w.r.t. \( \pi_k \) converges to a constant \( c \) as \( k \to \infty \). If \( \delta_0(X) \) is a decision rule with \( R(\delta_0, \theta) \leq c \) for all \( \theta \in \Theta \), then \( \delta_0 \) is minimax.

**Example 4.4.5:** As an application of Theorem 4.4.7, we can prove that the sample mean of \( iid \) observations with a normal distribution is minimax under SEL.

For the sake of simplicity, let us assume that we have a single observation \( X \) following \( N(\theta, 1) \). The usual estimator of \( \theta \) is \( \delta_0(X) = X \) (MLE, UMVUE, and admissible under SEL). Suppose now we take a sequence of priors \( \{\pi_k\} \) such that \( \pi_k(\theta) = N(0, k) \) distribution. Under SEL, the Bayes rule corresponding to \( \pi_k \) is \( \delta_k^X(X) = \frac{k}{k+1} X \) with Bayes risk \( r(\delta_k^X, \pi_k) = \frac{k}{k+1} \). The limiting Bayes risk is thus

\[
\lim_{k \to \infty} r(\delta_k^X, \pi_k) = 1 = R(\delta_0, \theta) \quad \forall \theta
\]

By Theorem 4.4.7, \( \delta_0 \) is minimax under SEL.

Using Theorem 4.4.7, one can find a minimax interval estimator for \( \theta \) under a loss function when \( X \) follows \( N(\theta, 1) \) as illustrated below.

For an interval estimate \( C = C(X) \) of \( \theta \) under a loss \( L(C, \theta) = al(C) - bI_C(\theta) \) with suitable \( a \) and \( b \), we consider the proper prior \( \pi(\theta) = N(0, k) \). The posterior distribution \( \pi(\theta|X) \) is \( N\left(\frac{k}{k+1} X, \frac{k}{k+1}\right) \). The Bayes rule w.r.t. \( \pi(\theta) \) which minimizes the Bayes expected loss (BEL) has the form

\[
\text{BEL} = aE_{\pi(\theta|X)}[l(C)] - bE_{\pi(\theta|X)}[I_C(\theta)]
\]

Since \( l(C) \), the length of \( C \), is a constant in \( \theta \), we can obtain

\[
\text{BEL} = a \cdot l(C) - b \cdot P_{\pi(\theta|X)}(\theta \in C)
\]

where the second term in the above BEL indicates a posterior probability coverage of \( C \). The above BEL is minimized if and only if the posterior
Chapter 4 Elements of Decision Theory

probability coverage of $C$ is maximized and that too happens when $C$ has the form: $C(X) = \frac{k}{k+1} X \pm d$. Thus

$$\text{BEL} = 2ad - 2b \cdot P\left(0 < Z < d/\sqrt{k/(k+1)} \right)$$

(4.4.16)

where $Z$ is a standard normal random variable. One can now minimize the $\text{BEL}$ (4.4.16) to obtain an optimal $d$, say $d_k^\pi$. Thus, the Bayes interval estimate for $\theta$ (w.r.t. $\pi(\theta)$) is $C_k^\pi(X) = \frac{k}{k+1} X \pm d_k^\pi$. Note that the interval $C_0(X) = X \pm d$ can be approximated by $C_k^\pi(X)$ by taking $k \to \infty$ and $d$ is chosen suitably. Using Theorem 4.4.7, it now can be shown that $C_0(X)$ is minimax for certain values of $a$, $b$, and $d$.

Over the last several decades Statistical Decision Theory has produced some interesting results which have counter intuitive features. The most prominent among these is the famous ‘Stein’s effect’ (Stein (1981)) as described below.

Suppose we have independent observations $X_1, \ldots, X_n$ such that each $X_i$ follows $N(\theta_i, 1)$, $1 \leq i \leq n$. For estimating each $\theta_i$ separately under the SEL $(\delta_i - \theta_i)^2$ the usual estimator $\delta_i^0 = X_i$ is admissible. But for simultaneous estimation of $\theta = (\theta_1, \ldots, \theta_n)$ under the generalized SEL $\sum_{i=1}^n (\delta_i - \theta_i)^2$, the combined estimator $\delta^0 = (\delta_1^0, \ldots, \delta_n^0)'$ is admissible only for $n \leq 2$. For $n \geq 3$, the above $\delta^0$ is inadmissible and can be dominated by a host of ‘shrinkage’ estimators including the famous ‘James-Stein estimator’ given as

$$\delta^{JS} = (1 - (p - 2)/||\delta^0||^2)\delta^0.$$  

This feature of inadmissibility of a vector of individual admissible estimators when the dimension is greater than a critical value is called ‘Stein’s effect’. The Stein’s effect is found to be applicable for many other distributions, like Gamma, Poisson, etc.

Another interesting decision-theoretic result derived by Stein is the one for normal variance estimation which is discussed in detail in the next chapter.
Chapter 5

Computational Aspects

5.1 Preliminaries

In earlier chapters, especially in Chapter 3 and Chapter 4, we have seen how to evaluate an estimator in terms of bias, variance, and risk. For instance, given iid observations \( X_1, X_2, \ldots, X_n \) from a \( N(\mu, \sigma^2) \) population where both \( \mu \) and \( \sigma \) are unknown, the variance \( \sigma^2 \) can be estimated by its MLE

\[
\hat{\sigma}^2_{ML} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n} S_X \text{ (say)}
\]

where \( \bar{X} \) is the sample mean. Using the fact that \( S_X/\sigma^2 \) follows \( \chi^2_{n-1} \) distribution, it is now easy to find the bias and variance of \( \hat{\sigma}^2_{ML} \) as

\[
\text{Bias of } \hat{\sigma}^2_{ML} = E \left[ \hat{\sigma}^2_{ML} \right] - \sigma^2 \\
= \left( \frac{n-1}{n} \right) \sigma^2 - \sigma^2 \\
= -\frac{\sigma^2}{n} \tag{5.1.1}
\]

\[
\text{Variance of } \hat{\sigma}^2_{ML} = E \left[ \left( \hat{\sigma}^2_{ML} - E \left[ \hat{\sigma}^2_{ML} \right] \right)^2 \right] \\
= \frac{2(n-1)}{n^2} \sigma^4
\]

Using \( S_X/\sigma^2 \) as a pivot, one can find a \((1 - \alpha)\)-level confidence interval for \( \sigma^2 \) as

\[
\left[ \frac{S_X}{\chi^2_{n-1, \alpha/2}}, \frac{S_X}{\chi^2_{n-1, (1-\alpha)/2}} \right]
\]
where $\chi^2_{n-1}(\alpha/2)$ and $\chi^2_{n-1}(1-\alpha/2)$ are respectively the upper and lower tail $(\alpha/2)$-probability cutoff points of $\chi^2_{n-1}$ distribution. The length of this confidence interval is

$$S_X \left( \frac{1}{\chi^2_{n-1}, (1-\alpha/2)} - \frac{1}{\chi^2_{n-1}, \alpha/2} \right)$$

which is a random variable. Since the length plays an important role in evaluating a confidence interval, one can find the average (or expected) length of the above confidence interval which turns out to be

$$(n-1) \left( \frac{1}{\chi^2_{n-1}, (1-\alpha/2)} - \frac{1}{\chi^2_{n-1}, \alpha/2} \right) \sigma^2$$

and this can be compared with the expected length of any other $(1-\alpha)$-level confidence interval for $\sigma^2$.

For testing $H_0 : \sigma^2 = \sigma^2_0$ against $H_1 : \sigma^2 > \sigma^2_0$ (for some given $\sigma^2_0$), we can consider a test procedure at $\alpha$-level of significance as:

- reject $H_0$ if $\frac{S_X}{\sigma^2_0} > \chi^2_{n-1, \alpha}$
- accept $H_0$ if $\frac{S_X}{\sigma^2_0} \leq \chi^2_{n-1, \alpha}$

The power of the above test procedure can be found as:

$$\text{Power} = P\left( \frac{S_X}{\sigma^2} > \chi^2_{n-1, \alpha} \bigg| \sigma^2 > \sigma^2_0 \right) = P \left( \chi^2_{n-1} > \left( \frac{\sigma^2}{\sigma^2_0} \right) \chi^2_{n-1, \alpha} \bigg| \sigma^2 > \sigma^2_0 \right)$$

which is an increasing function of $\sigma^2$ (increases from $\alpha$ to 1 as $\sigma^2$ goes from $\sigma^2_0$ to $\infty$).

The above derivations involving $\sigma^2$ (point as well as interval estimation, and hypothesis testing) were relatively easy because of the involvement of chi-square distribution. But there are cases where it is difficult to obtain a closed expression as illustrated by the following example.

For estimating the normal variance $\sigma^2$ as described above, the MLE of $\sigma^2$ is certainly not the optimal estimator in terms of mean squared error (MSE). The optimal estimator of the form (constant)$S_X$ (i.e., a
constant multiple of \( S_X \) having the smallest MSE is given by

\[
\hat{\sigma}_{\text{opt}}^2 = \frac{S_X}{(n + 1)} \tag{5.1.4}
\]

In other words, the (optimal) estimator \( \hat{\sigma}_{\text{opt}}^2 \) has the smallest risk under the squared error loss (SEL) in the class of estimators \( \{ cS_X, c > 0 \} \).

Similar to (5.1.1), one can find the bias, variance, and risk of \( \hat{\sigma}_{\text{opt}}^2 \). However, an improved estimator which is better than \( \hat{\sigma}_{\text{opt}}^2 \) (in terms of lower risk under SEL) is

\[
\hat{\sigma}_S^2 = \min \left\{ \frac{S_X}{(n + 1)}, \frac{S_X + n(\bar{X} - \mu_0)^2}{(n + 2)} \right\} \tag{5.1.5}
\]

where \( \mu_0 \) is any known real number. Even though the estimator \( \hat{\sigma}_S^2 \) looks a bit uncommon, it has a nice interpretation. The estimator \( \hat{\sigma}_S^2 \) can be expressed as

\[
\hat{\sigma}_S^2 = \begin{cases} 
S_X/(n + 1) & \text{if } n(\bar{X} - \mu_0)^2/S_X > 1/(n + 1) \\
(S_X + n(\bar{X} - \mu_0)^2)/(n + 2) & \text{if } n(\bar{X} - \mu_0)^2/S_X \leq 1/(n + 1) 
\end{cases} \tag{5.1.6}
\]

The region of \( \{ n(\bar{X} - \mu_0)^2/S_X > 1/(n + 1) \} \) can be treated as a critical (or rejection) region for testing the null hypothesis \( H_0 : \mu = \mu_0 \) against \( H_1 : \mu \neq \mu_0 \). Note that the type-I error of the test is not predetermined unlike the regular hypothesis testing case. The test statistic \( n(\bar{X} - \mu_0)^2/S_X \), under \( H_0 : \mu = \mu_0 \), follows \( F_{1, (n-1)} \)-distribution, and hence the type-I error of the test is

\[
\text{type-I error} = P(\text{reject } H_0 | H_0 \text{ is true}) = P \left( F_{1, (n-1)} > \frac{1}{(n + 1)} \right) \tag{5.1.7}
\]

which is dependent on the sample size \( n \).

If \( H_0 \) is rejected, then \( \mu \) is treated as completely unknown and \( \sigma^2 \) is estimated by \( S_X/(n + 1) \) (i.e., \( \hat{\sigma}_{\text{opt}}^2 \)). On the other hand, if \( H_0 \) is accepted, then \( \mu \) is treated as known (and equal to \( \mu_0 \)) and \( \sigma^2 \) is estimated by

\[
\frac{S_X + n(\bar{X} - \mu_0)^2}{(n + 2)} = \frac{1}{(n + 2)} \sum_{i=1}^{n} (X_i - \mu_0)^2
\]

\(^1\)C. Stein (1964) first proposed this estimator.
Chapter 5  Computational Aspects

which is $\hat{\sigma}_{opt}^2$ with iid observations $X_1, X_2, \ldots, X_n$ from $N(\mu_0, \sigma^2)$ distribution.

Finding the bias or risk expressions for $\hat{\sigma}_s^2$ is not as easy as it would be for $\hat{\sigma}_{ml}^2$ or $\hat{\sigma}_{opt}^2$. In fact, it is not possible to get any closed expression of the bias (or risk) of $\hat{\sigma}_s^2$. To see this, we use the notations $U = \sqrt{n}(X - \mu_0)/\sigma$ (which follows $N(\mu - \mu_0)/\sigma, 1)$, and $V = S_X/\sigma^2$ (which follows $\chi^2_{n-1}$). Let

$$F = \frac{U^2}{V} = \frac{n(X - \mu_0)^2}{S_X}$$

we can then obtain

$$E[\hat{\sigma}_s^2] = E\left[\frac{S_X}{(n+1)}I\left(F > \frac{1}{(n+1)}\right) + \frac{(S_X + \sigma^2U^2)}{(n+2)}I\left(F \leq \frac{1}{(n+1)}\right)\right]$$

(5.1.8)

The random variable $F$ follows a noncentral $F_{1,(n-1)}(\lambda)$-distribution with the noncentrality parameter $\lambda$ (say) $= n(\mu - \mu_0)^2/\sigma^2$. Therefore, the bias of $\hat{\sigma}_s^2$ is given by

$$\text{Bias of } \hat{\sigma}_s^2 = E[\hat{\sigma}_s^2] - \sigma^2 = \sigma^2E\left[\frac{S_X/\sigma^2}{(n+1)}I\left(F > \frac{1}{(n+1)}\right) + \frac{(S_X + \sigma^2U^2)/\sigma^2}{(n+2)}I\left(F \leq \frac{1}{(n+1)}\right) - 1\right]$$

(5.1.9)

From the above bias expression (5.1.9), it is not clear how the estimator $\hat{\sigma}_s^2$ fares compared to $\hat{\sigma}_{ml}^2$ or $\hat{\sigma}_{opt}^2$ in terms of bias.

Similar to the above bias expression (5.1.9), one can also derive the risk expression under SEL (i.e., MSE) for the estimator $\hat{\sigma}_s^2$. Although $\hat{\sigma}_s^2$ is known to be superior to $\hat{\sigma}_{opt}^2$, it is important to study the risk function of $\hat{\sigma}_s^2$ and see how much improvement it offers over $\hat{\sigma}_{opt}^2$. If the risk improvement of $\hat{\sigma}_s^2$ over $\hat{\sigma}_{opt}^2$ turns out insignificant (over a large section of the parameter space), then one may ignore $\hat{\sigma}_s^2$ and use $\hat{\sigma}_{opt}^2$ instead.

Another example where deriving bias (or variance or risk) is more complicated than the one discussed above involves gamma distribution. Suppose that we have iid observations $X_1, X_2, \ldots, X_n$ from $G(\alpha, \beta)$ distribution with the pdf

$$f(x | \alpha, \beta) = \frac{1}{\beta \Gamma(\alpha)} e^{-x/\beta} x^{\alpha-1}, \quad x > 0$$

(5.1.10)

where both $\alpha$ and $\beta$ are unknown. To estimate $\alpha$ and $\beta$ by the maximum
5.1 Preliminaries

likelihood estimation method, we first note the likelihood function

\[ L(\alpha, \beta \mid X) = \frac{1}{\beta^n \Gamma(\alpha)} e^{-nX/\beta} \left( \prod_{i=1}^{n} X_i \right)^{\alpha-1} \]

where \( \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \); or the logarithmic likelihood function

\[ \ln L(\alpha, \beta \mid X) = -n\alpha \ln \beta - n \ln \Gamma(\alpha) - \frac{n\overline{X}}{\beta} + (\alpha - 1) \sum_{i=1}^{n} \ln X_i \] (5.1.11)

Differentiating the above log-likelihood function w.r.t. \( \alpha \) and \( \beta \) and setting them equal to 0 yields

\[ \ln \beta + \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{1}{n} \sum_{i=1}^{n} \ln X_i \quad \text{and} \quad \alpha \beta = \overline{X} \] (5.1.12)

where \( \Gamma'(\alpha) = \frac{\partial \Gamma(\alpha)}{\partial \alpha} \). The maximum likelihood estimators \( \hat{\alpha}_{ML} \) and \( \hat{\beta}_{ML} \) of \( \alpha \) and \( \beta \), respectively, are the solutions of the system of equations in (5.1.12).

Note that the above MLEs of \( \alpha \) and \( \beta \) do not have closed expressions, and hence it is impossible to derive bias, variance, and risk expressions of these estimators. Yet it is important for real life problems to study the performance of \( \hat{\alpha}_{ML} \) and \( \hat{\beta}_{ML} \) (in terms of bias, variance, or risk), and evaluate them in the presence of other estimators. For example, method of moments estimators (MMEs) have closed expressions for above \( \alpha \) and \( \beta \) estimation.

The two examples discussed above were for point estimation. But similar difficulties can also arise either in set (interval) estimation where one may not have a closed expression for the volume (length) of the confidence set (interval) under consideration, or in hypothesis testing where one is interested in studying the complicated power function of a particular test procedure. Essentially, all these problems can be unified to look as an expectation of a suitable functional, say \( G(X, \theta) \), i.e., one needs to study \( E[G(X, \theta)] \) where \( X \) follows a probability distribution \( f(x \mid \theta) \).

In Section 5.2, we first talk about numerical integration technique which is becoming popular with fast computers and handy computing packages. Section 5.3 discusses Monte-Carlo simulation method which can be convenient when numerical integration is either too time consuming or not so accurate. Section 5.4 introduces a popular nonparametric
method based on a resampling technique. Finally, in Section 5.5 we will see simple simulation techniques useful in testing and interval estimation.

5.2 Numerical Integration

Given the random vector $X = (X_1, X_2, \ldots, X_n)'$ following a joint probability distribution $f(x|\theta)$, we would like to study $E[G(X, \theta)]$ where $G$ is a suitable functional.\(^2\) Usually, the functional $G$ depends on the data $X$ through the minimal sufficient statistic, say $T = T(X)$. The dimension of $T$ is much less than that of $X$. Therefore, for convenience we can write

$$E[G(X, \theta)] = E[G(T, \theta)] \quad (5.2.1)$$

where the distribution of $T$ is $g(t|\theta)$. The distribution $g$ of $T$ is assumed to be known except the parameter $\theta$. The expression (5.2.1) can be written as

$$E[G(T, \theta)] = \int_{R_T} G(t, \theta) g(t|\theta) \, dt \quad (5.2.2)$$

where $R_T$ is the suitable range for $T$ and $dt = dt_1 dt_2 \ldots dt_k$ with $k$ being the dimension of $T$.

To study $E[G(X, \theta)]$ for $\theta \in \Theta$, we select several possible values of $\theta$ from $\Theta$, and for each given value of $\theta$ in $\Theta$ the integration in (5.2.2) is done numerically. Finally, numerical integration values are plotted against $\theta$ to get a better understanding of $E[G(X, \theta)]$.

As an example, we now compute the bias of $\hat{\sigma}_s^2$ through numerical integration.

Bias of $\hat{\sigma}_s^2 = \sigma^2 E \left[ \frac{V}{(n+1)} \left( \frac{W}{V} > \frac{1}{(n+1)} \right) + \frac{(V+W)}{(n+2)} \left( \frac{W}{V} < \frac{1}{(n+1)} \right) - 1 \right] \quad (5.2.3)$

where $W$ follows a noncentral $\chi^2_1(\lambda)$ distribution with noncentrality parameter $\lambda = n(\mu - \mu_0)^2/\sigma^2$, and $W$ is independent of $V$. The pdfs of $V$ and $W$ are respectively

$$V \sim g_1(v) = \frac{1}{2^{(n-1)/2} \Gamma((n-1)/2)} e^{-v/2} v^{(n-1)/2 - 1}$$

\(^2\)Here $G$ is a function of the data and the parameter $\theta$. Such a $G$ is called a ‘functional’.

© 2006 by Taylor & Francis Group, LLC
\[ W \sim g_2(w|\lambda) = \sum_{j=0}^{\infty} e^{-\lambda/2} \left( \frac{\lambda}{2} \right)^j \frac{1}{j!} e^{-w/2} w^{(j+1)/2 - 1} \]

(5.2.4)

As one can see from above, a noncentral \( \chi^2_k(\lambda) \) distribution can be characterized as a Poisson mixture of central \( \chi^2_{k+2J} \) distributions with \( J \) following Poisson(\( \lambda/2 \)) distribution.

Let

\[ A_1(\lambda) = \frac{V}{(n+1)} \int \left( \frac{W}{V} > \frac{1}{(n+1)} \right) \left( v_{g1}g_2(w|\lambda) \right) dv \]

and

\[ A_2(\lambda) = \frac{(V+W)}{(n+2)} \int \left( \frac{W}{V} < \frac{1}{(n+1)} \right) \left( v_{g1}g_2(w|\lambda) \right) dv \]

(5.2.6)

From (5.2.3), (5.2.5), and (5.2.6), it is easy to see that the bias of \( \hat{\sigma}_S^2 \) is a function of \( \sigma^2 \) as well as \( \lambda = n(\mu - \mu_0)^2/\sigma^2 \) (apart from \( n \)), and hence denoting it by \( B_n(\hat{\sigma}_S^2|\sigma^2,\lambda) \) we get

\[ \text{Bias of } \hat{\sigma}_S^2 \quad = \quad B_n(\hat{\sigma}_S^2|\sigma^2,\lambda) \]

\[ = \sigma^2 \left\{ A_1(\lambda) + A_2(\lambda) - 1 \right\} \]

(5.2.7)

where \( A_1(\lambda) \) and \( A_2(\lambda) \) are given in (5.2.5) and (5.2.6), respectively. The expression \( B_n(\hat{\sigma}_S^2|\sigma^2,\lambda) \) is directly proportional to \( A_1(\lambda) + A_2(\lambda) - 1 \) and hence we can plot

\[ B_n(\hat{\sigma}_S^2|\sigma^2,\lambda)/\sigma^2 = B_n(\hat{\sigma}_S^2|1,\lambda) \]

(5.2.8)

against \( \lambda \) to get an idea about the bias of \( \hat{\sigma}_S^2 \). This is illustrated below with specific choices of \( n \) and \( \lambda \).

First, using the computer package Mathematica, we compute

\[ B_n(\hat{\sigma}_S^2|1,\lambda) = A_1(\lambda) + A_2(\lambda) - 1 \]

for selected values of \( \lambda \) with \( n = 5 \) and 10. The numerical results are provided in Table 5.2.1. Figure 5.2.1 shows the plots of

\[ B_5(\hat{\sigma}_S^2|1,\lambda) = B_5 \quad \text{and} \quad B_{10}(\hat{\sigma}_S^2|1,\lambda) = B_{10} \]
Table 5.2.1: Computed Values of $B_n(\hat{\sigma}_S^2 \mid 1, \lambda)$.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$n$</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-3.77</td>
<td>-214</td>
<td>1.0</td>
</tr>
<tr>
<td>0.1</td>
<td>-3.75</td>
<td>-213</td>
<td>1.5</td>
</tr>
<tr>
<td>0.3</td>
<td>-3.72</td>
<td>-210</td>
<td>2.5</td>
</tr>
<tr>
<td>0.5</td>
<td>-3.69</td>
<td>-208</td>
<td>5.0</td>
</tr>
<tr>
<td>0.7</td>
<td>-3.66</td>
<td>-206</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Figure 5.2.1: Plots of $B_5(\hat{\sigma}_S^2 \mid 1, \lambda)$ and $B_{10}(\hat{\sigma}_S^2 \mid 1, \lambda)$. 

© 2006 by Taylor & Francis Group, LLC
Denoting the bias of $\hat{\sigma}^2_{opt}$ (see (5.1.4)) by $B_n(\hat{\sigma}^2_{opt}|\sigma^2)$ which depends on the parameter $\sigma^2$ only except $n$, we have

$$B_n(\hat{\sigma}^2_{opt}|\sigma^2) = \sigma^2 \left( \frac{n-1}{n+1} \right) - \sigma^2 = -\frac{2}{(n+1)} \sigma^2$$

(5.2.9)

It is straightforward that

$$B_n(\hat{\sigma}^2_{opt}|1) = B_n(\hat{\sigma}^2_{opt}|\sigma^2)/\sigma^2$$

with

$$B_5(\hat{\sigma}^2_{opt}|1) = -0.33 \text{ and } B_{10}(\hat{\sigma}^2_{opt}|1) = -0.18$$

Remark 5.2.1 From the above numerical result, it is clear that $\hat{\sigma}^2_{opt}$ has always larger absolute bias than $\hat{\sigma}^2_{S}$ does. Indeed, it is true that $B_n(\hat{\sigma}^2_{S}|1,\lambda) \rightarrow B_n(\hat{\sigma}^2_{opt}|1)$ as $\lambda \rightarrow \infty$ for fixed $n$. Both $\hat{\sigma}^2_{S}$ and $\hat{\sigma}^2_{opt}$ underestimate $\sigma^2$.

Remark 5.2.2 The computer package ‘Mathematica’ works fine for single or double integration(s). But for triple or higher order integrations, it may not be very effective.

5.3 Monte-Carlo Simulation

Monte-Carlo simulation is an alternative approach (alternative to numerical integration) to evaluate the expectation $E[G(X,\theta)]$ where the random variable $X$ follows the distribution $f(x|\theta)$, $\theta \in \Theta$. The distribution $f(x|\theta)$ is assumed to have a known form. The technique of simulation depends heavily on artificial data generated from the distribution $f(x|\theta)$ for selected values of $\theta$. Most of the latest standard statistical software packages have a built-in mechanism to generate data from commonly used probability distributions.

The Monte-Carlo simulation is conducted through a large number (say, $M$) of steps (or cycles). First, fix a suitable value of $\theta \in \Theta$. Then at the $j$th step, $1 \leq j \leq M$, generate a random sample $X^{(j)}$ of size $n$ from the known distribution $f(x|\theta)$. For this data $X^{(j)} = (X_1^{(j)}, \ldots, X_n^{(j)})$, compute $G(X,\theta)$ and call it $G_j$, i.e., $G_j = G(X^{(j)},\theta)$. After the $M^{th}$ step, we have generated the simulated (or replicated) values of $G(X,\theta)$
as \( G_1, G_2, \ldots, G_M \). The expectation of the functional \( G(X, \theta) \) is then approximated by the average of \( G_1, G_2, \ldots, G_M \), or

\[
E[G(X, \theta)] \approx \frac{1}{M} \sum_{j=1}^{M} G_j = \frac{1}{M} \sum_{j=1}^{M} G(X^{(j)}, \theta) = E^*_G(\theta) \text{ (say)} \quad (5.3.1)
\]

Since the functional \( G(X, \theta) \) often depends on the data \( X \) through the minimal sufficient statistic \( T = T(X) \) with dimension \( k < n \), it is convenient to conduct the simulation by generating observations on \( T \) rather than \( X \). To be precise, let (5.2.1) hold and \( T \) has distribution \( g(t \mid \theta) \), \( \theta \in \Theta \), then for a selected value of \( \theta \in \Theta \), independently generate \( T^{(1)}, T^{(2)}, \ldots, T^{(M)} \) following \( g(t \mid \theta) \). The expectation of \( G(X, \theta) \) is then approximated by

\[
E[G(X, \theta)] = E[G(T, \theta)] \approx \frac{1}{M} \sum_{j=1}^{M} G(T^{(j)}, \theta) = E^*_G(\theta) \text{ (say)} \quad (5.3.2)
\]

For several values of \( \theta \) in \( \Theta \), say, \( \theta_1, \theta_2, \ldots, \theta_k \), one can thus approximate \( E[G(X, \theta)] \) by using (5.3.2), and obtain \( E^*_G(\theta_1), E^*_G(\theta_2), \ldots, E^*_G(\theta_k) \) which can be plotted against \( \theta \) (for \( \theta = \theta_1, \theta_2, \ldots, \theta_k \)) to get a better understanding of \( E[G(X, \theta)] \).

To approximate \( E[G(X, \theta)] \), one can write a program code based upon the following steps:

**Step-1**: Input \( \theta \), (say \( \theta = \theta_l, l = 1, 2, \ldots, k \))

**Step-2**: Generate \( T^{(j)} \) following \( g(t \mid \theta) \)

**Step-3**: Compute \( G(T^{(j)}, \theta) = G_j \) (say)

**Step-4**: Compute \( \overline{G}_j = \left( \frac{j-1}{j} \right) \overline{G}_{j-1} + \left( \frac{1}{j} \right) G_j \), where \( \overline{G}_0 = 0 \)

**Step-5**: Repeat above steps 2 through 4 for \( j = 1, 2, \ldots, M \)

**Step-6**: \[ E^*_G(\theta) = \overline{G}_M \] (5.3.3)

To demonstrate the above simulation steps, we now compute the bias of \( \hat{\sigma}^2_S \). This can be done by using either (5.1.6) or (5.2.3). First, the expression (5.1.6) will be used to demonstrate the bias computation, and
then show that the same can also be achieved by (5.2.3) with less complications.

**Use of the expression (5.1.6):**

For fixed $n$ and $\mu_0$:

**Step-1:** Select $\mu \in (-\infty, \infty)$ and $\sigma^2 > 0$

**Step-2:** Generate iid observations $X^{(1)}_1, X^{(1)}_2, \ldots, X^{(1)}_n$ from $N(\mu, \sigma^2)$

**Step-3:** Compute $\hat{\sigma}^2_S$ for the above set of observations and call it $\hat{\sigma}^2_{S(1)}$

**Step-4:** Repeat above steps 2 and 3 for a large number (say, $M$) of times and obtain $\hat{\sigma}^2_{S(1)}, \hat{\sigma}^2_{S(2)}, \ldots, \hat{\sigma}^2_{S(M)}$

**Step-5:** The bias of $\hat{\sigma}^2_S$ is approximated as

$$\frac{1}{M} \sum_{j=1}^{M} \hat{\sigma}^2_{S(j)} - \sigma^2 = \frac{1}{M} \sum_{j=1}^{M} (\hat{\sigma}^2_{S(j)} - \sigma^2) \quad (5.3.4)$$

Theoretically (from (5.2.7)), we know that the above simulated bias depends on $n, \mu, \sigma^2$ only through $\sigma^2$ and $\lambda$.

**Remark 5.3.1** To avoid computational errors and other difficulties such as ‘over-flow’ errors, it is suggested that the average in (5.3.4) can be done through the recursive relation (Step-4) in (5.3.3). Also, the idea that the average of a large number of simulated values (i.e., $\sum_{j=1}^{M} \hat{\sigma}^2_{S(j)}/M$) approximates the true expectation (i.e., $E[\hat{\sigma}^2_S]$) stems from the ‘Strong Law of Large Numbers’, i.e., $\sum_{i=1}^{M} Y_i/M$ converges to $\mu_Y$ almost surely as $M \to \infty$ where $Y_i$’s are iid random variables with mean $\mu_Y$.

We now show how to use the expression (5.2.3) to approximate the bias of $\hat{\sigma}^2_S$.

**Use of the expression (5.2.3):**

Note that in the expression (5.2.3) the random variables $V$ and $W$ are independent, $V \sim \chi^2_{n-1}$ and $W \sim \chi^2(\lambda)$ with $\lambda = n(\mu - \mu_0)^2/\sigma^2$. The random variable $W$ can be characterized as $W = U^2$ where $U \sim N(\pm \sqrt{\lambda}, 1)$. Without loss of generality, we can use $U \sim N(\sqrt{\lambda}, 1)$ to characterize $W$. Therefore, for fixed $n$:

**Step-1:** Select $\sigma^2 > 0, \lambda > 0$
Chapter 5  Computational Aspects

Step-2: Generate independent random observations \( V^{(1)} \sim \chi^2_{n-1} \) and \( U^{(1)} \sim N(\sqrt{\lambda}, 1) \)

Step-3: Compute
\[
g(V^{(1)}, W^{(1)}) = \frac{V^{(1)}}{(n+1)} I \left( \frac{W^{(1)}}{V^{(1)}} > \frac{1}{(n+1)} \right) + \frac{(V^{(1)} + W^{(1)})}{(n+2)} I \left( \frac{W^{(1)}}{V^{(1)}} < \frac{1}{(n+1)} \right) - 1
\]
where \( W^{(1)} = (U^{(1)})^2 \) and call \( g(V^{(1)}, W^{(1)}) = g^{(1)} \)

Step-4: Repeat above steps 2 and 3 for a large number of times (say, \( M \)), and obtain \( g^{(1)}, g^{(2)}, \ldots, g^{(M)} \)

Step-5: The bias of \( \hat{\sigma}^2_S \) is then approximated by
\[
\sigma^2 \left( \frac{1}{M} \sum_{j=1}^{M} g^{(j)} \right) = \sigma^2 \left( \frac{1}{M} \sum_{j=1}^{M} g(V^{(j)}, W^{(j)}) \right) \quad (5.3.5)
\]

Remark 5.3.2  The advantage in using the expression (5.2.3) is that it uses the minimal sufficient statistics and hence uses fewer computational steps. For example, at every data generating step (Step-2) it directly generates \( V^{(j)} \) and \( U^{(j)} \) instead of \( X_1^{(j)}, X_2^{(j)}, \ldots, X_n^{(j)} \) which leads to substantial savings in computational time when \( n \) is large.

Note that a simulated average
\[
\frac{1}{M} \sum_{j=1}^{M} G_j = \frac{1}{M} \sum_{j=1}^{M} G(T^{(j)}, \theta)
\]
which approximates \( E(G(T, \theta)) \) (see (5.3.2)) is based on a very large sample (of size \( M \)) of artificially generated observations (\( G_j = G(T^{(j)}, \theta) \) values are all iid) and therefore subject to sampling variations. As a result, two separate runs of simulations (with same \( n, M, \theta \)) will not produce exactly the same average. Hence, the magnitude of the fluctuation from one simulation run to another is a key feature in measuring the accuracy of the simulated average. The standard error (SE) which measures the accuracy of the simulated average can be estimated by
\[
SE(M) = \frac{S(M)}{\sqrt{M}}, \quad \text{where} \quad S(M) = \left( \frac{1}{(M-1)} \sum_{j=1}^{M} (G_j - \bar{G}_M)^2 \right)^{1/2} \quad (5.3.6)
\]
with $\bar{G}_M = \sum_{j=1}^M G_j/M$. For computational convenience, the value of $S_{(M)}$ can be computed through a recursive relation as shown below.

After generating $G_k = G(T^k, \theta)$, $k = 1, 2, 3, \ldots, j$, the SE is estimated by $SE_{(j)}$ as

$$SE_{(j)} = \frac{S_{(j)}}{\sqrt{j}}$$

where

$$S_{(j)} = \left(\frac{(j-2)S_{(j-1)}^2 + (j-1)\bar{G}_{j-1}^2 + G_j^2 - j\bar{G}_j^2}{(j-1)}\right)^{1/2}$$ (5.3.7)

One can see how to calculate $G_j$ and $G_{j-1}$ in Step-4 of (5.3.3). At the end of $M^{th}$ replication, one gets $SE_{(M)} = S_{(M)}/\sqrt{M}$, the desired estimated standard error of the result.

**Remark 5.3.3** The estimated standard error ($SE_{(M)}$) of the simulated average plays an important role in determining the value of $M$, the number of replications, which is to be determined by the investigator. Before undertaking a full fledged simulation study one usually conducts a pilot study with a few possible values of $M$ (typically $M$ ranges from 5,000 to 100,000) to get an idea about $SE_{(M)}$. It is suggested that the value of $M$ should be chosen such that $SE_{(M)}$ should not exceed a fixed percentage (say, 1 or 0.5%) of $\bar{G}_M$ to maintain a high accuracy in approximation of $E(G(T, \theta))$ by $\bar{G}_M$.

As an example of the simulation, the bias of $\hat{\sigma}^2$ is once again computed for $\sigma^2 = 1$ and the selected values of $\lambda$ with $n = 5$ and 10. The expression (5.2.3) has been used to carry out the simulation with $M = 10,000$ replications. The results are summarized in Table 5.3.1 which can be compared with the corresponding numerical results listed in Table 5.2.1. The estimated standard errors are reported in parentheses.

**Remark 5.3.4** Using the concepts of the central limit theorem, one can expect that the simulated mean $\bar{G}_M$ to be within two standard errors of the actual mean $E(G(T, \theta))$. In Table 5.3.1, the maximum estimated $SE$ is approximately 0.005. Therefore, the reported values of $\bar{G}_M$ are expected to be correct up to the second decimal place.
5.4 Bootstrap Method of Resampling

In Section 5.2 and Section 5.3, we have discussed methods to evaluate the performance of an estimator computationally. But both numerical integration and simulation methods were based on the underlying probability distribution of the data. The assumed probability distribution of the data is called the model of the data. A model is usually selected based on some expert knowledge or some past experience where similar situations had been encountered.

For instance, it is well-known that the normal model (or the bell curve) is widely used in modeling IQ (intelligent quotient) scores of individuals of a fixed population. To be precise, let iid observations \( X_1, X_2, \ldots, X_n \) be available where \( X_i \) is the IQ score (obtained from a test) of the \( i \)th subject in the sample, \( 1 \leq i \leq n \). It is assumed that \( X_1, X_2, \ldots, X_n \) are iid \( N(\mu, \sigma^2) \) where \( \mu \in (-\infty, \infty) \) and \( \sigma^2 > 0 \). Note that our assumed normal model is truly a family of distributions. Suppose the variance \( \sigma^2 \) is estimated by the estimator \( \hat{\sigma}_g^2 \) (given in (5.1.5)). The bias or the risk (under a suitable loss function) of \( \hat{\sigma}_g^2 \) depends heavily on the normal model. Though \( \hat{\sigma}_g^2 \) is written in terms of the minimal sufficient statistics, the choice of the minimal sufficient statistics and the corresponding probability distributions are dependent on the assumed model.

Even though the maximum possible care is taken in a model selection

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>( n )</th>
<th>( \hat{\sigma}_g^2 )</th>
<th>( n )</th>
<th>( \hat{\sigma}_g^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>5</td>
<td>−.3737</td>
<td>10</td>
<td>−.3584</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0043)</td>
<td></td>
<td>(0.0045)</td>
</tr>
<tr>
<td>0.1</td>
<td>5</td>
<td>−.3716</td>
<td>10</td>
<td>−.3531</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0043)</td>
<td></td>
<td>(0.0045)</td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>−.3681</td>
<td>10</td>
<td>−.3513</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0044)</td>
<td></td>
<td>(0.0046)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>−.3649</td>
<td>10</td>
<td>−.3353</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0044)</td>
<td></td>
<td>(0.0047)</td>
</tr>
<tr>
<td>0.7</td>
<td>5</td>
<td>−.3621</td>
<td>10</td>
<td>−.3302</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0044)</td>
<td></td>
<td>(0.0047)</td>
</tr>
</tbody>
</table>

Table 5.3.1: Simulated Bias of \( \hat{\sigma}_g^2 \).
(the applicability of which must be validated by the given data set), there is always plenty of room for making errors which may go undetected. Therefore, one should be very careful in explaining the bias or the risk of an estimator (say, $\hat{\sigma}^2$) of a parameter ($\sigma^2$) under the assumed model (say, the normal) which may be misspecified. For example, if an IQ data set really comes from a $t$-distribution with location parameter $\mu$, scale parameter $\sigma$, and 20 degrees of freedom, i.e., if the true pdf of each observation is $f_k(x | \mu, \sigma)$ with $k = 20$ where

$$f_k(x | \mu, \sigma) = \frac{\Gamma((k + 1)/2)}{\sigma \sqrt{k\pi} \Gamma(k/2)} \left(1 + \frac{(x - \mu)^2}{k\sigma^2}\right)^{-(k+1)/2}$$  \hspace{1cm} (5.4.1)

then the standard tests for normality may not detect the difference between the $t$-distribution (5.4.1) and the $N(\mu, \sigma^2)$ distribution unless the sample size is large. This is one of the major drawbacks of parametric inference where the investigator makes a decision based upon the assumed model. To avoid the problem of model misspecification, one can follow a nonparametric approach where the true distribution of the data is assumed to lie in a much larger family, the family of all probability distributions. One then uses large sample (asymptotic) results to evaluate an estimator of a parameter. But for small to moderate sample sizes, the large sample results are hardly applicable and may lead to inappropriate inferences. One, therefore, should keep an open mind in choosing a model and estimating a parameter by carefully employing parametric as well as nonparametric approaches thereby arriving at a meaningful conclusion.

Bootstrap method is a nonparametric resampling technique where the data is allowed to speak for itself. The basic technique is described below in a fairly general setup.

Suppose we have iid observations $X_1, X_2, \ldots, X_n$ from a probability distribution $F(x)$ (cdf) which is completely unknown. Our parameter of interest is $\theta = \theta(F)$ (i.e., $\theta$ is a function of the unknown probability distribution $F$). The parameter $\theta$ can be the population mean, population variance, or population standard deviation, etc. Let $\theta$ be estimated by an estimator $\hat{\theta} = \hat{\theta}(X)$ which may be available from the empirical distribution of the data or may be obtained from a rough parametric consideration. The probability distribution of $\hat{\theta}$, called the sampling distribution of $\hat{\theta}$, depends on the distribution of $X$ which is $F$. Therefore, let $G_F(\cdot)$ be the true probability distribution of the parameter estimator.
\( \hat{\theta} \), and bias or risk of \( \hat{\theta} \) will be based upon \( G_F \) which is dependent on \( F \). Since \( F \) is unknown, so is the distribution \( G_F(\cdot) \). But the beauty of the bootstrap method is that it enables us, through computations, to approximate the sampling distribution \( G_F(\cdot) \) of \( \hat{\theta} \).

**The Bootstrap Procedure:**

Let \{\( X_1, X_2, \ldots, X_n \)\} be a random sample selected from an unknown probability distribution \( F \). We treat this original sample as a small new population.

**Step-1:** Draw a random sample of size \( n \) from \{\( X_1, X_2, \ldots, X_n \)\} with replacement and let the resampled observations be \( X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)} \) where the superscript denotes the first replication. Based on \( \{X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}\} = X^{(1)} \) (say), compute \( \hat{\theta} \) and obtain \( \hat{\theta}^{(1)} = \hat{\theta}(X^{(1)}) \).

**Step-2:** Repeat the above Step-1 independently for a large number of times (say, \( M \)). At the \( j \)th replication, \( 1 \leq j \leq M \), we have resampled data set \( X^{(j)} = \{X_1^{(j)}, X_2^{(j)}, \ldots, X_n^{(j)}\} \) and the corresponding parameter estimate is given by \( \hat{\theta}^{(j)} = \hat{\theta}(X^{(j)}) \).

**Step-3:** The computed values \( \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(M)} \) are the replicated values of \( \hat{\theta} \), and can be used to approximate the true distribution \( G_F \) of \( \hat{\theta} \). The approximate distribution of \( \hat{\theta} \), say \( \tilde{G} \), is obtained as

\[
\tilde{G}(y) = \frac{\# \{\hat{\theta}^{(j)}'s \leq y\}}{M} = \frac{1}{M} \sum_{j=1}^{M} I_{(-\infty, y]}(\hat{\theta}^{(j)}) \tag{5.4.2}
\]

where \( I_{(-\infty, y]}(\cdot) \) is the indicator function defined for the interval \( (-\infty, y] \).

The reason why \( \tilde{G}(y) \) approximates the true distribution \( G_F(y) \) of \( \hat{\theta} \) is not difficult to justify. As mentioned before, the distribution \( G_F \) of \( \hat{\theta} \) is completely unknown because \( F \), the distribution of \( X \), is assumed (in a nonparametric setup) to be so. However, an estimate of \( G_F \), say \( \hat{G} = G_{\hat{F}} \), where \( \hat{F} \) is the empirical distribution based on \( X \), i.e.,

\[
\hat{F}(x) = \frac{\# X_i's \leq x}{n} = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i) \tag{5.4.3}
\]

In fact, \( \hat{G} = G_{\hat{F}} \) is the MLE of \( G \) since \( \hat{F} \) is so for \( F \). Now, in the bootstrap method, the \( j \)th replicated sample \( X_1^{(j)}, X_2^{(j)}, \ldots, X_n^{(j)} \) i.i.d. from \( \hat{F} \) gives \( \hat{\theta}^{(j)} \) which has true probability distribution \( G_{\hat{F}} \), \( 1 \leq j \leq M \). Thus, after generating \( \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(M)} \) where \( M \) is sufficiently large, we are trying to capture the distribution \( \tilde{G} = G_{\hat{F}} \) based on the
empirical distribution $\tilde{G}$ constructed from $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$, \ldots, $\hat{\theta}^{(M)}$. When $M$ is sufficiently large and $n$ is not too small,

$$\tilde{G}(y) \approx \hat{G}(y) = G_{F}(y) \approx G_{F}(y)$$ (5.4.4)

If $M \to \infty$, then $\tilde{G}(y) \to \hat{G}(y) = G_{F}(y)$ almost surely; and if further $n \to \infty$, then $\tilde{G}(y) = G_{F}(y) \to G_{F}(y)$ almost surely.

Using the bootstrap method, one can estimate the bias of an estimator $\hat{\theta}$ by $\hat{B}_{b}(\hat{\theta})$ where

$$\hat{B}_{b}(\hat{\theta}) = \frac{1}{M} \sum_{j=1}^{M} (\hat{\theta}^{(j)} - \hat{\theta}) = \overline{\hat{\theta}^{(\cdot)}} - \hat{\theta}$$

with

$$\overline{\hat{\theta}^{(\cdot)}} = \frac{1}{M} \sum_{j=1}^{M} \hat{\theta}^{(j)}$$ (5.4.5)

The standard error (SE) of $\hat{\theta}$ is estimated by

$$\hat{SE}_{b}(\hat{\theta}) = \frac{1}{\sqrt{M}} \left( \frac{1}{(M-1)} \sum_{j=1}^{M} (\hat{\theta}^{(j)} - \overline{\hat{\theta}^{(\cdot)}})^{2} \right)^{1/2}$$ (5.4.6)

where $\overline{\hat{\theta}^{(\cdot)}}$ is given by (5.4.5). Similarly, one can estimate the risk of $\hat{\theta}$ under the squared error loss (SEL) by

$$\hat{R}_{b}(\hat{\theta}) = \frac{1}{M} \sum_{j=1}^{M} (\hat{\theta}^{(j)} - \hat{\theta})^{2}$$ (5.4.7)

The subscript ‘$b$’ in above equations indicates that these estimates are obtained by the bootstrap method.

Now we apply the bootstrap method to a real life data set as a demonstration.

**Example 5.4.1:** The results of the blood tests, showing nearest milligrams of glucose per 100 milliliters of blood, were obtained from 53 nonpregnant women at Boston City Hospital. The data set (Source: *The American Journal of Clinical Nutrition*, Vol. 19, pp. 345–351) is displayed as follows:

The following Figure 5.4.1 is a percentage (or a relative frequency) histogram of the glucose data set.
Table 5.4.1: Glucose per 100 Milliliters of Blood (in mg).

<table>
<thead>
<tr>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
</tr>
<tr>
<td>0.1</td>
</tr>
<tr>
<td>0.15</td>
</tr>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.25</td>
</tr>
<tr>
<td>0.3</td>
</tr>
<tr>
<td>0.35</td>
</tr>
<tr>
<td>0.4</td>
</tr>
<tr>
<td>0.45</td>
</tr>
<tr>
<td>0.5</td>
</tr>
</tbody>
</table>

Figure 5.4.1: A Histogram of the Glucose Data Set.
Our goal is to estimate $\sigma^2$, the variance of the population. The parameter $\sigma^2$ is the variance of blood glucose of the population of all women from which the above sample of 53 was taken. The available estimators of $\sigma^2$ are $\hat{\sigma}_{ML}^2$, $\hat{\sigma}_{opt}^2$, and $\hat{\sigma}_S^2$ (see (5.1.1), (5.1.4), and (5.1.6)). The following table gives bootstrap estimates of bias and SE of the above estimates with $M = 10,000$ replications.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Estimated Bias</th>
<th>Estimated SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\sigma}_{ML}^2$</td>
<td>-1.2455</td>
<td>0.1171</td>
</tr>
<tr>
<td>$\hat{\sigma}_{opt}^2$</td>
<td>-1.2220</td>
<td>0.1149</td>
</tr>
<tr>
<td>$\hat{\sigma}_S^2$</td>
<td>-1.3601</td>
<td>0.1130</td>
</tr>
</tbody>
</table>

Table 5.4.2: Bootstrap Estimates of Bias and SE.

5.5 Testing and Interval Estimation Based on Computations

Hypothesis Testing

One can consider a hypothesis testing procedure based on statistical simulation and numerical computation. Since fast computing has become convenient, it can be implemented easily at a relatively low cost.

Note that one needs to come up with a suitable test statistic and its cut-off point(s) to test a hypothesis. In this regard the LRT comes in handy due to its asymptotic distribution.

The following test procedure based on statistical simulation and numerical computation is similar to LRT, but does not use the asymptotic chi-square distribution to obtain the cut-off point(s) for the test statistic. This is discussed for two possible cases for testing a hypothesis on a scalar valued parameter.

**Test** $H_0$: $\theta = \theta_0$ vs. $H_1$: $\theta < \theta_0$ or $\theta > \theta_0$ or $\theta \neq \theta_0$.

**Case 1: Nuisance parameter is absent**

**Step-1**: Obtain $\hat{\theta}_{ML}$, the MLE of $\theta$.

**Step-2**: Set $\theta = \theta_0$ (specified by $H_0$). Generate the data $X$ from $f(x|\theta_0)$ for a large number of (say $M$) times. For each of these
replicated data, recalculate the MLE of \( \theta \) (pretending that it is unknown). Let these recalculated MLEs be \( \hat{\theta}_{01}, \hat{\theta}_{02}, \ldots, \hat{\theta}_{0M} \).

**Step-3**: Order the above recalculated MLEs as \( \hat{\theta}_{0(1)} \leq \hat{\theta}_{0(2)} \leq \ldots \leq \hat{\theta}_{0(M)} \).

**Step-4**:

(a) For testing against \( H_1 : \theta < \theta_0 \), let \( \hat{\theta}_L = \hat{\theta}_{0(\alpha M)} \).
Reject \( H_0 \) if \( \hat{\theta}_{ML} < \hat{\theta}_L \); accept \( H_0 \) if \( \hat{\theta}_{ML} \geq \hat{\theta}_L \).

(b) For testing against \( H_1 : \theta > \theta_0 \), let \( \hat{\theta}_U = \hat{\theta}_{0(1-\alpha M)} \).
Reject \( H_0 \) if \( \hat{\theta}_{ML} > \hat{\theta}_U \); accept \( H_0 \) if \( \hat{\theta}_{ML} \leq \hat{\theta}_U \).

(c) For testing against \( H_1 : \theta \neq \theta_0 \), let \( \hat{\theta}_U = \hat{\theta}_{0(1-\alpha/2 M)} \) and \( \hat{\theta}_L = \hat{\theta}_{0(\alpha/2 M)} \).
Reject \( H_0 \) if \( \hat{\theta}_{ML} > \hat{\theta}_U \) or \( \hat{\theta}_{ML} < \hat{\theta}_L \). Accept otherwise.

As an application of the above method assume that we have iid observations \( X_1, X_2, \ldots, X_n \) from \( N(\theta, 1) \) and we have to test \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta \neq \theta_0 \).

1. In Step-1, we have \( \hat{\theta}_{ML} = \bar{X} \).

2. In Step-2, we generate

   (1st replication): \( X_1^{(1)}, \ldots, X_n^{(1)} \) iid \( N(\theta_0, 1) \); get \( \hat{\theta}_{01} = \bar{X}^{(1)} \)

   

   : 

   (Mth replication): \( X_1^{(M)}, \ldots, X_n^{(M)} \) iid \( N(\theta_0, 1) \); get \( \hat{\theta}_{0M} = \bar{X}^{(M)} \)

[Note that the values \( \hat{\theta}_{01}, \ldots, \hat{\theta}_{0M} \) are representing the \( N(\theta_0, 1/n) \) distribution; i.e., a histogram of these values clearly follows the \( N(\theta_0, 1/n) \) pdf.]

3. In Step-3, obtain \( \hat{\theta}_L = \hat{\theta}_{0((\alpha/2)M)} \) and \( \hat{\theta}_U = \hat{\theta}_{0(1-\alpha/2)M} \). \( \hat{\theta}_L \) and \( \hat{\theta}_U \) are approximately \( (\theta_0 - z(\alpha/2)/\sqrt{n}) \) and \( (\theta_0 + z(\alpha/2)/\sqrt{n}) \) respectively.

4. In Step-4, reject \( H_0 \) if \( \hat{\theta}_{ML} > \hat{\theta}_U \) or \( \hat{\theta}_{ML} < \hat{\theta}_L \); accept otherwise. [In classical theory, we reject \( H_0 \) if \( \hat{\theta}_{ML} > \theta_0 + z(\alpha/2)/\sqrt{n} \) or \( \hat{\theta}_{ML} < \theta_0 - z(\alpha/2)/\sqrt{n} \).]
5.5 Testing and Interval Estimation Based on Computations

Case-2 : Nuisance parameter is present

Suppose the data $X$ follows a distribution $f(X|\theta)$ where $\theta = (\theta^{(1)}, \theta^{(2)}) \in \Theta$, and $\theta^{(2)}$ can be vector valued. The pdf (or pmf) $f(X|\theta)$ is known except $\theta$ and our interest lies in $\theta^{(1)}$ only. Hence $\theta^{(2)}$ is a nuisance parameter (vector). To test $H_0: \theta^{(1)} = \theta^{(1)}_0$ against a suitable alternative $H_1: (\theta^{(1)} < \theta^{(1)}_0$ or $\theta^{(1)} > \theta^{(1)}_0$ or $\theta^{(1)} \neq \theta^{(1)}_0$) at level $\alpha$, the following steps are followed.

Step-1 : Obtain $\hat{\theta}_{ML} = (\hat{\theta}_{ML}^{(1)}, \hat{\theta}_{ML}^{(2)})$, the MLE of $\theta$.

Step-2 : (a) Set $\theta^{(1)} = \theta^{(1)}_0$, then find the MLE of $\theta^{(2)}$. Call this ‘restricted MLE’ of $\theta^{(2)}$, and denote it by $\hat{\theta}_{RML}^{(2)}$.

(b) Generate the data $X$ from $f(x|\theta^{(1)}_0, \hat{\theta}_{RML}^{(2)})$ for a large number of times (say $M$). For each of these replicated data, recalculate the MLE of $\theta$ (pretending that it is unknown) and retain the component relevant for $\theta^{(1)}$. Let these recalculated MLEs of $\theta^{(1)}$ be $\hat{\theta}^{(1)}_{01}, \hat{\theta}^{(1)}_{02}, \ldots, \hat{\theta}^{(1)}_{0M}$.

Step-3 : Let $\hat{\theta}^{(1)}_{0(1)} \leq \hat{\theta}^{(1)}_{0(2)} \leq \ldots \leq \hat{\theta}^{(1)}_{0(M)}$ be the ordered values of $\hat{\theta}^{(1)}_{01}, \hat{\theta}^{(1)}_{02}, \ldots, \hat{\theta}^{(1)}_{0M}$.

Step-4 : Similar to that of Case-1 where $\hat{\theta}_{0(i)}$’s are replaced by $\hat{\theta}^{(1)}_{0(i)}$’s and $\hat{\theta}_{ML}$ is used in place of $\hat{\theta}_{ML}$.

Interval Estimation

Earlier we discussed about interval estimation of a parameter based on a pivot. But in many problems a pivot is not readily available. In such a case one tests a mock hypothesis on the parameter, and then inverts the acceptance region to obtain an interval estimate of the parameter. Since we have discussed about hypothesis testing based on simulation and computation, we can take advantage of the same to come up with an interval estimate of the parameter under consideration. In the following we provide the procedure only for the above first case (where there is no nuisance parameter), but can be extended for the second case (where there are nuisance parameters) easily with slight notational changes. Also, the procedure is given for a two sided interval estimate, but can be modified suitably for one sided interval estimates as well. We follow the notations used above for hypothesis testing.
**Step-1** : Take several values \( \theta_0^1, \theta_0^2, \ldots, \theta_0^K \), over the parameter space of \( \theta \), suitably spaced. [It is suggested that these values of \( \theta \), numbered somewhere between five and eight, be taken equally spaced over \( \hat{\theta}_{\text{ML}} \pm 2(SE) \), where \( SE \) is the standard error or estimated standard deviation of \( \hat{\theta}_{\text{ML}} \).]

**Step-2** : Perform a hypothesis testing of \( H^0_1 : \theta = \theta_0 \) against \( H^1_1 : \theta \neq \theta_0 \) \((j = 1, 2, \ldots, K)\) at level \( \alpha \). For each \( j \), obtain the cut off points \((\hat{\theta}_L^j, \hat{\theta}_U^j)\) based on the simulation and computation described earlier, \( j = 1, 2, \ldots, K \). Remember that the bounds \((\hat{\theta}_L^j, \hat{\theta}_U^j)\) are for \( \hat{\theta}_{\text{ML}} \), the MLE of \( \theta \).

**Step-3** : Plot the lower bounds \( \hat{\theta}_L^j \), \( 1 \leq j \leq K \), against \( \theta_0 \) and then approximate the plotted curve by a suitable smooth function say, \( g_L(\theta_0) \). Similarly, plot the upper bounds \( \hat{\theta}_U^j \), \( 1 \leq j \leq K \), against \( \theta_0 \) and then approximate the plotted curve by a suitable smooth function say, \( g_U(\theta_0) \).

**Step-4** : Finally, solve for \( \theta_0 \) from the equations \( g_L(\theta_0) = \hat{\theta}_{\text{ML}} \) and \( g_U(\theta_0) = \hat{\theta}_{\text{ML}} \). The two solutions of \( \theta_0 \) thus obtained set the boundaries of the interval estimate of \( \theta \) with minimum confidence level \((1 - \alpha)\).

[In fact, the curves \( g_L^{-1}(\hat{\theta}_{\text{ML}}) \) and \( g_U^{-1}(\hat{\theta}_{\text{ML}}) \) can act as \((1 - \alpha)\)-level two sided confidence bounds for the parameter \( \theta \) when \( \hat{\theta}_{\text{ML}} \) is the MLE of \( \theta \).] The above computer intensive approach of interval estimation comes in handy in several applications as we will see in later chapters.
Part-II
Exponential and Other
Positively Skewed
Distributions with
Applications
Chapter 6

Exponential Distribution

6.1 Preliminaries

In Chapter 2, the exponential distribution has been mentioned briefly. A random variable $X$ is said to have the exponential distribution with scale parameter $\beta$ if its pdf is given by

$$f(x | \beta) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0 \text{ and } \beta > 0 \quad (6.1.1)$$

The mean of $X$ or $E(X)$ is $\beta$, and hence if $X$ represents the life span of a randomly selected object in hours, then $\beta$ stands for the mean life of all such objects having the exponential (6.1.1) life time distribution. One can use the reparameterization $\lambda = 1/\beta$ and then the pdf (6.1.1) of $X$ becomes

$$f(x | \lambda) = \lambda e^{-\lambda x}, \quad x > 0 \text{ and } \lambda > 0 \quad (6.1.2)$$

Note that with the pdf (6.1.2), $E(X) = 1/\lambda$, and $\lambda = 1/E(X)$ is called the failure rate.

Henceforth an exponential distribution with mean $\beta$ (i.e., the representation (6.1.1)) will be denoted as ‘$\text{Exp}(\beta)$’. At the same token, the reparameterized version (6.1.2) will be denoted as ‘$\text{Exp}(1/\lambda)$’.

Figure 6.1.1 plots $f(x | \beta)$ (6.1.1) for various values of $\beta$.

In queuing theory the exponential distribution is commonly used to model the probability distribution of a random variable that represents service time. For examples, the time is taken to complete service at a grocery checkout counter, at a doctor’s office, at an airline check-in counter,
at a bank teller, and so on. In reliability applications, the exponential distribution is frequently used to model the lifetimes of components which are subject to wear out. For examples, lifetime of electric bulbs, batteries, appliances, and transistors, etc., are modeled by the exponential distribution. This distribution has also been used successfully to model the distribution of the length of time between successive random events such as the time between two auto accidents on a particular stretch of a highway; the time between the arrivals of two customers at a ticket counter; and the time between breakdowns of a machine, etc.

From (6.1.1) or (6.1.2), it is seen that the distribution’s support is the positive side of the real line, i.e., $\mathbb{R}^+$. Since ‘time’, in general, is a positive
measure, an exponential random variable is often well-suited to represent quantitative characteristics measured in time units. Also, from the plots in Figure 6.1.1, it is evident that the distribution is heavily skewed to the right. Since very few service times for customers, life times of appliances, or failure times of machines exceed their expected values, the exponential distribution seems natural to model the probability distribution of these types of random variables.

One appealing feature of the exponential distribution is its mathematical simplicity. If a random variable \( X \) follows \( \text{Exp}(\beta) \), then \( Y = X/\beta \) follows \( \text{Exp}(1) \). Therefore,

\[
P(a \leq X \leq b) = P\left(\frac{a}{\beta} \leq Y \leq \frac{b}{\beta}\right) = \int_{a/\beta}^{b/\beta} e^{-y} \, dy = e^{-a/\beta} - e^{-b/\beta}
\]

(6.1.3)

**Example 6.1.1:** The service time at a drive-through window of a fast-food restaurant is assumed to be exponentially distributed with a mean of 2 minutes. What percentage of customers end up waiting more than 5 minutes?

Let the random variable \( X \) be the service time (the time from placing an order till delivery of the food) of a customer selected at random. Since \( X \) follows \( \text{Exp}(2) \), the pdf of \( X \) is

\[
f(x) = \frac{1}{2} e^{-x/2}, \quad x \geq 0
\]

The probability that \( X \) is more than 5 minutes is

\[
P(X > 5) = 1 - P(0 \leq X \leq 5) = e^{-5/2} \quad \text{(from (6.1.3))} \quad (6.1.4)
\]

Thus 8.2% of the customers wait for more than 5 minutes.

**Example 6.1.2:** Suppose that the radiator of a particular type of imported cars has lifetime distributed as \( \text{Exp}(\beta) \) with \( \beta = 5 \) years. If five of these cars are sold by a local dealer recently, what is the probability that at least two will have original radiators functioning at the end of 8 years?
Let $p$ be the probability that a given radiator of a car mentioned above functions after 8 years. If $X$ represents the life of a radiator, then $X$ follows $\text{Exp}(\beta)$ with $\beta = 5$ years. Hence,

$$p = P(X > 8) = \int_{8}^{\infty} \frac{1}{5} e^{-x/5} dx = e^{-8/5} \approx 0.2$$

The dealer sells five cars lately, and let $Y$ be the number of cars (out of five) with the original radiator still functioning after eight years. It is seen that $Y$ follows a binomial distribution, $\text{B}(n, p)$, with $n = 5$ and $p = 0.2$. Therefore,

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - [P(Y = 0) + P(Y = 1)]$$

where

$$P(Y = y) = \binom{5}{y} (0.2)^y (0.8)^{5-y}, \quad y = 0, 1$$

Straightforward computation yields that

$$P(Y \geq 2) \approx 0.263$$

Therefore, 26.3% of the time at least two cars out of five will have original radiators functioning after 8 years.

In the following, we provide two real life data sets which tend to follow exponential distributions. These data sets are obtained from Sakai and Burris (1985).

**Example 6.1.3:** Differences in physical characteristics between males and females are well known for animals. Similarly, sex-related differences in vegetative growth of dioecious plants have long been recognized and are of interest lately. For example, vegetative growth of females may be reduced because of nutrient drain of fruit production, or female plants may require larger size to attain the energy necessary for fruit production. Data from field studies have shown that females are associated with larger size in many short-lived species. In perennial species, females tend to exhibit reduced vigor, growth rate, and vegetative spread. Two varieties of ‘trembling aspen’ (Populus tremuloides) have been studied and males and females have been compared in terms of ramets (stems). Table 6.1.1 shows ramet size distributions (the frequency and relative frequency distributions) for 4289 females and 3498 males of Pellston Plain clones, and their sample probability histograms (or the relative frequency distributions) are displayed in Figure 6.1.2 and Figure 6.1.3.
6.1 Preliminaries

<table>
<thead>
<tr>
<th>Ramet Size (in cm) (dbh)</th>
<th>Sex of the Plant</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 – 1.3</td>
<td>1450</td>
<td>0.338</td>
<td></td>
<td>1350</td>
<td>0.386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 – 3.8</td>
<td>1363</td>
<td>0.318</td>
<td></td>
<td>1177</td>
<td>0.336</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8 – 6.3</td>
<td>725</td>
<td>0.169</td>
<td></td>
<td>500</td>
<td>0.143</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.3 – 8.9</td>
<td>300</td>
<td>0.070</td>
<td></td>
<td>178</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.9 – 11.4</td>
<td>175</td>
<td>0.041</td>
<td></td>
<td>80</td>
<td>0.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4 – 14.0</td>
<td>100</td>
<td>0.023</td>
<td></td>
<td>70</td>
<td>0.020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.0 – 16.5</td>
<td>60</td>
<td>0.014</td>
<td></td>
<td>50</td>
<td>0.014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.5 – 19.1</td>
<td>47</td>
<td>0.011</td>
<td></td>
<td>35</td>
<td>0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.1 – 21.6</td>
<td>35</td>
<td>0.008</td>
<td></td>
<td>30</td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.6 or more</td>
<td>34</td>
<td>0.008</td>
<td></td>
<td>28</td>
<td>0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4289</td>
<td>1.000</td>
<td></td>
<td>3498</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1.1: Ramet Size Distributions.

A further generalization of the exponential distribution (6.1.1) is given by the two-parameter exponential distribution \( \text{Exp}(\beta, \mu) \) with pdf

\[
f(x \mid \beta, \mu) = \frac{1}{\beta} e^{-(x-\mu)/\beta}, \quad x > \mu \text{ and } \mu \in \mathbb{R}, \beta > 0 \quad (6.1.5)
\]

In the expression (6.1.5), the parameter \( \mu \) is a location (or shift) parameter, and \( \beta \) is a scale parameter. If, for example, the life span of a particular brand of electric bulbs follows a two-parameter exponential distribution as mentioned earlier, then \( \mu \) indicates the minimum guarantee time. Figure 6.1.4 shows plots of the two-parameter exponential pdf with \((\beta, \mu) = (1, 1)\) and \((2, 3)\).

In many real life applications, it is reasonable to assume \( \mu = 0 \). This may be due to some prior or expert knowledge, or after testing and accepting a null hypothesis \( H_0 : \mu = 0 \) against \( H_1 : \mu \neq 0 \). The special case thus obtained is the one-parameter exponential distribution \( \text{Exp}(\beta) \) discussed earlier which is essentially \( \text{Exp}(\beta, 0) \).

If \( \mu = 0 \) and \( \beta = 1 \), then the two-parameter exponential distribution \( \text{Exp}(1, 0) \) is called the standard exponential distribution.

Notational Clarification: From now on \( \text{Exp}(\beta, \mu) \) indicates a two-parameter exponential distribution with scale parameter \( \beta \) and location

© 2006 by Taylor & Francis Group, LLC
Figure 6.1.2: The relative frequency histogram and a matching exponential pdf of female ramet size with $\beta = 3.0$.

parameter $\mu$. If not mentioned otherwise, $\text{Exp}(\beta)$ indicates $\text{Exp}(\beta, 0)$, i.e., the one parameter exponential distribution with scale parameter $\beta$.

6.2 Characterization of Exponential Distribution

The easiest and the most widely accepted way to characterize the exponential distribution is through the memory-less property (see the equation (6.2.1) below). Suppose that the continuous random variable $X > 0$ represents the lifetime of a component. Also, let the future lifetime have the same probability distribution no matter how old the com-
Figure 6.1.3: The relative frequency histogram of male ramet size and a matching exponential pdf with $\beta = 2.5$.

ponent is, i.e.,

$$P(X \leq x + c \mid X > c) = P(X \leq x) \quad (6.2.1)$$

for all $c > 0$ and all $x > 0$. If $f(x)$ and $F(x)$ denote the pdf and cdf of $X$ respectively, then (6.2.1) yields

$$\frac{P(c < X \leq x + c)}{P(X > c)} = P(X \leq x)$$

or,

$$\frac{F(x + c) - F(c)}{1 - F(c)} = F(x) \quad \forall \ x > 0, \ c > 0 \quad (6.2.2)$$

Dividing both sides of (6.2.2) by $x$, and then taking the limit as $x$ tending to 0 gives

$$\lim_{x \to 0} \frac{F(x + c) - F(c)}{x} = \frac{1}{1 - F(c)} \lim_{x \to 0} \frac{F(x)}{x}$$
Using $F(0) = 0$, the above equation can be written as

$$\frac{F'(c)}{1 - F(c)} = F'(0)$$

Since $F'(0)$ is a constant, we can denote it as $F'(0) = a$. Now the above equation becomes

$$\frac{-F'(c)}{1 - F(c)} = -a \quad \forall c > 0 \quad (6.2.3)$$

Solving the above differential equation\(^1\) (6.2.3), one gets

$$1 - F(c) = be^{-ac} \quad (6.2.4)$$

\(^1\)For a function $g(x) = y$ and the differential equation $y'/y = k$ (constant), one can get the solution as $g(x) \propto \exp(kx)$. 
6.2 Characterization of Exponential Distribution

for some suitable constant \( b \). Using the condition that \( \lim_{c \to 0} F(c) = 0 \), and replacing \( c \) by \( x \), one can obtain the following result,

\[
1 - F(x) = e^{-ax} \quad \text{or} \quad f(x) = ae^{-ax}
\]

which is the same as \( \text{Exp}(\beta, \mu) \) with \( \mu = 0 \) and \( \beta = 1/a \).

Conversely, it can be seen easily that if \( X \) follows \( \text{Exp}(\beta, 0) \), then the condition (6.2.1) is satisfied. Further, \( X \) follows \( \text{Exp}(\beta, \mu) \) if and only if (6.2.1) is satisfied for all \( c > \mu \) and \( x > 0 \).

Another characterization of the exponential distribution is given in terms of the order statistics.

**Theorem 6.2.1** Let \( X_1, X_2, \ldots, X_n \) be continuous iid random variables and \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \) be the order statistics, then if the conditional expectation \( E(X_{(i+1)} - X_{(i)} \mid X_{(i)} = x) \) is independent of \( x \) for fixed \( i, 1 \leq i < n \), then \( X_i \)'s have exponential distribution.

Desu (1971) gave a different type of characterization of \( \text{Exp}(\beta, 0) \), the key idea of which was to present a function of the order statistics having the same distribution as the one sampled. This is presented in the following theorem.

**Theorem 6.2.2** Let \( X_1, X_2, \ldots, X_n \) be iid nonnegative random variables with a nondegenerate\(^2\) cdf \( F(\cdot) \) and \( X_{(1)} = \min \{X_1, X_2, \ldots, X_n\} \). Then \( nX_{(1)} \) follows the same cdf \( F(\cdot) \) if and only if \( F(x) = 1 - \exp(-x/\beta) \), for \( x \geq 0 \), where \( \beta \) is a positive constant.\(^3\)

While Theorem 6.2.2 characterizes an exponential distribution in terms of the distribution of the smallest observed value, the following result, due to Alasamullah (1977), characterizes it in terms of the distribution of the spacings of order statistics when the \( iid \) observations are thought to have a particular type of probability distributions.

Let \( F \) be the cdf of a nonnegative random variable and \( \overline{F}(x) = 1 - F(x) \) for \( x \geq 0 \). The cdf \( F \) is called (i) **new better than used** (NBU) if

\[
\overline{F}(x + y) \leq \overline{F}(x)\overline{F}(y) \quad \text{for} \ x, y \geq 0
\]

\(^2\)That is, the range (or support) has more than one point.

\(^3\)\( F(x) = 1 - \exp(-x/\beta) \) is the cdf of \( \text{Exp}(\beta, 0) \) distribution.
and (ii) new worse than used (NWU) if
\[ F(x + y) \geq F(x)F(y) \quad \text{for } x, y \geq 0 \] (6.2.7)

**Remark 6.2.1** If we think of a random variable \( X \) representing the lifetime of some unit and has cdf \( F \), then \( X \) being NBU (or NWU) means \( P(X > x + y | X > x) \leq (or \geq) P(X > y) \), i.e., the probability of additional life of any used item is less (or greater) than that of a new item. In other words, NBU (or NWU) implies that survivability of an old item is lower (or higher) than that of a new one.

**Theorem 6.2.3** Let \( X_1, X_2, \ldots, X_n \) be iid random variables having cdf \( F(x) \) that is strictly increasing on \([0, \infty)\). Then the following properties are equivalent:

1. \( F(x) = 1 - \exp(-x/\beta) \) for some \( \beta > 0 \) and \( x > 0 \)
2. The variables \( Y_i = (n-i) \left( X_{i+1} - X_i \right) \), \( 1 \leq i \leq n \), are identically distributed with cdf \( F(\cdot) \), and \( F(\cdot) \) is either NBU or NWU

Seshadri and Csörgő (1969) exploited further the idea of spacings of order statistics and characterized exponential distributions in terms of the distribution of spacings. Such characterizations eventually help us in testing the goodness of fit of an exponential distribution.

**Theorem 6.2.4** Let \( X_1, X_2, \ldots, X_n \) be iid nonnegative random variables with mean \( \beta > 0 \). Define \( S = \sum_{i=1}^{n} X_i \) and \( Z_j = \sum_{i=1}^{j} X_i / S \), \( 1 \leq j \leq n-1 \). Then the random variables \( Z_j \)'s act like \((n-1)\) ordered random variables from continuous Uniform over the interval \([0, 1]\) (i.e., CU\([0, 1]\)) if and only if the random variables \( X_i \)'s are \( \text{Exp}(\beta, 0) \).

**Theorem 6.2.5** Let \( X_1, X_2, \ldots, X_n \) be iid nonnegative random variables. Define \( Y_i = (n+1-i) \left( X_i - X_{i-1} \right) \), \( 1 \leq i \leq n \), where \( X_0 = 0 \). The random variables \( Y_i \)'s follow \( \text{Exp}(\beta, 0) \) if and only if \( X_i \)'s are so.

The extension of Theorem 6.2.4 to a more general case when \( X_i \)'s are \( \text{Exp}(\beta, \mu) \) is given in the next theorem.

**Theorem 6.2.6** Let \( X_1, X_2, \ldots, X_n \) be iid random variables with mean \( \beta \) and \( X_i > \mu \) for all \( 1 \leq i \leq n \). Define \( Y_i = (n+i-1) \left( X_i - X_{i-1} \right) \), \( 1 \leq i \leq n \), where \( X_0 = \mu \). Let \( S = \sum_{i=1}^{n} Y_i \) and \( Z_j = \sum_{i=1}^{j} Y_i / S \), \( 1 \leq j \leq n-1 \). Then the random variables \( Z_j \)'s act like \((n-1)\) ordered CU\([0,1]\) random variables if and only if \( X_i \)'s are \( \text{Exp}(\beta, \mu) \).
6.3 Estimation of Parameter(s)

We first consider estimation of the scale parameter \( \beta \) in \( \text{Exp}(\beta) \) distribution.

Given iid observation \( X_1, X_2, \ldots, X_n \) with common pdf (6.1.1) the likelihood function of the data is

\[
L(\beta | X_1, X_2, \ldots, X_n) = \beta^{-n} \exp \left\{ -\frac{n}{\beta} \sum_{i=1}^{n} X_i \right\}, \quad \beta > 0 \tag{6.3.1}
\]

and hence the minimal sufficient statistic for \( \beta \) is \( \sum_{i=1}^{n} X_i \) (or \( \bar{X} \)). The maximum likelihood estimator (MLE) as well as the method of moment estimator of \( \beta \) is

\[
\hat{\beta}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X} \tag{6.3.2}
\]

It can be shown that

\[
E(\hat{\beta}_{\text{ML}}) = 0, \quad \text{Var}(\hat{\beta}_{\text{ML}}) = \frac{\beta^2}{n} \tag{6.3.3}
\]

i.e., \( \hat{\beta}_{\text{ML}} \) is an unbiased estimator of \( \beta \) for the model (6.1.1). Since \( \hat{\beta}_{\text{ML}} \) is a function of the minimal sufficient statistic, it is also the unique minimum variance unbiased estimator (UMVUE) of \( \beta \) (see Theorem 3.4.4, Rao-Blackwell Theorem). Also, if \( \eta = \eta(\beta) \) is a function of \( \beta \), then the MLE of \( \eta \) is

\[
\hat{\eta}_{\text{ML}} = \eta(\hat{\beta}_{\text{ML}}) \tag{6.3.4}
\]

However, \( \hat{\eta}_{\text{ML}} \) is not necessarily an unbiased estimator of \( \eta \).

For example, as a special case, let \( \xi_c \) be the 100\( c \)th percentile point \( (0 < c < 1) \) of the \( \text{Exp}(\beta) \)-distribution; i.e., \( c = F(\xi_c) = P(X \leq \xi_c) = 1 - \exp(-\xi_c/\beta) \). Then \( \xi_c = -\beta \ln(1-c) \). Hence, the MLE of \( \xi_c \) is \( \hat{\xi}_c(\text{ML}) = -\hat{\beta}_{\text{ML}} \ln(1-c) \).

**Example 6.3.1:** Consider the given ordered data set

\[
\begin{array}{cccc}
48.96 & 54.67 & 100.75 & 120.68 \\
148.34 & 228.43 & 293.45 & 415.82 \\
\end{array}
\]

which is to be modeled by a suitable exponential distribution. Estimate \( \beta \) by \( \hat{\beta}_{\text{ML}} = \bar{X} = 177.3875 \). Then get \( \hat{\xi}_c = -(176.3875) \ln (1-c) \), and this is plotted in Figure 6.3.1.
Chapter 6  Exponential Distribution

Note that the risk function of $\hat{\beta}_{\text{ML}}$ under the squared error loss (or the mean squared error) is same as the variance of $\hat{\beta}_{\text{ML}}$ (due to its unbiasedness). Therefore

$$\text{Var}(\hat{\xi}_c(\text{ML})) = (\ln (1-c))^2 \text{Var}(\hat{\beta}_{\text{ML}}) = (\ln (1-c))^2 \beta^2 / n \quad (6.3.5)$$

The following two diagrams plot $\text{Var}(\hat{\xi}_c(\text{ML}))$ against $\beta$ (with fixed $c = 0.90$) and $c$ (with fixed $\beta = 1$), for $n = 8$. 

Figure 6.3.1: The graph of $\hat{\xi}_c$ against $c \in (0, 1)$. 

© 2006 by Taylor & Francis Group, LLC
Now we consider a two-parameter exponential distribution with scale parameter \( \beta \) and location parameter \( \mu \).

Given iid observations \( X_1, X_2, \ldots, X_n \) with pdf (6.1.5), the likelihood
function of the data is

\[
L(\beta, \mu \mid X_1, X_2, \ldots, X_n) = \beta^{-n} \exp \left\{ -\frac{1}{\beta} \sum_{i=1}^{n} (X_i - \mu) \right\}
\]

\[
= \beta^{-n} \exp \left\{ -\frac{1}{\beta} \sum_{i=1}^{n} (X_i - X_{(1)}) \right\}
- n(X_{(1)} - \mu)/\beta
\]  

(6.3.6)

where \( \beta > 0 \) and \( X_i \geq \mu \). Therefore, the minimal sufficient statistics for \((\beta, \mu)\) is \((\sum_{i=1}^{n} (X_i - X_{(1)}), X_{(1)})\). The MLEs of \( \beta \) and \( \mu \) are

\[
\hat{\beta}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - X_{(1)}) \quad \text{and} \quad \hat{\mu}_{\text{ML}} = X_{(1)}
\]  

(6.3.7)

respectively. The statistics \( \hat{\beta}_{\text{ML}} \) and \( \hat{\mu}_{\text{ML}} \) are mutually independent, and

\[
2n\hat{\beta}_{\text{ML}}/\beta \sim \chi^2_{2(n-1)}, \quad \hat{\mu}_{\text{ML}} \sim \text{Exp}(\beta/n, \mu)
\]  

(6.3.8)

Using the above properties (6.3.8), it is easy to see that

\[
E(\hat{\beta}_{\text{ML}}) = \beta(n - 1)/n \quad \text{and} \quad E(\hat{\mu}_{\text{ML}}) = \mu + \beta/n
\]  

(6.3.9)

Therefore, both \( \hat{\beta}_{\text{ML}} \) and \( \hat{\mu}_{\text{ML}} \) are biased estimators of their respective parameters under the model (6.1.5). Hence, the best unbiased estimators (UEs) of \( \beta \) and \( \mu \) respectively are

\[
\hat{\beta}_{\text{U}} = n\hat{\beta}_{\text{ML}}/(n - 1) \quad \text{and} \quad \hat{\mu}_{\text{U}} = \hat{\mu}_{\text{ML}} - \hat{\beta}_{\text{U}}/n
\]  

(6.3.10)

Under the model (6.1.5), one can find the first two population moments as

\[
E(X) = \mu + \beta \quad \text{and} \quad E(X^2) = \beta^2 + (\mu + \beta)^2
\]  

(6.3.11)

Hence, one can find the method of moments estimators (MMEs) of \( \mu \) and \( \beta \) by equating the population moments (in (6.3.11)) with their sample counterparts as

\[
\hat{\beta}_{\text{MM}} = \left\{ \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right\}^{1/2} \quad \text{and} \quad \hat{\mu}_{\text{MM}} = \bar{X} - \hat{\beta}_{\text{MM}}
\]  

(6.3.12)

But since the MMEs are not functions of minimal sufficient statistics, they (or the unbiased estimators based on the MMEs) are worse than the
MLE (6.3.9) as well as the UE (6.3.10) in terms of variance. Therefore, one is better off by focusing only on the MLEs or UEs.

From (6.3.9) it is easy to see that

\[
\text{Bias} (\hat{\beta}_{\text{ML}}) = E(\hat{\beta}_{\text{ML}} - \beta) = -\beta/n, \quad \text{Var}(\hat{\beta}_{\text{ML}}) = (n-1)\beta^2/n^2
\]

and

\[
\text{Bias} (\hat{\mu}_{\text{ML}}) = E(\hat{\mu}_{\text{ML}} - \mu) = \beta/n, \quad \text{Var}(\hat{\mu}_{\text{ML}}) = \beta^2/n^2
\]  

(6.3.13)

From (6.3.13), we have the following MSE expressions of the MLEs:

\[
\text{MSE}(\hat{\beta}_{\text{ML}}) = (n-1)\beta^2/n^2 + \beta^2/n^2 = \beta^2/n
\]

\[
\text{MSE}(\hat{\mu}_{\text{ML}}) = 2\beta^2/n^2
\]

(6.3.14)

For the unbiased estimators in (6.3.10), we have

\[
\text{MSE}(\hat{\beta}_{\text{U}}) = \text{Var}(\hat{\beta}_{\text{U}}) = \beta^2/(n-1)
\]

\[
\text{MSE}(\hat{\mu}_{\text{U}}) = \text{Var}(\hat{\mu}_{\text{U}}) = \beta^2/(n(n-1))
\]

(6.3.15)

Thus, in term of MSE (i.e., risk under the squared error loss), for estimating \( \beta \), the unbiased estimator \( \hat{\beta}_u \) is comparable to the MLE \( \hat{\beta}_{\text{ML}} \) except for small \( n \); and for estimating \( \mu \), the unbiased estimator \( \hat{\mu}_u \) gives almost 50% improvement over the MLE \( \hat{\mu}_{\text{ML}} \) for large \( n \). The following figures provide the MSE curves of the above estimators.
Figure 6.3.3: MSE curves plotted against $\beta$ for $n = 8$: (a) for $\beta$ estimation; (b) for $\mu$ estimation.
6.4 Goodness of Fit Tests for Exponential Distribution

In parametric inference where the probability distribution of the data is assumed to fall in a known family of distributions (called the probability model of the data), it is imperative that the practitioner needs to verify the validity of the model before going into further inferences. If the assumed model itself is doubtful in explaining the data, then such an assumption can lead to incorrect inferences.

In the case of exponential distribution (or model), one needs to test the validity of it when a data set is available. Given a data set, the easiest thing one can do is to draw a probability histogram of the data and if the histogram shows a pattern similar to those in Figure 6.1.1, then an exponential model may seem plausible. However, one should go through a rigorous checking of the model validity known as a goodness of fit test and this is described below. We first consider the Exp($\beta, 0$) (or Exp($\beta$)) model, and then we generalize it to the Exp($\beta, \mu$) model.

Suppose that a random sample $X_1, X_2, \ldots, X_n$ is to be tested to see if it could come from the Exp($\beta, 0$) distribution. We consider testing the null hypothesis

$$H_0 : X_i's \text{ are iid } \text{Exp}(\beta, 0) \text{ random variables} \quad (6.4.1)$$

against the alternative hypothesis

$$H_1 : X_i's \text{ are not iid } \text{Exp}(\beta, 0) \text{ random variables} \quad (6.4.2)$$

Using the characterization in Theorem 6.2.4, testing $H_0$ in (6.4.1) is equivalent to test

$$H_0^* : Z_i's \ (1 \leq i \leq n - 1) \text{ are ordered statistics from CU}[0, 1] \quad (6.4.3)$$

where $Z_i's$ are defined in Theorem 6.2.4. Therefore, the problem now becomes to test the goodness of fit of the ordered observations $Z_1, Z_2, \ldots, Z_n$ from CU[0, 1].

A well-known test statistic for testing $H_0^*$ in (6.4.3) is

$$T^* = -2 \sum_{i=1}^{n-1} \ln Z_i \quad (6.4.4)$$
Chapter 6  Exponential Distribution

which is due to Pearson (1933). If \( T^* \) is ‘too small’ or ‘too large’, then \( H_0^* \) is rejected. The probability distribution of \( T^* \) under \( H_0^* \) is \( \chi^2_{2(n-1)} \) (i.e., a chi-square distribution with \( 2(n-1) \) degrees of freedom). Large values of \( T^* \) mean the points have moved toward 0 and small values of \( T^* \) imply the points have shifted toward 1. One can thus reject \( H_0^* \) (6.4.3), and hence \( H_0 \) (6.4.1) at level \( \alpha \), i.e.,

\[
T^* > \chi^2_{2(n-1),\alpha/2} \text{ or } T^* < \chi^2_{2(n-1),1-(\alpha/2)}
\]

and accept \( H_0 \) otherwise.

For testing a more general hypothesis

\[ H_0 : X_i \text{'s are iid Exp}(\beta,\mu) \text{ random variables} \quad (6.4.5) \]

against a suitable \( H_1 \) (i.e., \( X_i \text{'s are not iid Exp}(\beta,\mu) \)), we can use the characterization provided by Theorem 6.2.6. In practical applications, however, note that both \( \mu \) and \( \beta \) are unknown, and hence \( Y_1 \) (as defined in Theorem 6.2.6) cannot be observed. Therefore, one should only use the spacings \( Y_2, Y_3, \ldots, Y_n \) and then define

\[
S^* = \sum_{i=2}^{n} Y_i \quad \text{and} \quad Z_j^{**} = \frac{1}{S^*} \sum_{i=2}^{j} Y_i, \quad \text{for } 2 \leq j \leq n - 1 \quad (6.4.6)
\]

If the \( \text{Exp}(\beta,\mu) \) model is correct for original random variables \( X_1, X_2, \ldots, X_n \), then \( Z_j^{**} \)'s should act like \( (n-2) \) ordered observations from \( \text{CU}[0,1] \). Similar to (6.4.4), a test statistic for testing \( H_0 \) in (6.4.5) is

\[
T^{**} = -2 \sum_{i=2}^{n-2} \ln Z_i^{**} \quad (6.4.7)
\]

which, under \( H_0 \) in (6.4.5), follows \( \chi^2_{2(n-2)} \) distribution. Therefore, reject \( H_0 \) if \( T^{**} > \chi^2_{2(n-2),\alpha/2} \) or \( T^{**} < \chi^2_{2(n-2),1-(\alpha/2)} \); and accept \( H_0 \) otherwise.

Example 6.4.1: Consider the given ordered data set

\[
\begin{array}{cccc}
48.96 & 54.67 & 100.75 & 120.68 \\
148.34 & 228.43 & 293.45 & 415.82
\end{array}
\]

which is to be modeled by a suitable exponential distribution. Test the goodness of fit for exponential distributions.

First, we would like to see whether the given data are from \( \text{Exp}(\beta,0) \) by testing the hypotheses,

\[ H_0 : X_i \text{'s are iid Exp}(\beta,0) \text{ random variables} \]

\(^4\)Cutoff points are available from \( \chi^2 \)-distribution tables in the Appendix.
H$_1$: $X_i$’s are not *iid* $\text{Exp}(\beta, 0)$ random variables

As shown in (6.4.4), the test statistic for testing the above statements is

$$T^* = -2 \sum_{i=1}^{n-1} \ln Z_i$$

where $Z_i$’s are defined in Theorem 6.2.4. The following table provides necessary computational results which are needed for calculating $T^*$. It is noted that $S = 1411.10$.

\[
\begin{array}{|c|c|c|c|c|}
\hline
i & X_i & \sum_{k=1}^{i} X_k & Z_i = \frac{\sum_{k=1}^{i} X_k}{S} & \ln Z_i \\
\hline
1 & 48.96 & 48.96 & 0.034696 & -3.36112 \\
2 & 54.67 & 103.63 & 0.038743 & -3.25081 \\
3 & 100.75 & 204.38 & 0.071398 & -2.63948 \\
4 & 120.68 & 325.06 & 0.085522 & -2.45898 \\
5 & 148.34 & 473.40 & 0.105124 & -2.25262 \\
6 & 228.43 & 701.83 & 0.161881 & -1.82090 \\
7 & 293.45 & 995.28 & 0.207958 & -1.57042 \\
8 & 415.82 & 1411.10 & & \\
\hline
\end{array}
\]

Table 6.4.1: Computational Results for Calculating $T^*$.

Adding the last column of Table 6.4.1, we can obtain the test statistic as

$$T^* = -2 \sum_{i=1}^{n-1} \ln Z_i = (-2) \times (-17.3543) = 34.70865$$

At the significant level $\alpha = 0.05$, the Chi-square critical values with $2(n-1) = 14$ degrees of freedom can be found from the $\chi^2$-distribution table in the Appendix as $\chi^2_{14,(\alpha/2)} = 26.119$. Therefore, we reject the statement $H_0$ because $T^* = 34.70865 > \chi^2_{14,(\alpha/2)} = 26.119$. Hence, it is unlikely that the given data are from the exponential distribution $\text{Exp}(\beta, 0)$.

Now check whether the data could have come from a suitable $\text{Exp}(\beta, \mu)$ distribution. Adding the last column of Table 6.4.2, we can obtain the test statistic as

$$T^{**} = -2 \sum_{i=1}^{n-1} \ln Z_i = (-2) \times (-11.51538) = 23.03076$$
Table 6.4.2: Computational Results for Calculating $T^{**}$.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$Y_i$</th>
<th>$Z_i^{**}$</th>
<th>$\ln Z_i^{**}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>51.39</td>
<td>0.0105951</td>
<td>-4.54736</td>
</tr>
<tr>
<td>3</td>
<td>460.8</td>
<td>0.105599</td>
<td>-2.24811</td>
</tr>
<tr>
<td>4</td>
<td>219.23</td>
<td>0.150798</td>
<td>-1.89181</td>
</tr>
<tr>
<td>5</td>
<td>331.92</td>
<td>0.21923</td>
<td>-1.51762</td>
</tr>
<tr>
<td>6</td>
<td>1041.17</td>
<td>0.433889</td>
<td>-0.83497</td>
</tr>
<tr>
<td>7</td>
<td>910.28</td>
<td>0.621563</td>
<td>-0.47551</td>
</tr>
<tr>
<td>8</td>
<td>1835.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At the significant level $\alpha = 0.05$, the Chi-square critical value with $2(n - 2) = 12$ degrees of freedom can be found from the $\chi^2$-distribution table in Appendix as $\chi^2_{12, (\alpha/2)} = 23.337$ and $\chi^2_{12, 1-(\alpha/2)} = 4.404$. Therefore, we accept the statement $H_0$ because $T^{**} = 23.03076$ falls between the two cut-off points. Hence, it seems plausible that the given data are obtained from a two-parameter exponential distribution $\text{Exp}(\beta, \mu)$. 
Chapter 7

Gamma Distribution

7.1 Preliminaries

In Chapter 2, the gamma distribution and its basic properties were introduced. A random variable $X$ is said to have a (two parameter) gamma distribution with scale parameter $\beta$ and shape parameter $\alpha$ provided the pdf of $X$ is given as

$$f(x|\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0 \quad (7.1.1)$$

The above pdf includes the following special cases:

(i) If $\alpha = k/2$, for any integer $k$, and $\beta = 2$, then we get the $\chi^2_k$ (Chi-square with $k$ degrees of freedom) pdf as

$$f(x|k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1}e^{-x/2}, \quad 0 < x < \infty, \quad k > 0 \quad (7.1.2)$$

(ii) If $\alpha = 1$, then we obtain the exponential distribution with scale parameter $\beta$ as discussed in Chapter 6.

(iii) If $\beta = 1$, then the gamma pdf (7.1.1) is called the standard gamma distribution with shape parameter $\alpha$ as

$$f(x|\alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1}e^{-x}, \quad 0 < x < \infty, \quad \alpha > 0 \quad (7.1.3)$$

From (7.1.3) one can have

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \quad (7.1.4)$$
which, by integration by parts, yields

\[ \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) = (\alpha - 1)(\alpha - 2)\Gamma(\alpha - 2) = \ldots \] (7.1.5)

As a result, if \(\alpha\) is an integer, then

\[ \Gamma(\alpha) = (\alpha - 1)! = (\alpha - 1)(\alpha - 2) \ldots 3 \cdot 2 \cdot 1 \] (7.1.6)

The quantity \(\Gamma_c(\alpha)\) defined as

\[ \Gamma_c(\alpha) = \int_0^c x^{\alpha-1}e^{-x}dx \] (7.1.7)

is called an incomplete gamma function. Thus, if \(X\) follows a standard gamma distribution with shape parameter \(\alpha\), then the cdf of \(X\) is

\[ P(X \leq x) = \frac{1}{\Gamma(\alpha)} \int_0^x u^{\alpha-1}e^{-u}du = \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)} \] (7.1.8)

A generalization of (7.1.1) known as a three-parameter gamma distribution is given as

\[ f(x | \alpha, \beta, \mu) = \frac{1}{\beta^\alpha\Gamma(\alpha)}(x - \mu)^{\alpha-1}e^{-(x-\mu)/\beta}, \] (7.1.9)

where the extra parameter \(\mu\) is called the shift or location parameter and \(\alpha\) and \(\beta\) are the same as discussed in (7.1.1). This three-parameter gamma distribution is often called a **shifted gamma distribution**.

In most of the real life applications where gamma distribution is well suited one usually applies the regular two-parameter gamma distribution given in (7.1.1). If there is a compelling reason to use a three-parameter gamma distribution, then one should test a null hypothesis \(H_0 : \mu = 0\) against \(H_1 : \mu \neq 0\). If the null hypothesis is accepted, then the three-parameter gamma model can be reduced to the regular two-parameter one, otherwise the three-parameter gamma model is retained.

The following figure plots the pdf of a two-parameter gamma distribution (as given in (7.1.1)) for various values of \(\alpha\) and \(\beta\).

It can be seen from **Figure 7.1.1** that, as \(\alpha\) increases, the shape of the pdf becomes similar to the normal pdf. In fact, for the three-parameter gamma distribution (7.1.9), it can be verified that
7.1 Preliminaries

Figure 7.1.1: Two-parameter Gamma pdfs.

\[
\lim_{\alpha \to \infty} P \left[ \left( \frac{(X - \mu)}{\beta} - \alpha \right) \sqrt{\alpha} \leq u \right] = \Phi(u) \quad (7.1.10)
\]

and in particular if \( X \) follows a \( \chi_k^2 \)-distribution, then

\[
\lim_{k \to \infty} P \left[ \frac{(X - k)}{\sqrt{2k}} \leq u \right] = \Phi(u)
\]

where \( \Phi(\cdot) \) is the standard normal cdf.

In real life problems, gamma distribution gives useful representation of many physical situations. It can be used to make realistic adjustments to exponential distribution in representing lifetime in ‘life-testing’
experiments. In the following, we provide some examples where gamma distribution seems to give a good fit. The shape parameter depends on such factors as the applied pressure and the composite wall thickness.

**Example 7.1.1:** Lifetimes of pressure vessels constructed of fiber/epoxy composite materials wrapped around metal liners may be modeled by a gamma distribution. Keating, Glaser, and Ketchum (1990) reported the failure times (in hours) of 20 similarly constructed vessels subjected to a certain constant pressure with the following observations:

<table>
<thead>
<tr>
<th>Lifetime of Vessels (hr)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 – 298.0</td>
<td>10</td>
<td>0.50 (50%)</td>
</tr>
<tr>
<td>298.0 – 596.0</td>
<td>2</td>
<td>0.10 (10%)</td>
</tr>
<tr>
<td>596.0 – 894.0</td>
<td>3</td>
<td>0.15 (15%)</td>
</tr>
<tr>
<td>894.0 – 1192.0</td>
<td>1</td>
<td>0.05 (5%)</td>
</tr>
<tr>
<td>1192.0 – 1490.0</td>
<td>2</td>
<td>0.10 (10%)</td>
</tr>
<tr>
<td>1490.0 – 1788.0</td>
<td>2</td>
<td>0.10 (10%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>20</strong></td>
<td>1.00 (100%)</td>
</tr>
</tbody>
</table>

Table 7.1.1: Frequency Table for Pressure Vessel Lifetime Data.

The following histogram (Figure 7.1.2) along with a matching gamma *pdf* are drawn with the relative frequency values in Table 7.1.1.
Figure 7.1.2: Histogram of the vessel lifetime data and a matching gamma pdf with $\alpha = 1.45$ and $\beta = 300$.

**Example 7.1.2:** The following data on the times between successive failures of air conditioning (A/C) equipment in Boeing 720 aircrafts have been provided by Proschan (1963). The data in operating hours are as follows:

\[
\begin{align*}
90 & \quad 10 & \quad 60 & \quad 186 & \quad 61 & \quad 49 \\
14 & \quad 24 & \quad 56 & \quad 20 & \quad 79 & \quad 84 \\
44 & \quad 59 & \quad 29 & \quad 118 & \quad 25 & \quad 156 \\
310 & \quad 76 & \quad 26 & \quad 44 & \quad 23 & \quad 62 \\
130 & \quad 208 & \quad 70 & \quad 101 & \quad 208
\end{align*}
\]
Chapter 7  Gamma Distribution

The following frequency table is made with four classes over the range from 9.0 to 311.0 hours. The endpoints are chosen conveniently to get class boundaries which do not coincide with the observations.

<table>
<thead>
<tr>
<th>Operating Hours</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 – 59.5</td>
<td>13</td>
<td>0.448 (44.8%)</td>
</tr>
<tr>
<td>59.5 – 119.5</td>
<td>10</td>
<td>0.345 (34.5%)</td>
</tr>
<tr>
<td>119.5 – 179.5</td>
<td>2</td>
<td>0.07 (7%)</td>
</tr>
<tr>
<td>179.5 – 239.5</td>
<td>3</td>
<td>0.10 (10.%)</td>
</tr>
<tr>
<td>239.5 – 311.0</td>
<td>1</td>
<td>0.03 (3%)</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>≈ 1.00 (≈ 100%)</td>
</tr>
</tbody>
</table>

Table 7.1.2: Frequency Table for Operating Hours of A/C Equipment.

The following histogram (Figure 7.1.3) is drawn with the relative frequencies shown in Table 7.1.2.

Apart from the lifetime data, gamma distribution is widely used in seismology, hydrology, and economics. Such examples are given by McDonald and Jensen (1979), Bougeault (1982), Coe and Stern (1982), Singh and Singh (1985), among a host of other authors. The next example shows the use of gamma distribution in entomology.

Example 7.1.3: Matis, Rubink, and Makela (1992) used the gamma distribution to model the ‘transit-time distribution’ of Africanized honey bee spread through northern Guatemala and Mexico. Data were collected on the first capture times of the honey bee at various monitoring transects in northern Guatemala and on the eastern and western coasts of Mexico. The time intervals between consecutive sightings (in months) are reported where the distances between transects are fixed and the random variable is the time interval between consecutive observations. The transit time data (in months/ 100km) consists of 45 observations.

5.3  1.8  4.2  5.7  3.8  0.8  1.4  3.5  17.5
4.6  0.8  6.3  2.9  0.6  1.9  2.0  6.7  5.5
2.5  2.2  6.7  5.7  10.0  3.3  3.5  20.0  1.6
8.3  4.8  20.0  3.6  8.2  1.3  4.0  5.0  1.7
2.0  2.9  19.2  1.1  1.4  1.5  3.2  8.6  2.2

The following histogram (Figure 7.1.4) is drawn with the relative frequencies shown in Table 7.1.3.
Figure 7.1.3: Histogram of the A/C equipment operating hours and a matching pdf with $\alpha = 2.0$ and $\beta = 49.0$.

<table>
<thead>
<tr>
<th>Transit Time</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50 – 2.95</td>
<td>19</td>
<td>0.422 (42.2%)</td>
</tr>
<tr>
<td>2.95 – 5.40</td>
<td>12</td>
<td>0.267 (26.7%)</td>
</tr>
<tr>
<td>5.40 – 7.85</td>
<td>6</td>
<td>0.133 (13.3%)</td>
</tr>
<tr>
<td>7.85 – 10.30</td>
<td>4</td>
<td>0.089 (8.9%)</td>
</tr>
<tr>
<td>10.30 – 12.75</td>
<td>0</td>
<td>0.000 (0.0%)</td>
</tr>
<tr>
<td>12.75 – 15.20</td>
<td>0</td>
<td>0.000 (0.0%)</td>
</tr>
<tr>
<td>15.20 – 17.65</td>
<td>1</td>
<td>0.022 (2.2%)</td>
</tr>
<tr>
<td>17.65 – 20.10</td>
<td>3</td>
<td>0.067 (6.7%)</td>
</tr>
</tbody>
</table>

Total 45 $\approx 1.00$ ($\approx 100\%$)

Table 7.1.3: Frequency Table for Transit Time of Africanized Honey Bee.
Figure 7.1.4: Histogram of Transit Time of Africanized Honey Bee and a matching pdf with $\alpha = 1.15$ and $\beta = 4.0$.

**Notational Clarification:** A two-parameter gamma distribution of the form (6.1.1) will be called $G(\alpha, \beta)$-distribution. And, a three-parameter gamma distribution of the form (7.1.9) will be called a $G(\alpha, \beta, \mu)$-distribution. Therefore, $G(\alpha, \beta)$ implies the $G(\alpha, \beta, 0)$-distribution, and vice-versa. Though $\alpha$ is used to denote the shape parameter, it is also used to indicate the significance level in a hypothesis testing problem and the distinction should be clear from the context.
7.2 Characterization of Gamma Distribution

The classical result on characterizing Gamma distribution was due to Lukacs (1955), part of which is widely used in statistical inference. This well-known result says that if $X_1$ and $X_2$ are two independent random variables with $X_i$ having a Gamma($\alpha_i, \beta$) distribution, $i = 1, 2$, then $U = (X_1 + X_2)$ and $V = X_1/(X_1 + X_2)$ are independent. Moreover, $U$ follows Gamma($\alpha_1 + \alpha_2, \beta$) distribution and $V$ follows Beta($\alpha_1, \alpha_2$) distribution. The proof of this result is straightforward. One needs to find the joint pdf of $(U, V)$ using the method of transformation which factorizes into independent pieces with respective structures. In the following, Lukacs’ (1955) original characterization result is presented.

**Theorem 7.2.1** Let $X_1$ and $X_2$ be independent nonnegative random variables with cdfs $F_1$ and $F_2$ (or pdfs $f_1$ and $f_2$) respectively. Then $U_1 = (X_1/X_2)$ and $U_2 = (X_1 + X_2)$ are independent if and only if $F_1$ and $F_2$ are gamma distributions (cdfs) with the same scale parameter.

Khatri and Rao (1968) obtained the following characterization of the gamma distribution based on constancy of various regression functions.

**Theorem 7.2.2** (a) If $X_1, X_2, \ldots, X_n$ ($n \geq 3$) are independent positive random variables and

\[ Y_1 = \sum_{i=1}^{n} b_{1i} X_i, \quad Y_j = \prod_{i=1}^{n} X_i^{b_{ji}}, \quad j = 2, 3, \ldots, n, \quad b_{1i} \neq 0, \quad 1 \leq i \leq n \]

with the $(n-1) \times n$ matrix \( \left( b_{ji} \right) \) (j = 2, 3, ... , n; i = 1, 2, ... , n) nonsingular, then the constancy of $E[Y_1|Y_2, \ldots, Y_n]$ ensures that each $X_i$ must have a common gamma distribution.

(b) In the above conditions of (a), if $E[1/X_j] \neq 0$ (j = 1, 2, ... , n) and $Z_1 = \sum_{i=1}^{n} b_{1i}/X_i$, $Z_j = \sum_{i=1}^{n} b_{ji} X_j$, $j = 2, 3, \ldots, n$ with $b_{ji}$’s satisfying the same conditions as in (a), then the constancy of $E[Z_1|Z_2, \ldots, Z_n]$ ensures that each $X_i$ has a gamma distribution, not necessarily the same for all $i$.

(c) Under the same conditions as in (a), if $E[X_1 \ln X_1] < \infty$, then the constancy of $E\left[ \sum_{j=1}^{n} a_j X_j \prod_{i=1}^{n} X_i^{b_{ji}} \right]$ with $\sum_{j=1}^{n} a_j b_j = 0$, $|b_n| > \max_{1 \leq j \leq n-1}(|b_j|)$, and $a_j b_j/(a_n b_n) < 0$ for all $j = 1, 2, \ldots, (n-1)$ ensures that $X_1$ has a gamma distribution.
(d) If \(X_1, X_2, \ldots, X_n\) are iid positive random variables with \(E[1/X_i] \neq 0\), \(i = 1, 2, \ldots, n\), then constancy of
\[
E\left[\sum_{j=1}^{n} \frac{a_j}{X_j} \left| \sum_{j=1}^{n} b_j X_j \right.\right]
\]
with same conditions on \(a_j\)'s and \(b_j\)'s as in (c), ensures that the common distribution of \(X_i\)'s is a gamma distribution.

**Corollary 7.2.1** (i) In Theorem 7.2.2 (parts (a) and (b)), if one takes \(b_{11} = \ldots = b_{1n} = 1, b_{j,j-1} = -1, b_{j,j} = 1\) and all other \(b_{ji}\)'s zero, then the condition in part (a) becomes the constancy of
\[
E\left[\sum_{j=1}^{n} X_j \left| X_2/X_1, X_3/X_1, \ldots, X_n/X_1 \right.\right]
\]
and the condition of part (b) becomes the constancy of
\[
E\left[\sum_{j=1}^{n} 1/X_j \left| X_2 - X_1, X_3 - X_1, \ldots, X_n - X_1 \right.\right]
\]

(ii) In Theorem 7.2.2 (parts (b) and (c)), if one takes \(a_1 = a_2 = \ldots = a_n = 1, b_n = n-1, b_1 = b_2 = \ldots = b_{n-1} = -1\), then the condition in part (c) reduces to the constancy of
\[
E\left[\sum_{j=1}^{n} X_j \left| \prod_{j=1}^{n-1} (X_n/X_j) \right.\right]
\]
and the condition in part (d) reduces to the constancy of
\[
E\left[\sum_{j=1}^{n} 1/X_j \left| X_n - \bar{X} \right.\right]
\]
where \(\bar{X} = \sum_{j=1}^{n} X_j/n\).

Hall and Simons (1969) gave a characterization of the gamma distribution in terms of quadratic expression of two random variables as given below.

**Theorem 7.2.3** Suppose \(X_1\) and \(X_2\) are independent random variables with finite second moments and characteristic functions \(\psi_1(t)\) and \(\psi_2(t)\), respectively. Then for some \(a_1, a_2\),
\[
E\left[X_i^2 \left| X_1 + X_2 \right.\right] = a_i(X_1 + X_2)^2 \quad a.s., \quad i = 1, 2
\]
7.3 Estimation of Parameters

We first consider estimation of parameters $\alpha$ and $\beta$ in a two-parameter gamma distribution. If $X_1, X_2, \ldots, X_n$ are independent random variables each having pdf (7.1.1), then the maximum likelihood estimators of $\alpha$ and $\beta$ are $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ respectively obtained by solving the system of equations:

$$
\hat{\alpha}_{ML} \hat{\beta}_{ML} = \bar{X} \quad \text{and} \quad \ln \hat{\beta}_{ML} + \psi(\hat{\alpha}_{ML}) = \frac{1}{n} \sum_{i=1}^{n} \ln X_i \quad (7.3.1)
$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $\psi(\alpha) = \frac{\partial}{\partial \alpha} \ln \Gamma(\alpha)$, called the di-gamma function. Substituting $\hat{\beta}_{ML} = \frac{\bar{X}}{\hat{\alpha}_{ML}}$ in the second equation of (7.3.1), we obtain the following equation for $\hat{\alpha}_{ML}$

$$
\ln \hat{\alpha}_{ML} - \psi(\hat{\alpha}_{ML}) = \ln \left[ \frac{\text{arithmetic mean (AM) of } X_1, X_2, \ldots, X_n}{\text{geometric mean (GM) of } X_1, X_2, \ldots, X_n} \right] \quad (7.3.2)
$$

Given the values of $X_i$, $i = 1, 2, \ldots, n$, one can now find the value of $\hat{\alpha}_{ML}$ by solving (7.3.2) which then in turn gives the value of $\hat{\beta}_{ML}$ as $\hat{\beta}_{ML} = \frac{\bar{X}}{\hat{\alpha}_{ML}}$.

From the above derivation it is obvious that $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ do not have closed expressions, and as a result, it is impossible to study analytical sampling properties of these estimators for fixed sample sizes. However, some asymptotic properties (as $n \to \infty$) of $\sqrt{n}\hat{\alpha}_{ML}$ and $\sqrt{n}\hat{\beta}_{ML}$ are known to be as

$$
\lim_{n \to \infty} E \left[ \sqrt{n} (\hat{\alpha}_{ML} - \alpha) \right] = 0
$$

$$
\lim_{n \to \infty} E \left[ \sqrt{n} (\hat{\beta}_{ML} - \beta) \right] = 0
$$

$$
\lim_{n \to \infty} \text{Var} (\sqrt{n}\hat{\alpha}_{ML}) = \frac{\alpha}{(\alpha \psi'(\alpha) - 1)} \quad (7.3.3)
$$
Chapter 7 Gamma Distribution

\[
\begin{align*}
\lim_{n \to \infty} \text{Var} \left( \sqrt{n} \hat{\beta}_{\text{ML}} \right) &= \frac{\beta^2 \psi'(\alpha)}{(\alpha \psi'(\alpha) - 1)} \\
\text{and} \quad \lim_{n \to \infty} \text{Corr} \left( \hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}} \right) &= -\frac{1}{\sqrt{\alpha \psi'(\alpha)}}
\end{align*}
\]

where \( \psi(\alpha) \) is the di-gamma function mentioned above. In other words, asymptotically, both \( \hat{\alpha}_{\text{ML}} \) and \( \hat{\beta}_{\text{ML}} \) are unbiased and

\[
\begin{align*}
\text{Var} \left( \hat{\alpha}_{\text{ML}} \right) &= \frac{\alpha}{n(\alpha \psi'(\alpha) - 1)} + O(n^{-2}) \\
\text{Var} \left( \hat{\beta}_{\text{ML}} \right) &= \frac{\beta^2 \psi'(\alpha)}{n(\alpha \psi'(\alpha) - 1)} + O(n^{-2})
\end{align*}
\]

\( (7.3.4) \)

Berman (1981) proved that for fixed sample size \( n \geq 2 \), the estimator \( \hat{\alpha}_{\text{ML}} \) is always positively biased. Using the reparameterization \( \lambda = \frac{1}{\beta} \), it was also shown that \( \hat{\lambda}_{\text{ML}} = \frac{1}{\hat{\beta}_{\text{ML}}} \) is also positively biased, i.e., \( E \left( \hat{\lambda}_{\text{ML}} \right) > \lambda \) for all \( \alpha, \lambda > 0 \). Shenton and Bowman (1972) have given the mean expressions of \( \hat{\alpha}_{\text{ML}} \) and \( \hat{\lambda}_{\text{ML}} \) as (assuming that \( n > 3 \))

\[
\begin{align*}
E \left[ \hat{\alpha}_{\text{ML}} - \alpha \right] &= \frac{3\alpha}{(n - 3)} \left\{ 1 - \frac{2}{9\alpha} + \frac{(n - 1)}{27\alpha^2} + \frac{7(n^2 - 1)}{162(n + 3)n^2} + \ldots \right\} \\
E \left[ \hat{\lambda}_{\text{ML}} - \lambda \right] &= \frac{n\alpha\lambda}{(n - 3)(n\alpha - 1)} \left\{ 3 + \frac{1}{3\alpha} - \frac{3}{n\alpha} + \frac{(n - 1)}{9n\alpha^2} + \ldots \right\}
\end{align*}
\]

\( (7.3.5) \)

From the expressions in (7.3.5), Anderson and Ray (1975) suggested the modified MLEs of \( \alpha \) and \( \lambda \) as

\[
\begin{align*}
\hat{\alpha}_{\text{MML}} &= \frac{(n - 3)}{n} \hat{\alpha}_{\text{ML}} + \frac{2}{3n} \text{ and} \\
\hat{\lambda}_{\text{MML}} &= \hat{\lambda}_{\text{ML}} g(\hat{\alpha}_{\text{MML}})
\end{align*}
\]

\( (i.e., \hat{\beta}_{\text{MML}} = \frac{\hat{\beta}_{\text{ML}}}{g(\hat{\alpha}_{\text{MML}})} \) \)

where

\[
g(\alpha) = \left\{ 1 + \frac{3n\alpha}{(n - 3)(n\alpha - 1)} \left[ 1 + \frac{1}{9\alpha} - \frac{1}{n\alpha} + \frac{(n - 1)}{27n\alpha^2} \right] \right\}^{-1}
\]

The above modified MLEs have smaller bias, and this reduction of bias is not gained at the cost of an increase in MSE (mean squared error). Extensive numerical study shows that \( \hat{\alpha}_{\text{MML}} \) and \( \hat{\lambda}_{\text{MML}} \) can provide substantial improvements over \( \hat{\alpha}_{\text{ML}} \) and \( \hat{\lambda}_{\text{ML}} \), respectively, in terms of bias and MSE for small sample sizes \( (n \leq 20) \).
7.3 Estimation of Parameters

Using the parameters $\alpha$ and $\beta$ (as in (7.1.1)), Shenton and Bowman (1969) have shown that

$$
E\left[\left(\frac{\hat{\alpha}_{ML}}{\alpha}\right)^s\right] \approx \frac{n^s}{(n-3)(n-5)\cdots(n-2s-1)} \left(1 - \frac{s(s+1)}{3n\alpha}\right) \quad \text{and}
$$
$$
E\left[\frac{\hat{\beta}_{ML}}{\beta}\right] = 1 - \frac{1}{n} + O(\alpha^{-2})
$$

(7.3.7)

It is seen that though $\hat{\alpha}_{ML}$ is positively biased, the estimator $\hat{\beta}_{ML}$ is negatively biased.

Since the MLEs (and hence the modified MLEs) of the gamma parameters $\alpha$ and $\beta$ do not have closed expressions, a simple and convenient way of estimating parameters is through the method of moments. The method of moments estimators (MMEs) of $\alpha$ and $\beta$ are obtained by equating the first two population moments with those sample analogues, i.e., $\hat{\alpha}_{MM}$ and $\hat{\beta}_{MM}$ are obtained by solving the following equations:

$$
\hat{\alpha}_{MM} \hat{\beta}_{MM} = \bar{X} \quad \text{and} \quad \hat{\alpha}_{MM}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2
$$

which produce

$$
\hat{\alpha}_{MM} = \frac{n(\bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{and} \quad \hat{\beta}_{MM} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{nX}
$$

(7.3.8)

Though $\hat{\alpha}_{MM}$ and $\hat{\beta}_{MM}$ have simpler expressions, the MLEs (or the modified MLEs) perform consistently better than the MMEs in terms of bias as well as MSE even for small sample sizes.

We now consider estimation of parameters for a three-parameter gamma distribution.

If we have iid observations $X_1, X_2, \ldots, X_n$ from a three-parameter gamma distribution with parameters $\alpha, \beta$, and $\mu$ (as shown in (7.1.9)), then the MLEs $\hat{\alpha}_{ML}, \hat{\beta}_{ML}$ and $\hat{\mu}_{ML}$ are obtained by solving the following system of equations:

$$
\sum_{i=1}^{n} \ln(X_i - \hat{\mu}_{ML}) - n\ln\hat{\beta}_{ML} - n\psi(\hat{\alpha}_{ML}) = 0
$$
$$
\sum_{i=1}^{n} (X_i - \hat{\mu}_{ML}) - n\hat{\alpha}_{ML}\hat{\beta}_{ML} = 0
$$
$$
\sum_{i=1}^{n} (X_i - \hat{\mu}_{ML})^{-1} - n\left(\hat{\beta}_{ML}(\hat{\alpha}_{ML} - 1)\right)^{-1} = 0
$$

(7.3.9)
From the third equation in (7.3.9), it is seen that if \( \hat{\alpha}_{\text{ML}} \) is less than 1, then some \( X_i \)'s must be less than \( \hat{\mu}_{\text{ML}} \) which is contradictory since \( X_i \)'s are all greater than \( \mu \) and hence \( \hat{\mu}_{\text{ML}} \) must be less than \( X_i \)'s. Therefore, the system of equations in (7.3.9) has to be solved iteratively. One can use MME estimators of \( \alpha, \beta, \mu \) as starting (or initial) values; then use the first equation to determine a new value of \( \hat{\beta}_{\text{ML}} \), given \( \hat{\alpha}_{\text{ML}} \) and \( \hat{\mu}_{\text{ML}} \); then the second equation for a new value of \( \hat{\mu}_{\text{ML}} \), given \( \hat{\alpha}_{\text{ML}} \) and \( \hat{\beta}_{\text{ML}} \); and then the third equation for a new value of \( \hat{\alpha}_{\text{ML}} \), given \( \hat{\beta}_{\text{ML}} \) and \( \hat{\mu}_{\text{ML}} \). Continue this process until the stabilized values of \( \hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}}, \) and \( \hat{\mu}_{\text{ML}} \) are achieved.

**Example 7.3.1:** Demonstrate the MLE computation by drawing a random sample of size \( n = 10 \) from a gamma distribution with \( \alpha = 5, \beta = 1, \) and \( \mu = 0 \), and then follow the iterative process until \( \hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}}, \) and \( \hat{\mu}_{\text{ML}} \) values (stabilized) are achieved.

First, a random sample of size \( n = 10 \) is generated from \( G(5.0, 1.0, 0.0) \) distribution. The data set is:

\[
\begin{align*}
1.8671 & \quad 3.9899 & \quad 5.4569 & \quad 2.7314 & \quad 3.7961 & \quad 5.2041 & \quad 7.4768 & \quad 3.0366 \\
3.7333 & \quad 5.1969 \\
\end{align*}
\]

Then calculate the MLEs using (7.3.9) to get \( \hat{\alpha}_{\text{ML}} = 5.7375, \hat{\beta}_{\text{ML}} = 0.6586, \) and \( \hat{\mu}_{\text{ML}} = 0.4700 \).

If we equate the first three population (i.e., \( G(\alpha, \beta, \mu) \)) moments with their sample counterparts, then we have the method of moments estimators \( \hat{\mu}_{\text{MM}}, \hat{\alpha}_{\text{MM}}, \) and \( \hat{\beta}_{\text{MM}} \) obtained by solving the following three equations:

\[
\begin{align*}
(i) & \quad \mu + \alpha \beta = \bar{X} \\
(ii) & \quad \alpha \beta^2 = S/n \\
(iii) & \quad 2\beta(4\alpha^2 + 3\alpha + 1) = T/S \\
\end{align*}
\]

where \( S = \sum_{i=1}^{n} (X_i - \bar{X})^2 \) and \( T = \sum_{i=1}^{n} (X_i - \bar{X})^3 \). (The third equation above is obtained by equating the third sample moment with its population counterpart, and then using the second equation above.)

In the following, we provide an approach to extend the two-parameter modified maximum likelihood estimation for the three-parameter case.
After obtaining $\hat{\mu}_{\text{ML}}, \hat{\alpha}_{\text{ML}}, \text{and } \hat{\beta}_{\text{ML}}$ (solving (7.3.9) as discussed earlier), get
\[
X_i^* = X_i - \hat{\mu}_{\text{ML}}, \quad i = 1, 2, \ldots, n \tag{7.3.11}
\]
The new observations $X_1^*, \ldots, X_n^*$ are now modelled by a $G(\alpha, \beta)$ distribution (two parameter gamma model).

Obtain $\hat{\alpha}_{\text{MM}}$ and $\hat{\beta}_{\text{MM}}$ based on $X_i^*$’s as described in (7.3.6). Then, plug in $\hat{\alpha}_{\text{MM}}$ and $\hat{\beta}_{\text{MM}}$ in the third equation in (7.3.9) and solve for $\mu$ to get $\hat{\mu}_{\text{MM}}$.

The above three methods (i.e., MM estimation, ML estimation, and modified ML estimation) are now used for estimating the parameters of a three-parameter gamma distribution used to model the data set in Example 7.1.2. The estimates are given below.

\[
\hat{\mu}_{\text{MM}} = -7.89; \quad \hat{\alpha}_{\text{MM}} = 1.73; \quad \text{and } \hat{\beta}_{\text{MM}} = 52.96
\]
\[
\hat{\mu}_{\text{ML}} = 2.76; \quad \hat{\alpha}_{\text{ML}} = 1.85; \quad \text{and } \hat{\beta}_{\text{ML}} = 42.73
\]
\[
\hat{\mu}_{\text{MML}} = 0.63; \quad \hat{\alpha}_{\text{MML}} = 1.38; \quad \text{and } \hat{\beta}_{\text{MML}} = 59.86
\]

In the following we provide the bias and MSE of the above three estimators through simulation.

Generate $X_1, \ldots, X_n$ from $G(\alpha, \beta, \mu)$ $M = 20,000$ times with various $\alpha$ values, $\beta = 1$ and $\mu = 0$. Estimate the bias and MSE of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$ (following each of the methods discussed above).

\[
\text{Bias}(\hat{\alpha}) \approx \left( \frac{\sum_{j=1}^{M} \hat{\alpha}^{(j)} / M }{M} \right) - \alpha
\]
\[
\text{Bias}(\hat{\beta}) \approx \left( \frac{\sum_{j=1}^{M} \hat{\beta}^{(j)} / M }{M} \right) - \beta \tag{7.3.12}
\]
\[
\text{Bias}(\hat{\mu}) \approx \left( \frac{\sum_{j=1}^{M} \hat{\mu}^{(j)} / M }{M} \right) - \mu
\]

and

\[
\text{MSE}(\hat{\alpha}) \approx \sum_{j=1}^{M} \left( \hat{\alpha}^{(j)} - \bar{\alpha} \right)^2 / M
\]
\[
\text{MSE}(\hat{\beta}) \approx \sum_{j=1}^{M} \left( \hat{\beta}^{(j)} - \bar{\beta} \right)^2 / M \tag{7.3.13}
\]
\[
\text{MSE}(\hat{\mu}) \approx \sum_{j=1}^{M} \left( \hat{\mu}^{(j)} - \bar{\mu} \right)^2 / M
\]

where $\hat{\alpha}^{(j)}$, $\hat{\beta}^{(j)}$, $\hat{\mu}^{(j)}$ are the estimated values of $\alpha$, $\beta$, and $\mu$ based on the $j$th ($j = 1, \ldots, M$) replicated sample, and $\bar{\alpha} = \sum_{j=1}^{M} \hat{\alpha}^{(j)}/M$. 

© 2006 by Taylor & Francis Group, LLC
\bar{\beta} = \sum_{j=1}^{M} \hat{\beta}^{(j)}/M, \bar{\mu} = \sum_{j=1}^{M} \hat{\mu}^{(j)}/M. The biases and MSEs are plotted against \alpha.

(a) Bias of \alpha estimators (n = 10).

(b) Bias of \beta estimators (n = 10).
7.3 Estimation of Parameters

Figure 7.3.1: Graphs of bias of various estimators plotted against \( \alpha \) when \( n = 10 \) and data come from \( G(\alpha, 1, 0) \) distribution: (a) for \( \alpha \) estimation, (b) for \( \beta \) estimation, (c) for \( \mu \) estimation.
(a) MSE of $\alpha$ estimators ($n = 10$).

(b) MSE of $\beta$ estimators ($n = 10$).
7.3 Estimation of Parameters

Figure 7.3.2: Graphs of MSE of various estimators plotted against \( \alpha \) when \( n = 10 \) and data come from \( G(\alpha, 1, 0) \) distribution: (a) for \( \alpha \) estimation, (b) for \( \beta \) estimation, (c) for \( \mu \) estimation.
(a) Bias of $\alpha$ estimators ($n = 20$).

(b) Bias of $\beta$ estimators ($n = 20$).
7.3 Estimation of Parameters

Figure 7.3.3: Graphs of bias of various estimators plotted against $\alpha$ when $n = 20$ and data come from $G(\alpha, 1, 0)$ distribution: (a) for $\alpha$ estimation, (b) for $\beta$ estimation, (c) for $\mu$ estimation.
(a) MSE of $\alpha$ estimators ($n = 20$).

(b) MSE of $\beta$ estimators ($n = 20$).
Figure 7.3.4: Graphs of MSE of various estimators plotted against $\alpha$ when $n = 20$ and data come from $G(\alpha, 1, 0)$ distribution: (a) for $\alpha$ estimation, (b) for $\beta$ estimation, (c) for $\mu$ estimation.

**Remark 7.3.1** From our extensive simulation study it appears that the modified maximum likelihood (MML) estimation has the best overall performance. Hence the MML estimation method is suggested for three parameter gamma distribution.

### 7.4 Goodness of Fit Tests for Gamma Distribution

Dahiya and Gurland (1972) provided a general goodness of fit test for a large sample by making use of the central limit theorem. In the following, we provide a simplified version of the goodness of fit test applicable to a two-parameter gamma distribution (7.1.1).

To test whether a random sample $X_1, X_2, \ldots, X_n$ is following a $G(\alpha, \beta)$
distribution, we adopt the following notations:

\[
\begin{align*}
  k_1 &= \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X} \\
  k_2 &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \\
  k_3 &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^3 \\
  h_1 &= k_1 = \bar{X} \\
  h_2 &= \frac{k_2}{k_1} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n\bar{X}} \\
  h_3 &= \frac{k_3}{k_2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^3}{\sum_{i=1}^{n} (X_i - \bar{X})^2}
\end{align*}
\]  

(7.4.1)

Also, define

\[
\begin{align*}
  \hat{\theta}^* &= (k_1, k_2/k_1)' = (\alpha \beta, \beta)' \\
  \hat{\theta}^* &= \left( k_1, \frac{k_2}{k_1} \right)' = \left( \bar{X}, \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 / \bar{X} \right)'
\end{align*}
\]  

(7.4.2)

Note that

\[
\xi = W' \theta^*, \quad \text{where} \\
W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix}
\]  

(7.4.3)

Define the matrices \( J \) and \( G \) as

\[
J = 3 \begin{pmatrix} 1 & 0 & 0 \\ -\frac{(2\alpha+1)}{\alpha} & 1 & 0 \\ \frac{(3\alpha+2)}{\alpha^2} & -\frac{(3\alpha+2)}{\alpha} & 1 \end{pmatrix}
\]  

(7.4.4)

and

\[
G = ((g_{ij}))_{3\times3}
\]  

(7.4.5)

where

\[
\begin{align*}
  g_{11} &= \alpha \beta^2 \\
  g_{12} &= 2\alpha(\alpha + 1)\beta^3 \\
  g_{13} &= 3\alpha(\alpha + 1)(\alpha + 2)\beta^4
\end{align*}
\]
7.4 Goodness of Fit Tests for Gamma Distribution

\[ g_{21} = 2\alpha(\alpha + 1)\beta^3 \]
\[ g_{22} = 2\alpha(\alpha + 1)(2\alpha + 3)\beta^4 \]
\[ g_{23} = 6\alpha(\alpha + 1)(\alpha + 2)^2\beta^5 \]
\[ g_{31} = 3\alpha(\alpha + 1)(\alpha + 2)\beta^4 \]
\[ g_{32} = 6\alpha(\alpha + 1)(\alpha + 2)^2\beta^5 \]
\[ g_{33} = 3\alpha(\alpha + 1)(\alpha + 2)(3\alpha^2 + 15\alpha + 20)\beta^6 \]

The matrix \( \Sigma \) is defined as

\[ \Sigma = J G J' \] (7.4.6)

and then \( \Sigma \) is estimated by \( \hat{\Sigma} \) where

\[ \hat{\Sigma} = [\Sigma]_{\theta^* = \hat{\theta}^*} \] (7.4.7)

In other words, the matrix \( \Sigma \), whose elements are functions of \( \alpha \) and \( \beta \), is estimated by replacing \( \alpha \) and \( \beta \) by \( \hat{\alpha}_{ML} \) and \( \hat{\beta}_{ML} \) respectively. Now define the test statistic \( \hat{Q} \) as

\[ \hat{Q} = n h' \hat{A} h \] (7.4.8)

where

\[ \hat{A} = \hat{\Sigma}^{-1}(I - \hat{R}) \]
\[ \hat{R} = W'(W\hat{\Sigma}^{-1}W')^{-1}W\hat{\Sigma}^{-1} \]

and \( h = (h_1, h_2, h_3)' \) (as in (7.4.1))

**Theorem 7.4.1** The asymptotic null distribution\(^1\) of \( \hat{Q} \) is that of \( \chi^2_1 \).

Using the above result, for a large \( n \), we reject gamma distribution as a suitable one for the data if \( \hat{Q} \) exceeds the upper \( \alpha \)-cutoff point of \( \chi^2_1 \) distribution, i.e., reject gamma model if \( \hat{Q} > \chi^2_{1, \alpha} \); and retain the gamma model otherwise. [Here \( \alpha \) indicates the level of significance.]

As a demonstration of this goodness of fit test, we use the data set in Example 7.1.2 to justify the applicability of a Gamma distribution to model the data.

For simplicity, one can get \( \hat{\Sigma} \) by directly using \( \hat{J} \) and \( \hat{G} \) obtained from \( J \) and \( G \) respectively, by replacing \( \alpha \) and \( \beta \) by \( \hat{\alpha}_{ML} \) and \( \hat{\beta}_{ML} \) respectively.

\(^1\)The distribution of \( Q \) when \( X_1, X_2, \ldots, X_n \) are assumed to follow (7.1.1).
For the data set in Example 7.1.2, we have

\[ \hat{\alpha}_{ML} = 1.67 \quad \text{and} \quad \hat{\beta}_{ML} = 49.98 \]

\[ \hat{J} = \begin{bmatrix} 3 & 0 & 0 \\ -7.79641 & 0.0359425 & 0 \\ 18.03 & -0.251957 & 0.000719 \end{bmatrix} \]

\[ \hat{G} = \begin{bmatrix} 4171.66 & 1113390 & 306337000 \\ 1113390 & 352803000 & 112381000000 \\ 306337000 & 112381000000 & 40876100000000 \end{bmatrix} \]

Also,

\[ \hat{h} = \begin{bmatrix} 83.5172 \\ 57.9592 \\ 105.909 \end{bmatrix} \]

which gives

\[ \hat{Q} = 0.002692 \]

Since \( \chi^2_{1,0.05} \) (with \( \alpha = 0.05 \)) = \( \chi^2_{1,0.05} = 3.841 \) and \( \hat{Q} < \chi^2_{1,0.05} \), we accept the fact that a two-parameter gamma model is suitable for the data in Example 7.1.2.

**Remark 7.4.1** To verify the goodness of fit of a three parameter gamma model, first obtain the \( \hat{\mu}_{ML} \) (as given in the previous section). Transform the original data \( X_1, X_2, \ldots, X_n \) (iid \( G(\alpha, \beta, \mu) \)) into \( X_1^*, X_2^*, \ldots, X_n^* \) where \( X_i^* = X_i - \hat{\mu}_{ML} \). Carry out the goodness of fit test of a two parameter gamma model on the transformed data \( X_1^*, X_2^*, \ldots, X_n^* \). If a \( G(\alpha, \beta) \) is found suitable for the transformed data \( X_1^*, X_2^*, \ldots, X_n^* \), then it indicates that a \( G(\alpha, \beta, \mu) \) model would be reasonable for the original data \( X_1, X_2, \ldots, X_n \).

**Remark 7.4.2** It is always desirable to fit a simpler model (probability distribution) to the data so that the inferences become easier. In the case of a two-parameter gamma model one may wish to test \( H_0 : \alpha = 1 \) against \( H_1 : \alpha \neq 1 \). Note that \( \alpha = 1 \) gives the \( \text{Exp}(\beta) \) distribution. We apply the computational approach (discussed in Section 5.5) in testing the above \( H_0 \) for the data set in Example 7.1.2 as shown below. It should be noted that the same approach can be followed for testing \( H_0 : \alpha = \alpha_0 \) for any \( \alpha_0(>0) \) other than 1.
7.4 Goodness of Fit Tests for Gamma Distribution

Testing $H_0 : \alpha = 1$ vs. $H_1 : \alpha \neq 1$ for the data set in Example 7.1.2.

**Step-1**: Obtain $\hat{\alpha}_{ML} = 1.67$ and $\hat{\beta}_{ML} = 49.98$ (using (7.3.1)).

**Step-2**: (a) Set $\alpha = 1$ (the $H_0$ value of $\alpha$). Then find the restricted MLE of $\beta$ (with $H_0$ restriction on $\alpha$) as $\hat{\beta}_{RML} = 83.51724$. (Notice that when $\alpha = 1$, the $G(\alpha, \beta)$ model reduces to the $\text{Exp}(\beta)$ model for which the MLE of $\beta$ is $\bar{X}$.)

(b) Now generate $(X_1, \ldots, X_{29})$ iid from $G(1, 83.51724)$, i.e., $\text{Exp}(83.51724)$. For this data recalculate $\hat{\beta}_{ML}$ and $\hat{\alpha}_{ML}$ following (7.3.1). Repeat this $M = 10,000$ times and retain only the $\hat{\alpha}_{ML}$ values, denoted as $\hat{\alpha}_{01}, \hat{\alpha}_{02}, \ldots, \hat{\alpha}_{0M}$.

**Step-3**: Order the above and get $\hat{\alpha}_{0(1)} \leq \ldots \leq \hat{\alpha}_{0(M)}$; and using the 5% significance level get the cutoff points $\hat{\alpha}_L = \hat{\alpha}_{0(250)} = 0.6874$ and $\alpha_U = \hat{\alpha}_{0(9750)} = 1.7287$.

**Step-4**: Since $\hat{\alpha}_{ML} = 1.67$ falls between the two cut-off points given above, we accept the $H_0 : \alpha = 1$ at 5% level, i.e., the same data set can be modelled by a one-parameter exponential ($\text{Exp}(\beta)$) distribution.

**Remark 7.4.3** One can invert the acceptance region of the hypothesis testing procedure discussed above to find a confidence interval for $\alpha$ as shown below. A similar approach can be followed to find a confidence interval for $\beta$ also.

Finding a 95% confidence interval for $\alpha$ for the data set in Example 7.1.2.

**Step-1**: Consider $\alpha_0 = 0.5, 1.0, 1.5, \ldots, 5.0$.

**Step-2**: For each $\alpha_0$ above, pretend to test $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$.

Follow the steps after Remark 7.4.2 and obtain the cutoff points $\hat{\alpha}_L$ and $\hat{\alpha}_U$. Note that $\hat{\alpha}_L$ and $\hat{\alpha}_U$ are very much dependent on $\alpha_0$ (the value of $\alpha$ used in $H_0$).

**Step-3**: Plot $\hat{\alpha}_L$ and $\hat{\alpha}_U$ simultaneously against $\alpha_0$. 

© 2006 by Taylor & Francis Group, LLC
Chapter 7  Gamma Distribution

Figure 7.4.1: Plots of $\hat{\alpha}_L$ and $\hat{\alpha}_U$ against $\alpha_0$ for the data set in Example 7.1.2.

If necessary, one can approximate both $\hat{\alpha}_L$ and $\hat{\alpha}_U$ by suitable non-linear regression functions of $\alpha_0$ as $\hat{\alpha}_L \approx g_1(\alpha_0)$ and $\hat{\alpha}_U \approx g_2(\alpha_0)$.

**Step-4**: Draw the horizontal line at $\hat{\alpha}_U = \hat{\alpha}_L = \hat{\alpha}_{ML} = 1.67$. The values of $\hat{\alpha}_L$ and $\hat{\alpha}_U$ intersected by the horizontal line provide a 95% confidence interval for $\alpha$. If one uses $g_1$ and $g_2$ as defined in Step-3, then the bounds of the confidence interval are $g_1^{-1}(\hat{\alpha}_{ML}) \approx 0.88$ and $g_2^{-1}(\hat{\alpha}_{ML}) \approx 2.56$. 
Chapter 8

Weibull Distribution

8.1 Preliminaries

In Chapter 2, we have seen the basic properties of a Weibull distribution. A random variable \( X \) is said to have a (two-parameter) Weibull distribution, or Weibull(\( \alpha, \beta \)), with parameters \( \alpha \) and \( \beta \) provided the pdf of \( X \) is given as

\[
f(x|\alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0 \quad (8.1.1)
\]

When \( \beta = 1 \), the pdf (8.1.1) is reduced to

\[
f(x|\alpha, 1) = \alpha x^{\alpha-1} e^{-x^\alpha}, \quad x > 0, \quad \alpha > 0 \quad (8.1.2)
\]

and this is known as the Standard Weibull distribution.

The distribution is named after Waloddi Weibull, a Swedish physicist, who used it to represent the probability distribution of the breaking strength of materials. The use of the Weibull distribution in reliability and quality control work has been advocated by Kao (1958, 1959).

Note that if \( X \) follows (8.1.1), and one makes a transformation of

\[
Y = \left(\frac{X}{\beta}\right)^\alpha \quad (8.1.3)
\]

then the random variable \( Y \) follows a standard exponential distribution, i.e., \( Y \) has pdf \( e^{-y} \), \( y > 0 \). Conversely, if \( Y \) (in (8.1.3)) follows a standard exponential distribution, then \( X = \beta Y^{1/\alpha} \) has a Weibull(\( \alpha, \beta \)) distribution.
The following diagram provides graphs of Weibull($\alpha, \beta$) pdfs for various values of $\alpha$ and $\beta$.

![Figure 8.1.1: Two-parameter Weibull pdf's.](image)

From the above figure, it is clear that the Weibull($\alpha, \beta$) family includes pdfs from highly positively skewed to almost symmetric ones. When $\alpha = 1$, the distribution becomes exponential distribution with scale parameter $\beta$. Differentiating (8.1.1), one can get

$$f'(x | \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-2} e^{-\left(x/\beta\right)^\alpha} \left[(\alpha - 1) - \frac{\alpha}{\beta^\alpha} x^\alpha\right] \quad (8.1.4)$$

It is seen that for $0 < \alpha \leq 1$, the mode of the pdf is at zero, and $f(x | \alpha, \beta)$ is monotonically decreasing in $x > 0$. For $\alpha > 1$, the pdf has a unique mode at $x_0 = \beta \left(\frac{\alpha - 1}{\alpha}\right)^{1/\alpha}$ which tends to $\beta$ as $\alpha \to \infty$. 

© 2006 by Taylor & Francis Group, LLC
8.1 Preliminaries

The cumulative distribution function based on the pdf (8.1.1) is
\[
F(x) = P(X \leq x) = 1 - \exp\left\{-\left(\frac{x}{\beta}\right)^\alpha\right\}
\] (8.1.5)
and the hazard rate, \( r(x) \), the instantaneous failure rate (if the random variable \( X \) denotes the lifetime) is
\[
r(x) = \frac{f(x|\alpha, \beta)}{1 - F(x)} = \frac{\alpha}{\beta^\alpha} x^{\alpha-1}
\] (8.1.6)
This hazard rate can also be obtained by differentiating the hazard function \( R(x) \) defined as
\[
R(x) = -\ln(1 - F(x))
\] (8.1.7)
Note that, in the special case, where \( \alpha = 1 \), one obtains the exponential distribution where the hazard rate is a constant.

A further generalization of the model in (8.1.1) is possible by introducing a shift or location parameter which gives rise to a three-parameter Weibull distribution with pdf
\[
f(x|\alpha, \beta, \mu) = \frac{\alpha}{\beta^\alpha}(x - \mu)^{\alpha-1} \exp\left\{-((x-\mu)/\beta)^\alpha\right\}, \quad x > \mu, \ -\infty < \mu < \infty, \ \alpha > 0, \ \beta > 0 \] (8.1.8)
and its cumulative distribution function (cdf)
\[
F(x) = 1 - \exp\left\{-((x-\mu)/\beta)^\alpha\right\}, \quad x > \mu
\] (8.1.9)
Note that for any value of \( \beta > 0 \),
\[
F(\beta + \mu) = 1 - e^{-1} \approx 0.6321 = 63.21\%
\] (8.1.10)
It means that \((\beta + \mu)\) is approximately the 63rd percentile point of the Weibull distribution given in (8.1.8). The above three-parameter Weibull distribution will be denoted by Weibull(\(\alpha, \beta, \mu\)) or simply by W(\(\alpha, \beta, \mu\)) distribution.

If \( X \) follows a W(\(\alpha, \beta, \mu\)) distribution, then \( Y' = \left\{(X - \mu)/\beta\right\}^\alpha \) follows a standard exponential distribution (similar to the arguments for (8.1.3)).

In the following, we present some data sets where the Weibull distribution seems to provide a good fit.
Example 8.1.1: Lieblein and Zelen (1956) gave the results of the test of endurance, in millions of revolutions (m.o.r.), of 23 ball bearings with the following ordered observations:

17.88  28.92  33.00  41.52  42.12  
45.60  48.48  51.84  51.96  54.12  
55.56  67.80  68.63  68.64  68.88  
84.12  93.12  98.64  105.12  105.84  
127.92 128.04  173.40

The following table gives the frequency distribution of the above observations.

<table>
<thead>
<tr>
<th>Endurance (in m.o.r.)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.5 – 48.7</td>
<td>7</td>
<td>0.304 (30.4%)</td>
</tr>
<tr>
<td>48.7 – 79.9</td>
<td>8</td>
<td>0.348 (34.8%)</td>
</tr>
<tr>
<td>79.9 – 111.1</td>
<td>5</td>
<td>0.217 (21.7%)</td>
</tr>
<tr>
<td>111.1 – 142.3</td>
<td>2</td>
<td>0.087 (8.7%)</td>
</tr>
<tr>
<td>142.3 – 173.5</td>
<td>1</td>
<td>0.043 (4.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>1.000 (100%)</td>
</tr>
</tbody>
</table>

Table 8.1.1: Frequency Table of the Test of Endurance Data.

The histogram in Figure 8.1.2 is drawn with the above relative frequencies.

Example 8.1.2: The following data set is taken from Cox and Oakes (1984) and consists of ten values of the number of cycles to failure when springs are subjected to various stress levels. For the following data, the stress level is 950 N/mm², and the values, given in units of 1000 cycles, are:

225  171  198  189  189
135  162  135  117  162

Since the above data set is small (with $n = 10$), we can make the frequency table with three classes only as given in Table 8.1.2.
Figure 8.1.2: Histogram of the Test of Endurance Data and a matching $W(\alpha, \beta, \mu)$ pdf with $\alpha = 1.7$, $\beta = 70.0$, $\mu = 16.0$.

<table>
<thead>
<tr>
<th>Cycles to Failure (in '000 of Cycles)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>117 – 153</td>
<td>3</td>
<td>0.3 (30%)</td>
</tr>
<tr>
<td>153 – 189</td>
<td>5</td>
<td>0.5 (50%)</td>
</tr>
<tr>
<td>189 – 225</td>
<td>2</td>
<td>0.2 (20%)</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>1.0 (100%)</td>
</tr>
</tbody>
</table>

Table 8.1.2: Frequency Table for Springs’ Failure Data.

The histogram in Figure 8.1.3 is drawn with the frequency distribution in Table 8.1.2 which is then approximated by a three-parameter Weibull
pdf with $\alpha = 2.4, \beta = 78.0$, and $\mu = 99.0$.

Example 8.1.3: Suppose that the service life, in years, of a particular type of camera battery is a random variable whose probability distribution can be modeled by a two-parameter $W(\alpha, \beta)$ distribution with $\alpha = 0.5$ and $\beta = 2.0$.

(a) What is the probability that such a battery life will exceed 2 years?

(b) How long such a battery is expected to last?

Let the random variable $X$ represent the service life (in years) of a camera battery. It is given that $X$ follows a $W(\alpha, \beta)$-distribution.

(a) $P(\text{battery life will exceed 2 years}) = P(X > 2)$

\begin{align*}
&= \exp\left\{-\left(\frac{2}{\beta}\right)^{\alpha}\right\} \quad \text{(by (8.1.5))}
\end{align*}
8.2 Characterization of Weibull Distribution

\[ = \exp\left\{-(2/2) \right\}^{0.5} \] (since \( \alpha = 0.5 \) and \( \beta = 2.0 \))

\[ = 36.79\% \approx 37\% \]

(b) Expected lifetime of a battery \( = E[X] \)

\[ = \beta \Gamma(1 + 1/\alpha) \]

\[ = (2.0) \Gamma(1 + 1/0.5) \]

\[ = 2 \Gamma(3) = 4 \text{ years} \]

Apart from modeling lifetimes of components, Weibull distribution has applications in many other real life problems ranging from wind power generation to rainfall intensity. For such applications, one can see Carlin and Haslett (1982), Barros and Estevan (1983), Tuller and Brett (1984), Wilks (1989), and Johnson and Haskell (1983).

**Notational Clarification:** A two-parameter Weibull distribution of the form (8.1.1) will be called a \( W(\alpha, \beta) \)-distribution, whereas a three-parameter Weibull distribution of the form (8.1.8) will be called a \( W(\alpha, \beta, \mu) \)-distribution. Hence, \( W(\alpha, \beta, 0) \)-distribution is nothing but the \( W(\alpha, \beta) \)-distribution and vice-versa. Again, as mentioned in the previous chapter, here \( \alpha \) is primarily used to denote the shape parameter, and it should be clear from the context if it is used to indicate the level of significance in testing a hypothesis.

8.2 Characterization of Weibull Distribution

Characterizations of the Weibull distribution are best done through the hazard function and failure rate distribution as shown by Roy and Mukherjee (1986).

Recall that the hazard function \( R \) of a cdf \( F \) of a nonnegative random variable is defined as

\[ R(x) = -\ln(1 - F(x)), \quad x > 0 \quad (8.2.1) \]

Note that by differentiating \( R(x) \) w.r.t. \( x \), one gets the failure rate, i.e.,

\[ R'(x) = f(x)/(1 - F(x)) = r(x) \]

If \( F(x) = 1 - \exp\{-x/\beta\}^\alpha \), \( x > 0 \); i.e., the cdf of a Weibull distribution, then \( R(x) = (x/\beta)^\alpha \). The following two results characterize a two-parameter Weibull distribution through \( R \).
Chapter 8  Weibull Distribution

Theorem 8.2.1  The hazard function \( R \) of a random variable \( X \) satisfies the condition:

\[
R(xy) \cdot R(1) = R(x) \cdot R(y) \quad \text{for all } x > 0 \text{ and } y > 0 \tag{8.2.2}
\]

if and only if \( X \) follows a \( W(\alpha, \beta) \) distribution.

Another variant of the above result is given in the next theorem.

Theorem 8.2.2  A nonnegative random variable \( X \) follows a \( W(\alpha, \beta) \) distribution, if and only if its hazard function \( R \) satisfies the following conditions:

1. \( R(xy) \cdot R(1) = R(x) \cdot R(y) \), with \( R(1) > 0 \) for all \( x > 0 \) and at least one value of \( y > 0 \), \( y \neq 1 \); and
2. \( \frac{xR'(x)}{R(x)} \) is nondecreasing in \( x > 0 \).

One may be interested in characterizing a Weibull distribution through the hazard rate (or the failure rate) instead of the hazard function (see (8.1.6)).

If \( X \) is a random variable with cdf \( F \), then it is well-known that \( F(X) \) follows continuous uniform (CU[0, 1]) distribution. It is easy to see that the probability distribution of \( R(X) \), where \( R \) being the hazard function, is completely known as \( P(R(X) \leq t) = 1 - \exp(-t) \). Therefore, the probability distribution of \( R(X) \) cannot characterize that of \( X \). However, the failure rate of \( X \) may characterize the probability distribution of \( X \). For instance, \( X \) follows an exponential distribution if and only if the failure rate of \( X \) is a constant (i.e., \( R'(X) \) has a degenerate distribution). This exponential characterization can be extended for the Weibull distribution as follows.

Theorem 8.2.3  Let \( X \) be a random variable with failure rate \( r(\cdot) \) such that

1. \( r \) is strictly increasing with \( r(0) = 0 \); and
2. \( r(X) \) follows \( W(\alpha', \beta') \) with \( \alpha' > 1 \) if and only if \( X \) follows \( W(\alpha, \beta) \) where \( \alpha > 1 \), \( (1/\alpha) + (1/\alpha') = 1 \), and \( \beta' = (\alpha/\beta)^{1/(\alpha-1)} \).

Dubey (1966) has given a different type of characterization of the Weibull distribution in terms of the smallest order statistics.
8.3 Estimation of Parameters

**Theorem 8.2.4** Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables. Then $\min_{1 \leq i \leq n} (X_i)$ has a Weibull distribution if and only if the common distribution of the $X_i$'s is a Weibull distribution. Moreover, if $X_i$'s have pdf (8.1.8), then $\min_{1 \leq i \leq n} (X_i)$ has the same pdf with $\beta$ replaced by $\beta^* = \beta n^{-1/\alpha}$.

Further characterizations of the Weibull distribution in terms of entropy maximization and Fisher-information minimization are given by Roy and Mukherjee (1986), but these are omitted here because they involve generalized gamma distribution which is beyond the scope of this handbook.

### 8.3 Estimation of Parameters

First we consider estimation of parameters $\alpha$ and $\beta$ of a two-parameter Weibull($\alpha, \beta$) distribution. Given independent and identically distributed observations $X_1, X_2, \ldots, X_n$ having the common pdf (8.1.1), the maximum likelihood estimators $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ of $\alpha$ and $\beta$, respectively, satisfy the equations

\[
\frac{\hat{\beta}_{ML}}{\hat{\alpha}_{ML}} = \left( \frac{1}{n} \sum_{i=1}^{n} X_i^{\hat{\alpha}_{ML}} \right)^{1/\hat{\alpha}_{ML}},
\]

\[
\hat{\alpha}_{ML} = \left[ \left( \frac{\sum_{i=1}^{n} X_i^{\hat{\alpha}_{ML}} \ln X_i}{\sum_{i=1}^{n} X_i^{\hat{\alpha}_{ML}}} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} \ln X_i \right) \right]^{-1}.
\]

If there is an unknown shift parameter $\mu$ involved, i.e., if the random sample comes from a three-parameter Weibull distribution with pdf (8.1.8), then apart from $\alpha$ and $\beta$, $\mu$ has to be estimated from the data. In this case, in (8.3.1) one needs to replace $X_i$ by $(X_i - \hat{\mu}_{ML})$, and there is an additional third equation

\[
\left( \hat{\alpha}_{ML} - 1 \right) \sum_{i=1}^{n} \left( X_i - \hat{\mu}_{ML} \right)^{-1} = \frac{\hat{\alpha}_{ML}}{\hat{\beta}_{ML}} \sum_{i=1}^{n} \left( X_i - \hat{\mu}_{ML} \right)^{\hat{\alpha}_{ML} - 1}.
\]

If the value of $\hat{\mu}_{ML}$ (after simplifying (8.3.2) and (8.3.1) with $X_i$ replaced by $(X_i - \hat{\mu}_{ML})$) is smaller than $\min_{1 \leq i \leq n} (X_i)$, then it is the MLE of $\mu$. © 2006 by Taylor & Francis Group, LLC
Otherwise, the MLE of \( \mu \) is taken as \( \hat{\mu}_{ML} = \min_{1 \leq i \leq n} (X_i) = X_{(1)} \), and one needs to find \( \hat{\alpha}_{ML} \) and \( \hat{\beta}_{ML} \) by solving (8.3.2) and (8.3.1) with \( X_i \) replaced by \( (X_i - X_{(1)}) \).

The aforementioned MLEs are known to have the desirable properties of consistency and asymptotic normality under fairly general conditions. In particular, if \( \alpha > 2 \), the usual regularity conditions are satisfied for \( \hat{\alpha}_{ML} \) and \( \hat{\beta}_{ML} \) and some weaker ones for \( \hat{\mu}_{ML} \). Weibull parameter MLEs are ‘regular’ (i.e., have usual asymptotic distribution) if and only if at least one of the following conditions holds: (a) \( \alpha > 2 \), (b) the shift parameter \( \mu \) is known, and (c) the sample is censored from below. (See Harter (1971) for more on this topic.) When \( 0 < \alpha \leq 1 \), the estimate \( \hat{\mu} = X_{(1)} \) is a super-efficient estimator of \( \mu \) (efficient when \( \alpha = 1 \)), but no true MLE for the other two parameters exist. When \( 1 < \alpha < 2 \), MLEs exist but their asymptotic variance-covariance matrix is meaningless because the determinant of the information matrix becomes negative, thereby yielding negative values for asymptotic variances of the MLEs. When \( \alpha = 2 \) (Rayleigh distribution), the determinant of the information matrix vanishes.

Dubey (1965) has shown that for a two-parameter Weibull(\( \alpha, \beta \)) distribution, assuming that MLEs of \( \alpha \) and \( \beta \) are regular and the sample size \( n \) is sufficiently large,

\[
\lim_{n \to \infty} \text{Var} \left( \sqrt{n} \hat{\beta}_{ML} \right) = \left[ 1 + \left( \frac{\psi(2)}{\psi'(1)} \right)^2 \right] \frac{1}{\alpha^2} = 1.087 \frac{1}{\alpha^2}
\]

\[
\lim_{n \to \infty} \text{Var} \left( \sqrt{n} \hat{\alpha}_{ML} \right) = \frac{\alpha^2}{\psi'(1)} = 0.608\alpha^2 \tag{8.3.3}
\]

\[
\lim_{n \to \infty} \text{Corr} \left( \hat{\alpha}_{ML}, \hat{\beta}_{ML} \right) = \frac{\psi(2)}{\psi'(1) + (\psi(2))^2 \psi''(1)} = 0.313
\]

where \( \psi(u) = \frac{\partial}{\partial u} \ln \Gamma(u) \) is the di-gamma function.

For the two-parameter case, the method of moments estimators (MMEs) of \( \alpha \) and \( \beta \) are obtained by solving the equations:

\[
\bar{X} = \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \beta^2 \Gamma \left( 1 + \frac{2}{\alpha} \right) \tag{8.3.4}
\]

The values of \( \alpha \) and \( \beta \) thus obtained are denoted by \( \hat{\alpha}_{MM} \) and \( \hat{\beta}_{MM} \), respectively. Another way of getting MMEs is to compute \( \hat{\alpha}_{MM} \) by solving
the following equation first

\[
\left( \frac{\Gamma (1 + 1/\hat{\alpha}_{MM})}{\Gamma (1 + 2/\hat{\alpha}_{MM})} \right)^2 = \frac{n \bar{X}^2}{\sum_{i=1}^{n} X_i^2}
\]  
(8.3.5)

and then get \( \hat{\beta}_{MM} \) by

\[
\hat{\beta}_{MM} = \frac{\bar{X}}{\Gamma (1 + 1/\hat{\alpha}_{MM})}
\]  
(8.3.6)

In the presence of a shift parameter \( \mu \) (i.e., for a three-parameter
Weibull model), the parameter MMEs are found by solving the following
three equations for \( \alpha, \beta, \) and \( \mu, \) and denoted by \( \hat{\alpha}_{MM}, \hat{\beta}_{MM}, \) and \( \hat{\mu}_{MM}, \) respectively:

\[
\bar{X} = \mu + \beta \Gamma \left( 1 + \frac{1}{\alpha} \right)
\]

\[
\frac{1}{n} \sum_{i=1}^{n} X_i^2 = \mu^2 + 2\mu \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) + \beta^2 \Gamma \left( 1 + \frac{2}{\alpha} \right)
\]  
(8.3.7)

\[
\frac{1}{n} \sum_{i=1}^{n} X_i^3 = \mu^3 + 3\mu \beta^2 \Gamma \left( 1 + \frac{2}{\alpha} \right) + 3\mu^2 \beta \Gamma \left( 1 + \frac{1}{\alpha} \right) + \beta^3 \Gamma \left( 1 + \frac{3}{\alpha} \right)
\]

subject to the restrictions that \( \hat{\mu}_{MM} < X_{(1)}, \hat{\alpha}_{MM} > 0 \) and \( \hat{\beta}_{MM} > 0. \)

Although maximum likelihood estimators (MLEs) are more computationally tedious to obtain than the other convenient estimators, MLEs perform better than the other estimators as shown by Thoman, Bain, and Antle (1969).

Maximum likelihood estimators for Weibull parameters with censored and grouped data sets can also be obtained. Zanakis and Kyparisis (1986) have discussed the computational aspects of obtaining MLEs when one has only:

(i) \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(m)} \) (\( 1 \leq m \leq n \)), i.e., first \( m \), out of \( n \), ordered observations (i.e., right censoring)

(ii) \( X_{(r+1)} \leq X_{(r+2)} \leq \ldots \leq X_{(m)} \) (\( 1 \leq r \leq m \leq n \)), i.e., all but lower \( r \) and upper \( (n - m) \) ordered observations (i.e., left and right censored)

(iii) Left and right censored grouped data, i.e., the censored observations \( X_{(r+1)} \leq X_{(r+2)} \leq \ldots \leq X_{(m)} \) are not observed individually, but
are instead divided into \( k \) groups with the group boundaries given by
\[ d_0 = X_{(r+1)} < d_1 < \ldots < d_k = X_{(m)} \]

The proposed numerical methods tend to be either fast but not very reliable, or sufficiently reliable but too slow. This is expected in view of the intrinsic complexity involved in maximum likelihood estimation which is essentially a nonlinear optimization problem.

**Example 8.3.1:** To demonstrate how the above estimation techniques could be followed, we use the data set presented in Example 8.1.1 (by Lieblein and Zelen (1956)).

First we assume a two-parameter Weibull model for the data (i.e., Weibull(\( \alpha, \beta \)), \( \mu \) is assumed to be 0). Using (8.3.1) and (8.3.5) to (8.3.6), our MLEs and MMEs are
\[
\hat{\alpha}_{ML} = 2.10204, \quad \hat{\beta}_{ML} = 81.8779 \\
\hat{\alpha}_{MM} = 2.01568, \quad \hat{\beta}_{MM} = 81.5069
\]

Now let us assume a three-parameter Weibull model (i.e., Weibull(\( \alpha, \beta, \mu \))) and using the system of equations discussed earlier, the MLEs and MMEs are
\[
\hat{\alpha}_{ML} = 1.5943, \quad \hat{\beta}_{ML} = 63.8793, \quad \hat{\mu}_{ML} = 14.8761 \\
\hat{\alpha}_{MM} = 2.0027, \quad \hat{\beta}_{MM} = 81.0261, \quad \hat{\mu}_{MM} = 0.4181
\]

**Remark 8.3.1** Maximum likelihood estimation of Weibull parameters is always a bit tricky. It is clear that if \( \alpha < 1 \) is allowed, then the likelihood function becomes unbounded. To avoid this problem, one usually restricts \( \alpha \) to be greater than or equal to 1. With \( \alpha > 1 \), the failure rate of a three-parameter Weibull distribution is
\[
r(x) = \frac{\alpha(x - \mu)^{\alpha-1}}{\beta^\alpha}, \quad x > \mu
\]
which is an increasing function. Increasing failure rates are associated with common materials that wear out with usage. On the other hand, the case \( \alpha < 1 \), which implies a decreasing failure rate, represents work-hardened materials.
8.3 Estimation of Parameters

**Theorem 8.3.1** (Johnson and Haskell (1983)) Assume that \((\alpha, \beta, \mu) \in \Theta\) where \(\Theta\) is given by

\[
\Theta = \{1 < \alpha < \infty, 0 < \beta < \infty, -\infty < \mu < \infty\}
\]

for suitable constants \(\overline{\alpha}, \overline{\beta}, \overline{\mu},\) and \(\overline{\mu}\. Then \((\hat{\alpha}_{ML}, \hat{\beta}_{ML}, \hat{\mu}_{ML})\) is a consistent estimator of \((\alpha, \beta, \mu)\); in other words, as \(n \to \infty\), \((\hat{\alpha}_{ML}, \hat{\beta}_{ML}, \hat{\mu}_{ML})\) converges to \((\alpha, \beta, \mu)\) in probability.

In many applied problems, as shown by Johnson and Haskell (1983), one is interested in a lower tolerance bound for a Weibull distribution. Let \(\xi_{c}\) denote the population \(100c^{th}\) percentile point \((0 < c < 1)\); i.e.,

\[
c = F(\xi_{c}) = 1 - \exp\left\{-\left(\frac{\xi_{c} - \mu}{\beta}\right)^{\alpha}\right\}
\]

and \(\xi_{c}\) can be obtained as

\[
\xi_{c} = \mu + \beta \left\{-\ln(1 - c)\right\}^{1/\alpha}
\]  

(8.3.10)

Therefore, the MLE of \(\xi_{c}\) is given by

\[
\hat{\xi}_{c(ML)} = \hat{\mu}_{ML} + \hat{\beta}_{ML} \left\{-\ln(1 - c)\right\}^{1/\hat{\alpha}_{ML}}
\]  

(8.3.11)

Due to the complexities associated with maximum likelihood estimation as discussed above, several modifications have been suggested to avoid computational problems. In the following, we provide one such technique based upon the profile likelihood (PL) suggested by Lockhart and Stephens (1994).

For a given random sample \(X_{1}, X_{2}, \ldots, X_{n}\) following a three-parameter Weibull distribution (8.1.8), the likelihood function is

\[
L(\alpha, \beta, \mu) = \prod_{i=1}^{n} \left[\frac{\alpha}{\beta} \left(\frac{X_{i} - \mu}{\beta}\right)^{\alpha-1} \exp\left\{-\left(\frac{X_{i} - \mu}{\beta}\right)^{\alpha}\right\}\right]
\]  

(8.3.12)

The profile likelihood \(L^{*}(\mu_{0})\) is defined for given \(\mu_{0}\) as

\[
L^{*}(\mu_{0}) = \sup_{\alpha, \beta} L(\alpha, \beta, \mu = \mu_{0})
\]  

(8.3.13)

A plot of \(\ln L^{*}(\mu_{0})\) against \(\mu_{0}\) usually takes one of the following three forms.
Case (i) There is a local minimum for \( \mu_0 \) close to \( X_{(1)} \) which gives a saddlepoint for the likelihood, and a local maximum for \( \mu_0 \) further from \( X_{(1)} \), giving the desired maximum likelihood solution for \( \mu \).

Case (ii) There is a local minimum for \( \mu_0 \), but no local maximum occurs.

Case (iii) The likelihood (or the log-likelihood) steadily decreases as \( \mu_0 \) moves away from \( X_{(1)} \) toward \(-\infty\).

By eliminating \( \beta \) from the equations in (8.3.1) (with \( X_i \) replaced by \( (X_i - \mu) \)) and (8.3.2), the two equations in \( \mu \) and \( \alpha \) can be written as

\[
\frac{1}{\alpha} - \frac{\left\{ \sum_{i=1}^{n} (X_i - \mu)^\alpha \ln(X_i - \mu) \right\}}{\left\{ \sum_{i=1}^{n} (X_i - \mu)^\alpha \right\}} + \frac{1}{n} \sum_{i=1}^{n} \ln(X_i - \mu) = 0 \tag{8.3.14}
\]

and

\[
\left( \frac{\alpha - 1}{\alpha} \right) \sum_{i=1}^{n} (X_i - \mu)^{-1} - \frac{\left\{ \sum_{i=1}^{n} (X_i - \mu)^{\alpha - 1} \right\}}{\left\{ \sum_{i=1}^{n} (X_i - \mu)^\alpha \right\}} = 0 \tag{8.3.15}
\]

The above two equations produce \( \hat{\alpha}_{PL} \) and \( \hat{\mu}_{PL} \) which are then used to get \( \hat{\beta}_{PL} \) by the equation

\[
\hat{\beta}_{PL} = \left\{ \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_{PL})^{\hat{\alpha}_{PL}} \right\}^{1/\hat{\alpha}_{PL}} \tag{8.3.16}
\]

To get the solutions \( \hat{\alpha}_{PL} \) and \( \hat{\mu}_{PL} \), we now fix \( \mu = \mu_0 \) and plot the solutions \( \alpha^* \) of Equation (8.3.14) and \( \alpha_* \) of Equation (8.3.15) against \( \mu_0 \). It is seen that as \( \mu_0 \to X_{(1)} \) from below, \( \alpha^* \to 0 \) and \( \alpha_* \to 1 \). If the graphs of \( \alpha^* \) and \( \alpha_* \) cross, then there is either the minimum or the maximum in case (i), or the minimum in case (ii); and if the graphs do not cross then it is case (iii) (which is observed to happen rarely).

Given a data set, it will be helpful to decide whether the situation is case (iii) without trying to solve the likelihood equations. For case (iii), as \( \mu_0 \to -\infty \), the plots of \( \alpha^* \) and \( \alpha_* \) have parallel asymptotes. Let the limiting difference between \( \alpha_* \) and \( \alpha^* \) as

\[
\Delta = \lim_{\mu_0 \to -\infty} (\alpha_* - \alpha^*) \tag{8.3.17}
\]
To find the value of $\Delta$, define $X = \sum_{i=1}^{n} X_i / n$, $S_0 = \sum_{i=1}^{n} X_i^2 / n$ and $T_r = \sum_{i=1}^{n} (X_i)^r \exp\left(-\gamma X_i\right)$, $r = 0, 1, 2$; where the quantity $\gamma$ is the solution of $\gamma(X - T_1 / T_0) = 1$. The value of $\gamma$ can easily be found by iteration, starting with $\gamma = -1$ in the expressions of $T_0$ and $T_1$. Define $D = XT_0 + \gamma(T_2 - XT_1); \gamma$ is the limiting slope of the lines, and $\Delta$ is given by

$$\Delta = \left\{ \frac{X T_0 - \gamma(S_0 T_0 - T_2)/2}{D} \right\}$$

(8.3.18)

A negative value of $\Delta$ means that we have case (iii), in which case a three parameter Weibull model with $\mu$ unknown looks questionable.

For case (i) and case (ii) it is important to establish whether or not a saddlepoint exists; and if it does then we have case (i). The saddlepoint is usually very close to $X(1)$. Thus it can be detected by starting with $\mu_0 = X(1) - \epsilon$, where $\epsilon$ is sufficiently small so that $\alpha_\ast - \alpha^\ast$ is positive, and then decreasing $\mu_0$ in very small steps until the saddlepoint is passed (i.e., $(\alpha_\ast - \alpha^\ast)$ becomes negative). The steps in $\mu_0$ can then be increased until once again $(\alpha_\ast - \alpha^\ast) = 0$, and at that point the values of $\mu_0$ and $\alpha$ become the PL solutions $\hat{\mu}_{PL}$ and $\hat{\alpha}_{PL}$ respectively. If no saddlepoint exists then we are in case (ii), for which $(\alpha_\ast - \alpha^\ast)$ will decrease to a minimum positive value and then start to increase again. In this case the MLE of $\mu$ is $X(1)$ which is biased. The bias in the estimate $\hat{\mu}_{PL} = X(1)$ is approximately $\beta n^{-1/\alpha}$. Since $\alpha$ and $\beta$ are unknown, they must be estimated and then used to reduce the bias in the estimate $\hat{\mu}_{PL} = X(1)$. The whole process of estimating $\alpha, \beta,$ and $\mu$ now goes through an iteration process. The likelihood equation $\frac{\partial (\ln L)}{\partial \alpha} = 0$ is needed, i.e.,

$$\frac{n}{\alpha} + \sum_{i=1}^{n} \ln(X_i - \mu) - \beta^{-\alpha} \sum_{i=1}^{n} (X_i - \mu)^\alpha \ln(X_i - \mu) - \left\{ n - \beta^{-\alpha} \sum_{i=1}^{n} (X_i - \mu)^\alpha \right\} \ln \beta = 0$$

(8.3.19)

Suppose that $\hat{\mu}(l), \hat{\beta}(l)$, and $\hat{\alpha}(l)$ are the estimates at the $l^{th}$ iteration. Then the improved estimates $\hat{\mu}(l+1), \hat{\beta}(l+1), \hat{\alpha}(l+1)$ are obtained as follows.

**Step 1**: $\hat{\mu}(l+1) = X(1) - \hat{\beta}(l)n^{-1/\hat{\alpha}(l)}$

**Step 2**: Solve the Equation (8.3.19) for $\hat{\alpha}(l+1)$ using $\mu = \hat{\mu}(l+1)$ and $\beta = \hat{\beta}(l)$
Step 3: Use Equation (8.3.16) to get $\hat{\beta}_{(l+1)}$ using $\hat{\mu}_{PL} = \hat{\mu}_{(l+1)}$ and $\hat{\alpha}_{PL} = \hat{\alpha}_{(l+1)}$.

Iteration of these three steps continues until the required accuracy for $\hat{\alpha}_{PL}$ is obtained. Initial estimates $\hat{\alpha}_0$ and $\hat{\beta}_0$ may be found by setting $\hat{\mu}_0 = X_{(1)}$ and continuing with Steps (2) and (3) given above, but using only the $(n - 1)$ sample values $X_{(2)}, X_{(3)}, \ldots, X_{(n)}$ (since otherwise the term $\ln(X_{(1)} - \mu)$ with $\mu = \hat{\mu}_0 = X_{(1)}$ in (8.3.19) causes problem). When the iteration stops, we get the estimates $\hat{\mu}_{PL}$, $\hat{\beta}_{PL}$, and $\hat{\alpha}_{PL}$ for case (ii).

Example 8.3.2: Here we need to estimate the parameters of a three-parameter Weibull distribution for the data sets in Example 8.1.1 and Example 8.1.2. The parameter estimates are given below.

### Example 8.1.1

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usual MLE</td>
<td>14.876</td>
<td>1.5943</td>
<td>63.8793</td>
</tr>
<tr>
<td>Lockhart and Stephens’ PLE</td>
<td>9.1717</td>
<td>1.5759</td>
<td>63.6851</td>
</tr>
</tbody>
</table>

### Example 8.1.2

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usual MLE</td>
<td>99.0109</td>
<td>2.3755</td>
<td>78.2417</td>
</tr>
<tr>
<td>Lockhart and Stephens’ PLE</td>
<td>100.161</td>
<td>1.5648</td>
<td>73.3424</td>
</tr>
</tbody>
</table>

8.4 Goodness of Fit Tests for Weibull Distribution

In this section we discuss the goodness of fit tests for two-parameter and three-parameter Weibull distributions as suggested by Coles (1989) and Lockhart and Stephens (1994).
First we discuss the goodness of fit test for a two parameter Weibull distribution with pdf (8.1.1) (i.e., with \( cdf \) (8.1.5)). Coles’ (1989) goodness-of-fit-test is based on Michael’s (1983) stabilized probability plot and Blom’s (1958) procedure of parameter estimation.

To determine whether an ordered sample \( Y_1 \leq \ldots \leq Y_n \) has been drawn from a population with \( cdf G_Y((y-\delta)/\eta) \), of location-scale family, the plotting coordinates \((d_i,e_i)\), \( 1 \leq i \leq n \), of the stabilized probability plot are calculated as

\[
\begin{align*}
    d_i &= (2/\pi) \sin^{-1} \left\{ \sqrt{(i-0.5)/n} \right\} \quad \text{and} \\
    e_i &= (2/\pi) \sin^{-1} \left\{ \sqrt{G_Y((Y_i)-\delta)/\eta} \right\}
\end{align*}
\]

(8.4.1)

Note that, if the ordered sample \( Y_1 \leq \ldots \leq Y_n \) really comes from the distribution with \( cdf G_Y((y-\delta)/\eta) \) then:

**Case** (i) The coordinates \((d_i,e_i)\), \( 1 \leq i \leq n \), are expected to be close to the line joining the points \((0,0)\) and \((1,1)\)

**Case** (ii) As shown by Michael (1983), the plotted points \((d_i,e_i)\) have approximately equal variances due to the asymptotic distribution of \( e_i \)’s

The two-parameter Weibull distribution with \( cdf \) (8.1.5) is not a location-scale type, however, a log transformation yields a variable with a location-scale type \( cdf \). In other words, if \( X_1, \ldots, X_n \) are iid \( W(\alpha, \beta) \), if and only if \( Y_1 = \ln X_1, \ldots, Y_n = \ln X_n \) are iid with \( cdf \)

\[
G_Y(y) = G_Y((y-\delta)/\eta) = 1 - \exp \left[ -\exp \left\{ (y-\delta)/\eta \right\} \right]
\]

(8.4.2)

where \( \delta = \ln \beta \) and \( \eta = 1/\alpha \). Thus, instead of checking whether an ordered sample \( X_1 \leq \ldots \leq X_n \) has been drawn from a \( W(\alpha, \beta) \) distribution with unknown parameters, one can test whether \( Y_1 \leq \ldots \leq Y_n \) has been drawn from the distribution with \( cdf \) (8.4.2).

The first test statistic to test the claim that the \( cdf \) (8.4.2) is the true distribution is

\[
\Delta^* = \max_{1 \leq i \leq n} |d_i - \hat{e}_i|, \quad \text{where} \quad \hat{e}_i = e_i \quad \text{with } \delta \text{ and } \eta \text{ replaced by } \hat{\delta} \text{ and } \hat{\eta}
\]

(8.4.4)

The estimates \( \hat{\delta} \) and \( \hat{\eta} \) in (8.4.4) are due to Blom (1958) and obtained through the steps as given below.
Chapter 8  Weibull Distribution

Step 1: Calculate

\[ u_i = - \left( \frac{n - i + 1}{n + 1} \right) \ln \left( \frac{n - i + 1}{n + 1} \right) ; \]

\[ 1 \leq i \leq n - 1, \quad 0 = u_o = u_n \]

and

\[ \nu_i = u_i \frac{n!}{(i - 1)!(n - i)!} \int_0^1 \ln(-\ln(1 - t)) t^{i-1}(1 - t)^{n-i} \, dt \]

\[ 1 \leq i \leq n \]

Step 2: Calculate

\[ c_{1i} = u_i - u_{i+1} \quad \text{and} \quad c_{2i} = \nu_i - \nu_{i+1}, \quad 0 \leq i \leq n - 1 \]

Step 3: Invert the matrix

\[ C = \begin{bmatrix} \sum_{i=1}^n c_{1i}^2 & \sum_{i=1}^n c_{1i}c_{2i} \\
\sum_{i=1}^n c_{1i}c_{2i} & \sum_{i=1}^n c_{2i}^2 \end{bmatrix} \]

and get

\[ C^{-1} = \begin{bmatrix} c_{11} & c_{12} \\
c_{21} & c_{22} \end{bmatrix} \]

Step 4: Finally, the unbiased nearly best linear estimates of \( \delta \) and \( \eta \) are obtained as

\[ \hat{\delta} = \sum_{i=1}^n d_{1i} Y(i) \quad \text{and} \quad \hat{\eta} = \sum_{i=1}^n d_{2i} Y(i) \]

where

\[ d_{ri} = u_i \left\{ c^{r1}(c_{1i} - c_{1(i-1)}) + c^{r2}(c_{2i} - c_{2(i-1)}) \right\}, \quad r = 1, 2 \]

Coles (1989) provided simulated critical values for \( \Delta^* \) (in (8.4.3)) based on 6,000 replications for selected values of \( n \) and significance level. Note that large values of \( \Delta^* \) indicate lack-of-fit. In the following we have provided refined simulated critical values based on 10,000 replications.

The second test statistic to test the claim that the cdf (8.4.2) is the true distribution is

\[ \Delta^{**} = \text{correlation coefficient of} \ (d_i, \hat{e}_i), \ 1 \leq i \leq n \quad (8.4.5) \]
8.4 Goodness of Fit Tests for Weibull Distribution

<table>
<thead>
<tr>
<th>$n$</th>
<th>Level of significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.2048</td>
</tr>
<tr>
<td>10</td>
<td>0.1889</td>
</tr>
<tr>
<td>15</td>
<td>0.1446</td>
</tr>
<tr>
<td>20</td>
<td>0.1440</td>
</tr>
<tr>
<td>25</td>
<td>0.1431</td>
</tr>
<tr>
<td>30</td>
<td>0.1319</td>
</tr>
<tr>
<td>35</td>
<td>0.1250</td>
</tr>
<tr>
<td>40</td>
<td>0.1183</td>
</tr>
<tr>
<td>45</td>
<td>0.1150</td>
</tr>
<tr>
<td>50</td>
<td>0.1099</td>
</tr>
<tr>
<td>75</td>
<td>0.0966</td>
</tr>
<tr>
<td>100</td>
<td>0.0861</td>
</tr>
<tr>
<td>300</td>
<td>0.0554</td>
</tr>
<tr>
<td>500</td>
<td>0.0438</td>
</tr>
</tbody>
</table>

Table 8.4.1: Simulated Critical Values for $\Delta^*$ with 10,000 Replications.

Since stabilized probability plots from distributions other than (8.4.2) tend to be nonlinear, the statistic $\Delta^{**}$ is a reasonable measure of goodness of fit. Hence small values of $\Delta^{**}$, as opposed to large values of $\Delta^*$, are significant, i.e., if $\Delta^{**}$ value is smaller than the one listed in Table 8.4.2, then it indicates a lack-of-fit.

Instructions to calculate simulated critical values of the test statistics $\Delta^*$ and $\Delta^{**}$.

**Step-1**: (a) Fix the value of $n$, say, $n = 5$.

(b) Calculate $u_i$, $\nu_i$, $C^{-1}$ etc. i.e., Steps 1 through 3 after (8.4.4) (since they do not depend on the observations).

(c) Generate observations $Y_1^{(1)}, \ldots, Y_n^{(1)}$ from (8.4.2) with $\delta = 0$ and $\eta = 1$; or generate observations $X_1^{(1)}, \ldots, X_n^{(1)}$ from (8.1.1) with $\alpha = \beta = 1$ and then transform to $Y_1^{(1)} = \ln X_1^{(1)}, \ldots, Y_n^{(1)} = \ln X_n^{(1)}$. The superscript $(1)$ indicates the first replication.
Table 8.4.2: Simulated Critical Values for $\Delta^{**}$ with 10,000 Replications.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Level of significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.8428</td>
</tr>
<tr>
<td>10</td>
<td>0.9050</td>
</tr>
<tr>
<td>15</td>
<td>0.9330</td>
</tr>
<tr>
<td>20</td>
<td>0.9480</td>
</tr>
<tr>
<td>25</td>
<td>0.9584</td>
</tr>
<tr>
<td>30</td>
<td>0.9654</td>
</tr>
<tr>
<td>35</td>
<td>0.9705</td>
</tr>
<tr>
<td>40</td>
<td>0.9738</td>
</tr>
<tr>
<td>45</td>
<td>0.9771</td>
</tr>
<tr>
<td>50</td>
<td>0.9790</td>
</tr>
<tr>
<td>75</td>
<td>0.9864</td>
</tr>
<tr>
<td>100</td>
<td>0.9898</td>
</tr>
<tr>
<td>300</td>
<td>0.9966</td>
</tr>
<tr>
<td>500</td>
<td>0.9978</td>
</tr>
</tbody>
</table>

(d) Calculate $\hat{\delta}(1) = \sum_{i=1}^{n} d_1 Y^{(1)}(i)$, $\hat{\eta}(1) = \sum_{i=1}^{n} d_2 Y^{(1)}(i)$.

(e) Calculate $d_i$ and $\tilde{e}_i(1) = (\frac{2}{\pi}) \sin^{-1} \left\{ \sqrt{G_Y((Y^{(1)}(i) - \hat{\delta}(1))/\hat{\eta}(1))} \right\}$ where $G_Y((y - \delta)/\eta)$ expression is given in (8.4.2). $1 \leq i \leq n$.

(f) Calculate

$$\Delta^{*(1)} = \max_{1 \leq i \leq n} \left| d_i - \tilde{e}_i(1) \right|$$

$$\Delta^{**(1)} = \text{correlation coefficient of } (d_i, \tilde{e}_i(1)), \; 1 \leq i \leq n$$

These are the test statistic values from the first replication.

**Step-2**: Repeat the above steps (1)(c) to (1)(f) for a large number of replications, say, $M$. Thus we get

$$\Delta^{*(1)}, \Delta^{*(2)}, \ldots, \Delta^{*(M)}$$

and

$$\Delta^{**(1)}, \Delta^{**(2)}, \ldots, \Delta^{**(M)}$$
8.4 Goodness of Fit Tests for Weibull Distribution

Step-3: Order the values $\Delta^*(1), \Delta^*(2), \ldots, \Delta^*(M)$ (smallest to largest) and with $M = 10,000$ get $\Delta^*$ value with rank $(10,000)(1 - \alpha)$. This is the $\alpha$-level critical value for $\Delta^*$. If $\alpha = 0.05$, then get the value of $\Delta^*$ (from the ordered replicated values) that has rank 9500. Similarly, for $\Delta^{**}$, get the value with rank $(10,000)(\alpha)$, i.e., 500.

Example 8.4.1: To demonstrate the usage of the goodness of fit test procedures for a two-parameter Weibull distribution, we consider the data set (with $n = 23$) presented in Example 8.1.1. We obtain

$$\Delta^* = 0.08562 \quad \text{and} \quad \Delta^{**} = 0.9860$$

Note that the critical values for $\Delta^*$ (Table 8.4.1) are monotonically decreasing with $n$. On the other hand, the critical values for $\Delta^{**}$ (Table 8.4.2) are monotonically increasing with $n$. For $n = 23$, $\Delta^* = 0.08562 < 0.1190 = \text{critical value for } (n = 25) < \text{critical value for } (n = 23)$. Thus Weibull seems to be a plausible model. The same conclusion is arrived at using the statistic $\Delta^{**}$ with similar argument.

In the following we now discuss about the goodness of fit test for a three-parameter Weibull distribution based on the empirical distribution function.

Let $X_{(1)} \leq \ldots \leq X_{(n)}$ be the order statistics of the $iid$ observation $X_1, \ldots, X_n$ which is thought to be coming from the distribution with pdf (8.1.8). The following steps are to be followed.

Step 1: First estimate the unknown parameters $\alpha, \beta, \mu$ by $\hat{\alpha}, \hat{\beta}, \hat{\mu}$ as described in Section 8.3 based on the profile likelihood approach of Lockhart and Stephens (1994).

Step 2: Make the transformation

$$Z_{(i)} = F(X_{(i)} \mid \hat{\alpha}, \hat{\beta}, \hat{\mu}) = 1 - \exp \left[ - \left\{ \frac{X_{(i)} - \hat{\mu}}{\hat{\beta}} \right\}^{\hat{\alpha}} \right], \, i = 1, 2, \ldots, n$$

Step 3: The three test statistics based on the empirical distribution are calculated as

$$\Delta_1 = \sum_{i=1}^{n} \left\{ Z_{(i)} - \left( \frac{2i - 1}{2n} \right) \right\}^2 + \frac{1}{12n}$$
\[
\Delta_2 = \Delta_1 - \eta(Z - 0.5)^2, \quad Z = \frac{1}{n} \sum_{i=1}^{n} Z_{(i)}/n
\]
\[
\Delta_3 = -n - \frac{1}{\eta} \sum_{i=1}^{n} (2i - 1) \left[ \ln Z_{(i)} + \ln \left\{ 1 - Z_{(n-i+1)} \right\} \right]
\]

**Step 4:** Let \( \hat{\eta} = 1/\hat{\alpha} \). The following table gives the critical values for the above three test statistics at different significance levels. The null hypothesis that the data \( X_1, \ldots, X_n \) comes from the model (8.1.8) is rejected if the test statistic used is greater than the critical value for a specific level. When \( 0 < \hat{\eta} < 0.5 \), the critical point is obtained from the table corresponding to \( \hat{\eta} \). But for \( \hat{\eta} \geq 0.5 \), the last line labelled as \( \hat{\eta} = 0.5 \) should be used. Also note that the table does not use the sample size \( n \). For a value of \( \hat{\eta} \), \( 0 < \hat{\eta} < 0.5 \), which is not listed in the table, linear interpolation can be done to get approximate critical values.

<table>
<thead>
<tr>
<th>( \hat{\eta} )</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \Delta_1 )</td>
<td>( \Delta_2 )</td>
<td>( \Delta_3 )</td>
</tr>
<tr>
<td>0.00</td>
<td>0.144</td>
<td>0.143</td>
<td>0.836</td>
</tr>
<tr>
<td>0.05</td>
<td>0.145</td>
<td>0.144</td>
<td>0.845</td>
</tr>
<tr>
<td>0.10</td>
<td>0.147</td>
<td>0.146</td>
<td>0.856</td>
</tr>
<tr>
<td>0.15</td>
<td>0.149</td>
<td>0.148</td>
<td>0.869</td>
</tr>
<tr>
<td>0.20</td>
<td>0.152</td>
<td>0.150</td>
<td>0.885</td>
</tr>
<tr>
<td>0.25</td>
<td>0.154</td>
<td>0.152</td>
<td>0.902</td>
</tr>
<tr>
<td>0.30</td>
<td>0.157</td>
<td>0.154</td>
<td>0.923</td>
</tr>
<tr>
<td>0.35</td>
<td>0.161</td>
<td>0.157</td>
<td>0.947</td>
</tr>
<tr>
<td>0.40</td>
<td>0.165</td>
<td>0.159</td>
<td>0.974</td>
</tr>
<tr>
<td>0.45</td>
<td>0.170</td>
<td>0.162</td>
<td>1.006</td>
</tr>
<tr>
<td>0.50</td>
<td>0.175</td>
<td>0.166</td>
<td>1.043</td>
</tr>
</tbody>
</table>

Table 8.4.3: Critical Values for the Test Statistics \( \Delta_1, \Delta_2, \Delta_3 \).

**Example 8.4.2:** To demonstrate the usage of the above goodness of fit test procedures for a three-parameter Weibull distribution we consider the data set (with \( n = 10 \)) presented in Example 8.1.2. For the data set
we obtain
\[ \Delta_1 = 0.0842, \quad \Delta_2 = 0.0827, \quad \text{and} \quad \Delta_3 = 0.4664 \]

With \( \hat{\eta} = 0.6391 > 0.5 \) we should compare the above \( \Delta \)-values with the critical values corresponding to \( \hat{\eta} = 0.5 \) in Table 8.4.3.

Note that the values of \( \Delta_1, \Delta_2, \) and \( \Delta_3 \) are smaller than the respective critical values available from the Table 8.4.3. Hence a three-parameter Weibull model seems plausible for the given data set.

**Remark 8.4.1** As discussed in Remark 7.4.2, here also we seek to find out whether a simpler Weibull model can be appropriate for a given data set. Once a two-parameter \( W(\alpha, \beta) \) is found suitable, one can further test whether \( \alpha = 1 \) or not. Notice that \( \alpha = 1 \) reduces \( W(\alpha, \beta) \) to \( \text{Exp}(\beta) \) which is much easier to work with. In the following we demonstrate the usage of our computational approach to test \( H_0 : \alpha = 1 \) vs. \( H_1 : \alpha \neq 1 \).

**Testing \( H_0 : \alpha = 1 \) vs. \( H_1 : \alpha \neq 1 \) for the data set in Example 8.1.1.**

**Step-1**: Obtain \( \hat{\alpha}_{\text{ML}} = 2.10204 \) and \( \hat{\beta}_{\text{ML}} = 81.8779 \) (using (8.3.1)).

**Step-2**: (a) Set \( \alpha = 1 \) (the \( H_0 \) value of \( \alpha \)). Then find the restricted MLE of \( \beta \) (with the \( H_0 \) restriction on \( \alpha \)) as \( \hat{\beta}_{\text{RML}} = 72.22391 \).

(b) Now generate \((X_1, X_2, \ldots, X_{23})\) iid from \( W(1, 72.22391) \), i.e., \( \text{Exp}(72.22391) \). For this data recalculate \( \hat{\beta}_{\text{ML}} \) and \( \hat{\alpha}_{\text{ML}} \) following (8.3.1). Repeat this \( M = 10,000 \) times and retain only the \( \hat{\alpha}_{\text{ML}} \) values, denoted as \( \hat{\alpha}_0(1), \hat{\alpha}_0(2), \ldots, \hat{\alpha}_0(M) \).

**Step-3**: Order the above and get \( \hat{\alpha}_0(1), \hat{\alpha}_0(2), \ldots, \hat{\alpha}_0(M) \); and using the 5% significance level get the cutoff points \( \hat{\alpha}_L = \hat{\alpha}_0(250) = 0.7628 \) and \( \hat{\alpha}_U = \hat{\alpha}_0(9750) = 1.4847 \).

**Step-4**: Since \( \hat{\alpha}_{\text{ML}} \) exceeds the above upper bound, we reject \( H_0 : \alpha = 1 \) at 5% level, and accept \( H_1 : \alpha \neq 1 \), i.e., the Weibull model cannot be reduced to an exponential one.
Remark 8.4.2 Similar to what we did in the earlier chapter we can now invert the acceptance region for testing a hypothesis on \( \alpha \) to find a confidence interval for \( \alpha \). A similar approach can be followed to test a hypothesis on \( \beta \), and then invert the acceptance region to find a suitable confidence interval for \( \beta \).

Finding a 95% confidence interval for \( \alpha \) for the data set in Example 8.1.1.

Step-1: Consider \( \alpha_0 = 1.0, 1.5, 2.0, \ldots, 5.0 \).

Step-2: For each \( \alpha_0 \) above, pretend to test \( H_0 : \alpha = \alpha_0 \) vs. \( H_1 : \alpha \neq \alpha_0 \). Follow the steps after Remark 8.4.1 and obtain the cutoff points \( \hat{\alpha}_L \) and \( \hat{\alpha}_U \) which are functions of \( \alpha_0 \) (the value of \( \alpha \) under \( H_0 \)).

Step-3: Plot \( \hat{\alpha}_L \) and \( \hat{\alpha}_U \) simultaneously against \( \alpha_0 \). If necessary, approximate both \( \hat{\alpha}_L \) and \( \hat{\alpha}_U \) as functions of \( \alpha_0 \) by nonlinear regression functions as \( \hat{\alpha}_L \approx g_1(\alpha_0) \) and \( \hat{\alpha}_U \approx g_2(\alpha_0) \). See Figure 8.4.1 below.

Step-4: The curves of \( \hat{\alpha}_L \) and \( \hat{\alpha}_U \) (as functions of \( \alpha_0 \)) intersected by the horizontal line \( \hat{\alpha}_U = \hat{\alpha}_L = \hat{\alpha}_{ML} \) provide a 95% confidence interval for \( \alpha \). In other words, a 95% two sided confidence interval for \( \alpha \) is \((g_1^{-1}(\hat{\alpha}_{ML}), g_2^{-1}(\hat{\alpha}_{ML})) \approx (1.37, 2.86)\).
Figure 8.4.1: Plots of $\hat{\alpha}_L$ and $\hat{\alpha}_U$ against $\alpha_0$ for the data set in Example 8.1.1.
Chapter 9

Extreme Value Distributions

9.1 Preliminaries

A random variable $X$ is said to have a Type-I Extreme Value (EV) distribution with parameters $\mu$ and $\beta$ provided the pdf of $X$ is given as

$$f_1(x \mid \mu, \beta) = \frac{1}{\beta} e^{-(x-\mu)/\beta} \exp \left\{ -e^{-(x-\mu)/\beta} \right\}, \ x \in \mathbb{R}, \ \beta > 0, \ \mu \in \mathbb{R} \quad (9.1.1)$$

The above positively skewed pdf with range over the whole real line has the cdf

$$F_1(x) = P(X \leq x) = \exp \left\{ -e^{-(x-\mu)/\beta} \right\} \quad (9.1.2)$$

and henceforth the distribution will be called as EV-I($\beta, \mu$) distribution. When $\beta = 1$ and $\mu = 0$, the above pdf (9.1.1) reduces to

$$f_1(x \mid 0, 1) = e^{-x} \exp \left\{ -e^{-x} \right\} \quad (9.1.3)$$

and this is known as EV-I(1,0) or standard Type I Extreme Value Distribution.
Chapter 9  Extreme Value Distributions

Figure 9.1.1: Type-I EV distribution pdfs for various $\beta$ and $\mu$.

The above Type-I Extreme Value Distribution is also known as **double exponential** distribution due to the structure of the distribution function (9.1.2) (presence of two ‘e’ s).

There are two other types of extreme value distributions widely used by the researchers. These are Type-II and Type-III extreme value distributions with three parameters given as:

(a) Type-II extreme value distribution (EV-II $(\alpha, \beta, \mu)$) with pdf

$$f_2(x|\alpha, \beta, \mu) = \left(\frac{\alpha}{\beta}\right) \left(\frac{x - \mu}{\beta}\right)^{-(\alpha+1)} \exp\left\{-\left(\frac{x - \mu}{\beta}\right)^{-\alpha}\right\},$$

$$x \geq \mu, \beta > 0, \mu \in \mathbb{R}, \alpha > 0 \quad (9.1.4)$$
9.1 Preliminaries

(b) Type-III Extreme Value Distribution (EV-III($\alpha, \beta, \mu$)) with pdf

\[ f_3(x|\alpha, \beta, \mu) = \left(\frac{\alpha}{\beta}\right) \left(\frac{\mu - x}{\beta}\right)^{(\alpha-1)} \exp\left\{ - \left(\frac{\mu - x}{\beta}\right)^{\alpha}\right\} \],

\[ x \leq \mu, \; \beta > 0, \; \mu \in \mathbb{R}, \; \alpha > 0 \] (9.1.5)

The cdfs corresponding to above Type-II and Type-III pdfs are respectively

\[ F_2(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < \mu \\ \exp\left\{ -\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right\} & \text{if } x \geq \mu \end{cases} \] (9.1.6)

![Figure 9.1.2: Type-II EV distribution pdfs for various $\alpha$, $\beta$, and $\mu$.](image)

and

\[ F_3(x) = P(X \leq x) = \begin{cases} \exp\left\{-\left(\frac{x-\mu}{\beta}\right)^\alpha\right\} & \text{if } x \leq \mu \\ 1 & \text{if } x > \mu \end{cases} \] (9.1.7)
Chapter 9  Extreme Value Distributions

The standard Type-II Extreme Value Distribution is obtained when $\alpha = 1$, $\beta = 1$, and $\mu = 0$ (i.e., EV-II $(1, 1, 0)$) with pdf

$$f_2(x|1, 1, 0) = x^{-2}e^{-1/x}, \quad x > 0$$

(9.1.8)

Similarly, the standard Type-III Extreme Value Distribution (i.e., EV-III $(1, 1, 0)$) has pdf

$$f_3(x|1, 1, 0) = e^x, \quad x \leq 0$$

(9.1.9)

The name ‘extreme value’ is attached to the above mentioned distributions because they can be obtained, under certain conditions, as limiting distributions (as $n \to \infty$) of the largest value of $n$ independent random variables each having the same continuous distribution.

Figure 9.1.3: Type-III EV distribution pdfs for various $\alpha$, $\beta$, and $\mu$.  

© 2006 by Taylor & Francis Group, LLC
The Type-I limiting distribution of the smallest value of $n$ iid random variables has the cdf $G_1(x)$ given as

$$G_1(x) = 1 - \exp\left\{-e^{(x-\mu)/\beta}\right\} \quad (9.1.10)$$

with pdf

$$g_1(x|\mu, \beta) = \frac{1}{\beta} e^{(x-\mu)/\beta} \exp\left\{-e^{(x-\mu)/\beta}\right\}, \quad x \in \mathbb{R}, \ \beta > 0, \ \mu \in \mathbb{R} \quad (9.1.11)$$

Note that, since $\min\{X^*_1, \ldots, X^*_n\} = -\max\{-X^*_1, \ldots, -X^*_n\}$, the above pdf (9.1.11) can be obtained from (9.1.1) by replacing $X$ by $(-X)$ and $\mu$ by $(-\mu)$. In a similar way one can easily derive Type-II and Type-III limiting distributions of the smallest value of iid observations from (9.1.4) and (9.1.5).

If $X^*_1, X^*_2, \ldots, X^*_n$ are independent random variables with common pdf $f^*(x)$, then the cumulative distribution function (cdf) of $X^*_n = \max\{X^*_1, \ldots, X^*_n\}$ is $F^*_n(x) = [F^*(x)]^n$, where $F^*(x)$ is the common cdf of $X^*_i$'s, $1 \leq i \leq n$. It is obvious that

$$\lim_{n \to \infty} F^*_n(x) = \begin{cases} 1 & \text{if } F^*(x) = 1 \\ 0 & \text{if } F^*(x) < 1 \end{cases}$$

which is of very little interest. The limiting distribution is thus found by considering a sequence $\{(X^*_n - \mu^*_n)/\sigma^*_n\}$ where the constants $\mu^*_n$ and $\sigma^*_n$ may depend on $n$ only ($\mu^*_n$ and $\sigma^*_n$ are not necessary measures of location and scale of the distribution of $X^*_n$). The EV-I distribution is found by taking $\sigma^*_n = 1$, and the other two types are derived by taking $\sigma^*_n \neq 1$.

In a fundamental paper Gnedenko (1943) established certain correspondences between the parent distribution $F^*(x)$ (from which the observations $X^*_1, \ldots, X^*_n$ come) and the type of EV distribution to which the limiting distribution belongs. It was shown that the conditions relate essentially to the behavior of $F^*(x)$ for large (or small) values of $x$ if the limiting distribution of the largest (or smallest) value of $\{X^*_1, \ldots, X^*_n\}$ is to be considered. It is quite possible that $X^*_n = \max\{X^*_1, \ldots, X^*_n\}$ and $X^*_1 = \min\{X^*_1, \ldots, X^*_n\}$ may have a different limiting distributions.

Gnedenko’s (1943) conditions, which are necessary and sufficient, do characterize the extreme value distributions described above. The conditions are given in the following:
(i) For Type-I EV distribution: Let $\xi^*_c$ be the $(100c)^{th}$ percentile point of the distribution $F^*(\cdot)$, i.e., $F^*(\xi^*_c) = c \ (0 < c < 1)$. Then the condition is:

$$\lim_{n \to \infty} n \left[ 1 - F^* \left( \xi^*_{1-1/n} + x \left( \xi^*_{1-1/(ne)} - \xi^*_{1-1/n} \right) \right) \right] = e^{-x} \quad (9.1.12)$$

(ii) For Type-II EV distribution:

$$\lim_{x \to \infty} \frac{1 - F^*(x)}{1 - F^*(bx)} = b^\alpha \quad \text{for any } b > 0, \alpha > 0 \quad (9.1.13)$$

(iii) For Type-III EV distribution: There exists $d \in \mathbb{R}$ such that

$$\lim_{x \to \infty} \frac{1 - F^*(bx + d)}{1 - F^*(x)} = b^\alpha \quad \text{for any } b > 0, \alpha > 0 \quad (9.1.14)$$

where $F^*(d) = 1$, $F^*(x) < 1$ for $x < d$

The extreme value distributions are used widely in many applied problems ranging from industrial engineering to clinical studies. Fuller (1914) first considered the concept of ‘extreme value’ in connection with flood control. Later Gumbel (1941, 1944, 1945, 1949), while studying the return of flood flows, gave the theory of EV distribution a solid foundation. In hydrological studies one is concerned about the height of a dam or a river-bank wall which is exceeded by annual maximum hourly or daily or weekly water level with a (small) probability $p$, which can vary from site to site depending upon the potential amount of damage from flooding. While one looks at maximum value of water level recorded over a fixed period for flood control, the opposite, i.e., minimum value of water level recorded over a fixed period is used for building ports or docks to maintain a certain level of navigation. For applications of EV distributions in sea-level studies see Tawn (1992) and many other references therein.

Gumbel (1937a, 1937b) used extreme value distributions for modelling human life times radioactive emissions. Other areas where EV distributions have found applications include meteorological data (Thom (1954), Jenkinson (1955)), corrosion studies (Aziz (1955), Eldredge (1957)), environmental studies (DasGupta and Bhaumik (1995)), and analyzing the results of horse races (Henery (1984)). Also, in industrial engineering problems the job characteristics like ovality, eccentricity, etc. follow an Extreme Value Distribution (see e.g., DasGupta, Ghosh, and Rao (1981),
DasGupta (1993)).

**Example 9.1.1.** DasGupta and Bhaumik (1995) reproduced a data set (Table 9.1.1) on ozone concentration in stratosphere which protects the earth from ultraviolet (UV) radiation emitted by the sun. The data set was originally reported by Pyle (1985). It has been observed that the ozone level is most stable at midnight and varies greatly during the day time. Also, the ozone level and its fluctuation are dependent on the altitude. The following table gives minimum of four observations (at an interval of 15 minutes) per hour recorded in terms of percentage deviation from midnight values for a diurnal cycle for two different altitudes (40 and 48 km from the earth’s surface).

<table>
<thead>
<tr>
<th>Hour</th>
<th>Altitude 40 km</th>
<th>Altitude 48 km</th>
<th>Hour</th>
<th>Altitude 40 km</th>
<th>Altitude 48 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>13</td>
<td>2.64</td>
<td>-5.21</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
<td>14</td>
<td>3.43</td>
<td>-5.86</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td>0.15</td>
<td>15</td>
<td>4.14</td>
<td>-6.00</td>
</tr>
<tr>
<td>3</td>
<td>0.15</td>
<td>0.15</td>
<td>16</td>
<td>4.43</td>
<td>-5.71</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
<td>0.20</td>
<td>17</td>
<td>4.14</td>
<td>-4.14</td>
</tr>
<tr>
<td>5</td>
<td>0.25</td>
<td>0.25</td>
<td>18</td>
<td>1.00</td>
<td>-1.43</td>
</tr>
<tr>
<td>6</td>
<td>0.30</td>
<td>0.30</td>
<td>19</td>
<td>0.00</td>
<td>-0.14</td>
</tr>
<tr>
<td>7</td>
<td>-1.00</td>
<td>-0.21</td>
<td>20</td>
<td>-0.25</td>
<td>-0.25</td>
</tr>
<tr>
<td>8</td>
<td>-1.14</td>
<td>0.86</td>
<td>21</td>
<td>-0.20</td>
<td>-0.20</td>
</tr>
<tr>
<td>9</td>
<td>-0.71</td>
<td>-0.57</td>
<td>22</td>
<td>-0.15</td>
<td>-0.15</td>
</tr>
<tr>
<td>10</td>
<td>0.21</td>
<td>-2.71</td>
<td>23</td>
<td>-0.10</td>
<td>-0.10</td>
</tr>
<tr>
<td>11</td>
<td>1.43</td>
<td>-4.14</td>
<td>24</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Table 9.1.1: Ozone Level Values (in percentage deviation from midnight = 0 hour) at Two Altitudes.

The following two frequency tables with five subgroups each have been created for the ozone level data for altitudes 40 and 48 km respectively.
Chapter 9  Extreme Value Distributions

The following two histograms are drawn with the above relative frequencies.

Figure 9.1.4: (a) Histogram for altitude 40 km and a matching EV-I pdf.
Example 9.1.2. The following data set (Table 9.1.3) was reported by Press (1950) where frequency distribution of maximum gust-velocity measurements per traverse was obtained and extreme value theory was applied to model this observed frequency distribution.

Note that in the table gust-velocities are all even values. For a better representation of the data and to draw a relative frequency histogram, each gust-velocity $x$ can be thought of representing the interval $(x-1, x+1)$. This is given in the following Table 9.1.4, and the corresponding histogram is given in Figure 9.1.5.

An interesting relationship among the three EV distributions discussed earlier is that by simple ‘$\ln$’ (natural logarithm) transformations we can transform Type-II and Type-III EV distributions to Type-I EV
Table 9.1.3: Frequency Distribution of 485 Maximum Gust-Velocity Measurements.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0–3.0</td>
<td>0.82%</td>
</tr>
<tr>
<td>3.0–5.0</td>
<td>2.26%</td>
</tr>
<tr>
<td>5.0–7.0</td>
<td>5.57%</td>
</tr>
<tr>
<td>7.0–9.0</td>
<td>9.90%</td>
</tr>
<tr>
<td>9.0–11.0</td>
<td>12.78%</td>
</tr>
<tr>
<td>11.0–13.0</td>
<td>11.96%</td>
</tr>
<tr>
<td>13.0–15.0</td>
<td>11.34%</td>
</tr>
<tr>
<td>15.0–17.0</td>
<td>12.37%</td>
</tr>
<tr>
<td>17.0–19.0</td>
<td>12.58%</td>
</tr>
<tr>
<td>19.0–21.0</td>
<td>7.42%</td>
</tr>
<tr>
<td>21.0–23.0</td>
<td>3.51%</td>
</tr>
<tr>
<td>23.0–25.0</td>
<td>3.71%</td>
</tr>
<tr>
<td>25.0–27.0</td>
<td>1.65%</td>
</tr>
<tr>
<td>27.0–29.0</td>
<td>1.44%</td>
</tr>
<tr>
<td>29.0–31.0</td>
<td>1.24%</td>
</tr>
<tr>
<td>31.0–33.0</td>
<td>0.62%</td>
</tr>
<tr>
<td>33.0–35.0</td>
<td>0.21%</td>
</tr>
<tr>
<td>35.0–37.0</td>
<td>0.41%</td>
</tr>
<tr>
<td>37.0–39.0</td>
<td>0.21%</td>
</tr>
</tbody>
</table>

Table 9.1.4: Relative Frequency Distribution of Gust-Velocity Measurements.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0–3.0</td>
<td>0.82%</td>
</tr>
<tr>
<td>3.0–5.0</td>
<td>2.26%</td>
</tr>
<tr>
<td>5.0–7.0</td>
<td>5.57%</td>
</tr>
<tr>
<td>7.0–9.0</td>
<td>9.90%</td>
</tr>
<tr>
<td>9.0–11.0</td>
<td>12.78%</td>
</tr>
<tr>
<td>11.0–13.0</td>
<td>11.96%</td>
</tr>
<tr>
<td>13.0–15.0</td>
<td>11.34%</td>
</tr>
<tr>
<td>15.0–17.0</td>
<td>12.37%</td>
</tr>
<tr>
<td>17.0–19.0</td>
<td>12.58%</td>
</tr>
<tr>
<td>19.0–21.0</td>
<td>7.42%</td>
</tr>
<tr>
<td>21.0–23.0</td>
<td>3.51%</td>
</tr>
<tr>
<td>23.0–25.0</td>
<td>3.71%</td>
</tr>
<tr>
<td>25.0–27.0</td>
<td>1.65%</td>
</tr>
<tr>
<td>27.0–29.0</td>
<td>1.44%</td>
</tr>
<tr>
<td>29.0–31.0</td>
<td>1.24%</td>
</tr>
<tr>
<td>31.0–33.0</td>
<td>0.62%</td>
</tr>
<tr>
<td>33.0–35.0</td>
<td>0.21%</td>
</tr>
<tr>
<td>35.0–37.0</td>
<td>0.41%</td>
</tr>
<tr>
<td>37.0–39.0</td>
<td>0.21%</td>
</tr>
</tbody>
</table>

distribution as shown below.

If $X$ follows EV-II $(\alpha, \beta, \mu)$ (i.e., has pdf (9.1.4)) then $Y = \ln(X - \mu)$ follows EV-I $(\beta_*, \mu_*)$ (with pdf (9.1.1)) where $\beta_* = 1/\alpha$ and $\mu_* = \ln \beta$. Note that the range of the new random variable $Y$ is the whole real line.

Similarly, if $X$ follows EV-III $(\alpha, \beta, \mu)$ (i.e., has pdf (9.1.5)) then $Y = -\ln(\mu - X)$ follows EV-I $(\beta_*, \mu_*)$ (with pdf (9.1.1)) where $\beta_* = 1/\alpha$.
and $\mu_* = -ln\beta$. Again, note that the new random variable $Y$ has the whole real line as its range.

Converse of the above two properties are also true. If $X$ follows EVT-1 ($\beta, \mu$) (as given in (9.1.1)) then:

(a) $Y_1 = \mu + exp(X)$ follows EVT-II($\alpha_*, \beta_*, \mu_*$) where $\alpha_* = 1/\beta$, $\beta_* = \exp(\mu)$ and $\mu_* = \mu$

(b) $Y_2 = \mu - exp(-X)$ follows EVT-III($\alpha_*, \beta_*, \mu_*$), where $\alpha_* = 1/\beta$, $\beta_* = \exp(-\mu)$ and $\mu_* = \mu$

Also, note that if $X$ follows Type-I EV distribution (with pdf (9.1.1)) then $U = \exp(-\frac{(X - \mu)}{\beta})$ has the standard exponential distribution,
i.e., the pdf of $U$ is $e^{-u}$, $u \geq 0$. Using this property of EV-I $(\beta, \mu)$, the moment generating function (mgf) of $X$ can be found as

$$E(e^{tX}) = e^{t\mu \Gamma(1 - \beta t)}, \quad \text{for } |\beta t| < 1 \quad (9.1.15)$$

From (9.1.15) one can find the mean of $X$ as

$$E(X) = \mu - \beta \Gamma'(1), \quad \text{where } \Gamma'(1) = \frac{\partial}{\partial x} \Gamma(x)|_{x=1} \quad (9.1.16)$$

and the variance of $X$ as

$$Var(X) = \beta^2 \left[ \Gamma''(1) - (\Gamma'(1))^2 \right], \quad \text{where } \Gamma''(1) = \frac{\partial^2}{\partial x^2} \Gamma(x)|_{x=1} \quad (9.1.17)$$

The EV-I $(\beta, \mu)$ is unimodal with mode at $\mu$. From the distribution function (9.1.2) the $100\gamma^{\text{th}}$ ($0 < \gamma < 1$) percentile point $\xi_\gamma$ can be found by solving the equation

$$\exp \left\{ -e^{-\frac{(\xi_\gamma - \mu)}{\beta}} \right\} = \gamma$$

which yields

$$\xi_\gamma = \mu - \beta \ln(-\ln\gamma) \quad (9.1.18)$$

In particular, the median (i.e., $\xi_{0.5}$) is obtained by using $\gamma = 0.5$, i.e., $\xi_{0.5} \approx \mu + (0.366)\beta$.

The following table gives the percentiles for standard Type-I EV distribution (i.e., EV-I $(1, 0)$).

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\xi_\gamma$</th>
<th>$\gamma$</th>
<th>$\xi_\gamma$</th>
<th>$\gamma$</th>
<th>$\xi_\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>-1.957</td>
<td>0.100</td>
<td>-1.100</td>
<td>0.950</td>
<td>1.866</td>
</tr>
<tr>
<td>0.005</td>
<td>-1.750</td>
<td>0.250</td>
<td>-0.705</td>
<td>0.975</td>
<td>2.416</td>
</tr>
<tr>
<td>0.010</td>
<td>-1.640</td>
<td>0.500</td>
<td>-0.164</td>
<td>0.990</td>
<td>3.137</td>
</tr>
<tr>
<td>0.025</td>
<td>-1.468</td>
<td>0.750</td>
<td>0.521</td>
<td>0.995</td>
<td>3.679</td>
</tr>
<tr>
<td>0.050</td>
<td>-1.306</td>
<td>0.900</td>
<td>1.305</td>
<td>0.999</td>
<td>4.936</td>
</tr>
</tbody>
</table>

Table 9.1.5: Percentiles of the Standard Type-I Extreme Value Distribution.

9.2 Characterizations of Extreme Value Distributions

In the previous section it has been discussed how Type-II and Type-III EV distributions are related to the Type-I EV distribution (through
the simple ‘\(\ln\)’ transformation). Therefore, in this section we talk about characterization of Type-I EV distribution only. However, note that, Gnedenko’s (1943) necessary and sufficient conditions do serve as most general characterization of the three types of EV distributions.

It is seen that a random variable \(X\) has a Type-I EV distribution, if and only if \(\exp\{-(X-\mu)/\beta\}\) has the standard exponential distribution. Therefore, the characterization results for exponential distribution can also be used for Type-I EV distribution, by applying them to \(\exp\{-(X-\mu)/\beta\}\).

Dubey (1966) characterized the Type-I EV distribution by the distribution of the smallest order statistics from the same distribution.

**Theorem 9.2.1** Let \(X_1, X_2, \ldots, X_n\) be iid observations and \(X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}\). Then, \(X_{(1)}\) follows a Type-I EVD distribution with pdf (9.1.11) if and only if \(X_i\)’s are so.

Ballerini (1987) provided a characterization of the Type-I EV distribution based on record analysis as given below.

**Theorem 9.2.2** Suppose \(\{X_i\}_{i \geq 0}\) be a sequence of iid nondegenerate continuous random variables with cdf \(F\). Assume that \(EX_0^+ < \infty\) (where \(X_0^+ = X_0I(X_0 > 0)\)). Define \(Y_i = (X_i + ic)\) where \(c > 0\) and \(i \geq 0\). Let \(Y_{(n)} = \max\{Y_1, \ldots, Y_n\}\), and \(1_n = I[Y_n > Y_{(n-1)}] = I[\text{record occurs at time} \ n]\). If for every \(c > 0\), \(Y_{(n)}\) is independent of \(1_n\) for each \(n\), then the cdf \(F\) has the form (9.1.2) for some \(\mu\) real and \(\beta > 0\).

Among the other characterizations of the Type-I EV distribution, prominent are Leadbetter, Lindgren, and Rootzén (1983) where it has been shown that EV-I \((\beta, \mu)\) is the only ‘max-stable’ distribution with the entire real line as its support. Also see Sethuraman (1965) for characterization of all three types of EV distributions in terms of ‘complete confounding’ of random variables.

Before we go to the next section, we would like to observe the close relationship between the Weibull distribution and the Type-I Extreme Value Distribution.

**Result 9.2.1** Suppose \(X\) follows a two parameter Weibull distribution \((W(\alpha, \beta))\) with pdf (8.1.1). Then (i) \(Y = \ln X\) has the EV-I \((\beta_*, \mu_*)\) of smallest values with pdf (9.1.11) where the location parameter \(\mu_* = \ln \beta\) and scale parameter \(\beta_* = 1/\alpha\); (ii) \(Y = -\ln X\) has the EV-I \((\beta_*, \mu_*)\) of...
Chapter 9  Extreme Value Distributions

Chapter 9

largest values with pdf (9.1.1) where the location parameter \(\mu_* = -\ln\beta\) and scale parameter \(\beta_* = 1/\alpha\).

Conversely, if \(X\) follows EV-I \((\beta, \mu)\) with pdf (9.1.1) (or \(9.1.11\)), then \(Y = \exp(X)\) (or \(\exp(-X)\)) follows a two-parameter Weibull distribution \((W(\alpha_*, \beta_*))\) with pdf (8.1.1) where \(\alpha_* = 1/\beta_*\) and \(\beta_* = \exp(\mu)\) (or \(\exp(-\mu)\)).

**Remark 9.2.1** A similar relationship between the Weibull distribution and the other two types of EV distributions can be established easily through the relationships that exist between Type-I EV distribution and the Type-II and Type-III EV distributions as discussed earlier.

Another interesting result is the relationship between the Type-I EV distribution and the logistic distribution (see Section 2.2) as stated below.

**Result 9.2.2** Let \(X_1\) and \(X_2\) be two independent random variables having the Type-I EV distributions with common scale parameter \(\beta\) but different (possibly) location parameters \(\mu_1\) and \(\mu_2\) respectively (i.e., \(X_i\) follows EV-I \((\beta, \mu_i)\), \(i = 1, 2\)). Then \(Y = (X_1 - X_2)\) follows a two-parameter logistic distribution with location parameter \(\mu = (\mu_1 - \mu_2)\) and scale parameter \(\beta\).

### 9.3 Estimation of Parameters

First we consider estimation of parameters for the Type-I EV distribution only. Then we consider the other two types of EV distribution.

Given iid observations \(X_1, X_2, \ldots, X_n\) having the common pdf (9.1.1), the maximum likelihood estimators \(\hat{\beta}_{ML}\) and \(\hat{\mu}_{ML}\) of \(\beta\) and \(\mu\) respectively satisfy the equations:

\[
\sum_{i=1}^{n} \exp\{- (X_i - \hat{\mu}_{ML})/\hat{\beta}_{ML}\} = n \tag{9.3.1}
\]

\[
\bar{X} - \hat{\mu}_{ML} - \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_{ML}) \exp\{- (X_i - \hat{\mu}_{ML})/\hat{\beta}_{ML}\} = \hat{\beta}_{ML} \tag{9.3.2}
\]

where \(\bar{X} = \sum_{i=1}^{n} X_i/n\). Further simplification of (9.3.1) gives \(\hat{\mu}_{ML}\) in terms of \(\hat{\beta}_{ML}\) as

\[
\hat{\mu}_{ML} = -\hat{\beta}_{ML} \ln\left\{\frac{1}{n} \sum_{i=1}^{n} \exp(-X_i/\hat{\beta}_{ML})\right\} \tag{9.3.3}
\]
which after plugging in (9.3.2) gives \( \hat{\beta}_{ML} \) in terms of the following equation:

\[
\hat{\beta}_{ML} = X - \left\{ \sum_{i=1}^{n} X_i \exp(-X_i/\hat{\beta}_{ML}) \right\} / \left\{ \sum_{i=1}^{n} \exp(-X_i/\hat{\beta}_{ML}) \right\} \quad (9.3.4)
\]

The above MLEs, which are asymptotically unbiased, are obtained through numerical methods and have the following asymptotically properties (Downton (1966)).

\[
\lim_{n \to \infty} \text{Var}(\sqrt{n}\hat{\mu}_{ML}) = \left\{ 1 + 6(1 - \gamma)^2 / \pi^2 \right\} \beta^2 \approx 1.1087 \beta^2
\]

\[
\lim_{n \to \infty} \text{Var}(\sqrt{n}\hat{\beta}_{ML}) = 6\beta^2 / \pi^2 \approx 0.6079 \beta^2
\]

\[
\lim_{n \to \infty} \text{Corr}(\hat{\mu}_{ML}, \hat{\beta}_{ML}) = (\gamma - 1) \sqrt{\frac{6}{n^2}} / \sqrt{1 + 6(1 - \gamma)^2 / \pi^2} \approx 0.313 \quad (9.3.5)
\]

where \( \gamma = 0.57721566 \ldots \) is Euler’s constant. However, for fixed \( n \), the above MLEs are biased.

In many practical problems one does not have a complete data set due to censoring. For instance, out of a random sample of size \( n \) the smallest \( r_1 \) and the highest \( r_2 \) sample values are not available, i.e., we observe only the ordered values \( X_{(r_1+1)} \leq X_{(r_1+2)} \leq \ldots \leq X_{(n-r_2)} \). Harter and Moore (1968) provided the method of deriving the MLEs of \( \mu \) and \( \beta \) based on the doubly censored observations \( X_{(r_1+1)} \leq \ldots \leq X_{(n-r_2)} \) as given below.

Define \( Z_i = (X_{(i)} - \mu) / \beta, \quad F(Z_i) = \frac{1}{2} \pm \frac{1}{2} \exp\{-\exp(\pm Z_i)\}, \quad i = (r_1+1), \ldots, (n-r_2) \). Also, \( f(Z_i) = \exp\{\pm Z_i - \exp(\pm Z_i)\} \), and the upper (lower) signs being those for the distribution of smallest (largest) values (with cdf (9.1.10) (9.1.2) as mentioned in Section 9.1). The MLEs of \( \mu \) and \( \beta \) based on the doubly censored (DC) observations, denoted by \( \hat{\mu}_{ML}^{DC} \) and \( \hat{\beta}_{ML}^{DC} \), are obtained by solving the following two equations for \( \mu \) and \( \beta \):

\[
\mp \left( \frac{n - r_1 - r_2}{\beta} \right) \frac{1}{\beta} \sum_{i=r_1+1}^{n-r_2} \exp(\pm Z_i) - r_1 f(Z_{r_1+1}) / [\beta F(Z_{r_1+1})] + r_2 f(Z_{n-r_2}) / [\beta \{1 - F(Z_{n-r_2})\}] = 0 \quad (9.3.6)
\]

and

\[
\frac{1}{\beta} \sum_{i=r_1+1}^{n-r_2} Z_i \pm \frac{1}{\beta} \sum_{i=r_1+1}^{n-r_2} Z_i \exp(\pm Z_i) - \left( \frac{n - r_1 - r_2}{\beta} \right) - r_1 Z_{r_1+1}
\]
\[ f(Z_{r_1+1}/|\beta F(Z_{r_1+1})| + r_2Z_{n-r_2}f(Z_{n-r_2})/|\beta(1 - F(Z_{n-r_2}))|) = 0 \] (9.3.7)

To find the asymptotic dispersion matrix of \((\hat{\beta}_{ML}^{DC}, \hat{\mu}_{ML}^{DC})\) we define the following terms:
\[
\Gamma_\omega(y) = \int_0^\omega u^{y-1}e^{-u}du \\
\Gamma'_\omega(a) = \frac{d}{dy}\left[\Gamma_\omega(y)\right]_{y=a} \\
\Gamma''_\omega(a) = \frac{d^2}{dy^2}\left[\Gamma_\omega(y)\right]_{y=a}
\]

Let \(q_1 = r_1/n\), \(q_2 = r_2/n\), and \(p = 1 - (q_1 + q_2) = 1 - (r_1 + r_2)/n\). Keeping \(q_1\) and \(q_2\) fixed, as \(n \to \infty\), \(Z_{r_1+1} \to \tilde{z}_1\) and \(Z_{n-r_2} \to \tilde{z}_2\) such that
\[
F(\tilde{z}_1) = \int_{-\infty}^{\tilde{z}_1} f(t) dt = q_1 \\
1 - F(\tilde{z}_2) = \int_{\tilde{z}_2}^{\infty} f(t) dt = q_2
\]

For the distribution of the smallest values, the information matrix is
\[
I = \left(\begin{array}{cc}
I_{11} & I_{12} \\
I_{21} & I_{22}
\end{array}\right)
\]
where
\[
I_{11} = \frac{n}{\beta^2} \left[ -p - 2 \left\{ \Gamma'_{-lnq_2}(1) - \Gamma'_{-ln(1-q_1)}(1) \right\} + \Gamma''_{-lnq_2}(2) - \Gamma''_{-ln(1-q_1)}(2) \\
+ 2 \left\{ \Gamma'_{-lnq_2}(2) - \Gamma'_{-ln(1-q_1)}(2) \right\} - q_2lnq_2 ln(-lnq_2) \{2 + ln(-lnq_2)\} \\
+ (1 - q_1)ln(1 - q_1)ln \{-ln(1 - q_1)\} \{2 + ln \{-ln(1 - q_1)\}\} \\
+ ln(1 - q_1)ln \{-ln(1 - q_1)\} / q_1 \right] \\
I_{22} = \frac{n}{\beta^2} \left[ p + q_2ln q_2 - (1 - q_1)ln(1 - q_1) \right] \\
I_{12} = I_{21} = \frac{n}{\beta^2} \left[ \Gamma'_{-lnq_2}(2) - \Gamma'_{-ln(1-q_1)}(2) - q_2lnq_2 ln(-lnq_2) \\
+ (1 - q_1)ln(1 - q_1)ln \{-ln(1 - q_1)\} \\
+ 1/1 - (1 - q_1)^2 ln \{-ln(1 - q_1)\} \right] (9.3.8)
\]
The asymptotic dispersion matrix of \((\hat{\beta}_{ML}^{DC}, \hat{\mu}_{ML}^{DC})\) is \(I^{-1}\).

If one considers the distribution of the largest values instead of the distribution of smallest values, then it can be shown that by interchanging \(q_1\) and \(q_2\), the values of \(I_{11}\) and \(I_{22}\) remain the same, and \(I_{12}\) (the \((1,2)\)-element of the matrix \(I^{-1}\)) retains the same absolute value but changes the sign.
In the case of single (or one-sided) censoring, where we observe only the \( r \) smallest observations \( X(1) \leq X(2) \leq \ldots \leq X(r) \) from a sample of size \( n \), Schüpbach and Hüsler (1983) proposed simple looking unbiased estimators of \( \mu \) and \( \beta \) for the model (9.1.11) given as

\[
\hat{\beta}_{SH} = \left\{ -\sum_{i=1}^{l} X(i) + \left( \frac{l}{r-l} \right) \sum_{i=l+1}^{r} X(i) \right\} / (nk_{r,n})
\]

\[
\hat{\mu}_{SH} = \bar{X}_{(r)} + d_{r,n} \hat{\beta}_{SH}
\]

where \( \bar{X}_{(r)} = \sum_{i=1}^{r} X(i)/r \), \( r = \lfloor pn \rfloor < n \), \( l = \lfloor qn \rfloor < r \), \( 0 < p < 1 \), \( 0 < q < 1 \). Also, \( l \) is chosen such that the variance of \( \hat{\beta}_{SH} \) is minimized, and \( k_{r,n} \) and \( d_{r,n} \) are suitable constants. It has been observed that for \( 0 < p < 0.8 \), the optimal choice of \( l \) is \((r - 1)\). For \( 0.8 < p < 1.0 \), Schüpbach and Hüsler (1983) provided the values of optimal \( l \) and \( k_{r,n} \) as well as \( d_{r,n} \) as given below in Table 9.3.1.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( r )</th>
<th>( l )</th>
<th>( k_{r,n} )</th>
<th>( d_{r,n} )</th>
<th>( n )</th>
<th>( r )</th>
<th>( l )</th>
<th>( k_{r,n} )</th>
<th>( d_{r,n} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>0.800</td>
<td>0.894</td>
<td>14</td>
<td>12</td>
<td>1.27</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>7</td>
<td>1.067</td>
<td>0.751</td>
<td>13</td>
<td>11</td>
<td>0.982</td>
<td>0.907</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>0.962</td>
<td>0.918</td>
<td>17</td>
<td>16</td>
<td>13</td>
<td>1.240</td>
<td>0.687</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>8</td>
<td>1.120</td>
<td>0.737</td>
<td>15</td>
<td>13</td>
<td>1.158</td>
<td>0.789</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>1.016</td>
<td>0.890</td>
<td>14</td>
<td>12</td>
<td>1.016</td>
<td>0.889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>9</td>
<td>1.166</td>
<td>0.726</td>
<td>18</td>
<td>17</td>
<td>14</td>
<td>1.268</td>
<td>0.681</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>1.065</td>
<td>0.866</td>
<td>16</td>
<td>14</td>
<td>1.187</td>
<td>0.778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>10</td>
<td>1.207</td>
<td>0.716</td>
<td>15</td>
<td>13</td>
<td>1.048</td>
<td>0.873</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>1.013</td>
<td>0.846</td>
<td>19</td>
<td>18</td>
<td>15</td>
<td>1.294</td>
<td>0.676</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>11</td>
<td>1.244</td>
<td>0.707</td>
<td>17</td>
<td>15</td>
<td>1.214</td>
<td>0.760</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>1.055</td>
<td>0.829</td>
<td>16</td>
<td>14</td>
<td>1.078</td>
<td>0.859</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>12</td>
<td>1.277</td>
<td>0.699</td>
<td>20</td>
<td>19</td>
<td>16</td>
<td>1.318</td>
<td>0.672</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>1.093</td>
<td>0.814</td>
<td>18</td>
<td>16</td>
<td>1.239</td>
<td>0.760</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>1.019</td>
<td>0.927</td>
<td>17</td>
<td>15</td>
<td>1.105</td>
<td>0.846</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>15</td>
<td>13</td>
<td>1.307</td>
<td>0.693</td>
<td>16</td>
<td>14</td>
<td>0.9905</td>
<td>0.932</td>
<td></td>
</tr>
</tbody>
</table>

Table 9.3.1: Values of \( l \), \( k_{r,n} \), and \( d_{r,n} \) for Selected \( n \) and \( r \) (0.8 \leq r/n < 1.0).

In the case of complete sample \( X(1) \leq X(2) \leq \ldots \leq X(n) \), the estimators in (9.3.9) remain same except that
$l = [0.84n] = \text{largest integer not exceeding } 0.84n$

d_{r,n} \text{ is replaced by } \gamma = 0.57721566 \text{ (Euler’s constant)}$

$k_{r,n} \text{ is replaced by } k_n; \text{ given in the following Table 9.3.2 (Engelhardt and Bain (1977))}$

<table>
<thead>
<tr>
<th>(n)</th>
<th>(k_n)</th>
<th>(n)</th>
<th>(k_n)</th>
<th>(n)</th>
<th>(k_n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.6931</td>
<td>22</td>
<td>1.4609</td>
<td>42</td>
<td>1.5208</td>
</tr>
<tr>
<td>3</td>
<td>0.9808</td>
<td>23</td>
<td>1.4797</td>
<td>43</td>
<td>1.5303</td>
</tr>
<tr>
<td>4</td>
<td>1.1507</td>
<td>24</td>
<td>1.4975</td>
<td>44</td>
<td>1.4891</td>
</tr>
<tr>
<td>5</td>
<td>1.2674</td>
<td>25</td>
<td>1.5142</td>
<td>45</td>
<td>1.4984</td>
</tr>
<tr>
<td>6</td>
<td>1.3545</td>
<td>26</td>
<td>1.4479</td>
<td>46</td>
<td>1.5075</td>
</tr>
<tr>
<td>7</td>
<td>1.1828</td>
<td>27</td>
<td>1.4642</td>
<td>47</td>
<td>1.5163</td>
</tr>
<tr>
<td>8</td>
<td>1.2547</td>
<td>28</td>
<td>1.4796</td>
<td>48</td>
<td>1.5248</td>
</tr>
<tr>
<td>9</td>
<td>1.3141</td>
<td>29</td>
<td>1.4943</td>
<td>49</td>
<td>1.5331</td>
</tr>
<tr>
<td>10</td>
<td>1.3644</td>
<td>30</td>
<td>1.5083</td>
<td>50</td>
<td>1.5411</td>
</tr>
<tr>
<td>11</td>
<td>1.4079</td>
<td>31</td>
<td>1.5216</td>
<td>51</td>
<td>1.5046</td>
</tr>
<tr>
<td>12</td>
<td>1.4461</td>
<td>32</td>
<td>1.4665</td>
<td>52</td>
<td>1.5126</td>
</tr>
<tr>
<td>13</td>
<td>1.3332</td>
<td>33</td>
<td>1.4795</td>
<td>53</td>
<td>1.5204</td>
</tr>
<tr>
<td>14</td>
<td>1.3686</td>
<td>34</td>
<td>1.4920</td>
<td>54</td>
<td>1.5279</td>
</tr>
<tr>
<td>15</td>
<td>1.4004</td>
<td>35</td>
<td>1.5040</td>
<td>55</td>
<td>1.5352</td>
</tr>
<tr>
<td>16</td>
<td>1.4293</td>
<td>36</td>
<td>1.5156</td>
<td>56</td>
<td>1.5424</td>
</tr>
<tr>
<td>17</td>
<td>1.4556</td>
<td>37</td>
<td>1.5266</td>
<td>57</td>
<td>1.5096</td>
</tr>
<tr>
<td>18</td>
<td>1.4799</td>
<td>38</td>
<td>1.4795</td>
<td>58</td>
<td>1.5167</td>
</tr>
<tr>
<td>19</td>
<td>1.3960</td>
<td>39</td>
<td>1.4904</td>
<td>59</td>
<td>1.5236</td>
</tr>
<tr>
<td>20</td>
<td>1.4192</td>
<td>40</td>
<td>1.5009</td>
<td>60</td>
<td>1.5304</td>
</tr>
<tr>
<td>21</td>
<td>1.4408</td>
<td>41</td>
<td>1.5110</td>
<td>(\infty)</td>
<td>1.5692</td>
</tr>
</tbody>
</table>

Table 9.3.2: Values of \(k_n\) for \(2 \leq n < \infty\).

Another simple way of estimating the Type-I EV Distribution parameters is to match the log-log (natural) transformation of the empirical distribution function with that of the \(cdf\) and then use a simple regression method.

Note that the \(cdf\) in (9.1.2) (distribution of largest values) yields

\[-\ln(-\ln F_1(x)) = (x - \mu)/\beta\]
9.3 Estimation of Parameters

\[ x = \mu + \beta \{ -\ln(-\ln F_1(x)) \} \]  \hspace{1cm} (9.3.10)

On the other hand, the empirical distribution function based on the doubly censored observations \( X_{(r_1+1)}, \ldots, X_{(n-r_2)} \) is \( \hat{F}_1(x) \) where

\[ \hat{F}_1(x) = \{ \# X_{(i)} s \leq x \} / (n - r_1 - r_2) \]  \hspace{1cm} (9.3.11)

For \( i = (r_1 + 1), (r_1 + 2), \ldots, (n - r_2) \), we have from (9.3.11),

\[ \hat{F}_1(X_{(i)}) = (i - r_1) / (n - r_1 - r_2) \]  \hspace{1cm} (9.3.12)

Now in (9.3.10) replace \( F_1(x) \) by \( \hat{F}_1(\cdot) \) for \( x = X_{(r_1+1)}, \ldots, X_{(n-r_2)} \), yielding

\[ X_{(i)} = \mu + \beta \{ -\ln(-\ln((i - r_1)/(n - r_1 - r_2))) \} \]

i.e.,

\[ X_{(i)} = \mu + \beta \{ -\ln(\ln(n - r_1 - r_2) - \ln(i - r_1)) \} \]  \hspace{1cm} (9.3.13)

Use a simple linear regression approach (where \( X_{(i)} \)'s play the role of dependent variable values and \( \{ -\ln(\ln(n - r_1 - r_2) - \ln(i - r_1)) \} \)'s are the independent variable values) to estimate \( \mu \) and \( \beta \), and call the resultant estimators \( \hat{\mu}_R \) and \( \hat{\beta}_R \) respectively.

Identical steps can be followed for the distribution of smallest values (i.e., \( \text{cdf} \) (9.1.10)).

Also note that when \( r_1 = 0 \) and \( r_2 = n \), we have the full sample (uncensored).

In the following we use the data set in Example 9.1.2 to estimate the parameters of a Type-I EV distribution.

<table>
<thead>
<tr>
<th>Method</th>
<th>( \hat{\beta} )</th>
<th>( \hat{\mu} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>5.2446</td>
<td>11.7128</td>
</tr>
<tr>
<td>Schüpbach &amp; Huster</td>
<td>5.21717</td>
<td>11.6112</td>
</tr>
<tr>
<td>Regression</td>
<td>4.81539</td>
<td>11.8520</td>
</tr>
</tbody>
</table>

Table 9.3.3: Estimated Parameters with Various Methods.

Since this data set is large (with \( n = 485 \)), we apply the bootstrap method to estimate the bias and MSE of the above three types of estimates. Number of replications used here is 10,000.
## 9.4 Goodness of Fit Tests for Extreme Value Distributions

In this section we discuss the goodness of fit tests for extreme value distributions of Type-I (i.e., EV-I(\(\beta, \mu\))) only.

Shapiro and Brain (1987) proposed a goodness of fit test for the Type-I limiting distribution of the smallest value of iid random variables, i.e., for the distribution with pdf (9.1.11). To test that a random sample \(X_1, X_2, \ldots, X_n\) is following the EV-I(\(\beta, \mu\)) with pdf (9.1.11), we first order the observations \(X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}\) and then define the following quantities:

\[
\omega_i = \ln\left(\frac{n + 1}{n - i + 1}\right), \quad i = 1, 2, \ldots, (n - 1)
\]

\[
\omega_n = n - \sum_{i=1}^{n-1} \omega_i
\]

\[
\omega_{n+i} = \omega_i(1 + \ln \omega_i) - 1, \quad i = 1, 2, \ldots, (n - 1)
\]

\[
\omega_{2n} = 0.422n - \sum_{i=1}^{n-1} \omega_{n+i}
\]

\[
A_1 = \sum_{i=1}^{n} \omega_i X_{(i)} \quad \text{and} \quad A_2 = \sum_{i=1}^{n} \omega_{n+i} X_{(i)}
\]

The test statistic to test goodness of fit is

\[
W_n = \frac{(0.6079A_2 - 0.2570A_1)^2}{n \sum_{i=1}^{n} (X_{(i)} - \bar{X})^2}
\]
Goodness of Fit Tests for Extreme Value Distributions

which provides a two-tail test of the null hypothesis (i.e., the data follows (9.1.11)). The distribution of $W_n$ under the null hypothesis is not analytically tractable. However, the above authors carried out a comprehensive simulation work based on which percentile points of the null distribution of $W_n$ were expressed as a function of the sample size. For $0 < \alpha < 1$, let $W_{n,\alpha}$ be the $(100\alpha)^{th}$ percentile point, i.e.,

$$\alpha = P(W \leq W_{n,\alpha} \mid \text{null hypothesis is true}) \quad (9.4.3)$$

Shapiro and Brain (1987) showed that

$$W_{n,\alpha} \approx \alpha_0 + \alpha_1 (\ln(n)) + \alpha_2 (\ln(n))^2 \quad (9.4.4)$$

The values of $\alpha$, $\alpha_0$, $\alpha_1$ and $\alpha_2$ are provided in the following.

<table>
<thead>
<tr>
<th>$0 &lt; \alpha &lt; 1$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>0.10102</td>
<td>0.04249</td>
<td>0.005882</td>
</tr>
<tr>
<td>0.025</td>
<td>0.11787</td>
<td>0.08550</td>
<td>-0.002048</td>
</tr>
<tr>
<td>0.050</td>
<td>0.13200</td>
<td>0.10792</td>
<td>-0.006487</td>
</tr>
<tr>
<td>0.950</td>
<td>1.46218</td>
<td>-0.21111</td>
<td>0.012914</td>
</tr>
<tr>
<td>0.975</td>
<td>1.64869</td>
<td>-0.26411</td>
<td>0.016840</td>
</tr>
<tr>
<td>0.995</td>
<td>1.91146</td>
<td>-0.31361</td>
<td>0.017669</td>
</tr>
</tbody>
</table>

Table 9.4.1: Values of $\alpha$, $\alpha_0$, $\alpha_1$, and $\alpha_2$.

Using the formula (9.4.4) and the Table 9.4.1, the percentile points of $W_n$ under the null hypothesis can be found as given below.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.005</td>
</tr>
<tr>
<td>10</td>
<td>0.18464</td>
</tr>
<tr>
<td>15</td>
<td>0.23004</td>
</tr>
<tr>
<td>20</td>
<td>0.25922</td>
</tr>
<tr>
<td>25</td>
<td>0.28110</td>
</tr>
<tr>
<td>30</td>
<td>0.29873</td>
</tr>
<tr>
<td>35</td>
<td>0.31358</td>
</tr>
<tr>
<td>40</td>
<td>0.32644</td>
</tr>
</tbody>
</table>

Table 9.4.2: Values of $W_{n,\alpha}$ for Selected Values of $n$ and $\alpha$. 

© 2006 by Taylor & Francis Group, LLC
To test that a random sample \( X_1, X_2, \ldots, X_n \) is following the EV-I(\( \beta, \mu \)) with pdf (9.1.1) (i.e., the limiting distribution of the largest value of iid observations), one needs to work with \( X^*_1, X^*_2, \ldots, X^*_n \) where \( X^*_i = -X_i \). Note that \( X^*_i \) follows (9.1.11) if and only if \( X_i \) follows (9.1.1). Hence, one works with \( X^*_i \)'s, defines the \( W_n \) statistic based on \( X^*_i(1) \leq X^*_i(2) \leq \ldots \leq X^*_i(n) \), and make use of \( W_{n,\alpha} \) as discussed above.

At level \( \alpha \), one accepts the null hypothesis if \( W_{n,\alpha/2} \leq W_n \leq W_{n,1-(\alpha/2)} \); and rejects otherwise.

As an application, we consider the data set (of size \( n = 24 \)) given in the Example 9.1.1. The values for the altitude of 40 km are used to see if the pdf (9.1.11) is appropriate for the data. Our ordered observations are (row wise):

\[
\begin{align*}
-1.14, & \quad -1.00, \quad -0.71, \quad -0.25, \quad -0.20, \quad -0.15, \\
-0.10, & \quad -0.05, \quad 0.00, \quad 0.00, \quad 0.10, \quad 0.15, \\
0.20, & \quad 0.20, \quad 0.21, \quad 0.25, \quad 0.30, \quad 1.00, \\
1.43, & \quad 2.64, \quad 3.43, \quad 4.14, \quad 4.14, \quad 4.43.
\end{align*}
\]

<table>
<thead>
<tr>
<th>i</th>
<th>( \omega_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0408</td>
</tr>
<tr>
<td>2</td>
<td>0.0834</td>
</tr>
<tr>
<td>3</td>
<td>0.1278</td>
</tr>
<tr>
<td>4</td>
<td>0.1744</td>
</tr>
<tr>
<td>5</td>
<td>0.2231</td>
</tr>
<tr>
<td>6</td>
<td>0.2744</td>
</tr>
<tr>
<td>7</td>
<td>0.3285</td>
</tr>
<tr>
<td>8</td>
<td>0.3857</td>
</tr>
</tbody>
</table>

Table 9.4.3: The Coefficients \( W_i \)'s for \( n = 24 \).

Since \( n = 24 \), we first calculate the \( W_i \) values, \( 1 \leq i \leq (n - 1) = 23 \). We now have \( A_1 = 48.1970, A_2 = 69.6708, W_n = 0.585897 \). With \( \alpha = 0.1 \), \( W_{n,(\alpha/2)} = 0.409457 \) and \( W_{n,1-(\alpha/2)} = 0.921693 \). Since \( W_{n,(\alpha/2)} < W_n < W_{n,1-(\alpha/2)} \), we accept the fact that the above data set can be modelled by EV-I(\( \beta, \mu \)).
Chapter 10

System Reliability

10.1 Preliminaries

Complex machines, henceforth referred to as ‘systems’, are made of smaller devices, called components. The life span of the whole system depends on the life span of the individual components which are often independent of each other. Depending on the nature of the system (i.e., how the components are connected with each other), the system works if some or all of the components work. On the other hand, the system fails as soon as some or all of the components fail. In this chapter we shall deal mainly with reliability of two types of systems — Series System and Parallel System.

One can handle a system reliability in two ways as described below.

<table>
<thead>
<tr>
<th>Study of a System Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case-A</strong></td>
</tr>
<tr>
<td>We have lifetime data available on the whole system, or it is possible to generate data on the lifetime of the whole system conveniently</td>
</tr>
<tr>
<td><strong>Case-B</strong></td>
</tr>
<tr>
<td>We do not have data on the whole system life, or it is not feasible (or cost effective) to generate data on the lifetime of the whole system</td>
</tr>
</tbody>
</table>

Figure 10.1.1: Two possible ways of handling a system reliability.
Chapter 10  System Reliability

Case-A: It is possible to have lifetime data on the whole system.

In this case usually the data tend to follow a positively skewed distribution, and hence can be modelled by one of the distributions discussed in the earlier chapters (Chapter 6 to Chapter 9).

As an example of this case, consider the problem of studying the lifetime of a particular brand of fax machines. Though a fax machine has numerous smaller components in it, it is possible to build several fax machines at a relatively low cost and to study how long they can survive. The data thus generated can then be modelled by a suitable positively skewed distribution.

Case-B: It is difficult to have lifetime data on the whole system.

In many real life cases it is not possible to generate lifetime data on the whole system at an acceptable cost. But it is possible to test individual components and to generate lifetime data on these individual components. If one knows how the individual components are connected (or placed) to make the whole system work, then individual component data can be used (along with the structural knowledge of the whole system) to study the lifetime of the whole system.

As an example of this case, consider the problem of studying the lifetime of a communication satellite. Various components like battery charger, solar panel, and data transmitter, etc. are integral parts of the satellite, and they must operate to make the whole satellite functional. Instead of launching a satellite and then observing its lifetime to collect a single costly observation, it makes sense to test the component prototypes in a simulated environment at a much lower cost, and then use that information to draw inferences about the satellite’s lifetime.

For notational and other convenience one can unify the above two cases (A and B) theoretically under Case-B. When lifetime data is available on the whole system then just pretend that the whole system consists of a single component (the system itself), and then we can draw inferences just as we do under the above Case-B. On the other hand, under Case-B, think of a system with $k$ components and we have lifetime data on the individual components. When $k = 1$, we are under the above Case-A. Therefore, in the rest of the chapter, we consider systems with $k$ components, assuming that the lifetime data are available on the individual components.
10.2 Single Component Systems

Let the lifetime of a component be represented by a nonnegative random variable $X$ with pdf $f(x)$ and cdf $F(x)$. The reliability of the component at time $t$ is the probability that the component does not fail on or before time $t$, i.e.,

$$
\hat{F}(t) = P(X > t) = \int_t^\infty f(x) \, dx = 1 - P(X \leq t) = 1 - F(t) \quad (10.2.1)
$$

Often, we assume a known structure for $f(x)$, the pdf of lifetime, which may involve one or more unknown parameters. In such a case one can estimate the unknown parameter, and hence the component reliability from a sample of observations on $X$. For example, one can test $n$ prototypes of the component and observe their life spans to obtain the data $X_1, X_2, \ldots, X_n$, where $X_i$ represents the life span of the $i^{th}$ prototype. In the following we discuss inferences on reliability using some of the distributions discussed in earlier chapters.

Exponential Distribution

Example 10.2.1. A random sample of size 8 has been obtained on the lifetime (in hours) of a component with the following observations:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>48.96</td>
<td>54.67</td>
<td>100.75</td>
<td>120.68</td>
</tr>
<tr>
<td>148.34</td>
<td>228.43</td>
<td>293.45</td>
<td>415.82</td>
</tr>
</tbody>
</table>

Assuming an exponential model estimate the component reliability.

To address the above problem we will use both the one-parameter exponential model (i.e., $\text{Exp}(\beta)$ model (6.1.1)) and the two-parameter exponential model (i.e., $\text{Exp}(\beta, \mu)$ model (6.1.5)).

Case I: One Parameter Exponential Model

Step 1: Goodness of fit test (see Section 6.4):

(i) $S = \sum_{i=1}^{n} X_i = 1411.10$

(ii) $Z_j = \sum_{i=1}^{j} X_i / S, \quad 1 \leq j \leq (n - 1)$, values are given respectively as (in increasing order)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.034696</td>
<td>0.073439</td>
<td>0.144837</td>
<td>0.230359</td>
</tr>
<tr>
<td>0.335483</td>
<td>0.497364</td>
<td>0.705322</td>
<td></td>
</tr>
</tbody>
</table>

© 2006 by Taylor & Francis Group, LLC
(iii) \( T^* = -2 \sum_{i=1}^{n-1} \ln Z_i = 23.024823 \)

(iv) \( \chi^2_{2(n-1), (\alpha/2)} \) and \( \chi^2_{2(n-1), 1-(\alpha/2)} \) cut-off points with significance level \( \alpha = 0.05 \) are respectively 26.12 and 5.63. Since \( T^* \) falls in between these two values, at 5% level, the model \( \text{Exp}(\beta) \) seems plausible for the above data set.

**Step 2**: Parameter estimation:

(i) The MLE (also the UMVUE) of the model parameter \( \beta \) is \( \hat{\beta}_{\text{ML}} = 176.3875 \).

(ii) The reliability of a component at time \( t > 0 \) is

\[
\bar{F}(t) = P(X > t) = \exp(-t/\beta) \quad (10.2.2)
\]

(iii) The MLE of \( \bar{F}(t) \) is

\[
\hat{F}_{\text{ML}} = \hat{F}_{\text{ML}}(t|\beta) = \exp(-t/\hat{\beta}_{\text{ML}}) = \exp(-t/176.3875) \quad (10.2.3)
\]

which is biased.

The bias of \( \hat{F}_{\text{ML}} \), which is a function of \( \beta \), is

\[
B(\hat{F}_{\text{ML}}|\beta) = E(\hat{F}_{\text{ML}}) - \bar{F} = \int_0^\infty \left\{ e^{-nt/y} - e^{-t/\beta} \right\} \{\beta^n \Gamma(n)\}^{-1} e^{-x/\beta} x^{n-1} \, dx
\]

\[
= \int_0^\infty \left\{ e^{nt/y} - e^{-t_*} \right\} e^{-y} y^{n-1} / \Gamma(n) \, dy, \quad t_* = t/\beta \quad (10.2.4)
\]

since \( n\hat{\beta}_{\text{ML}} = \sum_{i=1}^n X_i \) follows \( G(n, \beta) \) distribution (with pdf \( (7.1.1) \)). In the present example, \( n = 8 \).

**Figure 10.2.1** gives the plot of \( B(\hat{F}_{\text{ML}}) \) as a function of \( t_* = t/\beta \).

Similarly, the MSE of \( \hat{F}_{\text{ML}} \) is given as

\[
\text{MSE}(\hat{F}_{\text{ML}}|\beta) = E(\hat{F}_{\text{ML}} - \bar{F})^2 = \int_0^\infty \left\{ e^{nt/y} - e^{-t_*} \right\}^2 e^{-y} y^{n-1} / \Gamma(n) \, dy, \quad t_* = t/\beta \quad (10.2.5)
\]

The UMVUE of \( \bar{F}(t) \) is given as

\[
\hat{F}_{\text{U}} = \hat{F}_{\text{U}}(t|\beta) = \begin{cases} 
1 - \left\{ t/(n\hat{\beta}_{\text{ML}}) \right\}^{-1} & \text{if } n\hat{\beta}_{\text{ML}} > t \\
0 & \text{otherwise}
\end{cases} \quad (10.2.6)
\]
10.2 Single Component Systems

Figure 10.2.1: Bias of $\hat{F}_{ML}$ as a function of $t_\ast = t/\beta$ with $n = 8$.

(Note that the expression in Bain (1978) (Theorem 2.2.2, page-124) is incorrect, presumably due to a typographical error.)

The $MSE$ of $\hat{F}_U$ is given as

$$MSE(\hat{F}_U|\beta) = E(\hat{F}_U - \bar{F})^2$$
$$= \int_{t_\ast}^{\infty} (1 - t_\ast/y)^2(n-1)e^{-y}y^{n-1}/\Gamma(n) dy - (e^{-t_\ast})^2,$$
$$t_\ast = t/\beta \quad (10.2.7)$$

The following figure gives the plots of the $MSE$ curves of $\hat{F}_{ML}$ and $\hat{F}_U$ as a function of $t_\ast = t/\beta$. 

© 2006 by Taylor & Francis Group, LLC
In reality, since $\beta$ is unknown, so is $t^*$; and hence the bias and $MSE$ are unknown (except for the zero bias of UMVUE). However, these can be estimated by replacing $t^*$ by $\hat{t}^* = t/\hat{\beta}_{ML}$.

In Example 10.2.1, $\hat{t}^* = t/(176.3875)$. If we want to study the component reliability at time $t = 300$ hours, then $\hat{t}^* = 300/(176.3875) = 1.700800 \approx 1.7$. Then

\[
\begin{align*}
\hat{B}(\hat{F}_{ML}|\beta) &= \text{estimated bias of } \hat{F}_{ML} = -0.00403, \quad \text{and} \\
\hat{MSE}(\hat{F}_{ML}|\beta) &= \text{estimated } MSE \text{ of } \hat{F}_{ML} = 0.00980
\end{align*}
\]

On the other hand, the bias of $\hat{F}_U$ is 0, and

\[
\hat{MSE}(\hat{F}_U|\beta) = \text{estimated } MSE \text{ of } \hat{F}_U = 0.01219
\]
10.2 Single Component Systems

Case II: Two Parameter Exponential Model

Step 1: Goodness of fit test (see Section 6.4):

(i) First we obtain the ‘spacings’ defined as \( Y_i = (n+i-1)(X_i - X_{(i-1)}) \) for \( i = 2, 3, \ldots, n \). Hence

\[
Y_2 = 39.97, \quad Y_3 = 276.48, \quad Y_4 = 99.65, \quad Y_5 = 110.64
\]

\[
Y_6 = 240.27, \quad Y_7 = 130.04 \quad \text{and} \quad Y_8 = 122.37
\]

(ii) \( S^* = \sum_{i=2}^{n} Y_i = 1019.42 \)

(iii) \( Z_{j*} = \sum_{i=2}^{j} Y_i / S^*, \ 2 \leq j \leq (n-1) \), values are given respectively as (in increasing order)

0.039209, 0.310422, 0.408173, 0.516709, 0.752398, 0.879961

(iv) \( T^{**} = -2 \sum_{i=2}^{(n-1)} \ln Z_i^{**} = 12.754769 \)

(v) \( \chi^2_{2(n-2), (\alpha/2)} \) and \( \chi^2_{2(n-2), (1-\alpha/2)} \) cut off points with \( \alpha = 0.05 \) are respectively 23.34 and 4.40. Since \( T^{**} \) falls between these two values, at 5% level, the model \( \text{Exp}(\beta, \mu) \) seems plausible for the above data set.

Step 2: Parameter estimation:

(i) The \( MLEs \) of the model parameters \( \mu \) and \( \beta \) are \( \hat{\mu}_{ML} = X_{(1)} = \min\{X_1, \ldots, X_n\} = 48.96 \) and \( \hat{\beta}_{ML} = \sum_{i=1}^{n} (X_i - X_{(1)}) / n = 127.4275 \).

(ii) The reliability of a component at time \( t(> \mu) \) is

\[
\tilde{F}(t) = P(X > t) = P(X - \mu > t - \mu) = \exp(-(t - \mu) / \beta) \quad (10.2.8)
\]

(iii) Therefore, the \( MLE \) of \( \tilde{F}(t) \) is

\[
\hat{\tilde{F}}_{ML}(t|\beta, \mu) = \left\{ \begin{array}{ll}
\exp(-(t - \hat{\mu}_{ML}) / \hat{\beta}_{ML}) & \text{if } t > \hat{\mu}_{ML} \\
1 & \text{if } t \leq \hat{\mu}_{ML}
\end{array} \right.
\]

\[
= \left\{ \begin{array}{ll}
\exp(-(t - 48.96) / 127.4275) & \text{if } t > 48.96 \\
1 & \text{if } t \leq 48.96
\end{array} \right.
\quad (10.2.9)
\]

© 2006 by Taylor & Francis Group, LLC
It can be shown that the probability distribution of \( \hat{F}_{ML} \) depends on the term \((t - \mu)/\beta\), and hence the bias and MSE of \( \hat{F}_{ML} \) are functions of \((t - \mu)/\beta\) only where \((t > \mu)\).

To study the bias and MSE of \( \hat{F}_{ML} \) (see 10.2.9), it would be helpful to look at the following representation. Define \( t^* = (t - \mu)/\beta \). Then \( \hat{F}_{ML} \) can be characterized as

\[
\hat{F}_{ML} = \begin{cases} 
\exp\left\{(-nt^* + W_1)/W_2\right\} & \text{if } t^* > W_1 \\
1 & \text{if } t^* < W_1 
\end{cases}
\] (10.2.10)

where \( W_1 \) and \( W_2 \) are independent random variables with \( W_1 \sim \text{Exp}(1) \) and \( W_2 \sim G(n-1, 1) \). Hence,

\[
B(\hat{F}_{ML}|\beta, \mu) = E(\hat{F}_{ML} - \bar{F})
\]

\[
= \int_0^\infty \int_0^\infty \left[ \exp\left\{(-nt^* + w_1)/w_2\right\} - \exp(-t^*) \right] \times e^{-w_1}e^{-w_2}w_2^{n-2}/\Gamma(n-1) \, dw_1 dw_2
\]

\[
+ \int_0^\infty \int_{t^*}^\infty \left[ 1 - \exp(-t^*) \right] \times e^{-w_1}e^{-w_2}w_2^{n-2}/\Gamma(n-1) \, dw_1 dw_2
\]

The Figure 10.2.3 gives the plot of \( B(\hat{F}_{ML}|\beta, \mu) \) as a function of \( t^* = (t - \mu)/\beta > 0 \).

Similarly, the MSE of \( \hat{F}_{ML} \) is given as

\[
MSE(\hat{F}_{ML}|\beta, \mu) = E((\hat{F}_{ML} - \bar{F})^2)
\]

\[
= \int_0^\infty \int_0^\infty \left[ \exp\left\{(-nt^* + w_1)/w_2\right\} - \exp(-t^*) \right]^2 \times e^{-w_1}e^{-w_2}w_2^{n-2}/\Gamma(n-1) \, dw_1 dw_2
\]

\[
+ \int_0^\infty \int_{t^*}^\infty \left[ 1 - \exp(-t^*) \right]^2 \times e^{-w_1}e^{-w_2}w_2^{n-2}/\Gamma(n-1) \, dw_1 dw_2
\]

(10.2.12)

An analogue to the estimator (10.1.6) for estimating \( \bar{F}(t) \) (for the two parameter exponential model) is the unbiased estimator given as

\[
\hat{F}_U = \hat{F}_U(t)
\]

\[
= \begin{cases} 
0 & \text{if } t > n\hat{\beta}_{ML} + \hat{\mu}_{ML} \\
(1 - 1/n)\left[1 - \left\{(t - \hat{\mu}_{ML})/(n\hat{\beta}_{ML})\right\}\right]^{n-2} & \text{if } \hat{\mu}_{ML} \leq t < n\hat{\beta}_{ML} + \hat{\mu}_{ML} \\
1 & \text{if } t < \hat{\mu}_{ML}
\end{cases}
\]
Figure 10.2.3: Bias of $\hat{F}_{ML}(t)$ as a function of $t_{**} = (t - \mu)/\beta$ (n = 8).

The $MSE$ of $\hat{F}_U$ can be obtained through the following expression. Define $g(w_1, w_2|t_{**})$ as

$$g(w_1, w_2|t_{**}) = \begin{cases} 
0 & \text{if } t_{**} > w_2 + \frac{w_1}{n} \\
(1 - 1/n)(1 - (t_{**} - \frac{w_1}{n})w_2)^{n-2} & \text{if } \frac{w_1}{n} \leq t_{**} \leq w_2 + \frac{w_1}{n} \\
1 & \text{if } t_{**} < \frac{w_1}{n}
\end{cases}$$

(10.2.13)

Then the $MSE$ of $\hat{F}_U$ is

$$MSE(\hat{F}_U|\beta, \mu) = \int_0^{\infty} \int_0^{\infty} [g(w_1, w_2|t_{**}) - \exp(-t_{**})]^2 \times e^{-w_1} e^{-w_2} w_2^{n-2}/\Gamma(n - 1) \, dw_1 dw_2$$

(10.2.14)
The Figure 10.2.4 gives the plots of the \( MSE \) curves of \( \hat{F}_{ML} \) and \( \hat{F}_{U} \) as functions of \( t^* = (t - \mu)/\beta \).

![Plot of MSE for ML and U estimators](image)

Figure 10.2.4: MSEs of \( \hat{F}_{ML}(t) \) and \( \hat{F}_{U}(t) \) as functions of \( t^* = (t - \mu)/\beta \) \( (n = 8) \).

**Remark 10.2.1** Estimating the reliability of a component with exponential lifetime and studying the sampling distribution of these estimators is relatively easy. But this may not be so if the component lifetime follows a more general distribution as shown below. In such a case, Monte-Carlo simulation studies come handy in studying the reliability estimators.

**Gamma Distribution**

**Example 10.2.2.** Consider the data set of Example 7.1.1 where it is assumed that a gamma distribution might be suitable to model the
10.2 Single Component Systems

lifetime (in hours) of pressure vessels. For convenience the data set has been reproduced below.

\[
\begin{align*}
274.00 & & 1.70 & & 871.00 & & 1311.00 & & 236.00 \\
458.00 & & 54.90 & & 1787.00 & & 0.75 & & 776.00 \\
28.50 & & 20.80 & & 363.00 & & 1661.00 & & 828.00 \\
290.00 & & 175.00 & & 970.00 & & 1278.00 & & 126.00
\end{align*}
\]

We shall estimate the reliability of a pressure vessel at time \( t \) using both the two-parameter gamma model (i.e. \( G(\alpha, \beta) \) model (7.1.1)) as well as the three-parameter gamma model (i.e., \( G(\alpha, \beta, \mu) \) model (7.1.9)).

**Case I: Two-Parameter Gamma Model**

Denoting the observations as \( X_1, \ldots, X_n \) (\( n = 20 \)), the maximum likelihood estimates of \( \alpha \) and \( \beta \) are found by solving the equations \( \hat{\alpha}_{ML} \hat{\beta}_{ML} = \bar{X} \) and \( ln(\hat{\beta}_{ML}) + \psi(\hat{\alpha}_{ML}) = (\sum_{i=1}^{n} lnX_i)/n \), where \( \bar{X} = \sum_{i=1}^{n} X_i/n \) and \( \psi(\alpha) = \frac{\partial}{\partial \alpha} \ln \Gamma(\alpha) \); and these are

\[
\hat{\alpha}_{ML} = 0.579182; \quad \hat{\beta}_{ML} = 993.699 \quad (10.2.15)
\]

The reliability of a pressure vessel at time \( t(>0) \) is

\[
\hat{F} = \bar{F}(t|\alpha, \beta) = P(X > t) = \int_{t}^{\infty} \exp(-x/\beta)x^{\alpha-1}/(\beta^{\alpha}\Gamma(\alpha)) \, dx \quad (10.2.16)
\]

which does not have any further closed expression. The MLE of \( \bar{F} \) is

\[
\hat{\bar{F}}_{ML} = \bar{F}(t|\hat{\alpha}_{ML}, \hat{\beta}_{ML}) = \int_{t}^{\infty} \exp(-x/\hat{\beta}_{ML})x^{\hat{\alpha}_{ML}-1}/(\hat{\beta}_{ML}^{\hat{\alpha}_{ML}}\Gamma(\hat{\alpha}_{ML})) \, dx \quad (10.2.17)
\]

The sampling distribution of \( \hat{\bar{F}}_{ML} \) is impossible to study analytically for fixed sample sizes. For the data set in the above Example 10.2.2, with \( t = 300 \) hours, \( \hat{\alpha}_{ML} = 0.579182 \), and \( \hat{\beta}_{ML} = 993.699 \),

\[
\hat{\bar{F}}_{ML} = \frac{1}{(993.699)^{0.579182}\Gamma(0.579182)} \times \int_{300}^{\infty} \exp(-x/993.699)x^{0.579182-1} \, dx = 0.496069 \quad (10.2.18)
\]
The Figure 10.2.4 plots $\hat{F}_{ML}$ (in (10.2.18)) against $t > 0$ for the above data set.

In general, one can use the Monte-Carlo simulation to study the behavior of the estimator $\hat{F}_{ML}$ as described in Chapter 5 and given in the following.

For the sake of simplicity we generate data from a gamma distribution with $\alpha > 0$ and $\beta = 1$. Note, $\hat{F} = \int_{t}^{\infty} \exp(-x/\beta) x^{\alpha-1}/(\beta^{\alpha} \Gamma(\alpha)) \, dx$.

**Step 1**: Generate $X_1, \ldots, X_n$ iid from $G(\alpha, \beta)$ with a specific $\alpha > 0$ and $\beta = 1$. Fix $t > 0$.

**Step 2**: Calculate $\hat{\alpha}_{ML}, \hat{\beta}_{ML}, \hat{F}_{ML}$.

**Step 3**: The error in $\hat{F}$ estimation is calculated as

$$E_{ML} = (\hat{F}_{ML} - \hat{F})$$
Step 4: Repeat the above Step 1 through Step 3 for a large number of (say $M$) times. The resultant values of $E_{ML}$ for these replications are termed as $E_{ML}^{(1)}, \ldots, E_{ML}^{(M)}$.

Step 5: The bias and MSE of $\hat{F}_{ML}$ are approximated as

$$B(\hat{F}_{ML}\mid \alpha, \beta) \approx \frac{1}{M} \sum_{j=1}^{M} E_{ML}^{(j)}$$

and

$$MSE(\hat{F}_{ML}\mid \alpha, \beta) \approx \frac{1}{M} \sum_{j=1}^{M} (E_{ML}^{(j)})^2$$

One can carry out such a simulation for different combinations of $(\alpha, \beta)$ and $t$.

The Figure 10.2.6 (a) to Figure 10.2.6 (b), Figure 10.2.7 (a) to Figure 10.2.7 (b) give the bias and MSE (computed through the above simulation steps) curves of $\hat{F}_{ML}$ plotted against $\alpha$ (for fixed $\beta = 1$ and $t > 0$) and $t$ (for fixed $\alpha$ and $\beta$).

![Figure 10.2.6: (a) Simulated bias of $\hat{F}_{ML}$ plotted against $\alpha$ ($\beta = 1, t = 1$).](image)

The above mentioned figures, though they provide an idea about the sampling distribution of $\hat{F}_{ML}$, can do little for a specific data set. Since the parameters $\alpha, \beta$ are unknown, we can only plot the bias and MSE curves of $\hat{F}_{ML}$ for limited combinations of $(\alpha, \beta)$. A more realistic approach would
Figure 10.2.6: (b) Simulated MSE of $\hat{F}_{ML}$ plotted against $\alpha$ ($\beta = 1, t = 1$).

Figure 10.2.7: (a) Simulated bias of $\hat{F}_{ML}$ plotted against $t > 0$ ($\alpha = 5, \beta = 1$).

be to follow the bootstrap method as described in Chapter 5 to study the bias and $MSE$ of $\hat{F}_{ML}$. This is demonstrated below with the data set in Example 10.2.2.
Semiparametric Bootstrap Method

Fix the observed sample \(\{X_1, \ldots, X_n\}\), call it ‘original sample’.

**Step 1**: Draw a sample of size \(n\) with replacement from the original sample, call it a ‘bootstrap sample’.

**Step 2**: Calculate \(\hat{\alpha}_{ML}, \hat{\beta}_{ML},\) and \(\hat{F}_{ML}\) based on the bootstrap sample, and call them \(\hat{\alpha}^*_ML, \hat{\beta}^*_ML,\) and \(\hat{F}^*_ML\) respectively.

**Step 3**: The observed error in \(\hat{F}_{ML}\) for the new sample is calculated as \(E^*_ML = (\hat{F}^*_ML - \hat{F}_{ML})\).

**Step 4**: Repeat the above Step 1 through Step 3 for a large number of (say, \(M\)) times. The resultant values of \(E^*_ML\) for these replications are termed as \(E^*_ML(1), \ldots, E^*_ML(M)\).

**Step 5**: The estimated bias (called, semiparametric bootstrap bias) and the estimated \(MSE\) (called, semiparametric bootstrap \(MSE\)) of \(\hat{F}_{ML}\) are obtained as (abbreviated as ‘SB-Bias’ and ‘SB-\(MSE\)’ respectively)

\[
SB - Bias(\hat{F}_{ML}|\alpha, \beta) = \frac{1}{M} \sum_{j=1}^{M} E^*_ML(j)
\]
and

\[ SB - MSE(\hat{F}_{ML}|\alpha, \beta) = \frac{1}{M} \sum_{j=1}^{M} (E_{ML}^{*j})^2 \]

Using \( M = 20,000 \), the following SB-bias and SB-MSE curves have been generated for the data set in Example 10.2.2.

![Figure 10.2.8: (a) SB-Bias of \( \hat{F}_{ML} \) (Example 10.2.2).](image)

**Remark 10.2.2** (a) A true nonparametric bootstrap method does not require any specific probability model (parametric) for the data set. Yet we have assumed the \( G(\alpha, \beta) \) model for the data in Example 10.2.2 to begin with. Then we have followed the bootstrap approach to study the behavior of the parameter estimates \( \hat{\alpha}_{ML} \) and \( \hat{\beta}_{ML} \), and hence we call it a semiparametric bootstrap (SB) method.

(b) It has been observed that the semiparametric (or nonparametric) bootstrap method can give satisfactory results for moderate to large sample sizes \( n \). For small \( n \), such bootstrap results (bias and MSE) tend to be very unstable (fluctuate too much from one set of replications to another). Hence, a parametric bootstrap is suggested as an extra tool to evaluate an estimator.
10.2 Single Component Systems

Figure 10.2.8: (b) SB-MSE of \( \hat{F}_{ML} \) (Example 10.2.2).

**Parametric Bootstrap Method**

For the observed sample \( \{X_1, \ldots, X_n\} \) (original sample), calculate \( \hat{\alpha}_{ML}, \hat{\beta}_{ML}, \) and \( \hat{F}_{ML} \).

**Step 1:** Generate a new sample (of size \( n \)) iid from \( G(\hat{\alpha}_{ML}, \hat{\beta}_{ML}) \) (i.e., the \( G(\alpha, \beta) \) distribution with (\( \alpha = \hat{\alpha}_{ML}, \beta = \hat{\beta}_{ML} \)).

**Step 2:** Calculate the MLEs of \( \alpha, \beta, \) and \( F \) based on the new sample, and call them \( \hat{\alpha}^{**}_{ML}, \hat{\beta}^{**}_{ML}, \) and \( \hat{F}_{ML}^{**} \) respectively.

**Step 3:** The observed error in \( \hat{F}_{ML}^{**} \) for the new sample is calculated as \( E_{ML}^{**} = (\hat{F}_{ML}^{**} - \hat{F}_{ML}) \).

**Step 4:** Repeat the above Step 1 through Step 3 for a large number (say, \( M \)) of times. The resultant values of \( E_{ML}^{**} \) for these replications are termed as \( E_{ML}^{**(1)}, \ldots, E_{ML}^{**(M)} \).

**Step 5:** The estimated bias and MSE (called, parametric bootstrap bias and MSE) of \( \hat{F}_{ML} \) are obtained as (and denoted by ‘PB-Bias’ and ‘PB-MSE’ respectively)

\[
P_{B} - \text{Bias} (\hat{F}_{ML} | \alpha, \beta) = \frac{1}{M} \sum_{j=1}^{M} E_{ML}^{**(j)}
\]
and
\[ PB - MSE(\hat{F}_{ML}(\alpha, \beta)) = \frac{1}{M} \sum_{j=1}^{M} (E_{ML}^{**}(j))^2 \]

Using \( M = 20,000 \), the following PB-Bias and PB-MSE curves have been generated for the data set in Example 10.2.2 and these are plotted against \( t > 0 \) in Figure 10.2.9 (a) and Figure 10.2.9 (b).

The other estimators \( \bar{F}(t|\alpha, \beta) \) are obtained from the modified MLEs (called MMLEs) and MMEs of \( \alpha \) and \( \beta \) as discussed in Section 7.3.

Using the reparametrization \( \beta = 1/\lambda \) and (7.3.6), the MMLEs of \( \alpha \) and \( \beta \) are
\[
\hat{\alpha}_{MML} = \left( \frac{n-3}{n} \hat{\alpha}_{ML} + \frac{2}{3n} \right)
\] and
\[
\hat{\beta}_{MML} = \hat{\beta}_{ML} g_0(\hat{\alpha}_{MML})
\]

where \( g_0(\alpha) = \{1 + \frac{3\alpha}{(n-3)(n\alpha-1)}(1 + \frac{1}{n\alpha} - \frac{1}{n\alpha} + \frac{(n-1)}{27n\alpha})\}. \]

On the other hand, the MMEs of \( \alpha \) and \( \beta \) are
\[
\hat{\alpha}_{MM} = \frac{n\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{and} \quad \hat{\beta}_{MM} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n\bar{X}}
\]

Thus, the two new estimators of \( \bar{F}(t|\alpha, \beta) \) are
\[
\hat{F}_{MML} = \bar{F}(t|\hat{\alpha}_{MML}, \hat{\beta}_{MML}) \quad \text{and} \quad \hat{F}_{MM} = \bar{F}(t|\hat{\alpha}_{MM}, \hat{\beta}_{MM})
\]

For the data set in Example 10.2.2, with \( t = 300 \),
\[
\hat{\alpha}_{MML} = 0.5256, \quad \hat{\beta}_{MML} = 1234.7
\] \[
\hat{\alpha}_{MM} = 1.0559, \quad \hat{\beta}_{MM} = 545.061
\]

and
\[
\hat{F}_{MML} = \int_{300}^{\infty} \exp(-x/1234.7)x^{0.5256-1}/(1234.7)^{0.5256}\Gamma(0.5256) \, dx = 0.506012
\] \[
\hat{F}_{MM} = \int_{300}^{\infty} \exp(-x/545.061)x^{1.0559-1}/(545.061)^{1.0031}\Gamma(1.0559) \, dx = 0.603638
\]
In the following (Figure 10.2.10 (a) and Figure 10.2.10 (b), Figure 10.2.11 (a) and Figure 10.2.11 (b)) we provide the bias and \(MSE\) curves of \(\hat{F}_{MML}\) and \(\hat{F}_{MM}\) through simulations for various \(n\).

Remark 10.2.3 Compare the bias and \(MSE\) of \(\hat{F}_M\) in Figure 10.2.6 (a)
and Figure 10.2.6 (b) with those of $\hat{F}_{MML}$ and $\hat{F}_{MM}$ in Figure 10.2.10 (a) and Figure 10.2.10 (b). Apparently, both $\hat{F}_{ML}$ and $\hat{F}_{MML}$ have similar bias (mostly negative), whereas $\hat{F}_{MM}$ has opposite (mostly positive) bias. In term of MSE, $\hat{F}_{MML}$ seems to have an overall best performance. Also note that the MMEs of $\alpha$ and $\beta$ are not functions of the sufficient statistics, and hence $\hat{F}_{MM}$ is inferior to $\hat{F}_{ML}$ or $\hat{F}_{MML}$ intuitively.

![Figure 10.2.10: (a) Simulated Biases of $\hat{F}_{MML}$ and $\hat{F}_{MM}$ plotted against $\alpha$ ($\beta=1, t=1$).](image-url)
Figure 10.2.10: (b) Simulated $MSE$s of $\hat{F}_{\text{MML}}$ and $\hat{F}_{\text{MM}}$ plotted against $\alpha$ ($\beta = 1, t = 1$).

Figure 10.2.11: (a) Simulated Biases of $\hat{F}_{\text{MML}}$ and $\hat{F}_{\text{MM}}$ plotted against $t > 0$ ($\alpha = 5.0, \beta = 1.0$).
Remark 10.2.4  One can also follow the parametric bootstrap approach to get estimated bias and \( \text{MSE} \) of the reliability estimators \( \hat{F}_{\text{MM}} \) and \( \hat{F}_{\text{MML}} \). But one has to be extra cautious in comparing the PB-Bias and PB-MSE curves of various estimators. Note that while SB-Bias and SB-MSE are computed based on resampling from the same ‘original sample’, PB-Bias and PB-MSE are computed based on sampling from different distributions (\( G(\hat{\alpha}_{\text{MML}}, \hat{\beta}_{\text{MML}}) \) and \( G(\hat{\alpha}_{\text{MM}}, \hat{\beta}_{\text{MM}}) \)) and hence comparisons may be criticized.
Case II: Three-parameter gamma model

If we assume a three-parameter gamma model $G(\alpha, \beta, \mu)$ (as given in (7.1.9)) for the data in Example 10.2.2, the reliability of a pressure vessel
at time $t(>\mu)$ is

$$
\bar{F} = \bar{F}(t|\alpha, \beta, \mu) = P(X > t) = \int_t^\infty \exp \left\{-(x - \mu)/\beta \right\} \times (x - \mu)^{\alpha-1}/(\beta^\alpha \Gamma(\alpha)) \, dx
$$

(10.2.22)

The following estimates of $\bar{F}$ are found by employing suitable estimates of the model parameters as

$$
\hat{\bar{F}}_{\text{ML}} = \begin{cases} 
1 & \text{if } t \leq \hat{\mu}_{\text{ML}} \\
\bar{F}(t|\hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}}, \hat{\mu}_{\text{ML}}) & \text{if } t > \hat{\mu}_{\text{ML}}
\end{cases}
$$

(using the MLEs in (10.2.23));

$$
\hat{\bar{F}}_{\text{MML}} = \begin{cases} 
1 & \text{if } t \leq \hat{\mu}_{\text{MML}} \\
\bar{F}(t|\hat{\alpha}_{\text{MML}}, \hat{\beta}_{\text{MML}}, \hat{\mu}_{\text{MML}}) & \text{if } t > \hat{\mu}_{\text{MML}}
\end{cases}
$$

(using the MMLEs in (10.2.23));

$$
\hat{\bar{F}}_{\text{MM}} = \begin{cases} 
1 & \text{if } t \leq \hat{\mu}_{\text{MM}} \\
\bar{F}(t|\hat{\alpha}_{\text{MM}}, \hat{\beta}_{\text{MM}}, \hat{\mu}_{\text{MM}}) & \text{if } t > \hat{\mu}_{\text{MM}}
\end{cases}
$$

(using the MMEs in (10.2.23)).

For the data set in Example 10.2.2, using $t = 300$ hours,

$$
\hat{\bar{F}}_{\text{ML}} = 0.667; \hat{\bar{F}}_{\text{MML}} = 0.644; \hat{\bar{F}}_{\text{MM}} = 0.652 \quad (10.2.23)
$$

Also, the above three estimators have been plotted against $t > 0$ for the above data set in Figure 10.2.13.

Similar to the case of a two-parameter gamma model, simulated bias and MSE of $\hat{\bar{F}}_{\text{ML}}$, $\hat{\bar{F}}_{\text{MML}}$, and $\hat{\bar{F}}_{\text{MM}}$ are plotted against $t > 0$ as given in Figure 10.2.14 (a) and Figure 10.2.14 (b).

The data set in Example 10.2.2 is now used to get the semiparametric bootstrap bias and $MSE$ (i.e., SB-bias and SB-$MSE$) of the estimates $\hat{\bar{F}}_{\text{ML}}$, $\hat{\bar{F}}_{\text{MML}}$, and $\hat{\bar{F}}_{\text{MM}}$ for various values of $t > 0$ (and then plotted against $t$) as given in Figure 10.2.16 (a) and Figure 10.2.16 (b). Again, $\hat{\bar{F}}_{\text{MML}}$ seems to perform best among all the three estimators.
10.2 Single Component Systems

Figure 10.2.13: The $\hat{F}_{ML}$, $\hat{F}_{MML}$, and $\hat{F}_{MM}$ curves for the data set in Example 10.2.2 with the $G(\alpha, \beta, \mu)$ model.

Figure 10.2.14: (a) Simulated Biases of $\hat{F}_{ML}$, $\hat{F}_{MML}$, and $\hat{F}_{MM}$ plotted against $t > 0$ ( $\alpha = 5.0, \beta = 1.0, \mu = 0.0$) for $n = 10$ with a $G(\alpha, \beta, \mu)$ model.
Chapter 10  System Reliability

Figure 10.2.14: (b) Simulated \( \text{MSEs of } \hat{F}_{\text{ML}}, \hat{F}_{\text{MML}}, \text{and } \hat{F}_{\text{MM}} \) plotted against \( t > 0 \) (\( \alpha = 5.0, \beta = 1.0, \mu = 0.0 \)) for \( n = 10 \) with a \( G(\alpha, \beta, \mu) \) model.

Figure 10.2.15: (a) Simulated Biases of \( \hat{F}_{\text{ML}}, \hat{F}_{\text{MML}}, \text{and } \hat{F}_{\text{MM}} \) plotted against \( t > 0 \) (\( \alpha = 5.0, \beta = 1.0, \mu = 0.0 \)) for \( n = 20 \) with a \( G(\alpha, \beta, \mu) \) model.
10.2 Single Component Systems

Weibull Distribution

Apart from exponential and gamma distributions, Weibull distribution is widely used to model lifetime data sets. In the following we study the
estimation of a system reliability using a Weibull distribution.

**Example 10.2.3.** Consider the data set of Example 8.1.1 where it is assumed that a Weibull distribution could be used to model the endurance (in m.o.r.) of ball bearings. The sample observations from 23 ball bearings are:

\[
\begin{array}{cccccccccc}
17.88 & 41.52 & 48.48 & 54.12 & 68.64 & 84.12 & 105.12 & 128.04 \\
28.92 & 42.12 & 51.84 & 55.56 & 68.64 & 93.12 & 105.84 & 173.40 \\
33.00 & 45.60 & 51.96 & 67.80 & 68.88 & 98.64 & 127.92
\end{array}
\]

We shall estimate the reliability of a ball bearing at \( t \) m.o.r. using both a two-parameter Weibull model (i.e., \( W(\alpha, \beta) \) model (8.1.1)) as well as a three-parameter Weibull model (i.e., \( W(\alpha, \beta, \mu) \) model (8.1.8)). We will see that in terms of simpler closed expression of the cdf Weibull model has certain advantages over a gamma model.

**Case I: Two-parameter Weibull model**

With observations \( X_1, X_2, \ldots, X_n \) \( (n = 23 \text{ in Example 10.2.3}) \), the maximum likelihood estimates of \( \alpha \) and \( \beta \) are found by solving the equa-
10.2 Single Component Systems

\[
\hat{\beta}_{ML} = \left\{ \sum_{i=1}^{n} X_i^{\hat{\alpha}_{ML}} / n \right\}^{1/\hat{\alpha}_{ML}}
\]

\[
\hat{\alpha}_{ML} = \left[ \left\{ \sum_{i=1}^{n} X_i^{\hat{\alpha}_{ML}} \ln X_i \right\} / \left\{ \sum_{i=1}^{n} X_i^{\hat{\alpha}_{ML}} \right\} - \left\{ \sum_{i=1}^{n} \ln X_i / n \right\} \right]^{-1}
\]

For the above data set, we have

\[
\hat{\alpha}_{ML} = 2.10 \quad \text{and} \quad \hat{\beta}_{ML} = 81.86 \quad (10.2.24)
\]

The reliability at time (or m.o.r. for the above data set) \( t > 0 \) is

\[
\hat{F} = \hat{F}(t|\hat{\alpha}, \hat{\beta}) = P(X > t) = \exp\{-t/\hat{\beta}\} \quad (10.2.25)
\]

Thus, the MLE of \( \hat{F} \) is

\[
\hat{F}_{ML} = \hat{F}(t|\hat{\alpha}_{ML}, \hat{\beta}_{ML}) = \exp\{-t/\hat{\beta}_{ML}\} \quad (10.2.26)
\]

which for the data set in Example 10.2.3, with \( t = 100 \) m.o.r., yields

\[
\hat{F}_{ML} = 0.21817 \quad (10.2.27)
\]

The graph of \( \hat{F}_{ML} \) is plotted against \( t > 0 \) in the following figure.

Figure 10.2.17: The \( \hat{F}_{ML} \) curve for the data set in Example 10.2.3.
On the other hand, the method of moments estimates of $\alpha$ and $\beta$ are obtained from

$$\left( \frac{\Gamma \left( 1 + \frac{1}{\alpha_{\text{MM}}} \right)}{\Gamma \left( 1 + \frac{2}{\alpha_{\text{MM}}} \right)} \right)^2 = \frac{n \bar{X}^2}{\sum_{i=1}^{n} X_i^2} \quad \text{and} \quad \hat{\beta}_{\text{MM}} = \frac{\bar{X}}{\Gamma \left( 1 + \frac{1}{\alpha_{\text{MM}}} \right)}$$

which for the above data set (Example 10.2.3) gives

$$\hat{\alpha}_{\text{MM}} = 2.066, \quad \hat{\beta}_{\text{MM}} = 81.5342$$

The resultant estimator of $\bar{F}$ is

$$\hat{\bar{F}}_{\text{MM}} = \bar{F}(t|\hat{\alpha}_{\text{MM}}, \hat{\beta}_{\text{MM}}) = \exp \left\{ -(t/81.5342)^{2.066} \right\} \quad (10.2.28)$$

![Figure 10.2.18: The $\hat{\bar{F}}_{\text{MM}}$ curve for the data set in Example 10.2.3.](image)

To study how the estimates $\hat{\bar{F}}_{\text{ML}}$ and $\hat{\bar{F}}_{\text{MM}}$ perform under the above mentioned $W(\alpha, \beta)$ model, we carry out a simulation study as described below.

**Step-1 :** Generate $X_1, X_2, \ldots, X_n$ *i.i.d* from $W(\alpha, \beta)$ with a specific $\alpha > 0$ and $\beta = 1$. Fix $t > 0$. 

© 2006 by Taylor & Francis Group, LLC
**Step-2**: Calculate $\hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}}, \hat{F}_{\text{ML}}$ and $\hat{\alpha}_{\text{MM}}, \hat{\beta}_{\text{MM}}, \hat{F}_{\text{MM}}$.

**Step-3**: The error in estimation by each method is calculated as

$$E_{\text{ML}} = (\hat{F}_{\text{ML}} - \bar{F}) \quad \text{and} \quad E_{\text{MM}} = (\hat{F}_{\text{MM}} - \bar{F})$$

**Step-4**: Repeat the above Step 1 through Step 3 for a large number of (say $M$) times. The resultant values of $E_{\text{ML}}$ and $E_{\text{MM}}$ for these replications are termed as $E_{\text{ML}}^{(1)}, \ldots, E_{\text{ML}}^{(M)}$ and $E_{\text{MM}}^{(1)}, \ldots, E_{\text{MM}}^{(M)}$, respectively.

**Step-5**: The bias and MSE of $\hat{F}_{\text{ML}}$ and $\hat{F}_{\text{MM}}$ are approximated as

$$B(\hat{F}_{\text{ML}}|\alpha, \beta) \approx \frac{1}{M} \sum_{j=1}^{M} E_{\text{ML}}^{(j)}; \quad B(\hat{F}_{\text{MM}}|\alpha, \beta) \approx \frac{1}{M} \sum_{j=1}^{M} E_{\text{MM}}^{(j)}$$

and

$$\text{MSE}(\hat{F}_{\text{ML}}|\alpha, \beta) \approx \frac{1}{M} \sum_{j=1}^{M} (E_{\text{ML}}^{(j)})^2; \quad \text{MSE}(\hat{F}_{\text{MM}}|\alpha, \beta) \approx \frac{1}{M} \sum_{j=1}^{M} (E_{\text{MM}}^{(j)})^2$$

(10.2.29)

First we plot the bias curves against $\alpha$ for $\beta = 1.0$ and $t = 1$, and then against $t$ for $\beta = 1.0$ and $\alpha = 5.0$. These are shown in Figure 10.2.19 (a) and Figure 10.2.19 (b).

Similarly the MSE curves are plotted in Figure 10.2.20 (a) and Figure 10.2.20 (b).
Figure 10.2.19: (a) Simulated Biases of $\hat{F}_{\text{ML}}$ and $\hat{F}_{\text{MM}}$ plotted against $\alpha$ with $n = 10$ and 20.

Following the steps of semiparametric bootstrap discussed earlier we now study $\hat{F}_{\text{ML}}$ and $\hat{F}_{\text{MM}}$ for the data set in Example 10.2.3.

**Semiparametric Bootstrap Method**

Fix the observe sample $\{X_1, \ldots, X_n\}$ with $n = 23$. This is our ‘original sample’.

**Step-1**: Draw a sample of size $n$ with replacement from the original sample, and call it a ‘bootstrap sample’.

**Step-2**: Calculate $\hat{\alpha}_{\text{ML}}$, $\hat{\beta}_{\text{ML}}$, and $\hat{F}_{\text{ML}}$ based on the bootstrap sample, and call them $\hat{\alpha}^*_\text{ML}$, $\hat{\beta}^*_\text{ML}$, and $\hat{F}^*_\text{ML}$, respectively. Similarly get $\hat{\alpha}^*_\text{MM}$, $\hat{\beta}^*_\text{MM}$, and $\hat{F}^*_\text{MM}$.

**Step-3**: The observed error in $\hat{F}_{\text{ML}}$ for the bootstrap sample is $E^*_{\text{ML}} = (\hat{F}^*_\text{ML} - \hat{F}_{\text{ML}})$. Similarly $E^*_{\text{MM}} = (\hat{F}^*_\text{MM} - \hat{F}_{\text{MM}})$.

**Step-4**: Repeat the above Step 1 through Step 3 for a large number of (say $M$) times. The resultant values of $E^*_{\text{ML}}$ for these replicated values are $E^*_{\text{ML}}^{(1)}, \ldots, E^*_{\text{ML}}^{(M)}$. Similarly get $E^*_{\text{MM}}^{(1)}, \ldots, E^*_{\text{MM}}^{(M)}$. 

© 2006 by Taylor & Francis Group, LLC
10.2 Single Component Systems

Figure 10.2.19: (b) Simulated Biases of $\hat{F}_{\text{ML}}$ and $\hat{F}_{\text{MM}}$ plotted against $t$ with $n = 10$ and 20.

Figure 10.2.20: (a) Simulated MSEs of $\hat{F}_{\text{ML}}$ and $\hat{F}_{\text{MM}}$ plotted against $\alpha$ ($\beta = 1$, $t = 1.0$) with $n = 10$ and 20.
Figure 10.2.20: (b) Simulated MSEs of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ plotted against $t > 0$ ($\alpha = 5.0$, $\beta = 1.0$) with $n = 10$ and 20.

**Step-5**: The semiparametric bootstrap bias and MSE are

$$
SB - Bias(\hat{F}_{ML}|\alpha, \beta) = \frac{1}{M} \sum_{j=1}^{M} E^{*}(j)
$$

and

$$
SB - MSE(\hat{F}_{ML}|\alpha, \beta) = \frac{1}{M} \sum_{j=1}^{M} (E^{*(j)})^2
$$

Similarly get SB-Bias ($\hat{F}_{MM}|\alpha, \beta$)) and SB-MSE ($\hat{F}_{MM}|\alpha, \beta$)).

The semiparametric bias and MSE of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ are plotted against $t$ as shown in Figure 10.2.21 (a) and Figure 10.2.21 (b).

**Remark 10.2.5** From the above simulation studies it is clear that overall performance of $\hat{F}_{ML}$ is better than that of $\hat{F}_{MM}$. This is not surprising for small to moderate sample sizes since the MLE $\hat{F}_{ML}$ of $\hat{F} = \hat{F}(t|\alpha, \beta)$ depends on $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ which in turn depend on the data through the minimal sufficient statistics thereby ensuring that no information is lost. On the other hand, the MME $\hat{F}_{MM}$ of $\hat{F} = \hat{F}(t|\alpha, \beta)$ depends only on the sample moments (through $\hat{\alpha}_{MM}$ and $\hat{\beta}_{MM}$) which may not retain all the information necessary for $(\alpha, \beta)$ and hence for $\hat{F}$. However, for large
10.2 Single Component Systems

Figure 10.2.21: (a) SB-Biases of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ (Example 10.2.3).

sample sizes this gap between the information in the data and the information used through sample moments diminishes, and as a result $\hat{F}_{MM}$ can perform as good as $\hat{F}_{ML}$.

**Parametric bootstrap method**

In the following we provide the parametric bootstrap bias and MSE of $\hat{F}_{ML}$ for the data set in Example 10.2.3.

**Step-1**: Obtain $\hat{\alpha}_{ML} = 2.10$ and $\hat{\beta}_{ML} = 81.86$ from the original data set of size $n = 23$. Also, $\hat{F}_{ML} = \hat{F}(t|\hat{\alpha}_{ML}, \hat{\beta}_{ML})$.

**Step-2**: Generate new data \{\(X_{1}^{**}, \ldots, X_{n}^{**}\)\} \((n = 23) iid\) from $W(2.10, 81.86)$. Based on this new data recalculate $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$, and call them $\hat{\alpha}_{ML}^{**}$ and $\hat{\beta}_{ML}^{**}$, respectively. Then $\hat{F}_{ML}^{**} = \hat{F}(t|\hat{\alpha}_{ML}^{**}, \hat{\beta}_{ML}^{**})$.

**Step-3**: The observed error in estimating $\bar{F}$ by $\hat{F}_{ML}$ for the new sample is $E_{ML}^{**} = (\hat{F}_{ML}^{**} - \hat{F}_{ML})$.

**Step-4**: Repeat the above Step 2 through Step 3 a large number of (say $M$) times. The resultant values of $E_{ML}^{**}$ for these replications are $E_{ML}^{**(1)}, \ldots, E_{ML}^{**(M)}$.

© 2006 by Taylor & Francis Group, LLC
Figure 10.2.21: (b) SB-MSEs of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ (Example 10.2.3).

**Step-5**: The parametric bootstrap bias (‘PB-bias’) and MSE (‘PB-MSE’) of $\hat{F}_{ML}$ are

$$PB - Bias(\hat{F}_{ML}|\alpha, \beta) = \frac{1}{M} \sum_{j=1}^{M} E_{ML}^{**(j)}$$

and

$$PB - MSE(\hat{F}_{ML}|\alpha, \beta) = \frac{1}{M} \sum_{j=1}^{M} (E_{ML}^{**(j)})^2$$

Using $M = 20,000$, PB-Bias and PB-MSE are plotted in Figure 10.2.22 (a) and Figure 10.2.22 (b) against $t$.

**Case-II**: Three-Parameter Weibull Model

We now assume a three-parameter Weibull model $W(\alpha, \beta, \mu)$ (as given in (8.1.8)) for the data in Example 10.2.3. The reliability of ball-bearing at $t$ m.o.r. ($t > \mu$) is (from (8.1.9))

$$\bar{F} = \bar{F}(t|\alpha, \beta, \mu) = P(X > t) = \exp\left\{-((x - \mu)/\beta)^\alpha\right\} \quad (10.2.30)$$
10.2 Single Component Systems

Figure 10.2.22: (a) PB-Bias of $\hat{F}_{ML}$ (Example 10.2.3).

Figure 10.2.22: (b) PB-MSE of $\hat{F}_{ML}$ (Example 10.2.3).
The following estimates of $\bar{F}$ are obtained by using suitable estimates of $\alpha$, $\beta$, and $\mu$, and these are plotted in Figure 10.2.23.

\[
\hat{F}_{ML} = \begin{cases} 
1 & \text{if } t \leq \hat{\mu}_{ML} \\
\bar{F}(t \mid \hat{\alpha}_{ML}, \hat{\beta}_{ML}, \hat{\mu}_{ML}) & \text{if } t > \hat{\mu}_{ML}
\end{cases}
\]

\[
\hat{F}_{MM} = \begin{cases} 
1 & \text{if } t \leq \hat{\mu}_{MM} \\
\bar{F}(t \mid \hat{\alpha}_{MM}, \hat{\beta}_{MM}, \hat{\mu}_{MM}) & \text{if } t > \hat{\mu}_{MM}
\end{cases}
\]

![Graph showing \(\hat{F}_{ML}\) and \(\hat{F}_{MM}\) plotted against \(t\).](image)

Figure 10.2.23: $\hat{F}_{ML}$ and $\hat{F}_{MM}$ plotted against $t$ (Example 10.2.3).

Similar to the case of two-parameter Weibull model, simulated bias and MSE of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ are plotted against $t$ for fixed $\alpha$, $\beta$, $\mu$ (in Figure 10.2.24 (a) and Figure 10.2.25 (a)), and against $\alpha$ for fixed $t$, $\beta$, $\mu$ (in Figures 10.2.24 (b) and Figure 10.2.25 (b)).

**Remark 10.2.6** From our extensive simulation study it appears that $\hat{F}_{ML}$ has overall better performance than $\hat{F}_{MM}$ for estimating $\bar{F}$ using both a two-parameter as well as a three-parameter Weibull model.
10.2 Single Component Systems

Figure 10.2.24: (a) Simulated Biases of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ plotted against $t > \mu$ ($\alpha = 5.0$, $\beta = 1.0$, $\mu = 0.0$).

Figure 10.2.24: (b) Simulated Biases of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ plotted against $\alpha$ ($t = 1.0$, $\beta = 1.0$, $\mu = 0.0$).
Figure 10.2.25: (a) Simulated MSEs of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ plotted against $t > \mu$ ($\alpha = 5.0$, $\beta = 1.0$, $\mu = 0.0$).

Figure 10.2.25: (b) Simulated MSEs of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ plotted against $t > \mu$ ($\alpha = 5.0$, $\beta = 1.0$, $\mu = 0.0$).
Remark 10.2.7 Again, for the data set in Example 10.2.3, one can carry out bootstrap bias and MSE estimation of $\hat{F}_{ML}$ and $\hat{F}_{MM}$ for a three-parameter Weibull model, but we skip it for brevity.

10.3 Reliability of a Series System with Componentwise Data

Consider a system which consists of $k$ independent components, say $C_1, C_2, \ldots, C_k$, connected in series. The whole system works as long as all components work, and the whole system fails as soon as any one of the components fails.

Let the lifetime of the component $C_i$ ($1 \leq i \leq k$) follow a probability distribution, independent of the other components, which can be modelled by a pdf $f_i(\cdot)$ with corresponding cdf $F_i(\cdot)$; i.e., if $X_i (\geq 0)$ represents the lifetime of $C_i$, then

$$P(X_i \leq t) = \int_{0}^{t} f_i(x) \, dx = F_i(t)$$

i.e., $P(X_i > t) = 1 - F_i(t) = \bar{F}_i(t)$ (10.3.1)

The series system fails as soon as any one of the components fails, and hence the reliability of the series system at time $t$, denoted by $\bar{F}_s(t)$, is

$$\bar{F}_s(t) = P(\text{Series System life} > t)$$

$$= P(X_1 > t, \ldots, X_k > t)$$ (10.3.2)

On the assumption that the failure time distributions of the components are independent,

$$\bar{F}_s(t) = P(X_1 > t) \ldots P(X_k > t) = \prod_{i=1}^{k} \bar{F}_i(t)$$ (10.3.3)
where $\bar{F}_i(t)$ is the reliability of $C - i$ at time $t$. Note that:

(i) Since each $\bar{F}_i(t) \leq 1$, $\bar{F}_s(t) \leq \min_{1 \leq i \leq k} \bar{F}_i(t)$; i.e, the reliability of a series system is much smaller than the reliability of the individual components

(ii) If the components have identical probability distribution then $\bar{F}_s(t) = (\bar{F}(t))^k$, where $\bar{F}(\cdot)$ is the common reliability function of the individual components

From the expression (10.3.3) it is clear that if each $\bar{F}_i(t)$ is not simple (or mathematically trackable), then $\bar{F}_s(t)$ gets even more complicated. This is probably one reason why most of the published research has focused on on $\bar{F}_s(t)$ with each $X_i$ following an exponential distribution. It is also customary to assume that the distributions of $X_i$’s belong to the same family of probability distributions. In the following we will study $\bar{F}_s(t)$ for a few families of distributions for individual $X_i$’s.

As mentioned earlier (in Section 10.1), we assume that we have data on the lifetime of each component (instead of the whole system); i.e., we have

$$X_{i1}, \ldots, X_{in_i}, \text{ iid with pdf } f_i(x), 1 \leq i \leq k$$  \hspace{1cm} (10.3.4)

where $n = n_1 + \ldots + n_k = \text{total sample size}$.

### Components Following an Exponential Distribution

If we assume that each $X_i$ follows an $\text{Exp}(\beta_i)$ distribution, then $\bar{F}_i(t) = \exp(-t/\beta_i)$; i.e.,

$$\bar{F}_s(t) = \exp\left\{ -t \left( \frac{1}{\beta_1} + \frac{1}{\beta_2} + \ldots + \frac{1}{\beta_k} \right) \right\}$$  \hspace{1cm} (10.3.5)

**Remark 10.3.1** If the components are all identically distributed with a common $\text{Exp}(\beta)$ distribution, then $\bar{F}_s(t) = \exp\{-kt/\beta\}$. The problem then becomes drawing inferences on the reliability function $\exp(-t_*/\beta)$, $t_* = kt$, based on $n$ iid observations with pdf $\left( \frac{1}{\beta} \right) \exp(-x/\beta)$; and one can follow the method discussed in Section 10.2 with $\text{Exp}(\beta)$ model.

For the independent (and possibly nonidentical) exponential components, one can summarize the data by looking at the minimal sufficient
10.3 Reliability of a Series System with Componentwise Data

statistics \( \mathbf{T} = (T_1, \ldots, T_k) \), where

\[
T_i = \sum_{j=1}^{n_i} X_{ij} = \text{total observed lifetime of the prototypes of C-}i,
\]

\[ i = 1, 2, \ldots, k \]  

(10.3.6)

Also, it can be shown that \( T_i \) follows \( G(n_i, \beta_i) \) distribution, \( 1 \leq i \leq k \); and they are independent. Since the MLE of \( \beta_i \) is \( \hat{\beta}_{i(\text{ML})} = T_i/n_i \), the MLE of \( \bar{F}_s(t) \) is

\[
\hat{F}_{s(\text{ML})}(t) = \exp\left\{ -t \sum_{i=1}^{k} \left( n_i/T_i \right) \right\}
\]

(10.3.7)

However, the UMVUE of \( \bar{F}_s(t) \) is

\[
\hat{F}_{s(\text{U})} = \prod_{i=1}^{k} \left\{ \left( 1 - t/T_i \right)^{n_i-1} I \left( T_i > t \right) \right\}
\]

(10.3.8)

where \( I(\cdot) \) denotes an indicator function. In a recent study Chandhuri, Pal, and Sarkar (1998) derived the asymptotic bias and MSE expressions of \( \hat{F}_{s(\text{ML})} \) and \( \hat{F}_{s(\text{U})} \), and it was shown that \( \hat{F}_{s(\text{ML})} \) has smaller first order (i.e., in terms of expansion up to order \( O(n_i^{-1}) \) MSE than that of \( \hat{F}_{s(\text{U})} \).

The first order bias and MSE of \( \hat{F}_{s(\text{ML})} \) (hence called ‘FO-Bias’ and ‘FO-MSE’) are

\[
\text{FO-Bias}(\hat{F}_{s(\text{ML})} \mid \beta_1, \ldots, \beta_k) \approx \left( \bar{F}_s(t) \right) \sum_{i=1}^{k} \frac{t(t - 2\beta_i)/(2n_i\beta_i^2)}{t_i^{n_i-1} dt_i} - \bar{F}_s(t)
\]

\[
\text{FO-MSE}(\hat{F}_{s(\text{ML})} \mid \beta_1, \ldots, \beta_k) \approx \left( \bar{F}_s(t) \right) \sum_{i=1}^{k} \frac{1/(n_i\beta_i^2)}{t_i^{n_i-1} dt_i}
\]

(10.3.9)

On the other hand, the exact bias and MSE are

\[
\text{Bias}(\hat{F}_{s(\text{ML})} \mid \beta_1, \ldots, \beta_k) = \prod_{i=1}^{k} \left[ \int_{0}^{\infty} \left\{ \Gamma(n_i) \beta_i^{n_i} \right\}^{-1} \exp\left\{ -\left( tn_i/t_i \right) - (t_i/\beta_i) \right\} t_i^{n_i-1} dt_i \right] - \bar{F}_s(t)
\]

and

\[
\text{MSE}(\hat{F}_{s(\text{ML})} \mid \beta_1, \ldots, \beta_k) = \prod_{i=1}^{k} \left[ \int_{0}^{\infty} \left\{ \Gamma(n_i) \beta_i^{n_i} \right\}^{-1} \exp\left\{ -\left( 2tn_i/t_i \right) - (t_i/\beta_i) \right\} t_i^{n_i-1} dt_i \right] - 2\bar{F}_s(t) \cdot \text{Bias}(\hat{F}_{s(\text{ML})} \mid \beta_1, \ldots, \beta_k) - (\bar{F}_s(t))^2
\]

(10.3.10)
Chapter 10  System Reliability

\[
(n_1, n_2) \quad \begin{array}{c|c|c|c|c} \hline \beta_1, \beta_2 \quad \text{FO-Bias} & \text{Bias} & \text{FO-MSE} & \text{MSE} \\ \hline (5, 5) & (1.0, 0.1) & 0.0001 & 0.0002 & 0.0000 & 0.0000 \\ & (1.0, 0.5) & -0.0050 & -0.0035 & 0.0025 & 0.0019 \\ & (1.0, 1.0) & -0.0270 & -0.0220 & 0.0073 & 0.0063 \\ & (1.0, 2.0) & -0.0390 & -0.0340 & 0.0124 & 0.0113 \\ \hline (5, 10) & (1.0, 0.1) & 0.0000 & 0.0000 & 0.0001 & 0.0000 \\ & (1.0, 0.5) & -0.0050 & -0.0039 & 0.0015 & 0.0013 \\ & (1.0, 1.0) & -0.0203 & -0.0167 & 0.0055 & 0.0048 \\ & (1.0, 1.5) & -0.0273 & -0.0229 & 0.0087 & 0.0077 \\ & (1.0, 2.0) & -0.0307 & -0.0259 & 0.0112 & 0.0099 \\ \hline \end{array}
\]

Table 10.3.1: A Comparison of First Order Bias and MSE of \( \hat{F}_s(\text{ML}) \) with the Exact Ones \((k = 2)\).

For a complex system it is often desired to have a \((1 - \alpha)\) level lower confidence bound for the system reliability, i.e., one seeks a bound say, \( \hat{F}_{ss} \) such that

\[
(1 - \alpha) \leq P(\hat{F}_{ss} \leq \bar{F}_s(t) \leq 1) \quad \text{(10.3.11)}
\]

One can find a two-sided interval estimate of \( \bar{F}_s(t) \), but since \( \bar{F}_s(t) \) is always bounded above by 1, a lower confidence bound as given in (10.3.11) is enough for the interval estimate. Since \( \bar{F}_s(t) \) is a monotone function of \( \sum_{i=1}^{k} (1/\beta_i) \), it is sufficient to find a \((1 - \alpha)\) level upper confidence bound for it. Arzayan and Pal (2005) derived a conditional test procedure for testing a hypothesis on \( \sum_{i=1}^{k} (1/\beta_i) \), and then inverted the acceptance region to get a \((1 - \alpha)\) level upper confidence bound for \( \sum_{i=1}^{k} (1/\beta_i) \).

The following notations will be useful in deriving an interval estimate for \( \sum_{i=1}^{k} (1/\beta_i) \), and subsequently that for \( \bar{F}_s(t) \).

\[\text{Notations:} \quad \lambda = \sum_{i=1}^{k} (1/\beta_i), \quad n = \sum_{i=1}^{k} n_i; \quad k_i = n/n_i \]

\[\bar{T} = \sum_{i=1}^{k} n_i T_i/n; \quad \text{and} \quad U_i = n(\bar{T} - T_i)/n_i^2, \quad i = 1, 2, \ldots, k \]

© 2006 by Taylor & Francis Group, LLC
10.3 Reliability of a Series System with Componentwise Data

Result 10.3.1 The conditional pdf of $\bar{T}$ given $(U_1, \ldots, U_{k-1})$ is $h_{\lambda}(\bar{t} \mid u_1, \ldots, u_{k-1})$ where

\[
h_{\lambda}(\bar{t} \mid u_1, \ldots, u_{k-1}) \propto \prod_{i=1}^{k-1} \left( \bar{t} - \frac{n_i^2}{n} u_i \right)^{(n_i-1)} \left( \bar{t} + \frac{1}{mn_k} \sum_{i=1}^{k-1} n_i^3 u_i \right)^{(n_k-1)} \times \exp(-\lambda \bar{t})\]

where $\bar{t} > \max \left\{ -1 \frac{1}{mn_k} \sum_{i=1}^{k-1} n_i^2 u_i, n_i^2 u_1, \ldots, n_{k-1}^2 u_{k-1} \right\} = t_0$ (say).

Result 10.3.2 To test $H_0 : \lambda = \lambda_0$, a conditional test procedure is:

1. reject $H_0$ if (given $U_1, \ldots, U_{k-1}$) $\bar{T} > c$ for $H_1 : \lambda < \lambda_0$
2. $\bar{T} < c$ for $H_1 : \lambda > \lambda_0$

The cut-off point $c$ is such that

\[
\alpha = \text{level of the test} = \begin{cases} \int_{t_0}^{c} h_{\lambda_0}(\bar{t} \mid u_1, \ldots, u_{k-1}) \, d\bar{t} & \text{for } H_1 : \lambda < \lambda_0 \\ \int_{c}^{\infty} h_{\lambda_0}(\bar{t} \mid u_1, \ldots, u_{k-1}) \, d\bar{t} & \text{for } H_1 : \lambda > \lambda_0 \end{cases}\]

Also, the cut-off point $c = c(\lambda_0 \mid \alpha, u_1, \ldots, u_{k-1})$ is a decreasing function of $\lambda_0$.

The above procedure is now implemented, as a demonstration, for a two component series system with componentwise data presented by Liberman and Ross (1971).

Example 10.3.1. Independent random samples of size 4 are obtained on each of two components as given below:

Component-1 (or C-1): 2.0 7.1 6.0 3.0
Component-2 (or C-2): 5.0 4.0 2.0 3.1

Assuming exponential lifetime distribution ($\text{Exp}(\beta_i)$) for each component, draw inferences on the reliability of a series system with the above two components.

For the above data set, we have $T_1 = 18.1$ and $T_2 = 14.1$. Since $\hat{\lambda}_{\text{ML}} = 0.504882$ the MLE of $\bar{F}_s(t)$ is

\[
\hat{F}_{s(\text{ML})} = \exp\{-t(0.504882)\} \quad (10.3.12)
\]
We now get the cut-off point $c = c(\lambda_0|\alpha, u_1, \ldots, u_{k-1})$ through numerical integration for selected values of $\lambda_0$. As a rule of thumb one can select values of $\lambda_0$ over the interval $(LB, LB + 3SE)$, where $LB$ = lower bound $= \max_i(1/\hat{\beta}_i(ML)) = \max_i(n_i/T_i)$, and $SE$ is the standard error of $\hat{\lambda}_{ML} = \sum_{i=1}^{k}(1/\hat{\beta}_i(ML)) = \sum_{i=1}^{k}(n_i/T_i)$. (All we need is a rough estimate of $SE$ so that we have an interval for $\lambda_0$.) For large $n_i$ values, $SE = \sqrt{\text{Var}(\hat{\lambda}_{ML})} \approx \left\{ \sum_{i=1}^{k}(n_i\beta_i^2)^{-1} \right\}^{1/2}$ (10.3.13)

Using (10.3.13), we now obtain the values of $c = c(\lambda_0|\alpha, U_1)$ where $U_1 = -1.0$ and $\lambda_0 = 0.2837(0.01)1.185$.

We need a $(1-\alpha)$ level upper confidence bound for $\lambda$. So, we test $H_0 : \lambda = \lambda_0 \text{ vs. } H_A : \lambda < \lambda_0$, and the cut-off point $c = c(\lambda_0|\alpha, U_1)$ is such that

$$\alpha = \int_c^{\infty} h_{\lambda_0}(\bar{t}|u_1 = -1.0) \, d\bar{t} \quad (10.3.14)$$

where $h_{\lambda_0}(\bar{t}|u_1 = -1.0)$ is available from Result 10.3.1. Also, $\bar{T} = 16.1$.

For $\alpha = 0.05$, the values of $c$ satisfying (10.3.14) can be approximated fairly well by the nonlinear function

$$c(\lambda_0|\alpha = 0.05, u_1 = -1.0) \approx c_*(\lambda_0) = b_1\exp(-b_2\lambda_0) + b_3 \quad (10.3.15)$$

with $b_1 = 97.33$, $b_2 = 3.83$ and $b_3 = 9.0$. By inverting $c_*(\lambda_0)$ we get a $(1-\alpha) = 0.95$ level upper confidence bound for $\lambda$ as $(0, c_*^{-1}(\bar{T}))$. [Since the conditional acceptance region for $H_0$ is $\bar{T} \leq c \approx c_*(\lambda_0)$, and $c_*$ is decreasing in $\lambda_0$, we have $\lambda_0 \leq c_*^{-1}(\bar{T})$. ] So, our 95% upper confidence bound for $\lambda$ is 0.6827, which gives the lower 95% confidence bound for $\bar{F}_s(t)$ as $\bar{F}_{s*} = \exp(-0.6827t)$.

The following Figure 10.3.2 plots $\hat{F}_{s(ML)} = \exp(-0.504882t)$ and its 95% lower confidence bound $\hat{F}_{s*} = \exp(-0.6827t)$ against $t > 0$. 

© 2006 by Taylor & Francis Group, LLC
10.3 Reliability of a Series System with Componentwise Data

Figure 10.3.2: Graphs of $\hat{F}_{s(\hat{ML})}$ and its 95% lower confidence bound $\hat{F}_{s*}$ plotted against $t$ (Example 10.3.1).

The general methodology is given through the following steps:

**Step-1**: Get the minimal sufficient statistic $T = (T_1, \ldots, T_k)$ where $T_i = \sum_{j=1}^{k} X_{ij}$, $i = 1, 2, \ldots, k$.

**Step-2**: (a) $\hat{\lambda}_{ML} = \text{MLE of } \lambda = \sum_{i=1}^{k} (1/\beta_i) = \sum_{i=1}^{k} (n_i/T_i)$.

(b) Obtain $LB$ (lower bound) and $UB$ (upper bound) for $\lambda$ as $LB = \max_i (n_i/T_i)$ and $UB = LB + \{\sum_{i=1}^{k} (n_i/\beta_i)^2\}^{1/2}$, with $\beta_i$ replaced by $\hat{\beta}_i(\hat{ML})$.

**Step-3**: For $\lambda_0 = (LB)(0.1)(UB)$, calculate $c = c(\lambda_0|\alpha, u_1, \ldots, u_{k-1})$ such that

$$\alpha = \frac{\int_{\infty}^{\infty} h^*(\bar{t}|u_1, \ldots, u_{k-1}) \, d\bar{t}}{\int_{t_0}^{\infty} h^*(\bar{t}|u_1, \ldots, u_{k-1}) \, d\bar{t}}$$

where

$$h^*(\bar{t}|u_1, \ldots, u_{k-1}) = \prod_{i=1}^{k-1} \left( \bar{t} - \frac{n_i^2}{n} u_i \right)^{n_i-1} \left( \bar{t} + \frac{1}{nn_k} \sum_{i=1}^{k} n_i^3 u_i \right)^{(n_k-1)} \exp(-\bar{t} \lambda)$$

The expression for $t_0$ is given in Result 10.3.1.
Step-4: Plot the \( c \) values against \( \lambda_0 \). Note that \( c \), as a function of \( \lambda_0 \), is decreasing. Approximate \( c(\lambda_0; \alpha, u_1, \ldots, u_{k-1}) \) by \( c^*(\lambda_0) = b_1 \exp(-b_2 \lambda_0) + b_3 \), and obtain the values of \( b_1, b_2, \) and \( b_3 \).

Step-5: The \( (1 - \alpha) \) level upper confidence bound for \( \lambda \) is obtained as \( \lambda \leq c^{-1}(\bar{T}) = -\ln((\bar{T} - b_3)/b_1)/b_2 \). The \( (1 - \alpha) \) level lower confidence bound for \( \bar{F}_s(t) \) is \( \bar{F}_s(t) \geq \exp\{-tc^{-1}(_s)(\bar{T})\} = \hat{\bar{F}}_s \).

Components Following Gamma Distribution

If we assume that each \( X_i \), i.e., the lifetime of the \( i \)th component (C-\( i \)) follows a \( G(\alpha_i, \beta_i) \) model (see (7.1.1)) then \( \bar{F}_i(t) = \int_t^\infty \{G(\alpha_i, \beta_i) \text{ pdf at } x\} \, dx \), and the reliability of the whole system at time \( t \) is

\[
\bar{F}_s(t) = \prod_{i=1}^{k} \left[ \int_t^\infty \{G(\alpha_i, \beta_i) \text{ pdf at } x\} \, dx \right] \tag{10.3.16}
\]

Further, the MLE of \( \bar{F}_s(t) \) is

\[
\hat{\bar{F}}_{s(ML)}(t) = \prod_{i=1}^{k} \left[ \int_t^\infty \{G(\hat{\alpha}_{i(ML)}, \hat{\beta}_{i(ML)}) \text{ pdf at } x\} \, dx \right] \tag{10.3.17}
\]

(One can use \( \hat{\alpha}_{i(MML)} \) and \( \hat{\beta}_{i(MML)} \) instead of \( \hat{\alpha}_{i(ML)} \) and \( \hat{\beta}_{i(ML)} \). The resultant estimate of \( \bar{F}_s(t) \) will be denoted by \( \hat{\bar{F}}_{s(MML)} \).)

The sampling distribution of \( \hat{\bar{F}}_{s(ML)} \) (or that of \( \hat{\bar{F}}_{s(MML)} \)) is not possible to derive due to the complexity involved. Even if one tries to study the bias and MSE of \( \hat{\bar{F}}_{s(ML)} \) through simulation, a complete picture is hard to comprehend if \( k \geq 3 \) due to the large number of parameters involved. When \( n_i \)'s are sufficiently large, one can follow the bootstrap method to study the sampling distribution of \( \hat{\bar{F}}_{s(ML)} \). Also, for large \( n_i \)'s, \( \hat{\bar{F}}_{s(ML)} \) is approximately unbiased and follows an asymptotic normal distribution which can be used for interval estimating as given below.

Define \( \eta_{i1}, \eta_{i2}, \) and \( \eta_{i3} \) (which are functions of \( \alpha_i, \beta_i, \) and \( t \)) for \( i = 1, 2, \ldots, k \) as

\[
\eta_{i1} = \int_0^t \{G(\alpha_i, \beta_i) \text{ pdf at } x\} \, dx = 1 - \bar{F}_i(t) \tag{10.3.18}
\]

\[
\eta_{i2} = \int_0^t (\ln x) \{G(\alpha_i, \beta_i) \text{ pdf at } x\} \, dx \tag{10.3.19}
\]
10.3 Reliability of a Series System with Componentwise Data

\[ \eta_{i3} = \int_0^t (\ln x)^2 \left\{ G(\alpha_i, \beta_i) \right\} \text{pdf at } x \, dx \quad (10.3.20) \]

Define \( \sigma_i^{*2} \) (which is a function of \( \alpha_i \), \( \beta_i \), and \( t \)) as

\[
\sigma_i^{*2} = \left\{ n_i (\alpha_i \psi'(\alpha_i) - 1) \right\}^{-1} \left[ \alpha_i \left\{ \eta_{i2} - (\ln \beta_i + \psi(\alpha_i)) \eta_{i1} \right\}^2 
+ \psi'(\alpha_i) \left\{ (\alpha_i - \beta_i^{-2}) \eta_{i1} \right\}^2 
+ 2 \left\{ \eta_{i2} - (\ln \beta_i + \psi(\alpha_i)) \eta_{i1} \right\} \left\{ (\alpha_i - \beta_i^{-2}) \eta_{i1} \right\} \right\}^{1/2} \quad (10.3.21)
\]

where \( \psi(\alpha_i) = \Gamma'(\alpha_i)/\Gamma(\alpha_i) \) and \( \psi'(\alpha_i) = (\partial/\partial \alpha_i) \psi(\alpha_i) \) are di-gamma and tri-gamma functions respectively. Using the standard asymptotic properties of \( \hat{\alpha}_{i(ML)} \) and \( \hat{\beta}_{i(ML)} \), \( i = 1, 2, \ldots, k \), it can be shown that as each \( n_i \to \infty \),

\[
\left[ \ln \hat{F}_{s(ML)} \right] \overset{d}{\rightarrow} N \left( \ln \hat{F}_s(t), \sum_{i=1}^k \left( \frac{\sigma_i^{*2}}{\eta_{i1}^2} \right) \right) \quad (10.3.22)
\]

Since \( \sigma_i^{*2} \) and \( \eta_{i1}^2 \) are unknown (they have unknown parameters \( \alpha_i \)'s and \( \beta_i \)'s in them), obtain their estimates \( \hat{\sigma}_i^{*2} \) and \( \hat{\eta}_{i1}^2 \) respectively by replacing \( \alpha_i \) and \( \beta_i \) by \( \hat{\alpha}_{i(ML)} \) and \( \hat{\beta}_{i(ML)} \). For convenience, define

\[
\hat{\delta}^2 = \sum_{i=1}^k \left( \frac{\hat{\sigma}_i^{*2}}{\hat{\eta}_{i1}^2} \right) \quad (10.3.23)
\]

Then, an approximate \( (1 - \alpha) \) level confidence bound for \( \hat{F}_s(t) \) is found as (using (10.3.23))

\[
\hat{F}_{ss} = \hat{F}_{s(ML)} \exp \{ -z_{\alpha} \hat{\delta} \} \quad (10.3.24)
\]

where \( z_{\alpha} \) is the right tail \( \alpha \)-probability cut-off point of the standard normal distribution.

Though the confidence bound in (10.3.24) is meant for large \( n_i \) values, it can still be used as a crude bound for small sample sizes. In the following we compute the value of \( \hat{F}_{ss} \) for a small data set used by Draper and Guttman (1972).

**Example 10.3.2.** Independent samples of size 3 and 4 are obtained on lifetime (in 1000 hours) of each of two components of a series system with the following results:
Component-1: 28.56, 22.42, 180.24
Component-2: 35.76, 226.58, 160.08, 556.06

Assuming a $G(\alpha_i, \beta_i)$ model for component-$i$ lifetime ($i = 1, 2$) infer about the system reliability at time $t = 100$.

**Step-1**: Using the observations \{28.56, 22.42, 180.24\} obtain $\hat{\alpha}_1(\text{ML}) = 1.228$, $\hat{\beta}_1(\text{ML}) = 62.742$ (solving the system of equations given in Chapter 7). Similarly, using the observations \{35.76, 226.58, 160.08, 556.06\} obtain $\hat{\alpha}_2(\text{ML}) = 1.39$, $\hat{\beta}_2(\text{ML}) = 175.823$.

**Step-2**: Plot $\hat{F}_s(\text{ML})$ (given in (10.3.17)) against $t > 0$. At $t = 100$, $\hat{F}_s(\text{ML}) = 0.1986$.

![Plot of $\hat{F}_s(\text{ML})$ against $t > 0$ (Example 10.3.2).](image)

**Step-3**: Using $t = 100$, we have $\hat{\eta}_{11} = 0.7284$, $\hat{\eta}_{21} = 0.2688$; $\hat{\eta}_{12} = 2.5013$, $\hat{\eta}_{22} = 1.0165$; $\hat{\eta}_{13} = 9.2758$, $\hat{\eta}_{23} = 4.0058$. Also, $\hat{\delta}^2 = 1.0816$.

**Step-4**: The lower confidence bound for $\hat{F}_s(t)$ at $t = 100$ is $\hat{F}_{s*} = \hat{F}_s(\text{ML}) \exp\{-z_\alpha \hat{\delta}\} = 0.1986 \exp\{-1.645(1.04)\} = 0.03589$ for $\alpha = 0.05$. 

© 2006 by Taylor & Francis Group, LLC
10.4 Reliability of a Parallel System with Componentwise Data

Step-5: Plot $\widehat{F}_s^*$ (in Step-4) against $\alpha$, $0 < \alpha < 1$. (Here, $(1 - \alpha)$ is the confidence level.)

Figure 10.3.4: Plot of $\widehat{F}_s^*$ against $\alpha$ (Example 10.3.2).

10.4 Reliability of a Parallel System with Componentwise Data

In this Section we consider a system consisting of $k$--independent components say C-1, C-2, ... , C-k, connected parallelly. The whole system works as long as at least one component works, and the whole system fails as soon as all the components fail.
If the life time $X_i$ of C-$i$ follows a pdf $f_i(\cdot)$ (with cdf $F_i(\cdot)$), then assuming that the components are all independent, the reliability of the parallel system at time $t$, denoted by $\bar{F}_p(t)$, is

\[
\bar{F}_p(t) = P(\text{Parallel System life} > t) = 1 - P(\text{Parallel System life} \leq t) = 1 - P(X_1 \leq t, \ldots, X_k \leq t)
\]

\[
= 1 - \prod_{i=1}^{k} F_i(t) = 1 - \prod_{i=1}^{k} (1 - \bar{F}_i(t)) \quad (10.4.1)
\]

Note that

(a) $\bar{F}_p(t) \geq \max_{1 \leq i \leq k} \bar{F}_i(t)$; i.e., the reliability of a parallel system is larger than the reliability of individual components

(b) If the components have identical probability distribution (with common cdf $F(\cdot)$ and reliability function $\bar{F}(\cdot)$), then $\bar{F}_p(t) = 1 - (1 - \bar{F}(t))^k$

Again, similar to the series system case, the expression (10.4.1) gets very complicated for most of the probability distributions. So most of the work that is available in the literature is only for the exponential distribution. We adopt the componentwise data set as presented in (10.3.4).
10.4 Reliability of a Parallel System with Componentwise Data

Components Following an Exponential Distribution

If we follow that each $X_i$ follows an $\text{Exp}(\beta_i)$ distribution, then $\bar{F}_i(t) = \exp(-t/\beta_i)$; i.e.,

$$\bar{F}_i(t) = 1 - \prod_{i=1}^{k} \left\{ 1 - \exp(-t/\beta_i) \right\} \tag{10.4.2}$$

For (i) $k = 2$,

$$\bar{F}_p(t) = (e^{-t/\beta_1} + e^{-t/\beta_2}) - e^{-t(1/\beta_1 + 1/\beta_2)}$$

(ii) $k = 3$,

$$\bar{F}_p(t) = \left( e^{-t/\beta_1} + e^{-t/\beta_2} + e^{-t/\beta_3} \right) - \left( e^{-t(1/\beta_1 + 1/\beta_2)} + e^{-t(1/\beta_1 + 1/\beta_3)} + e^{-t(1/\beta_2 + 1/\beta_3)} \right)$$

In general, the expansion of (10.4.2) looks like

$$\bar{F}_p(t) = \sum_{l=1}^{k} (-1)^{l-1} \sum_{(i_1, \ldots, i_l) \in J} \exp \left\{ -t \sum_{i=1}^{l} (1/\beta_{i_1} + 1/\beta_{i_2} + \ldots + 1/\beta_{i_l}) \right\} \tag{10.4.3}$$

Using the componentwise data and the notations used in (10.3.6), the MLE of $\bar{F}_p(t)$ is

$$\hat{\bar{F}}_{p(\text{ML})}(t) = \sum_{l=1}^{k} (-1)^{l-1} \sum_{(i_1, \ldots, i_l) \in J} \exp \left\{ -t \sum_{i=1}^{l} \left( \frac{n_{ij}}{T_{ij}} \right) \right\} \tag{10.4.4}$$

where $n_{ij}$ and $T_{ij}$ are respectively the $n_i$ and $T_i$ values corresponding to subscript value $i_j$.

Similar to the case of the series system, the UMVUE of $\bar{F}_p(t)$ can be found as

$$\hat{\bar{F}}_{p(\text{U})} = \sum_{l=1}^{k} (-1)^{l-1} \sum_{(i_1, \ldots, i_l) \in J} \prod_{j=1}^{l} \left\{ (1 - \frac{t}{T_{ij}})^{n_{ij}-1} I(T_{ij} > t) \right\} \tag{10.4.5}$$

For the sake of simplicity we focus on the case of $k = 2$ (i.e., a two-component parallel system with exponential lifetime for each component).
With \( k = 2 \),
\[
\hat{F}_{p(ML)} = \left\{ e^{-tn_1/T_1} + e^{-tn_2/T_2} \right\} - e^{t((n_1/T_1)+(n_1/T_2))} \quad (10.4.6)
\]

To derive the bias and MSE of \( \hat{F}_{p(ML)} \) in (10.4.6), it is convenient to adopt the following notations. Let
\[
t_{i*} = t/\beta_i, \quad i = 1, 2
\]
\[
E_{1i} = \int_0^{\infty} \exp\left\{ (-n_it_{i*}/u_i) - u_i \right\} du_i/\Gamma(n_i), \quad i = 1, 2
\]
\[
(10.4.7)
\]

Then the bias of \( \hat{F}_{p(ML)} \) is
\[
B(\hat{F}_{p(ML)}|\beta_1, \beta_2) = \sum_{i=1}^{2} \left( E_{1i} - \exp(-t_{i*}) \right) - \left\{ \prod_{i=1}^{2} E_{1i} - \prod_{i=1}^{2} \exp(-t_{i*}) \right\}
\]
\[
(10.4.8)
\]

In order to write the MSE expansion of \( \hat{F}_{p(ML)} \) define
\[
E_{2i} = \int_0^{\infty} \exp\left\{ (-2n_it_{i*}/u_i) - u_i \right\} du_i/\Gamma(n_i), \quad i = 1, 2
\]
\[
(10.4.9)
\]

Then,
\[
MSE(\hat{F}_{p(ML)}|\beta_1, \beta_2) = \sum_{i=1}^{2} E_{2i} + 2 \prod_{i=1}^{2} E_{1i} + \prod_{i=1}^{2} E_{2i} - 2E_{21}E_{12} - 2E_{11}E_{22} - 2\hat{F}_p(t)B(\hat{F}_{p(ML)}|\beta_1, \beta_2) - (\hat{F}_p(t))^2
\]

Note that, using the notations \( t_{i*} = t/\beta_i, \quad i = 1, 2 \),
\[
\hat{F}_p(t) = \sum_{i=1}^{2} \exp(-t_{i*}) - \prod_{i=1}^{2} \exp(-t_{i*}) \quad (10.4.10)
\]

Thus, both the bias and the MSE of \( \hat{F}_{p(ML)} \) are dependent on \( \beta_1 \) and \( \beta_2 \) only through \( t_{i*} \) and \( t_{2*} \).

In the following we plot three dimensional diagrams of \( B(\hat{F}_{p(ML)}|\beta_1, \beta_2) \) and \( MSE(\hat{F}_{p(ML)}|\beta_1, \beta_2) \) against \( t_{1*} \) and \( t_{2*} \).
Figure 10.4.2: (a) Bias of $\hat{F}_{p}^{\text{ML}}$ for $k = 2$ and $(n_1, n_2) = (5, 10)$.

Figure 10.4.2: (b) Bias of $\hat{F}_{p}^{\text{ML}}$ for $k = 2$ and $(n_1, n_2) = (10, 10)$. 
Figure 10.4.3: (a) MSE of $\hat{F}_{p(ML)}$ for $k = 2$ and $(n_1, n_2) = (5, 10)$.

Figure 10.4.3: (b) MSE of $\hat{F}_{p(ML)}$ for $k = 2$ and $(n_1, n_2) = (10, 10)$. 
Components Following Gamma Distribution

If we assume that $X_i$, i.e., the lifetime of the $i^{th}$ component $(C_i)$ follows a $G(\alpha_i, \beta_i)$ model (in (7.1.1)) then $\bar{F}_i(t) = \int_0^\infty \{G(\alpha_i, \beta_i) \text{ pdf at } x\} \, dx$, and the reliability of the whole parallel at time $t$ is

$$\bar{F}_p(t) = 1 - \prod_{i=1}^k \left[ \int_0^t \{G(\alpha_i, \beta_i) \text{ pdf at } x\} \, dx \right]$$  \hspace{1cm} (10.4.11)

The MLE of $\bar{F}_p(t)$ is

$$\hat{F}_p(ML) = 1 - \prod_{i=1}^k \left[ \int_0^t \{G(\hat{\alpha}_{i(ML)}, \hat{\beta}_{i(ML)}) \text{ pdf at } x\} \, dx \right]$$  \hspace{1cm} (10.4.12)

One can also use $\hat{\alpha}_{i(MML)}$ and $\hat{\beta}_{i(MML)}$ instead of $\hat{\alpha}_{i(ML)}$ and $\hat{\beta}_{i(ML)}$. The resultant estimate of $\bar{F}_p(t)$ will be denoted by $\hat{F}_p(MML)(t)$.

Again, similar to the series system, the sampling distribution of $\hat{F}_p(ML)(t)$ (or $\hat{F}_p(MML)(t)$) is nearly impossible. However, for large $n_i$'s, $\hat{F}_p(ML)(t)$ (and $\hat{F}_p(MML)(t)$) is (are) approximately unbiased and has (have) asymptotic normal distribution. We will use the notations adopted in (10.3.18) to (10.3.21). It can be shown that as each $n_i \to \infty$,

$$\ln(1 - \hat{F}_p(ML)) \overset{d}{\to} N\left(\ln \left[ \prod_{i=1}^k q_i \sum_{i=1}^k \left( \sigma_i^2 / \eta_i^2 \right) \right] \right)$$  \hspace{1cm} (10.4.13)

Note $\hat{\delta}$ as defined in (10.3.23). Then an approximate $(1 - \alpha)$ level lower confidence bound for $\hat{F}_p(t)$ is found as

$$\hat{F}_{pl} = 1 - (1 - \hat{F}_p(ML)) \exp\{z_\alpha \hat{\delta}\}$$  \hspace{1cm} (10.4.14)

where $z_\alpha$ is the right tail $\alpha$-probability cut-off point of the standard normal distribution.

Example 10.4.1. We use the data set of Example 10.3.2 for two components assuming that they are parallelly connected and each follows a gamma model.

We already have (from Step-2 of Example 10.3.2) $\hat{\alpha}_{1(ML)} = 1.228$, $\hat{\beta}_{1(ML)} = 62.742$; $\hat{\alpha}_{2(ML)} = 1.390$, $\hat{\beta}_{2(ML)} = 175.823$.

Also, using $t = 25$, we have $\hat{\eta}_{11} = 0.2324$, $\hat{\eta}_{21} = 0.0496$; $\hat{\eta}_{12} = 0.5401$, $\hat{\eta}_{22} = 0.1277$; $\hat{\eta}_{13} = 1.4253$, $\hat{\eta}_{23} = 0.3304$; and $\hat{\delta}^2 = 3.100$. So $\hat{F}_p(ML) =$
0.9884. For the above specific data set, the plot of the expression $\hat{F}_{p(ML)}$ as a function of $t$ is given in Figure 10.4.4.

The lower confidence bound for $\bar{F}_p(t)$ at $t = 25$ is

$$\hat{F}_{p*} = 1 - (1 - \hat{F}_{p(ML)}) exp(-z_\alpha \delta) = 1 - 0.0116 \times \exp(z_\alpha 1.76) = 0.7914$$ (using $\alpha = 0.05$)

[We would like to point out that the above lower bound of the parallel system reliability is not very effective for small sample sizes, and can be meaningful only for moderate to large sample sizes.]

Remark 10.4.1 Recently Chandhury, Hu, and Afshar (2001) have provided a nice algorithm to calculate the confidence bounds for system reliability which is applicable for general systems (beyond the series and parallel systems discussed here).
Figure 10.4.5: Plot of the lower confidence bound $\bar{F}_{ps}$ against $\alpha$. 
Selected Statistical Tables

1. The Standard Normal Distribution Table
2. (Student’s) t-Distribution Table
3. Chi-Square Distribution Table
4. (Snedecor’s) F-Distribution Table
1. The Standard Normal Distribution Table

The table entry corresponding to $z$ denotes $P(0 \leq Z \leq z)$ for a standard normal random variable $Z$ (i.e., the shaded region = area under the $N(0,1)$ curve between 0 and $z$).

<table>
<thead>
<tr>
<th>$z$</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.0000</td>
<td>.0040</td>
<td>.0080</td>
<td>.0120</td>
<td>.0160</td>
<td>.0199</td>
<td>.0239</td>
<td>.0279</td>
<td>.0319</td>
<td>.0359</td>
</tr>
<tr>
<td>0.1</td>
<td>.0398</td>
<td>.0438</td>
<td>.0478</td>
<td>.0517</td>
<td>.0557</td>
<td>.0596</td>
<td>.0636</td>
<td>.0675</td>
<td>.0714</td>
<td>.0753</td>
</tr>
<tr>
<td>0.2</td>
<td>.0793</td>
<td>.0832</td>
<td>.0871</td>
<td>.0910</td>
<td>.0948</td>
<td>.0987</td>
<td>.1026</td>
<td>.1064</td>
<td>.1103</td>
<td>.1141</td>
</tr>
<tr>
<td>0.3</td>
<td>.1179</td>
<td>.1217</td>
<td>.1255</td>
<td>.1293</td>
<td>.1331</td>
<td>.1368</td>
<td>.1406</td>
<td>.1443</td>
<td>.1480</td>
<td>.1517</td>
</tr>
<tr>
<td>0.4</td>
<td>.1554</td>
<td>.1591</td>
<td>.1628</td>
<td>.1664</td>
<td>.1700</td>
<td>.1736</td>
<td>.1772</td>
<td>.1808</td>
<td>.1844</td>
<td>.1879</td>
</tr>
<tr>
<td>0.5</td>
<td>.1915</td>
<td>.1950</td>
<td>.1985</td>
<td>.2019</td>
<td>.2054</td>
<td>.2088</td>
<td>.2123</td>
<td>.2157</td>
<td>.2190</td>
<td>.2224</td>
</tr>
<tr>
<td>0.6</td>
<td>.2257</td>
<td>.2291</td>
<td>.2324</td>
<td>.2357</td>
<td>.2389</td>
<td>.2422</td>
<td>.2454</td>
<td>.2486</td>
<td>.2517</td>
<td>.2549</td>
</tr>
<tr>
<td>0.7</td>
<td>.2580</td>
<td>.2611</td>
<td>.2642</td>
<td>.2673</td>
<td>.2703</td>
<td>.2734</td>
<td>.2764</td>
<td>.2793</td>
<td>.2823</td>
<td>.2852</td>
</tr>
<tr>
<td>0.8</td>
<td>.2881</td>
<td>.2910</td>
<td>.2939</td>
<td>.2967</td>
<td>.2995</td>
<td>.3023</td>
<td>.3051</td>
<td>.3078</td>
<td>.3106</td>
<td>.3133</td>
</tr>
<tr>
<td>0.9</td>
<td>.3159</td>
<td>.3186</td>
<td>.3212</td>
<td>.3238</td>
<td>.3264</td>
<td>.3289</td>
<td>.3315</td>
<td>.3340</td>
<td>.3365</td>
<td>.3389</td>
</tr>
<tr>
<td>1.0</td>
<td>.3413</td>
<td>.3438</td>
<td>.3461</td>
<td>.3485</td>
<td>.3508</td>
<td>.3531</td>
<td>.3554</td>
<td>.3577</td>
<td>.3599</td>
<td>.3621</td>
</tr>
<tr>
<td>1.1</td>
<td>.3643</td>
<td>.3665</td>
<td>.3686</td>
<td>.3708</td>
<td>.3729</td>
<td>.3749</td>
<td>.3770</td>
<td>.3790</td>
<td>.3810</td>
<td>.3830</td>
</tr>
<tr>
<td>1.2</td>
<td>.3849</td>
<td>.3869</td>
<td>.3888</td>
<td>.3907</td>
<td>.3925</td>
<td>.3943</td>
<td>.3962</td>
<td>.3980</td>
<td>.3997</td>
<td>.4015</td>
</tr>
<tr>
<td>1.3</td>
<td>.4032</td>
<td>.4049</td>
<td>.4066</td>
<td>.4082</td>
<td>.4099</td>
<td>.4115</td>
<td>.4131</td>
<td>.4147</td>
<td>.4162</td>
<td>.4177</td>
</tr>
<tr>
<td>1.4</td>
<td>.4192</td>
<td>.4207</td>
<td>.4222</td>
<td>.4236</td>
<td>.4251</td>
<td>.4265</td>
<td>.4279</td>
<td>.4292</td>
<td>.4306</td>
<td>.4319</td>
</tr>
<tr>
<td>1.5</td>
<td>.4332</td>
<td>.4345</td>
<td>.4357</td>
<td>.4370</td>
<td>.4382</td>
<td>.4394</td>
<td>.4406</td>
<td>.4418</td>
<td>.4429</td>
<td>.4441</td>
</tr>
<tr>
<td>1.6</td>
<td>.4452</td>
<td>.4463</td>
<td>.4474</td>
<td>.4484</td>
<td>.4495</td>
<td>.4505</td>
<td>.4515</td>
<td>.4525</td>
<td>.4535</td>
<td>.4545</td>
</tr>
<tr>
<td>1.7</td>
<td>.4554</td>
<td>.4564</td>
<td>.4573</td>
<td>.4582</td>
<td>.4591</td>
<td>.4599</td>
<td>.4608</td>
<td>.4616</td>
<td>.4625</td>
<td>.4633</td>
</tr>
<tr>
<td>1.8</td>
<td>.4641</td>
<td>.4649</td>
<td>.4656</td>
<td>.4664</td>
<td>.4671</td>
<td>.4678</td>
<td>.4686</td>
<td>.4693</td>
<td>.4699</td>
<td>.4706</td>
</tr>
<tr>
<td>1.9</td>
<td>.4713</td>
<td>.4719</td>
<td>.4726</td>
<td>.4732</td>
<td>.4738</td>
<td>.4744</td>
<td>.4750</td>
<td>.4756</td>
<td>.4761</td>
<td>.4767</td>
</tr>
<tr>
<td>2.0</td>
<td>.4772</td>
<td>.4778</td>
<td>.4783</td>
<td>.4788</td>
<td>.4793</td>
<td>.4798</td>
<td>.4803</td>
<td>.4808</td>
<td>.4812</td>
<td>.4817</td>
</tr>
<tr>
<td>2.1</td>
<td>.4821</td>
<td>.4826</td>
<td>.4830</td>
<td>.4834</td>
<td>.4838</td>
<td>.4842</td>
<td>.4846</td>
<td>.4850</td>
<td>.4854</td>
<td>.4857</td>
</tr>
<tr>
<td>2.2</td>
<td>.4861</td>
<td>.4864</td>
<td>.4868</td>
<td>.4871</td>
<td>.4875</td>
<td>.4878</td>
<td>.4881</td>
<td>.4884</td>
<td>.4887</td>
<td>.4890</td>
</tr>
<tr>
<td>2.3</td>
<td>.4893</td>
<td>.4896</td>
<td>.4898</td>
<td>.4901</td>
<td>.4904</td>
<td>.4906</td>
<td>.4909</td>
<td>.4911</td>
<td>.4913</td>
<td>.4916</td>
</tr>
<tr>
<td>2.4</td>
<td>.4918</td>
<td>.4920</td>
<td>.4922</td>
<td>.4924</td>
<td>.4927</td>
<td>.4929</td>
<td>.4931</td>
<td>.4932</td>
<td>.4934</td>
<td>.4936</td>
</tr>
<tr>
<td>2.5</td>
<td>.4938</td>
<td>.4940</td>
<td>.4941</td>
<td>.4943</td>
<td>.4945</td>
<td>.4946</td>
<td>.4948</td>
<td>.4949</td>
<td>.4951</td>
<td>.4952</td>
</tr>
<tr>
<td>2.6</td>
<td>.4953</td>
<td>.4955</td>
<td>.4956</td>
<td>.4957</td>
<td>.4959</td>
<td>.4960</td>
<td>.4961</td>
<td>.4962</td>
<td>.4963</td>
<td>.4964</td>
</tr>
<tr>
<td>2.7</td>
<td>.4965</td>
<td>.4966</td>
<td>.4967</td>
<td>.4968</td>
<td>.4969</td>
<td>.4970</td>
<td>.4971</td>
<td>.4972</td>
<td>.4973</td>
<td>.4974</td>
</tr>
<tr>
<td>2.8</td>
<td>.4974</td>
<td>.4975</td>
<td>.4976</td>
<td>.4977</td>
<td>.4977</td>
<td>.4978</td>
<td>.4979</td>
<td>.4979</td>
<td>.4980</td>
<td>.4981</td>
</tr>
<tr>
<td>2.9</td>
<td>.4981</td>
<td>.4982</td>
<td>.4982</td>
<td>.4983</td>
<td>.4984</td>
<td>.4984</td>
<td>.4985</td>
<td>.4985</td>
<td>.4986</td>
<td>.4986</td>
</tr>
<tr>
<td>3.0</td>
<td>.4986</td>
<td>.4987</td>
<td>.4987</td>
<td>.4988</td>
<td>.4988</td>
<td>.4989</td>
<td>.4989</td>
<td>.4989</td>
<td>.4990</td>
<td>.4990</td>
</tr>
<tr>
<td>3.1</td>
<td>.4990</td>
<td>.4991</td>
<td>.4991</td>
<td>.4992</td>
<td>.4992</td>
<td>.4992</td>
<td>.4992</td>
<td>.4993</td>
<td>.4993</td>
<td>.4993</td>
</tr>
<tr>
<td>3.2</td>
<td>.4993</td>
<td>.4993</td>
<td>.4994</td>
<td>.4994</td>
<td>.4994</td>
<td>.4994</td>
<td>.4994</td>
<td>.4994</td>
<td>.4995</td>
<td>.4995</td>
</tr>
</tbody>
</table>
2. (Student’s) \( t \)-Distribution Table
The table entries (i.e., \( t_{\nu, \alpha} \) values) are the right tail \( \alpha \) probability cut-off points for the \( t_{\nu} \)-distribution, where \( \nu = df \) (i.e., area under the \( t_{\nu} \)-pdf on the right of \( t_{\nu, \alpha} \) is \( \alpha \)).

<table>
<thead>
<tr>
<th>( df )</th>
<th>0.150</th>
<th>0.100</th>
<th>0.075</th>
<th>0.050</th>
<th>0.025</th>
<th>0.010</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6962</td>
<td>3.0777</td>
<td>4.0366</td>
<td>6.1280</td>
<td>12.7062</td>
<td>31.8210</td>
<td>63.6559</td>
</tr>
<tr>
<td>2</td>
<td>1.8862</td>
<td>2.9777</td>
<td>3.8776</td>
<td>5.8400</td>
<td>10.2150</td>
<td>27.0540</td>
<td>54.0776</td>
</tr>
<tr>
<td>3</td>
<td>2.0424</td>
<td>2.3533</td>
<td>3.1799</td>
<td>4.5400</td>
<td>8.6108</td>
<td>21.3190</td>
<td>42.9769</td>
</tr>
<tr>
<td>4</td>
<td>2.2099</td>
<td>2.1015</td>
<td>2.7407</td>
<td>3.7469</td>
<td>7.4537</td>
<td>18.1240</td>
<td>34.9145</td>
</tr>
<tr>
<td>5</td>
<td>2.3299</td>
<td>2.0126</td>
<td>2.5082</td>
<td>3.4108</td>
<td>6.9452</td>
<td>16.2380</td>
<td>31.5987</td>
</tr>
<tr>
<td>7</td>
<td>2.5328</td>
<td>1.8771</td>
<td>2.2682</td>
<td>3.1058</td>
<td>6.0601</td>
<td>13.5820</td>
<td>26.7979</td>
</tr>
<tr>
<td>8</td>
<td>2.6181</td>
<td>1.8209</td>
<td>2.1964</td>
<td>3.0000</td>
<td>5.7059</td>
<td>12.5840</td>
<td>25.0474</td>
</tr>
<tr>
<td>9</td>
<td>2.6954</td>
<td>1.7673</td>
<td>2.1385</td>
<td>2.9004</td>
<td>5.3836</td>
<td>11.6460</td>
<td>23.7104</td>
</tr>
<tr>
<td>10</td>
<td>2.7644</td>
<td>1.7173</td>
<td>2.0910</td>
<td>2.8185</td>
<td>5.1502</td>
<td>10.8200</td>
<td>22.5637</td>
</tr>
<tr>
<td>11</td>
<td>2.8251</td>
<td>1.6701</td>
<td>2.0507</td>
<td>2.7426</td>
<td>4.9406</td>
<td>10.1030</td>
<td>21.5929</td>
</tr>
<tr>
<td>12</td>
<td>2.8817</td>
<td>1.6256</td>
<td>1.9996</td>
<td>2.6753</td>
<td>4.7526</td>
<td>9.4742</td>
<td>20.7270</td>
</tr>
<tr>
<td>13</td>
<td>2.9337</td>
<td>1.5839</td>
<td>1.9563</td>
<td>2.6152</td>
<td>4.5860</td>
<td>8.9263</td>
<td>19.9158</td>
</tr>
<tr>
<td>14</td>
<td>2.9815</td>
<td>1.5446</td>
<td>1.9185</td>
<td>2.5587</td>
<td>4.4322</td>
<td>8.4345</td>
<td>19.1730</td>
</tr>
<tr>
<td>15</td>
<td>3.0262</td>
<td>1.5071</td>
<td>1.8835</td>
<td>2.5035</td>
<td>4.3027</td>
<td>8.0101</td>
<td>18.5101</td>
</tr>
<tr>
<td>16</td>
<td>3.0676</td>
<td>1.4715</td>
<td>1.8513</td>
<td>2.4516</td>
<td>4.1879</td>
<td>7.6208</td>
<td>17.9110</td>
</tr>
<tr>
<td>17</td>
<td>3.1064</td>
<td>1.4376</td>
<td>1.8223</td>
<td>2.4027</td>
<td>4.0861</td>
<td>7.2617</td>
<td>17.3819</td>
</tr>
<tr>
<td>18</td>
<td>3.1436</td>
<td>1.4054</td>
<td>1.7960</td>
<td>2.3566</td>
<td>3.9950</td>
<td>6.9335</td>
<td>16.9208</td>
</tr>
<tr>
<td>19</td>
<td>3.1797</td>
<td>1.3747</td>
<td>1.7721</td>
<td>2.3125</td>
<td>3.9112</td>
<td>6.6343</td>
<td>16.5272</td>
</tr>
<tr>
<td>21</td>
<td>3.2486</td>
<td>1.3173</td>
<td>1.7296</td>
<td>2.2300</td>
<td>3.7663</td>
<td>6.1082</td>
<td>15.9292</td>
</tr>
<tr>
<td>22</td>
<td>3.2817</td>
<td>1.2902</td>
<td>1.7104</td>
<td>2.1900</td>
<td>3.7047</td>
<td>5.8730</td>
<td>15.6991</td>
</tr>
<tr>
<td>23</td>
<td>3.3139</td>
<td>1.2640</td>
<td>1.6924</td>
<td>2.1513</td>
<td>3.6497</td>
<td>5.6538</td>
<td>15.5063</td>
</tr>
<tr>
<td>24</td>
<td>3.3454</td>
<td>1.2385</td>
<td>1.6755</td>
<td>2.1137</td>
<td>3.5999</td>
<td>5.4486</td>
<td>15.3390</td>
</tr>
<tr>
<td>25</td>
<td>3.3763</td>
<td>1.2136</td>
<td>1.6595</td>
<td>2.0771</td>
<td>3.5564</td>
<td>5.2563</td>
<td>15.1920</td>
</tr>
<tr>
<td>40</td>
<td>3.4942</td>
<td>1.0850</td>
<td>1.5567</td>
<td>1.9390</td>
<td>3.1646</td>
<td>4.3629</td>
<td>13.5459</td>
</tr>
<tr>
<td>50</td>
<td>3.5363</td>
<td>1.0298</td>
<td>1.5223</td>
<td>1.8835</td>
<td>3.0381</td>
<td>4.0725</td>
<td>12.8922</td>
</tr>
<tr>
<td>60</td>
<td>3.5775</td>
<td>0.9830</td>
<td>1.4923</td>
<td>1.8360</td>
<td>2.9310</td>
<td>3.8070</td>
<td>12.3452</td>
</tr>
<tr>
<td>70</td>
<td>3.6179</td>
<td>0.9448</td>
<td>1.4661</td>
<td>1.7960</td>
<td>2.8324</td>
<td>3.5640</td>
<td>11.8937</td>
</tr>
<tr>
<td>80</td>
<td>3.6576</td>
<td>0.9141</td>
<td>1.4435</td>
<td>1.7631</td>
<td>2.7417</td>
<td>3.3430</td>
<td>11.5362</td>
</tr>
<tr>
<td>90</td>
<td>3.6968</td>
<td>0.8888</td>
<td>1.4240</td>
<td>1.7359</td>
<td>2.6586</td>
<td>3.1388</td>
<td>11.2591</td>
</tr>
<tr>
<td>100</td>
<td>3.7356</td>
<td>0.8667</td>
<td>1.4072</td>
<td>1.7129</td>
<td>2.5836</td>
<td>2.9509</td>
<td>11.0387</td>
</tr>
<tr>
<td>120</td>
<td>3.7848</td>
<td>0.8371</td>
<td>1.3865</td>
<td>1.6930</td>
<td>2.5152</td>
<td>2.7606</td>
<td>10.7799</td>
</tr>
<tr>
<td>( \infty )</td>
<td>3.8416</td>
<td>0.8155</td>
<td>1.3727</td>
<td>1.6746</td>
<td>2.4505</td>
<td>2.5955</td>
<td>10.5969</td>
</tr>
</tbody>
</table>
3. Chi-Square Distribution Table

The table entries (i.e., $\chi^2_{\nu, \alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $\chi^2_{\nu}$-distribution, where $\nu = df$ (i.e., area under the $\chi^2_{\nu}$-pdf on the right of $\chi^2_{\nu, \alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>0.150</th>
<th>0.100</th>
<th>0.075</th>
<th>0.050</th>
<th>0.025</th>
<th>0.010</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0723</td>
<td>2.7055</td>
<td>3.1701</td>
<td>3.8415</td>
<td>5.0239</td>
<td>6.6349</td>
<td>7.8794</td>
</tr>
<tr>
<td>2</td>
<td>3.7942</td>
<td>4.6052</td>
<td>5.1805</td>
<td>5.9915</td>
<td>7.3778</td>
<td>9.2103</td>
<td>10.5666</td>
</tr>
<tr>
<td>15</td>
<td>20.6030</td>
<td>22.3070</td>
<td>23.4522</td>
<td>24.9958</td>
<td>27.4883</td>
<td>30.5779</td>
<td>32.8013</td>
</tr>
<tr>
<td>17</td>
<td>22.9770</td>
<td>24.7960</td>
<td>25.9705</td>
<td>27.5871</td>
<td>30.1910</td>
<td>33.4087</td>
<td>35.7185</td>
</tr>
<tr>
<td>18</td>
<td>24.1555</td>
<td>25.9894</td>
<td>27.2178</td>
<td>28.8693</td>
<td>31.5264</td>
<td>34.8053</td>
<td>37.1565</td>
</tr>
<tr>
<td>19</td>
<td>25.3289</td>
<td>27.2036</td>
<td>28.4581</td>
<td>30.1435</td>
<td>32.8523</td>
<td>36.1909</td>
<td>38.5823</td>
</tr>
<tr>
<td>20</td>
<td>26.4976</td>
<td>28.4120</td>
<td>29.6920</td>
<td>31.4104</td>
<td>34.1696</td>
<td>37.5662</td>
<td>39.9969</td>
</tr>
<tr>
<td>21</td>
<td>27.6620</td>
<td>29.6151</td>
<td>30.9200</td>
<td>32.6706</td>
<td>35.4789</td>
<td>38.9322</td>
<td>41.4011</td>
</tr>
<tr>
<td>22</td>
<td>28.8225</td>
<td>30.8133</td>
<td>32.1424</td>
<td>33.9244</td>
<td>36.7807</td>
<td>40.2894</td>
<td>42.7957</td>
</tr>
<tr>
<td>23</td>
<td>29.9792</td>
<td>32.0069</td>
<td>33.3597</td>
<td>35.1725</td>
<td>38.0756</td>
<td>41.6384</td>
<td>44.1813</td>
</tr>
<tr>
<td>24</td>
<td>31.1325</td>
<td>33.1962</td>
<td>34.5723</td>
<td>36.4150</td>
<td>39.3641</td>
<td>42.9789</td>
<td>45.5855</td>
</tr>
<tr>
<td>25</td>
<td>32.2825</td>
<td>34.3816</td>
<td>35.7803</td>
<td>37.6525</td>
<td>40.6465</td>
<td>44.3141</td>
<td>46.9279</td>
</tr>
<tr>
<td>26</td>
<td>33.4295</td>
<td>35.6323</td>
<td>36.9841</td>
<td>38.8851</td>
<td>41.9323</td>
<td>45.6417</td>
<td>48.2899</td>
</tr>
<tr>
<td>27</td>
<td>34.5736</td>
<td>36.7412</td>
<td>38.1840</td>
<td>40.1133</td>
<td>43.1945</td>
<td>46.9629</td>
<td>49.6449</td>
</tr>
<tr>
<td>28</td>
<td>35.7150</td>
<td>37.9159</td>
<td>39.3801</td>
<td>41.3371</td>
<td>44.4608</td>
<td>48.2782</td>
<td>50.9934</td>
</tr>
<tr>
<td>29</td>
<td>36.8538</td>
<td>39.0875</td>
<td>40.5727</td>
<td>42.5570</td>
<td>45.7223</td>
<td>49.8579</td>
<td>52.3536</td>
</tr>
<tr>
<td>30</td>
<td>37.9903</td>
<td>40.2560</td>
<td>41.7619</td>
<td>43.7730</td>
<td>46.9792</td>
<td>50.8922</td>
<td>53.6720</td>
</tr>
<tr>
<td>40</td>
<td>49.2439</td>
<td>51.8501</td>
<td>53.5010</td>
<td>55.7585</td>
<td>59.3417</td>
<td>63.6907</td>
<td>66.7660</td>
</tr>
<tr>
<td>50</td>
<td>50.3460</td>
<td>63.1671</td>
<td>65.0303</td>
<td>67.5048</td>
<td>71.4202</td>
<td>76.1539</td>
<td>79.4900</td>
</tr>
<tr>
<td>60</td>
<td>61.3411</td>
<td>74.3970</td>
<td>76.4113</td>
<td>79.0819</td>
<td>83.2977</td>
<td>88.3794</td>
<td>91.9517</td>
</tr>
<tr>
<td>70</td>
<td>72.5548</td>
<td>85.5270</td>
<td>87.6802</td>
<td>90.5312</td>
<td>95.0232</td>
<td>100.4252</td>
<td>104.2149</td>
</tr>
<tr>
<td>80</td>
<td>83.1058</td>
<td>96.5782</td>
<td>98.8606</td>
<td>101.8795</td>
<td>106.6286</td>
<td>112.3288</td>
<td>116.3211</td>
</tr>
<tr>
<td>90</td>
<td>93.2041</td>
<td>107.5650</td>
<td>109.9688</td>
<td>113.1453</td>
<td>118.1359</td>
<td>124.1163</td>
<td>128.2989</td>
</tr>
<tr>
<td>100</td>
<td>114.6588</td>
<td>118.4980</td>
<td>121.0166</td>
<td>124.3421</td>
<td>129.5612</td>
<td>135.8067</td>
<td>140.1695</td>
</tr>
<tr>
<td>120</td>
<td>136.0620</td>
<td>140.2326</td>
<td>142.9646</td>
<td>146.5674</td>
<td>152.2114</td>
<td>158.9502</td>
<td>163.6482</td>
</tr>
</tbody>
</table>
3. Chi-Square Distribution Table (cont.)
The table entries (i.e., $\chi^2_{\nu, \alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $\chi^2_{\nu}$-distribution, where $\nu = df$ (i.e., area under the $\chi^2_{\nu}$-pdf on the right of $\chi^2_{\nu, \alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>df = $\nu$</th>
<th>0.995</th>
<th>0.990</th>
<th>0.975</th>
<th>0.950</th>
<th>0.925</th>
<th>0.900</th>
<th>0.850</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0010</td>
<td>0.0039</td>
<td>0.0089</td>
<td>0.0158</td>
<td>0.0358</td>
</tr>
<tr>
<td>2</td>
<td>0.0100</td>
<td>0.0201</td>
<td>0.0506</td>
<td>0.1026</td>
<td>0.1559</td>
<td>0.2107</td>
<td>0.3250</td>
</tr>
<tr>
<td>3</td>
<td>0.0717</td>
<td>0.1148</td>
<td>0.2158</td>
<td>0.3519</td>
<td>0.4720</td>
<td>0.5844</td>
<td>0.7978</td>
</tr>
<tr>
<td>4</td>
<td>0.2070</td>
<td>0.2971</td>
<td>0.4844</td>
<td>0.7107</td>
<td>0.8969</td>
<td>1.0636</td>
<td>1.3665</td>
</tr>
<tr>
<td>5</td>
<td>0.4117</td>
<td>0.5543</td>
<td>0.8312</td>
<td>1.1455</td>
<td>1.3937</td>
<td>1.6103</td>
<td>1.9938</td>
</tr>
<tr>
<td>6</td>
<td>0.6757</td>
<td>0.8721</td>
<td>1.2373</td>
<td>1.6354</td>
<td>1.9415</td>
<td>2.2041</td>
<td>2.6613</td>
</tr>
<tr>
<td>7</td>
<td>0.9893</td>
<td>1.2390</td>
<td>1.6899</td>
<td>2.1673</td>
<td>2.5277</td>
<td>2.8331</td>
<td>3.3583</td>
</tr>
<tr>
<td>8</td>
<td>1.3444</td>
<td>1.6465</td>
<td>2.1797</td>
<td>2.7326</td>
<td>3.1440</td>
<td>3.4895</td>
<td>4.0782</td>
</tr>
<tr>
<td>9</td>
<td>1.7349</td>
<td>2.0879</td>
<td>2.7004</td>
<td>3.3251</td>
<td>3.8787</td>
<td>4.1682</td>
<td>4.8165</td>
</tr>
<tr>
<td>10</td>
<td>2.1559</td>
<td>2.5582</td>
<td>3.2470</td>
<td>3.9403</td>
<td>4.4459</td>
<td>4.8652</td>
<td>5.5701</td>
</tr>
<tr>
<td>11</td>
<td>2.6032</td>
<td>3.0535</td>
<td>3.8158</td>
<td>4.5748</td>
<td>5.1234</td>
<td>5.5778</td>
<td>6.3364</td>
</tr>
<tr>
<td>12</td>
<td>3.0738</td>
<td>3.5706</td>
<td>4.4038</td>
<td>5.2260</td>
<td>5.8175</td>
<td>6.3038</td>
<td>7.1138</td>
</tr>
<tr>
<td>13</td>
<td>3.5650</td>
<td>4.0969</td>
<td>5.0088</td>
<td>5.8919</td>
<td>6.5238</td>
<td>7.0415</td>
<td>7.9008</td>
</tr>
<tr>
<td>16</td>
<td>5.1422</td>
<td>5.8122</td>
<td>6.9077</td>
<td>7.9617</td>
<td>8.7067</td>
<td>9.3122</td>
<td>10.3090</td>
</tr>
<tr>
<td>50</td>
<td>27.9907</td>
<td>29.7067</td>
<td>32.3574</td>
<td>34.7643</td>
<td>36.3971</td>
<td>37.6868</td>
<td>30.8563</td>
</tr>
<tr>
<td>60</td>
<td>35.5345</td>
<td>37.4849</td>
<td>40.4817</td>
<td>43.1880</td>
<td>45.0165</td>
<td>46.4589</td>
<td>48.7587</td>
</tr>
<tr>
<td>70</td>
<td>43.2752</td>
<td>45.4117</td>
<td>48.7576</td>
<td>51.7393</td>
<td>53.7478</td>
<td>55.3289</td>
<td>57.8443</td>
</tr>
<tr>
<td>80</td>
<td>51.1719</td>
<td>53.5401</td>
<td>57.1532</td>
<td>60.3915</td>
<td>62.5676</td>
<td>64.2778</td>
<td>66.9938</td>
</tr>
<tr>
<td>90</td>
<td>59.1963</td>
<td>61.7541</td>
<td>65.6466</td>
<td>69.1260</td>
<td>71.4596</td>
<td>73.2911</td>
<td>76.1954</td>
</tr>
<tr>
<td>100</td>
<td>67.3276</td>
<td>70.0649</td>
<td>74.2219</td>
<td>77.9295</td>
<td>80.4139</td>
<td>82.3581</td>
<td>85.4406</td>
</tr>
<tr>
<td>120</td>
<td>83.8516</td>
<td>86.9233</td>
<td>91.5726</td>
<td>95.7046</td>
<td>98.4641</td>
<td>100.6236</td>
<td>104.0374</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 =$ numerator $df$ and $\nu_2 =$ denominator $df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>161.4476</td>
<td>199.5000</td>
<td>215.7073</td>
<td>224.5832</td>
<td>230.1619</td>
<td>233.9860</td>
</tr>
<tr>
<td>4</td>
<td>6.6079</td>
<td>5.7861</td>
<td>5.4095</td>
<td>5.1922</td>
<td>5.0503</td>
<td>4.9503</td>
<td>4.8759</td>
</tr>
<tr>
<td>5</td>
<td>5.9874</td>
<td>5.1433</td>
<td>4.7571</td>
<td>4.5337</td>
<td>4.3874</td>
<td>4.2839</td>
<td>4.2067</td>
</tr>
<tr>
<td>6</td>
<td>5.5914</td>
<td>4.7374</td>
<td>4.3468</td>
<td>4.1203</td>
<td>3.9715</td>
<td>3.8660</td>
<td>3.7870</td>
</tr>
<tr>
<td>7</td>
<td>5.3177</td>
<td>4.4590</td>
<td>4.0662</td>
<td>3.8379</td>
<td>3.6875</td>
<td>3.5806</td>
<td>3.5005</td>
</tr>
<tr>
<td>8</td>
<td>5.1174</td>
<td>4.2565</td>
<td>3.8625</td>
<td>3.6331</td>
<td>3.4817</td>
<td>3.3738</td>
<td>3.2927</td>
</tr>
<tr>
<td>10</td>
<td>4.8443</td>
<td>3.9823</td>
<td>3.5874</td>
<td>3.3567</td>
<td>3.2039</td>
<td>3.0946</td>
<td>3.0123</td>
</tr>
<tr>
<td>12</td>
<td>4.6672</td>
<td>3.8056</td>
<td>3.4105</td>
<td>3.1791</td>
<td>3.0254</td>
<td>2.9153</td>
<td>2.8321</td>
</tr>
<tr>
<td>13</td>
<td>4.6001</td>
<td>3.7389</td>
<td>3.3439</td>
<td>3.1122</td>
<td>2.9582</td>
<td>2.8477</td>
<td>2.7642</td>
</tr>
<tr>
<td>14</td>
<td>4.5431</td>
<td>3.6823</td>
<td>3.2874</td>
<td>3.0556</td>
<td>2.9013</td>
<td>2.7905</td>
<td>2.7066</td>
</tr>
<tr>
<td>15</td>
<td>4.4940</td>
<td>3.6337</td>
<td>3.2389</td>
<td>3.0069</td>
<td>2.8524</td>
<td>2.7413</td>
<td>2.6572</td>
</tr>
<tr>
<td>16</td>
<td>4.4513</td>
<td>3.5915</td>
<td>3.1968</td>
<td>2.9647</td>
<td>2.8100</td>
<td>2.6987</td>
<td>2.6143</td>
</tr>
<tr>
<td>17</td>
<td>4.4139</td>
<td>3.5546</td>
<td>3.1599</td>
<td>2.9277</td>
<td>2.7729</td>
<td>2.6613</td>
<td>2.5767</td>
</tr>
<tr>
<td>18</td>
<td>4.3807</td>
<td>3.5219</td>
<td>3.1274</td>
<td>2.8951</td>
<td>2.7401</td>
<td>2.6283</td>
<td>2.5435</td>
</tr>
<tr>
<td>19</td>
<td>4.3512</td>
<td>3.4928</td>
<td>3.0984</td>
<td>2.8661</td>
<td>2.7109</td>
<td>2.5990</td>
<td>2.5140</td>
</tr>
<tr>
<td>20</td>
<td>4.1799</td>
<td>3.3158</td>
<td>2.9223</td>
<td>2.6896</td>
<td>2.5336</td>
<td>2.4205</td>
<td>2.3343</td>
</tr>
<tr>
<td>30</td>
<td>4.0847</td>
<td>3.2317</td>
<td>2.8387</td>
<td>2.6060</td>
<td>2.4495</td>
<td>2.3359</td>
<td>2.2490</td>
</tr>
<tr>
<td>50</td>
<td>4.0343</td>
<td>3.1826</td>
<td>2.7900</td>
<td>2.5572</td>
<td>2.4004</td>
<td>2.2864</td>
<td>2.1992</td>
</tr>
<tr>
<td>60</td>
<td>4.0012</td>
<td>3.1504</td>
<td>2.7581</td>
<td>2.5252</td>
<td>2.3683</td>
<td>2.2541</td>
<td>2.1665</td>
</tr>
<tr>
<td>80</td>
<td>3.9604</td>
<td>3.1108</td>
<td>2.7188</td>
<td>2.4859</td>
<td>2.3287</td>
<td>2.2142</td>
<td>2.1263</td>
</tr>
<tr>
<td>100</td>
<td>3.9361</td>
<td>3.0873</td>
<td>2.6955</td>
<td>2.4626</td>
<td>2.3053</td>
<td>2.1906</td>
<td>2.1025</td>
</tr>
<tr>
<td>120</td>
<td>3.9201</td>
<td>3.0718</td>
<td>2.6802</td>
<td>2.4472</td>
<td>2.2899</td>
<td>2.1750</td>
<td>2.0868</td>
</tr>
<tr>
<td>$\infty$</td>
<td>3.8415</td>
<td>2.9957</td>
<td>2.6049</td>
<td>2.3719</td>
<td>2.2141</td>
<td>2.0986</td>
<td>2.0096</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) *F*-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1, \nu_2, \alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1, \nu_2}$-distribution, where $\nu_1 = \text{numerator df}$ and $\nu_2 = \text{denominator df}$ (i.e., area under the $F_{\nu_1, \nu_2}$-pdf on the right of $F_{\nu_1, \nu_2, \alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator df $= \nu_2$</th>
<th>( \alpha = 0.05 )</th>
<th>Numerator df $= \nu_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>238.83</td>
<td>240.54</td>
</tr>
<tr>
<td>3</td>
<td>8.8452</td>
<td>8.8123</td>
</tr>
<tr>
<td>4</td>
<td>6.0410</td>
<td>5.9988</td>
</tr>
<tr>
<td>10</td>
<td>3.0717</td>
<td>3.0204</td>
</tr>
<tr>
<td>11</td>
<td>2.9480</td>
<td>2.8962</td>
</tr>
<tr>
<td>12</td>
<td>2.8486</td>
<td>2.7964</td>
</tr>
<tr>
<td>13</td>
<td>2.7669</td>
<td>2.7144</td>
</tr>
<tr>
<td>14</td>
<td>2.6987</td>
<td>2.6458</td>
</tr>
<tr>
<td>15</td>
<td>2.6408</td>
<td>2.5876</td>
</tr>
<tr>
<td>16</td>
<td>2.5911</td>
<td>2.5377</td>
</tr>
<tr>
<td>17</td>
<td>2.5480</td>
<td>2.4943</td>
</tr>
<tr>
<td>18</td>
<td>2.5102</td>
<td>2.4563</td>
</tr>
<tr>
<td>19</td>
<td>2.4768</td>
<td>2.4227</td>
</tr>
<tr>
<td>20</td>
<td>2.4471</td>
<td>2.3928</td>
</tr>
<tr>
<td>30</td>
<td>2.2662</td>
<td>2.2107</td>
</tr>
<tr>
<td>40</td>
<td>2.1802</td>
<td>2.1240</td>
</tr>
<tr>
<td>50</td>
<td>2.1299</td>
<td>2.0734</td>
</tr>
<tr>
<td>60</td>
<td>2.0970</td>
<td>2.0401</td>
</tr>
<tr>
<td>80</td>
<td>2.0564</td>
<td>1.9991</td>
</tr>
<tr>
<td>100</td>
<td>2.0323</td>
<td>1.9748</td>
</tr>
<tr>
<td>120</td>
<td>2.0164</td>
<td>1.9588</td>
</tr>
<tr>
<td>100</td>
<td>1.9384</td>
<td>1.8799</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 =$ numerator $df$ and $\nu_2 =$ denominator $df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>245.949</td>
<td>246.463</td>
<td>246.918</td>
<td>247.323</td>
<td>247.686</td>
<td>248.013</td>
<td>250.095</td>
</tr>
<tr>
<td>3</td>
<td>8.7029</td>
<td>8.6923</td>
<td>8.6829</td>
<td>8.6745</td>
<td>8.6670</td>
<td>8.6602</td>
<td>8.6166</td>
</tr>
<tr>
<td>4</td>
<td>5.8578</td>
<td>5.8441</td>
<td>5.8320</td>
<td>5.8211</td>
<td>5.8114</td>
<td>5.8025</td>
<td>5.7459</td>
</tr>
<tr>
<td>9</td>
<td>3.0061</td>
<td>2.9890</td>
<td>2.9737</td>
<td>2.9600</td>
<td>2.9477</td>
<td>2.9365</td>
<td>2.8637</td>
</tr>
<tr>
<td>10</td>
<td>2.8450</td>
<td>2.8276</td>
<td>2.8120</td>
<td>2.7980</td>
<td>2.7854</td>
<td>2.7740</td>
<td>2.6996</td>
</tr>
<tr>
<td>11</td>
<td>2.7186</td>
<td>2.7009</td>
<td>2.6851</td>
<td>2.6709</td>
<td>2.6581</td>
<td>2.6464</td>
<td>2.5705</td>
</tr>
<tr>
<td>12</td>
<td>2.6169</td>
<td>2.5989</td>
<td>2.5828</td>
<td>2.5684</td>
<td>2.5554</td>
<td>2.5436</td>
<td>2.4663</td>
</tr>
<tr>
<td>13</td>
<td>2.5331</td>
<td>2.5149</td>
<td>2.4987</td>
<td>2.4841</td>
<td>2.4709</td>
<td>2.4589</td>
<td>2.3803</td>
</tr>
<tr>
<td>14</td>
<td>2.4630</td>
<td>2.4446</td>
<td>2.4282</td>
<td>2.4134</td>
<td>2.4000</td>
<td>2.3879</td>
<td>2.3082</td>
</tr>
<tr>
<td>15</td>
<td>2.4034</td>
<td>2.3849</td>
<td>2.3680</td>
<td>2.3533</td>
<td>2.3398</td>
<td>2.3275</td>
<td>2.2468</td>
</tr>
<tr>
<td>16</td>
<td>2.3522</td>
<td>2.3335</td>
<td>2.3167</td>
<td>2.3016</td>
<td>2.2880</td>
<td>2.2756</td>
<td>2.1938</td>
</tr>
<tr>
<td>17</td>
<td>2.3077</td>
<td>2.2888</td>
<td>2.2719</td>
<td>2.2567</td>
<td>2.2429</td>
<td>2.2304</td>
<td>2.1477</td>
</tr>
<tr>
<td>18</td>
<td>2.2686</td>
<td>2.2496</td>
<td>2.2325</td>
<td>2.2172</td>
<td>2.2033</td>
<td>2.1906</td>
<td>2.1071</td>
</tr>
<tr>
<td>19</td>
<td>2.2341</td>
<td>2.2149</td>
<td>2.1977</td>
<td>2.1823</td>
<td>2.1683</td>
<td>2.1555</td>
<td>2.0712</td>
</tr>
<tr>
<td>20</td>
<td>2.2033</td>
<td>2.1840</td>
<td>2.1667</td>
<td>2.1511</td>
<td>2.1370</td>
<td>2.1242</td>
<td>2.0391</td>
</tr>
<tr>
<td>30</td>
<td>2.0148</td>
<td>1.9946</td>
<td>1.9765</td>
<td>1.9601</td>
<td>1.9452</td>
<td>1.9317</td>
<td>1.8409</td>
</tr>
<tr>
<td>40</td>
<td>1.9245</td>
<td>1.9037</td>
<td>1.8851</td>
<td>1.8682</td>
<td>1.8529</td>
<td>1.8389</td>
<td>1.7444</td>
</tr>
<tr>
<td>50</td>
<td>1.8714</td>
<td>1.8503</td>
<td>1.8313</td>
<td>1.8141</td>
<td>1.7985</td>
<td>1.7841</td>
<td>1.6872</td>
</tr>
<tr>
<td>60</td>
<td>1.8364</td>
<td>1.8151</td>
<td>1.7959</td>
<td>1.7784</td>
<td>1.7625</td>
<td>1.7480</td>
<td>1.6491</td>
</tr>
<tr>
<td>80</td>
<td>1.7932</td>
<td>1.7716</td>
<td>1.7520</td>
<td>1.7342</td>
<td>1.7180</td>
<td>1.7032</td>
<td>1.6017</td>
</tr>
<tr>
<td>100</td>
<td>1.7675</td>
<td>1.7456</td>
<td>1.7259</td>
<td>1.7079</td>
<td>1.6915</td>
<td>1.6764</td>
<td>1.5733</td>
</tr>
<tr>
<td>120</td>
<td>1.7505</td>
<td>1.7285</td>
<td>1.7085</td>
<td>1.6904</td>
<td>1.6739</td>
<td>1.6587</td>
<td>1.5543</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.6664</td>
<td>1.6435</td>
<td>1.6228</td>
<td>1.6038</td>
<td>1.5865</td>
<td>1.5705</td>
<td>1.4591</td>
</tr>
</tbody>
</table>
The table entries (i.e., $F_{\nu_1, \nu_2, \alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1, \nu_2}$-distribution, where $\nu_1 = \text{numerator df}$ and $\nu_2 = \text{denominator df}$ (i.e., area under the $F_{\nu_1, \nu_2}$-pdf on the right of $F_{\nu_1, \nu_2, \alpha}$ is $\alpha$).

$$\begin{array}{ccccccccc}
\text{Denominator df} = \nu_2 & 40 & 50 & 60 & 80 & 100 & 120 & \infty \\
\hline
4 & 5.7170 & 5.6995 & 5.6877 & 5.6641 & 5.6581 & 5.6281 & & \\
8 & 3.0428 & 3.0204 & 2.9862 & 2.9747 & 2.9669 & 2.9276 & & \\
9 & 2.8259 & 2.8028 & 2.7675 & 2.7556 & 2.7475 & 2.7067 & & \\
10 & 2.6609 & 2.6371 & 2.6111 & 2.5884 & 2.5801 & 2.5379 & & \\
11 & 2.5309 & 2.5066 & 2.4901 & 2.4692 & 2.4566 & 2.4045 & & \\
12 & 2.4259 & 2.4010 & 2.3842 & 2.3628 & 2.3498 & 2.3106 & & \\
13 & 2.3392 & 2.3138 & 2.2966 & 2.2747 & 2.2614 & 2.2064 & & \\
14 & 2.2664 & 2.2405 & 2.2229 & 2.2006 & 2.1870 & 2.1307 & & \\
15 & 2.2043 & 2.1780 & 2.1601 & 2.1373 & 2.1234 & 2.0658 & & \\
16 & 2.1507 & 2.1240 & 2.1058 & 2.0826 & 2.0685 & 2.0096 & & \\
17 & 2.1040 & 2.0769 & 2.0584 & 2.0348 & 2.0204 & 1.9604 & & \\
18 & 2.0629 & 2.0354 & 2.0166 & 1.9927 & 1.9780 & 1.9168 & & \\
19 & 2.0264 & 1.9986 & 1.9795 & 1.9552 & 1.9403 & 1.8780 & & \\
20 & 1.9938 & 1.9656 & 1.9464 & 1.9217 & 1.9066 & 1.8432 & & \\
21 & 1.9718 & 1.9609 & 1.9396 & 1.9121 & 1.8965 & 1.8223 & & \\
22 & 1.9518 & 1.9600 & 1.9373 & 1.9077 & 1.8912 & 1.8183 & & \\
23 & 1.9283 & 1.9595 & 1.9357 & 1.9045 & 1.8879 & 1.8148 & & \\
24 & 1.9054 & 1.9590 & 1.9343 & 1.9019 & 1.8843 & 1.8113 & & \\
25 & 1.8832 & 1.9581 & 1.9327 & 1.9001 & 1.8825 & 1.8094 & & \\
26 & 1.8619 & 1.9531 & 1.9271 & 1.8944 & 1.8767 & 1.8037 & & \\
27 & 1.8413 & 1.9482 & 1.9216 & 1.8886 & 1.8708 & 1.7977 & & \\
28 & 1.8214 & 1.9433 & 1.9164 & 1.8834 & 1.8655 & 1.7925 & & \\
29 & 1.8022 & 1.9385 & 1.9113 & 1.8783 & 1.8604 & 1.7893 & & \\
30 & 1.7837 & 1.9337 & 1.9061 & 1.8731 & 1.8552 & 1.7841 & & \\
31 & 1.7659 & 1.9290 & 1.8913 & 1.8583 & 1.8395 & 1.7684 & & \\
32 & 1.7488 & 1.9244 & 1.8865 & 1.8535 & 1.8346 & 1.7633 & & \\
33 & 1.7324 & 1.9199 & 1.8818 & 1.8487 & 1.8298 & 1.7582 & & \\
34 & 1.7166 & 1.9155 & 1.8777 & 1.8446 & 1.8258 & 1.7544 & & \\
35 & 1.7015 & 1.9112 & 1.8797 & 1.8465 & 1.8278 & 1.7539 & & \\
36 & 1.6869 & 1.9070 & 1.8718 & 1.8387 & 1.8198 & 1.7491 & & \\
37 & 1.6729 & 1.9030 & 1.8639 & 1.8308 & 1.8119 & 1.7435 & & \\
38 & 1.6595 & 1.8991 & 1.8560 & 1.8238 & 1.8049 & 1.7393 & & \\
39 & 1.6466 & 1.8953 & 1.8481 & 1.8158 & 1.7969 & 1.7354 & & \\
40 & 1.6341 & 1.8916 & 1.8404 & 1.8124 & 1.7935 & 1.7329 & & \\
45 & 1.5930 & 1.8614 & 1.7988 & 1.7660 & 1.7470 & 1.6858 & & \\
50 & 1.5737 & 1.8522 & 1.7895 & 1.7564 & 1.7374 & 1.6761 & & \\
60 & 1.5553 & 1.8432 & 1.7805 & 1.7470 & 1.7280 & 1.6666 & & \\
80 & 1.5374 & 1.8343 & 1.7716 & 1.7380 & 1.7190 & 1.6575 & & \\
100 & 1.5203 & 1.8256 & 1.7628 & 1.7289 & 1.7100 & 1.6484 & & \\
120 & 1.5041 & 1.8170 & 1.7540 & 1.7200 & 1.6913 & 1.6297 & & \\
\infty & 1.3940 & 1.3501 & 1.3180 & 1.2735 & 1.2434 & 1.2214 & 1.0000
\end{array}
4. (Snedecor’s) F-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 =$ numerator $df$ and $\nu_2 =$ denominator $df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>647.7890</td>
<td>799.5000</td>
<td>864.1630</td>
<td>899.5833</td>
<td>921.8479</td>
<td>937.1111</td>
<td>948.2169</td>
</tr>
<tr>
<td>6</td>
<td>8.8131</td>
<td>7.2599</td>
<td>6.5988</td>
<td>6.2272</td>
<td>5.9876</td>
<td>5.8198</td>
<td>5.6955</td>
</tr>
<tr>
<td>7</td>
<td>8.0727</td>
<td>6.5415</td>
<td>5.8898</td>
<td>5.5226</td>
<td>5.2852</td>
<td>5.1186</td>
<td>4.9949</td>
</tr>
<tr>
<td>8</td>
<td>7.5709</td>
<td>6.0595</td>
<td>5.4160</td>
<td>5.0526</td>
<td>4.8173</td>
<td>4.6517</td>
<td>4.5286</td>
</tr>
<tr>
<td>19</td>
<td>5.9216</td>
<td>4.5075</td>
<td>3.9034</td>
<td>3.5587</td>
<td>3.3327</td>
<td>3.1718</td>
<td>3.0509</td>
</tr>
<tr>
<td>20</td>
<td>5.8715</td>
<td>4.4613</td>
<td>3.8587</td>
<td>3.5147</td>
<td>3.2891</td>
<td>3.1283</td>
<td>3.0074</td>
</tr>
<tr>
<td>30</td>
<td>5.5675</td>
<td>4.1821</td>
<td>3.5894</td>
<td>3.2499</td>
<td>3.0265</td>
<td>2.8667</td>
<td>2.7460</td>
</tr>
<tr>
<td>50</td>
<td>5.4239</td>
<td>4.0510</td>
<td>3.4633</td>
<td>3.1261</td>
<td>2.9037</td>
<td>2.7444</td>
<td>2.6238</td>
</tr>
<tr>
<td>100</td>
<td>5.2184</td>
<td>3.8643</td>
<td>3.2841</td>
<td>2.9504</td>
<td>2.7295</td>
<td>2.6736</td>
<td>2.5502</td>
</tr>
<tr>
<td>120</td>
<td>5.1523</td>
<td>3.8046</td>
<td>3.2269</td>
<td>2.8943</td>
<td>2.6740</td>
<td>2.5154</td>
<td>2.3948</td>
</tr>
<tr>
<td>$\infty$</td>
<td>5.0239</td>
<td>3.6889</td>
<td>3.1161</td>
<td>2.7858</td>
<td>2.5665</td>
<td>2.4082</td>
<td>2.2875</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) F-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1, \nu_2, \alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1, \nu_2}$-distribution, where $\nu_1 =$ numerator df and $\nu_2 =$ denominator df (i.e., area under the $F_{\nu_1, \nu_2}$-pdf on the right of $F_{\nu_1, \nu_2, \alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>$\nu_2$</th>
<th>$\alpha = 0.025$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>965.6562 963.2846 968.6274 973.0252 976.7079</td>
</tr>
<tr>
<td>11</td>
<td>8.9796 8.9047 8.8439 8.7935 8.7512</td>
</tr>
<tr>
<td>14</td>
<td>5.5996 5.5234 5.4613 5.4098 5.3662</td>
</tr>
<tr>
<td>20</td>
<td>3.8549 3.7790 3.7168 3.6649 3.6209</td>
</tr>
<tr>
<td>50</td>
<td>3.2853 3.2093 3.1469 3.0946 3.0502</td>
</tr>
<tr>
<td>60</td>
<td>3.1987 3.1227 3.0602 3.0078 2.9633</td>
</tr>
<tr>
<td>80</td>
<td>3.1248 3.0488 2.9862 2.9337 2.8890</td>
</tr>
<tr>
<td>100</td>
<td>3.0610 2.9849 2.9222 2.8696 2.8249</td>
</tr>
<tr>
<td>120</td>
<td>3.0053 2.9291 2.8664 2.8137 2.7689</td>
</tr>
<tr>
<td>150</td>
<td>2.9563 2.8801 2.8172 2.7645 2.7196</td>
</tr>
<tr>
<td>200</td>
<td>2.9128 2.8365 2.7737 2.7209 2.6758</td>
</tr>
<tr>
<td>300</td>
<td>2.6513 2.5746 2.5112 2.4577 2.4120</td>
</tr>
<tr>
<td>500</td>
<td>2.5289 2.4519 2.3882 2.3343 2.2882</td>
</tr>
<tr>
<td>1000</td>
<td>2.4579 2.3808 2.3168 2.2627 2.2162</td>
</tr>
<tr>
<td>10000</td>
<td>2.4117 2.3344 2.2702 2.2159 2.1692</td>
</tr>
<tr>
<td>$\infty$</td>
<td>2.3549 2.2775 2.2130 2.1584 2.1115</td>
</tr>
<tr>
<td>10000</td>
<td>2.3215 2.2439 2.1793 2.1245 2.0773</td>
</tr>
<tr>
<td>12000</td>
<td>2.2994 2.2217 2.1570 2.1021 2.0548</td>
</tr>
<tr>
<td>$\infty$</td>
<td>2.1918 2.1136 2.0483 1.9927 1.9447</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 =$ numerator $df$ and $\nu_2 =$ denominator $df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>984.8668</td>
<td>986.9187</td>
<td>988.7331</td>
<td>990.3490</td>
<td>991.7973</td>
<td>993.1028</td>
<td>1001.414</td>
</tr>
<tr>
<td>4</td>
<td>8.6565</td>
<td>8.6326</td>
<td>8.6113</td>
<td>8.5924</td>
<td>8.5753</td>
<td>8.5599</td>
<td>8.4613</td>
</tr>
<tr>
<td>6</td>
<td>5.2687</td>
<td>5.2439</td>
<td>5.2218</td>
<td>5.2021</td>
<td>5.1844</td>
<td>5.1684</td>
<td>5.0652</td>
</tr>
<tr>
<td>7</td>
<td>4.5678</td>
<td>4.5428</td>
<td>4.5206</td>
<td>4.5008</td>
<td>4.4829</td>
<td>4.4667</td>
<td>4.3624</td>
</tr>
<tr>
<td>13</td>
<td>3.0527</td>
<td>3.0269</td>
<td>3.0039</td>
<td>2.9832</td>
<td>2.9646</td>
<td>2.9477</td>
<td>2.8372</td>
</tr>
<tr>
<td>14</td>
<td>2.9493</td>
<td>2.9234</td>
<td>2.9003</td>
<td>2.8795</td>
<td>2.8607</td>
<td>2.8437</td>
<td>2.7324</td>
</tr>
<tr>
<td>15</td>
<td>2.8621</td>
<td>2.8360</td>
<td>2.8128</td>
<td>2.7919</td>
<td>2.7730</td>
<td>2.7559</td>
<td>2.6437</td>
</tr>
<tr>
<td>16</td>
<td>2.7875</td>
<td>2.7614</td>
<td>2.7380</td>
<td>2.7170</td>
<td>2.6980</td>
<td>2.6808</td>
<td>2.5678</td>
</tr>
<tr>
<td>17</td>
<td>2.7230</td>
<td>2.6968</td>
<td>2.6733</td>
<td>2.6522</td>
<td>2.6331</td>
<td>2.6158</td>
<td>2.5020</td>
</tr>
<tr>
<td>18</td>
<td>2.6667</td>
<td>2.6404</td>
<td>2.6168</td>
<td>2.5956</td>
<td>2.5764</td>
<td>2.5590</td>
<td>2.4445</td>
</tr>
<tr>
<td>19</td>
<td>2.6171</td>
<td>2.5907</td>
<td>2.5670</td>
<td>2.5457</td>
<td>2.5265</td>
<td>2.5089</td>
<td>2.3937</td>
</tr>
<tr>
<td>20</td>
<td>2.5731</td>
<td>2.5465</td>
<td>2.5228</td>
<td>2.5014</td>
<td>2.4821</td>
<td>2.4645</td>
<td>2.3486</td>
</tr>
<tr>
<td>30</td>
<td>2.3072</td>
<td>2.2799</td>
<td>2.2554</td>
<td>2.2334</td>
<td>2.2134</td>
<td>2.1952</td>
<td>2.0739</td>
</tr>
<tr>
<td>40</td>
<td>2.1819</td>
<td>2.1542</td>
<td>2.1293</td>
<td>2.1068</td>
<td>2.0864</td>
<td>2.0677</td>
<td>1.9429</td>
</tr>
<tr>
<td>50</td>
<td>2.1090</td>
<td>2.0810</td>
<td>2.0558</td>
<td>2.0330</td>
<td>2.0122</td>
<td>1.9933</td>
<td>1.8659</td>
</tr>
<tr>
<td>60</td>
<td>2.0613</td>
<td>2.0330</td>
<td>2.0076</td>
<td>1.9846</td>
<td>1.9636</td>
<td>1.9445</td>
<td>1.8152</td>
</tr>
<tr>
<td>80</td>
<td>2.0226</td>
<td>1.9741</td>
<td>1.9483</td>
<td>1.9250</td>
<td>1.9037</td>
<td>1.8843</td>
<td>1.7523</td>
</tr>
<tr>
<td>100</td>
<td>1.9679</td>
<td>1.9391</td>
<td>1.9132</td>
<td>1.8897</td>
<td>1.8682</td>
<td>1.8486</td>
<td>1.7148</td>
</tr>
<tr>
<td>120</td>
<td>1.9450</td>
<td>1.9161</td>
<td>1.8900</td>
<td>1.8663</td>
<td>1.8474</td>
<td>1.8249</td>
<td>1.6899</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.8326</td>
<td>1.8028</td>
<td>1.7759</td>
<td>1.7515</td>
<td>1.7291</td>
<td>1.7085</td>
<td>1.5660</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 =$ numerator $df$ and $\nu_2 =$ denominator $df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>$\alpha = 0.025$</th>
<th>Numerator $df = \nu_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>1005.596</td>
<td>1008.098</td>
</tr>
<tr>
<td>4</td>
<td>8.4111</td>
<td>8.3808</td>
</tr>
<tr>
<td>6</td>
<td>5.0125</td>
<td>4.9804</td>
</tr>
<tr>
<td>8</td>
<td>3.8398</td>
<td>3.8067</td>
</tr>
<tr>
<td>9</td>
<td>3.5055</td>
<td>3.4719</td>
</tr>
<tr>
<td>11</td>
<td>3.0613</td>
<td>3.0268</td>
</tr>
<tr>
<td>12</td>
<td>2.9063</td>
<td>2.8714</td>
</tr>
<tr>
<td>13</td>
<td>2.7797</td>
<td>2.7443</td>
</tr>
<tr>
<td>14</td>
<td>2.6740</td>
<td>2.6384</td>
</tr>
<tr>
<td>15</td>
<td>2.5850</td>
<td>2.5488</td>
</tr>
<tr>
<td>16</td>
<td>2.5085</td>
<td>2.4719</td>
</tr>
<tr>
<td>17</td>
<td>2.4422</td>
<td>2.4053</td>
</tr>
<tr>
<td>18</td>
<td>2.3842</td>
<td>2.3468</td>
</tr>
<tr>
<td>19</td>
<td>2.3329</td>
<td>2.2952</td>
</tr>
<tr>
<td>20</td>
<td>2.2873</td>
<td>2.2493</td>
</tr>
<tr>
<td>30</td>
<td>2.0089</td>
<td>1.9681</td>
</tr>
<tr>
<td>40</td>
<td>1.8752</td>
<td>1.8324</td>
</tr>
<tr>
<td>50</td>
<td>1.7963</td>
<td>1.7520</td>
</tr>
<tr>
<td>60</td>
<td>1.7440</td>
<td>1.6985</td>
</tr>
<tr>
<td>80</td>
<td>1.6790</td>
<td>1.6318</td>
</tr>
<tr>
<td>100</td>
<td>1.6401</td>
<td>1.5917</td>
</tr>
<tr>
<td>120</td>
<td>1.6141</td>
<td>1.5649</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.4835</td>
<td>1.4284</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 = \text{numerator } df$ and $\nu_2 = \text{denominator } df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>$\alpha = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4052.185</td>
</tr>
<tr>
<td>3</td>
<td>34.1162</td>
</tr>
<tr>
<td>10</td>
<td>10.0443</td>
</tr>
<tr>
<td>11</td>
<td>9.6460</td>
</tr>
<tr>
<td>12</td>
<td>9.3302</td>
</tr>
<tr>
<td>14</td>
<td>8.8616</td>
</tr>
<tr>
<td>19</td>
<td>8.1849</td>
</tr>
<tr>
<td>30</td>
<td>7.5625</td>
</tr>
<tr>
<td>60</td>
<td>7.0771</td>
</tr>
<tr>
<td>80</td>
<td>6.9627</td>
</tr>
<tr>
<td>100</td>
<td>6.8553</td>
</tr>
<tr>
<td>120</td>
<td>6.8509</td>
</tr>
<tr>
<td>$\infty$</td>
<td>6.6349</td>
</tr>
</tbody>
</table>
4. (Snedecor's) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 = \text{numerator df}$ and $\nu_2 = \text{denominator df}$ (i.e., area under the $F_{\nu_1,\nu_2}\text{-pdf}$ on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>(\alpha = 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Numerator $df = \nu_1$</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>5980.954</td>
</tr>
<tr>
<td>3</td>
<td>27.4892</td>
</tr>
<tr>
<td>6</td>
<td>8.1017</td>
</tr>
<tr>
<td>8</td>
<td>6.0289</td>
</tr>
<tr>
<td>9</td>
<td>5.4671</td>
</tr>
<tr>
<td>15</td>
<td>4.0045</td>
</tr>
<tr>
<td>30</td>
<td>3.1726</td>
</tr>
<tr>
<td>40</td>
<td>2.9930</td>
</tr>
<tr>
<td>50</td>
<td>2.8900</td>
</tr>
<tr>
<td>60</td>
<td>2.8233</td>
</tr>
<tr>
<td>80</td>
<td>2.7420</td>
</tr>
<tr>
<td>100</td>
<td>2.6943</td>
</tr>
<tr>
<td>120</td>
<td>2.6629</td>
</tr>
<tr>
<td>$\infty$</td>
<td>2.5113</td>
</tr>
</tbody>
</table>
The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 = numerator df$ and $\nu_2 = denominator df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>$\alpha = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>6156.974</td>
</tr>
<tr>
<td>6</td>
<td>9.7222</td>
</tr>
<tr>
<td>8</td>
<td>5.5151</td>
</tr>
<tr>
<td>20</td>
<td>3.0880</td>
</tr>
<tr>
<td>30</td>
<td>2.7002</td>
</tr>
<tr>
<td>40</td>
<td>2.5216</td>
</tr>
<tr>
<td>50</td>
<td>2.4190</td>
</tr>
<tr>
<td>60</td>
<td>2.3523</td>
</tr>
<tr>
<td>80</td>
<td>2.2709</td>
</tr>
<tr>
<td>100</td>
<td>2.2230</td>
</tr>
<tr>
<td>120</td>
<td>2.1915</td>
</tr>
<tr>
<td>$\infty$</td>
<td>2.0385</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) F-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1$ = numerator $df$ and $\nu_2$ = denominator $df$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Numerator $df = \nu_1$</th>
<th>$\alpha = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40</td>
</tr>
<tr>
<td>Denominator $df = \nu_2$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6286.427</td>
</tr>
<tr>
<td>7</td>
<td>5.9084</td>
</tr>
<tr>
<td>8</td>
<td>5.1156</td>
</tr>
<tr>
<td>16</td>
<td>3.0182</td>
</tr>
<tr>
<td>17</td>
<td>2.9205</td>
</tr>
<tr>
<td>18</td>
<td>2.9205</td>
</tr>
<tr>
<td>19</td>
<td>2.7608</td>
</tr>
<tr>
<td>20</td>
<td>2.6947</td>
</tr>
<tr>
<td>21</td>
<td>2.2992</td>
</tr>
<tr>
<td>22</td>
<td>2.1142</td>
</tr>
<tr>
<td>23</td>
<td>2.0066</td>
</tr>
<tr>
<td>24</td>
<td>1.9360</td>
</tr>
<tr>
<td>25</td>
<td>1.8489</td>
</tr>
<tr>
<td>30</td>
<td>1.7628</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.5923</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) F-Distribution Table (cont.)

The table entries (i.e., \(F_{\nu_1,\nu_2,\alpha}\) values) are the right tail \(\alpha\) probability cut-off points for the \(F_{\nu_1,\nu_2}\)-distribution, where \(\nu_1 = \text{numerator } df\) and \(\nu_2 = \text{denominator } df\) (i.e., area under the \(F_{\nu_1,\nu_2}-pdf\) on the right of \(F_{\nu_1,\nu_2,\alpha}\) is \(\alpha\)).

<table>
<thead>
<tr>
<th>Denominator</th>
<th>Numerator (df = \nu_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha = 0.005)</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>16210.72</td>
</tr>
<tr>
<td>2</td>
<td>198.5013</td>
</tr>
<tr>
<td>3</td>
<td>55.5520</td>
</tr>
<tr>
<td>12</td>
<td>11.7542</td>
</tr>
<tr>
<td>13</td>
<td>11.3735</td>
</tr>
<tr>
<td>14</td>
<td>11.0602</td>
</tr>
<tr>
<td>15</td>
<td>10.7980</td>
</tr>
<tr>
<td>16</td>
<td>10.5754</td>
</tr>
<tr>
<td>17</td>
<td>10.3841</td>
</tr>
<tr>
<td>18</td>
<td>10.2181</td>
</tr>
<tr>
<td>19</td>
<td>10.0725</td>
</tr>
<tr>
<td>50</td>
<td>8.6258</td>
</tr>
<tr>
<td>60</td>
<td>8.4946</td>
</tr>
<tr>
<td>80</td>
<td>8.3346</td>
</tr>
<tr>
<td>100</td>
<td>8.2406</td>
</tr>
<tr>
<td>120</td>
<td>8.1788</td>
</tr>
<tr>
<td>(\infty)</td>
<td>7.8829</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 = \text{numerator df}$ and $\nu_2 = \text{denominator df}$ (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>$\alpha = 0.005$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>23925.41</td>
</tr>
<tr>
<td>2</td>
<td>199.3746</td>
</tr>
<tr>
<td>3</td>
<td>44.1256</td>
</tr>
<tr>
<td>7</td>
<td>8.6781</td>
</tr>
<tr>
<td>8</td>
<td>7.4959</td>
</tr>
<tr>
<td>10</td>
<td>6.1159</td>
</tr>
<tr>
<td>11</td>
<td>5.6821</td>
</tr>
<tr>
<td>40</td>
<td>3.3497</td>
</tr>
<tr>
<td>50</td>
<td>3.2188</td>
</tr>
<tr>
<td>60</td>
<td>3.1344</td>
</tr>
<tr>
<td>80</td>
<td>3.0320</td>
</tr>
<tr>
<td>100</td>
<td>2.9721</td>
</tr>
<tr>
<td>120</td>
<td>2.9329</td>
</tr>
<tr>
<td>$\infty$</td>
<td>2.7465</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1,\nu_2,\alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1,\nu_2}$-distribution, where $\nu_1 =$ numerator df and $\nu_2 =$ denominator df (i.e., area under the $F_{\nu_1,\nu_2}$-pdf on the right of $F_{\nu_1,\nu_2,\alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24630.21</td>
<td>24681.46</td>
<td>24726.79</td>
<td>24767.17</td>
<td>24803.35</td>
<td>24835.97</td>
<td>25043.62</td>
</tr>
<tr>
<td>2</td>
<td>199.4329</td>
<td>199.4370</td>
<td>199.4407</td>
<td>199.4440</td>
<td>199.4469</td>
<td>199.4495</td>
<td>199.4662</td>
</tr>
<tr>
<td>3</td>
<td>43.0846</td>
<td>43.0082</td>
<td>42.9406</td>
<td>42.8803</td>
<td>42.8262</td>
<td>42.7775</td>
<td>42.4658</td>
</tr>
<tr>
<td>7</td>
<td>7.9677</td>
<td>7.9148</td>
<td>7.8678</td>
<td>7.8258</td>
<td>7.7881</td>
<td>7.7539</td>
<td>7.5344</td>
</tr>
<tr>
<td>9</td>
<td>6.0324</td>
<td>5.9828</td>
<td>5.9388</td>
<td>5.8994</td>
<td>5.8639</td>
<td>5.8318</td>
<td>5.6247</td>
</tr>
<tr>
<td>10</td>
<td>5.4706</td>
<td>5.4221</td>
<td>5.3789</td>
<td>5.3402</td>
<td>5.3054</td>
<td>5.2740</td>
<td>5.0705</td>
</tr>
<tr>
<td>11</td>
<td>5.0489</td>
<td>5.0011</td>
<td>4.958</td>
<td>4.9205</td>
<td>4.8862</td>
<td>4.8552</td>
<td>4.6543</td>
</tr>
<tr>
<td>30</td>
<td>3.0057</td>
<td>2.9610</td>
<td>2.9211</td>
<td>2.8852</td>
<td>2.8526</td>
<td>2.8230</td>
<td>2.6278</td>
</tr>
<tr>
<td>40</td>
<td>2.7810</td>
<td>2.7365</td>
<td>2.6966</td>
<td>2.6607</td>
<td>2.6281</td>
<td>2.5984</td>
<td>2.4014</td>
</tr>
<tr>
<td>50</td>
<td>2.6531</td>
<td>2.6085</td>
<td>2.5686</td>
<td>2.5326</td>
<td>2.4999</td>
<td>2.4701</td>
<td>2.2716</td>
</tr>
<tr>
<td>60</td>
<td>2.5704</td>
<td>2.5258</td>
<td>2.4859</td>
<td>2.4498</td>
<td>2.4170</td>
<td>2.3872</td>
<td>2.1874</td>
</tr>
<tr>
<td>80</td>
<td>2.4700</td>
<td>2.4254</td>
<td>2.3853</td>
<td>2.3491</td>
<td>2.3162</td>
<td>2.2862</td>
<td>2.0844</td>
</tr>
<tr>
<td>100</td>
<td>2.4112</td>
<td>2.3666</td>
<td>2.3264</td>
<td>2.2901</td>
<td>2.2571</td>
<td>2.2270</td>
<td>2.0239</td>
</tr>
<tr>
<td>120</td>
<td>2.3727</td>
<td>2.3279</td>
<td>2.2877</td>
<td>2.2514</td>
<td>2.2183</td>
<td>2.1881</td>
<td>1.9839</td>
</tr>
<tr>
<td>$\infty$</td>
<td>2.1889</td>
<td>2.1438</td>
<td>2.1032</td>
<td>2.0664</td>
<td>2.0328</td>
<td>2.0020</td>
<td>1.7913</td>
</tr>
</tbody>
</table>
4. (Snedecor’s) $F$-Distribution Table (cont.)

The table entries (i.e., $F_{\nu_1, \nu_2, \alpha}$ values) are the right tail $\alpha$ probability cut-off points for the $F_{\nu_1, \nu_2}$-distribution, where $\nu_1 =$ numerator $df$ and $\nu_2 =$ denominator $df$ (i.e., area under the $F_{\nu_1, \nu_2}$-pdf on the right of $F_{\nu_1, \nu_2, \alpha}$ is $\alpha$).

<table>
<thead>
<tr>
<th>Denominator $df = \nu_2$</th>
<th>Numerator $df = \nu_1$</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>80</th>
<th>100</th>
<th>120</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25148.15</td>
<td>25211.09</td>
<td>25253.13</td>
<td>25305.79</td>
<td>25337.45</td>
<td>25358.57</td>
<td>25463.18</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>199.4745</td>
<td>199.4795</td>
<td>199.4829</td>
<td>199.4870</td>
<td>199.4895</td>
<td>199.4912</td>
<td>199.4994</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>42.3082</td>
<td>42.2130</td>
<td>42.1494</td>
<td>42.0696</td>
<td>42.0215</td>
<td>41.9894</td>
<td>41.8502</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.4224</td>
<td>7.3544</td>
<td>7.3087</td>
<td>7.2512</td>
<td>7.2165</td>
<td>7.1933</td>
<td>7.0774</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5.5185</td>
<td>5.4539</td>
<td>5.4104</td>
<td>5.3555</td>
<td>5.3223</td>
<td>5.3001</td>
<td>5.1889</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3.7560</td>
<td>3.6975</td>
<td>3.6552</td>
<td>3.6017</td>
<td>3.5692</td>
<td>3.5473</td>
<td>3.4372</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>3.3107</td>
<td>3.2482</td>
<td>3.2058</td>
<td>3.1519</td>
<td>3.1191</td>
<td>3.0970</td>
<td>2.9853</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3.1057</td>
<td>3.0430</td>
<td>3.0003</td>
<td>2.9461</td>
<td>2.9130</td>
<td>2.8907</td>
<td>2.7775</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.0215</td>
<td>2.9586</td>
<td>2.9158</td>
<td>2.8614</td>
<td>2.8282</td>
<td>2.8058</td>
<td>2.6918</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>2.5240</td>
<td>2.4594</td>
<td>2.4151</td>
<td>2.3583</td>
<td>2.3234</td>
<td>2.2997</td>
<td>2.1775</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>2.2958</td>
<td>2.2295</td>
<td>2.1838</td>
<td>2.1248</td>
<td>2.0884</td>
<td>2.0635</td>
<td>1.9334</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>2.1644</td>
<td>2.0967</td>
<td>2.0498</td>
<td>1.9890</td>
<td>1.9512</td>
<td>1.9253</td>
<td>1.7880</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>2.0788</td>
<td>2.0100</td>
<td>1.9621</td>
<td>1.8998</td>
<td>1.8608</td>
<td>1.8341</td>
<td>1.6904</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>1.9739</td>
<td>1.9033</td>
<td>1.8539</td>
<td>1.7892</td>
<td>1.7484</td>
<td>1.7292</td>
<td>1.5654</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.9119</td>
<td>1.8400</td>
<td>1.7896</td>
<td>1.7230</td>
<td>1.6808</td>
<td>1.6516</td>
<td>1.4875</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>1.8709</td>
<td>1.7981</td>
<td>1.7468</td>
<td>1.6789</td>
<td>1.6356</td>
<td>1.6055</td>
<td>1.4335</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.6715</td>
<td>1.5923</td>
<td>1.5351</td>
<td>1.4567</td>
<td>1.4046</td>
<td>1.3668</td>
<td>1.0528</td>
<td></td>
</tr>
</tbody>
</table>
Bibliography


© 2006 by Taylor & Francis Group, LLC


[55] Pearson, K. (1933). On a Method of Determining Whether a Sample of Size $n$ Supposed to have been Drawn from a Parent Population Having a Known Probability Integral has been Drawn at Random. *Biometrika*, 25, 379–410.


========================================

**Special Acknowledgement**

The materials (data sets, partial tables, equations or other mathematical expressions) taken from:

(i) Dahiya and Gurland (1972); Engelhardt and Bain (1977); Harter and Moore (1968); Keating, Glaser and Ketchum (1990); Proschan (1963); Schüpbach and Hüsler (1983) - have been reprinted with permission from the relevant journals. Copyright by the American Statistical Association. All rights reserved.

(ii) Lockhart and Stephens (1994) - have been reprinted with permission from the Royal Statistical Society, London, UK.

(iii) Coles (1989) - have been reprinted with permission from the Oxford University Press, Oxford, UK.

========================================