Mathematical Statistics: Exercises and Solutions

Jun Shao

Springer

Mathematical Statistics: Exercises and Solutions

Jun Shao

Mathematical Statistics: Exercises and Solutions

Springer

Jun Shao
Department of Statistics
University of Wisconsin
Madison, WI 52706
USA
shao@stat.wisc.edu

Library of Congress Control Number: 2005923578

ISBN-10: 0-387-24970-2 Printed on acid-free paper.
ISBN-13: 978-0387-24970-4

Printed in the United States of America. (EB)

9 8 7 6 5 4 3 2 1

springeronline.com

To My Parents

Preface

Since the publication of my book *Mathematical Statistics* (Shao, 2003), I have been asked many times for a solution manual to the exercises in my book. Without doubt, exercises form an important part of a textbook on mathematical statistics, not only in training students for their research ability in mathematical statistics but also in presenting many additional results as complementary material to the main text. Written solutions to these exercises are important for students who initially do not have the skills in solving these exercises completely and are very helpful for instructors of a mathematical statistics course (whether or not my book *Mathematical Statistics* is used as the textbook) in providing answers to students as well as finding additional examples to the main text. Motivated by this and encouraged by some of my colleagues and Springer-Verlag editor John Kimmel, I have completed this book, *Mathematical Statistics: Exercises and Solutions*.

This book consists of solutions to 400 exercises, over 95% of which are in my book *Mathematical Statistics*. Many of them are standard exercises that also appear in other textbooks listed in the references. It is only a partial solution manual to *Mathematical Statistics* (which contains over 900 exercises). However, the types of exercise in *Mathematical Statistics* not selected in the current book are (1) exercises that are routine (each exercise selected in this book has a certain degree of difficulty), (2) exercises similar to one or several exercises selected in the current book, and (3) exercises for advanced materials that are often not included in a mathematical statistics course for first-year Ph.D. students in statistics (e.g., Edgeworth expansions and second-order accuracy of confidence sets, empirical likelihoods, statistical functionals, generalized linear models, nonparametric tests, and theory for the bootstrap and jackknife, etc.). On the other hand, this is a stand-alone book, since exercises and solutions are comprehensible independently of their source for likely readers. To help readers not using this book together with *Mathematical Statistics*, lists of notation, terminology, and some probability distributions are given in the front of the book.

All notational conventions are the same as or very similar to those in *Mathematical Statistics* and so is the mathematical level of this book. Readers are assumed to have a good knowledge in advanced calculus. A course in real analysis or measure theory is highly recommended. If this book is used with a statistics textbook that does not include probability theory, then knowledge in measure-theoretic probability theory is required.

The exercises are grouped into seven chapters with titles matching those in *Mathematical Statistics*. A few errors in the exercises from *Mathematical Statistics* were detected during the preparation of their solutions and the corrected versions are given in this book. Although exercises are numbered independently of their source, the corresponding number in *Mathematical Statistics* is accompanied with each exercise number for convenience of instructors and readers who also use *Mathematical Statistics* as the main text. For example, Exercise 8 (#2.19) means that Exercise 8 in the current book is also Exercise 19 in Chapter 2 of *Mathematical Statistics*.

A note to students/readers who have a need for exercises accompanied by solutions is that they should not be completely driven by the solutions. Students/readers are encouraged to try each exercise first without reading its solution. If an exercise is solved with the help of a solution, they are encouraged to provide solutions to similar exercises as well as to think about whether there is an alternative solution to the one given in this book. A few exercises in this book are accompanied by two solutions and/or notes of brief discussions.

I would like to thank my teaching assistants, Dr. Hansheng Wang, Dr. Bin Cheng, and Mr. Fang Fang, who provided valuable help in preparing some solutions. Any errors are my own responsibility, and a correction of them can be found on my web page http://www.stat.wisc.edu/~shao.

Madison, Wisconsin — Jun Shao
April 2005

Contents

Notation

$\mathcal{R}$: The real line.

$\mathcal{R}^k$: The k-dimensional Euclidean space.

$c = (c_1, ..., c_k)$: A vector (element) in $\mathcal{R}^k$ with jth component $c_j \in \mathcal{R}$; c is considered as a $k \times 1$ matrix (column vector) when matrix algebra is involved.

c^τ: The transpose of a vector $c \in \mathcal{R}^k$ considered as a $1 \times k$ matrix (row vector) when matrix algebra is involved.

$\|c\|$: The Euclidean norm of a vector $c \in \mathcal{R}^k$, $\|c\|^2 = c^\tau c$.

$|c|$: The absolute value of $c \in \mathcal{R}$.

A^τ: The transpose of a matrix A.

$\mathrm{Det}(A)$ or $|A|$: The determinant of a matrix A.

$\mathrm{tr}(A)$: The trace of a matrix A.

$\|A\|$: The norm of a matrix A defined as $\|A\|^2 = \mathrm{tr}(A^\tau A)$.

A^{-1}: The inverse of a matrix A.

A^-: The generalized inverse of a matrix A.

$A^{1/2}$: The square root of a nonnegative definite matrix A defined by $A^{1/2}A^{1/2} = A$.

$A^{-1/2}$: The inverse of $A^{1/2}$.

$\mathcal{R}(A)$: The linear space generated by rows of a matrix A.

I_k: The $k \times k$ identity matrix.

J_k: The k-dimensional vector of 1's.

$\emptyset$: The empty set.

(a, b): The open interval from a to b.

$[a, b]$: The closed interval from a to b.

$(a, b]$: The interval from a to b including b but not a.

$[a, b)$: The interval from a to b including a but not b.

$\{a, b, c\}$: The set consisting of the elements a, b, and c.

$A_1 \times \cdots \times A_k$: The Cartesian product of sets $A_1, ..., A_k$, $A_1 \times \cdots \times A_k = \{(a_1, ..., a_k) : a_1 \in A_1, ..., a_k \in A_k\}$.

$\sigma(\mathcal{C})$: The smallest σ-field that contains $\mathcal{C}$.

$\sigma(X)$: The smallest σ-field with respect to which X is measurable.

$\nu_1 \times \cdots \times \nu_k$: The product measure of ν_1,...,ν_k on $\sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_k)$, where ν_i is a measure on $\mathcal{F}_i$, $i = 1, ..., k$.

$\mathcal{B}$: The Borel σ-field on $\mathcal{R}$.

$\mathcal{B}^k$: The Borel σ-field on $\mathcal{R}^k$.

A^c: The complement of a set A.

$A \cup B$: The union of sets A and B.

$\cup A_i$: The union of sets $A_1, A_2, ...$.

$A \cap B$: The intersection of sets A and B.

$\cap A_i$: The intersection of sets $A_1, A_2, ...$.

I_A: The indicator function of a set A.

$P(A)$: The probability of a set A.

$\int f d\nu$: The integral of a Borel function f with respect to a measure ν.

$\int_A f d\nu$: The integral of f on the set A.

$\int f(x)dF(x)$: The integral of f with respect to the probability measure corresponding to the cumulative distribution function F.

$\lambda \ll \nu$: The measure λ is dominated by the measure ν, i.e., $\nu(A) = 0$ always implies $\lambda(A) = 0$.

$\frac{d\lambda}{d\nu}$: The Radon-Nikodym derivative of λ with respect to ν.

$\mathcal{P}$: A collection of populations (distributions).

a.e.: Almost everywhere.

a.s.: Almost surely.

a.s. $\mathcal{P}$: A statement holds except on the event A with $P(A) = 0$ for all $P \in \mathcal{P}$.

δ_x: The point mass at $x \in \mathcal{R}^k$ or the distribution degenerated at $x \in \mathcal{R}^k$.

$\{a_n\}$: A sequence of elements $a_1, a_2, ...$.

$a_n \to a$ or $\lim_n a_n = a$: $\{a_n\}$ converges to a as n increases to ∞.

$\limsup_n a_n$: The largest limit point of $\{a_n\}$, $\limsup_n a_n = \inf_n \sup_{k \geq n} a_k$.

$\liminf_n a_n$: The smallest limit point of $\{a_n\}$, $\liminf_n a_n = \sup_n \inf_{k \geq n} a_k$.

$\to_p$: Convergence in probability.

$\to_d$: Convergence in distribution.

g': The derivative of a function g on $\mathcal{R}$.

g'': The second-order derivative of a function g on $\mathcal{R}$.

$g^{(k)}$: The kth-order derivative of a function g on $\mathcal{R}$.

$g(x+)$: The right limit of a function g at $x \in \mathcal{R}$.

$g(x-)$: The left limit of a function g at $x \in \mathcal{R}$.

$g_+(x)$: The positive part of a function g, $g_+(x) = \max\{g(x), 0\}$.

$g_-(x)$: The negative part of a function g, $g_-(x) = \max\{-g(x), 0\}$.

$\partial g/\partial x$: The partial derivative of a function g on $\mathcal{R}^k$.

$\partial^2 g/\partial x \partial x^\tau$: The second-order partial derivative of a function g on $\mathcal{R}^k$.

$\exp\{x\}$: The exponential function e^x.

$\log x$ or $\log(x)$: The inverse of e^x, $\log(e^x) = x$.

$\Gamma(t)$: The gamma function defined as $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$, $t > 0$.

$F^{-1}(p)$: The pth quantile of a cumulative distribution function F on $\mathcal{R}$, $F^{-1}(t) = \inf\{x : F(x) \geq t\}$.

$E(X)$ or EX: The expectation of a random variable (vector or matrix) X.

$\mathrm{Var}(X)$: The variance of a random variable X or the covariance matrix of a random vector X.

$\mathrm{Cov}(X, Y)$: The covariance between random variables X and Y.

$E(X|\mathcal{A})$: The conditional expectation of X given a σ-field $\mathcal{A}$.

$E(X|Y)$: The conditional expectation of X given Y.

$P(A|\mathcal{A})$: The conditional probability of A given a σ-field $\mathcal{A}$.

$P(A|Y)$: The conditional probability of A given Y.

$X_{(i)}$: The ith order statistic of $X_1, ..., X_n$.

$\bar{X}$ or $\bar{X}_\cdot$: The sample mean of $X_1, ..., X_n$, $\bar{X} = n^{-1}\sum_{i=1}^n X_i$.

$\bar{X}_{\cdot j}$: The average of X_{ij}'s over the index i, $\bar{X}_{\cdot j} = n^{-1}\sum_{i=1}^n X_{ij}$.

S^2: The sample variance of $X_1, ..., X_n$, $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

F_n: The empirical distribution of $X_1, ..., X_n$, $F_n(t) = n^{-1}\sum_{i=1}^n \delta_{X_i}(t)$.

$\ell(\theta)$: The likelihood function.

H_0: The null hypothesis in a testing problem.

H_1: The alternative hypothesis in a testing problem.

$L(P, a)$ or $L(\theta, a)$: The loss function in a decision problem.

$R_T(P)$ or $R_T(\theta)$: The risk function of a decision rule T.

r_T: The Bayes risk of a decision rule T.

$N(\mu, \sigma^2)$: The one-dimensional normal distribution with mean μ and variance σ^2.

$N_k(\mu, \Sigma)$: The k-dimensional normal distribution with mean vector μ and covariance matrix Σ.

$\Phi(x)$: The cumulative distribution function of $N(0, 1)$.

z_α: The $(1-\alpha)$th quantile of $N(0, 1)$.

χ_r^2: The chi-square distribution with degrees of freedom r.

$\chi_{r,\alpha}^2$: The $(1-\alpha)$th quantile of the chi-square distribution χ_r^2.

$\chi_r^2(\delta)$: The noncentral chi-square distribution with degrees of freedom r and noncentrality parameter δ.

t_r: The t-distribution with degrees of freedom r.

$t_{r,\alpha}$: The $(1-\alpha)$th quantile of the t-distribution t_r.

$t_r(\delta)$: The noncentral t-distribution with degrees of freedom r and noncentrality parameter δ.

$F_{a,b}$: The F-distribution with degrees of freedom a and b.

$F_{a,b,\alpha}$: The $(1-\alpha)$th quantile of the F-distribution $F_{a,b}$.

$F_{a,b}(\delta)$: The noncentral F-distribution with degrees of freedom a and b and noncentrality parameter δ.

▮: The end of a solution.

Terminology

σ-field: A collection $\mathcal{F}$ of subsets of a set Ω is a σ-field on Ω if (i) the empty set $\emptyset \in \mathcal{F}$; (ii) if $A \in \mathcal{F}$, then the complement $A^c \in \mathcal{F}$; and (iii) if $A_i \in \mathcal{F}$, $i = 1, 2, ...$, then their union $\cup A_i \in \mathcal{F}$.

σ-finite measure: A measure ν on a σ-field $\mathcal{F}$ on Ω is σ-finite if there are $A_1, A_2, ...$ in $\mathcal{F}$ such that $\cup A_i = \Omega$ and $\nu(A_i) < \infty$ for all i.

Action or decision: Let X be a sample from a population P. An action or decision is a conclusion we make about P based on the observed X.

Action space: The set of all possible actions.

Admissibility: A decision rule T is admissible under the loss function $L(P, \cdot)$, where P is the unknown population, if there is no other decision rule T_1 that is better than T in the sense that $E[L(P, T_1)] \leq E[L(P, T)]$ for all P and $E[L(P, T_1)] < E[L(P, T)]$ for some P.

Ancillary statistic: A statistic is ancillary if and only if its distribution does not depend on any unknown quantity.

Asymptotic bias: Let T_n be an estimator of θ for every n satisfying $a_n(T_n - \theta) \to_d Y$ with $E|Y| < \infty$, where $\{a_n\}$ is a sequence of positive numbers satisfying $\lim_n a_n = \infty$ or $\lim_n a_n = a > 0$. An asymptotic bias of T_n is defined to be EY/a_n.

Asymptotic level α test: Let X be a sample of size n from P and $T(X)$ be a test for $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$. If $\lim_n E[T(X)] \leq \alpha$ for any $P \in \mathcal{P}_0$, then $T(X)$ has asymptotic level α.

Asymptotic mean squared error and variance: Let T_n be an estimator of θ for every n satisfying $a_n(T_n - \theta) \to_d Y$ with $0 < EY^2 < \infty$, where $\{a_n\}$ is a sequence of positive numbers satisfying $\lim_n a_n = \infty$. The asymptotic mean squared error of T_n is defined to be EY^2/a_n^2 and the asymptotic variance of T_n is defined to be $\mathrm{Var}(Y)/a_n^2$.

Asymptotic relative efficiency: Let T_n and T_n' be estimators of θ. The asymptotic relative efficiency of T_n' with respect to T_n is defined to be the asymptotic mean squared error of T_n divided by the asymptotic mean squared error of T_n.

Asymptotically correct confidence set: Let X be a sample of size n from P and $C(X)$ be a confidence set for θ. If $\lim_n P(\theta \in C(X)) = 1 - \alpha$, then $C(X)$ is $1 - \alpha$ asymptotically correct.

Bayes action: Let X be a sample from a population indexed by $\theta \in \Theta \subset \mathcal{R}^k$. A Bayes action in a decision problem with action space A and loss function $L(\theta, a)$ is the action that minimizes the posterior expected loss $E[L(\theta, a)]$ over $a \in A$, where E is the expectation with respect to the posterior distribution of θ given X.

Bayes risk: Let X be a sample from a population indexed by $\theta \in \Theta \subset \mathcal{R}^k$. The Bayes risk of a decision rule T is the expected risk of T with respect to a prior distribution on Θ.

Bayes rule or Bayes estimator: A Bayes rule has the smallest Bayes risk over all decision rules. A Bayes estimator is a Bayes rule in an estimation problem.

Borel σ-field $\mathcal{B}^k$: The smallest σ-field containing all open subsets of $\mathcal{R}^k$.

Borel function: A function f from Ω to $\mathcal{R}^k$ is Borel with respect to a σ-field $\mathcal{F}$ on Ω if and only if $f^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{B}^k$.

Characteristic function: The characteristic function of a distribution F on $\mathcal{R}^k$ is $\int e^{\sqrt{-1}t^\tau x} dF(x)$, $t \in \mathcal{R}^k$.

Complete (or bounded complete) statistic: Let X be a sample from a population P. A statistic $T(X)$ is complete (or bounded complete) for P if and only if, for any Borel (or bounded Borel) f, $E[f(T)] = 0$ for all P implies $f = 0$ except for a set A with $P(X \in A) = 0$ for all P.

Conditional expectation $E(X|\mathcal{A})$: Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A}$ be a σ-field contained in $\mathcal{F}$. The conditional expectation of X given $\mathcal{A}$, denoted by $E(X|\mathcal{A})$, is defined to be the a.s.-unique random variable satisfying (a) $E(X|\mathcal{A})$ is Borel with respect to $\mathcal{A}$ and (b) $\int_A E(X|\mathcal{A}) dP = \int_A X dP$ for any $A \in \mathcal{A}$.

Conditional expectation $E(X|Y)$: The conditional expectation of X given Y, denoted by $E(X|Y)$, is defined as $E(X|Y) = E(X|\sigma(Y))$.

Confidence coefficient and confidence set: Let X be a sample from a population P and $\theta \in \mathcal{R}^k$ be an unknown parameter that is a function of P. A confidence set $C(X)$ for θ is a Borel set on $\mathcal{R}^k$ depending on X. The confidence coefficient of a confidence set $C(X)$ is $\inf_P P(\theta \in C(X))$. A confidence set is said to be a $1 - \alpha$ confidence set for θ if its confidence coefficient is $1 - \alpha$.

Confidence interval: A confidence interval is a confidence set that is an interval.

Consistent estimator: Let X be a sample of size n from P. An estimator $T(X)$ of θ is consistent if and only if $T(X) \to_p \theta$ for any P as $n \to \infty$. $T(X)$ is strongly consistent if and only if $\lim_n T(X) = \theta$ a.s. for any P. $T(X)$ is consistent in mean squared error if and only if $\lim_n E[T(X) - \theta]^2 = 0$ for any P.

Consistent test: Let X be a sample of size n from P. A test $T(X)$ for testing $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ is consistent if and only if $\lim_n E[T(X)] = 1$ for any $P \in \mathcal{P}_1$.

Decision rule (nonrandomized): Let X be a sample from a population P. A (nonrandomized) decision rule is a measurable function from the range of X to the action space.

Discrete probability density: A probability density with respect to the counting measure on the set of nonnegative integers.

Distribution and cumulative distribution function: The probability measure corresponding to a random vector is called its distribution (or law). The cumulative distribution function of a distribution or probability measure P on $\mathcal{B}^k$ is $F(x_1, ..., x_k) = P((-\infty, x_1] \times \cdots \times (-\infty, x_k])$, $x_i \in \mathcal{R}$.

Empirical Bayes rule: An empirical Bayes rule is a Bayes rule with parameters in the prior estimated using data.

Empirical distribution: The empirical distribution based on a random sample $(X_1, ..., X_n)$ is the distribution putting mass n^{-1} at each X_i, $i = 1, ..., n$.

Estimability: A parameter θ is estimable if and only if there exists an unbiased estimator of θ.

Estimator: Let X be a sample from a population P and $\theta \in \mathcal{R}^k$ be a function of P. An estimator of θ is a measurable function of X.

Exponential family: A family of probability densities $\{f_\theta : \theta \in \Theta\}$ (with respect to a common σ-finite measure ν), $\Theta \subset \mathcal{R}^k$, is an exponential family if and only if $f_\theta(x) = \exp\big\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\big\}h(x)$, where T is a random p-vector with a fixed positive integer p, η is a function from Θ to $\mathcal{R}^p$, h is a nonnegative Borel function, and $\xi(\theta) = \log\big\{\int \exp\{[\eta(\theta)]^\tau T(x)\}h(x)d\nu\big\}$.

Generalized Bayes rule: A generalized Bayes rule is a Bayes rule when the prior distribution is improper.

Improper or proper prior: A prior is improper if it is a measure but not a probability measure. A prior is proper if it is a probability measure.

Independence: Let $(\Omega, \mathcal{F}, P)$ be a probability space. Events in $\mathcal{C} \subset \mathcal{F}$ are independent if and only if for any positive integer n and distinct events A_1,...,A_n in $\mathcal{C}$, $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$. Collections $\mathcal{C}_i \subset \mathcal{F}$, $i \in \mathcal{I}$ (an index set that can be uncountable), are independent if and only if events in any collection of the form $\{A_i \in \mathcal{C}_i : i \in \mathcal{I}\}$ are independent. Random elements X_i, $i \in \mathcal{I}$, are independent if and only if $\sigma(X_i)$, $i \in \mathcal{I}$, are independent.

Integration or integral: Let ν be a measure on a σ-field $\mathcal{F}$ on a set Ω. The integral of a nonnegative simple function (i.e., a function of the form $\varphi(\omega) = \sum_{i=1}^k a_i I_{A_i}(\omega)$, where $\omega \in \Omega$, k is a positive integer, $A_1, ..., A_k$ are in $\mathcal{F}$, and $a_1, ..., a_k$ are nonnegative numbers) is defined as $\int \varphi d\nu = \sum_{i=1}^k a_i \nu(A_i)$. The integral of a nonnegative Borel function is defined as $\int f d\nu = \sup_{\varphi \in S_f} \int \varphi d\nu$, where S_f is the collection of all nonnegative simple functions that are bounded by f. For a Borel function f, its integral exists if and only if at least one of $\int \max\{f, 0\} d\nu$ and $\int \max\{-f, 0\} d\nu$ is finite, in which case $\int f d\nu = \int \max\{f, 0\} d\nu - \int \max\{-f, 0\} d\nu$. f is integrable if and only if both $\int \max\{f, 0\} d\nu$ and $\int \max\{-f, 0\} d\nu$ are finite. When ν is a probability measure corresponding to the cumulative distribution function F on $\mathcal{R}^k$, we write $\int f d\nu = \int f(x) dF(x)$. For any event A, $\int_A f d\nu$ is defined as $\int I_A f d\nu$.

Invariant decision rule: Let X be a sample from $P \in \mathcal{P}$ and $\mathcal{G}$ be a group of one-to-one transformations of X ($g_i \in \mathcal{G}$ implies $g_1 \circ g_2 \in \mathcal{G}$ and $g_i^{-1} \in \mathcal{G}$). $\mathcal{P}$ is invariant under $\mathcal{G}$ if and only if $\bar{g}(P_X) = P_{g(X)}$ is a one-to-one transformation from $\mathcal{P}$ onto $\mathcal{P}$ for each $g \in \mathcal{G}$. A decision problem is invariant if and only if $\mathcal{P}$ is invariant under $\mathcal{G}$ and the loss $L(P, a)$ is invariant in the sense that, for every $g \in \mathcal{G}$ and every $a \in A$ (the collection of all possible actions), there exists a unique $\bar{g}(a) \in A$ such that $L(P_X, a) = L\left(P_{g(X)}, \bar{g}(a)\right)$. A decision rule $T(x)$ in an invariant decision problem is invariant if and only if, for every $g \in \mathcal{G}$ and every x in the range of X, $T(g(x)) = \bar{g}(T(x))$.

Invariant estimator: An invariant estimator is an invariant decision rule in an estimation problem.

LR (Likelihood ratio) test: Let $\ell(\theta)$ be the likelihood function based on a sample X whose distribution is P_θ, $\theta \in \Theta \subset \mathcal{R}^p$ for some positive integer p. For testing $H_0 : \theta \in \Theta_0 \subset \Theta$ versus $H_1 : \theta \notin \Theta_0$, an LR test is any test that rejects H_0 if and only if $\lambda(X) < c$, where $c \in [0, 1]$ and $\lambda(X) = \sup_{\theta \in \Theta_0} \ell(\theta) / \sup_{\theta \in \Theta} \ell(\theta)$ is the likelihood ratio.

LSE: The least squares estimator.

Level α test: A test is of level α if its size is at most α.

Level $1 - \alpha$ confidence set or interval: A confidence set or interval is said to be of level $1 - \alpha$ if its confidence coefficient is at least $1 - \alpha$.

Likelihood function and likelihood equation: Let X be a sample from a population P indexed by an unknown parameter vector $\theta \in \mathcal{R}^k$. The joint probability density of X treated as a function of θ is called the likelihood function and denoted by $\ell(\theta)$. The likelihood equation is $\partial \log \ell(\theta) / \partial \theta = 0$.

Location family: A family of Lebesgue densities on $\mathcal{R}$, $\{f_\mu : \mu \in \mathcal{R}\}$, is a location family with location parameter μ if and only if $f_\mu(x) = f(x-\mu)$, where f is a known Lebesgue density.

Location invariant estimator. Let $(X_1, ..., X_n)$ be a random sample from a population in a location family. An estimator $T(X_1, ..., X_n)$ of the location parameter is location invariant if and only if $T(X_1+c, ..., X_n+c) = T(X_1, ..., X_n) + c$ for any X_i's and $c \in \mathcal{R}$.

Location-scale family: A family of Lebesgue densities on $\mathcal{R}$, $\{f_{\mu,\sigma} : \mu \in \mathcal{R}, \sigma > 0\}$, is a location-scale family with location parameter μ and scale parameter σ if and only if $f_{\mu,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, where f is a known Lebesgue density.

Location-scale invariant estimator. Let $(X_1, ..., X_n)$ be a random sample from a population in a location-scale family with location parameter μ and scale parameter σ. An estimator $T(X_1, ..., X_n)$ of the location parameter μ is location-scale invariant if and only if $T(rX_1 + c, ..., rX_n + c) = rT(X_1, ..., X_n) + c$ for any X_i's, $c \in \mathcal{R}$, and $r > 0$. An estimator $S(X_1, ..., X_n)$ of σ^h with a fixed $h \neq 0$ is location-scale invariant if and only if $S(rX_1 + c, ..., rX_n + c) = r^h T(X_1, ..., X_n)$ for any X_i's and $r > 0$.

Loss function: Let X be a sample from a population $P \in \mathcal{P}$ and A be the set of all possible actions we may take after we observe X. A loss function $L(P, a)$ is a nonnegative Borel function on $\mathcal{P} \times A$ such that if a is our action and P is the true population, our loss is $L(P, a)$.

MRIE (minimum risk invariant estimator): The MRIE of an unknown parameter θ is the estimator has the minimum risk within the class of invariant estimators.

MLE (maximum likelihood estimator): Let X be a sample from a population P indexed by an unknown parameter vector $\theta \in \Theta \subset \mathcal{R}^k$ and $\ell(\theta)$ be the likelihood function. A $\hat{\theta} \in \Theta$ satisfying $\ell(\hat{\theta}) = \max_{\theta \in \Theta} \ell(\theta)$ is called an MLE of θ (Θ may be replaced by its closure in the above definition).

Measure: A set function ν defined on a σ-field $\mathcal{F}$ on Ω is a measure if (i) $0 \leq \nu(A) \leq \infty$ for any $A \in \mathcal{F}$; (ii) $\nu(\emptyset) = 0$; and (iii) $\nu\left(\cup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \nu(A_i)$ for disjoint $A_i \in \mathcal{F}$, $i = 1, 2, ...$.

Measurable function: a function from a set Ω to a set Λ (with a given σ-field $\mathcal{G}$) is measurable with respect to a σ-field $\mathcal{F}$ on Ω if $f^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{G}$.

Minimax rule: Let X be a sample from a population P and $R_T(P)$ be the risk of a decision rule T. A minimax rule is the rule minimizes $\sup_P R_T(P)$ over all possible T.

Moment generating function: The moment generating function of a distribution F on $\mathcal{R}^k$ is $\int e^{t^\tau x} dF(x)$, $t \in \mathcal{R}^k$, if it is finite.

Monotone likelihood ratio: The family of densities $\{f_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathcal{R}$ is said to have monotone likelihood ratio in $Y(x)$ if, for any $\theta_1 < \theta_2$, $\theta_i \in \Theta$, $f_{\theta_2}(x)/f_{\theta_1}(x)$ is a nondecreasing function of $Y(x)$ for values x at which at least one of $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ is positive.

Optimal rule: An optimal rule (within a class of rules) is the rule has the smallest risk over all possible populations.

Pivotal quantity: A known Borel function R of (X, θ) is called a pivotal quantity if and only if the distribution of $R(X, \theta)$ does not depend on any unknown quantity.

Population: The distribution (or probability measure) of an observation from a random experiment is called the population.

Power of a test: The power of a test T is the expected value of T with respect to the true population.

Prior and posterior distribution: Let X be a sample from a population indexed by $\theta \in \Theta \subset \mathcal{R}^k$. A distribution defined on Θ that does not depend on X is called a prior. When the population of X is considered as the conditional distribution of X given θ and the prior is considered as the distribution of θ, the conditional distribution of θ given X is called the posterior distribution of θ.

Probability and probability space: A measure P defined on a σ-field $\mathcal{F}$ on a set Ω is called a probability if and only if $P(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Probability density: Let $(\Omega, \mathcal{F}, P)$ be a probability space and ν be a σ-finite measure on $\mathcal{F}$. If $P \ll \nu$, then the Radon-Nikodym derivative of P with respect to ν is the probability density with respect to ν (and is called Lebesgue density if ν is the Lebesgue measure on $\mathcal{R}^k$).

Random sample: A sample $X = (X_1, ..., X_n)$, where each X_j is a random d-vector with a fixed positive integer d, is called a random sample of size n from a population or distribution P if $X_1, ..., X_n$ are independent and identically distributed as P.

Randomized decision rule: Let X be a sample with range $\mathcal{X}$, A be the action space, and $\mathcal{F}_A$ be a σ-field on A. A randomized decision rule is a function $\delta(x, C)$ on $\mathcal{X} \times \mathcal{F}_A$ such that, for every $C \in \mathcal{F}_A$, $\delta(X, C)$ is a Borel function and, for every $X \in \mathcal{X}$, $\delta(X, C)$ is a probability measure on $\mathcal{F}_A$. A nonrandomized decision rule T can be viewed as a degenerate randomized decision rule δ, i.e., $\delta(X, \{a\}) = I_{\{a\}}(T(X))$ for any $a \in A$ and $X \in \mathcal{X}$.

Risk: The risk of a decision rule is the expectation (with respect to the true population) of the loss of the decision rule.

Sample: The observation from a population treated as a random element is called a sample.

Scale family: A family of Lebesgue densities on $\mathcal{R}$, $\{f_\sigma : \sigma > 0\}$, is a scale family with scale parameter σ if and only if $f_\sigma(x) = \frac{1}{\sigma} f(x/\sigma)$, where f is a known Lebesgue density.

Scale invariant estimator. Let $(X_1, ..., X_n)$ be a random sample from a population in a scale family with scale parameter σ. An estimator $S(X_1, ..., X_n)$ of σ^h with a fixed $h \neq 0$ is scale invariant if and only if $S(rX_1, ..., rX_n) = r^h T(X_1, ..., X_n)$ for any X_i's and $r > 0$.

Simultaneous confidence intervals: Let $\theta_t \in \mathcal{R}$, $t \in \mathcal{T}$. Confidence intervals $C_t(X)$, $t \in \mathcal{T}$, are $1-\alpha$ simultaneous confidence intervals for θ_t, $t \in \mathcal{T}$, if $P(\theta_t \in C_t(X), t \in \mathcal{T}) = 1 - \alpha$.

Statistic: Let X be a sample from a population P. A known Borel function of X is called a statistic.

Sufficiency and minimal sufficiency: Let X be a sample from a population P. A statistic $T(X)$ is sufficient for P if and only if the conditional distribution of X given T does not depend on P. A sufficient statistic T is minimal sufficient if and only if, for any other statistic S sufficient for P, there is a measurable function ψ such that $T = \psi(S)$ except for a set A with $P(X \in A) = 0$ for all P.

Test and its size: Let X be a sample from a population $P \in \mathcal{P}$ and $\mathcal{P}_i$ $i = 0, 1$, be subsets of $\mathcal{P}$ satisfying $\mathcal{P}_0 \cup \mathcal{P}_1 = \mathcal{P}$ and $\mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$. A randomized test for hypotheses $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ is a Borel function $T(X) \in [0, 1]$ such that after X is observed, we reject H_0 (conclude $P \in \mathcal{P}_1$) with probability $T(X)$. If $T(X) \in \{0, 1\}$, then T is nonrandomized. The size of a test T is $\sup_{P \in \mathcal{P}_0} E[T(X)]$, where E is the expectation with respect to P.

UMA (uniformly most accurate) confidence set: Let $\theta \in \Theta$ be an unknown parameter and Θ' be a subset of Θ that does not contain the true value of θ. A confidence set $C(X)$ for θ with confidence coefficient $1 - \alpha$ is Θ'-UMA if and only if for any other confidence set $C_1(X)$ with significance level $1 - \alpha$, $P\big(\theta' \in C(X)\big) \leq P\big(\theta' \in C_1(X)\big)$ for all $\theta' \in \Theta'$.

UMAU (uniformly most accurate unbiased) confidence set: Let $\theta \in \Theta$ be an unknown parameter and Θ' be a subset of Θ that does not contain the true value of θ. A confidence set $C(X)$ for θ with confidence coefficient $1 - \alpha$ is Θ'-UMAU if and only if $C(X)$ is unbiased and for any other unbiased confidence set $C_1(X)$ with significance level $1-\alpha$, $P\big(\theta' \in C(X)\big) \leq P\big(\theta' \in C_1(X)\big)$ for all $\theta' \in \Theta'$.

UMP (uniformly most powerful) test: A test of size α is UMP for testing $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ if and only if, at each $P \in \mathcal{P}_1$, the power of T is no smaller than the power of any other level α test.

UMPU (uniformly most powerful unbiased) test: An unbiased test of size α is UMPU for testing $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ if and only if, at each $P \in \mathcal{P}_1$, the power of T is no larger than the power of any other level α unbiased test.

UMVUE (uniformly minimum variance estimator): An estimator is a UMVUE if it has the minimum variance within the class of unbiased estimators.

Unbiased confidence set: A level $1 - \alpha$ confidence set $C(X)$ is said to be unbiased if and only if $P(\theta' \in C(X)) \leq 1-\alpha$ for any P and all $\theta' \neq \theta$.

Unbiased estimator: Let X be a sample from a population P and $\theta \in \mathcal{R}^k$ be a function of P. If an estimator $T(X)$ of θ satisfies $E[T(X)] = \theta$ for any P, where E is the expectation with respect to P, then $T(X)$ is an unbiased estimator of θ.

Unbiased test: A test for hypotheses $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ is unbiased if its size is no larger than its power at any $P \in \mathcal{P}_1$.

Some Distributions

1. Discrete uniform distribution on the set $\{a_1, ..., a_m\}$: The probability density (with respect to the counting measure) of this distribution is
$$f(x) = \begin{cases} m^{-1} & x = a_i, \ i = 1, ..., m \\ 0 & \text{otherwise,} \end{cases}$$
where $a_i \in \mathcal{R}$, $i = 1, ..., m$, and m is a positive integer. The expectation of this distribution is $\bar{a} = \sum_{j=1}^{m} a_j/m$ and the variance of this distribution is $\sum_{j=1}^{m}(a_j - \bar{a})^2/m$. The moment generating function of this distribution is $\sum_{j=1}^{m} e^{a_j t}/m, \ t \in \mathcal{R}$.

2. The binomial distribution with size n and probability p: The probability density (with respect to the counting measure) of this distribution is
$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, ..., n \\ 0 & \text{otherwise,} \end{cases}$$
where n is a positive integer and $p \in [0, 1]$. The expectation and variance of this distributions are np and $np(1-p)$, respectively. The moment generating function of this distribution is $(pe^t + 1 - p)^n$, $t \in \mathcal{R}$.

3. The Poisson distribution with mean θ: The probability density (with respect to the counting measure) of this distribution is
$$f(x) \begin{cases} \frac{\theta^x e^{-\theta}}{x!} & x = 0, 1, 2, ... \\ 0 & \text{otherwise,} \end{cases}$$
where $\theta > 0$ is the expectation of this distribution. The variance of this distribution is θ. The moment generating function of this distribution is $e^{\theta(e^t - 1)}, \ t \in \mathcal{R}$.

4. The geometric with mean p^{-1}: The probability density (with respect to the counting measure) of this distribution is
$$f(x) = \begin{cases} (1-p)^{x-1} p & x = 1, 2, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $p \in [0,1]$. The expectation and variance of this distribution are p^{-1} and $(1-p)/p^2$, respectively. The moment generating function of this distribution is $pe^t/[1-(1-p)e^t], \; t < -\log(1-p)$.

5. Hypergeometric distribution: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} \frac{\binom{n}{x}\binom{m}{r-x}}{\binom{N}{r}} & x = 0,1,...,\min\{r,n\}, \; r-x \le m \\ 0 & \text{otherwise,} \end{cases}$$

where r, n, and m are positive integers, and $N = n+m$. The expectation and variance of this distribution are equal to rn/N and $rnm(N-r)/[N^2(N-1)]$, respectively.

6. Negative binomial with size r and probability p: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & x = r, r+1, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $p \in [0,1]$ and r is a positive integer. The expectation and variance of this distribution are r/p and $r(1-p)/p^2$, respectively. The moment generating function of this distribution is equal to $p^r e^{rt}/[1-(1-p)e^t]^r, \; t < -\log(1-p)$.

7. Log-distribution with probability p: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} -(\log p)^{-1} x^{-1} (1-p)^x & x = 1,2,... \\ 0 & \text{otherwise,} \end{cases}$$

where $p \in (0,1)$. The expectation and variance of this distribution are $-(1-p)/(p \log p)$ and $-(1-p)[1+(1-p)/\log p]/(p^2 \log p)$, respectively. The moment generating function of this distribution is equal to $\log[1-(1-p)e^t]/\log p, \; t \in \mathcal{R}$.

8. Uniform distribution on the interval (a,b): The Lebesgue density of this distribution is

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x),$$

where a and b are real numbers with $a < b$. The expectation and variance of this distribution are $(a+b)/2$ and $(b-a)^2/12$, respectively. The moment generating function of this distribution is equal to $(e^{bt}-e^{at})/[(b-a)t], \; t \in \mathcal{R}$.

9. Normal distribution $N(\mu, \sigma^2)$: The Lebesgue density of this distribution is
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},$$
where $\mu \in \mathcal{R}$ and $\sigma^2 > 0$. The expectation and variance of $N(\mu, \sigma^2)$ are μ and σ^2, respectively. The moment generating function of this distribution is $e^{\mu t + \sigma^2 t^2/2}$, $t \in \mathcal{R}$.

10. Exponential distribution on the interval (a, ∞) with scale parameter θ: The Lebesgue density of this distribution is
$$f(x) = \frac{1}{\theta} e^{-(x-a)/\theta} I_{(a,\infty)}(x),$$
where $a \in \mathcal{R}$ and $\theta > 0$. The expectation and variance of this distribution are $\theta + a$ and θ^2, respectively. The moment generating function of this distribution is $e^{at}(1 - \theta t)^{-1}$, $t < \theta^{-1}$.

11. Gamma distribution with shape parameter α and scale parameter γ: The Lebesgue density of this distribution is
$$f(x) = \frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} I_{(0,\infty)}(x),$$
where $\alpha > 0$ and $\gamma > 0$. The expectation and variance of this distribution are $\alpha\gamma$ and $\alpha\gamma^2$, respectively. The moment generating function of this distribution is $(1 - \gamma t)^{-\alpha}$, $t < \gamma^{-1}$.

12. Beta distribution with parameter (α, β): The Lebesgue density of this distribution is
$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1} I_{(0,1)}(x),$$
where $\alpha > 0$ and $\beta > 0$. The expectation and variance of this distribution are $\alpha/(\alpha + \beta)$ and $\alpha\beta/[(\alpha + \beta + 1)(\alpha + \beta)^2]$, respectively.

13. Cauchy distribution with location parameter μ and scale parameter σ: The Lebesgue density of this distribution is
$$f(x) = \frac{\sigma}{\pi[\sigma^2 + (x - \mu)^2]},$$
where $\mu \in \mathcal{R}$ and $\sigma > 0$. The expectation and variance of this distribution do not exist. The characteristic function of this distribution is $e^{\sqrt{-1}\mu t - \sigma|t|}$, $t \in \mathcal{R}$.

14. Log-normal distribution with parameter (μ, σ^2): The Lebesgue density of this distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2} I_{(0,\infty)}(x),$$

where $\mu \in \mathcal{R}$ and $\sigma^2 > 0$. The expectation and variance of this distribution are $e^{\mu+\sigma^2/2}$ and $e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$, respectively.

15. Weibull distribution with shape parameter α and scale parameter θ: The Lebesgue density of this distribution is

$$f(x) = \frac{\alpha}{\theta} x^{\alpha-1} e^{-x^\alpha/\theta} I_{(0,\infty)}(x),$$

where $\alpha > 0$ and $\theta > 0$. The expectation and variance of this distribution are $\theta^{1/\alpha}\Gamma(\alpha^{-1} + 1)$ and $\theta^{2/\alpha}\{\Gamma(2\alpha^{-1} + 1) - [\Gamma(\alpha^{-1} + 1)]^2\}$, respectively.

16. Double exponential distribution with location parameter μ and scale parameter θ: The Lebesgue density of this distribution is

$$f(x) = \frac{1}{2\theta} e^{-|x-\mu|/\theta},$$

where $\mu \in \mathcal{R}$ and $\theta > 0$. The expectation and variance of this distribution are μ and $2\theta^2$, respectively. The moment generating function of this distribution is $e^{\mu t}/(1 - \theta^2 t^2)$, $|t| < \theta^{-1}$.

17. Pareto distribution: The Lebesgue density of this distribution is

$$f(x) = \theta a^\theta x^{-(\theta+1)} I_{(a,\infty)}(x),$$

where $a > 0$ and $\theta > 0$. The expectation this distribution is $\theta a/(\theta - 1)$ when $\theta > 1$ and does not exist when $\theta \leq 1$. The variance of this distribution is $\theta a^2/[(\theta - 1)^2(\theta - 2)]$ when $\theta > 2$ and does not exist when $\theta \leq 2$.

18. Logistic distribution with location parameter μ and scale parameter σ: The Lebesgue density of this distribution is

$$f(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma[1 + e^{-(x-\mu)/\sigma}]^2},$$

where $\mu \in \mathcal{R}$ and $\sigma > 0$. The expectation and variance of this distribution are μ and $\sigma^2\pi^2/3$, respectively. The moment generating function of this distribution is $e^{\mu t}\Gamma(1 + \sigma t)\Gamma(1 - \sigma t)$, $|t| < \sigma^{-1}$.