A Collection of Exercises in Advanced Mathematical Statistics

The Solution Manual of All Odd-Numbered Exercises from "Mathematical Statistics" (2000)

Mohsen Soltanifar
Ph.D Candidate, Dalla Lana School of Public Health
University of Toronto, Canada

Keith Knight
Professor of Statistics, Department of Statistical Science
University of Toronto, Canada

(July 1, 2018)

Published by:

CRC Press
6000 Broken Sound Parkway, NW Suite 300, Boca Raton FL

USA office: 270 Madison Avenue, New york, NY 10016

UK office: 2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Web site: www.crcpress.com

A Collection of Exercises in Advanced Mathematical Statistics

Preface

We are pleased that this solution manual for the textbook[1] is now available. The authors hope that the textbook readers find this manual useful as they proceed through learning and teaching advanced mathematical statistics. For the student readers, it is expected the manual to be consulted after their own earlier attempts to solve the problems. For the course instructors, it is hoped the manual to be an aid in offering extra solved lecture examples and extra help.

These offered solutions are for the 110 odd-numbered problems to have a balanced situation. In one hand, it helps those self-studying readers to get some help with content and, on the other hand it allows the instructors to choose assignments and doctoral comprehensive exam questions from unsolved even-numbered problems.

Throughout the solutions, the same notational conventions as those in the textbook have been used. Furthermore, it has been heavily emphasized to the content of the textbook by frequent referral to theorems, examples and pages numbers. The goal was to help the reader to learn the textbook content by frequent referrals. In some areas, we put some gaps named "(Exercise !)" to make the readers involved in the process of solutions. On some other cases, some references were used in the solutions and were cited in the "Reference Section" at the end of manual.

This solution manual has modified some typos in the textbook content aimed to be addressed for the second edition of the textbook. The solutions themselves may have some errors. In case of any potential error, then please e-mail Professor Keith Knight in order to amend them as soon as possible. Finally, extra solutions for the solved problems are welcomed for consideration in the subsequent editions of this solution manual.

Mohsen Soltanifar,MSc, PhD(c)
Toronto, Canada. 2018
mohsen.soltanifar@mail.utoronto.ca

Professor Keith Knight,PhD
Toronto, Canada. 2018
keith@utstat.toronto.edu

[1]Knight, K. (2000) *Mathematical Statistics*, CRC Press, Boca Raton, ISBN: 1-58488-178-X

Contents

Chapter 1

Introduction to Probability

Problem 1.1. Show that

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= \sum_{i=1}^{n} P(A_i) - \sum\sum_{i<j} P(A_i \cap A_j) \\ &= \sum\sum\sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots - (-1)^n . P(A_1 \cap \cdots \cap A_n). \end{aligned}$$

Solution. We prove the assertion by induction on n. For the case $n = 1$ it trivially holds. Assume for the case $n = N$ it holds (induction hypothesis). Then, using Proposition 1.1.(c) for $A = \bigcup_{k=1}^{N} A_k$ and $B = A_{N+1}$ and two applications of induction hypothesis it follows that:

$$\begin{aligned} P(\bigcup_{k=1}^{N+1} A_k) &= P(\bigcup_{k=1}^{N} A_k) + P(A_{N+1}) - P(\bigcup_{k=1}^{N} A_k \cap A_{N+1}) \\ &= P(\bigcup_{k=1}^{N} A_k) + P(A_{N+1}) - P(\bigcup_{k=1}^{N} (A_k \cap A_{N+1})) \\ &= \sum_{k=1}^{N} (-1)^{k-1} \sum_{i_1 < \cdots < i_k} P(A_{i_1} \cap \cdots \cap A_{i_k}) + P(A_{N+1}) \\ &\quad - \sum_{k=1}^{N} (-1)^{k-1} \sum_{i_1 < \cdots < i_k} P(A_{i_1} \cap \cdots \cap A_{i_k} \cap A_{N+1}) \\ &= \sum_{k=1}^{N+1} (-1)^{k-1} \sum_{i_1 < \cdots < i_k} P(A_{i_1} \cap \cdots \cap A_{i_k}) \end{aligned}$$

proving the assertion for the case $n = N + 1$.
$\square$

Problem 1.3. Consider an experiment where a coin is tossed an infinite number of times ; the probability of heads on the kth toss is exactly one head $(1/2)^k$.

(a) Calculate (as accurately as possible) the probability that at least one head is observed.

(b) Calculate (as accurately as possible) the probability that exactly one head is observed.

Solution. (a) First, let A_k $(k \geq 1)$ be the event of head outcome on the k-th toss with $P(A_k) = \frac{1}{2^k}$.

Second, let A denote the event of at least one head in infinite number of times. Then, using $\log(1+x) \approx x \ \ (|x|<1)$ it follows that:

$$\begin{aligned} P(A) &= 1-P(A^c) = 1-P(\bigcap_{k=1}^{\infty} A_k^c) = 1-\prod_{k=1}^{\infty} P(A_k^c) = 1-\prod_{k=1}^{\infty}(1-\frac{1}{2^k}) = 1-\exp(\log(\prod_{k=1}^{\infty}(1-\frac{1}{2^k}))) \\ &= 1-\exp(\sum_{k=1}^{\infty}\log(1-\frac{1}{2^k})) \approx 1-\exp(-\sum_{k=1}^{\infty}\frac{1}{2^k}) = 1-e^{-1}. \end{aligned}$$

(b) First, let B_k $(k \geq 1)$ be the event of observing one head on the $k-$th toss and no head in the other times with

$$P(B_k) = \frac{\prod_{k=1}^{\infty}(1-\frac{1}{2^k})}{(1-\frac{1}{2^k})}.(\frac{1}{2^k}) = \frac{\prod_{k=1}^{\infty}(1-\frac{1}{2^k})}{2^k-1}.$$

Second, let B be the event of exactly one head in infinite number of times. Then, using $\sum_{k=1}^{10^6}\frac{1}{2^k-1} = 1.6067$ it follows that:

$$P(B) = P(\bigcup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} P(B_k) = \sum_{k=1}^{\infty}\frac{\prod_{k=1}^{\infty}(1-\frac{1}{2^k})}{2^k-1} = (\prod_{k=1}^{\infty}(1-\frac{1}{2^k}))(\sum_{k=1}^{\infty}\frac{1}{2^k-1}) \approx 1.6067e^{-1}.$$

□

Problem 1.5. (a) Suppose that $\{A_n\}$ is a decreasing sequence of events with limit A; that is $A_{n+1} \subset A_n$ for all $n \geq 1$ with

$$A = \bigcap_{n=1}^{\infty} A_n$$

Using Axioms of Probability show that

$$\lim_{n\to\infty} P(A_n) = P(A).$$

(b)Let X be a random variable and suppose that $\{x_n\}$ is strictly increasing sequence of numbers(that is, $x_n > x_{n+1}$ for all n)whose limit is x_0. Define $A_n = [X \leq x_n]$. Show that

$$\bigcap_{n=1}^{\infty} A_n = [X \leq x_0]$$

and hence (using part (a)) that $P(X_n \leq x_n) \to P(X \leq x_0)$.
(c) Now let $\{x_n\}$ be strictly increasing sequence of numbers (that is, $x_n < x_{n+1}$ for all n)whose limit is x_0. Again defining $A_n = [X \leq x_n]$ Show that

$$\bigcup_{n=1}^{\infty} A_n = [X < x_0]$$

and hence that $P(X_n \leq x_n) \to P(X < x_0)$.

Solution. (a) Using Proposition 1.1.(d) for the increasing sequence $\{A_n^c\}$ it follows that:

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} 1-P(A_n^c) = 1-\lim_{n\to\infty} P(A_n^c) = 1-P(\bigcup_{n=1}^{\infty} A_n^c) = P(\bigcap_{n=1}^{\infty} A_n) = P(A).$$

(b) First, as $x_n \downarrow x_0$ and $[X \leq x_0] \subseteq [X_n \leq x_n]$ $(n \geq 1)$ it follows that:

$$[X \leq x_0] \subseteq \bigcap_{n=1}^{\infty} A_n. \qquad (*)$$

Second, let $w \in \bigcap_{n=1}^{\infty} A_n$, then $X(w) \leq x_n$ for all $n \geq 1$; and, consequently, $X(w) \leq \inf_{n\geq 1}(x_n) = \lim_{n\to\infty} x_n = x_0$ implying $w \in [X \leq x_0]$. Hence:

$$\bigcap_{n=1}^{\infty} A_n \subseteq [X \leq x_0]. \qquad (**)$$

Now, by $(*)$ and $(**)$ the assertion follows.

(c) First, as $x_n \uparrow x_0$ and $[X \leq x_n] \subseteq [X < x_0]$ $(n \geq 1)$ it follows that:

$$\bigcup_{n=1}^{\infty} A_n \subseteq [X < x_0]. \qquad (***)$$

Second, let $w \in [X < x_0]$, then $X(w) < x_0$; but, $\sup_{n\geq 1}(x_n) = \lim_{n\to\infty} x_n = x_0$ and hence for some N we have $X(w) \leq x_N < x_0$ yielding $w \in [X \leq x_N] = A_N \subseteq \bigcup_{n=1}^{\infty} A_n$. Thus,

$$[X < x_0] \subseteq \bigcup_{n=1}^{\infty} A_n. \qquad (****)$$

Now, by $(***)$ and $(****)$ the assertion follows.
$\square$

Problem 1.7. Suppose that $F_1(x), \cdots, F_k(x)$ are distribution functions.
(a) Show that $G(x) = p_1.F_1(x) + \cdots + p_k.F_k(x)$ is a distribution function provided that $p_i \geq 0 (i = 1, \cdots, k)$ and $p_1 + \cdots + p_k = 1$.
(b)If $F_1(x), \cdots, F_k(x)$ have density (frequency) functions $f_1(x), \cdots, f_k(x)$, show that $G(x)$ defined in (a) has density (frequency) function $g(x) = f_1(x) + \cdots + f_k(x)$.

Solution. (a) It is sufficient to prove basic properties of the distribution function for G. First, let $x \leq y$ then $F_i(x) \leq F_i(y)$, $(1 \leq i \leq k)$ implying:

$$G(x) = \sum_{i=1}^{k} p_i.F_i(x) \leq \sum_{i=1}^{k} p_i.F_i(y) = G(y).$$

Second, as $\lim_{y\downarrow x} F_i(y) = F_i(x)$ $(1 \leq i \leq k)$, it follows that:

$$\lim_{y\downarrow x} G(y) = \lim_{y\downarrow x} \sum_{i=1}^{k} p_i.F_i(y) = \sum_{i=1}^{k} p_i. \lim_{y\downarrow x} F_i(y) = \sum_{i=1}^{k} p_i.F_i(x) = G(x).$$

Third, as $\lim_{y\uparrow\infty} F_i(y) = 1$ $(1 \leq i \leq k)$, it follows that:

$$\lim_{y\uparrow\infty} G(y) = \lim_{y\uparrow\infty} \sum_{i=1}^{k} p_i.F_i(y) = \sum_{i=1}^{k} p_i. \lim_{y\uparrow\infty} F_i(y) = \sum_{i=1}^{k} p_i.1 = 1.$$

Finally, the case $\lim_{y\downarrow -\infty} G(y) = 0$ is left for the reader as Exercise.

(b) First, let X be a continuous random variable. Then, as $f_i(x) \geq 0 (1 \leq i \leq k)$ it follows that $g(x) = \sum_{i=1}^k p_i.f_i(x) \geq 0$ for all x. Next, let $-\infty < a < b < \infty$, then:

$$\begin{aligned} P_G(a \leq X \leq b) &= G(b) - G(a) = \sum_{i=1}^k p_i.F_i(b) - \sum_{i=1}^k p_i.F_i(a) = \sum_{i=1}^k p_i.(F_i(b) - F_i(a)) \\ &= \sum_{i=1}^k p_i.P_{F_i}(a \leq X \leq b) = \sum_{i=1}^k p_i. \int_a^b f_i(x)dx = \int_a^b \sum_{i=1}^k p_i.f_i(x)dx = \int_a^b g(x)dx. \end{aligned}$$

Second, let X be a discrete random variable. Since $f_i(x) = P_{F_i}(X = x)\ \ (1 \leq x \leq k)$ for all x it follows that:

$$\begin{aligned} g(x) &= \sum_{i=1}^k p_i.f_i(x) = \sum_{i=1}^k p_i.P_{F_i}(X = x) = \sum_{i=1}^k p_i.(F_i(x) - \lim_{y\uparrow x} F_i(y)) \\ &= \sum_{i=1}^k p_i.F_i(x) - \lim_{y\uparrow x} \sum_{i=1}^k p_i.F_i(y) = G(x) - \lim_{y\uparrow x} G(y) = P_G(X = x). \end{aligned}$$

□

Problem 1.9. Suppose that X is a random variable with distribution function F and inverse (or quantile function) F^{-1}. Show that

$$E(X) = \int_0^1 F^{-1}(t)dt$$

if $E(X)$ is well-defined.

Solution. It is sufficient to prove the assertion for $X \geq 0$ (Exercise). First, assume for some $0 < b < \infty$ to have $F_X(b) = 1$. Then, an application of Charles-Ange Laisant formulae for function $F_X : [0, b] \to [0, 1]$ yields:

$$\int_0^1 F^{-1}(t)dt + \int_0^b F(x)dx = b,$$

or equivalently:

$$\int_0^1 F^{-1}(t)dt = \int_0^b (1 - F(x))dx. \qquad (*)$$

On the other hand, by definition:

$$E(X) = \int_0^b (1 - F(x))dx, \qquad (**)$$

and a comparison of $(*)$ and $(**)$ proves the assertion for this case. Second, let for all $0 < b < \infty$ to have $F_X(b) \neq 1$. Consider an increasing sequence $\{x_n\}$ such that $\int_{x_n}^{\infty} (1 - F(x))dx \leq \frac{1}{n}$ $(n \geq 1)$ and $x_n \uparrow \infty$. Define $X_n = X.1_{[0,x_n]}$ $(n \geq 1)$, the by above conditions:

$$\begin{aligned} F_{X_n}(x) &= F_X(x).1_{[0,x_n]}(x) + 1_{(x_n,\infty)}(x), \ \ (n \geq 1) \\ \int_0^1 (F_X^{-1}(t) - F_{X_n}^{-1}(t))dt &= \int_{x_n}^{\infty} (1 - F(x))dx \leq \frac{1}{n}\ (n \geq 1). \end{aligned}$$

Consequently, an application of the first case yields:

$$E(X) = \lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} \int_0^1 F_{X_n}^{-1}(t)dt = \int_0^1 F_X^{-1}(t)dt.$$

□

Problem 1.11. Let X be a random variable with finite expected value $E(X)$ and suppose that $g(x)$ is a convex function:

$$g(tx + (1-t)y) \leq tg(x) + (1-t)g(y)$$

for $0 \leq t \leq 1$.
(a) Show that for any x_0, there exists a linear function $h(x) = ax + b$ such that $h(x_0) = g(x_0)$ and $h(x) \leq g(x)$ for all x.
(b) Prove Jensen's inequality: $g(E(X)) \leq E(g(X))$.

Solution. (a) For any x_0, the left derivative $g^{'}(x_0^-)$ exists and it is sufficient to consider the left tangent line at x_0 given by $h(x) = g^{'}(x_0^-).(x - x_0) + g(x_0)$. Next, considering:

$$\begin{aligned} g(x) &\geq \frac{g(t.x + (1-t).x_0) - (1-t)g(x_0)}{t} = \frac{g(t.x + (1-t).x_0) - g(x_0)}{t} + g(x_0) \\ &= \frac{g(t.x + (1-t).x_0) - g(x_0)}{tx + (1-t)x_0 - x_0}(x - x_0) + g(x_0) \qquad (*) \end{aligned}$$

and taking limit from both sides of $(*)$ as $t \downarrow 0$, it follows that $g(x) \geq h(x)$.
(b) For $h(x) = ax + b$ we have:

$$E(g(X)) \geq E(h(X)) = E(aX + b) = aE(X) + b = h(E(X)) = g(E(X)).$$

□

Problem 1.13. Suppose $X \sim Gamma(\alpha, \lambda)$. Show that
(a) $E(X^r) = \Gamma(r+\alpha)/(\lambda^r\Gamma(\alpha))$for $r > -\alpha$;
(b) $Var(X) = \alpha/\lambda^2$.

Solution.(a)

$$E(X^r) = \int_0^\infty x^r.f(x)dx = \int_0^\infty \frac{\lambda^\alpha.x^{\alpha+r-1}}{\Gamma(\alpha)}.e^{-\lambda.x}dx = \frac{\Gamma(r+\alpha)}{\lambda^r.\Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+r}.x^{\alpha+r-1}}{\Gamma(\alpha+r)}.e^{-\lambda.x}dx = \frac{\Gamma(r+\alpha)}{\lambda^r.\Gamma(\alpha)}.$$

(b)

$$Var(X) = E(X^2) - E^2(X) = \frac{\Gamma(2+\alpha)}{\lambda^2.\Gamma(\alpha)} - (\frac{\Gamma(1+\alpha)}{\lambda^1.\Gamma(\alpha)})^2 = \frac{(\alpha+1)(\alpha)\Gamma(\alpha)}{\lambda^2.\Gamma(\alpha)} - (\frac{(\alpha)\Gamma(\alpha)}{\lambda^1.\Gamma(\alpha)})^2 = \frac{\alpha}{\lambda^2}.$$

□

Problem 1.15. Suppose that $X \sim N(0,1)$.
(a) Show that $E(X^k) = 0$ if k is odd.

(b) Show that $E(X^k) = 2^{k/2}\Gamma((k+1)/2)/\Gamma(1/2)$ if k is even.

Solution.(a) For odd k, the function $g(x) = x^k.f_X(x)$ is an integrable odd function over real line and hence, $E(X) = \int_{-\infty}^{\infty} g(x)dx = 0$.

(b) Let $k = 2m$. Since $\frac{X^2}{2} = W \sim Gamma(1/2, 1)$, it follows from Problem 1.13(a) that:

$$E(X^k) = E(X^{2m}) = 2^m.E(W^m) = 2^m\frac{\Gamma(1/2+m)}{\Gamma(1/2)} = 2^{k/2}\frac{\Gamma((1+k)/2)}{\Gamma(1/2)}.$$

□

Problem 1.17. Let $m(t) = E[exp(tX)]$ be the moment generating function of X. $c(t) = ln(m_X(t))$ is often called the cumulant generating function of X.
(a) Show that $c'(0) = E(X)$ and $c"(0) = Var(X)$.
(b) Suppose that X has a Poisson distribution with parameter λ as in Example 1.33. Use the cumulant generating function of X to show that $E(X) = Var(X) = \lambda$.
(c) The mean and variance are the first two cumulants of a distribution; in general, the k-th cumulant is defined to be $c^{(k)}(0)$. Show that the third and fourth cumulants are

$$\begin{aligned} c^{(3)}(0) &= E(X^3) - 3E(X)E(X^2) + 2[E(X)]^3, \\ c^{(4)}(0) &= E(X^4) - 4E(X^3)E(X) + 12E(X^2)[E(X)]^2 - 3[E(X^2)]^2 - 6[E(X)]^4 \end{aligned}$$

(d) Suppose that $X \sim N(\mu, \sigma^2)$. Show that all but the first two cumulants of X are exactly 0.

Solution. (a)

$$c^{'}(0) = \frac{d}{dt}\log(m_X(t))|_{t=0} = \frac{m_X^{'}(t)}{m_X(t)}|_{t=0} = \frac{E(X)}{1} = E(X),$$

$$c^{"}(0) = \frac{d^2}{dt^2}\log(m_X(t))|_{t=0} = \frac{m_X^{"}(t)m_X(t) - m_X^{'}(t)m_X^{'}(t)}{m_X^2(t)}|_{t=0} = \frac{E(X^2) - E^2(X)}{1} = Var(X).$$

(b) Since, $c_X(t) = \log(m_X(t)) = \log(\exp(\lambda.(e^t - 1))) = \lambda.e^t - 1$, by part (a) it follows that:

$$E(X) = c^{'}(0) = \lambda.e^t|_{t=0} = \lambda,$$

$$E(X) = c^{"}(0) = \lambda.e^t|_{t=0} = \lambda.$$

(c) First, we have:

$$\begin{aligned} c(t) &= \log(m_X(t)), \\ c^{(1)}(t) &= m_X^{(1)}(t).m_X^{-1}(t), \\ c^{(2)}(t) &= m_X^{(2)}(t).m_X^{-1}(t) - (m_X^{(1)}(t))^2.m_X^{-2}(t), \\ c^{(3)}(t) &= m_X^{(3)}(t).m_X^{-1}(t) - 3.m_X^{(2)}(t).m_X^{(1)}(t).m_X^{-2}(t) + 2(m_X^{(1)}(t))^3.m_X^{-3}(t), \\ c^{(4)}(t) &= m_X^{(4)}(t).m_X^{-1}(t) - 4.m_X^{(3)}(t).m_X^{(1)}(t).m_X^{-2}(t) - 3(m_X^{(2)}(t))^2.m_X^{-2}(t) \\ &+ 12.(m_X^{(1)}(t))^2.m_X^{(2)}(t).m_X^{-3}(t) - 6(m_X^{(1)}(t))^4.m_X^{-4}(t), \end{aligned}$$

and using $m_X^{(i)}(t) = E(X^i)$, $(0 \leq i \leq 4)$ the assertion follows.

(d)Since $c(t) = \log(m_X(t)) = \log(\exp(\mu.t + \frac{\sigma^2.t^2}{2})) = \mu.t + \frac{\sigma^2.t^2}{2}$, it follows that:

$$c^{(1)}(t) = \mu + \sigma^2.t, \quad c^{(2)}(t) = \sigma^2, \quad c^{(n)}(t) = 0, \quad (n \geq 3) \qquad (*)$$

and letting $t = 0$ in $(*)$ the assertion is proved.
□

Problem 1.19. The Gompertz distribution is sometimes used as a model for the length of human life; this model is particular good for modelling survival beyond 40 years. Its distribution function is:

$$F(x) = 1 - \exp[-\beta(\exp(\alpha x) - 1)] \qquad for x \geq 0$$

where $\alpha, \beta > 0$.
(a) Find the hazard function for this distribution.
(b) Suppose that X has distribution function F. Show that

$$E(X) = \frac{\exp(\beta)}{\alpha} \int_1^\infty \frac{\exp(-\beta.t)}{t} dt$$

while the median of F is

$$F^{-1}(1/2) = \frac{1}{\alpha} ln(1 + ln(2)/\beta).$$

(c) Show that $F^{-1}(1/2) \geq E(X)$ for all $\alpha > 0, \beta > 0$.

Solution. (a)

$$\lambda(x) = \frac{\frac{d}{dx}F(x)}{1 - F(x)} = \frac{\beta.\alpha.\exp(\alpha x) * \exp[-\beta(\exp(\alpha x) - 1)]}{\exp[-\beta(\exp(\alpha x) - 1)]} = \beta.\alpha.\exp(\alpha x). \quad (x \geq 0)$$

(b)First, using change of variable technique with $t = \exp(\alpha.x)$ and $dt = \alpha.tdx$ it follows that:

$$\begin{aligned} E(X) &= \int_0^\infty (1 - F(x))dx = \int_0^\infty \exp[-\beta(\exp(\alpha x) - 1)]dx = \exp(\beta) \int_0^\infty \exp[-\beta(\exp(\alpha x))]dx \\ &= \exp(\beta) \int_1^\infty \exp(-\beta.t) \frac{dt}{\alpha.t} = \frac{\exp(\beta)}{\alpha} \int_1^\infty \frac{\exp(-\beta.t)}{t} dt. \end{aligned}$$

Second, solving equation $F(x) = \frac{1}{2}$, one concludes:

$$\exp[-\beta(\exp(\alpha x) - 1)] = \frac{1}{2} \Leftrightarrow \exp(\alpha x) - 1 = \frac{\log(2)}{\beta} \Leftrightarrow Med(x) = \frac{\log(1 + \frac{\log(2)}{\beta})}{\alpha}.$$

(c) Fix $\alpha > 0$, and define :

$$\begin{aligned} H(\beta) &= \alpha.(F^{-1}(1/2) - E(X)) \\ &= \alpha.(\frac{1}{\alpha} ln(1 + ln(2)/\beta) - \frac{\exp(\beta)}{\alpha} \int_1^\infty \frac{\exp(-\beta.t)}{t} dt) \\ &= \log(1 + \frac{\log(2)}{\beta}) - \int_0^\infty \frac{e^{-\beta.t}}{t + 1} dt. \end{aligned}$$

The following plot of H shows that it takes both positive and negative values. Hence, the given inequality does not hold for all $\beta > 0$.

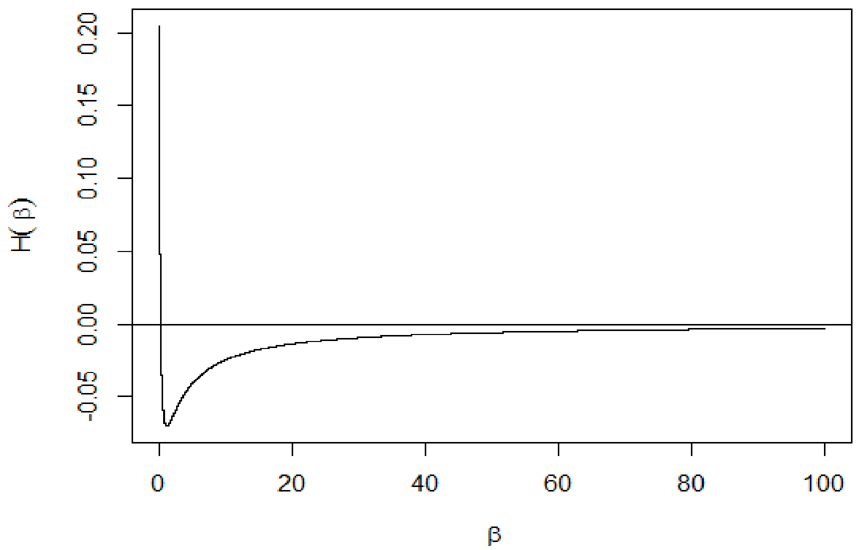

Figure 1.1 Plot of function $H(\beta) = \log(1 + \frac{\log(2)}{\beta}) - \int_0^{\infty} \frac{e^{-\beta.t}}{t+1} dt$

□

Problem 1.21. Suppose that X is a non-negative random variable where $E(X^r)$ is finite for some $r > 0$. Show that $E(X^s)$ is finite for $0 \leq s \leq r$.

Solution. For given $0 \leq s \leq r$, we have $x^{s-1} \leq x^{r-1} \quad (x \geq 1)$. Hence, by Problem 1.20(a) and the given assumption it follows that:

$$
\begin{aligned}
E(X^s) &= s\int_0^{\infty} x^{s-1}(1 - F(x))dx = s(\int_0^1 x^{s-1}(1 - F(x))dx + \int_1^{\infty} x^{s-1}(1 - F(x))dx) \\
&= s(\int_0^1 x^{s-1}(1 - F(x))dx) + s(\int_1^{\infty} x^{s-1}(1 - F(x))dx) \leq s(\int_0^1 x^{s-1}dx) + r(\int_1^{\infty} x^{r-1}(1 - F(x))dx) \\
&\leq 1 + r(\int_0^{\infty} x^{r-1}(1 - F(x))dx) = 1 + E(X^r) < \infty.
\end{aligned}
$$

□

Problem 1.23. Suppose that X is a non-negative random variable with distribution function $F(x) = P(X \leq x)$. Show that

$$E(X^r) = r\int_0^{\infty} x^{r-1}(1 - F(x))dx,$$

for any $r > 0$.

Solution. Using Fubini's Theorem for $0 < x < t < \infty$ we have:

$$
\begin{aligned}
\int_0^{\infty} r.x^{r-1}(1 - F(x))dx &= \int_0^{\infty} r.x^{r-1}(\int_x^{\infty} f(t)dt)dx = \int_0^{\infty}\int_x^{\infty}(r.x^{r-1}f(t))dtdx \\
&= \int_0^{\infty}\int_0^{t}(r.x^{r-1}f(t))dxdt = \int_0^{\infty} t^r f(t)dt = E(X^r).
\end{aligned}
$$

□

Problem 1.25. Suppose that X has a distribution function $F(x)$ with inverse $F^{-1}(t)$.
(a) Suppose also that $E(|X|) < \infty$ and define $g(t) = E(|X-t|)$. Show that g is minimized at $t = F^{-1}(1/2)$.
(b) The assumption that $E(|X|) < \infty$ in (a) is unnecessary if we define $g(t) = E[|X-t| - |X|]$. Show that $g(t)$ is finite for all t and that $t = F^{-1}(1/2)$ minimizes $g(t)$.
(c) Define $\rho_\alpha(x) = \alpha.x.I(x \geq 0) + (\alpha - 1).x.I(x < 0)$ for some $0 < \alpha < 1$. Show that $g(t) = E[\rho_\alpha(X-t) - \rho_\alpha(X)]$ is minimized at $t = F^{-1}(\alpha)$.

Solution.(a) Since:

$$\begin{aligned} g(t) &= E(|X-t|) = \int_{-\infty}^{\infty} |x-t| f(x)dx = \int_{-\infty}^{t} -(x-t)f(x)dx + \int_{t}^{\infty}(x-t)f(x)dx \\ &= -\int_{-\infty}^{t} x.f(x)dx + t.\int_{-\infty}^{t} f(x)dx + \int_{t}^{\infty} x.f(x)dx - t\int_{t}^{\infty} f(x)dx, \end{aligned}$$

we have:

$$\frac{d}{dt}g(t) = -t.f(t) + F(t) + t.f(t) + t.f(t) - (1-F(t)) - t.f(t) = 2.F(t) - 1 = 0,$$

and consequently $t = F^{-1}(1/2)$ minimizes g.

(b) First,

$$|g(t)| = |E(|X-t| - |X|)| \leq E(||X-t| - |X||) \leq E(|X-t-X|) = |t| < \infty.$$

Second,

$$\begin{aligned} g(t) &= E(|X-t| - |X|) = \int_{-\infty}^{\infty} (|x-t| - |x|)f(x)dx \\ &= \int_{-\infty}^{\infty} [-1_{(-\infty,t)}(x)(x-t) + 1_{(t,\infty)}(x)(x-t) - (-1_{(-\infty,0)}(x).x + 1_{(0,\infty)}(x).x)]f(x)dx \\ &= \int_{-\infty}^{\infty} [x(-1_{(-\infty,t)}(x) + 1_{(t,\infty)}(x) + 1_{(-\infty,0)}(x) - 1_{(0,\infty)}(x)) + t(1_{(-\infty,t)}(x) - 1_{(t,\infty)}(x))]f(x)dx \\ &= \int_{-\infty}^{\infty} [x((1 - 2.1_{(-\infty,t)}(x)) + (2.1_{(-\infty,0)}(x) - 1)) + t(2.1_{(-\infty,t)}(x) - 1)]f(x)dx \\ &= 2(\int_{-\infty}^{0} x.f(x)dx - \int_{-\infty}^{t} x.f(x)dx) + t.(2F(t) - 1), \end{aligned}$$

yields:

$$\frac{d}{dt}g(t) = -2t.f(t) + 2F(t) - 1 + 2t.f(t) = 2F(t) - 1 = 0,$$

and thus $t = F^{-1}(1/2)$ minimizes g.

(c) Given $\frac{d}{dt}(\rho_\alpha(x-t) - \rho_\alpha(x))) = -\alpha + 1_{(-\infty,t)}(x)$, we may generalize the solution in part (b) as follows:

$$\frac{d}{dt}g(t) = E(\frac{d}{dt}(\rho_\alpha(X-t) - \rho_\alpha(X))) = E(-\alpha + 1_{(-\infty,t)}(X)) = -\alpha + F(t) = 0,$$

and hence $t = F^{-1}(\alpha)$ minimizes g.
□

Problem 1.27. Let X be a positive random variable with distribution function F. Show that $E(X) < \infty$ if, and only if,

$$\sum_{k=1}^{\infty} P(X > k\epsilon) < \infty$$

for any $\epsilon > 0$.

Solution. First, to prove the necessity, let $\epsilon = \frac{1}{2}$, and consider $[X] \leq X < [X] + 1$, then by rearrangement of sums:

$$\begin{aligned} E([X]) &= \sum_{l=1}^{\infty} l.P([X] = l) = \sum_{l=1}^{\infty} P(l \leq X < l+1) \\ &= \sum_{k=1}^{\infty} \sum_{m=0}^{\infty} P(k+m \leq X < k+m+1) = \sum_{k=1}^{\infty} P(X \geq k) \\ &< \sum_{k=1}^{\infty} P(X > \frac{k}{2}) < \infty, \end{aligned}$$

and hence by $E(X) < E([X]) + 1$, the assertion follows.

Second, to prove the sufficiency, let $\epsilon > 0$, then :

$$\sum_{k=1}^{\infty} P(X > k.\epsilon) \leq \sum_{k=1}^{\infty} P(X \geq k.\epsilon) = \sum_{k=1}^{\infty} P(\frac{X}{\epsilon} \geq k) = E([\frac{X}{\epsilon}]) \leq E(\frac{X}{\epsilon}) = \frac{1}{\epsilon}.E(X) < \infty.$$

□

Chapter 2

Random Vector and Joint Distribution

Problem 2.1. Suppose that X and Y are independent Geometric random variables with frequency function $f(x) = \theta.(1-\theta)^x$ for $x = 0, 1, 2, \cdots$

(a) Show that $Z = X + Y$ has a Negative Binomial distribution and identify the parameters of Z.
(b) Extend the result of part (a): If $X_1, \cdots, X_n$ are i.i.d. Geometric random variables, show that $S = X_1 + \cdots + X_n$ has a Negative Binomial distribution and identify the parameters of S.

Solution. (a) Using $C(n,k)$ as the notation for binomial coefficient we have:

$$
\begin{aligned}
P(Z = z) &= P(X + Y = z) \\
&= \sum_{y=0}^{\infty} P(X + Y = z | Y = y) P(Y = y) \\
&= \sum_{y=0}^{\infty} P(X = z - y) P(Y = y) \\
&= \sum_{y=0}^{z} P(X = z - y) P(Y = y) \\
&= \sum_{y=0}^{z} (\theta.(1-\theta)^{z-y}.\theta.(1-\theta)^y) \\
&= (z+1)\theta^2.(1-\theta)^z \\
&= C(2 + z - 1, z).\theta^2.(1-\theta)^z.
\end{aligned}
$$

So, $Z \sim NB(2, \theta)$.

(b)We claim $S_n \sim NB(n, \theta)$ $(n \geq 1)$. For the case $n = 1$ as $S_1 = X_1$ it trivially holds. Let it hold for the case $n > 1$(induction hypothesis). Then, for $S_{n+1} = S_n + X_{n+1}$ it follows that:

$$\begin{aligned}
P(S_{n+1} = s) &= P(S_n + X_{n+1} = s) = \sum_{x=0}^{\infty} P(S_n + X_{n+1} = s|X_{n+1} = x)P(X_{n+1} = x) \\
&= \sum_{x=0}^{\infty} P(S_n = s - x)P(X_{n+1} = x) = \sum_{x=0}^{s} P(S_n = s - x)P(X_{n+1} = x) \\
&= \sum_{x=0}^{s} C(n + s - x - 1, s - x).\theta^n.(1 - \theta)^{s-x}.\theta.(1 - \theta)^x \\
&= \sum_{x=0}^{s} C(n + s - x - 1, s - x).\theta^{n+1}.(1 - \theta)^s \\
&= C(n + 1 + s - 1, s)\theta^{n+1}.(1 - \theta)^s, \quad s = 0, 1, \cdots .
\end{aligned}$$

where in the last equation, the equality $\sum_{x=0}^{s} C(n + s - x - 1, s - x) = C(n + s, s)$ was used in which can be proved by induction on s and Pascal's rule for binomial coefficients. Consequently, $S_{n+1} \sim NB(n + 1, \theta)$.
□

Problem 2.3. If $f_1(x), \cdots, f_k(x)$ are density(frequency) functions then

$$g(x) = p_1.f_1(x) + \cdots + p_k.f_k(x)$$

is also a density (frequency) function provided that $p_i \geq 0(i = 1, \cdots, k)$ and $p_1 + \cdots + p_k = 1$. We can thin of sampling from $g(x)$ as first sampling a discrete random variable Y taking values 1 through k with probabilities $p_1, \cdots, p_k$ and then, conditional on $Y = i$, sampling from $f_i(x)$. The distribution whose density or frequency function is $g(x)$ is called a mixture distribution.
(a) Suppose that X has frequency function $g(x)$. Show that

$$P(Y = i|X = x) = \frac{p_i f_i(x)}{g(x)}$$

provided that $g(x) > 0$.
(b) Suppose that X has a density function $g(x)$. Show that we can reasonably define

$$P(Y = i|X = x) = \frac{p_i f_i(x)}{g(x)}$$

in the sense that $P(Y_i = i) = E(P(Y = i|X))$.

Solution.(a) Using Bayes Theorem (Proposition 1.5.) it follows that:

$$P(Y = i|X = x) = \frac{P(X = x|Y = i)P(Y = i)}{P(X = x)} = \frac{f_i(x).p_i}{g(x)}, (g(x) > 0).$$

(b)We prove the assertion for discrete random variable X :

$$\begin{aligned}
E(P(Y = i|X)) &= \sum_x P(Y = i|X = x)P(X = x) = \sum_x \frac{p_i.f_i(x)}{g(x)} g(x) \\
&= p_i.\sum_x f_i(x) = p_i = P(Y = i) \ (1 \leq i \leq k).
\end{aligned}$$

The proof for the continuous random variable X is similar by replacing sums in above by integrals.
□

Problem 2.5. Mixture distributions can be extended in the following way. Suppose that $f(x;\theta)$ is a density or a frequency function where θ lies in some set $\Theta \subseteq R$. Let $p(\theta)$ be a density function on Θ and define

$$g(x) = \int_{\Theta} f(x;\theta)p(\theta)d\theta.$$

Then $g(x)$ is itself a density or frequency function. As before, we can review sampling from $g(x)$ as first sampling from $p(\theta)$ and that given θ, sampling from $f(x;\theta)$.
(a) Suppose that X has the mixture density or frequency function $g(x)$. Show that

$$E(X) = E(E(X|\theta))$$

and

$$Var(X) = Var(E(X|\theta)) + E(Var(X|\theta))$$

where $E(X|\theta)$ and $Var(X|\theta)$ are the mean and variance of a random variable with density or frequency function $f(x;\theta)$.
(b) The Negative Binomial distribution introduced in Example 1.12 can be obtained as a Gamma mixture of Poisson distributions. Let $f(x;\lambda)$ be a Poisson frequency function with mean λ and $p(\lambda)$ be a Gamma distribution with mean μ and variance μ^2/α. Show that the mixture distribution has frequency function

$$g(x) = \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)}(\frac{\alpha}{\alpha+\mu})^{\alpha}(\frac{\mu}{\alpha+\mu})^{x}$$

for $x = 0, 1, 2, \cdots$. Note that this form of the Negative Binomial is richer than the form given in Example 1.12.
(c) Suppose that X has a Negative Binomial distribution as given in part (b). Find the mean and variance of X.
(d) Show that the moment generating function of the Negative Binomial distribution in (b) is

$$m(t) = (\frac{\alpha}{\alpha+\mu(1-\exp(t))})^{\alpha}. \quad \text{for} \quad t < ln(1+\alpha/\mu)$$

Solution.(a) First, using Fubini's Theorem it follows that:

$$\begin{aligned} E(E(X|\theta)) &= \int_{\Theta} E(X|\theta)p(\theta)d\theta = \int_{\Theta}(\int_{\chi} x.f(x;\theta)dx)p(\theta)d\theta \\ &= \int_{\Theta}\int_{\chi}(x.f(x;\theta)p(\theta))dxd\theta = \int_{\chi} x.(\int_{\Theta} f(x;\theta)p(\theta)d\theta)dx \\ &= \int_{\chi} x.g(x)dx = E(X). \end{aligned}$$

Second, using $E(Y) = E(E(Y|\theta))$ for $Y = (X - E(X))^2$ we have:

$$\begin{aligned} Var(X) &= E((X-E(X))^2) = E(E((X-E(X))^2|\theta)) = E(E(X^2 - 2.E(X).X + E^2(X)|\theta)) \\ &= E(E(X^2|\theta) - 2.E(X.E(X)|\theta) + E^2(X)) = E(E(X^2|\theta) - 2E(X).E(X|\theta) + E^2(X)) \\ &= E((E(X^2|\theta) - E^2(X|\theta)) + (E^2(X|\theta) - 2E(X).E(X|\theta) + E^2(X))) \\ &= E(E(X^2|\theta) - E^2(X|\theta)) + E((E(X|\theta) - E(X))^2) \\ &= E(Var(X|\theta)) + E((E(X|\theta) - E(E(X|\theta)))^2) \\ &= E(Var(X|\theta)) + Var(E(X|\theta)). \end{aligned}$$

(b)

$$\begin{aligned}
g(x) &= \int_0^\infty f(x,\lambda)p(\lambda)d\lambda = \int_0^\infty (\frac{e^{-\lambda}\lambda^x}{x!} * \frac{\lambda^{\alpha-1}.e^{-\alpha.\lambda/\mu}}{\Gamma(\alpha)(\mu/\alpha)^\alpha})d\lambda \\
&= \frac{1}{x!\Gamma(\alpha)}\int_0^\infty \frac{\lambda^{x+\alpha-1}}{(\mu/\alpha)^\alpha}.e^{-(1+\alpha/\mu)\lambda}d\lambda \\
&= \frac{1}{x!\Gamma(\alpha)}.\frac{1}{(\mu/\alpha)^\alpha}\int_0^\infty \frac{((1+\alpha/\mu)\lambda)^{x+\alpha-1}}{(1+\alpha/\mu)^{x+\alpha}}.e^{-(1+\alpha/\mu)\lambda}d((1+\alpha/\mu)\lambda) \\
&= \frac{1}{x!\Gamma(\alpha)}.\frac{\alpha^\alpha}{\mu^\alpha}.\frac{\mu^{x+\alpha}}{(\mu+\alpha)^{x+\alpha}}.\Gamma(1) \\
&= \frac{1}{x!\Gamma(\alpha)}.(\frac{\alpha}{\alpha+\mu})^\alpha.(\frac{\mu}{\alpha+\mu})^x \quad x = 0,1,\cdots.
\end{aligned}$$

(c)

$$E(X) = E(E(X|\lambda)) = E(\lambda) = \mu.$$

$$Var(X) = Var(E(X|\lambda)) + E(Var(X|\lambda)) = Var(\lambda) + E(\lambda) = \frac{\mu^2}{\alpha} + \mu.$$

(d)Using Poisson moment generating function and Gamma moment generating function we have:

$$\begin{aligned}
M_X(t) &= E(e^{t.X}) = E(E(e^{t.X}|\lambda)) = E(M_{X|\lambda}(t)) = E(e^{\lambda(e^t-1)}) \\
&= M_\lambda(e^t-1) = (\frac{1}{1-\frac{\mu}{\alpha}(e^t-1)})^\alpha = (\frac{\frac{\alpha}{\alpha+\mu}}{(\frac{\alpha}{\alpha+\mu})(1+\frac{\mu}{\alpha}-\frac{\mu}{\alpha}.e^t)})^\alpha \\
&= (\frac{\frac{\alpha}{\alpha+\mu}}{(1-\frac{\mu}{\alpha+\mu}).e^t)})^\alpha = (\frac{\frac{\alpha}{\alpha+\mu}}{(1-(1-\frac{\alpha}{\alpha+\mu})).e^t)})^\alpha. \text{ for } t < ln(1+\frac{\alpha}{\mu})
\end{aligned}$$

□

Problem 2.7. Suppose that $X_1,\cdots$ are i.i.d. random variables with moment generating function $m(t) = E(exp(t.X_i))$. Let N be a Poisson random variable (independent of X_i's) with parameter λ and define the compound Poisson random variable

$$S = \sum_{i=1}^N X_i$$

where $S = 0$ if $N = 0$.
(a) Show that the moment generating function of S is

$$E(\exp(tS)) = \exp(\lambda(m(t)-1)).$$

(b) Suppose that the X_i's are Exponential with $E(X_i) = 1$ and $\lambda = 5$. Evaluate $P(S > 5)$.

Solution.(a)

$$\begin{aligned}
M_S(t) &= E(e^{t.S}) = E(E(e^{t.S}|N)) = E(E(e^{t.\sum_{i=1}^N X_i}|N)) \\
&= E(E(\prod_{i=1}^N e^{t.X_i}|N)) = E(\prod_{i=1}^N E(e^{t.X_i}|N)) \\
&= E(\prod_{i=1}^N E(e^{t.X_i})) = E(m(t)^N) = E(e^{\log(m(t)).N}) \\
&= M_N(\log(m(t))) = e^{\lambda.(e^{\exp(\log(m(t))}-1)} = e^{\lambda.(m(t)-1)}.
\end{aligned}$$

(b) By Problem 1.14, $S|N \sim Gamma(N,1)$ and $P(S>s|N)=\sum_{j=0}^{N-1}\frac{e^{-s}.s^j}{j!}$. Then:

$$\begin{aligned}P(S>5) &= E(1_{S>5}) = E(E(1_{S>5}|N)) = E(P(S>5|N))\\ &= E(\sum_{j=0}^{N-1}\frac{e^{-5}.5^j}{j!}) = E(\sum_{j=0}^{N-1}P(N=j))\\ &= E(P(N<N)) = E(0) = 0.\end{aligned}$$

□

Problem 2.9. Consider the experiment in Problem 1.3. where a coin is tossed an infinite number of times where the probability of heads on the $k-$th toss is $(1/2)^k$. Define X to be the number of heads observed in the experiment.
(a) Show that the probability generating function of X is

$$p(t)=\prod_{k=1}^{\infty}(1-\frac{1-t}{2^k}).$$

(b) Use the result of part (a) to evaluate $P(X=x)$ for $x=0,\cdots,5$.

Solution. (a) Let $X=\sum_{k=1}^{\infty}1_{A_k}$ in which $1_{A_k}\sim Bernouli(1/2^k)$, $(k\geq 1)$ and $M_{1_{A_k}}(t)=(1-\frac{1}{2^k})+(\frac{1}{2^k})e^t$, $(k\geq 1)$. Then:

$$p_X(t)=M_X(\log(t))=\prod_{k=1}^{\infty}M_{1_{A_k}}(\log(t))=\prod_{k=1}^{\infty}((1-\frac{1}{2^k})+(\frac{1}{2^k})e^{\log(t)})=\prod_{k=1}^{\infty}(1-\frac{1-t}{2^k}).$$

(b) Using Problem 1.18(c):

$$P_X(X=x)=\frac{1}{x!}\frac{d^x}{dt^x}p(t)|_{t=0}\quad x=0,1,2,3,4,5.$$

Next, define $u(t)=\sum_{k=1}^{\infty}\log(1-\frac{1-t}{2^k})$, then $p(t)=e^{u(t)}$. Consequently:

$$\begin{aligned}P(X=0) &= \frac{1}{0!}(e^{u(0)}),\\ P(X=1) &= \frac{1}{1!}(u^{(1)}(0).e^{u(0)}),\\ P(X=2) &= \frac{1}{2!}((u^{(2)}(0)+{u^{(1)}}^2(0)).e^{u(0)}),\\ P(X=3) &= \frac{1}{3!}((u^{(3)}(0)+3u^{(1)}(0)u^{(2)}(0)+{u^{(1)}}^2(0)).e^{u(0)}),\\ P(X=4) &= \frac{1}{4!}((u^4(0)+3{u^{(2)}}^2(0)+4u^{(1)}(0)u^{(3)}(0)+6{u^{(1)}}^2(0)u^{(2)}(0)+{u^{(1)}}^4(0)).e^{u(0)}),\\ P(X=5) &= \frac{1}{5!}((u^{(5)}(0)+10.u^{(2)}(0).u^{(3)}(0)+5.u^{(1)}(0)u^{(4)}(0)+15.u^{(1)}(0){u^{(2)}}^2(0)\\ &\quad +10.{u^{(1)}}^2(0)u^{(3)}(0)+10.{u^{(1)}}^3(0).u^{(2)}(0)+{u^{(1)}}^5(0)).e^{u(0)}),\end{aligned}$$

where in which $u(0)=\sum_{k=1}^{\infty}\log(\frac{2^k-1}{2^k})$, and $u^{(x)}(0)=\sum_{k=1}^{\infty}\frac{(-1)^{x+1}.(x-1)!}{(2^k-1)^x}$ for $x=1,2,3,4,5$.
□

Problem 2.11. Suppose we want to generate random variables with a Cauchy distribution. As an alternative to the method described in Problem 1.24, we can generate independent random variables V and W where $P(V=1)=P(V=-1)=1/2$ and W has density

$$g(x)=\frac{2}{\pi(1+x^2)} \quad \text{for} \quad |x|\leq 1.$$

(W can be generated by using the rejection method in Problem 2.10) Then we define $X=W^V$; show that X has Cauchy distribution.

Solution. For Z=1/W, an application of Theorem 2.3 with $h^{-1}(Z)=1/Z$ implies $f_Z(x)=\frac{2.1_{|x|>1}}{\pi(1+x^2)}$. Consequently:

$$\begin{aligned}
f_X(x) &= P(W^V=x)=P(W^V=x|V=1)P(V=1)+P(W^V=x|V=-1)P(V=-1)\\
&= P(W=x)\frac{1}{2}+P(W^{-1}=x)\frac{1}{2}=\frac{f_W(x)+f_Z(x)}{2}\\
&= \frac{(\frac{2}{\pi(1+x^2)}).(1_{|x|\leq 1}+1_{|x|>1})}{2}=\frac{1}{\pi(1+x^2)}.
\end{aligned}$$

□

Problem 2.13. Suppose that $X_1,\cdots,X_n$ are i.i.d. Uniform random variables on $[0,1]$. Define $S_n=(X_1+\cdots+X_n) \mod 1$; S_n is simply the "decimal" part of $X_1+\cdots+X_n$.
(a) Show that $S_n=(S_{n-1}+X_n) \mod 1$ for all $n\geq 2$.
(b) Show that $S_n\sim Unif(0,1)$ for all $n\geq 1$.

Solution.(a) By definition $S_n=\{\sum_{i=1}^n X_i\}=\sum_{i=1}^n X_i-[\sum_{i=1}^n X_i]$. Hence:

$$\begin{aligned}
\{S_{n-1}+X_n\} &= (S_{n-1}+X_n)-[S_{n-1}+X_n]\\
&= (\sum_{i=1}^{n-1}X_i-[\sum_{i=1}^{n-1}X_i]+X_n)-[\sum_{i=1}^{n-1}X_i-[\sum_{i=1}^{n-1}X_i]+X_n]\\
&= \sum_{i=1}^{n}X_i-[\sum_{i=1}^{n-1}X_i]-[\sum_{i=1}^{n}X_i-[\sum_{i=1}^{n-1}X_i]]\\
&= \sum_{i=1}^{n}X_i-[\sum_{i=1}^{n-1}X_i]-[\sum_{i=1}^{n}X_i]+[[\sum_{i=1}^{n-1}X_i]]\\
&= \sum_{i=1}^{n}X_i-[\sum_{i=1}^{n}X_i]=S_n.
\end{aligned}$$

(b) We prove the assertion by induction on n. As for $n=1$, we have $S_1=X_1$ it trivially holds. Let it hold for case $n>1$(induction hypothesis). Then, by Part (a), $S_{n+1}=S_n+X_n-[S_n+X_n]$ and consequently:

$$
\begin{aligned}
F_{S_{n+1}}(t) &= P(S_{n+1} \le t) = P(S_n + X_{n+1} \le [S_n + X_{n+1}] + t) \\
&= \sum_{k=0}^{n+1} P(S_n + X_{n+1} \le [S_n + X_{n+1}] + t, [S_n + X_{n+1}] = k) \\
&= \sum_{k=0}^{n+1} P(S_n + X_{n+1} \le k + t, k \le S_n + X_{n+1} < k+1) \\
&= \sum_{k=0}^{n+1} P(k \le S_n + X_{n+1} \le k + t) \\
&= P(0 \le S_n + X_{n+1} \le t) + P(1 \le S_n + X_{n+1} \le 1 + t) \\
&= \frac{t^2}{2} + (t - \frac{t^2}{2}) = t. \ \text{ if } \ 0 \le t \le 1
\end{aligned}
$$

Also, by the first line above and the fact that $[S_n + X_{n+1}] \le S_n + X_{n+1} < [S_n + X_{n+1}] + 1$, it is clear that $F_{S_{n+1}}(t) = 0$ if $t < 0$ and $F_{S_{n+1}}(t) = 1$ if $t > 1$. Accordingly, $S_{n+1} \sim Unif(0,1)$.
□

Problem 2.15. Suppose $X_1, \cdots, X_n$ are independent nonnegative continuous random variables where X_i has hazard function $\lambda_i(x)$ $(i = 1, \cdots, n)$.
(a) If $U = \min_{1\le i\le n}(X_i)$, show that the hazard function of U is $\lambda_U(x) = \lambda_1(x) + \cdots + \lambda_n(x)$.
(b) If $V = \max_{1\le i\le n}(X_i)$, show that the hazard function of V satisfies $\lambda_V(x) \le \min(\lambda_1(x), \cdots, \lambda_n(x))$.
(c) Show that the result of (b) holds even if the X_is are not independent.

Solution.(a)Since:

$$
\begin{aligned}
F_U(x) &= 1 - S_U(x) = 1 - P(U \ge x) = 1 - P(\min_{1\le i\le n}(X_i) \ge x) \\
&= 1 - \prod_{i=1}^{n} P(X_i \ge x) = 1 - \prod_{i=1}^{n}(1 - F_{X_i}(x)),
\end{aligned}
$$

we have:

$$
f_U(x) = \frac{d}{dx}F_U(x) = -\sum_{j=1}^{n}(\prod_{j\ne i}(1 - F_{X_j}(x))(-f_{X_j}(x))) = \sum_{j=1}^{n}(\prod_{j\ne i}(1 - F_{X_j}(x))(f_{X_j}(x))).
$$

Hence:

$$
\lambda_U(x) = \frac{f_U(x)}{S_U(x)} = \frac{\sum_{j=1}^{n}(\prod_{j\ne i}(1 - F_{X_j}(x))(f_{X_j}(x)))}{\prod_{j=1}^{n}(1 - F_{X_j}(x))} = \sum_{j=1}^{n}\frac{f_{X_j}(x)}{(1 - F_{X_j}(x))} = \sum_{j=1}^{n}\lambda_j(x).
$$

(b),(c) We prove the assertion for cumulative hazard function $\Lambda(x) = \int_0^x \lambda(t)dt$. By,

$$
\exp(-\Lambda_V(x)) = S_V(x) = P(X \ge x) \ge P(X_i \ge x) = S_{X_i}(x) = \exp(-\Lambda_i(x)),
$$

and taking log it follows that $\Lambda_V(x) \le \Lambda_i(x)$ $(1 \le i \le n)$, and hence:

$$
\Lambda_V(x) \le \min_{1\le i\le n} \Lambda_i(x).
$$

As a counterexample for the hazard function case, let $n = 2$ and $X_i \sim exp(\lambda_i)$ $(i = 1, 2)$ with $\lambda_1 < \lambda_2$ be independent and $V = \max(X_1, X_2)$. Then:

$$\begin{aligned} F_V(x) &= 1 - \exp(-\lambda_1.x) - \exp(-\lambda_2.x) + \exp(-(\lambda_1 + \lambda_2).x), \\ S_V(x) &= \exp(-\lambda_1.x) + \exp(-\lambda_2.x) - \exp(-(\lambda_1 + \lambda_2).x), \\ f_V(x) &= \lambda_1. \exp(-\lambda_1.x) + \lambda_2. \exp(-\lambda_2.x) - (\lambda_1 + \lambda_2). \exp(-(\lambda_1 + \lambda_2).x), \end{aligned}$$

implying:

$$\lambda_V(x) = \frac{f_V(x)}{S_V(x)} = \frac{\lambda_1. \exp(\lambda_2.x) + \lambda_2. \exp(\lambda_1.x) - (\lambda_1 + \lambda_2)}{\exp(\lambda_2.x) + \exp(\lambda_1.x) - 1} \leq \lambda_1 = \min(\lambda_1(x), \lambda_2(x)),$$

and hence $x \leq \frac{1}{\lambda_1} \log(\frac{\lambda_2}{\lambda_2 - \lambda_1})$, a contradiction to unboundedness of range of x.
□

Problem 2.17. Suppose that X and Y are random variables such that both $E(X^2)$ and $E(Y^2)$ are finite. Define $g(t) = E((Y + t.X)^2)$.
(a) Show that $g(t)$ is minimized at $t = -\frac{E(XY)}{E(X^2)}$.
(b) Show that $(E(XY))^2 \leq E(X^2).E(Y^2)$; this is called the Cauchy-Schwarz inequality.
(c) Use part (b) to show that $|Corr(X, Y)| \leq 1$.

Solution. (a) Since:

$$g(t) = E((Y + t.X)^2) = E(Y^2 + 2tX.Y + t^2.X^2) = E(X^2)t^2 + 2E(XY)t + E(Y^2),$$

it follows that $g^{(1)}(t) = 2E(X^2)t + 2E(XY) = 0$, and hence $t = -\frac{E(XY)}{E(X^2)}$ minimizes g.

(b)Since $(Y + t.X)^2 \geq 0$, we have $g(t) = E((Y + t.X)^2) \geq 0$, and consequently:

$$\begin{aligned} 0 &\leq g(\frac{-E(XY)}{E(X^2)}) = E(X^2).(\frac{-E(XY)}{E(X^2)})^2 + 2E(XY).(\frac{-E(XY)}{E(X^2)}) + E(Y^2) \\ &= \frac{E^2(XY)}{E(X^2)} - \frac{2E^2(XY)}{E(X^2)} + E(Y^2) = \frac{-E^2(XY) + E(X^2)E(Y^2)}{E(X^2)}, \end{aligned}$$

implying: $E^2(XY) \leq E(X^2).E(Y^2)$.

(c) First, assume $E(X) = E(Y) = 0$, then, by Cauchy-Schwarz inequality in Part (b):

$$|Corr(X, Y)| = |\frac{Cov(X, Y)}{\sqrt{Var(X).Var(Y)}}| = |\frac{E(XY)}{\sqrt{E(X^2)E(Y^2)}}| \leq 1.$$

Second, for the case of $E(X) \neq 0$ or $E(Y) \neq 0$, define $X^* = X - E(X)$ and $Y^* = Y - E(Y)$. Then, $Cov(X, Y) = Cov(X^*, Y^*)$, $Var(X) = Var(X^*)$ and $Var(Y) = Var(Y^*)$. Consequently:

$$|Corr(X, Y)| = |\frac{Cov(X, Y)}{\sqrt{Var(X).Var(Y)}}| = |\frac{Cov(X^*, Y^*)}{\sqrt{Var(X^*).Var(Y^*)}}| = |Corr(X^*, Y^*)| \leq 1.$$

□
Problem 2.19. Suppose that X and Y are independent random variables with X discrete and Y

continuous. Define $Z = X + Y$.
(a) Show that Z is a continuous random variable with

$$P(Z \leq z) = \sum_x P(Y \leq z - x)P(X = x).$$

(b) If Y has a density function $f_Y(y)$, show that the density of Z is

$$f_Z(z) = \sum_x f_Y(z - x) f_X(x)$$

where $f_X(x)$ is the frequency function of X.

Solution.(a)

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) = \sum_x P(X + Y \leq z, X = x) \\ &= \sum_x P(X + Y \leq z | X = x)P(X = x) = \sum_x P(Y \leq z - x)P(X = x) \\ &= \sum_x F_Y(z - x)P(X = x). \end{aligned}$$

As Y is continuous random variable, $G_x(z) = F_Y(z - x)$ is continuous CDF and so is $H_x(z) = F_Y(z - x).P(X = x)$. Hence, F_Z as the sum of continuous functions H_x is continuous, as well.

(b) By Part (a):

$$\begin{aligned} F_Z(z) &= \sum_x F_Y(z - x).f_X(x) = \sum_x (\int_{-\infty}^{z-x} f_Y(y)dy) f_X(x) = \sum_x \int_{-\infty}^{z-x} (f_Y(y) f_X(x))dy \\ &= \sum_x \int_{-\infty}^{z} (f_Y(y^* - x) f_X(x))dy^* = \int_{-\infty}^{z} (\sum_x (f_Y(y^* - x) f_X(x)))dy^*, \end{aligned}$$

accordingly:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \sum_x (f_Y(z - x) f_X(x)).$$

□

Problem 2.21.(a) Show that

$$Cov(X, Y) = E(Cov(X, Y|Z)) + Cov(E(X|Z), E(Y|Z)).$$

(b) Suppose that $X_1, X_2, \cdots$ be i.i.d. Exponential random variables with parameter 1 and take N_1, N_2 to be independent Poisson random variables with parameters λ_1, λ_2 that are independent of the $X_i's$. Define compound Poisson random variables

$$S_1 = \sum_{i=1}^{N_1} X_i \qquad S_2 = \sum_{i=1}^{N_2} X_i$$

and evaluate $Cov(S_1, S_2)$ and $Corr(S_1, S_2)$. When is this correlation maximized ?

Solution. (a)

$$\begin{aligned}
E(Cov(X,Y|Z))+Cov(E(X|Z),E(Y|Z)) &= E(E(XY|Z)-E(X|Z)E(Y|Z)) \\
&+ E(E(X|Z)E(Y|Z))-E(E(X|Z)).E(E(Y|Z)) \\
&= E(E(XY|Z))-E(E(X|Z)E(Y|Z)) \\
&+ E(E(X|Z)E(Y|Z))-E(E(X|Z)).E(E(Y|Z)) \\
&= E(XY)-E(X).E(Y)=Cov(X,Y).
\end{aligned}$$

(b) We have:

$$Cov(S_1,S_2)=E(S_1.S_2)-E(S_1).E(S_2). \quad (*)$$

Using Theorem 2.8 the first term in the right hand side of (*) can be evaluated as follows:

$$\begin{aligned}
E(S_1.S_2) &= \sum_{n_1=1}^{\infty} E(S_1.S_2|N_1=n_1)P(N_1=n_1) \\
&= \sum_{n_1=1}^{\infty}(\sum_{n_2=1}^{\infty} E(S_1.S_2|N_1=n_1,N_2=n_2)P(N_2=n_2))P(N_1=n_1) \\
&= \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty} E((\sum_{i=1}^{n_1}X_i)(\sum_{j=1}^{n_2}X_j))P(N_2=n_2)P(N_1=n_1) \\
&= \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}(\sum_{i=1}^{n_1}\sum_{i=1}^{n_2}E(X_i.X_j))P(N_2=n_2)P(N_1=n_1) \\
&= \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}(\sum_{i=1}^{n_1}\sum_{i=1}^{n_2}(1+Cov(X_i,X_j)))P(N_2=n_2)P(N_1=n_1) \\
&= \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}(n_1.n_2+\sum_{i=1}^{n_1}\sum_{i=1}^{n_2}Cov(X_i,X_j))P(N_2=n_2)P(N_1=n_1) \\
&= \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}(n_1.n_2+\min(n_1,n_2))P(N_2=n_2)P(N_1=n_1) \\
&= \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}(n_1.n_2)P(N_2=n_2)P(N_1=n_1)+\sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}(\min(n_1,n_2))P(N_2=n_2)P(N_1=n_1) \\
&= E(N_1.N_2)+E(\min(N_1,N_2)) \\
&= \lambda_1.\lambda_2+E(\min(N_1,N_2)|N_1\le N_2).P(N_1\le N_2)+E(\min(N_1,N_2)|N_1\ge N_2).P(N_1\ge N_2) \\
&= \lambda_1.\lambda_2+E(N_1).P+E(N_2).(1-P) \quad (\text{define} P=P(N_1\le N_2)) \\
&= \lambda_1.\lambda_2+\lambda_1.P+\lambda_2.(1-P). \quad (**)
\end{aligned}$$

By Example 2.14 for $\mu=\sigma^2=1$ we have:

$$\begin{aligned}
E(S_1) &= \lambda_1 \quad Var(S_1)=2.\lambda_1 \\
E(S_2) &= \lambda_2 \quad Var(S_2)=2.\lambda_2. \quad (***)
\end{aligned}$$

Accordingly, by (*), (**) and (***) it follows that:

$$Cov(S_1,S_2)=\lambda_1.P+\lambda_2.(1-P):P=P(N_1\le N_2)=\sum_{n=0}^{\infty}F_{N_1}(n).f_{N_2}(n). \quad (\dagger)$$

Next, using (†) it follows that:

$$Corr(S_1, S_2) = \frac{Cov(S_1, S_2)}{\sqrt{Var(S_1).Var(S_2)}} = \frac{\lambda_1.P + \lambda_2.(1-P)}{2\sqrt{\lambda_1.\lambda_2}}. \quad (\dagger\dagger)$$

Finally, to find the maximum value of $Corr(S_1, S_2)$ using (††) we define:

$$H(\lambda_1, \lambda_2) = \frac{\lambda_1.P + \lambda_2.(1-P)}{2\sqrt{\lambda_1.\lambda_2}}. \quad (\dagger\dagger\dagger)$$

A simple calculus for the bivariate function H in († † †) shows that $\max(H) = \frac{1}{2}$, (Exercise!).
□

Problem 2.23. The mean residual life function $r(t)$ of a nonnegative random variable X is defined to be

$$r(t) = E(X - t | X \geq t).$$

($r(t)$ would be of interest, for example , to a life insurance company.)
(a) Suppose that F is the distribution function of X. Show that

$$r(t) = \frac{1}{1 - F(t)} \int_t^\infty (1 - F(x))dx.$$

(b) Show that $r(t)$ is constant if, and only if, X has an Exponential distribution.
(c) Show that

$$E(X^2) = 2\int_0^\infty r(t)(1 - F(t))dt.$$

(d) Suppose that X has a density function $f(x)$ that is different and $f(x) > 0$ for $x > 0$. Show that

$$\lim_{t\to\infty} r(t) = \lim_{t\to\infty} (-\frac{f(t)}{f'(t)}).$$

(e) Suppose that X has a Gamma distribution:

$$f(x) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}\exp(-\lambda.x) \quad \text{for} x > 0.$$

Evaluate the limit in part (c) for this distribution. Give an interpretation of this result.

Solution. (a) By Fubini's Theorem:

$$\begin{aligned} r(t) &= E(X - t|X \geq t) = \int_t^\infty (x - t)\frac{dF(x)}{P(X > t)} = \frac{\int_t^\infty \int_t^x dy dF(x)}{1 - F(t)} \\ &= \frac{\int_t^\infty (\int_y^\infty dF(x))dy)}{1 - F(t)} = \frac{\int_t^\infty (1 - F(y))dy}{1 - F(t)}. \end{aligned}$$

(b)

$$\begin{aligned} r(t) = c &\Leftrightarrow c(1 - F(t)) = \int_t^\infty (1 - F(x))dx \Leftrightarrow -c.f(t) = F(t) - 1 \\ &\Leftrightarrow \frac{f(t)}{1 - F(t)} = \frac{1}{c} \Leftrightarrow \lambda(x) = \frac{1}{c} \\ &\Leftrightarrow X \sim \exp(\frac{1}{c}). \end{aligned}$$

(c)

$$\begin{aligned}
E(X^2) &= \int_0^\infty x^2.f(x)dx = \int_0^\infty \int_0^x 2tdtf(x)dx = \int_0^\infty \int_0^x 2tf(x)dtdx \\
&= \int_0^\infty \int_t^\infty 2t.f(x)dxdt = \int_0^\infty 2t(\int_t^\infty f(x)dx)dt = \int_0^\infty 2t(1-F(t))dt \\
&= \int_0^\infty (2\int_0^t ds)(1-F(t))dt = 2\int_0^\infty \int_0^t (1-F(t))dsdt = 2\int_0^\infty \int_s^\infty (1-F(t))dtds \\
&= 2\int_0^\infty r(s)(1-F(s))ds = 2\int_0^\infty r(t)(1-F(t))dt.
\end{aligned}$$

(d) Using L'Hospital's Rule it follows:

$$\begin{aligned}
\lim_{t\to\infty} r(t) &=^{\text{definition}} \lim_{t\to\infty} \frac{\int_t^\infty (1-F(x))dx}{1-F(t)} =^{\text{LHR}} \lim_{t\to\infty} \frac{d/dt \int_t^\infty (1-F(x))dx}{d/dt(1-F(t))} \\
&= \lim_{t\to\infty} \frac{F(t)-1}{-f(t)} =^{\text{LHR}} \lim_{t\to\infty} \frac{d/dt(F(t)-1)}{d/dt(-f(t))} = \lim_{t\to\infty} (-\frac{f(t)}{f'(t)}).
\end{aligned}$$

(e) By Part (d) we have:

$$\begin{aligned}
\lim_{t\to\infty} r(t) &= \lim_{t\to\infty} (-\frac{f(t)}{f'(t)}) = \lim_{t\to\infty} \frac{-\frac{1}{\Gamma(\alpha)}\lambda^\alpha t^{\alpha-1}\exp(-\lambda.t)}{\frac{1}{\Gamma(\alpha)}\lambda^\alpha t^{\alpha-2}\exp(-\lambda.t)(\alpha-1-\lambda.t)} \\
&= \lim_{t\to\infty} -(\frac{t}{\alpha-1-\lambda.t}) = \frac{1}{\lambda} = \frac{1}{\alpha}.E(X).
\end{aligned}$$

The mean residual life function of Gamma distribution is asymptotically proportional to its mean.
□

Problem 2.25. Suppose that $X_1, \cdots, X_n$ are i.i.d. continuous random variables with distribution function $F(x)$ and density function $f(x)$; let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the order statistics.
(a) Show that the distribution function of $X_{(k)}$ is

$$G_k(x) = \sum_{j=k}^n C(n,j)F(x)^j(1-F(x))^{n-j}.$$

(b) Show that the density function of $X_{(k)}$ is

$$g_k(x) = \frac{n!}{(n-k)!(k-1)!}F(x)^{k-1}(1-F(x))^{n-k}f(x).$$

Solution.(a) For $S = \sum_{k=1}^n I_{(X_k \le x)} \sim Binomial(n, F(x))$ we have:

$$G_k(x) = P(X_{(k)} \le x) = P(S \ge k) = \sum_{j=k}^n P(S=j) = \sum_{j=k}^n C(n,j)F(x)^j(1-F(x))^{n-j}.$$

(b) As $C(n,j+1).(j+1) = C(n,j).(n-j)$, from Part (a) it follows:

$$\begin{aligned}
g_k(x) &= \frac{d}{dx}G_k(x) \\
&= \sum_{j=k}^{n} C(n,j)(jF(x)^{j-1}(1-F(x))^{n-j}.f(x) - (n-j)F(x)^j(1-F(x))^{n-j-1}f(x)) \\
&= C(n,k).(kF(x)^{k-1}(1-F(x))^{n-k}.f(x)) \\
&+ \sum_{j=k+1}^{n} C(n,j)(jF(x)^{j-1}(1-F(x))^{n-j}f(x)) \\
&- \sum_{j=k}^{n-1} C(n,j)(n-j)(F(x)^j(1-F(x))^{n-j-1}f(x)) \\
&= \frac{n!}{(n-k)!(k-1)!}F(x)^{k-1}(1-F(x))^{n-k}f(x) \\
&+ \sum_{j=k}^{n-1} C(n,j+1)(j+1)(F(x)^j(1-F(x))^{n-j-1}f(x)) \\
&- \sum_{j=k}^{n-1} C(n,j)(n-j)(F(x)^j(1-F(x))^{n-j-1}f(x)) \\
&= \frac{n!}{(n-k)!(k-1)!}F(x)^{k-1}(1-F(x))^{n-k}f(x).
\end{aligned}$$

□

Problem 2.27. Suppose that $X_1, \cdots, X_{n+1}$ be i.i.d. Exponential random variables with parameter $\lambda$and define

$$U_k = \frac{1}{T}\sum_{i=1}^{k} X_i \quad \text{for} k = 1, \cdots, n$$

where $T = X_1 + \cdots + X_{n+1}$.
(a) Find the joint density of $(U_1, \cdots, U_n, T)$. (Note that $0 < U_1 < U_2 < \cdots < U_n < 1$.)
(b) Show that the joint distribution of $(U_1, \cdots, U_n)$ is exactly the same as the joint distribution of the order statistics of an i.i.d. sample of n observations from a Uniform distribution on $[0,1]$.

Solution. (a) Since $U_k.T = \sum_{i=1}^{k} X_i$ we have:

$$X_k = \sum_{i=1}^{k} X_i - \sum_{i=1}^{k-1} X_i = U_k.T - U_{k-1}.T = (U_k - U_{k-1}).T. \quad (1 \le k \le n)$$

Defining $U_0 = 0$ and $U_{n+1} = 1$ there will be an extension of above equality to :

$$X_k = (U_k - U_{k-1}).T. \quad (1 \le k \le n+1)$$

Now, define transformation h via:

$$(U_1, \cdots, U_n, T) = h(X_1, \cdots, X_n, X_{n+1}).$$

Then, by Theorem 2.3. we have:

$$\begin{aligned}
f_{(U_1,\cdots,U_n,T)}(u_1,\cdots,u_n,t) &= f_{(X_1,\cdots,X_n,X_{n+1})}(h^{-1}(u_1,\cdots,u_n,t))|J(h^{-1}(u_1,\cdots,u_n,t))| \\
&= f_{(X_1,\cdots,X_n,X_{n+1})}((u_1-u_0)t,(u_2-u_1)t,\cdots,(u_n-u_{n-1})t,(u_{n+1}-u_n)t) \\
&\quad |\frac{d(X_1,\cdots,X_n,X_{n+1})}{d(u_1,\cdots,u_n,t)}| \\
&= \prod_{i=1}^{n+1} f_{X_i}((u_i-u_{i-1})t).det(\begin{pmatrix} +t & 0 & \cdots & 0 & 0 & u_1-u_0 \\ -t & +t & \cdots & 0 & 0 & u_2-u_1 \\ \cdots & \cdots & & & & \\ 0 & 0 & \cdots & -t & +t & u_n-u_{n-1} \\ 0 & 0 & \cdots & 0 & -t & 1-u_n \end{pmatrix}) \\
&= (\prod_{i=1}^{n+1} \lambda.e^{-\lambda(u_i-u_{i-1})t}).t^n.1_{0<u_1<\cdots<u_n<1}(u_1,\cdots,u_n) \\
&= \lambda^{n+1}.e^{-\lambda.t}t^n.1_{0<u_1<\cdots<u_n<1}(u_1,\cdots,u_n).
\end{aligned}$$

(b) By Part (a):

$$\begin{aligned}
f_{(U_1,\cdots,U_n)}(u_1,\cdots,u_n) &= \int_0^\infty f_{(U_1,\cdots,U_n,T)}(u_1,\cdots,u_n,t)dt \\
&= 1_{0<u_1<\cdots<u_n<1}(u_1,\cdots,u_n)\int_0^\infty \lambda^{n+1}.e^{-\lambda.t}t^n dt \\
&= \Gamma(n+1)\ 1_{0<u_1<\cdots<u_n<1}(u_1,\cdots,u_n) \\
&= n!\ 1_{0<u_1<\cdots<u_n<1}(u_1,\cdots,u_n).
\end{aligned}$$

□

Problem 2.29. Suppose that X and Y are independent Exponential random variables with parameters λ and μ respectively. Define random variables

$$T = \min(X,Y) \qquad \Delta = 1 \text{ if } X < Y, 0 \text{ otherwise.}$$

Note that T has a continuous distribution while Δ is discrete. (This is an example of type I censoring in reliability or survival analysis.)
(a) Find the density of T and the frequency function of Δ.
(b) Find the joint distribution function of (T,Δ).

Solution.(a) First,

$$\begin{aligned}
f_T(t) &= \frac{d}{dt}F_T(t) = \frac{d}{dt}(1-S_T(t)) = -\frac{d}{dt}S_T(t) = -\frac{d}{dt}P(T\geq t) \\
&= -\frac{d}{dt}(P(X\geq t)P(Y\geq t)) = -\frac{d}{dt}(S_X(t).S_Y(t)) = -\frac{d}{dt}(e^{-(\lambda+\mu)t}) \\
&= (\lambda+\mu)(e^{-(\lambda+\mu)t}),
\end{aligned}$$

and hence, $T \sim \exp(\lambda+\mu)$. Second,

$$\begin{aligned} P(\Delta=1) &= P(X<Y)=\int\int_{X<Y}\lambda e^{-\lambda.x}.\mu.e^{-\mu.y}dxdy \\ &= \int_0^\infty\int_x^\infty \lambda e^{-\lambda.x}.\mu.e^{-\mu.y}dydx=\int_0^\infty \lambda e^{-\lambda.x}(\int_x^\infty \mu.e^{-\mu.y}dy)dx \\ &= \int_0^\infty \lambda e^{-\lambda.x}e^{-\mu.x}dx=(\frac{\lambda}{\lambda+\mu})\int_0^\infty(\lambda+\mu)e^{-(\lambda+\mu).x}dx \\ &= \frac{\lambda}{\lambda+\mu}, \end{aligned}$$

implying $\Delta \sim Bernouli(\frac{\lambda}{\lambda+\mu})$.

(b) Since:

$$\begin{aligned} f_{T,\Delta}(t,1) &= P(\min(X,Y)=t, X<Y)=P(X=t, t<Y) \\ &= P_X(X=t).P_Y(t<Y)=\lambda.e^{-\lambda.t}.e^{-\mu.t} \\ &= \lambda.e^{-(\lambda+\mu)t}, \end{aligned}$$

and

$$\begin{aligned} f_{T,\Delta}(t,0) &= P(\min(X,Y)=t, X\geq Y)=P(Y=t, X\geq t) \\ &= P_Y(Y=t).P_X(X\geq t)=\mu.e^{-\mu.t}.e^{-\lambda.t} \\ &= \mu.e^{-(\lambda+\mu)t}, \end{aligned}$$

it follows that:

$$f_{T,\Delta}(t,\delta)=(\delta.\lambda+(1-\delta).\mu).e^{-(\lambda+\mu)t} \qquad \delta=0,1, \qquad 0<t.$$

□

Problem 2.31. Suppose that X has a Beta distribution with parameters α and β.
(a) Find the density function of $Y=X(1-X)^{-1}$.
(b) Suppose that $\alpha=m/2$ and $\beta=n/2$ and define Y as in part (a). Using the definition of F distribution, show that $nY/m \sim F(m,n)$.

Solution. (a) Define $Y=h(X)=\frac{X}{1-X}$, then $X=h^{-1}(Y)=\frac{Y}{1+Y}$. By Theorem 2.3:

$$\begin{aligned} f_Y(y) &= f_X(h^{-1}(y))|\frac{dh^{-1}(y)}{dy}|=f_X(\frac{y}{1+y})(\frac{1}{(1+y)^2}) \\ &= \frac{1}{B(\alpha,\beta)}(\frac{y}{y+1})^{\alpha-1}(\frac{1}{y+1})^{\beta-1}(\frac{1}{(1+y)^2}) \\ &= \frac{1}{B(\alpha,\beta)}y^{\alpha-1}(y+1)^{-(\alpha+\beta)}. \quad (0<y<\infty) \end{aligned}$$

(b) Let $U =^d \chi^2(m)$ and $V =^d \chi^2(n)$ be independent . Then, $\frac{U}{U+V} =^d B(\frac{m}{2},\frac{n}{2})$ and $F=\frac{U/m}{V/n} =^d F(m,n)$. Consequently:

$$\frac{nY}{m}=\frac{n}{m}\frac{X}{1-X}=\frac{X/m}{(1-X)/n} =^d \frac{\frac{U/(U+V)}{m}}{\frac{V/(U+V)}{n}}=\frac{U/m}{V/n} =^d F(m,n).$$

□

Problem 2.33. Suppose that $X \sim \chi^2(n)$.
(a) Show that $E(X^r) = 2^r\Gamma(r+n/2)/\Gamma(n/2)$ if $r > -n/2$.
(b) Using part (a), show that $E(X) = n$ and $Var(X) = 2n$.

Solution. (a)

$$\begin{aligned} E(X^r) &= \int_0^\infty x^r f_X(x)dx = \int_0^\infty x^r(\frac{1}{\Gamma(n/2)2^{n/2}}x^{n/2-1}e^{-x/2})dx \\ &= \frac{\Gamma(n/2+r)2^{n/2+r}}{\Gamma(n/2)2^{n/2}}\int_0^\infty \frac{x^{n/2+r-1}e^{-x/2}}{\Gamma(n/2+r)2^{n/2+r}}dx = \frac{\Gamma(n/2+r)}{\Gamma(n/2)}.2^r \quad \text{if} \quad r+n/2>0. \end{aligned}$$

(b)

$$\begin{aligned} E(X) &= \frac{\Gamma(n/2+1)}{\Gamma(n/2)}.2 = \frac{(n/2)\Gamma(n/2)}{\Gamma(n/2)}.2 = n. \\ E(X^2) &= \frac{\Gamma(n/2+2)}{\Gamma(n/2)}.2^2 = \frac{(n/2+1)(n/2)\Gamma(n/2)}{\Gamma(n/2)}.2^2 = (n+2)n. \\ Var(X) &= E(X^2) - E^2(X) = 2n. \end{aligned}$$

□

Problem 2.35. Suppose that $W \sim F(m,n)$. Show that

$$E(W^r) = (\frac{n}{m})^r\frac{\Gamma(r+m/2)\Gamma(-r+n/2)}{\Gamma(m/2)\Gamma(n/2)}$$

if $-m/2 < r < n/2$.

Solution. As for two independent U, V with $U =^d \chi^2(m)$ and $V =^d \chi^2(n)$ we have $W =^d \frac{U/m}{V/n} =^d F(m,n)$, two applications of Problem 2.33(a) imply:

$$\begin{aligned} E(W^r) &= E((\frac{U/m}{V/n})^r) = (\frac{n}{m})^r E(U^r V^{-r}) = (\frac{n}{m})^r.E(U^r).E(V^{-r}) \\ &= (\frac{n}{m})^r.(2^r\frac{\Gamma(r+m/2)}{\Gamma(m/2)}).(2^{-r}\frac{\Gamma(-r+n/2)}{\Gamma(n/2)}) \\ &= (\frac{n}{m})^r\frac{\Gamma(r+m/2)\Gamma(-r+n/2)}{\Gamma(m/2)\Gamma(n/2)} \quad \text{if} \quad -m/2<r<n/2. \end{aligned}$$

□

Problem 2.37. Suppose that $X \sim N_n(\mu, I)$; the elements of X are independent Normal random variables with variances equal to 1.
(a) Suppose that O is an orthogonal matrix whose first row is $\mu^T/\|\mu\|$ and let $Y = OX$. Show that $E(Y_1) = \|\mu\|$ and $E(Y_k) = 0$ for $k \geq 2$.
(b) Using part (a), show that the distribution of $\|X\|^2$ is the same as that of $\|Y\|^2$ and hence depends on μ only through its norm $\|\mu\|$.
(c) Let $\theta^2 = \|\mu\|^2$. Show that the density of $V = \|X\|^2$ is

$$f_V(x) = \sum_{k=0}^{\infty}\frac{\exp(-\theta^2/2)(\theta^2/2)^k}{k!}f_{2k+n}(x)$$

where $f_{2k+n}(x)$ is the density function of a χ^2 random variable with $2k+n$ degrees of freedom. (V has a non-central χ^2 distribution with n degrees of freedom and non-centrality parameter θ^2.)

Solution.(a) Let $O = \begin{pmatrix} a_{11}^T \cdots a_{1n}^T \\ \cdots \\ a_{n1}^T \cdots a_{nn}^T \end{pmatrix}$ and $X = \begin{pmatrix} X_1 \\ \cdots \\ X_n \end{pmatrix}$ Then:

$$Y_1 = (a_{11}^T, \cdots, a_{1n}^T) \begin{pmatrix} X_1 \\ \cdots \\ X_n \end{pmatrix} = \sum_{i=1}^{n} a_{1i}^T.X_i = \sum_{i=1}^{n} \frac{\mu_i^T}{\|\mu\|} X_i,$$

and hence:

$$E(Y_1) = \sum_{i=1}^{n} \frac{\mu_i^T}{\|\mu\|} E(X_i) = \sum_{i=1}^{n} \frac{\mu_i^T.\mu_i}{\|\mu\|} = \frac{(\|\mu\|)^2}{\|\mu\|} = \|\mu\|.$$

Next, as $(a_{k1}^T, \cdots, a_{kn}^T) = Y_k \perp Y_1 = (\mu_1, \cdots, \mu_n)$ $(k \geq 2)$ it follows that:

$$E(Y_k) = E(\sum_{i=1}^{n} a_{ki}^T.X_i) = \sum_{i=1}^{n} a_{ki}^T.E(X_i) = \sum_{i=1}^{n} a_{ki}^T.\mu_i = 0.$$

(b)

$$\|Y\|^2 = Y^T.Y = (OX)^T.(OX) = X^T.(O^T.O).X = X^T.X = \|X\|^2.$$

(c) First, $V = \|X\|^2 = \|Y\|^2 = \sum_{i=1}^n Y_i^2 = Y_1^2 + \sum_{i=2}^n Y_i^2 = U + W$ such that $U = Y_1^2 (Y_1 =^d N(0,1))$ and $W = \sum_{i=2}^n Y_i^2 =^d \chi^2(n-1)$. Furthermore:

$$f_U(t) = \frac{e^{\theta^2/2}}{2\sqrt{2\pi}\sqrt{t}} (e^{\theta\sqrt{t}} + e^{-\theta\sqrt{t}}).e^{-t/2} \quad (t > 0), \quad (*)$$

and

$$f_W(t) = \frac{t^{(n-1)/2-1}.e^{-t/2}}{2^{(n-1)/2}.\Gamma((n-1)/2)} \quad (t > 0). \quad (**)$$

Second, using (*) and (**) it follows that:

$$\begin{aligned}
f_V(x) &= \int_0^x f_U(x).f_W(x-t)dt \\
&= \int_0^x (\frac{e^{\theta^2/2}}{2\sqrt{2\pi}\sqrt{t}}(e^{\theta\sqrt{t}}+e^{-\theta\sqrt{t}}).e^{-t/2}) * (\frac{(x-t)^{(n-1)/2-1}.e^{-(x-t)/2}}{2^{(n-1)/2}.\Gamma((n-1)/2)})dt \\
&= e^{-\theta^2/2}.(\int_0^x (\frac{e^{\theta\sqrt{t}}+e^{-\theta\sqrt{t}}}{2})(\frac{(x-t)^{(n-1)/2-1}}{\sqrt{2\pi}\sqrt{t}2^{(n-1)/2}\Gamma((n-1)/2)})).e^{-x/2} \\
&= e^{-\theta^2/2}.(\int_0^x (\sum_{k=0}^{\infty}\frac{(\theta.\sqrt{t})^{2k}}{(2k)!})(\frac{(x-t)^{(n-1)/2-1}}{\sqrt{2\pi}\sqrt{t}2^{(n-1)/2}\Gamma((n-1)/2)})).e^{-x/2} \\
&= e^{-\theta^2/2}.(\sum_{k=0}^{\infty}(\frac{(\theta^2/2)^k}{k!})(\int_0^x \frac{t^{(2k+1)/2-1}(x-t)^{(n-1)/2-1}}{(\frac{(2k)!}{k!})\sqrt{2\pi}2^{(n-1)/2-k}\Gamma((n-1)/2)})).e^{-x/2} \\
&= e^{-\theta^2/2}.(\sum_{k=0}^{\infty}(\frac{(\theta^2/2)^k}{k!})(\frac{x^{(2k+n)/2-1}.B((2k+1)/2),(n-1)/2}{(\frac{(2k)!}{k!})\sqrt{2\pi}2^{(n-1)/2-k}\Gamma((n-1)/2)})).e^{-x/2} \\
&= e^{-\theta^2/2}.(\sum_{k=0}^{\infty}(\frac{(\theta^2/2)^k}{k!})(\frac{x^{(2k+n)/2-1}}{2^{(2k+n)/2}\Gamma((2k+n)/2)})).e^{-x/2} \\
&= \sum_{k=0}^{\infty}(e^{-\theta^2/2}.\frac{(\theta^2/2)^k}{k!})(\frac{x^{(2k+n)/2-1}}{2^{(2k+n)/2}\Gamma((2k+n)/2)}.e^{-x/2}) \\
&= \sum_{k=0}^{\infty}(e^{-\theta^2/2}.\frac{(\theta^2/2)^k}{k!})f_{2k+n}(x).
\end{aligned}$$

□

Problem 2.39. Consider the marked Poisson process in Example 2.22 where the call starting times arrive as a homogeneous Poisson process (with rate λ calls/minute) on the entire real line and the call lengths are continuous random variables with density function $f(x)$. In Example 2.22, we showed that the distribution of $N(t)$ is independent of t.

(a) Show that for any r,

$$Cov(N(t), N(t+r)) = \lambda\int_{|r|}^{\infty} xf(x)dx = \lambda[|r|(1-F(|r|)) + \int_{|r|}^{\infty}(1-F(x))dx]$$

and hence is independent of t and depends only on $|r|$.

(b) Suppose that the call lengths are Exponential random variables with mean μ. Evaluate $Cov(N(t), N(t+r))$. (This is called the autocovariance function of $N(t)$.)

(c) Suppose that the call lengths have a density function

$$f(x) = \alpha.x^{-\alpha-1} \quad \text{if} \quad x \geq 1.$$

Show that $E(X_i) < \infty$ if, and only if, $\alpha > 1$ and evaluate $Cov(N(t), N(t+r))$ in this case.

(d) Compare the autocovariance functions obtained in parts (b) and (c). For which distribution does $Cov(N(t), N(t+r))$ decay to 0 more slowly as $|r| \to \infty$?

Solution. (a) Fix $r > 0$, then:

$$\begin{aligned}
Cov(N(t), N(t+r)) &= Cov(\sum_{i=1}^{\infty} I_{(S_i \le t, t \le S_i + X_i < t+r)} + \sum_{i=1}^{\infty} I_{(S_i \le t, t+r < S_i + X_i)}, \\
&\quad \sum_{i=1}^{\infty} I_{(S_i \le t, t+r \le S_i + X_i)} + \sum_{i=1}^{\infty} I_{(t < S_i \le t+r, t+r \le S_i + X_i)}) \\
&= Cov(\sum_{i=1}^{\infty} I_{(S_i \le t, t+r \le S_i + X_i)}, \sum_{i=1}^{\infty} I_{(S_i \le t, t+r \le S_i + X_i)}) \\
&= Cov(\sum_{i=1}^{\infty} I_{(S_i \le t, t \le S_i + X_i.1_{X_i > r})}, \sum_{i=1}^{\infty} I_{(S_i \le t, t \le S_i + X_i.1_{X_i > r})}) \\
&= Var(\sum_{i=1}^{\infty} I_{(S_i \le t, t \le S_i + X_i.1_{X_i > r})}) \\
&= E(\sum_{i=1}^{\infty} I_{(S_i \le t, t \le S_i + X_i.1_{X_i > r})}) \\
&= \lambda.E(X.1_{X>r}) = \lambda. \int_0^{\infty} x.1_{x>r} f(x) dx = \lambda. \int_r^{\infty} x.f(x) dx.
\end{aligned}$$

(b)

$$\begin{aligned}
AF_1(r) &= Cov(N(t), N(t+r)) = \lambda.(|r|.S_X(|r|) + \int_{|r|}^{\infty} S_X(x) dx) \\
&= \lambda.(|r|.e^{-\mu |r|} + \int_{|r|}^{\infty} e^{-\mu.x} dx) \\
&= \lambda.|r|.e^{-\mu |r|}(|r| + \frac{1}{\mu}).
\end{aligned}$$

(c)First,

$$E(X) = \int_1^{\infty} x.f_X(x) dx = \int_1^{\infty} \frac{\alpha}{x^{\alpha}} dx < \infty \Leftrightarrow \alpha > 1.$$

Second,

$$\begin{aligned}
AF_2(r) &= Cov(N(t), N(t+r)) = \lambda. \int_{|r|}^{\infty} x.f(x) dx \\
&= \lambda. \int_{|r|}^{\infty} \frac{\alpha}{x^{\alpha}} dx = \lambda. \frac{\alpha}{\alpha - 1}.(1/|r|)^{\alpha - 1}.
\end{aligned}$$

(d)First, as

$$\lim_{r \to \infty} \frac{AF_1(r)}{AF_2(r)} = \lim_{r \to \infty} \frac{\lambda.|r|.e^{-\mu |r|}(|r| + \frac{1}{\mu})}{\lambda. \frac{\alpha}{\alpha-1}.(1/|r|)^{\alpha-1}} = \lim_{r \to \infty} (\frac{\alpha - 1}{\alpha})(\frac{|r|^{\alpha} + |r|^{\alpha-1}/|\mu|}{e^{\mu.|r|}}) = 0,$$

it follows that $AF_1 = o(AF_2)$ and AF_2 tends to 0 slower than AF_1.
Second, define:

$$G(r) =^{def} \frac{AF_1(r)}{AF_2(r)} = (\frac{\alpha - 1}{\alpha})(\frac{|r|^{\alpha} + |r|^{\alpha-1}/|\mu|}{e^{\mu.|r|}}) \quad -\infty < r < \infty.$$

Then, $\lim_{r \to \pm\infty} G(r) = 0 = \lim_{r \to 0} G(r)$ and $G(r) \ge 0 \quad (-\infty < r < \infty)$. Furthermore, G takes its maximum value at $r_0 = \frac{(\alpha-1) + \sqrt{(\alpha-1)^2 + 4(\alpha-1)}}{2\mu}$, (Exercise!). Hence: $AF_1 \le G(r_0).AF_2$.
□

Chapter 3

Convergence of Random Variables

Problem 3.1. (a) Suppose that $\{X_n^{(1)}\}, \cdots, \{X_n^{(k)}\}$ are sequences of random variables with $X_n^{(i)} \to_p 0$ as $n \to \infty$ for each $i = 1, \cdots, k$. Show that

$$\max_{1\leq i\leq k} |X_n^{(i)}| \to_p 0$$

as $n \to \infty$.
(b) Find an example to show that the conclusion of (a) is not necessarily true if the number of sequences $k = k_n \to \infty$.

Solution. (a) We prove the assertion by induction on k. For $k = 1$, it trivially holds. Let it hold for $k > 1$, (induction hypothesis). Then, for $\epsilon > 0$, as $\lim_{n\to\infty} P(\max_{1\leq i\leq k} |X_n^{(i)}| \leq \epsilon) = 1 = \lim_{n\to\infty} P(|X_n^{(k+1)}| \leq \epsilon)$, it follows that:

$$1 = \lim_{n\to\infty} P(\max_{1\leq i\leq k} |X_n^{(i)}| \leq \epsilon) \leq \lim_{n\to\infty} P((\max_{1\leq i\leq k} |X_n^{(i)}| \leq \epsilon) \cup (|X_n^{(k+1)}| \leq \epsilon)) \leq 1,$$

and consequently:

$$\lim_{n\to\infty} P((\max_{1\leq i\leq k} |X_n^{(i)}| \leq \epsilon) \cup (|X_n^{(k+1)}| \leq \epsilon)) = 1. \qquad (*)$$

Now, using (*) and another application of above assumptions and Proposition 1.1.(c), it follows that:

$$\begin{aligned}
\lim_{n\to\infty} P(\max_{1\leq i\leq k+1} |X_n^{(i)}| > \epsilon) &= \lim_{n\to\infty} P(\max(\max_{1\leq i\leq k} |X_n^{(i)}|, |X_n^{(k+1)}|) > \epsilon) \\
&= 1 - \lim_{n\to\infty} P(\max(\max_{1\leq i\leq k} |X_n^{(i)}|, |X_n^{(k+1)}|) \leq \epsilon) \\
&= 1 - \lim_{n\to\infty} P((\max_{1\leq i\leq k} |X_n^{(i)}| \leq \epsilon) \cap (|X_n^{(k+1)}| \leq \epsilon)) \\
&= 1 - \lim_{n\to\infty} [P(\max_{1\leq i\leq k} |X_n^{(i)}| \leq \epsilon) + P(|X_n^{(k+1)}| \leq \epsilon) \\
&- P((\max_{1\leq i\leq k} |X_n^{(i)}| \leq \epsilon) \cup (|X_n^{(k+1)}| \leq \epsilon))] \\
&= 1 - (1 + 1 - 1) = 0.
\end{aligned}$$

(b) Fix $1 \leq i$, and define $\{X_n^{(i)}\}_{n=1}^{\infty} = \{\frac{i}{n^2}\}_{n=1}^{\infty}$. Then, $\lim_{n\to\infty} X_n^{(i)} =^P 0 \ \ (1 \leq i)$. Furthermore, for $k(n) = n^2$, we have $\max_{1\leq i\leq k(n)} |X_n^{(i)}| = 1$, and consequently, $\lim_{n\to\infty} \max_{1\leq i\leq k(n)} |X_n^{(i)}| =^P 1 \neq 0$.
□

Problem 3.3. Suppose that $X_1, \cdots, X_n$ are i.i.d Exponential random variables with parameter λ and let $M_n = \max(X_1, \cdots, X_n)$. Show that $M_n - ln(n)/\lambda \to_d V$ where

$$P(V \leq x) = \exp[-\exp(-\lambda x)]$$

for all x.

Solution.

$$\begin{aligned}
\lim_{n\to\infty} P(M_n - \frac{ln(n)}{\lambda} \leq x) &= \lim_{n\to\infty} P(M_n \leq \frac{ln(n)}{\lambda} + x) = \lim_{n\to\infty} \prod_{i=1}^{n} P(X_i \leq \frac{ln(n)}{\lambda} + x) \\
&= \lim_{n\to\infty} \prod_{i=1}^{n} (1 - e^{-\lambda(\frac{ln(n)}{\lambda}+x))}) = \lim_{n\to\infty} (1 - e^{-(ln(n)+\lambda.x))})^n \\
&= \lim_{n\to\infty} (1 - \frac{-e^{-\lambda.x}}{n})^n = \exp(-\exp(-\lambda.x)) \quad -\infty < x < \infty.
\end{aligned}$$

□

Problem 3.5. Suppose that X_N has a Hyper-geometric distribution (see Example 1.13) with the following frequency function

$$f_N(x) = \frac{C(M_N, x)C(N - M_N, r_N - x)}{C(N, r_N)}$$

for $x = \max(0, r_N + M_N - N), \cdots, \min(M_N, r_N)$. When the population size N is large, it becomes somewhat difficult to compute probabilities using $f_N(x)$ so that it is desirable to find approximations to the distribution of X_N as $N \to \infty$.

(a) Suppose that $R_N \to r$(finite) and $M_N/N \to \theta$ for $0 < \theta < 1$. Show that $X_N \to_d Bin(r, \theta)$ as $N \to \infty$.

(b) Suppose that $r_N \to \infty$ with $r_N M_N/N \to \lambda > 0$. Show that $X_N \to_d Pois(\lambda)$ as $N \to \infty$.

Solution. (a) Using Stirling's formulae we have:

$\lim_{r_N\to r, \frac{M_N}{N}\to\theta} f_N(x) =$

$\lim_{r_N\to r, \frac{M_N}{N}\to\theta} \frac{C(M_N,x)C(N-M_N,r_N-x)}{C(N,r_N)} =$

$\lim_{r_N\to r, \frac{M_N}{N}\to\theta} \frac{(\frac{M_N!}{(M_N-x)!x!})(\frac{(N-M_N)!}{(N-M_N-(r_N-x))!(r_N-x)!})}{(\frac{N!}{(N-r_N)!r_N!})} =$

$\lim_{r_N\to r, \frac{M_N}{N}\to\theta}[\frac{r_N!}{(r_N-x)!x!} \frac{M_N!(N-M_N)!(N-r_N)!}{(M_N-x)!(N-M_N-(r_N-x))!N!}] =$

$C(r,x). \lim_{r_N\to r, \frac{M_N}{N}\to\theta} \frac{(M_N^{M_N+1/2})((N-M_N)^{N-M_N+1/2})((N-r_N)^{N-r_N+1/2})}{((M_N-x)^{M_N-x+1/2})((N-M_N-(r_N-x))^{N-M_N-(r_N-x)+1/2})(N^{N+1/2})} =$

$C(r,x). \lim_{r_N\to r, \frac{M_N}{N}\to\theta}[(\frac{M_N}{M_N-x})^{M_N} * (\frac{M_N}{M_N-x})^{1/2} * (\frac{N-M_N}{N-M_N-(r_N-x)})^{1/2} * (\frac{(M_N-x)^x(N-M_N-(r_N-x))^{r_N-x}}{(N-r_N)^{r_N}}) *$
$(\frac{N-r_N}{N})^N * (\frac{N-r_N}{N})^{1/2} * (\frac{N-M_N}{N-M_N-(r_N-x)})^{N-M_N}] =$

$C(r,x). \lim_{r_N\to r, \frac{M_N}{N}\to\theta}[e^x * 1 * 1 * ((\frac{M_N-x}{N-r_N})^x(\frac{N-M_N-(r_N-x)}{N-r_N})^{r_N-x}) * e^{-r} * 1 * e^{r-x}] =$

$C(r,x).\theta^x.(1-\theta)^{r-x}.$

(b) Using Stirling's formulae as in part (a) we have:

$$\begin{aligned}
\lim_{r_N\to\infty, r_N\frac{M_N}{N}\to\lambda} f_N(x) &= \lim_{r_N\to\infty, r_N\frac{M_N}{N}\to\lambda} \frac{C(M_N,x).C(N-M_N,r_N-x)}{C(N,r_N)} \\
&= \lim_{r_N\to\infty, r_N\frac{M_N}{N}\to\lambda} \frac{e^{-x}r_N^x e^x M_N^x e^x (N-M_N)^{r_N-x}e^{r_N-x}}{N^{r_N}e^{r_N}} \\
&= \frac{1}{x!}\lim_{r_N\to\infty, r_N\frac{M_N}{N}\to\lambda} (\frac{r_N M_N}{N})^x(\frac{N-M_N}{N})^{r_N-x} \\
&= \frac{1}{x!}.\lambda^x.\lim_{r_N\to\infty, r_N\frac{M_N}{N}\to\lambda}(1-\frac{r_N\frac{M_N}{N}}{r_N})^{r_N} \\
&= \frac{1}{x!}.\lambda^x.e^{-\lambda}.
\end{aligned}$$

□

Problem 3.7. (a) Let $\{X_n\}$ be a sequence of random variables. Suppose that $E(X_n)\to\theta$(where θ is finite) and $Var(X_n)\to 0$. Show that $X_n\to_p\theta$.
(b) A sequence of random variables $\{X_n\}$ converges in probability to infinity ($X_n\to_p\infty$) if for each $M>0$,

$$\lim_{n\to\infty} P(X_n\le M)=0.$$

Suppose that $E(X_n)\to\infty$ and $Var(X_n)\le k.E(X_n)$ for some $k<\infty$. Show that $X_n\to_p\infty$.

Solution. (a) Given $\epsilon>0$, then by Theorem 3.7:

$$\begin{aligned}
\lim_{n\to\infty} P(|X_n-\theta|>\epsilon) &\le \lim_{n\to\infty}\frac{E((X_n-\theta)^2)}{\epsilon^2} \\
&= \lim_{n\to\infty}\frac{Var(X_n)+(E(X_n)-\theta)^2}{\epsilon^2} \\
&= 0,
\end{aligned}$$

implying: $\lim_{n\to\infty}P(|X_n-\theta|>\epsilon)=0$.

(b) Given $M>0$, then there is $N\ge 1$ such that for any $n\ge N$ we have $M<(1-\frac{1}{M^2+1})E(X_n)$. Consequently, for $\epsilon=\frac{1}{M^2+1}$ an application of Theorem 3.7 yields:

$$\begin{aligned}
\lim_{n\to\infty}P(X_n\le M) &\le \lim_{n\to\infty}P(X_n\le(1-\epsilon)E(X_n)) = \lim_{n\to\infty}P(X_n-E(X_n)\le -\epsilon.E(X_n)) \\
&\le \lim_{n\to\infty}P(|X_n-E(X_n)|\ge\epsilon.E(X_n)) \le \lim_{n\to\infty}\frac{E(|X_n-E(X_n)|^2)}{\epsilon^2.E^2(X_n)} \\
&\le \lim_{n\to\infty}\frac{k}{\epsilon^2.E^2(X_n)} = 0,
\end{aligned}$$

and consequently, $\lim_{n\to\infty}P(X_n\le M)=0$.
□

Problem 3.9. Suppose that $X_1,\cdots,X_n$ are i.i.d. Poisson random variables with mean λ. By the CLT,

$$\sqrt{n}(\overline{X}_n-\lambda)\to_d N(0,\lambda).$$

(a) Find the limiting distribution of $\sqrt{n}(ln(\overline{X}_n) - ln(\lambda))$.
(b) Find a function g such that

$$\sqrt{n}(g(\overline{X}_n) - g(\lambda)) \to_d N(0,1).$$

Solution. (a) By Theorem 3.4. for $a_n = \sqrt{n}$, $g(x) = ln(x)$ $(g'(x) = \frac{1}{x})$ and $\overline{X}_n$ instead of X_n we have:

$$\lim_{n\to\infty} \sqrt{n}(ln(\overline{X}_n) - ln(\lambda)) =^d \frac{1}{\lambda}.N(0,\lambda) =^d N(0,\frac{1}{\lambda}).$$

(b) One non-trivial answer is $g(x) = 2\sqrt{x}$ $(g'(x) = \frac{1}{\sqrt{x}})$:

$$\lim_{n\to\infty} \sqrt{n}(2\sqrt{\overline{X}_n} - 2\sqrt{\lambda}) =^d \frac{1}{\sqrt{\lambda}}N(0,\lambda) =^d N(0,1).$$

□

Problem 3.11. The sample median of i.i.d. random variables is asymptotically Normal provided that the distribution function F has a positive derivative at the median; when this condition fails, an asymptotic distribution may still exist but will be non-Normal. To illustrate this, let $X_1, \cdots, X_n$ be i.i.d. random variables with density

$$f(x) = \frac{1}{6}|x|^{-2/3} \quad \text{for} |x| \leq 1.$$

(Notice this density has a singularity at 0.)
(a) Evaluate the distribution function X_i and its inverse (the quantile function).
(b) Let M_n be the sample median of $X_1, \cdots, X_n$. Find the limiting distribution of $n^{3/2}M_n$.

Solution. (a) As,

$$F_X(x) = P(X \leq x) = 1_{[-1,1]}(x).(\frac{1+x^{1/3}}{2}) + 1_{(1,\infty)}(x).$$

it follows: $F^{-1}(t) = (2t-1)^3$ $(0 < t < 1)$.

(b) First, for $U_1, \cdots, U_n \sim Unif[0,1]$ with $E(U_i) = 1/2$ and $Var(U_i) = 1/12$ as application of Theorem 3.8 yields:

$$\sqrt{n}(\overline{U_n} - 1/2) \to_d N(0, 1/12). \quad (*)$$

Second, by Problem 3.10(c) for $k \geq 1$:

$$a_n^k(g(X_n) - g(\theta)) \to_d \frac{1}{k!}g^{(k)}(\theta)Z^k \quad (**)$$

Now, in (*) and (**) take $g(\theta) = F^{-1}(\theta)$. Then, $g^{(1)}(1/2) = g^{(2)}(1/2) = 0$, and $g^{(3)}(1/2) = 48 \neq 0$. Consequently:

$$(\sqrt{n})^3.(F^{-1}(\overline{U_n}) - F^{-1}(1/2)) \to_d \frac{1}{3!}48Z^3, \quad (***)$$

and by $F^{-1}(\overline{U_n}) = M_n$ and $F^{-1}(1/2) = 0$ it follows from (***) that:

$$n^{3/2}M_n \to_d 8Z^3 : \quad Z \sim N(0, 1/12).$$

□

Problem 3.13. Suppose that $X_1, \cdots, X_n$ be i.i.d. discrete random variables with frequency function

$$f(x) = \frac{x}{21} \quad \text{for} \quad x = 1, 2, \cdots, 6.$$

(a) Let $S_n = \sum_{k=1}^{n} k.X_k$. Show that

$$\frac{(S_n - E(S_n))}{\sqrt{Var(S_n)}} \to_d N(0,1).$$

(b) Suppose $n = 20$. Use a Normal approximation to evaluate $P(S_{20} \geq 1000)$.
(c) Suppose $n = 5$. Compute the exact distribution of S_n using the probability generating function of S_n (See Problems 1.18 and 2.8).

Solution. (a) As $E(X) = \frac{13}{3}$ and $Var(X) = \frac{20}{9}$ it follows that $Var(S_n) = \frac{20}{9}\sum_{k=1}^{n} k^2$. Hence, by Theorem 3.9 for $X_k^* = \frac{X_k - E(X_k)}{\sqrt{20/9}}$ in which $E(X_k^*) = 0$ and $Var(X_k^*) = 1$ it follows that:

$$\begin{aligned}
\lim_{n\to\infty} \frac{(S_n - E(S_n))}{\sqrt{Var(S_n)}} &= \lim_{n\to\infty} \frac{\sum_{k=1}^{n} k.X_k - E(\sum_{k=1}^{n} k.X_k)}{\sqrt{20/9}\sum_{k=1}^{n} k^2} \\
&= \lim_{n\to\infty} \frac{1}{\sqrt{\sum_{k=1}^{n} k^2}} \sum_{k=1}^{n} k.X_k^* \\
&=^d N(0,1).
\end{aligned}$$

(b) As $E(S_n) = \frac{13}{6}n(n+1)$ and $Var(S_n) = \frac{10}{27}n(n+1)(2n+1)$ $(n \geq 1)$, it follows that:

$$P(S_{20} \geq 1000) = P(\frac{S_{20} - E(S_{20})}{\sqrt{Var(S_{20})}} \geq \frac{1000 - (13/6).20.21}{\sqrt{(10/27).20.21.41}}) = P(Z \geq 1.127) \approx 0.13.$$

(c) By:

$$\begin{aligned}
P_{S_5}(t) &= \prod_{k=1}^{5} P_{k.X_k}(t) = \prod_{k=1}^{5} E(t^{k.X}) = \prod_{k=1}^{5}(\sum_{x=1}^{6}(t^k)^x.\frac{x}{21}) \\
&= (\frac{1}{21})^5 \prod_{k=1}^{5}(\sum_{x=1}^{6} x.t^{k.x}) = \frac{\sum_{x_1=1}^{6}\sum_{x_2=1}^{6}\sum_{x_3=1}^{6}\sum_{x_4=1}^{6}\sum_{x_5=1}^{6}(x_1.x_2.x_3.x_4.x_5)t^{x_1+2.x_2+3.x_3+4.x_4+5.x_5}}{21^5},
\end{aligned}$$

it follows that:

$$P(S_5 = k) = \frac{P_{S_5}^{(k)}(0)}{k!} = \frac{\sum_{x_1+2.x_2+3.x_3+4.x_4+5.x_5=k:1\leq x_i\leq 6}(x_1.x_2.x_3.x_4.x_5)}{21^5}.$$

□

Problem 3.15. Suppose that $X_{n1}, X_{n2}, \cdots, X_{nn}$ are independent Bernoulli random variables with parameters $\theta_{n1}, \cdots, \theta_{nn}$ respectively. Define $S_n = X_{n1} + X_{n2} + \cdots + X_{nn}$.
(a) Show that the moment generating function of S_n is

$$m_n(t) = \prod_{i=1}^{n}(1 - \theta_{ni} + \theta_{ni}\exp(t)).$$

(b) Suppose that

$$\sum_{i=1}^{n}\theta_{ni} \to \lambda > 0 \text{ and } \max_{1\leq i\leq n}\theta_{ni} \to 0$$

as $n \to \infty$. Show that

$$ln(m_n(t)) = \lambda[\exp(t) - 1] + r_n(t)$$

where for each t, $r_n(t) \to 0$ as $n \to \infty$.
(c) Deduce from part (b) that $S_n \to_d Pois(\lambda)$.

Solution. (a)

$$m_n(t) = E(e^{t.S_n}) = E(e^{t.\sum_{i=1}^n X_{ni}}) = \prod_{i=1}^n E(e^{t.X_{ni}}) = \prod_{i=1}^n ((1-\theta_{ni}) + \theta_{ni}.e^t).$$

(b) By definition:

$$\begin{aligned}
\lim_{n\to\infty} ln(m_n(t)) &= \lim_{n\to\infty} \sum_{i=1}^n ln(1 + \theta_{ni}(e^t - 1)) \\
&= \lim_{n\to\infty} \sum_{i=1}^n (\sum_{k=1}^\infty \frac{(-1)^{k-1}(\theta_{ni}(e^t-1))^k}{k}) \\
&= \lim_{n\to\infty} \sum_{i=1}^n (\theta_{ni}.(e^t - 1) + \sum_{k=2}^\infty \frac{(-1)^{k-1}(\theta_{ni}(e^t-1))^k}{k}) \\
&= \lim_{n\to\infty} [\sum_{i=1}^n \theta_{ni}.(e^t - 1) + \sum_{i=1}^n (\sum_{k=2}^\infty \frac{(-1)^{k-1}(\theta_{ni}(e^t-1))^k}{k})] \\
&= \lambda.(e^t - 1) + \lim_{n\to\infty} r_n(t),
\end{aligned}$$

in which

$$\begin{aligned}
\lim_{n\to\infty} |r_n(t)| &\leq \lim_{n\to\infty} \sum_{i=1}^n (\sum_{k=2}^\infty \frac{(\theta_{ni}|(e^t-1))^k|}{k}) \\
&\leq \lim_{n\to\infty} [\sum_{i=1}^n \theta_{ni} * (\sum_{k=2}^\infty \frac{(\theta_{ni}|(e^t-1))^{k-1}|}{k} |e^t - 1|)] \\
&\leq \lim_{n\to\infty} [\sum_{i=1}^n \theta_{ni} * (\sum_{k=2}^\infty \frac{(\max_{1\leq i\leq n}(\theta_{ni})|(e^t-1))^{k-1}|}{k} |e^t - 1|)] \\
&= \lim_{n\to\infty} [\sum_{i=1}^n \theta_{ni}] * \lim_{n\to\infty} [(\sum_{k=2}^\infty \frac{(\max_{1\leq i\leq n}(\theta_{ni})|(e^t-1))^{k-1}|}{k} |e^t - 1|)] \\
&= \lambda * 0 = 0,
\end{aligned}$$

or $\lim_{n\to\infty} |r_n(t)| = 0$.

(c) As

$$\lim_{n\to\infty} m_{S_n}(t) = \lim_{n\to\infty} (e^{\lambda(\exp(t)-1)}.e^{r_n(t)}) = e^{\lambda(\exp(t)-1)} = m_{\text{Pois}(\lambda)}(t) \quad -\infty < t < \infty,$$

by second method described on page 126, it follows that:

$$\lim_{n\to\infty} S_n =^d \text{Pois}(\lambda).$$

□

Problem 3.17. Suppose that $X_1, \cdots, X_n$ are independent nonnegative random variables with hazard functions $\lambda_1(x), \cdots, \lambda_n(x)$ respectively. Define $U_n = \min(X_1, \cdots, X_n)$.
(a) Suppose that for some $\alpha > 0$,

$$\lim_{n\to\infty} \frac{1}{n^\alpha} \sum_{i=1}^{n} \lambda_i(t/n^\alpha) = \lambda_0(t)$$

for all $t > 0$ where $\int_0^\infty \lambda_0(t)dt = \infty$. Show that $n^\alpha U_n \to_d V$ where $P(V > x) = \exp(-\int_0^x \lambda_0(t)dt)$.
(b) Suppose that $X_1, \cdots, X_n$ are i.i.d. Weibull random variables (see Example 1.19) with density function

$$f(x) = \lambda.\beta.x^{\beta-1} \exp(-\lambda.x^\beta) \quad (x > 0)$$

where $\lambda, \alpha > 0$. Let $U_n = \min(X_1, \cdots, X_n)$ and find α such that $n^\alpha U_n \to_d V$.

Solution. (a) By fundamental relationship between survival and hazard functions in page 28 it follows:

$$\begin{aligned}
\lim_{n\to\infty} S_{n^\alpha.U_n}(t) &= \lim_{n\to\infty} \prod_{i=1}^{n} S_{n^\alpha.X_i}(t) = \lim_{n\to\infty} \prod_{i=1}^{n} S_{X_i}(t/n^\alpha) \\
&= \lim_{n\to\infty} \prod_{i=1}^{n} \exp(-\int_0^{t/n^\alpha} \lambda_i(u)du) = \lim_{n\to\infty} \exp(-\sum_{i=1}^{n} \int_0^t \lambda_i(u/n^\alpha)du) \\
&= \exp(-\int_0^t \lim_{n\to\infty} \frac{\sum_{i=1}^n \lambda_i(u/n^\alpha)}{n^\alpha} du) = \exp(-\int_0^t \lambda_0(u)du) = S_V(t). \quad (0 < t < \infty)
\end{aligned}$$

(b) By Part (a) it is sufficient to find $\alpha > 0$ such that $\lim_{n\to\infty} \frac{1}{n^\alpha} \sum_{i=1}^n \lambda_i(t/n^\alpha) = \lambda_0(t)$ in which $\int_0^\infty \lambda_0(t)dt = \infty$. But by Example 1.19, $\lambda_i(t) = \lambda.\beta.t^{\beta-1} \quad (t > 0)$, and furthermore:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n^\alpha} \sum_{i=1}^{n} \lambda_i(t/n^\alpha) &= \lim_{n\to\infty} \frac{n(\lambda.\beta.(t/n^\alpha)^{\beta-1})}{n^\alpha} \\
&= \lim_{n\to\infty} \lambda.\beta.t^{\beta-1} \frac{1}{n^{\alpha.\beta-1}} \\
&= \lambda.\beta.t^{\beta-1} \quad \text{if } \alpha = \frac{1}{\beta}.
\end{aligned}$$

Note that, $\int_0^\infty \lambda.\beta.t^{\beta-1}dt = \infty$, hence we may take $\lambda_0(t) = \lambda.\beta.t^{\beta-1} \quad (t > 0)$.
□

Problem 3.19. Suppose that $\{X_n\}$ is a sequence of random variables such that $X_n \to_d X$ where $E(X)$ is finite. We would like to investigate sufficient conditions under which $E(X_n) \to E(X)$ (assuming that $E(X_n)$ is well-defined). Note that in Theorem 3.5, we indicated that this convergence holds if the $X_n's$ are uniformly bounded.
(a) Let $\delta > 0$. Show that

$$E(|X_n|^{1+\delta}) = (1+\delta) \int_0^\infty x^\delta P(|X_n| > x)dx.$$

(b) Show that for any $M > 0$ and $\delta > 0$,

$$\int_0^M P(|X_n| > x)dx \le E(|X_n|) \le \int_0^M P(|X_n| > x)dx + \frac{1}{M^\delta} \int_M^\infty x^\delta P(|X_n| > x)dx.$$

(c) Again let $\delta > 0$ and suppose that $E(|X_n|^{1+\delta}) \leq K < \infty$ for all n. Assuming that $X_n \to_d X$, use results of parts (a) and (b) to show that $E(|X_n|) \to E(|X|)$ and $E(X_n) \to E(X)$.

Solution. (a) This follows from Problem 1.20 with replacing X with $|X_n|$ and $r = 1 + \delta$.

(b) Given $M > 0$. By definition in Page 33, we have:

$$\begin{aligned}
\int_0^M P(|X_n| > x)dx &\leq \int_0^\infty P(|X_n| > x)dx = E(|X_n|) \\
&= \int_0^M P(|X_n| > x)dx + \int_M^\infty P(|X_n| > x)dx \\
&\leq^{M<x} \int_0^M P(|X_n| > x)dx + \int_M^\infty (\frac{x}{M})^\delta P(|X_n| > x)dx.
\end{aligned}$$

(c) Fix, $M > 0$ and $n \geq 1$. Then, by Part (b):

$$\begin{aligned}
\int_0^M P(|X_n| > x)dx &\leq E(|X_n|) \leq \int_0^M P(|X_n| > x)dx + \frac{1}{M^\delta}\int_M^\infty x^\delta P(|X_n| > x)dx \\
&\leq \int_0^M P(|X_n| > x)dx + \frac{E(|X_n|^{1+\delta})}{M^\delta} \\
&\leq \int_0^M P(|X_n| > x)dx + \frac{K}{M^\delta}. \quad (*)
\end{aligned}$$

Taking limit as $n \to \infty$ from three sides of (*) it follows that:

$$\int_0^M P(|X| > x)dx \leq \lim_{n\to\infty}(E(|X_n|)) \leq \int_0^M P(|X| > x)dx + \frac{K}{M^\delta}. \quad (**)$$

Next, taking limit as $M \to \infty$ from three sides of (**) it follows that:

$$\int_0^\infty P(|X| > x)dx \leq \lim_{n\to\infty}(E(|X_n|)) \leq \int_0^\infty P(|X| > x)dx. \quad (***)$$

Consequently, by (***) and definition $E(|X|) = \int_0^\infty P(|X| > x)dx$, the assertion follows. Finally, the later assertion follows by considering $|E(X_n) - E(X)| \leq E(|X_n - X|)$ $(n \geq 1)$ and applying the first assertion for the case $X_n^* = X_n - X$ $(n \geq 1)$.

□

Problem 3.21. If $\{X_n\}$ is bounded in probability, we often write $X_n = O_P(1)$. Likewise, if $X_n \to_p 0$ then $X_n = o_P(1)$. This useful shorthand notation generalizes the big-oh and little-oh notation that is commonly used for sequences of numbers to sequences of random variables. If $X_n = O_P(Y_n)$ $(X_n = o_P(Y_n))$ then $X_n/Y_n = O_P(1)$ $(X_n/Y_n = o_P(1))$.

(a) Suppose that $X_n = O_P(1)$ and $Y_n = o_p(1)$. Show that $X_n + Y_n = O_P(1)$.

(b) Let $\{a_n\}$ and $\{b_n\}$ be sequences of constants where $a_n/b_n \to 0$ as $n \to \infty$ (that is, $a_n = o(b_n)$) and suppose that $X_n = O_P(a_n)$. Show that $X_n = o_P(b_n)$.

Solution. (a) Given $\epsilon > 0$, by assumption there is $N > 1$ and $M_\epsilon > 0$ such that:

$$P(|X_n| > \epsilon) < \frac{\epsilon}{2}\ (n \geq N), \qquad P(|Y_n| > M_\epsilon) < \frac{\epsilon}{2}\ (n \geq 1).$$

Then:

$$\begin{aligned} P(|X_n+Y_n|>M_\epsilon+\epsilon) &\le P(|X_n|+|Y_n|>M_\epsilon+\epsilon) \\ &= P(|X_n|+|Y_n|>M_\epsilon+\epsilon\cap|X_n|>\epsilon)+P(|X_n|+|Y_n|>M_\epsilon+\epsilon\cap|X_n|\le\epsilon) \\ &\le P(|X_n|>\epsilon)+P(|Y_n|\ge M_\epsilon) \\ &\le \epsilon/2+\epsilon/2=\epsilon. \quad (n\ge N) \end{aligned}$$

Next, for $1\le n\le N$, take $M_{n,\epsilon}>0$ such that $P(|X_n+Y_n|>M_{n,\epsilon})\le\epsilon$. Finally, take

$$M_\epsilon^*=\max(\max_{1\le n\le N}M_{n,\epsilon},M_\epsilon+\epsilon),$$

then:

$$\sup_{1\le n\le\infty}P(|X_n+Y_n|>M_\epsilon^*)\le\epsilon.$$

(b) Given $\epsilon>0$. There is $M_\epsilon>0$ such that $\sup_{1\le n<\infty}P(|\frac{X_n}{a_n}|>M_\epsilon)\le\epsilon$. Then, there is $N>1$ such that for any $n>N$ we have: $|\frac{a_n}{b_n}|<\frac{\epsilon}{M_\epsilon}$. Accordingly,

$$P(|\frac{X_n}{b_n}|>\epsilon)=P(|\frac{X_n}{a_n}|>\frac{\epsilon}{(|a_n/b_n|)})\le P(|\frac{X_n}{a_n}|>M_\epsilon)\le\epsilon \quad (n>N).$$

□

Problem 3.23. Suppose that $A_1,A_2,\cdots$ is a sequence of events. We are sometimes interested in determining the probability that infinitely many of the $A_k's$ occur. Define the event:

$$B=\cap_{n=1}^\infty\cup_{k=n}^\infty A_k.$$

It is possible to show that an outcome lies in B if, and only if, it belongs to infinitely many of the $A_k's$.
(a) Prove the first Borel-Cantelli Lemma: If $\sum_{k=1}^\infty P(A_k)<\infty$ then

$$P(A_k\ \text{ infinitely often})=P(B)=0.$$

(b) When the $A_k's$ are mutually independent, we can strengthen the first Borel-Cantelli Lemma. Suppose that

$$\sum_{k=1}^\infty P(A_k)=\infty$$

for mutually independent events $\{A_k\}$. Show that

$$P(A_k\ \text{ infinitely often})=P(B)=1;$$

this result is called the second Borel-Cantelli Lemma.

Solution.(a) By definition:

$$0\le P(B)=P(\cap_{n=1}^\infty\cup_{k=n}^\infty A_k)\le\inf_{n\ge1}P(\cup_{k=n}^\infty A_k)\le\inf_{n\ge1}\sum_{k=n}^\infty P(A_k)=0,$$

as $\sum_{k=1}^\infty P(A_k)<\infty$. Hence, $P(B)=0$.

(b) Given the assumption $\sum_{k=1}^{\infty} P(A_k) = \infty$ we have $\sum_{k=n}^{\infty} P(A_k) = \infty \ \ (n \geq 1)$ and hence:

$$\begin{aligned} 0 &\leq P(B^c) = P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k^c) \leq \sum_{n=1}^{\infty} P(\cap_{k=n}^{\infty} A_k^c) = \sum_{n=1}^{\infty} (\prod_{k=n}^{\infty} (1 - P(A_k))) \\ &\leq \sum_{n=1}^{\infty} (\prod_{k=n}^{\infty} \exp(-P(A_k))) = \sum_{n=1}^{\infty} \exp(-\sum_{k=n}^{\infty} P(A_k)) = \sum_{n=1}^{\infty} 0 = 0, \end{aligned}$$

implying $P(B^c) = 0$ or $P(B) = 1$.
□

Problem 3.25. Suppose that $X_1, X_2, \cdots$ are i.i.d. random variables with $E(X_i) = 0$ and $E(X_i^4) < \infty$. Define:

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

(a) Show that $E(|\overline{X_n}|^4) \leq k/n^2$ for some constant k.
(b) Using the first Borel-Cantelli Lemma, show that

$$\overline{X_n} \to_{wp1} 0.$$

(This gives a reasonably straightforward proof of the SLLN albeit under much stronger than necessary conditions.).

Solution. (a) By Problem 1.21, $E(X^2), E(X^3) < \infty$. Next, by Cauchy-Schwartz inequality in Problem 2.17 it follows:

$$\begin{aligned} E(|\overline{X_n}|^4) &= \frac{1}{n^4} E(|\sum_{i=1}^{n} X_i|^4) = \frac{1}{n^4} E(\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} X_{i_1} X_{i_2} X_{i_3} X_{i_4}) \\ &= \frac{1}{n^4} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} E(X_{i_1} X_{i_2} X_{i_3} X_{i_4}) \\ &= \frac{1}{n^4} [C(4,2) \sum_{i_1 \neq i_2} E(X_{i_1}^2 . X_{i_2}^2) + \sum_{i_1} E(X_{i_1}^4) \\ &\quad + C(4,1) \sum_{i_1 \neq i_2 = i_3 = i_4} E(X_{i_1} X_{i_2} X_{i_3} X_{i_4}) \\ &\quad + C(4,2) \sum_{i_1 \neq i_2 = i_3 \neq i_4} E(X_{i_1} X_{i_2} X_{i_3} X_{i_4}) \\ &\quad + \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} E(X_{i_1} X_{i_2} X_{i_3} X_{i_4})] \end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{n^4}[C(4,2)n^2 E(X_i^4) + n.E(X_i^4) \\
&\quad +C(4,1) \sum_{i_1 \neq i_2 = i_3 = i_4} E(X_{i_1})E(X_{i_2}X_{i_3}X_{i_4}) \\
&\quad +C(4,2) \sum_{i_1 \neq i_2 = i_3 \neq i_4} E(X_{i_1})E(X_{i_2}X_{i_3})E(X_{i_4}) \\
&\quad + \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} E(X_{i_1})E(X_{i_2})E(X_{i_3})E(X_{i_4})] \\
&\leq \frac{2.C(4,2).n^2}{n^4} E(X_i^4) = \frac{2.C(4,2).E(X_i^4)}{n^2},
\end{aligned}
$$

and hence for $k = 2.C(4,2).E(X_i^4)$ the assertion follows.

(b) Referring to discussion of Page 159, it is sufficient to prove that for any $\epsilon > 0$, we have

$$P(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} |\overline{X_k}| > \epsilon) = 0. \quad (*)$$

To do so, let $A_k = (|\overline{X_k}| > \epsilon)$ $(k \geq 1)$ and take $g(x) = x^4$ in Problem 3.8. Then:

$$\sum_{k=1}^{\infty} P(A_k) \leq \sum_{k=1}^{\infty} \frac{E(|\overline{X_k}|^4)}{\epsilon^4} \leq \frac{k}{\epsilon^4} \sum_{k^*=1}^{\infty} (\frac{1}{k^{*2}}) = \frac{k.\pi^2}{6.\epsilon^4} < \infty,$$

and the (*) follows by the first Borel-Cantelli Lemma.
□

Chapter 4

Principles of Point Estimation

Problem 4.1. Suppose that $X = (X_1, \cdots, X_n)$ has a one-parameter exponential family distribution with joint density or frequency function

$$f(x;\theta) = \exp[\theta.T(x) - d(\theta) + S(x)]$$

where the parameter space Θ is an open subset of R. Show that

$$E_\theta[\exp(sT(X))] = \exp(d(\theta + s) - d(\theta))$$

if s is sufficiently small.

Solution. Fix $\theta \in \Theta$. Given open $\Theta \subset R$, there is $\epsilon > 0$ such that for the open ball $B(\theta, \epsilon)$ we have $B(\theta, \epsilon) \subset \Theta$. Consequently, for any $s \in B(\theta, \epsilon)$:

$$\begin{aligned}
E_\theta[exp(sT(X))] &= \int_\chi exp(sT(x)).f(x;\theta)dx = \int_\chi exp(sT(x) + \theta.T(x) - d(\theta) + S(x))dx \\
&= (\int_\chi exp((s+\theta)T(x) - d(\theta+s) + S(x))dx)exp(d(\theta+s) - d(\theta)) \\
&= (\int_\chi f(x; s+\theta)dx)exp(d(\theta+s) - d(\theta)) \\
&= exp(d(\theta+s) - d(\theta)).
\end{aligned}$$

□

Problem 4.3. suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with density

$$f(x;\theta_1,\theta_2) = a(\theta_1,\theta_2)h(x) \quad \texttt{for} \;\; \theta_1 \le x \le \theta_2; \quad 0, \quad \texttt{otherwise}$$

where $h(x)$ is a known function defined on the real line.
(a) Show that

$$a(\theta_1,\theta_2) = (\int_{\theta_1}^{\theta_2} h(x)dx)^{-1}.$$

(b) Show that $(X_{(1)}, X_{(2)})$ is sufficient for (θ_1, θ_2).

Solution.(a) As

$$a(\theta_1,\theta_2).\int_{\theta_1}^{\theta_2} h(x)dx = \int_{\theta_1}^{\theta_2} a(\theta_1,\theta_2).h(x)dx = \int_\chi f(x;\theta_1,\theta_2)dx = 1,$$

it follows that: $a(\theta_1, \theta_2) = 1/(\int_{\theta_1}^{\theta_2} h(x)dx)$.

(b) Using Theorem 4.2 for the joint density function of $X = (X_1, \cdots, X_n)$ it follows:

$$\begin{aligned} f(\mathbf{x}; \theta_1, \theta_2) &= \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2) = \prod_{i=1}^{n} (a(\theta_1, \theta_2).h(x_i).1_{[\theta_1, \theta_2]}(x_i)) \\ &= (a(\theta_1, \theta_2)^n.1_{[\theta_1, +\infty)}(X_{(1)}).1_{(-\infty, \theta_2]}(X_{(n)})).(\prod_{i=1}^{n} h(x_i)) \\ &= g^*((X_{(1)}, X_{(n)}); (\theta_1, \theta_2)).h^*(\mathbf{x}), \end{aligned}$$

and accordingly, $(X_{(1)}, X_{(2)})$ is sufficient for (θ_1, θ_2).
□

Problem 4.5. Suppose that the lifetime of an electrical component is known to depend on some stress variable that varies over time; specifically, if U is the lifetime of the component, we have

$$\lim_{\Delta \downarrow 0} \frac{1}{\Delta} P(x \leq U \leq x + \Delta | U \geq x) = \lambda. \exp(\beta.\phi(x))$$

where $\phi(x)$ is the stress at time x. Assuming that we can measure $\phi(x)$ over time, we can conduct an experiment to estimate λ and β by replacing the component when it fails and observing the failure times of the components. Because $\phi(x)$ is not constant, the inter-failure times will not be i.i.d. random variables.
Define non-negative random variables $X_1 < \cdots < X_n$ such that X_1 has hazard function

$$\lambda_1(x) = \lambda. \exp(\beta.\phi(x))$$

and conditional on $X_i = x_i, X_{i+1}$ has hazard function

$$\lambda_{i+1}(x) = 0 \quad \texttt{if } x < x_i; \quad \lambda. \exp(\beta.\phi(x)), \quad \texttt{if } x \geq x_i$$

where λ, β are unknown parameters and $\phi(x)$ is a known function.
(a) Find the joint density of $(X_1, \cdots, X_n)$.
(b) Find sufficient statistics for (λ, β).

Solution. (a) Using fundamental relationship between density function and hazard function (page 29), it follows that:

$$\begin{aligned} f_{X_1, \cdots, X_n}(x_1, \cdots, x_n) &= \prod_{i=0}^{n-1} (\frac{f_{X_1, \cdots, X_{i+1}}(x_1, \cdots, x_{i+1})}{f_{X_1, \cdots, X_i}(x_1, \cdots, x_i)}) = \prod_{i=0}^{n-1} f_{X_{i+1}|X_1, \cdots, X_i}(x_{i+1}|x_1, \cdots, x_i) \\ &= \prod_{i=0}^{n-1} [\lambda_{X_{i+1}|X_1, \cdots, X_i}(x_{i+1}) exp(-\int_{x_i}^{x_{i+1}} \lambda_{X_{i+1}|X_1, \cdots, X_i}(t)dt)] \\ &= \prod_{i=0}^{n-1} [\lambda.1_{[x_i, \infty)}(x_{i+1}).exp(\beta.\phi(x_{i+1})).exp(-\int_{x_i}^{x_{i+1}} \lambda_{X_{i+1}|X_1, \cdots, X_i}(t)dt)] \\ &= 1_{0 < x_1 < \cdots < x_n}(x_1, \cdots, x_n).\lambda^n.exp(\sum_{i=1}^{n} \beta.\phi(x_i)).exp(-\lambda. \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} exp(\beta.\phi(t))dt) \\ &= 1_{0 < x_1 < \cdots < x_n}(x_1, \cdots, x_n).\lambda^n.exp(\sum_{i=1}^{n} (\beta.\phi(x_i)) - \lambda. \int_0^{x_n} exp(\beta\phi(t))dt). \end{aligned}$$

(b) Using Theorem 4.2 for the joint density function of $X = (X_1, \cdots, X_n)$ we have:

$$\begin{aligned} f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) &= \lambda^n.exp(\sum_{i=1}^{n}(\beta.\phi(x_i)) - \lambda.\int_0^{x_n} exp(\beta\phi(t))dt).1_{0<x_1<\cdots<x_n}(x_1,\cdots,x_n) \\ &= g^*((\sum_{i=1}^{n}\phi(x_i), x_n);(\beta,\lambda)).h^*(x_1,\cdots,x_n), \end{aligned}$$

and, thus $(\sum_{i=1}^{n}\phi(x_i), x_n)$ is sufficient for (β, λ).
□

Problem 4.7. Suppose that $X_1, \cdots, X_n$ are i.i.d. Uniform random variables on $[0, \theta]$:

$$f(x;\theta) = \frac{1}{\theta} \;\; \texttt{for} \;\; 0 \le x \le \theta.$$

Let $X_{(1)} = \min(X_1, \cdots, X_n)$ and $X_{(n)} = \max(X_1, \cdots, X_n)$.
(a) Define $T = X_{(n)}/X_{(1)}$. Is T ancillary for θ ?
(b) Find the joint distribution of T and $X_{(n)}$. Are T and $X_{(n)}$ independent ?

Solution. (a) First, let $X_1, \cdots, X_n$ be a random sample from a population with CDF F_X and pdf f_X. then, for the ordered statistics $X_{(1)} < \cdots < X_{(n)}$ and $1 \le i \ne j \le n$ we have (Casella & Berger, 2002):

$$f_{X_i,X_j}(u,v) = \frac{n!.f_X(u).f_X(v).F_X(u)^{i-1}.(F_X(v)-F_X(u))^{j-1-i}(1-F_X(v))^{n-j}}{(i-1)!(j-1-i)!(n-j)!}.1_{u<v}(u,v)$$

Thus, for our case of $i = 1$ and $j = n$ it follows that:

$$f_{X_{(1)},X_{(n)}}(u,v) = \frac{n(n-1)}{\theta.n}(\frac{v-u}{n})^{n-2}1_{0<u<v<\theta}(u,v) = \frac{n(n-1)(v-u)^{n-2}}{\theta^n}.1_{0<u<v<\theta}(u,v).$$

Consequently:

$$\begin{aligned} F_T(t) &= P(T \le t) = P(X_{(n)} \le t.X_{(1)}) = \int\int_{X_{(n)} \le t.X_{(1)}} f_{X_{(1)},X_{(n)}}(u,v)dudv \\ &= \int_0^\theta \int_{v/t}^{v} \frac{n(n-1)(v-u)^{n-2}}{\theta^n}dudv.1_{[1,\infty)}(t) = (1-\frac{1}{t})^{n-1}.1_{[1,\infty)}(t), \end{aligned}$$

implying:

$$f_T(t) = \frac{d}{dt}F_T(t) = \frac{(n-1).(t-1)^{n-2}}{t^n}1_{[1,\infty)}(t). \quad (*)$$

Thus, the density of T is independent of θ and hence T is ancillary statistics for it.

(b) Take $T = \frac{X_{(n)}}{X_{(1)}}$ and $W = X_{(n)}$; then, $X_{(1)} = \frac{W}{T}$, $X_{(n)} = W$, and $J = \frac{d(X_{(1)},X_{(n)})}{d(T,W)} = \frac{-W}{T^2}$. Now, by Theorem 2.3:

$$\begin{aligned} f_{T,W}(t,w) &= f_{X_{(1)},X_{(n)}}(\frac{w}{t},w).|J| = \frac{n(n-1)}{\theta^n}.(w-\frac{w}{t})^{n-2}.\frac{w}{t^2}.1_{\frac{w}{t}<w} \\ &= \frac{n(n-1)}{\theta^n}.w^{n-1}.(1-\frac{1}{t})^{n-2}.\frac{1}{t^2}.1_{1<t} = \frac{n(n-1)}{\theta^n}.\frac{w^{n-1}.(t-1)^{n-2}}{t^n}.1_{1<t}.\ (0 \le w \le \theta) \quad (**) \end{aligned}$$

Thus:

$$f_W(w) = \int_1^\infty f_{T,W}(t,w)dt = \frac{n.w^{n-1}}{\theta^n}\int_1^\infty \frac{(n-1).(t-1)^{n-2}}{t^n}dt = \frac{n.w^{n-1}}{\theta^n}. \quad (0 \le w \le \theta) \ (***)$$

Finally, by $(*), (**)$ and $(***)$ it follows that $f_{T,W}(t,w) = f_T(t).f_W(w)$, and thus W and T are independent.

□

Problem 4.9. Consider the Gini Index $\theta(F)$ as defined in example 4.21.
(a) Suppose that $X \sim F$ and let G be the distribution function of $Y = aX$ for some $a > 0$. Show that $\theta(G) = \theta(F)$.
(b) Suppose that F_p is a discrete distribution with probability p at 0 and probability $1-p$ at $x > 0$. Show that $\theta(F_p) \to 0$ as $p \to 0$ and $\theta(F_p) \to 1$ as $p \to 1$.
(c) Suppose that F is a Pareto distribution whose density is

$$f(x;\alpha) = \frac{\alpha}{x_0}(\frac{x}{x_0})^{-\alpha-1} \ \texttt{for} \ \ x > x_0 > 0 \ \ \alpha > 0,$$

(This is sometimes used as a model for income exceeding a threshold x_0). Show that $\theta(F) = (2.\alpha - 1)^{-1}$ $\texttt{for}\alpha > 1$. ($f(x;\alpha)$ is a density for $\alpha > 0$ but for $\alpha \le 1$, the expected value is infinite.)

Solution. (a) Referring to pages 191-192 we have:

$$\theta(F_X) = 1 - 2.\int_0^1 q_{F_X}(t)dt = 1 - 2.\int_0^1 (\frac{\int_0^t F_X^{-1}(s)ds}{\int_0^1 F_X^{-1}(s)ds})dt. \quad (*)$$

Next:

$$F_Y^{-1}(s) = \inf\{x : F_Y(x) \ge s\} = \inf\{x : F_X(\frac{x}{a}) \ge s\} = a.\inf\{\frac{x}{a} : F_X(\frac{x}{a}) \ge s\} = a.F_X^{-1}(s). \ (**)$$

Now, by $(*)$ and $(**)$ it follows that:

$$\theta(F_Y) = 1 - 2.\int_0^1 (\frac{\int_0^t F_Y^{-1}(s)ds}{\int_0^1 F_Y^{-1}(s)ds})dt = 1 - 2.\int_0^1 (\frac{\int_0^t a.F_X^{-1}(s)ds}{\int_0^1 a.F_X^{-1}(s)ds})dt = \theta(F_X).$$

(b)Using $(*)$ in part (a) and considering $F_p^{-1}(s) = x.1_{(p,1]}(s)$ it follows that:

$$\theta(F_p) = 1 - 2.\int_0^1 (\frac{\int_0^t x.1_{(p,1]}(s)ds}{\int_0^1 x.1_{(p,1]}(s)ds})dt = 1 - 2.\int_0^1 (\frac{\int_0^1 1_{[0,t]\cap(p,1]}(s)ds}{\int_0^1 1_{(p,1]}(s)ds})dt. \quad (\dagger)$$

Next, two times usage of $(\dagger)$, it follows that:

$$\lim_{p\to 0}\theta(F_p) = 1 - 2.\int_0^1 \lim_{p\to 0}(\frac{\int_0^1 1_{[0,t]\cap(p,1]}(s)ds}{\int_0^1 1_{(p,1)}(s)ds})dt = 1 - 2.\int_0^1 tdt = 1 - 1 = 0,$$

$$\lim_{p\to 1}\theta(F_p) = 1 - 2.\int_0^1 \lim_{p\to 1}(\frac{\int_0^1 1_{[0,t]\cap(p,1]}(s)ds}{\int_0^1 1_{(p,1)}(s)ds})dt = 1 - 2.\int_0^1 0dt = 1 - 0 = 1.$$

(c) As,

$$F(x;\alpha) = \int_{x_0}^x \frac{\alpha.x_0^\alpha}{t^{\alpha+1}}dt = 1 - (\frac{x_0}{x})^\alpha.1_{x>x_0}.$$

it follows that:

$$F^{-1}(s) = \inf\{x : F(x) \geq s\} = \inf\{x : 1 - (\frac{x_0}{x})^\alpha \geq s\} = x_0.(1-s)^{-\frac{1}{\alpha}},$$

implying:

$$\int_0^t F^{-1}(s)ds = \int_0^t x_0.(1-s)^{-\frac{1}{\alpha}}ds = x_0.\frac{\alpha}{\alpha-1}.[1-(1-t)^{1-\frac{1}{\alpha}}], \quad (0 \leq t \leq 1). \quad (\dagger\dagger)$$

Finally, by $(*)$ and $(\dagger\dagger)$ we have:

$$\theta(F) = 1 - 2.\int_0^1 (\frac{x_0.\frac{\alpha}{\alpha-1}.[1-(1-t)^{1-\frac{1}{\alpha}}]}{x_0.\frac{\alpha}{\alpha-1}})dt = 1 - 2[1 - \frac{\alpha}{2.\alpha - 1}] = \frac{1}{2\alpha - 1}.$$

□

Problem 4.11. the influence curve heuristic can be used to obtain the joint limiting distribution of a finite number of substitution principle estimators. Suppose that $\theta_1(F), \cdots, \theta_k(F)$ are functional parameters with influence curves $\phi_1(x : F), \cdots, \phi_k(x : F)$. The if $X_1, \cdots, X_n$ is an i.i.d. sample from F, we typically have:

$$\sqrt{n}(\theta_j(\widehat{F_n}) - \theta_j(F)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_j(X_i; F) + R_{nj} \quad (1 \leq j \leq k)$$

where $R_nj \to_p 0 \ \ (1 \leq j \leq k)$.
(a) Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables from a distribution F with mean μ and median θ; assume that $Var(X_i) = \sigma^2$ and $F'(\theta) > 0$. If $\widehat{\mu_n}$ is the sample mean and $\widehat{\theta_n}$ is the sample median , use the influence curve heuristic to show that

$$\sqrt{n}\begin{pmatrix} \widehat{\mu_n} - \mu \\ \widehat{\theta_n} - \theta \end{pmatrix} \to_d N_2(\mathbf{0}, \mathbf{C})$$

and give the elements of the variance-covariance matrix $\mathbf{C}$.

(b) Now assume that the $X_i's$ are i.i.d. with density

$$f(x;\theta) = \frac{p}{2\Gamma(1/p)}exp(-|x-\theta|^p)$$

where θ is the mean and median of the distribution and $p > 0$ is another parameter (that may be known or unknown). show that the matrix $\mathbf{C}$ in part (a) is:

$$\mathbf{C} = \begin{pmatrix} \Gamma(3/p)/\Gamma(1/p) & \Gamma(2/p)/p \\ \Gamma(2/p)/p & [\Gamma(1/p)/p]^2 \end{pmatrix}$$

(c) Consider estimators of the θ of the form $\widetilde{\theta_n} = s.\widehat{\mu_n} + (1-s).\widehat{\theta_n}$. For given s, find the limiting distribution of $\sqrt{n}(\widetilde{\theta_n} - \theta)$.

(d) For a given value of $p > 0$, find the value of s that minimizes the variance of this limiting distribution. For which value(s) of p is this optimal value equal to 0; for which values(s) is it equal to 1 ?

Solution. (a) Take

$$X_i^* = \begin{pmatrix} \phi_1(X_i, F) \\ \phi_2(X_i, F) \end{pmatrix} \ (1 \leq i \leq n) \text{ and } R_n^* = \begin{pmatrix} R_{n1} \\ R_{n2} \end{pmatrix} \ (n \geq 1).$$

Then,

$$\sqrt{n}\begin{pmatrix} \widehat{\mu_n} - \mu \\ \widehat{\theta}_n - \theta \end{pmatrix} = \frac{1}{\sqrt{n}}.\sum_{i=1}^{n} \begin{pmatrix} \phi_1(X_i, F) \\ \phi_2(X_i, F) \end{pmatrix} + \begin{pmatrix} R_{n1} \\ R_{n2} \end{pmatrix} = \frac{1}{\sqrt{n}}.\sum_{i=1}^{n} X_i^* + R_n^* = S_n^* + R_n^*,$$

in which $R_n^* \to_p \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. But, by Theorem 3.12, $S_n^* \to_d N_2(\mathbf{0}, \mathbf{C})$ in which $C_{ij} = Cov(\phi_1(X_i; F), \phi_2(X_j; F))$ $(1 \leq i, j \leq 2)$. Hence, by Theorem 3.3.

$$\sqrt{n}\begin{pmatrix} \widehat{\mu_n} - \mu \\ \widehat{\theta}_n - \theta \end{pmatrix} \to_d N_2(\mathbf{0}_{2\times 1}, \mathbf{C}_{2\times 2}) : C_{ij} = Cov(\phi_1(X_i; F), \phi_2(X_j; F)) \quad (1 \leq i, j \leq 2).$$

To compute the entries of the matrix $\mathbf{C}$, using Example 4.28 and argument in page 200, we have:

$$\begin{aligned}
C_{22} &= Var(\phi_2(X; F)) = Var(\frac{sgn(x-\theta(F))}{2.F'(\theta(F))}) = \int_{-\infty}^{\infty} (\frac{sgn(x-\theta(F))}{2.F'(\theta(F))})^2 dF(x) = \frac{1}{(2.F'(\theta))^2}, \quad (\theta(F) = \theta) \\
C_{11} &= Var(\phi_1(X; F)) = Var(x - \theta(F)) = \int_{-\infty}^{\infty} (x - \theta(F))^2 dF(x) = \sigma^2, \quad (\theta(F) = \mu) \\
C_{12} &= C_{21} = Cov(\phi_1(X; F), \phi_2(X; F)) = \int_{-\infty}^{\infty} (\frac{sgn(x-\theta)}{2F'(\theta)}).(x-\mu) dF(x) = E(\frac{sgn(X-\theta).(X-\mu)}{2.F'(\theta)}),
\end{aligned}$$

giving the following form of the variance-covariance matrix:

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & E(\frac{sgn(X-\theta).(X-\mu)}{2.F'(\theta)}) \\ E(\frac{sgn(X-\theta).(X-\mu)}{2.F'(\theta)}) & \frac{1}{(2.F'(\theta))^2} \end{pmatrix}.$$

(b) Using answer given in part (a) it follows that:

$$\begin{aligned}
C_{11} &= \sigma^2 = E((X-\theta)^2) = \int_{-\infty}^{\infty} \frac{p}{2.\Gamma(1/p)} |x-\theta|^2.e^{-|x-\theta|^p} dx \\
&= \frac{p}{\Gamma(1/p)} \int_{\theta}^{\infty} (x-\theta)^2.e^{-(x-\theta)^p} dx = \frac{p}{\Gamma(1/p)} \int_0^{\infty} y^{2/p}.e^{-y} \frac{dy}{p.y^{(p-1)/p}} \\
&= \frac{\int_0^{\infty} y^{3/p-1}.e^{-y} dy}{\Gamma(1/p)} = \frac{\Gamma(3/p)}{\Gamma(1/p)},
\end{aligned}$$

$$\begin{aligned}
C_{22} &= \frac{1}{(2.F'(\theta))^2} = \frac{1}{(2.f(\theta;\theta))^2} = \frac{1}{(p/\Gamma(1/p))^2} = (\frac{\Gamma(1/p)}{p})^2, \\
C_{12} &= E(\frac{sgn(X-\theta).(X-\theta)}{2.F'(\theta)}) = \frac{\Gamma(1/p)}{p} E(sgn(X-\theta).(X-\theta)) \\
&= \frac{\Gamma(1/p)}{p} \int_{-\infty}^{\infty} [\frac{p}{p\Gamma(1/p)} exp(-|x-\theta|^p).sgn(x-\theta).(x-\theta)] dx \\
&= \int_{\theta}^{\infty} (x-\theta).exp(-(x-\theta)^p) dx = \int_0^{\infty} y^{1/p}.e^{-y} \frac{dy}{p.y^{1-1/p}} \\
&= \frac{1}{p} \int_0^{\infty} y^{2/p-1}.e^{-y} dy = \frac{1}{p}\Gamma(\frac{2}{p}).
\end{aligned}$$

(c) By Theorem 3.2. for continuous function $g(\begin{pmatrix} U \\ V \end{pmatrix}) = s.U + (1-s).V$, $\widetilde{\theta_n} = s.\widehat{\mu_n} + (1-s).\widehat{\theta_n}$, and parts (a) and (b) we have:

$$\begin{aligned}
\lim_{n\to\infty} \sqrt{n}(\widetilde{\theta_n} - \theta) &= \lim_{n\to\infty} \sqrt{n}(\widehat{\mu_n} + (1-s).\widehat{\theta_n} - (s.\theta + (1-s).\theta)) \\
&= \lim_{n\to\infty} [s.(\sqrt{n}(\widehat{\mu_n} - \theta)) + (1-s).(\sqrt{n}(\widehat{\theta_n} - \theta))] \\
&=^d s.N(0, \frac{\Gamma(3/p)}{\Gamma(1/p)}) + (1-s).N(0, (\frac{\Gamma(1/p)}{p})^2) \\
&= s.X(p) + (1-s).Y(p),
\end{aligned}$$

in which $X(p) =^d N(0, \frac{\Gamma(3/p)}{\Gamma(1/p)})$ and $Y(p) =^d N(0, (\frac{\Gamma(1/p)}{p})^2)$, may have non-zero covariance.

(d) Define:

$$\begin{aligned}
Var_p(s) &= Var(s.X(p) + (1-s).Y(p)) \\
&= Var(X(p)).s^2 + Var(Y(p)).(1-s)^2 + 2.Cov(X(p), Y(p)).s.(1-s) \\
&= (Var(X(p)) + Var(Y(p)) - 2.Cov(X(p), Y(p))).s^2 \\
&\quad +2.(Cov(X(p), Y(p)) - Var(Y(p))).s + Var(Y(p)) \\
&= (Var(X(p) - Y(p))).s^2 \\
&\quad +2.Cov(X(p) - Y(p), Y(p)).s + Var(Y(p)) \\
&= a_p.s^2 + b_p.s + c_p : \\
&\quad a_p = Var(X(p) - Y(p)), \\
&\quad b_p = 2.Cov(X(p) - Y(p), Y(p)), \\
&\quad c_p = Var(Y(p)).
\end{aligned}$$

Then:

$$\begin{aligned}
\frac{d}{ds} Var_p(s) = 0 &\Rightarrow s_{min}(p) = \frac{-b_p}{2.a_p} = -\frac{Cov(X(p) - Y(p), Y(p))}{Var(X(p) - Y(p))}, \\
s_{min}(p) = 0 &\Rightarrow Cov(X(p) - Y(p), Y(p)) = 0, \\
s_{min}(p) = 1 &\Rightarrow Cov(X(p) - Y(p), X(p)) = 0.
\end{aligned}$$

□

Problem 4.13. Suppose that $X_1, \cdots$, are i.i.d. non-negative random variables with distribution function F and define the functional parameter

$$\theta(F) = \frac{(\int_0^\infty x dF(x))^2}{\int_0^\infty x^2 dF(x)}.$$

(Note that $\theta(F) = (E(X))^2/E(X^2)$ where $X \sim F$.)
(a) Find the influence curve of $\theta(F)$.
(b) Using $X_1, \cdots, X_n$, find a substitution principle estimator, $\widehat{\theta_n}$, of $\theta(F)$ and find the limiting distribution of $\sqrt{n}(\widehat{\theta_n} - \theta)$.

Solution. (a) Let $\theta_1(F)$, and $\theta_2(F)$ have corresponding influence curves $\phi_1(x;F)$ and $\phi_2(x;F)$, respectively. Then (Exercise !):

$$(i) \quad \phi_{\theta_1*\theta_2}(x;F) = \theta_1(F)*\phi_2(x;F) + \phi_1(x;F)*\theta_2(F),$$
$$(ii) \quad \phi_{\frac{\theta_1}{\theta_2}}(x;F) = \frac{\phi_1(x;F)*\theta_2(F) - \theta_1(F)*\phi_2(x;F)}{\theta_2^2(F)}. \quad (*)$$

By results of page 200 for $h_1(x)=x$ and $h_2(x)=x^2$ we have:

$$\theta(F) = \frac{E^2(X)}{E(X^2)} = \frac{\theta_1^2(F)}{\theta_2(F)} = \theta_1(F)*\frac{\theta_1}{\theta_2}(F). \quad (**)$$

Considering $\phi_1(x;F)=x-\mu_1$ and $\phi_2(x;F)=x^2-\mu_2$, an application of equations in $(*)$ and equation $(**)$ yields:

$$\begin{aligned}
\phi_{\theta_1*(\frac{\theta_1}{\theta_2})}(x;F) &= \phi_{\theta_1}(x;F).\frac{\theta_1(F)}{\theta_2(F)} + \theta_1(F).\phi_{\frac{\theta_1}{\theta_2}}(x;F)\\
&= (x-\mu_1)*\frac{\mu_1}{\mu_2} + \mu_1 * \frac{(x-\mu_1).\mu_2 - \mu_1(x^2-\mu_2)}{\mu_2^2}\\
&= \frac{-\mu_1^2}{\mu_2^2}.x^2 + 2.\frac{\mu_1}{\mu_2}.x - \frac{\mu_1^2}{\mu_2}.
\end{aligned}$$

(b) Let in solution to part (a), $\phi_{\theta_1*(\frac{\theta_1}{\theta_2})}(x;F) = A(\mu_1,\mu_2).x^2 + B(\mu_1,\mu_2).x + C(\mu_1,\mu_2)$ in which $A(\mu_1,\mu_2) = \frac{-\mu_1^2}{\mu_2^2}$, $B(\mu_1,\mu_2) = 2.\frac{\mu_1}{\mu_2}$ and $C(\mu_1,\mu_2) = -\frac{\mu_1^2}{\mu_2}$. then, by the argument on page 200, we have:

$$\sqrt{n}(\theta(\widehat{F_n}) - \theta(F)) \to_d N(0,\sigma^2(F))$$

where:

$$\begin{aligned}
\sigma^2(F) &= \int_{-\infty}^{\infty} \phi^2(x;F)dF(x) = E((A(\mu_1,\mu_2).X^2 + B(\mu_1,\mu_2).X + C(\mu_1,\mu_2))^2)\\
&= E(A(\mu_1,\mu_2)^2.X^4 + B(\mu_1,\mu_2)^2X^2 + C(\mu_1,\mu_2)^2\\
&\quad +2.A(\mu_1,\mu_2).B(\mu_1,\mu_2).X^3 + 2.A(\mu_1,\mu_2).C(\mu_1,\mu_2).X^2 + 2.B(\mu_1,\mu_2).C(\mu_1,\mu_2).X)\\
&= E(A(\mu_1,\mu_2)^2.X^4 + 2.A(\mu_1,\mu_2).B(\mu_1,\mu_2).X^3 + (B(\mu_1,\mu_2)^2 + 2.A(\mu_1,\mu_2).C(\mu_1,\mu_2)).X^2\\
&\quad +2.B(\mu_1,\mu_2).C(\mu_1,\mu_2).X + C(\mu_1,\mu_2)^2)\\
&= A(\mu_1,\mu_2)^2.E(X^4) + 2.A(\mu_1,\mu_2).B(\mu_1,\mu_2).E(X^3) + (B(\mu_1,\mu_2)^2 + 2.A(\mu_1,\mu_2).C(\mu_1,\mu_2)).E(X^2)\\
&\quad +2.B(\mu_1,\mu_2).C(\mu_1,\mu_2).E(X) + C(\mu_1,\mu_2)^2\\
&= A(\mu_1,\mu_2)^2.\mu_4 + 2.A(\mu_1,\mu_2).B(\mu_1,\mu_2).\mu_3 + (B(\mu_1,\mu_2)^2 + 2.A(\mu_1,\mu_2).C(\mu_1,\mu_2)).\mu_2\\
&\quad +2.B(\mu_1,\mu_2).C(\mu_1,\mu_2).\mu_1 + C(\mu_1,\mu_2)^2\\
&= (\frac{-\mu_1^2}{\mu_2^2})^2.\mu_4 + 2.(\frac{-\mu_1^2}{\mu_2^2}).(2.\frac{\mu_1}{\mu_2}).\mu_3 + ((2.\frac{\mu_1}{\mu_2})^2 + 2.(\frac{-\mu_1^2}{\mu_2^2}).(-\frac{\mu_1^2}{\mu_2})).\mu_2\\
&\quad +2.(2.\frac{\mu_1}{\mu_2}).(-\frac{\mu_1^2}{\mu_2}).\mu_1 + (-\frac{\mu_1^2}{\mu_2})^2.
\end{aligned}$$

□

Problem 4.15. Suppose that $X_1,\cdots,X_n$ are i.i.d. Normal random variables with mean 0 and unknown variance σ^2.

(a) Show that $E(|X_i|) = \sigma.\sqrt{2/\pi}$.
(b) Use the result of (a) to construct a method of moments estimator, $\widehat{\sigma_n}$, of σ. Find the limiting distribution of $\sqrt{n}(\widehat{\sigma_n} - \sigma)$.
(c) Another method of moments estimator of σ is:

$$\widetilde{\sigma_n} = (\frac{1}{n}\sum_{i=1}^{n} X_i^2)^{1/2}.$$

Find the limiting distribution of $\sqrt{n}(\widetilde{\sigma_n} - \sigma)$ and compare the results of parts (b) and (c).

Solution. (a)

$$\begin{aligned} E(|X_i|) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} |x|.e^{-x^2/2.\sigma^2}dx = \frac{2}{\sqrt{2\pi}\sigma}\int_{0}^{\infty} x.e^{-x^2/2.\sigma^2}dx \\ &= \frac{2\sigma}{\sqrt{2\pi}\sigma}\int_{0}^{\infty} x.e^{-x^2/2.\sigma^2}d(x^2/2.\sigma^2) = \sigma.\sqrt{2/\pi}.\int_{0}^{\infty} e^{-t}dt = \sigma.\sqrt{2/\pi}. \end{aligned}$$

(b) As $\sigma = \sqrt{\frac{\pi}{2}}.E(|X|) = \sqrt{\frac{\pi}{2}}.\int_{-\infty}^{\infty} |x|dF(x)$, it follows that:

$$\widehat{\sigma_n} = \sqrt{\frac{\pi}{2}}.(\frac{1}{n}.\sum_{i=1}^{n} |x_i|). \quad (n \geq 1)$$

Next, by Theorem 3.8, for $X_i^* = |X_i|, (1 \leq i \leq n)$, $\mu^* = \sigma.\sqrt{\frac{2}{\pi}}$ and $\sigma^{*2} = (1 - \frac{2}{\pi}).\sigma^2$, (Exercise !) we have: $\frac{\sqrt{n}(\overline{|X|}_n - \sigma.\sqrt{2/\pi})}{\sqrt{1-\frac{2}{\pi}}.\sigma} \to_d N(0,1)$, or equivalently:

$$\sqrt{n}(\widehat{\sigma_n} - \sigma) \to_d N(0, \frac{\pi - 2}{2}.\sigma^2). \quad (*)$$

(c)By Theorem 3.8 for $X_i^* = X_i^2, (1 \leq i \leq n)$, $\mu^* = \sigma^2$ and $\sigma^{*2} = 2.\sigma^4$ (Exercise!) we have $\sqrt{n}(\overline{X_n^2} - \sigma^2) \to_d N(0, 2\sigma^4)$. Then, by Theorem 3.4. for $g(x) = \sqrt{x}$, and $g'(x) = \frac{1}{2\sqrt{x}}$, it follows that:

$$\sqrt{n}(\widetilde{\sigma_n} - \sigma) \to_d \frac{1}{2\sigma^2}.N(0, 2.\sigma^4) =^d N(0, \frac{\sigma^2}{2}). \quad (**)$$

Finally, by $(*)$ and $(**)$ we have:

$$\mathtt{ARE}_\sigma(\widetilde{\sigma_n}, \widehat{\sigma_n}) = \frac{\frac{\pi-2}{2}.\sigma^2}{\frac{\sigma^2}{2}} = \pi - 2 > 1.$$

Thus, $\widetilde{\sigma_n}$ is more efficient than $\widehat{\sigma_n}$.
□

Problem 4.17. Let $U_1, \cdots, U_n$ be i.i.d. Uniform random variables on $[0, \theta]$. suppose that only the smallest r values are actually observed, that is the order statistics $U_{(1)} < U_{(2)} < \cdots < U_{(r)}$.
(a) Find the joint density of $U_{(1)}, U_{(2)}, \cdots, U_{(r)}$ and find a one-dimensional sufficient statistics for θ.
(b) Find a unbiased estimator of θ based on the sufficient statistics found in (a).

Solution. (a) Let $X_1, \cdots, X_n$ be i.i.d. continuous random variables with p.d.f $f(x)$ and survival function $S(x)$. Then, for the smallest r values $X_{(1)} < \cdots < X_{(r)}$ we have (David, 1981):

$$f_{X_{(1)},\cdots,X_{(r)}}(x_{(1)}, \cdots, x_{(r)}) = r!.C(n,r).[\prod_{i=1}^{r} f(x_{(i)})].[S(x_{(r)})]^{n-r}.1_{x_{(1)}<\cdots<x_{(r)}}. \quad (*)$$

Consequently, for $f(x) = \frac{1}{\theta}.1_{[0,\theta]}(x)$ and $S(x) = (1 - \frac{x}{\theta}).1_{[0,\theta]}(x)$ it follows from $(*)$ that:

$$\begin{aligned} f_{U_{(1)},\cdots,U_{(r)}}(u_{(1)},\cdots,u_{(r)}) &= r!.C(n,r).[\prod_{i=1}^{r}\frac{1}{\theta}.1_{[0,\theta]}(u_{(i)})][(1-\frac{u_{(r)}}{\theta}).1_{[0,\theta]}(u_{(r)})]^{n-r}.1_{u_{(1)}<\cdots<u_{(r)}} \\ &= [\frac{r!.C(n,r).1_{[0,\theta]}(u_{(r)}).(\theta-u_{(r)})^{n-r}}{\theta^n}]*[1_{u_{(1)}<\cdots<u_{(r)}}] \\ &= g^*(u_{(r)};\theta)*h^*(u_{(1)},\cdots,u_{(r)}), \end{aligned}$$

thus, by Theorem 4.2. $T(U_{(1)},\cdots,U_{(r)}) = U_{(r)}$ is sufficient statistics for θ.

(b) Using Problem 2.25(b), we have $f_{X_{(r)}}(x) = r.C(n,r).F(x)^{r-1}.S(x)^{n-r}.f(x)$, and therefore:

$$\begin{aligned} E(U_{(r)}) &= \int_0^\theta u.r.C(n,r).(\frac{u}{\theta})^{r-1}.(1-\frac{u}{\theta})^{n-r}.\frac{1}{\theta}du = \int_0^\theta \theta.r.C(n,r).(\frac{u}{\theta})^{r}.(1-\frac{u}{\theta})^{n-r}d(\frac{u}{\theta}) \\ &= \int_0^1 \theta.r.C(n,r).x^r.(1-x)^{n-r}dx = \theta.r.C(n,r).\int_0^1 x^{r+1-1}.(1-x)^{n-r+1-1}dx \\ &= \theta.r.C(n,r).B(r+1,n-r+1). \quad (**) \end{aligned}$$

Accordingly, by $(**)$, $\widehat{\theta} = \frac{U_{(r)}}{r.C(n,r).B(r+1,n-r+1)}$ is a unbiased estimator of θ.
□

Problem 4.19. Suppose that $X_1,\cdots,X_n$ are i.i.d. random variables with a continuous distribution function F. It can be shown that $g(t) = E(|X_i - t|)$ (or $g(t) = E(|X_i - t| - |X_i|)$) is minimized at $t = \theta$ where $F(\theta) = \frac{1}{2}$ (see Problem 1.25). This suggests that the median θ can be estimated by choosing $\widehat{\theta_n}$ to minimize

$$g_n(t) = \sum_{i=1}^{n} |X_i - t|.$$

(a) Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the order statistics. Show that if n is even then $g_n(t)$ is minimized for $X_{n/2} \leq t \leq X_{1+n/2}$ while if n is odd then $g_n(t)$ is minimized at $t = X_{(n+1)/2}$.
(b) Let $\widehat{F_n}(x)$ be the empirical distribution function. Show that $\widehat{F_n^{-1}} = X_{(n/2)}$ if n is even and $\widehat{F_n^{-1}} = X_{((n+1)/2)}$ if n is odd.

Solution. (a) As

$$g_n(t) = \sum_{i=1}^{n}|X_{(i)} - t| = \sum_{i=1}^{n}[(2i-n).t - \sum_{j=1}^{i}X_{(j)} + \sum_{j=i+1}^{n}X_{(j)}].1_{[X_{(i)},X_{(i+1)}]}(t), \quad (*)$$

it follows that g_n is a piecewise linear function of t that each linear piece is decreasing for $i < n/2$ and increasing for $i \geq n/2$. Let n be even, then $g_n^{'}(t) = 2i - n = 0$ if and only if $i = n/2$ with condition $X_{n/2} \leq t \leq X_{1+n/2}$, giving $X_{n/2} \leq t \leq X_{1+n/2}$ as the minimizing points for g_n. Next, let n be odd, then using $(*)$:

$$\begin{aligned} g_n(X_{(n+1)/2}) &= -X_{(n+1)/2} - \sum_{j=1}^{(n-1)/2} X_{(j)} + \sum_{j=(n+1)/2}^{n} X_{(j)} \\ &< -X_{(n-1)/2} - \sum_{j=1}^{(n-1)/2} X_{(j)} + \sum_{j=(n+1)/2}^{n} X_{(j)} \\ &= g_n(X_{(n-1)/2}), \end{aligned}$$

implying that $t = X_{(n+1)/2}$ is the minimizing point of g_n.

(b) Referring to page 204, we have $\widehat{F_n^{-1}} = X_{(i)}$ if $\frac{i-1}{n} \le t \le \frac{i}{n}$. Hence:

$$\begin{aligned} n = 2m \quad &: \quad \frac{i-1}{n} < \frac{1}{2} \le \frac{i}{n} \Leftrightarrow m \le i < m+1 \Leftrightarrow i = m = \frac{n}{2} \Rightarrow \widehat{F_n^{-1}} = X_{n/2} \\ n = 2m+1 \quad &: \quad \frac{i-1}{n} < \frac{1}{2} \le \frac{i}{n} \Leftrightarrow \frac{2m+1}{2} \le i < \frac{2m+3}{2} \Leftrightarrow i = \frac{2m+2}{2} = \frac{n+1}{2} \Rightarrow \widehat{F_n^{-1}} = X_{(n+1)/2}. \end{aligned}$$

□

Problem 4.21. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with distribution function. the substitution principle can be extended to estimating functional parameters of the form

$$\theta(F) = E[h(X_1, \cdots, X_k)]$$

where h is some special function. (We assume that this expected value is finite.) If $n \ge k$, a substitution principle estimator of $\theta(F)$ is

$$\widehat{\theta} = \frac{\sum_{i_1 < \cdots < i_k} h(X_{i_1}, \cdots, X_{i_k})}{C(n,k)}$$

where the summation extends over all combinations of k integers drawn from the integer 1 through n. The estimator $\widehat{\theta}$ is called a U-statistics.
(a) Show that $\widehat{\theta}$ is a unbiased estimator of $\theta(F)$.
(b) Suppose $Var(X_i) < \infty$. Show that

$$Var(X_i) = [E((X_1 - X_2)^2)]/2.$$

How does the "U-statistics" substitution principle estimator differ from the substitution principle estimator in Example 4.23?

Solution. (a) As $x_i's$ are i.i.d. it follows that fro any permutation $(i_1, \cdots, i_k)$ of $(1, \cdots, k)$ and any h, we have $E(h(X_{i_1}, \cdots, X_{i_k})) = E(h(X_1, \cdots, X_k)) = \theta(F)$. Thus:

$$E(\widehat{\theta}) = \frac{\sum_{i_1 < \cdots < i_k} E(h(X_{i_1}, \cdots, X_{i_k}))}{C(n,k)} = \frac{\sum_{i_1 < \cdots < i_k} \theta(F)}{C(n,k)} = \frac{C(n,k).\theta(F)}{C(n,k)} = \theta(F).$$

(b) First, let $\mu = E(X_i)$ $(i = 1, 2)$, then:

$$\begin{aligned} E((X_1 - X_2)^2) &= E(((X_1 - \mu) - (X_2 - \mu))^2) = E((X_1 - \mu)^2 - 2(X_1 - \mu)(X_2 - \mu) + (X_2 - \mu)^2) \\ &= Var(X_1) - 2(E(X_1) - \mu).(E(X_2) - \mu) + Var(X_2) = 2.Var(X_i), \end{aligned}$$

implying:

$$Var(X_i) = E(h(X_1, X_2)) : \quad h(X_1, X_2) = \frac{(X_1 - X_2)^2}{2}. \quad (*)$$

Second, by $(*)$ we have:

$$\widehat{\theta(F)}_{\texttt{U-Statistics}} = \frac{\sum_{i_1<i_2} \frac{(X_{i_1}-X_{i_2})^2}{2}}{C(n,2)}. \quad (**)$$

Comparing $(**)$ with

$$\widehat{\theta(F)} = \frac{n-1}{n} S^2 \quad (***)$$

given in Example 4.17 and Example 4.23 we observe that the one given by $(**)$ is a unbiased estimator of $\theta = \sigma^2$, while the other given by $(***)$ is a biased estimator of it.
□

Problem 4.23. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables and define an estimator $\widehat{\theta}$ by

$$\sum_{i=1}^{n} \psi(X_i - \widehat{\theta}) = 0$$

where ψ is an odd function ($\psi(x) = -\psi(-x)$) with derivative ψ'.
(a) Let $\widehat{\theta_{-j}}$ be the estimator computed from all the $X_i's$ except X_j. Show that:

$$\sum_{i=1}^{n} \psi(X_i - \widehat{\theta_{-j}}) = \psi(X_j - \widehat{\theta_{-j}}).$$

(b) Use approximation $\psi(X_i - \widehat{\theta_{-j}}) \approx \psi(X_i - \widehat{\theta}) + (\widehat{\theta} - \widehat{\theta_{-j}}).\psi'(X_i - \widehat{\theta})$ to show that

$$\widehat{\theta_{-j}} \approx \widehat{\theta} - \frac{\psi(X_j - \widehat{\theta})}{\sum_{i=1}^{n} \psi'(X_i - \widehat{\theta})}.$$

(c) Show that the jackknife estimator of $Var(\widehat{\theta})$ can be approximated by:

$$\frac{n-1}{n} \frac{\sum_{i=1}^{n} \psi^2(X_i - \widehat{\theta})}{(\sum_{i=1}^{n} \psi'(X_i - \widehat{\theta}))^2}.$$

Solution. (a) By definition,

$$\sum_{i \neq j} \psi(X_i - \widehat{\theta_{-j}}) = 0. \quad (*)$$

Adding $\psi(X_j - \widehat{\theta_{-j}})$ to both sides of $(*)$ yields the assertion.

(b)As $\psi(X_i - \widehat{\theta_{-j}}) \approx \psi(X_i - \widehat{\theta}) + (\widehat{\theta} - \widehat{\theta_{-j}}).\psi'(X_i - \widehat{\theta})$ $(1 \leq i \leq n)$, it follows that:

$$\sum_{i=1}^{n} \psi(X_i - \widehat{\theta_{-j}}) \approx \sum_{i=1}^{n} \psi(X_i - \widehat{\theta}) + (\widehat{\theta} - \widehat{\theta_{-j}}). \sum_{i=1}^{n} \psi'(X_i - \widehat{\theta}), \quad (**)$$

and, by part(a) and assumption it follows from $(**)$ that:

$$\psi(X_j - \widehat{\theta_{-j}}) \approx 0 + (\widehat{\theta} - \widehat{\theta_{-j}}). \sum_{i=1}^{n} \psi'(X_i - \widehat{\theta}),$$

implying

$$(\widehat{\theta}-\widehat{\theta_{-j}}) \approx \frac{\psi(X_j-\widehat{\theta_{-j}})}{\sum_{i=1}^{n}\psi'(X_i-\widehat{\theta})} \approx \frac{\psi(X_j-\widehat{\theta})}{\sum_{i=1}^{n}\psi'(X_i-\widehat{\theta})},$$

or equivalently the assertion.

(c)By definition on page 222 and result part (b) and given assumption in the problem it follows that: $\theta_{\bullet}=\frac{1}{n}\sum_{i=1}^{n}\widehat{\theta_{-i}} \simeq \theta-\frac{\sum_{i=1}^{n}\psi(X_i-\widehat{\theta})}{\sum_{i=1}^{n}\psi'(X_i-\widehat{\theta})}=\theta$, and; by another application of the result in part (b) we have:

$$\begin{aligned}\widehat{Var(\widehat{\theta})} &= \frac{n-1}{n}\sum_{j=1}^{n}(\widehat{\theta_{-j}}-\widehat{\theta}_{\bullet})^2 \simeq \frac{n-1}{n}\sum_{j=1}^{n}(\widehat{\theta_{-j}}-\widehat{\theta})^2\\ &\simeq \frac{n-1}{n}\sum_{j=1}^{n}\left(\frac{\psi(X_j-\widehat{\theta})}{\sum_{j=1}^{n}\psi'(X_j-\widehat{\theta}))}\right)^2=\frac{n-1}{n}\frac{\sum_{j=1}^{n}\psi^2(X_j-\widehat{\theta})}{(\sum_{j=1}^{n}\psi'(X_j-\widehat{\theta}))^2}.\end{aligned}$$

□

Chapter 5

Likelihood-Based Estimation

Problem 5.1. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with density

$$f(x;\theta_1,\theta_2) = a(\theta_1,\theta_2).h(x) \;\; \texttt{for} \;\; \theta_1 \le x \le \theta_2, \;\; 0, \;\; \texttt{otherwise}$$

where $h(x) > 0$ is a known continuous function defined on the real line.
(a) Show that the MLEs of θ_1 and θ_2 are $X_{(1)}$ and $X_{(n)}$ respectively.
(b) Let $\widehat{\theta_{1n}}$ and $\widehat{\theta_{2n}}$ be the MLEs of θ_1 and θ_2 and suppose that $h(\theta_1) = \lambda_1 > 0$ and $h(\theta_2) = \lambda_2 > 0$. Show that

$$n. \begin{pmatrix} \widehat{\theta_{1n}} - \theta_1 \\ \theta_2 - \widehat{\theta_{2n}} \end{pmatrix} \to_d \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

where Z_1 and Z_2 are independent Exponential random variables with parameters $\lambda_1.a(\theta_1,\theta_2)$ and $\lambda_2.a(\theta_1,\theta_2)$ respectively.

Solution. (a) As $\int_{\theta_1}^{\theta_2} a(\theta_1,\theta_2)h(x)dx = 1$, it follows that: $a(\theta_1,\theta_2) = 1/(\int_{\theta_1}^{\theta_2} h(x)dx)$. Consequently, substituting it in the following likelihood equation it follows that:

$$\begin{aligned}
L(\theta_1,\theta_2|\mathbf{x}) &= \prod_{i=1}^{n} f(x_i;(\theta_1,\theta_2)) \\
&= \prod_{i=1}^{n} (a(\theta_1,\theta_2) * h(x_i) * 1_{(-\infty,x_i]}(\theta_1) * 1_{[x_i,+\infty)}(\theta_2)) \\
&= a(\theta_1,\theta_2)^n * (\prod_{i=1}^{n} h(x_i)) * 1_{(-\infty,x_{(1)}]}(\theta_1) * 1_{[x_{(n)},+\infty}(\theta_2) \\
&= (\frac{1}{\int_{\theta_1}^{\theta_2} h(x)dx})^n * (\prod_{i=1}^{n} h(x_i))) * 1_{(-\infty,x_{(1)}]}(\theta_1) * 1_{[x_{(n)},+\infty)}(\theta_2). \quad (*)
\end{aligned}$$

Accordingly, by $(*)$ we have:

$$\begin{aligned}
\theta_2 \; fixed &: \; (\theta_1 \uparrow \propto a(\theta_1,\theta_2) \uparrow \propto L(\theta_1,\theta_2) \uparrow) \Rightarrow MLE(\theta_1) = X_{(1)}, \\
\theta_1 \; fixed &: \; (\theta_2 \downarrow \propto a(\theta_1,\theta_2) \uparrow \propto L(\theta_1,\theta_2) \uparrow) \Rightarrow MLE(\theta_2) = X_{(n)}.
\end{aligned}$$

(b)Let $u(n) = u_n(x,\theta)$ be a differentiable function of n such that $\lim_{n\to\infty} u_n = 0$. Then(Exercise !),

$$\lim_{n\to\infty} (1+u_n)^n = \exp(\lim_{n\to\infty} \frac{\frac{d}{dn}u(n)}{\frac{-1}{n^2}}). \quad (*)$$

Accordingly, three times usage of $(*)$ yields:

$$\begin{aligned}
\lim_{n\to\infty}(S_X(\theta_1+\frac{x}{n}))^n &= \lim_{n\to\infty}(1-\int_{\theta_1}^{\theta_1+\frac{x}{n}} f(t;\theta_1,\theta_2)dt)^n \\
&= \exp(\lim_{n\to\infty}\frac{-a(\theta_1,\theta_2)(\frac{-x}{n^2})h(\theta_1+\frac{x}{n})}{\frac{-1}{n^2}}) \\
&= \exp(-a(\theta_1,\theta_2).h(\theta_1).x) \\
\lim_{n\to\infty}(F_X(\theta_2-\frac{y}{n}))^n &= \lim_{n\to\infty}(1+\int_{\theta_1}^{\theta_2-\frac{y}{n}} f(t;\theta_1,\theta_2)dt-1)^n \\
&= \exp(\lim_{n\to\infty}\frac{a(\theta_1,\theta_2)(\frac{y}{n^2})h(\theta_2-\frac{y}{n})/(\int_{\theta_1}^{\theta_2-y/n}a(\theta_1,\theta_2)h(t)dt)}{\frac{-1}{n^2}}) \\
&= \exp(-a(\theta_1,\theta_2).h(\theta_2).y) \\
\lim_{n\to\infty}(F_X(\theta_2-\frac{y}{n})-F_X(\theta_1+\frac{x}{n}))^n &= \exp(-a(\theta_1,\theta_2).h(\theta_1).x-a(\theta_1,\theta_2).h(\theta_2).y),\ (Exercise!). \quad (**)
\end{aligned}$$

Next, considering $P(A^c\cap B^c)=1-(P(A)+P(B)-P(A\cap B))$, from $(**)$ it follows that:

$$\begin{aligned}
&\lim_{n\to\infty} F_{n(\widehat{\theta_{1n}}-\theta_1),n(\theta_2-\widehat{\theta_{n2}})}(x,y) = \\
&\lim_{n\to\infty} P(n(\widehat{\theta_{1n}}-\theta_1)\le x, n(\theta_2-\widehat{\theta_{n2}})\le y) = \\
&\lim_{n\to\infty} P(\widehat{\theta_{1n}}\le\theta_1+\frac{x}{n},\theta_2-\frac{y}{n}\le\widehat{\theta_{n2}}) = \\
&\lim_{n\to\infty} 1-[P(\widehat{\theta_{1n}}\ge\theta_1+\frac{x}{n})+P(\theta_2-\frac{y}{n}\ge\widehat{\theta_{n2}})-P(\widehat{\theta_{1n}}\ge\theta_1+\frac{x}{n},\theta_2-\frac{y}{n}\ge\widehat{\theta_{n2}}] = \\
&\lim_{n\to\infty} 1-[\prod_{i=1}^n P(X_i\ge\theta_1+\frac{x}{n})+\prod_{i=1}^n P(X_i\le\theta_2-\frac{y}{n})-\prod_{i=1}^n P(\theta_1+\frac{x}{n}\ge X_i\ge\theta_2-\frac{y}{n})] = \\
&\lim_{n\to\infty} 1-[(S_X(\theta_1+\frac{x}{n}))^n)+(F_X(\theta_2-\frac{y}{n}))^n)-(F_X(\theta_2-\frac{y}{n})-F_X(\theta_1+\frac{x}{n}))^n] = \\
&1-[\exp(-a(\theta_1,\theta_2).h(\theta_1).x)+\exp(-a(\theta_1,\theta_2).h(\theta_2).y)-\exp(-a(\theta_1,\theta_2).h(\theta_1).x-a(\theta_1,\theta_2).h(\theta_2).y)] = \\
&(1-\exp(-a(\theta_1,\theta_2).h(\theta_1).x))*(1-\exp(-a(\theta_1,\theta_2).h(\theta_2).y)) = \\
&F_{Z_1}(x)*F_{Z_2}(y),\ \ \texttt{for all}\ x,y.
\end{aligned}$$

□

Problem 5.3. Suppose that $X_1,\cdots,X_n,Y_1,\cdots,Y_n$ are independent Exponential random variables where the density of X_i is $f_i(x)=\lambda_i\theta.exp(-\lambda_i.\theta x)$ for $x\ge0$ and the density of Y_i is $g_i(x)=\lambda_i.exp(-\lambda_i x)$ for $x\ge0$ where $\lambda_1,\cdots,\lambda_n$ and θ are unknown parameters.
(a) Show that the MLE of θ (based on $X_1,\cdots,X_n,Y_1,\cdots,Y_n$) satisfies the equation

$$\frac{n}{\widehat{\theta}}-2\sum_{i=1}^n\frac{R_i}{1+\widehat{\theta}R_i}=0$$

where $R_i=X_i/Y_i$.
(b) Show that the density of R_i is

$$f_R(x;\theta)=\theta(1+\theta.x)^{-2}\ \ \texttt{for}\ x\ge0,$$

and show that the MLE for θ based on $R_1,\cdots,R_n$ is the same as that given in part (a).
(c) Let $\widehat{\theta_n}$ be the MLE in part (a). Find the limiting distribution of $\sqrt{n}(\widehat{\theta_n}-\theta)$.

(d) Use the data for $(X_i, Y_i), i = 1, \cdots, 20$ given in Table 5.7 to compute the maximum likelihood estimate of θ using either the Newton-Raphson or Fisher scoring algorithm. Find an approximate starting value for the iterations and justify your choice.

Table 5.7 Data for Problem 5.3.

x	y	x	y	x	y	x	y
0.7	3.8	20.2	2.8	1.1	2.8	15.2	8.8
11.3	4.6	0.3	1.9	1.9	3.2	0.2	7.6
2.1	2.1	0.9	1.4	0.5	8.5	0.7	1.3
30.7	5.6	0.7	0.4	0.8	14.5	0.4	2.2
4.6	10.3	2.3	0.9	1.2	14.4	2.3	4.0

(e) Give an estimate of the standard error for the maximum likelihood estimate computed in part (c).

Solution. (a) As

$$\begin{aligned} L(\theta, \lambda_1, \cdots, \lambda_n) &= f(\mathbf{x}, \mathbf{y}; \theta, \lambda_1, \cdots, \lambda_n) = (\prod_{i=1}^{n} f_{X_i}(x_i; \theta)) * (\prod_{i=1}^{n} f_{Y_i}(y_i; \theta)) \\ &= (\prod_{i=1}^{n} (\lambda_i.\theta.e^{-\lambda_i.\theta.x_i})) * (\prod_{i=1}^{n} (\lambda_i.e^{-\lambda_i.y_i})) = (\prod_{i=1}^{n} \lambda_i)^2.\theta^n.e^{-\sum_{i=1}^{n} \lambda_i.x_i.\theta - \sum_{i=1}^{n} \lambda_i.y_i}, \end{aligned}$$

it follows that:

$$\log(L(\theta, \lambda_1, \cdots, \lambda_n)) = 2.\sum_{i=1}^{n} \log(\lambda_i) + n.\log(\theta) - (\sum_{i=1}^{n} \lambda_i.x_i).\theta - \sum_{i=1}^{n} \lambda_i.y_i. \quad (*)$$

Consequently, by $(*)$ it follows that $\frac{d\log(L(\theta,\lambda_1,\cdots,\lambda_n))}{d\lambda_i} = \frac{2}{\lambda_i} - x_i.\theta - y_i = 0, \quad (1 \le i \le n)$, or equivalently,

$$\widehat{\lambda_i} = \frac{2}{x_i.\theta + y_i}, \quad (1 \le i \le n). \quad (**)$$

Finally, another usage of $(*)$ and substituting $(**)$ in the equation yields:

$$\begin{aligned} 0 &= \frac{d\log(L(\theta, \lambda_1, \cdots, \lambda_n))}{d\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} \widehat{\lambda}_i.x_i \\ &= \frac{n}{\theta} - \sum_{i=1}^{n} \frac{2.x_i}{x_i.\theta + y_i} = \frac{n}{\theta} - \sum_{i=1}^{n} \frac{2.(x_i/y_i)}{\theta.(x_i/y_i) + 1} \\ &= \frac{n}{\widehat{\theta}} - \sum_{i=1}^{n} \frac{2.R_i}{\widehat{\theta}.R_i + 1}. \end{aligned}$$

(b) First:

$$\begin{aligned} f_R(r;\theta) &= \frac{d}{dr}F_R(r;\theta) = \frac{d}{dr}(P(X \le r.Y)) = \frac{d}{dr}(\int_0^\infty \int_0^{ry} \lambda.\theta.e^{-\lambda.\theta.x}.\lambda.e^{-\lambda.x}dxdy) \\ &= \frac{d}{dr}(\int_0^\infty \lambda.e^{-\lambda.y}(1 - e^{-\lambda.\theta.r.y})dy = \frac{d}{dr}(1 - \frac{\lambda}{\lambda + \lambda.\theta.r}) = \theta.(1 + \theta.r)^{-2} \ \texttt{for} \ r \ge 0. \quad (***) \end{aligned}$$

Second, using $(***)$ it follows that:

$$L(\theta;\mathbf{r}) = \prod_{i=1}^{n}(f_R(r_i;\theta)) = \prod_{i=1}^{n}(\theta.(1+\theta.r_i)^{-2}) = \theta^n.[\prod_{i=1}^{n}(1+\theta.r_i)]^{-2}. \quad (\dagger)$$

Accordingly, it follows from $(\dagger)$ that:

$$\begin{aligned} 0 &= \frac{d}{d\theta}\log(L(\theta;\mathbf{r})) = \frac{d}{d\theta}[n.\log(\theta) - 2\sum_{i=1}^{n}\log(1+\theta.r_i)] \\ &= \frac{n}{\widehat{\theta}} - 2\sum_{i=1}^{n}(\frac{r_i}{1+\widehat{\theta}.r_i}). \end{aligned}$$

(c) All conditions A1-A6 page 245 are satisfied (Exercise !). Next, by Theorem 5.3 we have:

$$\sqrt{n}(\widehat{\theta_n}-\theta) \to_d N(0, \frac{I(\theta)}{J^2(\theta)}): \quad I(\theta) = Var_\theta(l'(x;\theta)),\ J(\theta) = -E_\theta(l''(x;\theta)),$$

in which

$$\begin{aligned} l(x;\theta) &= \log(f_R(x;\theta)) = \log(\theta) - 2.\log(1+\theta.x) \\ l'(x;\theta) &= \frac{1}{\theta} - \frac{2x}{1+\theta.x} : E_\theta(l'(x;\theta)) = 0, \\ l''(x;\theta) &= \frac{-1}{\theta^2} + \frac{2x^2}{(1+\theta.x)^2}, \\ I(\theta) &= E_\theta((l'(x;\theta))^2) = \int_0^\infty (\frac{1-\theta.x}{\theta.(1+\theta.x)})^2.\frac{\theta}{(1+\theta.x)^2}dx = \frac{1}{\theta^2}\int_0^\infty \frac{(1-y)^2}{(1+y)^4}dy = \frac{1}{3\theta^2}, \\ J(\theta) &= -\int_0^\infty (\frac{2(\theta.x)^2-(1+\theta.x)^2}{(1+\theta.x)^2.\theta^2})(\frac{\theta}{(1+\theta.x)^2})dx = \frac{-1}{\theta^2}.\int_0^\infty (\frac{2y^2-(1+y)^2}{(1+y)^4})dy = \frac{-1}{\theta^2}(\frac{-1}{3}) = \frac{1}{3.\theta^2}. \end{aligned}$$

Accordingly:

$$\sqrt{n}(\widehat{\theta_n}-\theta) \to_d N(0, 3\theta^2).$$

(d) Using data in Table 5.7 we may calculate $R_i = X_i/Y_i \ (1 \le i \le 20)$, in which:

Calculated Data for Problem 5.3.											
x	y	r	x	y	r	x	y	r	x	y	r
0.7	3.8	0.184	20.2	2.8	7.214	1.1	2.8	0.393	15.2	8.8	1.727
11.3	4.6	2.457	0.3	1.9	0.158	1.9	3.2	0.594	0.2	7.6	0.026
2.1	2.1	1.000	0.9	1.4	0.643	0.5	8.5	0.059	0.7	1.3	0.538
30.7	5.6	5.482	0.7	0.4	1.750	0.8	14.5	0.055	0.4	2.2	0.182
4.6	10.3	0.447	2.3	0.9	2.556	1.2	14.4	0.083	2.3	4.0	0.575

Next, plotting $S(\theta) = \sum_{i=1}^{20}(\frac{1-\theta.r_i}{\theta+r_i.\theta^2})$, in which $\{r_i\}_{i=1}^{20}$ are given by above table we have:

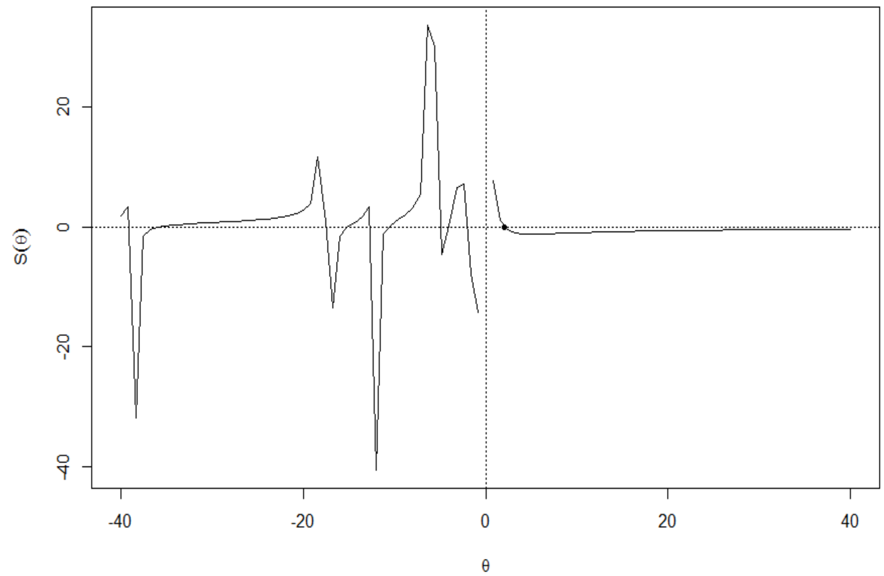

Figure 5.1 Plot of function $S(\theta) = \sum_{i=1}^{20}(\frac{1-\theta.r_i}{\theta+r_i.\theta^2})$

This suggests to take starting value $\theta_0 > 0$. On the other hand, $H(\theta) = \frac{-d}{d\theta}S(\theta) = \sum_{i=1}^{20}\frac{-r_i^2.\theta^2+2.r_i.\theta+1}{(\theta+r_i.\theta^2)^2}$. Thus, the Newton-Raphson algorithm (page 270) takes the following form:

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} + \frac{S(\widehat{\theta}^{(k)})}{H(\widehat{\theta}^{(k)})} = \widehat{\theta}^{(k)} + \frac{\sum_{i=1}^{20}(\frac{1-\widehat{\theta}^{(k)}.r_i}{\widehat{\theta}^{(k)}+r_i.(\widehat{\theta}^{(k)})^2})}{\sum_{i=1}^{20}(\frac{-r_i.(\widehat{\theta}^{(k)})^2+2.r_i.\widehat{\theta}^{(k)}+1}{\widehat{\theta}^{(k)}+r_i.(\widehat{\theta}^{(k)})^2)^2})} \quad (k \geq 0). \quad (\dagger\dagger)$$

Finally, using R software and $\theta_0 = 2.9000$ in $(\dagger\dagger)$ we have:

$$\theta_1 = 1.2298,\ \theta_2 = 1.6431,\ \theta_3 = 2.0184,\ \theta_4 = 2.0246,\ \theta_5 = 2.0246.$$

(e)As:

$$\widehat{s.e.}(\widehat{\theta_n}) = \frac{1}{\sqrt{n.I(\widehat{\theta_n})}} = \frac{1}{\sqrt{n.(\frac{1}{3.\widehat{\theta_n}^2})}} = \sqrt{\frac{3}{n}}.|\widehat{\theta_n}|, \quad (n \geq 1)$$

it follows that $\widehat{s.e.}(\widehat{\theta_5}) = \sqrt{3/5} * 2.0246 = 1.5683$.
□

Problem 5.5. Suppose that $X_1, \cdots, X_n$ are i.i.d. discrete random variables with frequency function

$$f(x;\theta) = \theta, \ \texttt{ for } \ x = -1, \ \ (1-\theta)^2.\theta^x \ \texttt{ for } \ x = 0, 1, 2, \cdots$$

where $0 < \theta < 1$.
(a) Show that the MLE of θ based on $X_1, \cdots, X_n$ is

$$\widehat{\theta_n} = \frac{2\sum_{i=1}^n I(X_i = -1) + \sum_{i=1}^n X_i}{2n + \sum_{i=1}^n X_i}$$

and show that $\{\widehat{\theta_n}\}$ is consistent for θ.
(b) Show that $\sqrt{n}(\widehat{\theta_n} - \theta) \to_d N(0, \sigma^2(\theta))$ and find the value of $\sigma^2(\theta)$.

Solution. (a)
$0 =$
$\frac{d}{d\theta}\log(L(\theta;\mathbf{x})) =$
$\frac{d}{d\theta}\log(\prod_{i=1}^n \log(f(x_i;\theta))) =$
$\sum_{i=1}^n(\frac{d}{d\theta}\log(f(x_i;\theta))) =$
$\sum_{i=1}^n(\frac{d}{d\theta}\log(\theta.1_{x_i=-1}+(1-\theta)^2.\theta^{x_i}.1_{x_i\geq 0})) =$
$\sum_{i=1}^n(\frac{1}{\theta}.1_{x_i=-1}+(\frac{2}{\theta-1}+\frac{x_i}{\theta})1_{x_i=-1}) =$
$(\frac{1}{\theta}).(\sum_{i=1}^n 1_{x_i=-1})+(\frac{2}{\theta-1}).(\sum_{i=1}^n 1_{x_i\geq 0}))+(\frac{1}{\theta}).(\sum_{i=1}^n(x_i.1_{x_i\geq 0})) =$
$(\frac{1}{\theta}).(\sum_{i=1}^n 1_{x_i=-1})+(\frac{2}{\theta-1}).(\sum_{i=1}^n(1-1_{x_i=-1}))+\frac{1}{\theta}\sum_{i=1}^n(x_i.(1-1_{x_i=-1})) =$
$\frac{1}{\theta}[\sum_{i=1}^n 1_{x_i=-1}+(2+\frac{2}{\theta-1})\sum_{i=1}^n(1-1_{x_i=-1})+\sum_{i=1}^n x_i(1-1_{x_i=-1})]$
$\Rightarrow$

$$2+\frac{2}{\widehat{\theta}-1} = -\frac{\sum_{i=1}^n 1_{x_i=-1}+\sum_{i=1}^n x_i.(1-1_{x_i=-1})}{\sum_{i=1}^n(1-1_{x_i=-1})}$$

or,

$$\frac{\widehat{\theta}-1}{2} = \frac{\sum_{i=1}^n(1-1_{X_i=-1})}{-\sum_{i=1}^n 1_{x_i=-1}-\sum_{i=1}^n x_i.(1-1_{x_i=-1})-2\sum_{i=1}^n(1-1_{x_i=-1})}$$

or,

$$\widehat{\theta} = \frac{2\sum_{i=1}^n 1_{x_i=-1}-\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i+2n}.$$

Next, as $E(1_{X=-1}) = P(X=-1) = \theta$, $\widehat{\theta_n} = \frac{2\sum_{i=1}^n(\frac{1_{X_i=-1}}{n})+\frac{\sum_{i=1}^n X_i}{n}}{2+\frac{\sum_{i=1}^n X_i}{n}}$, $\sum_{i=1}^n \frac{1_{X_i=-1}}{n} \to_p \theta$, $\sum_{i=1}^n \frac{X_i}{n} \to_p 0$, (by Theorem 3.6), for $X_n^* = (U_n, V_n) = (\sum_{i=1}^n \frac{1_{X_i=-1}}{n}, \sum_{i=1}^n \frac{X_i}{n})$ and $g(X^*) = g(U,V) = \frac{2U+V}{2+V}$ an application of Theorem 3.2 it follows that:

$$\widehat{\theta_n} = g(U_n, V_n) \to_p g(\theta, 0) = \theta.$$

(b) One may easily check that the conditions A1-A6 hold (Exercise !). Thus, by Theorem 5.3, it follows that $\sqrt{n}(\widehat{\theta_n}-\theta) \to_d N(0, \frac{I(\theta)}{J^2(\theta)})$.
Next, let $f(\mathbf{x};\theta)$ satisfies

$$\frac{d}{d\theta}(E_\theta(\frac{d}{d\theta}\log(f(\mathbf{x};\theta))) = \sum_{\mathbf{x}}\frac{d}{d\theta}[(\frac{d}{d\theta}\log(f(\mathbf{x};\theta))).f(\mathbf{x};\theta)],$$

then (Exercise!):

$$E_\theta((\frac{d}{d\theta}\log(f(\mathbf{x};\theta)))^2) = -E_\theta(\frac{d^2}{d\theta^2}\log(f(\mathbf{x};\theta))). \quad (*)$$

Consequently, as $E_\theta(l'(\theta)) = 0$, and the required condition for $(*)$ holds (0 in both sides), an application of $(*)$ yields:

$$I(\theta) = Var_\theta(l'(\theta)) = E_\theta((l'(\theta))^2) =^* -E_\theta(l''(\theta)) = J(\theta). \quad (**)$$

But,

$$\begin{aligned} J(\theta) &= -E_\theta(l''(\theta)) \\ &= -E_\theta((\frac{-2}{\theta^2}+\frac{2}{(1-\theta)^2}).1_{X=-1}-\frac{2}{(1-\theta)^2}-\frac{1}{\theta^2}.X) \\ &= -[(\frac{-2}{\theta^2}+\frac{2}{(1-\theta)^2}).\theta-\frac{2}{(1-\theta)^2}-\frac{1}{\theta^2}.0] \\ &= \frac{2}{\theta.(1-\theta)}. \quad (***) \end{aligned}$$

Finally, by $(**)$ and $(***)$ it follows that $\sigma^2(\theta) = \frac{\theta.(1-\theta)}{2}$, and:

$$\sqrt{n}(\widehat{\theta_n} - \theta) \to_d N(0, \frac{\theta.(1-\theta)}{2}).$$

$\square$

Problem 5.7. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ has a $k-$ parameter exponential family distribution with joint density or frequency function:

$$f(\mathbf{x}; \theta) = \exp[\sum_{i=1}^{k} c_i(\theta) T_i(x) - d(\theta) + S(x)]$$

where the parameter space Θ is an open subset of R^k and the function $\mathbf{c} = (c_1, \cdots, c_k)$ is one-to-one on Θ.

(a) Suppose that $E_\theta[T_i(\mathbf{X})] = b_i(\theta) \quad (i = 1, \cdots, k)$. Show that the MLE $\widehat{\theta}$ satisfies the equations

$$T_i(\mathbf{X}) = b_i(\widehat{\theta}) \quad (i = 1, \cdots, k).$$

(b) Suppose that the $X_i's$ are also i.i.d. so that $T_i(\mathbf{X})$ can be taken to be an average of i.i.d. random variables. If $\widehat{\theta_n}$ is the MLE, use the Delta Method to show that $\sqrt{n}(\widehat{\theta_n} - \theta)$ has the limiting distribution given in Theorem 5.4.

Solution. (a) Note that $c : \Theta \to R^k$ for $\theta = (\theta_1, \cdots, \theta_k)$ has the form

$$c(\theta_1, \cdots, \theta_k) = (c_1(\theta_1, \cdots, \theta_k), \cdots, c_k(\theta_1, \cdots, \theta_k))$$

and the matrix $(\frac{dc_i}{d\theta_j})_{i,j=1}^{n}$ is invertible. First, given $(l'(\mathbf{X}; \widehat{\theta_n}))_{1\times k} = 0_{1\times k}$ in which $(l'(\mathbf{X}; \widehat{\theta_n}))_{1\times k} = (\frac{dl(\mathbf{X};\widehat{\theta_n})}{d\theta_1}, \cdots, \frac{dl(\mathbf{X};\widehat{\theta_n})}{d\theta_k})$, we have:

$$\frac{dl(\mathbf{X}; \widehat{\theta_n})}{d\theta_j} = 0. \quad (1 \leq j \leq k) \quad (*)$$

But,

$$\frac{dl(\mathbf{X}; \widehat{\theta_n})}{d\theta_j} = \sum_{i=1}^{k} \frac{dc_i(\theta)}{d\theta_j}.T_i(\mathbf{X}) - \frac{dd(\theta)}{d\theta_j}. \quad (1 \leq j \leq k) \quad (**)$$

Thus, by $(*)$ and $(**)$ it follows:

$$\sum_{i=1}^{k} \frac{dc_i(\theta)}{d\theta_j}.T_i(\mathbf{X}) = \frac{dd(\theta)}{d\theta_j}. \quad (1 \leq j \leq k) \quad (***)$$

Second, taking expectation from both sides of $(***)$ and using the given assumption $E_\theta[T_i(\mathbf{X})] = b_i(\theta) \quad (i = 1, \cdots, k)$, we have:

$$\sum_{i=1}^{k} \frac{dc_i(\theta)}{d\theta_j}.b_i(\widehat{\theta}) = \frac{dd(\widehat{\theta})}{d\theta_j}. \quad (1 \leq j \leq k) \quad (****)$$

Third, a side by side subtraction from equations $(***)$ and $(****)$ implies:

$$\sum_{i=1}^{k} \frac{dc_i(\widehat{\theta})}{d\theta_j}.(T_i(\mathbf{X}) - b_i(\widehat{\theta})) = 0, \quad (1 \leq j \leq k)$$

or equivalently:

$$(\frac{dc_i(\widehat{\theta})}{d\theta_j})_{i,j=1}^k \times (T_i(\mathbf{X}) - b_i(\widehat{\theta}))_{1\times k}^t = 0_{k\times 1}. \quad (\dagger)$$

Finally, as the matrix $(\frac{dc_i(\widehat{\theta})}{d\theta_j})_{i,j=1}^k$ is invertible the only solution for $(\dagger)$ is $(T_i(\mathbf{X}) - b_i(\widehat{\theta}))_{1\times k}^t = 0_{1\times k}^t$, proving the assertion.

□

Problem 5.9. Let $X_1, \cdots, X_n$ be i.i.d. Exponential random variables with parameter λ. Suppose that the $X_i's$ are not observed exactly but rather we observe random variables $Y_1, \cdots, Y_n$ where $Y_i = k.\delta \texttt{ if } k\delta \leq X_i < (k+1)\delta$ for $k = 0, 1, 2, \cdots$ where $\delta > 0$ is known.
(a) Give the joint frequency function of $\mathbf{Y} = (Y_1, \cdots, Y_n)$ and show that $\sum_{i=1}^n Y_i$ is sufficient for λ.
(b) Find the MLE of λ based on $Y_1, \cdots, Y_n$.
(c) Let $\widehat{\lambda_n}$ be the MLE of λ in part (b). Show that

$$\sqrt{n}(\widehat{\lambda_n} - \lambda) \to_d N(0, \sigma^2(\lambda, \delta))$$

where $\sigma^2(\lambda, \delta) \to \lambda^2$ as $\delta \to 0$.

Solution. (a)

$$\begin{aligned}
P_{Y_1,\cdots,Y_n}(y_1, \cdots, y_n) &= P_{X_1,\cdots,X_n}(y_1 \leq X_1 < y_1 + \delta, \cdots, y_n \leq X_n < y_n + \delta) \\
&= \prod_{i=1}^n (P_{X_i}(y_i \leq X_i < y_i + \delta)) \\
&= \prod_{i=1}^n (\int_{y_i}^{y_i+\delta} \lambda.e^{-\lambda.t} dt) = \prod_{i=1}^n (e^{-\lambda.y_i}(1 - e^{-\lambda.\delta})) \\
&= (e^{-\lambda.\sum_{i=1}^n Y_i}(1 - e^{-\lambda.\delta})^n) * (1) = g^*(T(\mathbf{y}); \lambda) * h^*(\mathbf{y}).
\end{aligned}$$

Thus, by Theorem 4.2, $T(\mathbf{y}) = \sum_{i=1}^n Y_i$ is sufficient statistics for λ.

(b) As $l_n(\lambda) = \log(L(\lambda; y_1, \cdots, y_n)) = -\lambda.\sum_{i=1}^n Y_i + n.\log(1 - e^{-\lambda.\delta})$, it follows that:

$$0 = -\sum_{i=1}^n Y_i + n.\frac{\delta.e^{-\lambda.\delta}}{1 - e^{-\lambda.\delta}} \Rightarrow \widehat{\lambda_n} = \frac{1}{\delta}.\log(1 + \frac{\delta}{\overline{Y_n}}).$$

(c) By Theorem 5.3, $\sqrt{n}(\widehat{\lambda_n} - \lambda) \to_d N(0, \frac{I(\lambda)}{J^2(\lambda)})$. Next, using equalities

$$\sum_{n=0}^\infty x^n = \frac{1}{1-x}, \quad \sum_{n=0}^\infty n.x^n = \frac{x}{(1-x)^2}, \quad \sum_{n=0}^\infty n^2.x^n = \frac{x}{(1-x)^2} + \frac{2x^2}{(1-x)^3}, \quad |x| < 1,$$

we have:

$$\begin{aligned}
l(\lambda) &= \log(P(Y = k.\delta;\lambda)) = \log(e^{-\lambda.\delta.k}.(1-e^{-\lambda.\delta})) = -\lambda.\delta.k + \log(1-e^{-\lambda.\delta}) \quad (k \geq 0),\\
l^{'}(\lambda) &= -\delta.k + \frac{\delta.e^{-\lambda.\delta}}{1-e^{-\lambda.\delta}},\\
l^{''}(\lambda) &= \frac{-\delta^2.e^{-\lambda.\delta}}{(1-e^{-\lambda.\delta})^2},\\
E_\lambda(l^{'}(\lambda)) &= 0,\\
I(\lambda) &= Var_\lambda(l^{'}(\lambda)) = E_\lambda((l^{'}(\lambda))^2) = \frac{\delta^2.e^{\lambda.\delta}}{(e^{\lambda.\delta}-1)^2},\\
J(\lambda) &= -E_\lambda(l^{''}(\lambda)) = \frac{\delta^2.e^{-\lambda.\delta}}{(1-e^{-\lambda.\delta})^2} = \frac{\delta^2.e^{\lambda.\delta}}{(e^{\lambda.\delta}-1)^2},
\end{aligned}$$

implying: $\sigma^2(\lambda,\delta) = \frac{I(\lambda)}{J^2(\lambda)} = \frac{1}{I(\lambda)} = \frac{(e^{\lambda.\delta}-1)^2}{\delta^2.e^{\lambda.\delta}}$. Finally:

$$\lim_{\delta\to 0}\sigma^2(\lambda,\delta) = \lim_{\delta\to 0}[\lambda^2.(\frac{e^{\lambda.\delta}-1}{\lambda.\delta})^2.\frac{1}{e^{\lambda.\delta}}] = \lambda^2.(\frac{d}{dx}e^x|_{x=0})^2.1 = \lambda^2.$$

□

Problem 5.11. The key condition in Theorem 5.3. is (A6) as this allows us to approximate he likelihood equation by a linear equation in $\sqrt{n}(\widehat{\theta_n}-\theta)$. However, condition (A6) can be replaced by other similar conditions, some of which may be weaker than (A6).
Assume that $\widehat{\theta_n} \to_p \theta$ and that conditions (A1)-(A5) hold. Suppose that for some $\delta > 0$, there exists a function $K_\delta(x)$ and a constant $\alpha > 0$ such that:

$$|l^{(3)}(x;t) - l^{(3)}(x;\theta)| \leq K_\delta(x)|t-\theta|^\alpha$$

for $|t-\theta| \leq \delta$ where $E_\theta[K_\delta(X_i)] < \infty$. Show that the conclusion of Theorem 5.3. holds.

Solution. The given condition implies that $|l^{(3)}(x;t)| \leq |l^{(3)}(x;\theta)| + K_\delta(x).|t-\theta|^\alpha$, for all $|t-\theta| < \delta$. Next, returning to the proof of Theorem 5.3.(page 253) for any $0 < \delta^* < \delta, \theta < \theta_n^* < \widehat{\theta_n}$, and $|\widehat{\theta_n}-\theta| < \delta^* < \delta$ we have:

$$\begin{aligned}
|R_n| &= |(\widehat{\theta_n}-\theta).\frac{1}{2n}.\sum_{i=1}^n l^{(3)}(X_i;\theta_n^*)|\\
&\leq \frac{\delta^*}{2n}.\sum_{i=1}^n |l^{(3)}(X_i;\theta_n^*)|\\
&\leq \frac{\delta^*}{2n}.\sum_{i=1}^n [|l^{(3)}(x;\theta)| + K_\delta(x).|t-\theta|^\alpha]\\
&\leq \frac{\delta^*}{2n}.[n.|l^{(3)}(x;\theta)| + \sum_{i=1}^n |K_\delta(x)|.(\delta^*)^\alpha]\\
&= \frac{\delta^*}{2}.|l^{(3)}(x;\theta)| + \frac{(\delta^*)^{\alpha+1}}{2}.\frac{\sum_{i=1}^n |K_\delta(X_i)|}{n}. \quad (*)
\end{aligned}$$

Next, by Theorem 3.6.

$$\frac{\sum_{i=1}^n |K_\delta(X_i)|}{n} \to_p E_\theta(|K_\delta(X)|) < \infty. \quad (**)$$

Now, for given $\epsilon > 0$, by $(*)$ and $(**)$ there is sufficiently small $\delta^* > 0$ and $N_1 \geq 1$ such that:

$$P(|R_n| > \epsilon, |\widehat{\theta_n} - \theta| < \delta^*)\frac{\epsilon}{2}. \quad (n \geq N_1) \quad (***)$$

In addition, there is $N_2 \geq 1$ such that:

$$P(|R_n| > \epsilon, |\widehat{\theta_n} - \theta| > \delta^*) \leq P(|\widehat{\theta_n} - \theta| > \delta^*)) \leq \frac{\epsilon}{2}. \quad (n \geq N_2) \quad (****)$$

Take $N = \max(N_1, N_2)$, then by $(***)$ and $(****)$:

$$P(|R_n| > \epsilon) = P(|R_n| > \epsilon, |\widehat{\theta_n} - \theta| < \delta^*) + P(|R_n| > \epsilon, |\widehat{\theta_n} - \theta| > \delta^*) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \quad (n \geq N)$$

Accordingly: $R_n \to_p 0$.
□

Problem 5.13. The same approach used in Problem 5.12 can be used to determine the limiting distribution of the sample median under more general conditions. Again, let $X_1, \cdots, X_n$ be i.i.d. with distribution function F and median μ where now

$$\lim_{n\to\infty} \sqrt{n}[F(\mu + s/a_n) - F(\mu)] = \psi(s)$$

for some increasing function ψ and sequence of constants $a_n \to \infty$. The asymptotic distribution of $a_n(\widehat{\mu_n} - \mu)$ will be determined by considering the objective function

$$Z_n(u) = \frac{a_n}{\sqrt{n}} \sum_{i=1}^{n} [|X_i - \mu/a_n| - |X_i - \mu|].$$

(a) Show that $U_n = a_n(\widehat{\mu_n} - \mu)$ minimizes Z_n.
(b) Repeat the steps used in Problem 5.12 to show that

$$(Z_n(u_1), \cdots, Z_n(u_k)) \to_d (Z(u_1), \cdots, Z(u_k))$$

where $Z(u) = -uW + 2\int_0^u \psi(s)ds$ and $W\ N(0,1)$.
(c) Show that $a_n(\widehat{\mu_n} - \mu) \to_d \psi^{-1}(W/2)$.

Solution. (a) Referring to Problem 4.19, we have:

$$\begin{aligned} Z_n(u) &= \frac{a_n}{\sqrt{n}}(\sum_{i=1}^{n}(|X_i - \mu - \frac{u}{a_n}| - |X_i - \mu|)) = \frac{1}{\sqrt{n}}(\sum_{i=1}^{n}(|a_n(X_i - \mu) - u| - |a_n(X_i - \mu)|)) \\ &= \frac{1}{\sqrt{n}}(\sum_{i=1}^{n}|a_n(X_i - \mu) - u| - \sum_{i=1}^{n}|a_n(X_i - \mu)|) = \frac{1}{\sqrt{n}}.(g_n^*(u) - c): \ X_n^* =^{def} a_n.(X_i - \mu). \quad (*) \end{aligned}$$

By $(*)$, $arg(\min(Z_n)) = arg(\min(g_n^*))$ and by Problem 4.19, $arg(\min(g_n^*)) = \widehat{\mu_n^*}$. But, $\widehat{\mu_n^*} = a_n.(\widehat{\mu_n} - \mu)$. Hence, $arg(\min(Z_n)) = a_n.(\widehat{\mu_n} - \mu)$.

(b) By Theorem 3.8, for $X_i^* = sgn(X_i - \mu), \mu^* = 0$, and $\sigma^* = 1$, and by Theorem 3.6 for $X_i^{**} =$

$I_{X_i \leq \mu + \frac{s}{a_n}} - I_{X_i \leq \mu}$ it follows that:

$$\begin{aligned}
\lim_{n\to\infty} Z_n(u) &= \lim_{n\to\infty} \frac{a_n}{\sqrt{n}} (\sum_{i=1}^{n} (|X_i - \mu - \frac{u}{a_n}| - |X_i - \mu|)) \\
&= \lim_{n\to\infty} \frac{a_n}{\sqrt{n}} (\sum_{i=1}^{n} (\frac{-u}{a_n}.sgn(X_i - \mu) + 2\int_0^{\frac{u}{a_n}} I(X_i - \mu \leq s) - I(X_i - \mu \leq 0) ds)) \\
&= \lim_{n\to\infty} \frac{a_n}{\sqrt{n}} (\sum_{i=1}^{n} (\frac{-u}{a_n}.sgn(X_i - \mu) + \frac{2}{a_n}\int_0^{u} I(X_i - \mu \leq \frac{s}{a_n}) - I(X_i - \mu \leq 0) ds)) \\
&= \lim_{n\to\infty} \frac{-u}{\sqrt{n}} (\sum_{i=1}^{n} (sgn(X_i - \mu)) + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{u} I(X_i - \mu \leq \frac{s}{a_n}) - I(X_i - \mu \leq 0) ds)) \\
&= \lim_{n\to\infty} -u. \frac{\sum_{i=1}^{n} X_i^*}{\sqrt{n}} + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{u} X_i^{**} ds \\
&= -u.W + 2\int_0^{u} \lim_{n\to\infty} \sqrt{n} (\frac{\sum_{i=1}^{n} X_i^{**}}{n}) ds \\
&= -u.W + 2\int_0^{u} \lim_{n\to\infty} \sqrt{n} (F(\mu + \frac{s}{\sqrt{n}}) - F(\mu)) ds \\
&= -u.W + 2.\int_0^{u} \psi(s) ds.
\end{aligned}$$

(c) By Theorem 3.2 for $X_n^* = Z_n$ and $g^*(X^*) = arg(\min(X^*))$ and part (b) in which $Z_n(u) \to_d Z(u)$ it follows that:

$$a_n.(\widehat{\mu_n} - \mu) = arg(\min_{-\infty<u<\infty} Z_n(u)) \to_d arg(\min_{-\infty<u<\infty} Z(u)) = \psi^{-1}(\frac{W}{2}).$$

□

Problem 5.15. In Theorems 5.3. and 5.4, we assume that the parameter space Θ is an open subset of R^p. However, in many situations, this assumption is not valid; for example, the model may impose constrains on the parameter θ which effectively makes Θ a closed set. If Θ is not an open set then the MLE of θ need not satisfy the likelihood equations as the MLE $\widehat{\theta_n}$ may lie on the boundary of Θ. In determining the asymptotic distribution of $\widehat{\theta_n}$ the main concern is whether or not the true value of the parameter lies on the boundary of the parameter space. If θ lies in the interior of Θ then eventually (for sufficiently large n) $\widehat{\theta_n}$ will satisfy the likelihood equations and so Theorems 5.3 and 5.4 will still hold; however, the situation becomes more complicated if θ lies on the boundary of Θ.
Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with density or frequency function $f(x;\theta)$ (Satisfying conditions (B2)-(B6)) where θ lies on the boundary of Θ. Define (as in Problem 5.14) the function

$$Z_n(u) = \sum_{i=1}^{n} ln[f(X_i; \theta + u/\sqrt{n})/f(X_i;\theta)]$$

and the set

$$C_n = \{u : \theta + u/\sqrt{n} \in \Theta\}.$$

The limiting distribution of the MLE can be determined by the limiting behaviour of Z_n and C_n.
(a) Show that $\sqrt{n}(\widehat{\theta_n} - \theta)$ maximizes $Z_n(u)$ subject to the constraint $u \in C_n$.
(b) Suppose that $\{C_n\}$ is a decreasing sequence of sets whose limit is C. Show that C is non-empty.

(c) Parts (a) and (b) (together with Problem 5.14) suggest that $\sqrt{n}(\widehat{\theta_n} - \theta)$ converges in distribution to the maximizer of

$$Z(u) = u^T.W - \frac{1}{2}u^T J(\theta)u$$

(where $W \sim N_p(0, I(\theta))$) subject to $u \in C$. Suppose that $X_1, \cdots, X_n$ are i.i.d. Gamma random variables with shape parameter α and scale parameter λ where the parameter space is restricted so that $\alpha \geq \lambda > 0$ (that is, $E(X_i) \geq 1$.) If $\alpha = \lambda$, describe the limiting distribution of the MLEs.

Solution. (a) As:

$$\frac{dZ_n(u)}{du} = \sum_{i=1}^{n} \frac{d}{du}[\ln(f(X_i; \theta + \frac{u}{\sqrt{n}})) - \ln(f(X_i; \theta))] = \sum_{i=1}^{n} \frac{1}{\sqrt{n}}.l^{'}(X_i; \theta + \frac{u}{\sqrt{n}}) = 0,$$

it follows that $\widehat{\theta_n} = \theta + \frac{u}{\sqrt{n}}$, or $u = \sqrt{n}(\widehat{\theta_n} - \theta)$.

(b) Since $C_1 \supseteq \cdots \supseteq C_n \supseteq \cdots$, it follows that $C = \lim_{n\to\infty} C_n = \cap_{n=1}^{\infty} C_n$. Next, since Θ is closed and θ lies on its boundary,$\theta \in \Theta$, and hence, $0 \in C_n$ $(n \geq 1)$. Accordingly, by former result, $0 \in C$ and $C \neq \varnothing$.

(c) By Problem 5.14(b); and considering the fact that X_i is a two-parameter exponential family we have:

$$\sqrt{n}.(\widehat{\theta_n} - \theta) \to_d J^{-1}(\theta).N_2(0, I(\theta)) = N_2(0, J^{-1}(\theta).I(\theta).J^{-1}(\theta)) = N_2(0, I(\theta)) : \quad \theta = (\alpha, \lambda).$$

Hence, by Example 5.15 for $\alpha = \lambda = c$, we have:

$$\sqrt{n}(\widehat{\alpha_n} - \alpha) \to_d N(0, \frac{c}{c.\psi'(c) - 1})$$
$$\sqrt{n}(\widehat{\lambda_n} - \lambda) \to_d N(0, \frac{c^2}{c.\psi'(c) - 1}) : \quad \psi^{'}(c) = \frac{d^2}{dc^2}\log(\Gamma(c)).$$

□

Problem 5.17. Let $X_1, \cdots, X_n$ be i.i.d. random variables with density or frequency function $f(x; \theta)$ where θ is a real-valued parameter. Suppose that MLE of θ, $\widehat{\theta}$, satisfies the likelihood equation

$$\sum_{i=1}^{n} l^{'}(X_i; \widehat{\theta}) = 0$$

where $l^{'}(x; \theta)$ is the derivative with respect to θ of $\ln f(x; \theta)$.
(a) Let $\widehat{\theta_{-j}}$ be MLE of θ based on all the X_i's except X_j. Show that

$$\widehat{\theta_{-j}} \approx \widehat{\theta} + \frac{l^{'}(X_j; \widehat{\theta})}{\sum_{i=1}^{n} l^{"}(X_i; \widehat{\theta})}$$

(if n is reasonably large).
(b) Show that the jackknife estimator of $\widehat{\theta}$ satisfies

$$\widehat{Var(\widehat{\theta})} \approx \frac{n-1}{n} \frac{\sum_{j=1}^{n}[l^{'}(X_j; \widehat{\theta})]^2}{(\sum_{j=1}^{n} l^{"}(X_j; \widehat{\theta}))^2}.$$

(c) The result of part (b) suggests that the jackknife estimator of $Var(\widehat{\theta})$ is essentially the "sandwich" estimator; the later estimator is valid when the model is misspecified. Explain the apparent equivalences between these two estimators of $Var(\widehat{\theta})$.

Solution. (a) By given conditions:

$$\begin{aligned} 0 &= \sum_{1\le j\neq i\le n} l^{'}(X_i,\widehat{\theta_{-j}}) \approx \sum_{1\le j\neq i\le n} l^{'}(X_i,\theta) + (\widehat{\theta_{-j}}-\theta).\sum_{1\le j\neq i\le n} l^{''}(X_i,\theta) \quad (*) \\ 0 &= \sum_{1\le j\neq i\le n} l^{'}(X_i,\widehat{\theta}). \quad (**) \end{aligned}$$

Take $\theta=\widehat{\theta}$ in $(*)$, and add $l^{'}(X_j;\widehat{\theta})$ to both sides of it and then use $(**)$ to get:

$$l^{'}(X_j;\widehat{\theta}) \approx (\widehat{\theta_{-j}}-\widehat{\theta}).\sum_{1\le j\neq i\le n} l^{''}(X_i,\widehat{\theta}). \quad (***)$$

But as $n\uparrow\infty$, we have $\sum_{1\le j\neq i\le n} l^{''}(X_i,\widehat{\theta}) \approx \sum_{1\le i\le n} l^{''}(X_i,\widehat{\theta})$. Consequently, using the later result in $(***)$ it follows that:

$$l^{'}(X_j;\widehat{\theta}) \approx (\widehat{\theta_{-j}}-\widehat{\theta}).\sum_{1\le i\le n} l^{''}(X_i,\widehat{\theta}). \quad (****)$$

And $(****)$ is equivalent to $\frac{l^{'}(X_j;\widehat{\theta})}{\sum_{1\le i\le n} l^{''}(X_i,\widehat{\theta})} \approx \widehat{\theta_{-j}}-\widehat{\theta}$, and the assertion follows.

(b) First, an application of Part (a) and the given condition in the problem yield:

$$\widehat{\theta_{\bullet}} = \frac{1}{n}.\sum_{j=1}^{n}\widehat{\theta_{-j}} \approx \widehat{\theta} + \frac{\frac{1}{n}.\sum_{j=1}^{n} l^{'}(X_j;\widehat{\theta})}{\sum_{j=1}^{n} l^{''}(X_j;\widehat{\theta})} = \theta. \quad (\dagger)$$

Second, using $(\dagger)$ and another application of Part (a) it follows that:

$$\begin{aligned} \widehat{Var(\widehat{\theta})} &= \frac{n-1}{n}.\sum_{j=1}^{n}(\widehat{\theta_{-j}}-\theta_{\bullet})^2 \approx \frac{n-1}{n}.\sum_{j=1}^{n}(\widehat{\theta_{-j}}-\theta)^2 \\ &\approx \frac{n-1}{n}.\sum_{j=1}^{n}\left(\frac{l^{'}(X_j;\widehat{\theta})}{\sum_{1\le j\le n} l^{''}(X_j,\widehat{\theta})}\right)^2 = \frac{n-1}{n}\,\frac{\sum_{j=1}^{n}[l^{'}(X_j;\widehat{\theta})]^2}{(\sum_{j=1}^{n} l^{"}(X_j;\widehat{\theta}))^2}. \end{aligned}$$

(c) As $\widehat{\theta}$ is the solution for the equation $\sum_{i=1}^{n} l^{'}(X_i;\theta)=0$, it follows that $\widehat{\theta}$ is the substitution principle estimator of the functional parameter $\theta(F)$ defined by:

$$\int_{-\infty}^{\infty} l^{'}(x;\theta(F))dF(x) = 0: \quad \widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i\le x},$$

in which the influence curve of $\theta(F)$ is:

$$\phi(x;F) = \frac{l^{'}(x;\theta(F))}{\int_{-\infty}^{\infty} l^{''}(x;\theta(F))dF(x)}.$$

Consequently:

$$\sigma^2 = \int_{-\infty}^{\infty}\phi^2(x;F)dF(x) = \frac{\int_{-\infty}^{\infty}[l^{'}(x;\theta(F))]^2dF(x)}{[\int_{-\infty}^{\infty} l^{''}(x;\theta(F))dF(x)]^2},$$

and therefore:

$$\widehat{\sigma}^2_{spe}(\widehat{\theta}) = \frac{\sum_{i=1}^n [l'(X_i,\widehat{\theta})]^2}{[\sum_{i=1}^n l''(X_i,\widehat{\theta})]^2}. \quad (\dagger\dagger)$$

Next, by (††) and Part (b):

$$\lim_{n\to\infty}\left(\frac{\widehat{\sigma}^2_{jke}(\widehat{\theta})}{\widehat{\sigma}^2_{spe}(\widehat{\theta})}\right) = 1. \quad (\dagger\dagger\dagger)$$

Finally, the results in († † †) shows that the the jackknife estimator and the substitution principle estimator of $\widehat{Var}(\theta)$ are asymptotically equal.
□

Problem 5.19. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ has a joint density or frequency function $f(x;\theta)$ where θ has prior density $\pi(\theta)$. If $T = T(\mathbf{X})$ is sufficient for θ, show that the posterior density of θ given $\mathbf{X}$=$\mathbf{x}$ is the same as the posterior density of θ given $T = T(\mathbf{x})$.

Solution. By Sufficiency of $T = T(X)$ it follows that $f_{X|\theta,T}(x|\theta,t) = f_{X|T}(x|t)$ for all $x, t = T(x)$. Thus:

$$\begin{aligned}
\pi_{\theta|X}(\theta|x) &= \frac{f_{X|\theta}(x|\theta)}{\int_\Theta f_{X|\theta}(x|\theta)\pi(\theta)d\theta} = \frac{f_{X|\theta,T}(x|\theta,t).f_{T|\theta}(t|\theta)}{\int_\Theta f_{X|\theta,T}(x|\theta,t).f_{T|\theta}(t|\theta)\pi(\theta)d\theta} \\
&= \frac{f_{X|T}(x|t).f_{T|\theta}(t|\theta)}{\int_\Theta f_{X|T}(x|t).f_{T|\theta}(t|\theta)\pi(\theta)d\theta} = \frac{f_{T|\theta}(t|\theta)}{\int_\Theta f_{T|\theta}(t|\theta)\pi(\theta)d\theta} \\
&= \pi_{\theta|T(X)}(\theta|T(x)), \ \texttt{for all}\ \theta.
\end{aligned}$$

□

Problem 5.21. The Zeta distribution is sometimes used in insurance as a model for the number of policies held by a single person in an insurance portfolio. the frequency function for this distribution is

$$f(x;\alpha) = \frac{x^{-(\alpha+1)}}{\zeta(\alpha+1)}$$

for $x = 1, 2, 3, \cdots$ where $\alpha > 0$ and

$$\zeta(p) = \sum_{k=1}^{\infty} k^{-p}.$$

(The function $\zeta(p)$ is called the Riemann zeta function.)
(a) Suppose that $X_1, \cdots, X_n$ are i.i.d. Zeta random variables. Show that the MLE of α satisfies the equation

$$\frac{1}{n}\sum_{i=1}^n \ln(X_i) = -\frac{\zeta'(\widehat{\alpha_n}+1)}{\zeta(\widehat{\alpha_n}+1)}$$

and find the limiting distribution of $\sqrt{n}(\widehat{\alpha_n} - \alpha)$.
(b) Assume the following density for α :

$$\pi(\alpha) = \frac{1}{2}\alpha^2 exp(-\alpha) \ \texttt{for}\ \alpha > 0$$

A sample of 85 observations is collected; its frequency distribution is given in Table 5.8.

Table 5.8 Data for Problem 5.21.

Observation	1	2	3	4	5
Frequency	63	14	5	1	2

Find the posterior distribution of α. What is the mode (approximately) of this posterior distribution ?
(c) Repeat part (b) using the improper prior density

$$\pi(\alpha) = \frac{1}{\alpha} \ \texttt{for} \ \alpha > 0.$$

Compare the posterior densities in part (b) and (c).

Solution. (a) First,

$$\begin{aligned} l(\alpha) &= \log(f(\mathbf{x};\alpha)) = \log(\prod_{i=1}^{n} \frac{1}{\zeta(\alpha+1)}.x_i^{-(\alpha+1)}) \\ &= \sum_{i=1}^{n}[-\log(\zeta(\alpha+1)) - (\alpha+1).\log(x_i)] = -n.\log(\zeta(\alpha+1)) - (\alpha+1).\sum_{i=1}^{n}\log(x_i) \end{aligned}$$

implying:

$$\frac{dl(\alpha)}{d\alpha} = -n.\frac{\zeta^{'}(\alpha+1)}{\zeta(\alpha+1)} - \sum_{i=1}^{n}\log(x_i) = 0 \Rightarrow \overline{\log(X)_n} = -\frac{\zeta^{'}(\widehat{\alpha_n}+1)}{\zeta(\widehat{\alpha_n}+1)}.$$

Second, as $f(x) = \frac{1}{\zeta(\alpha+1)}.x^{-(\alpha+1)} = \exp[-(\alpha+1).\log(x) - \log(\zeta(\alpha+1))]$, it follows that X has an exponential family density with $c(\alpha) = -(\alpha+1), T(x) = \log(x), d(\alpha) = \log(\zeta(\alpha+1))$, and $S(x) = 0$. Hence, by Example 5.6, and theorem 5.3 it follows that $\sqrt{n}.(\widehat{\alpha_n} - \alpha) \to_d N(0, \frac{1}{I(\alpha)})$ in which

$$\begin{aligned} I(\alpha) &= d^{''}(\alpha) - c^{''}(\alpha).\frac{d^{'}(\alpha)}{c^{'}(\alpha)} \\ &= \frac{\zeta^{''}(\alpha+1)\zeta(\alpha+1) - (\zeta^{'}(\alpha+1))^2}{(\zeta(\alpha+1))^2} - 0.\frac{\frac{\zeta^{'}(\alpha+1)}{\zeta(\alpha+1)}}{-1} \\ &= \frac{\zeta^{''}(\alpha+1)\zeta(\alpha+1) - (\zeta^{'}(\alpha+1))^2}{(\zeta(\alpha+1))^2}. \end{aligned}$$

Thus:

$$\sqrt{n}.(\widehat{\alpha_n} - \alpha) \to_d N(0, \frac{(\zeta(\alpha+1))^2}{\zeta''(\alpha+1)\zeta(\alpha+1) - (\zeta'(\alpha+1))^2}).$$

(b) First, assuming $A = (\frac{1}{2})^{14}.(\frac{1}{3})^{5}.(\frac{1}{4}).(\frac{1}{5})^{2}$, we have:

$$\begin{aligned} \pi(\alpha|\mathbf{x}) &= \frac{f(\mathbf{x}|\alpha)\pi(\alpha)}{\int_0^\infty (f(\mathbf{x}|\alpha)\pi(\alpha))d\alpha} = \frac{\prod_{i=1}^{n} f(x_i|\alpha)\pi(\alpha)}{\int_0^\infty (\prod_{i=1}^{n} f(x_i|\alpha)\pi(\alpha))d\alpha} \\ &= \frac{A^{\alpha+1}.\frac{\pi(\alpha)}{(\zeta(\alpha+1))^{85}}}{\int_0^\infty (A^{\alpha+1}.\frac{\pi(\alpha)}{(\zeta(\alpha+1))^{85}})d\alpha}. \end{aligned}$$

And,

$$\begin{aligned} \pi_1(\alpha|\mathbf{x}) &= \frac{A^{\alpha+1}.(\zeta(\alpha+1))^{-85}.\alpha^2.e^{-\alpha}/2}{\int_0^\infty (A^{\alpha+1}.(\zeta(\alpha+1))^{-85}.\alpha^2.e^{-\alpha}/2)d\alpha} \\ &= \frac{\exp((\alpha+1)(\log(A)-1) + 2\log(\alpha) - 85.\log(\zeta(\alpha+1)))}{\int_0^\infty (\exp((\alpha+1)(\log(A)-1) + 2\log(\alpha) - 85.\log(\zeta(\alpha+1))))d\alpha}. \end{aligned}$$

Second, $\frac{d}{d\alpha}\pi_1(\alpha|x) = 0$ yields $(\log(A)-1) + \frac{2}{\alpha} - 85\frac{\zeta'(\alpha+1)}{\zeta(\alpha+1)} = 0$, or equivalently:

$$(\log(A)-1).\widehat{\alpha}.\zeta(\widehat{\alpha}+1) + 2.\zeta(\widehat{\alpha}+1) - 85.\widehat{\alpha}.\zeta'(\widehat{\alpha}+1) = 0.$$

(c) First,

$$\begin{aligned}
\pi_2(\alpha|\mathbf{x}) &= \frac{A^{\alpha+1}.(\zeta(\alpha+1))^{-85}.\frac{1}{\alpha}}{\int_0^\infty (A^{\alpha+1}.(\zeta(\alpha+1))^{-85}.\frac{1}{\alpha}d\alpha} \\
&= \frac{\exp((\alpha+1)\log(A) - 85\log(\zeta(\alpha+1)) - \log(\alpha))}{\int_0^\infty (\exp((\alpha+1)\log(A) - 85\log(\zeta(\alpha+1)) - \log(\alpha)))d\alpha}.
\end{aligned}$$

Thus, $\frac{d}{d\alpha}\pi_2(\alpha|\mathbf{x}) = 0$ implies $\log(A) - 85\frac{\zeta'(\alpha+1)}{\zeta(\alpha+1)} - \frac{1}{\alpha} = 0$, or equivalently:

$$\log(A).\widehat{\alpha}.\zeta(\widehat{\alpha}+1) - 85.\zeta'(\widehat{\alpha}+1).\widehat{\alpha} - \zeta(\widehat{\alpha}+1) = 0.$$

Second, to compare posterior densities in parts (b) and (c) define a function H via:

$$\begin{aligned}
H(\alpha) &= \frac{\pi_1(\alpha|\mathbf{x})}{\pi_2(\alpha|\mathbf{x})} = (\frac{c_1}{c_2}).\frac{\alpha^3}{e^\alpha} : \\
c_1 &= \int_0^\infty (\exp((\alpha+1)\log(A) - 85\log(\zeta(\alpha+1)) - \log(\alpha)))d\alpha \\
c_2 &= \int_0^\infty (\exp((\alpha+1)(\log(A)-1) + 2\log(\alpha) - 85.\log(\zeta(\alpha+1))))d\alpha.
\end{aligned}$$

It is clear that $\lim_{\alpha\to 0} H(\alpha) = 0 = \lim_{\alpha\to\infty} H(\alpha)$. Furthermore, H attains its maximum at $\alpha = 3$ (Exercise !). Consequently:

$$\pi_1(\alpha|\mathbf{x}) \le (\frac{27}{e^3}\frac{c_1}{c_2})\pi_2(\alpha|\mathbf{x}).$$

□

Problem 5.23. The concept of Jeffreys priors can be extended to derive "non-informative" priors for multiple parameters. Suppose that $\mathbf{X}$ has joint density or frequency function $f(\mathbf{x};\theta)$ and define the matrix

$$I(\theta) = E_\theta[S(\mathbf{X};\theta)S^T(\mathbf{X};T)]$$

where $S(\mathbf{x};\theta)$ is the gradient (vector of partial derivatives) of $\ln f(\mathbf{x};\theta)$ with respect to θ. The Jeffreys prior for θ is proportional to $\det(I(\theta))^{1/2}$.
(a) Show that the Jeffreys prior can be derived using the same considerations made in the single parameter space. That is, if $\phi = g(\theta)$ for some one-to-one function g such that $I(\phi)$ is constant then the Jeffreys prior for θ corresponds to a uniform prior for ϕ.
(b) Suppose that $X_1, \cdots, X_n$ are i.i.d. Normal random variables with mean μ and variance σ^2. Find the Jeffreys prior for (μ, σ).

Solution. (a) By assumption, $\pi_{\text{Jeffrey}}(\theta) = c_1.\sqrt{\det(I(\theta))}$ $(c_1 > 0)$. Also:

$$\begin{aligned}
I(\theta) &= (E_\theta(\frac{d\log(L)}{d\theta_i}\frac{d\log(L)}{d\theta_j}))_{i,j=1}^p \\
I(\phi) &= (E_\phi(\frac{d\log(L)}{d\phi_i}\frac{d\log(L)}{d\phi_j}))_{i,j=1}^p.
\end{aligned}$$

Now, by Theorem 2.3 for $\phi = g(\theta)$, it follows that:

$$\begin{aligned}
\pi_{\text{Jeffrey}}(\phi) &= \pi_{\text{Jeffrey}}(\theta).|\det(\frac{d\theta_i}{d\phi_j})_{i,j=1}^p| \\
&= c_1.\sqrt{\det(I(\theta))}.|\det(\frac{d\theta_i}{d\phi_j})_{i,j=1}^p| \\
&= c_1.\sqrt{\det(I(\theta)).|\det(\frac{d\theta_i}{d\phi_j})_{i,j=1}^p|^2} \\
&= c_1.\sqrt{\det((\frac{d\theta_k}{d\phi_i})_{i,k=1}^p).\det((E_\theta(\frac{d\log(L)}{d\theta_k}\frac{d\log(L)}{d\theta_l}))_{k,l=1}^p).\det((\frac{d\theta_l}{d\phi_j})_{l,j=1}^p)} \\
&= c_1.\sqrt{\det((E_\phi(\sum_{k,l}\frac{d\theta_k}{d\phi_i}\frac{d\log(L)}{d\theta_k}\frac{d\log(L)}{d\theta_l}\frac{d\theta_l}{d\phi_j}))_{i,j=1}^p)} \\
&= c_1.\sqrt{\det((E_\phi(\frac{d\log(L)}{d\phi_i}\frac{d\log(L)}{d\phi_j}))_{i,j=1}^p)} \\
&= c_1.\sqrt{\det(I(\phi))}. \quad (*)
\end{aligned}$$

But, $I(\phi) = \texttt{constant}$, and hence:

$$\sqrt{\det(I(\phi))} = c_2, \quad (c_2 > 0). \quad (**)$$

Consequently, by $(*)$ and $(**)$ it follows that:

$$\pi_{\text{Jeffrey}}(\phi) = c_1.c_2,$$

that is, the Jeffrey prior for θ corresponds to a uniform prior for ϕ.

(b)As $\log(f(x|(\mu,\sigma^2))) = constant - \frac{n}{2}\log(\sigma^2) - \frac{(n-1).s_x^2+n.(\overline{x}-\mu)^2}{2.\sigma^2}$, using $E(\overline{X}) = \mu, E(n(\overline{X}-\mu)^2) = \sigma^2$, and $E((n-1).s_x^2) = (n-1).\sigma^2$ it follows that:

$$\begin{aligned}
I((\mu,\sigma^2)) &= -\begin{pmatrix} \frac{d^2}{d\mu^2}\log(L) & \frac{d^2}{d\mu d\sigma^2}\log(L) \\ \frac{d^2}{d\sigma^2 d\mu}\log(L) & \frac{d^2}{d(\sigma^2)^2}\log(L) \end{pmatrix} \\
&= \begin{pmatrix} -E(\frac{-n}{\sigma^2}) & -E(\frac{-n(\overline{X}-\mu)}{\sigma^4}) \\ -E(\frac{-n(\overline{X}-\mu)}{\sigma^4}) & -E(\frac{n}{2\sigma^4} - \frac{(n-1)s_x^2+n(\overline{X}-\mu)^2}{\sigma^6}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2.\sigma^4} \end{pmatrix}.
\end{aligned}$$

Consequently:

$$\pi_{\text{Jeffrey}}((\mu,\sigma^2)|x) = c.\sqrt{\det(I(\mu,\sigma^2))} = c.\sqrt{\frac{n^2}{2.\sigma^6}} = \frac{c.(\frac{n}{\sqrt{2}})}{(\sigma^2)^{\frac{3}{2}}}. \quad (***)$$

Now, for $g(x,y) = (x,\sqrt{y})$, it follows from $(***)$ that:

$$\begin{aligned}
\pi_{\text{Jeffrey}}((\mu,\sigma)|x) &= \pi_{\text{Jeffrey}}((\mu,\sigma^2)|x).|det(\frac{d(\mu,\sigma^2)}{d(\mu,\sigma)})| \\
&= \frac{c.\frac{n}{\sqrt{2}}}{\sigma^3} * 2\sigma = \frac{c.\sqrt{2}.n}{\sigma^2} = \frac{c^*}{\sigma^2}. \quad (\dagger)
\end{aligned}$$

Note that a direct calculation yields:

$$\pi_{\text{Jeffrey}}((\mu,\sigma)|x) = c^{'}.\sqrt{\det(I(\mu,\sigma))} = \frac{c^{'}.\sqrt{n(2n-4)}}{\sigma^2} = \frac{c^*}{\sigma^2}. \quad (\dagger\dagger)$$

And, finally, both (†) and (††) show that $\pi_{\text{Jeffrey}}((\mu,\sigma)|x)$ is proportional only to $1/\sigma^2$.
□

Chapter 6

Optimality in Estimation

Problem 6.1. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ have joint density or frequency function $f(x;\theta)$ where θ is a real-valued parameter with a proper prior density function $\pi(\theta)$. For squared error loss, define the Bayes risk of an estimator $\widehat{\theta} = S(\mathbf{X})$:

$$R_B(\widehat{\theta}, \theta) = \int_\Theta E_\theta[(\widehat{\theta} - \theta)^2]\pi(\theta)d\theta.$$

The Bayes estimator minimizes the Bayes risk.
(a) Show that the Bayes estimator is the mean of the posterior distribution of θ.
(b) Suppose that the Bayes estimator in (a) is also an unbiased estimator. Show that the Bayes risk of this estimator must be 0. (This result implies that Bayes estimators and unbiased estimators agree only in pathological examples.)

Solution. (a) By $f(x|\theta).\pi(\theta) = \pi(\theta|x).f(x)$, we have:

$$\begin{aligned} R_B(\widehat{\theta}, \theta) &= \int_\Theta E_\theta((\widehat{\theta} - \theta)^2)\pi(\theta)d\theta = \int_\Theta (\int_\chi (\widehat{\theta(x)} - \theta)^2 f(x|\theta)dx)\pi(\theta)d\theta = \int_\Theta \int_\chi (\widehat{\theta(x)} - \theta)^2 f(x|\theta)\pi(\theta)dxd\theta \\ &= \int_\Theta \int_\chi (\widehat{\theta(x)} - \theta)^2 \pi(\theta|x).f(x)dxd\theta = \int_\chi (\int_\Theta (\widehat{\theta(x)} - \theta)^2 \pi(\theta|x)d\theta)f(x)dx. \quad (*) \end{aligned}$$

Now, by $(*)$ the Bayes risk is minimized when the posterior expected loss $\int_\Theta (\widehat{\theta(x)} - \theta)^2\pi(\theta|x)d\theta$ is minimized and it is minimized at $\widehat{\theta(x)} = E(\theta|x)$.

(b) First, given $E_\theta(\widehat{\theta}|\theta) = \theta$, it follows:

$$\begin{aligned} E_\theta((\widehat{\theta} - \theta)^2) &= E_\theta((\widehat{\theta})^2 - 2.\theta.\widehat{\theta} + (\theta)^2) = E_\theta((\widehat{\theta})^2) - 2.E_\theta(\theta.\widehat{\theta}) + E_\theta((\theta)^2) \\ &= E_\theta((\widehat{\theta})^2) - 2.E_\theta(E_\theta(\theta.\widehat{\theta}|\theta)) + E_\theta((\theta)^2) = E_\theta((\widehat{\theta})^2) - 2.E_\theta(\theta.E_\theta(\widehat{\theta}|\theta)) + E_\theta((\theta)^2) \\ &= E_\theta((\widehat{\theta})^2) - 2.E_\theta((\theta)^2) + E_\theta((\theta)^2) = E_\theta((\widehat{\theta})^2) - E_\theta((\theta)^2). \quad (**) \end{aligned}$$

Second, given $E_\theta(\theta|\widehat{\theta}) = \widehat{\theta}$, it follows:

$$\begin{aligned} E_\theta((\widehat{\theta} - \theta)^2) &= E_\theta((\widehat{\theta})^2 - 2.\widehat{\theta}.\theta + (\theta)^2) = E_\theta((\widehat{\theta})^2) - 2.E_\theta(\widehat{\theta}.\theta) + E_\theta((\theta)^2) \\ &= E_\theta((\widehat{\theta})^2) - 2.E_\theta(E_\theta(\widehat{\theta}.\theta|\widehat{\theta})) + E_\theta((\theta)^2) = E_\theta((\widehat{\theta})^2) - 2.E_\theta(\widehat{\theta}.E_\theta(\theta|\widehat{\theta})) + E_\theta((\theta)^2) \\ &= E_\theta((\widehat{\theta})^2) - 2.E_\theta((\widehat{\theta})^2) + E_\theta((\theta)^2) = -E_\theta((\widehat{\theta})^2) + E_\theta((\theta)^2). \quad (***) \end{aligned}$$

Now, comparing $(**)$ and $(***)$ it follows that $E_\theta((\widehat{\theta}-\theta)^2)=0$, and hence:

$$R_B(\widehat{\theta},\theta)=\int_\Theta 0\ \ \pi(\theta)d\theta=0.$$

□

Problem 6.3. Suppose that $X_1,\cdots,X_n$ are i.i.d. Poisson random variables with mean θ where θ has a Gamma (α,β) prior distribution.
(a) Show that

$$\widehat{\theta}=\frac{\alpha+\sum_{i=1}^n X_i}{\beta+n}$$

is the Bayes estimator of θ under squared error loss.
(b) Use the result of (a) to show that any estimator of the form $a\overline{X}+b$ for $0<a<1$ and $b>0$ is an admissible estimator of θ under squared error loss.

Solution. (a) By Problem 6.1,

$$\widehat{\theta}=E(\theta|\mathbf{x}).\quad (*)$$

Also:

$$\begin{aligned}\pi(\theta|\widehat{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_\Theta f(\mathbf{x}|t)\pi(t)dt}=\frac{\prod_{i=1}^n(\frac{e^{-\theta}.\theta^{x_i}}{x_i!})\frac{\beta^\alpha.\theta^{\alpha-1}}{\Gamma(\alpha)}exp(-\beta.\theta)}{\int_0^\infty\prod_{i=1}^n(\frac{e^{-t}.t^{x_i}}{x_i!})\frac{\beta^\alpha.t^{\alpha-1}}{\Gamma(\alpha)}exp(-\beta.t)dt}\\ &= \frac{\theta^{\sum_{i=1}^n x_i+\alpha-1}.e^{-(n+\beta)\theta}}{\int_0^\infty t^{\sum_{i=1}^n x_i+\alpha-1}.e^{-(n+\beta)t}dt}=\frac{(n+\beta)^{\sum_{i=1}^n x_i+\alpha}}{\Gamma(\sum_{i=1}^n x_i+\alpha)}.\theta^{\sum_{i=1}^n x_i+\alpha-1}.e^{-(n+\beta)\theta}.\quad (**)\end{aligned}$$

Consequently, by $(**)$ it follows that $\theta|\mathbf{x}\sim Gamma(\sum_{i=1}^n x_i+\alpha,n+\beta)$ implying:

$$E(\theta|\mathbf{x})=\frac{\sum_{i=1}^n x_i+\alpha}{n+\beta}.\quad (***)$$

Finally, a comparison of $(*)$ and $(***)$ proves the assertion.

(b) As $\widehat{\theta}=\frac{n.\overline{X}+\alpha}{n+\beta}=\frac{n}{n+\beta}.\overline{X}+\frac{\alpha}{n+\beta}$, fix $n>0$ and write $n/(n+\beta)=a$ and $\alpha/(n+\beta)=b$. Then, $\alpha=n(b/a)$, and $\beta=n((1-a)/a)$. Accordingly, taking

$$\theta_0=Gamma(n(b/a),n((1-a)/a))$$

implies $E(\theta_0|\mathbf{x})=a.\overline{X}+b$.

□

Problem 6.5. Given a loss function L, we want to find a minimax estimator of a parameter θ.
(a) Suppose that $\widehat{\theta}$ is a Bayes estimator of θ for some prior distribution $\pi(\theta)$ with

$$R_B(\widehat{\theta})=\sup_{\theta\in\Theta}R_\theta(\widehat{\theta}).$$

Show that $\widehat{\theta}$ is a minimax estimator. (The prior distribution π is called a least favourable distribution.)
(b) Let $\{\pi_n(\theta)\}$ be a sequence of prior density functions on Θ and suppose that $\widehat{\theta_n}$ are the corresponding Bayes estimators. If $\widehat{\theta_0}$ is an estimator with

$$\sup_{\theta\in\Theta}R_\theta(\widehat{\theta_0})=\lim_{n\to\infty}\int_\Theta R_\theta(\widehat{\theta_n})\pi_n(\theta)d\theta$$

show that $\widehat{\theta_0}$ is a minimax estimator.
(c) Suppose that $X \sim Bin(n, \theta)$. Assuming squared error loss, find a minimax estimator of θ.

Solution. (a) Given arbitrary estimator $\widetilde{\theta}$ of θ. Then:

$$\begin{aligned} \sup_{\theta\in\Theta} R_\theta(\widehat{\theta}) &= R_B(\widehat{\theta}) = \int_\Theta R_\theta(\widehat{\theta})\pi(\theta)d\theta \\ &\le \int_\Theta R_\theta(\widetilde{\theta})\pi(\theta)d\theta \le \int_\Theta \sup_{\theta\in\Theta}(R_\theta(\widetilde{\theta}))\pi(\theta)d\theta \\ &= (\sup_{\theta\in\Theta}(R_\theta(\widetilde{\theta}))).\int_\Theta \pi(\theta)d\theta = \sup_{\theta\in\Theta}(R_\theta(\widetilde{\theta})). \end{aligned}$$

Accordingly, $\widehat{\theta}$ is a minimax estimator of θ.

(b) Let $\widetilde{\theta}$ be an arbitrary estimator of θ. Then by:

$$\int_\Theta R_\theta(\widehat{\theta_n})\pi_n(\theta)d\theta \le \int_\Theta R_\theta(\widetilde{\theta})\pi_n(\theta)d\theta \quad (n \ge 1),$$

we have:

$$\begin{aligned} \sup_{\theta\in\Theta} R_\theta(\widehat{\theta_0}) &= \lim_{n\to\infty}\int_\Theta R_\theta(\widehat{\theta_n})\pi_n(\theta)d\theta \le \sup_{n\in\mathbb{N}}(\int_\Theta R_\theta(\widehat{\theta_n})\pi_n(\theta)d\theta) \\ &\le \sup_{n\in\mathbb{N}}(\int_\Theta R_\theta(\widetilde{\theta})\pi_n(\theta)d\theta) \le \sup_{n\in\mathbb{N}}(\int_\Theta (\sup_{\theta\in\Theta}(R_\theta(\widetilde{\theta})))\pi_n(\theta)d\theta) \\ &= \sup_{n\in\mathbb{N}}[(\int_\Theta \pi_n(\theta)d\theta).(\sup_{\theta\in\Theta}(R_\theta(\widetilde{\theta})))] = \sup_{\theta\in\Theta}(R_\theta(\widetilde{\theta})). \end{aligned}$$

Thus, $\widehat{\theta_0}$ is a minimax estimator of θ.

(c) Take $\theta \sim Beta(\alpha, \beta)$. Then, by Problem 6.2. $\widehat{\theta_{\alpha,\beta}}(X) = \frac{\alpha+X}{\alpha+\beta+n}$ and

$$\begin{aligned} R_\theta(\widehat{\theta_{\alpha,\beta}}(X)) &= Var(\widehat{\theta_{\alpha,\beta}}(X)) + Bias^2(\widehat{\theta_{\alpha,\beta}}(X)) = \frac{n.\theta.(1-\theta)}{(\alpha+\beta+n)^2} + \frac{(\alpha-\theta(\alpha+\beta))^2}{(\alpha+\beta+n)^2} \\ &= \frac{((\alpha+\beta)^2-n).\theta^2 + (n-2\alpha(\alpha+\beta)).\theta + c(\alpha,\beta,n)}{(\alpha+\beta+n)^2} = d(\alpha,\beta,n) \end{aligned}$$

if and only if $\alpha = \beta = \frac{\sqrt{n}}{2}$. Consequently, for:

$$\widehat{\theta_{\frac{\sqrt{n}}{2},\frac{\sqrt{n}}{2}}}(X) = \frac{X + \frac{\sqrt{n}}{2}}{n+\sqrt{n}}$$

we have:

$$R_B(\widehat{\theta_{\frac{\sqrt{n}}{2},\frac{\sqrt{n}}{2}}}(X)) = \int_\Theta R_\theta(\widehat{\theta_{\frac{\sqrt{n}}{2},\frac{\sqrt{n}}{2}}}(X))\pi(\theta)d\theta = \int_\Theta d(\alpha,\beta,n)\pi(\theta)d\theta = d(\alpha,\beta,n) = \sup_{\theta\in\Theta}(R_\theta(\widehat{\theta_{\frac{\sqrt{n}}{2},\frac{\sqrt{n}}{2}}}(X))),$$

and by Part (a) , $\widehat{\theta_{\frac{\sqrt{n}}{2},\frac{\sqrt{n}}{2}}}(X)$ is a minimax estimator of θ.
□

Problem 6.7. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ are random variables with joint density or frequency function $f(x;\theta)$ and suppose that $T = T(\mathbf{X})$ for θ.
Suppose that there exists no function $\phi(t)$ such that $\phi(T)$ is an unbiased estimator of $g(\theta)$. Show that no unbiased estimator of $g(\theta)$ (based on $\mathbf{X}$) exists.

Solution. Let for some $S = S(X)$, to have $E_\theta(S) = g(\theta)$. Then, for $\phi(T) = E_\theta(S|T)$ we have:

$$E_\theta(\phi(T)) = E_\theta(E_\theta(S|T)) = E_\theta(S) = g(\theta),$$

a contradiction to the assumption for $g(\theta)$.
$\square$
Problem 6.9. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ have a joint distribution depending on a parameter θ where $T = T(\mathbf{X})$ is sufficient for θ.
(a) Prove Basu's Theorem: If $S = S(\mathbf{X})$ is an ancillary statistic and the sufficient statistic T is complete then T and S are independent.
(b) Suppose that X and Y are independent Exponential random variables with parameter λ. Use Basu's theorem to show that $X + Y$ and $X/(X + Y)$ are independent.
(c) Suppose that $X_1, \cdots, X_n$ are i.i.d. Normal random variables with mean μ and variance σ^2. Let $T = T(X_1, \cdots, X_n)$ be a statistic such that

$$T(X_1 + a, \cdots, X_n + a) = T(X_1, \cdots, X_n) + a$$

and $E(T) = \mu$. Show that

$$Var(T) = Var(\overline{X}) + E[(T - \overline{X})^2].$$

Solution. (a) Fix given $-\infty < s < \infty$ and define:

$$g(t) = P(S(X) = s|T(X) = t) - P(S(X) = s) \quad -\infty < t < \infty.$$

Then in the right hand side of above equality, considering sufficient $T(X)$ (for the first component) and ancillary $S(X)$ (for the second component), it follows that g does not dependent to θ. Furthermore, another usage of ancillary assumption on $S(X)$ yields:

$$\begin{aligned}
E_\theta(g(T)) &= \int_{-\infty}^{\infty} g(t)dP_\theta(t) \\
&= \int_{-\infty}^{\infty} P(S(X) = s|T(X) = t)dP_\theta(T(X) = t) - P(S(X) = s) \\
&= \int_{-\infty}^{\infty} P_\theta(S(X) = s|T(X) = t)dP_\theta(T(X) = t) - P(S(X) = s) \\
&= P_\theta(S(X) = s) - P(S(X) = s) \\
&= P(S(X) = s) - P(S(X) = s) = 0, \quad \texttt{for all } \theta. \quad (*)
\end{aligned}$$

Next, by completeness of T it follows from $(*)$ that:

$$g(t) = 0, \quad -\infty < t < \infty \quad (**)$$

and $(**)$ is equivalent to :

$$P(S(X) = s|T(X) = t) = P(S(X) = s), \quad -\infty < s, t < \infty$$

or equivalently S, T are independent. This proves the Basu's Theorem for continuous random variables. For the discrete random variables one may simply substitute the integrals in above proof with sums.

(b) As $\lambda.X, \lambda.Y \sim \exp(1)$, $\frac{X}{X+Y} = \frac{\lambda.X}{\lambda.X+\lambda.Y}$ and $Z \sim \exp(1)$ does not depend to λ, it follows that $S = \frac{X}{X+Y}$ is an ancillary statistics. Next, using likelihood and an application of Theorem 4.2 show that $X + Y$ is sufficient statistics for λ. On the other hand, $T = X + Y \sim Gamma(2, \frac{1}{\lambda})$ with density (Exercise !):

$$f_{X+Y}(t) = \lambda^2.t.e^{-\lambda.t}. \ \ (t > 0)$$

Now, let $E_\lambda(g(T)) = 0$ for all λ. Then, $\int_0^\infty t.g(t).e^{-\lambda.t}dt = 0$ for all $\lambda > 0$. A change of variable $e^{-t} = x$ yields:

$$\int_0^1 [\frac{1}{x}\log(\frac{1}{x})g(\log(\frac{1}{x}))]x^\lambda dx = 0 \ \ \texttt{for all} \ \ \lambda > 0. \ \ (***)$$

Now an application of Stone-Weierstrass Theorem for the case of polynomials on (∗∗∗) yields $\frac{1}{x}\log(\frac{1}{x})g(\log(\frac{1}{x}))$ $= 0 \ \ (0 < x < 1)$ implying $g(\log(\frac{1}{x})) = 0 \ \ (0 < x < 1)$ or equivalently: $g \equiv 0$. Consequently, T is complete. Finally, by Basu's Theorem in Part (a) it follows that S and T are independent.

(c) As a special case of Example 6.9, $\overline{X}$ is a complete sufficient statistics for μ with σ^2 known. Next, we show that $T-\overline{X}$ is ancillary statistics for μ with σ^2 known. To see this, as $X_i-\mu \sim N(0, \sigma^2)$ $(1 \leq i \leq n)$ is ancillary for μ with σ^2 known, it follows that:

$$T(X_1, \cdots, X_n)-\frac{1}{n}.\sum_{i=1}^n X_i = (T(X_1, \cdots, X_n)-\mu)-(\frac{1}{n}.\sum_{i=1}^n X_i-\mu) = T(X_1-\mu, \cdots, X_n-\mu)-\frac{1}{n}.\sum_{i=1}^n (X_i-\mu),$$

is ancillary statistics for μ with σ^2 known. Now, by Basu's Theorem $\overline{X}$ and $T - \overline{X}$ are independent. Accordingly, by $E(T - \overline{X}) = \mu - \mu = 0$, we have:

$$Var(T) = Var(T - \overline{X}) + Var(\overline{X}) = E((T - \overline{X})^2) + Var(\overline{X}).$$

□

Problem 6.11. Suppose that $X_1, \cdots, X_n$ are i.i.d. Poisson random variables with mean λ.
(a)Use the fact that

$$\sum_{k=0}^\infty c_k.x^k = 0 \ \ \texttt{for all} \ \ a < x < b$$

if, and only if, $c_0 = c_1 = \cdots = 0$ to show that $T = \sum_{i=1}^n X_i$ is complete for λ.
(b) Find the unique UMVU estimator of λ^2.
(c) Find the unique UMVU estimator of λ^r for any integer $r > 2$.

Solution. (a) Suppose $E_\lambda(g(T)) = 0$ for all $\lambda > 0$ where $T = \sum_{i=1}^n X_i =^d Poisson(n\lambda)$. Then, by assumption:

$$\sum_{k=0}^\infty g(k)e^{-n\lambda}\frac{(n\lambda)^k}{k!} = e^{-n\lambda}[\sum_{k=0}^\infty g(k)\frac{(n\lambda)^k}{k!}] = 0, \ \ \texttt{for all} \ \ \lambda > 0.$$

Now, take $x = n\lambda$ in the given assumption, then by above equation:

$$\sum_{k=0}^\infty \frac{g(k)}{k!}x^k = 0, \ \ \texttt{for all} \ \ x > 0.$$

Hence, $\frac{g(k)}{k!} = 0, \ \ 0 \leq k \leq \infty$, implying $g \equiv 0$. Accordingly, T is complete statistics.

(b),(c) As

$$f(\mathbf{x}, \lambda) = \prod_{i=1}^{n}(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}) = (e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}).(\frac{1}{\prod_{i=1}^{n} x_i!}) = g^*(T(\mathbf{x}); \lambda).h^*(\mathbf{x})$$

by Theorem 4.2, it follows that $T = \sum_{i=1}^{n} X_i$ is sufficient for λ as well. Next, by Theorem 6.1. and Theorem 6.4 it is sufficient to find a function g such that $E(g(T)) = \lambda^r$, or $\sum_{k=0}^{\infty} g(k).\frac{e^{-n\lambda}(n.\lambda)^k}{k!} = \lambda^r$ or equivalently:

$$\sum_{k=0}^{\infty}(\frac{g(k)}{k!}n^k)\lambda^k = e^{n\lambda}.\lambda^r = \sum_{k=0}^{\infty}\frac{(n\lambda)^k}{k!}.\lambda^r = \sum_{k=r}^{\infty}(\frac{n^{k-r}}{(k-r)!})\lambda^k,$$

implying:

$$\begin{aligned}\frac{g(k)}{k!}n^k &= 0: \quad k = 0, \cdots, r-1\\ \frac{g(k)}{k!}n^k &= \frac{n^{k-r}}{(k-r)!}: \quad k = r, \cdots,\end{aligned}$$

and hence:

$$g(k) = \frac{r!.C(k,r)}{n^r} * 1_{[r,\infty)}(k): \quad k = 0, 1, 2, \cdots.$$

□

Problem 6.13. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ has a joint distribution that depends on an unknown parameter θ and define

$$\mathcal{U} = \{U : E_\theta(U) = 0, E_\theta(U^2) < \infty\}$$

to be the space of all statistics $U = U(\mathbf{X})$ that are unbiased estimators of 0 with finite variance.
(a) Suppose that $T = T(\mathbf{X})$ is an unbiased estimator of $g(\theta)$ with $Var_\theta(T) < \infty$. Show that any unbiased estimator S of $g(\theta)$ with $Var_\theta(S) < \infty$ can be written as

$$S = T + U$$

for some $U \in \mathcal{U}$.
(b) Let T be an unbiased estimator of $g(\theta)$ with $Var_\theta(T) < \infty$. Suppose that $cov_\theta(T, U) = 0$ for all $U \in \mathcal{U}$ (an all θ). Show that T is a UMVU estimator of $g(\theta)$.
(c) Suppose that T is a UMVU estimator of $g(\theta)$. Show that $Cov_\theta(T, U) = 0$ for all $U \in \mathcal{U}$.

Solution. (a) Let $U = S - T$. Then:

$$\begin{aligned}E_\theta(U) &= E_\theta(S) - E_\theta(T) = g(\theta) - g(\theta) = 0,\\ E_\theta(U^2) &= Var_\theta(U) = Var_\theta(S-T) = Var_\theta(S) - 2.Cov_\theta(S,T) + Var_\theta(T)\\ &= Var_\theta(S) - 2.Corr_\theta(S,T).\sqrt{Var_\theta(S).Var_\theta(T)} + Var_\theta(T)\\ &< \infty.\end{aligned}$$

Accordingly: $U \in \mathcal{U}$.

(b) Let S be another estimator satisfying $E_\theta(S) = g(\theta) = E_\theta(T)$. Then, by Part (a), $S - T \in \mathcal{U}$ and ;

furthermore:

$$\begin{aligned} Var_\theta(S) &= Var_\theta(T + (S - T)) \\ &= Var_\theta(T) + 2.Cov_\theta(T, (S - T)) + Var_\theta(S - T) \\ &= Var_\theta(T) + Var_\theta(S - T) \\ &\geq Var_\theta(T). \end{aligned}$$

Thus, UMVU$(g(\theta)) = T$.

(c) Let for some $U \in \mathcal{U}$, $Cov_\theta(T, U) \neq 0$, say $Cov_\theta(T, U) < 0$, (for the case $Cov_\theta(T, U) > 0$, the proof is the similar by replacing $-U$ instead of U.). Define $S_\lambda = T + \lambda.U$. then:

$$\begin{aligned} E_\theta(S_\lambda) &= E_\theta(T) + \lambda.E(U) = g(\theta) + \lambda.0 = g(\theta), \\ Var_\theta(S_\lambda) &= Var_\theta(T) + 2.\lambda.Cov_\theta(T, U) + \lambda^2.Var_\theta(U). \end{aligned}$$

Now, for $\lambda \in (0, \frac{-2Cov_\theta(U,V)}{Var_\theta(U)})$ we have:

$$Var_\theta(S_\lambda) < Var_\theta(T),$$

implying UMVU$(g(\theta)) = S_\lambda$, a contradiction.
□

Problem 6.15. Suppose that $X_1, \cdots, X_n$ are i.i.d. Normal random variables with mean θ and variance θ^2 where $\theta > 0$. Define:

$$\widehat{\theta_n} = \overline{X_n}(1 + \frac{\sum_{i=1}^n (X_i - \overline{X_n})^2 - n\overline{X_n}^2}{3\sum_{i=1}^n (X_i - \overline{X_n})^2})$$

where $\overline{X_n}$ is the sample mean of $X_1, \cdots, X_n$.
(a) Show that $\widehat{\theta_n} \to_p \theta$ as $n \to \infty$.
(b) Find the asymptotic distribution of $\sqrt{n}(\widehat{\theta_n} - \theta)$. Is $\widehat{\theta_n}$ asymptotically efficient?
(c) Find the Cramer-Rao lower bound for unbiased estimators of θ. (Assume all regularity conditions are satisfied.)
(d) Does there exist an unbiased estimator of θ that achieves the lower bound in (a)? Why or why not?

Solution. (a) We have

$$\widehat{\theta_n} = \overline{X_n}.[1 + \frac{S_n^2 - \frac{n}{n-1}\overline{X_n}^2}{3.S_n^2}], \quad (n \geq 1). \quad (*)$$

Next, by Example 4.19, $S_n^2 \to_p \theta^2$, by Theorem 3.6., $\overline{X_n} \to_p \theta$, and by Theorem $\frac{n}{n-1}.\overline{X_n}^2 \to_p \theta^2$. Accordingly, applying these results in $(*)$ it follows that $\widehat{\theta_n} \to_p \theta$.

(b) First, by $\overline{X_n} \to_p \theta$, $S_n^2 \to_p \theta^2$:

$$\begin{aligned}
\sqrt{n}(\widehat{\theta_n} - \theta) &= \sqrt{n}(\overline{X_n}.[1 + \frac{S_n^2 - \frac{n}{n-1}\overline{X_n}^2}{3.S_n^2}] - \theta) \\
&= \sqrt{n}.(\overline{X_n} - \theta) \\
&\quad -\sqrt{n}.(\overline{X_n}^2 - \theta^2)(\frac{1}{3\theta})(\frac{\frac{n}{n-1}\overline{X_n}^2 - \theta^2}{\overline{X_n}^2 - \theta^2})(\frac{\frac{\overline{X_n}}{3.S_n^2}}{\frac{1}{3\theta}}) \\
&\quad +\sqrt{n}(S_n^2 - \theta^2)).(\frac{1}{3\theta})(\frac{\frac{\overline{X_n}}{3.S_n^2}}{\frac{1}{3\theta}}) \\
&=^d \sqrt{n}.((\overline{X_n} - \frac{1}{3\theta}\overline{X_n}^2) - (\theta - \frac{\theta^2}{3\theta})) \\
&\quad +\sqrt{n}(\frac{1}{3\theta}S_n^2 - \frac{1}{3\theta}\theta^2)) \ \ (n \to \infty). \quad (**)
\end{aligned}$$

Next, using Example 5.14 for $\mu = \theta$ and $\sigma^2 = \theta^2$ we have:

$$\sqrt{n}(\overline{X_n} - \theta) \to_d N(0, \theta^2), \quad \sqrt{n}(S_n - \theta) \to_d N(0, \frac{\theta^2}{2}),$$

(with independent limits) and an application of Theorem 3.4, for $g_1(x) = x - \frac{1}{3\theta}x^2$ and $g_2(x) = \frac{1}{\theta}x^2$ on the last result yields:

$$\sqrt{n}.((\overline{X_n} - \frac{1}{3\theta}\overline{X_n}^2) - (\theta - \frac{\theta^2}{3\theta})) \to_d N(0, \frac{\theta^2}{9}), \quad \sqrt{n}(\frac{1}{3\theta}S_n^2 - \frac{1}{3\theta}\theta^2) \to_d N(0, \frac{2.\theta^2}{9}), \quad (***)$$

(with independent limits). Accordingly, by $(**)$, $(***)$, and an application of Theorem 3.3 it follows that:

$$\begin{aligned}
\sqrt{n}(\widehat{\theta_n} - \theta) &\to_d N(0, \frac{\theta^2}{9}) + N(0, \frac{2.\theta^2}{9}) \\
&=^d N(0, \frac{1}{3}\theta^2). \ \ (\dagger)
\end{aligned}$$

Finally, considering $(\dagger)$ with $\sigma^2(\theta) = \frac{1}{3}\theta^2$,and $I(\theta) = Var(l'(x;\theta)) = Var_\theta(-\theta^{-1} + \theta^{-3}.X^2 - \theta^{-2}.X) = \frac{11}{\theta^2}$ (Exercise!) we have:

$$\sigma^2(\theta) = \frac{\theta^2}{3} > \frac{\theta^2}{11} = \frac{1}{I(\theta)}.$$

Consequently, this sequence of estimators is not asymptotically efficient.

(c) First, let $X \sim N(\theta, \theta^2)$ then, $E(X) = \theta, E(X^2) = 2\theta^2, E(X^3) = 4.\theta^3, E(X^4) = 10.\theta^4$. Second, referring to Pages 324-325 let $X_1, \cdots, X_n$ be i.i.d. random variables with pdf $f(x;\theta)$ from exponential family. Then, for $E_\theta(T) = g(\theta)$:

$$CRLB_\theta(T(\mathbf{X})) = \frac{(g'(\theta))^2}{E_\theta(\frac{d}{d\theta}\log(f(\mathbf{x};\theta))^2)} = \frac{(g'(\theta))^2}{n.Var_\theta(\frac{d}{d\theta}\log(f(x;\theta)))} \quad (\dagger\dagger)$$

In particular, for $g(\theta) = \theta$, T with $E_\theta(T) = \theta$, $\log(f(x;\theta)) = -\frac{1}{2}(\log(2\pi) + 1) - \log(\theta) - \frac{1}{2}.\theta^{-2}x^2 + \theta^{-1}.x$

and $\frac{d}{d\theta}\log(f(x;\theta)) = -\theta^{-1} + \theta^{-3}.x^2 - \theta^{-2}.x$ it follows from (††) that:

$$\begin{aligned} CRLB_\theta(T(\mathbf{X})) &= \frac{1}{n.Var(-\theta^{-1} + \theta^{-3}.X^2 - \theta^{-2}.X)} \\ &= \frac{1/n}{\theta^{-4}Var(\theta^{-1}.X^2 + X)} \quad (Exercise!) \\ &= \frac{\frac{\theta^4}{n}}{11\theta^2} \\ &= \frac{\theta^2}{11n}. \end{aligned}$$

(d) Here, by Example 4.6.:

$$\log(f(\mathbf{x};\theta)) = \frac{-1}{2\theta^2}\sum_{i=1}^{n} x_i^2 + \frac{1}{\theta}.\sum_{i=1}^{n} x_i - \frac{n}{2}(1 + 2.\log(\theta) + \log(2\pi))$$

is a two parameter exponential family. Hence, as the one parameter exponential family representation in page 324 does not exist; and, it follows that there is no unbiased estimator T achieving the Cramer-Rao lower bound.
□

Problem 6.17. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with frequency function

$$f(x;\theta) = \theta, \ \ \texttt{for}\ x = -1, \ (1-\theta)^2.\theta^x \quad \texttt{for}\ x = 0, 1, 2, \cdots$$

where $0 < \theta < 1$.
(a) Find the Cramer-Rao lower bound for unbiased estimators of θ based on $X_1, \cdots, X_n$.
(b) Show that the maximum likelihood estimator of θ based on $X_1, \cdots, X_n$ is

$$\widehat{\theta_n} = \frac{2\sum_{i=1}^{n} I(X_i = -1) + \sum_{i=1}^{n} X_i}{2n + \sum_{i=1}^{n} X_i}$$

and show that $\{\widehat{\theta_n}\}$ is consistent for θ.
(c) Show that $\sqrt{n}(\widehat{\theta_n} - \theta) \to_d N(0, \sigma^2(\theta))$ and find the value of $\sigma^2(\theta))$. Compare $\sigma^2(\theta))$ to the Cramer-Rao lower bound found in part (a).

Solution. (a) By Page 327 for $g(\theta) = \theta$ and T with $E_\theta(T) = \theta$ we have:

$$CRLB_\theta(T) = \frac{(g^{'}(\theta))^2}{n.Var_\theta(\frac{d\log(f(x;\theta))}{d\theta})} = \frac{1}{n.Var_\theta(\frac{d\log(f(x;\theta))}{d\theta})}. \quad (*)$$

Next, we have:

$$\begin{aligned}
\log(f(x;\theta)) &= \log(\theta).1_{X=-1} + [2.\log(1-\theta) + x.\log(\theta)].1_{X\geq 0},\\
\frac{d\log(f(x;\theta))}{d\theta} &= (\frac{1}{\theta}).1_{X=-1} + (\frac{2}{\theta-1} + \frac{x}{\theta}).1_{X\geq 0},\\
E_\theta((\frac{d\log(f(x;\theta))}{d\theta})^2) &= (\frac{2}{\theta-1})^2 + \frac{4}{(\theta-1)\theta}E_\theta(X) + \frac{1}{\theta^2}.E_\theta(X^2) + (\frac{1}{\theta^2} - (\frac{2}{\theta-1} - \frac{1}{\theta})).\theta\\
&= (\frac{2}{\theta-1})^2 + \frac{4}{(\theta-1)\theta}.0 + \frac{1}{\theta^2}.\frac{2\theta}{1-\theta} + \frac{-4}{(\theta-1)^2}\\
&= \frac{2}{\theta.(1-\theta)},\\
E_\theta^2(\frac{d\log(f(x;\theta))}{d\theta}) &= 0,\\
Var_\theta(\frac{d\log(f(x;\theta))}{d\theta}) &= E_\theta((\frac{d\log(f(x;\theta))}{d\theta})^2) - E_\theta^2(\frac{d\log(f(x;\theta))}{d\theta}) = \frac{2}{\theta.(1-\theta)}. \quad (**)
\end{aligned}$$

Thus, by $(*)$ and $(**)$ it follows that $CRLB_\theta(T) = \frac{\theta.(1-\theta)}{2n}$.

(b) Refer to the solution of Problem 5.5(a).

(c) Refer to the solution of Problem 5.5(b). Also:

$$\sigma^2(\theta) = \frac{\theta.(1-\theta)}{2} \geq \frac{\theta.(1-\theta)}{2n} = CRLB(T) \quad (n \geq 1).$$

□

Problem 6.19. Suppose that $X_1, \cdots, X_n$ be i.i.d. Bernoulli random variables with parameter θ.
(a) Indicate why $S = X_1 + \cdots + X_n$ is a sufficient and complete statistics for θ.
(b) Find the UMVU estimator of $\theta.(1-\theta)$.

Solution. (a) As $f(x;\theta) = \exp(\log(\theta/(1-\theta)).x - \log(1/(1-\theta))), \quad (0 < \theta < 1)$ it follows that:

$$f(\mathbf{x};\theta) = \exp(\log(\theta/(1-\theta)).\sum_{i=1}^n x_i - \log(1/(1-\theta))), \quad (0 < \theta < 1)$$

and the assertion follows from Theorem 6.3. for $k = 1$, $c_1(\theta) = \log(\theta/(1-\theta))$, $T_1(\mathbf{X}) = \sum_{i=1}^n X_i$, $d(\theta) = n.\log(1/(1-\theta))$, and $C = (0,1)$.

(b) Let $S \sim Bin(n,\theta)$, then by Theorem 6.1 and Theorem 6.4. it is sufficient to find h such that $E_\theta(h(S)) = \theta.(1-\theta), \quad (0 < \theta < 1)$. Next, let $n = 2$ in the equation

$$\sum_{k=0}^n h(k).C(n,k).\theta^k.(1-\theta)^{n-k} = \theta.(1-\theta), \quad (0 < \theta < 1.$$

Then, after re-arranging powers of θ it follows that $h(0) = 0, h(1) = 1/2, h(2) = 0$. Hence, for this special case: $h(S) = \frac{S.(2-S)}{2}$. Now, let $n > 2$ and set $S^* = I(X_1 = 0, X_2 = 1)$ implying:

$$E_\theta(S) = P(X_1 = 0).P(X_1 = 1) = (1-\theta).\theta.$$

Accordingly, by Theorem 6.4., $S^{**} = E(S^*|S_n)$: $S_n = S$ is the unique UMVU estimator of $(1-\theta).\theta$. To calculate S^{**} we have:

$$\begin{aligned} S^{**}(s) &= E_\theta(S^*|S_n = s) = P_\theta(X_1 = 0, X_2 = 1|S_n = s) \\ &= \frac{P_\theta(X_1 = 0, X_2 = 1, X_3 + \cdots + X_n = s-1)}{P_\theta(S_n = s)} \\ &= 0 \quad \text{if} \quad s = 0, \\ &= \frac{P_\theta(X_1 = 0).P_\theta(X_2 = 1).P(s_{n-2} = s-1)}{P_\theta(S_n = s)} \\ &= \frac{(1-\theta).\theta.C(n-2, s-1).\theta^{s-1}.(1-\theta)^{n-2-(s-1)}}{C(n,s).\theta^s.(1-\theta)^{n-s}} \\ &= \frac{s.(n-s)}{n.(n-1)}, \end{aligned}$$

yielding:

$$S^{**} = \frac{S.(n-S)}{n.(n-1)}.$$

□

Problem 6.21. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with density or frequency function $f(x;\theta)$ satisfying the condition of Theorem 6.6. Let $\widehat{\theta_n}$ be the MLE of θ and $\widetilde{\theta_n}$ be another (regular)estimator of θ such that

$$\sqrt{n}\begin{pmatrix} \widehat{\theta_n} - \theta \\ \widetilde{\theta_n} - \theta \end{pmatrix} \to_d N_2(\mathbf{0}, C(\theta)).$$

Show that $C(\theta)$ must have the form

$$C(\theta) = \begin{pmatrix} I^{-1}(\theta) & I^{-1}(\theta) \\ I^{-1}(\theta) & \sigma^2(\theta) \end{pmatrix}.$$

Solution. By two times application of Theorem 6.6 for sequences of estimators $(\widehat{\theta_n})_{n=1}^\infty$ and $(\widetilde{\theta_n})_{n=1}^\infty$ we have:

$$\begin{aligned} \sqrt{n}(\widehat{\theta_n} - \theta) \to_d Z_1(\theta) \quad &: \quad Z_1(\theta) = N(0, \frac{1}{I(\theta)}) \\ \sqrt{n}(\widetilde{\theta_n} - \theta) \to_d Z_1(\theta) + Z_2(\theta) \quad &: \quad Z_2(\theta) \ \text{ independent } \ Z_1(\theta). \end{aligned}$$

Consequently, by assumption:

$$\begin{aligned} c_{11}(\theta) &= Var(Z_1(\theta)) = \frac{1}{I(\theta)}, \\ c_{22}(\theta) &= Var(Z_1(\theta) + Z_2(\theta)) = \sigma^2(\theta), \\ c_{12}(\theta) &= Cov(Z_1(\theta), Z_1(\theta) + Z_2(\theta)) = Var(Z_1(\theta)) = \frac{1}{I(\theta)}, \\ c_{21}(\theta) &= c_{12}(\theta) = \frac{1}{I(\theta)}. \end{aligned}$$

□

Chapter 7

Interval Estimation and Hypothesis Testing

Problem 7.1. Suppose that $X_1, \cdots, X_n$ are i.i.d. Normal random variables with unknown mean μ and variance σ^2.
(a) Using pivot $(n-1)S^2/\sigma^2$ where

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

we can obtain a 95% confidence interval $[k_1.S^2, k_2.S^2]$ for some constants k_1 and k_2. Find expressions for k_1 and k_2 if this confidence interval has minimum length. Evaluate k_1 and k_2 when $n = 10$.
(b) When n is sufficiently large, we can approximate the distribution of the pivot by a normal distribution (Why ?). Find approximations for k_1 and k_2 that are valid for large n.

Solution. (a) Let $X \sim \chi^2_{n-1}$ with p.d.f $f_{X,n-1}$ and the condition

$$P(a \le X \le b) = 1 - \alpha \quad (a < b). \qquad (*)$$

Then the interval $I_{(a,b)} = [a, b]$ with smallest length satisfying $(*)$ has the following constraint on its bounds (Tate & Klett, 1959):

$$f_{X,n+3}(a) = f_{X,n+3}(b).$$

Now, by Example 7.4. $\frac{(n-1).S^2}{\sigma^2} \sim \chi^2_{n-1}$ and for $k_1 = \frac{n-1}{b}, k_2 = \frac{n-1}{a}$ we have:

$$P(k_1.S^2 \le \sigma^2 \le k_2.S^2) = P(a \le \frac{(n-1).S^2}{\sigma^2} \le b) = P(a \le \chi^2_{n-1} \le b) = 0.95. \quad (**)$$

By $(*)$, the required a, b have the constraint $a^{\frac{n+1}{2}}.e^{-\frac{a}{2}} = b^{\frac{n+1}{2}}.e^{-\frac{b}{2}}$. Next, taking $n = 10$ we will have the following system of equations:

$$\begin{aligned} P(a \le \chi^2_9 \le b) &= 0.95 \\ a^{\frac{11}{2}}.e^{-\frac{a}{2}} &= b^{\frac{11}{2}}.e^{-\frac{b}{2}}, \end{aligned}$$

with solution $a = 3.284$, and $b = 26.077$. Thus, $k_1 = \frac{9}{26.077} = 0.345$ and $k_2 = \frac{9}{3.284} = 2.741$.

(b) By Example 5.14, for $\tilde{\sigma_n^2} = \frac{n-1}{n}S^2$ we have $\sqrt{n}.(\tilde{\sigma_n} \to \sigma) \to_d N(0, \frac{\sigma^2}{2})$. Define $g(x) = x^2$, then, by Theorem 3.4. $\sqrt{n}.(\tilde{\sigma_n^2} \to \sigma^2) \to_d N(0, 2.\sigma^4)$. Thus,

$$\frac{(n-1).S^2}{\sigma^2} = \frac{n.\tilde{\sigma_n^2}}{\sigma^2} \sim \frac{n}{\sigma^2}.N(\sigma^2, \frac{2.\sigma^4}{n}) =^d N_{(n,2n)}. \quad (n \uparrow \infty) \quad (***)$$

But, $\frac{(n-1).S^2}{\sigma^2} \sim \chi^2_{n-1}$, and by $(***)$ it follows that as $n \to \infty$:

$$P(a \le \chi^2_{n-1} \le b) = P(a \le N_{(n,2n)} \le b) = P(\frac{a-n}{\sqrt{2n}} \le N_{(0,1)} \le \frac{b-n}{\sqrt{2n}}) = 0.95. \quad (****)$$

On the other hand, $|I_{(a,b)}| = b - a = \sqrt{2n} * ((\frac{b-n}{\sqrt{2n}}) - (\frac{a-n}{\sqrt{2n}}))$ implying that the length of the desired interval is minimized when $\frac{a-n}{\sqrt{2n}} = -1.96$ and $\frac{b-n}{\sqrt{2n}} = 1.96$ or, equivalently, $a = -1.96.\sqrt{2n} + n$ and $b = 1.96.\sqrt{2n} + n$. Accordingly:

$$k_1 = \frac{n-1}{b} = \frac{n-1}{1.96.\sqrt{2n}+n}, \qquad k_2 = \frac{n-1}{a} = \frac{n-1}{-1.96.\sqrt{2n}+n}.$$

□

Problem 7.3. Suppose that $X_1, \cdots, X_n$ are i.i.d. continuous random variables with median θ.
(a) What is the distribution of $\sum_{i=1}^n I(X_i \le \theta)$?
(b) Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics of $X_1, \cdots, X_n$. Show that the interval $[X_{(l)}, X_{(u)}]$ is a $100.p\%$ confidence interval for θ and find an expression for p in terms of l and u.
(c) Suppose that for large n, we set

$$l = \lfloor \frac{n}{2} - 0.98 * \sqrt{n} \rfloor \quad \text{and} \quad u = \lceil \frac{n}{2} + 0.98 * \sqrt{n} \rceil.$$

Show that the confidence interval $[X_{(l)}, X_{(u)}]$ has coverage approximately 95%.

Solution. (a) As $P(I_{X_i \le \theta} = 1) = P(X_i \le \theta) = \frac{1}{2}$, it follows that $I_{X_i \le \theta} \sim Binomial(1, \frac{1}{2})$ $(1 \le i \le n)$, and by independence of $I_{X_i \le \theta}$ $(1eqi \le n)$ it follows that, $\sum_{i=1}^n I_{X_i \le \theta} \sim Binomial(n, \frac{1}{2})$.

(b) By an application of Problem 2.25 with $F_X(\theta) = 1 - F_X(\theta) = \frac{1}{2}$, we have:

$$\begin{aligned} p &= p(l,u) = P(X_{(l)} \le \theta \le X_{(u)}) = P(X_{(l)} \le \theta) - P(X_{(u)} < \theta) \\ &= \sum_{k=l}^{n} C(n,k).F_X(\theta)^k.(1-F_X(\theta))^{n-k} - \sum_{k=u}^{n} C(n,k).F_X(\theta)^k.(1-F_X(\theta))^{n-k} \\ &= \sum_{k=l}^{u-1} C(n,k).F_X(\theta)^k.(1-F_X(\theta))^{n-k} = \frac{\sum_{k=l}^{u-1} C(n,k)}{2^n}, \quad (l < u). \end{aligned}$$

(c) Let $Y = \sum_{i=1}^n I(X_i \le \theta) \sim Binomial(n, \frac{1}{2})$, then by part (a) and Theorem 3.8, $\frac{Y - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \sim N(0,1)$ as $n \to \infty$. Consequently:

$$\begin{aligned} P_\theta(X_{(l)} \le \theta \le X_{(u)}) &= P_\theta(X_{(l)} \le \theta) - P_\theta(X_{(u)} \le \theta) \\ &= P_\theta(l \le Y) - P_\theta(u \le Y) = (1 - P_\theta(Y < l)) - (1 - P_\theta(Y < u)) = P_\theta(l \le Y < u) \\ &= P_\theta(\frac{l - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \le Y < \frac{u - \frac{n}{2}}{\sqrt{\frac{n}{4}}}) \\ &\simeq P(-1.96 \le Z \le +1.96) = 0.95, \end{aligned}$$

implying

$$\frac{l - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \simeq -1.96, \qquad \frac{u - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \simeq +1.96,$$

or equivalently the assertion.
□

Problem 7.5. Suppose that $X_1, \cdots, X_n$ are i.i.d. Uniform random variables on $[0, \theta]$ and $X_{(1)}, \cdots, X_{(n)}$ be the order statistics.
(a) Show that for any r, $X_{(r)}/\theta$ is a pivot for θ.
(b) Use part (a) to derive a 95% confidence interval for θ based on $X_{(r)}$. Give the exact upper and lower confidence limits where $n = 10$ and $r = 5$.

Solution. (a) By Problem 2.25 (b) the p.d.f of $X_{(r)}/\theta$ is calculated as follows:

$$\begin{aligned} f_{X_{(r)}/\theta}(t) &= \theta.f_{X_{(r)}}(\theta.t) = \theta.r.C(n,r).(F_X(\theta.t))^{r-1}.(1 - F_X(\theta.t))^{n-r}.f_X(\theta.t) \\ &= \frac{\Gamma(n+1)}{\Gamma(r).\Gamma(n-r+1)}.t^{r-1}.(1-t)^{n-r+1-1}, \quad (0 \le t \le 1), \end{aligned}$$

yielding, $X_{(r)}/\theta \sim Beta(r, n-r+1)$.

(b) By part (a), $X \sim Beta(r, n-r+1)$ and $F_X(x) = I_x(r, n-r+1)$ we have:

$$0.95 = P_\theta(\frac{X_{(r)}}{b} \le \theta \le \frac{X_{(r)}}{a}) = P_\theta(a.\theta \le X_{(r)} \le b.\theta) = P_\theta(a \le \frac{X_{(r)}}{\theta} \le b) = I_b(r, n-r+1) - I_a(r, n-r+1).$$

In particular, for $n = 10$, and $r = 5$ we have: $0.95 = I_b(5,6) - I_a(5,6)$, and one choice for (a, b) can be $(a, b) = (0.20, 0.76)$, giving a 95% confidence interval for θ as $[\frac{X_{(5)}}{0.76}, \frac{X_{(5)}}{0.20}]$.
□

Problem 7.7. Suppose that $X_1, X_2, \cdots$ are i.i.d. Normal random variables with mean μ and variance σ^2, both unknown. With a fixed sample size, it is not possible to find a fixed length $100p.\%$ confidence interval for μ. However, it is possible to construct a fixed length confidence interval by allowing a random sample size. Suppose that $2d$ is the desired length of the confidence interval. Let n_0 be a fixed integer with $n_0 \ge 2$ and define

$$\overline{X_0} = \frac{1}{n_0}\sum_{i=1}^{n_0} X_i, \quad \text{and} \quad S_0^2 = \frac{1}{n_0 - 1}\sum_{i=1}^{n_0}(X_i - \overline{X_0})^2.$$

Now, given S_0^2, define a random integer N to be the smallest integer greater or equal than n_0 and greater than or equal to $[S_0.t_\alpha/d]^2$ where $\alpha = (1-p)/2$ and t_α is the $1-\alpha$ quantile of a t-distribution with $n_0 - 1$ degrees of freedom. Sample $N - n_0$ additional random variables and let $\overline{X} = N^{-1}.\sum_{i=1}^{N} X_i$.
(a) Show that $\sqrt{N}(\overline{X} - \mu)/S_0$ has $t-$ distribution with $n_0 - 1$ degrees of freedom.
(b) Use the result of part (a) to construct a $100p\%$ confidence interval for μ and show that this interval has length at most $2d$.

Solution. (a) Similar to Example 2.17, we have:

$$\begin{aligned}
\frac{\sqrt{N}(\overline{X_N}-\mu)}{S_0} &= \frac{(\sqrt{N}(\overline{X_N}-\mu))/(\sigma/\sqrt{N})}{\sqrt{S_0^2/\sigma^2}} \\
(\sqrt{N}(\overline{X_N}-\mu))/(\sigma/\sqrt{N}) &\sim N(0,1) \\
(n_0-1)S_0^2/\sigma^2 &\sim \chi^2_{n_0-1}
\end{aligned}$$

in which the later two distributions are independent. Hence, by definition of T distribution, $\sqrt{N}(\overline{X}-\mu)/S_0 \sim \mathcal{T}_{n_0-1}$.

(b) First, by Part (a):

$$\begin{aligned}
p &= P_\mu(-t_{(1-p)/2,n_0-1} \le \frac{\sqrt{N}(\overline{X}-\mu)}{S_0} \le t_{(1-p)/2,n_0-1}) \\
&= P_\mu(\overline{X} - \frac{t_{(1-p)/2,n_0-1}}{\sqrt{N}}.S_0 \le \mu \le \overline{X} + \frac{t_{(1-p)/2,n_0-1}}{\sqrt{N}}.S_0),
\end{aligned}$$

and take $I_{L,U} = [L(\overline{X}), U(\overline{X})] = [\overline{X} - \frac{t_{(1-p)/2,n_0-1}}{\sqrt{N}}.S_0, \overline{X} + \frac{t_{(1-p)/2,n_0-1}}{\sqrt{N}}.S_0]$.

Second, for $N = \max(n_0, [S_0.t_{(1-p)/2,n_0-1}/d]^2)$, it follows that:

$$|I_{L,U}| = 2 * \frac{t_{(1-p)/2,n_0-1}}{\sqrt{N}}.S_0 \le 2 * \frac{t_{(1-p)/2,n_0-1}}{S_0.t_{(1-p)/2,n_0-1}/d}.S_0 = 2d.$$

□

Problem 7.9. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with density function

$$f(x;\mu) = \lambda.\exp[-\lambda.(x-\mu)] \quad \text{for} \quad x \ge \mu.$$

Let $X_{(1)} = \min(X_1, \cdots, X_n)$.
(a) Show that

$$S(\lambda) = 2\lambda.\sum_{i=1}^{n}(X_i - X_{(1)}) \sim \chi^2(2(n-1))$$

and hence is a pivot for λ.
(b) Describe how to use the pivot in (a) to give an exact 95% confidence interval for λ.
(c) Give an approximate 95% confidence interval for λ based on $S(\lambda)$ for large n.

Solution. (a) As $X_i^* = 2\lambda(X_i-\mu) \sim \exp(\frac{1}{2})$ $(1 \le i \le n)$ are i.i.d., and $X_{(i)}^* =^d 2\lambda(X_{(i)}-\mu)$ $(1 \le i \le n)$, a re-arrangement of equations in Problem 2.26 yields:

$$\begin{aligned}
X_{(1)}^* &= \frac{1}{n}.Y_{(1)}^* \\
X_{(2)}^* &= \frac{1}{n}.Y_{(1)}^* + \frac{1}{n-1}.Y_{(2)}^* \\
\cdots & \cdots \cdots \\
X_{(n-1)}^* &= \frac{1}{n}.Y_{(1)}^* + \frac{1}{n-1}.Y_{(2)}^* + \cdots + \frac{1}{2}.Y_{(n-1)}^* \\
X_{(n)}^* &= \frac{1}{n}.Y_{(1)}^* + \frac{1}{n-1}.Y_{(2)}^* + \cdots + \frac{1}{2}.Y_{(n-1)}^* + Y_{(n)}^* \quad (*),
\end{aligned}$$

in which $Y^*_{(i)} \sim \chi^2_{(2)}$ $(1 \le i \le n)$. Consequently, by $(*)$ it follows that:

$$\begin{aligned} S(\lambda) &= \sum_{i=1}^{n}((2\lambda(X_i-\mu)) - (2\lambda.(X_{(1)}-\mu))) \\ &= \sum_{i=1}^{n}(X^*_{(i)} - X^*_{(1)}) = \sum_{i=1}^{n} X^*_{(i)} - n.X^*_{(1)} = \sum_{i=1}^{n} Y^*_{(i)} - Y^*_{(1)} \\ &= \sum_{i=2}^{n} Y^*_{(i)} \sim \chi^2(2(n-1)). \end{aligned}$$

(b) As

$$\begin{aligned} 0.95 &= P_\lambda(a \le S(\lambda) \le b) = P_\lambda(a \le 2\lambda.\sum_{i=1}^{n}(X_i - X_{(1)}) \le b) \\ &= P_\lambda(\frac{a}{2.\sum_{i=1}^{n}(X_i - X_{(1)})} \le \lambda \le \frac{b}{2.\sum_{i=1}^{n}(X_i - X_{(1)})}), \end{aligned}$$

for $a = \chi^2_{2(n-1),0.975}$ and $b = \chi^2_{2(n-1),0.025}$ it follows that:

$$[U(\mathbf{X}), V(\mathbf{X})] = [\frac{\chi^2_{2(n-1),0.975}}{2.\sum_{i=1}^{n}(X_i - X_{(1)})}, \frac{\chi^2_{2(n-1),0.025}}{2.\sum_{i=1}^{n}(X_i - X_{(1)})}].$$

(c) An application of Theorem 3.8 for i.i.d. random variables $X^*_i \sim \chi^2_{(1)}$ $(i = 1, 2, \cdots)$ with $\mu^* = 1$ and $(\sigma^*)^2 = 2$, yields $\frac{\chi^2_{(m)} - m}{\sqrt{2m}} \to_d N(0,1)$. Thus as $n \to \infty$:

$$\begin{aligned} 0.95 &= P_\lambda(a \le S(\lambda) \le b) = P_\lambda(\frac{a-2(n-1)}{\sqrt{2.2.(n-1)}} \le \frac{S(\lambda)-2(n-1)}{\sqrt{2.2.(n-1)}} \le \frac{b-2(n-1)}{\sqrt{2.2.(n-1)}}) \\ &\simeq P(\frac{a-2(n-1)}{\sqrt{2.2.(n-1)}} \le N(0,1) \le \frac{b-2(n-1)}{\sqrt{2.2.(n-1)}}). \end{aligned}$$

Next, one choice will be $\frac{a-2(n-1)}{\sqrt{2.2.(n-1)}} = -1.96$ and $\frac{b-2(n-1)}{\sqrt{2.2.(n-1)}} = +1.96$, or:

$$a = 2.\sqrt{n-1}.(\sqrt{n-1} - 1.96), \qquad b = 2.\sqrt{n-1}.(\sqrt{n-1} + 1.96).$$

Accordingly,

$$[U(\mathbf{X}), V(\mathbf{X})] = [\frac{\sqrt{n-1}.(\sqrt{n-1} - 1.96)}{\sum_{i=1}^{n}(X_i - X_{(1)})}, \frac{\sqrt{n-1}.(\sqrt{n-1} + 1.96)}{\sum_{i=1}^{n}(X_i - X_{(1)})}].$$

□

Problem 7.11. Consider a random sample of n individuals who are classified into one of three groups with probabilities $\theta^2, 2\theta.(1-\theta)$, and $(1-\theta)^2$. If Y_1, Y_2, Y_3 are the numbers in each group then $Y = (Y_1, Y_2, Y_3)$ has a Multinomial distribution:

$$f(y;\theta) = \frac{n!}{y_1!y_2!y_3!}\theta^{2y_1}[2\theta(1-\theta)]^{y_2}.(1-\theta)^{2y_3}$$

for $y_1, y_2, y_2 \ge 0; y_1 + y_2 + y_3 = n$ where $0 < \theta < 1$. (This model is the Hardy-Weinberg equilibrium model from genetics.)

(a) Find the maximum likelihood estimator of θ and give the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ as $n \to \infty$.
(b) Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. Suppose that for some k:

$$P_{\theta_0}[2Y_1 + Y_2 \geq k] = \alpha.$$

Then the test that rejects H_0 when $2Y_1 + Y_2 \geq k$ is a UMP level α test of H_0 versus H_1.
(c) Suppose that n is large and $\alpha = 0.05$. Find an approximate value for k in the UMP test in part (b).
(d) Suppose that $\theta_0 = 1/2$ in part (b). How large must n so that a 0.05 level test has power at least 0.80 when $\theta = 0.6$?

Solution. (a) Let $\theta^* = 1 - \theta$, then $\widehat{\theta^*} = 1 - \widehat{\theta}$. In addition:

$$\begin{aligned}
f(\mathbf{y};\theta^*) &= \frac{n!}{y_1!y_2!y_3!}(1-\theta^*)^{2y_1}.(2.\theta^*.(1-\theta^*))^{y_2}.(\theta^*)^{2y_3} : \quad y_1, y_2, y_3 \geq 0, y_1 + y_2 + y_3 = n, \\
l^{'}(\theta^*) &= -\frac{2.Y_1 + Y_2}{1-\theta^*} + \frac{Y_2 + 2.Y_3}{\theta^*} = 0 \to \widehat{\theta^*} = \frac{Y_2 + 2.Y_3}{2n} \to \widehat{\theta} = \frac{2.Y_1 + Y_2}{2n}, \\
l^{''}(\theta^*) &= -\frac{2.Y_1 + Y_2}{(1-\theta^*)^2} - \frac{Y_2 + 2.Y_3}{(\theta^*)^2} : \ E(Y_1) = n.(1-\theta^*)^2, E(Y_2) = 2n.\theta^*.(1-\theta^*),\ E(Y_3) = n.(\theta^*)^2, \\
E(l^{'}(\theta^*)) &= 0, \\
E(l^{''}(\theta^*)) &= \frac{-2n}{\theta^*.(1-\theta^*)}.
\end{aligned}$$

Now, by comments on page 254, $I(\theta^*) = -E_\theta(l^{''}(\theta^*))|_{n=1} = \frac{2}{\theta^*.(1-\theta^*)} = \frac{2}{\theta.(1-\theta)}$, and hence $I(\theta) = \frac{2}{\theta.(1-\theta)}$. Now, by theorem 5.3:

$$\sqrt{n}.(\widehat{\theta}_n - \widehat{\theta}) \to_d N(0, \frac{1}{I(\theta)}) = N(0, \frac{\theta.(1-\theta)}{2}).$$

(b) Using equation $Y_2 + Y_3 = -(2.Y_1 + Y_2) + 2n$, we have:

$$\begin{aligned}
f(\mathbf{y};\theta) &= \exp[(2.y_1 + y_2).\log(\theta) + (y_2 + 2.y_3).\log(1-\theta) + (y_2.\log(2) + \log(\frac{n!}{y_1!y_2!y_3!}))] \\
&= \exp[(2.y_1 + y_2).\log(\theta) + (-(2y_1 + y_2) + 2n).\log(1-\theta) + (y_2.\log(2) + \log(\frac{n!}{y_1!y_2!y_3!}))] \\
&= \exp[(2.y_1 + y_2).\log(\frac{\theta}{1-\theta}) + 2n.\log(1-\theta) + (y_2.\log(2) + \log(\frac{n!}{y_1!y_2!y_3!}))].
\end{aligned}$$

By Example 7.15 for $c(\theta) = \log(\frac{\theta}{1-\theta})$, $T(y) = 2.y_1 + y_2$, $b(\theta) = -2n.\log(1-\theta)$ and $S(y) = y_2.\log(2) + \log(\frac{n!}{y_1!y_2!y_3!})$ the assertion follows.

(c) By Part (a), $\sqrt{n}.(\widehat{\theta}_n - \theta) \sim N(0, \frac{\theta.(1-\theta)}{2})$ where $\widehat{\theta}_n = \frac{2.y_1 + y_2}{2n}$, as $n \to \infty$. Consequently:

$$\begin{aligned}
0.05 &= P_{\theta_0}(2.Y_1 + Y_2 \geq k) = P_{\theta_0}(\widehat{\theta}_n \geq \frac{k}{n}) \\
&= P_{\theta_0}(\frac{\sqrt{n}(\widehat{\theta}_n - \theta_0)}{\sqrt{\theta.(1-\theta)/2}} \geq \frac{\sqrt{n}(\frac{k}{n} - \theta_0)}{\sqrt{\theta_0.(1-\theta_0)/2}}) \simeq P(Z \geq \frac{\sqrt{n}(\frac{k}{n} - \theta_0)}{\sqrt{\theta_0.(1-\theta_0)/2}}),
\end{aligned}$$

or equivalently $P(Z \leq \frac{\sqrt{n}(\frac{k}{n}-\theta_0)}{\sqrt{\theta_0.(1-\theta_0)/2}}) \simeq 0.95$. Hence, $\frac{\sqrt{n}(\frac{k}{n}-\theta_0)}{\sqrt{\theta_0.(1-\theta_0)/2}} \simeq 1.645$, implying:

$$k(\theta_0, n) \simeq n.(\frac{1.645 * \sqrt{\theta_0.(1-\theta_0)/2}}{\sqrt{n}} + \theta_0).$$

(d) First, with $\theta_0 = \frac{1}{2}$ we have,

$$k(\theta_0, n) = \frac{n}{2} * (\frac{1.645}{\sqrt{2n}} + 1). \qquad (*)$$

Second, given power level, for $\theta_1 = 0.60$ we have:

$$0.80 \leq P_{\theta_1}(2.Y_1 + Y_2 \geq k(\theta_0, n)) \simeq P_{\theta_1}(Z \geq \frac{\sqrt{n}(\frac{k(\theta_0,n)}{n} - \theta_1)}{\sqrt{\theta_1.(1-\theta_1)/2}}) \qquad \rightarrow$$

$$P_{\theta_1}(Z \leq \frac{\sqrt{n}(\frac{k(\theta_0,n)}{n} - \theta_1)}{\sqrt{\theta_1.(1-\theta_1)/2}}) \leq 0.20 \rightarrow \frac{\sqrt{n}(\frac{k(\theta_0,n)}{n} - \theta_1)}{\sqrt{\theta_1.(1-\theta_1)/2}} \leq -0.84. \qquad (**)$$

Accordingly, plugging in $(*)$ in $(**)$ we get $n \geq 77$.
□

Problem 7.13. Suppose that $X \sim Bin(m, \theta)$ and $Y \sim Bin(n, \phi)$ are independent random variables and consdier testing:

$$H_0 : \theta \geq \phi \text{ versus } H_1 : \theta < \phi.$$

(a) Show that the joint frequency function of X and Y can be written in the form

$$f(x, y; \theta, \phi) = (\frac{\theta.(1-\phi)}{\phi.(1-\theta)})^x.(\frac{\phi}{1-\phi})^{x+y}.\exp[d(\theta, \phi) + S(x, y)]$$

and that H_0 is equivalent to

$$H_0 : \ln(\frac{\theta.(1-\phi)}{\phi.(1-\theta)}) \geq 0.$$

(b) The UMPU test of H_0 versus H_1 rejects H_1 at level α if $X \geq k$ where k is determined from conditional distribution of X given $X + Y = z$ (assuming that $\theta = \phi$). show that this conditional distribution is Hypergeometric . (This conditional test is called Fisher's exact test.)
(c) Show that the conditional frequency function of X given $X + Y = z$ is given by

$$P(X = x | X + Y = z) = \frac{C(m, x).C(n, z-x)\psi^x}{\sum_s C(m, s)C(n, z-s)\psi^s}$$

where the summation extends over s from $\max(0, z-n)$ to $\min(m, z)$ and $\psi = \frac{\theta(1-\phi)}{\phi(1-\theta)}$. (This is called a non-central Hypergeometric distribution.)

Solution. (a) First,

$$\begin{aligned}
f(x, y; \theta, \phi) &= f(x; \theta).f(y; \phi) = C(m, x).\theta^x.(1-\theta)^{m-x} * C(n, y).\phi^y.(1-\phi)^{n-y} \\
&= (\frac{\theta}{1-\theta})^x * (\frac{\phi}{1-\phi})^y * [C(m, x).(1-\theta)^m.C(n, y).(1-\phi)^n] \\
&= (\frac{\theta.(1-\phi)}{(1-\theta).\phi})^x.(\frac{\phi}{1-\phi})^{x+y}.\exp[m.\log(1-\theta) + n.\log(1-\phi) + \log(C(m, x).C(n, y))].
\end{aligned}$$

Take $d(\theta, \phi) = m.\log(1-\theta) + n.\log(1-\phi)$ and $S(x, y) = \log(C(m, x).C(n, y))$.

Second, as $\theta \geq \phi$, it follows that $\frac{\theta}{\phi} \geq 1$ and $\frac{1-\phi}{1-\theta} \geq 1$, implying $(\frac{\theta}{\phi}).(\frac{1-\phi}{1-\theta}) \geq 1$, or equivalently $\ln((\frac{\theta}{\phi}).(\frac{1-\phi}{1-\theta})) \geq 0$.

(b) Under null hypothesis $\theta = \phi = p$, and hence:

$$\begin{aligned} P(X=x|X+Y=z) &= \frac{P(X=x, X+Y=z)}{P(X+Y=z)} = \frac{P(X=x, Y=z-x)}{P(X+Y=z)} = \frac{P(X=x).P(Y=z-x)}{P(X+Y=z)} \\ &= \frac{C(m,x).p^x.(1-p)^{m-x}.C(n,z-x).p^{z-x}.(1-p)^{n-z+x}}{C(m+n,z).p^z.(1-p)^{n+m-z}} = \frac{C(m,x).C(n,z-x)}{C(m+n,z)}. \end{aligned}$$

(c) Using Part (a) representation we have:

$$\begin{aligned} P(X=x|X+Y=z) &= \frac{P(X=x, Y=z-x)}{P(X+Y=z)} = \frac{P(X=x, Y=z-x)}{\sum_s P(X=s, Y=z-s)} \\ &= \frac{\psi^x.(\frac{\phi}{1-\phi})^z.\exp[m.\log(1-\theta) + n.\log(1-\phi) + \log(C(m,x).C(n,z-x))]}{\sum_s \psi^s.(\frac{\phi}{1-\phi})^z.\exp[m.\log(1-\theta) + n.\log(1-\phi) + \log(C(m,s).C(n,z-s))]} \\ &= \frac{\psi^x.C(m,x).C(n,z-x)}{\sum_s[\psi^s.C(m,s).C(n,z-s)]}. \end{aligned}$$

□

Problem 7.15. Suppose that $X_1, \cdots, X_{10}$ are i.i.d. Uniform random variables on $[0, \theta]$ and consider testing

$$H_0 : \theta = 1 \quad \text{versus} \quad H_1 : \theta \neq 1$$

at the 5% level. Consider a test that rejects H_0 if $X_{(10)} < a$ or $X_{(10)} > b$ where $a < b \leq 1$.
(a) Show that a and b must satisfy the equation

$$b^{10} - a^{10} = 0.95.$$

(b) Does an unbiased test of H_0 versus H_1 of this form exist ? If so, find a and b to make the test unbiased.

Solution. (a) By Example 7.2. for $n = 10$ and $Y = \frac{X_{(10)}}{\theta}$ we have, $F_Y(y) = y^{10}$ $(0 \leq y \leq 1)$. Hence,

$$0.05 = P(X_{(10)} < a \cup X_{(10)} > b|\theta = 1) = 1 - P(a \leq X_{(10)} \leq b|\theta = 1)$$

or $P(\frac{a}{\theta} \leq Y \leq \frac{b}{\theta}|\theta = 1) = P(a \leq Y \leq b) = 0.95$ or $b^{10} - a^{10} = 0.95$.

(b) There is such an unbiased test if and only if $\inf_{1 \neq \theta > 0} \pi(\theta) \geq \pi(1)$. But,

$$\begin{aligned} \pi(\theta) &= P_\theta(X_{(10)} < a \cup X_{(10)} > b) = 1 - P_\theta(a \leq X_{(10)} \leq b) \\ &= 1 - P_\theta(a/\theta \leq X_{(10)}/\theta \leq b/\theta) = 1 - (\frac{b^{10} - a^{10}}{\theta^{10}}) \\ &= 1 - \frac{0.95}{\theta^{10}}, \quad (\theta > 0). \end{aligned}$$

Thus,

$$\pi(0.95^{\frac{1}{10}}) = 0 < 0.05 = \pi(1),$$

and consequently, such unbiased test does not exist. However, if one replace the alternative hypothesis with $H_1 : \theta \geq 1$, such an unbiased test will exist.
□

Problem 7.17. Suppose that $X = (X_1, \cdots, X_n)$ are continuous random variables with joint density $f(x)$ where $f = f_0$ or $f = f_1$ are the two possibilities for f. Based on X, we want to decide between f_0 and f_1 using a Non-Neyman-Pearson approach. Let $\phi(X)$ be an arbitrary test function where f_0 is chosen if $\phi = 0$ and f_1 is chosen if $\phi = 1$. Let $E_0(T)$ and $E_1(T)$ be expectations of a statistics $T = T(X)$ assuming the true joint densities are f_0 and f_1 respectively.
(a) Show that the test function ϕ that minimizes $\alpha.E_0[\phi(X)] + (1-\alpha)E_1[1-\phi(X)]$ (where $0 < \alpha < 1$ is a known constant) has the form

$$\phi(X) = 1 \ \text{ if } \ \frac{f_1(X)}{f_0(X)} \geq k$$

and 0 otherwise. Specify the value of k.
(b) Suppose that $X_1, \cdots, X_n$ are i.i.d. continuous random variables with common density f where $f = f_0$ or $f = f_1 (f_0 \neq f_1)$. Let $\phi_n(X)$ be the optimal test function (for some α) based on $X_1, \cdots, X_n$ as described in part (a). Show that

$$. \lim_{n\to\infty} (\alpha.E_0[\phi_n(X)] + (1-\alpha)E_1[1-\phi_n(X)]) = 0.$$

Solution. (a) Take $k = \frac{\alpha}{1-\alpha}$ and consider another test function $\psi(X)$. By conditions $\phi(x) = 1_{f_1(x)-k.f_0(x)\geq 0}$ and $0 \leq \psi(x) \leq 1$ it follows that:

$$(\phi(x) - \psi(x)).(f_1(x) - k.f_0(x)) \geq 0 \ \text{ for all } \ x.$$

Then, by integration we have:

$$\begin{aligned}
\int (\phi(x)-\psi(x)).(f_1(x) - (\frac{\alpha}{1-\alpha}).f_0(x))dx \geq 0 \quad &\Leftrightarrow \\
\int (\phi(x)-\psi(x)).((1-\alpha).f_1(x) - (\alpha).f_0(x))dx \geq 0 \quad &\Leftrightarrow \\
(1-\alpha).\int (\phi(x)-\psi(x)).f_1(x)dx - (\alpha).\int (\phi(x)-\psi(x)).f_0(x)dx \geq 0 \quad &\Leftrightarrow \\
\int (\alpha).(\phi(x)-\psi(x)).f_0(x)dx + \int (1-\alpha).(\psi(x)-\phi(x)).f_1(x)dx \leq 0 \quad &\Leftrightarrow \\
E_0((\alpha).(\phi(X)-\psi(X))) + E_1((1-\alpha).(\psi(X)-\phi(X))) \leq 0 \quad &\Leftrightarrow \\
((\alpha).E_0(\phi(X)) + (1-\alpha).E_1(1-\phi(X))) - ((\alpha).E_0(\psi(X)) + (1-\alpha).E_1(1-\psi(X))) \leq 0 \quad &\Leftrightarrow \\
G(\phi) - G(\psi) \leq 0 \quad &\Leftrightarrow \\
G(\phi) \leq G(\psi).&
\end{aligned}$$

(b) First, let the sequence of random variables $X_n^* \quad (i = 1, 2, \cdots)$ and sequence of real numbers $a_n \quad (n = 1, 2, \cdots)$ satisfy the conditions $\lim_{n\to\infty} X_n^* =^d X^*$, and $\lim_{n\to\infty} a_n = a$, respectively. Then, given corresponding C.D.F's $F_{X_n^*}$, and F_{X^*}, it follows that (Exercise !):

$$\lim_{n\to\infty} F_{X_n^*}(a_n) = F_{X^*}(a). \qquad (*)$$

Second, by definition

$$E_0(\log(\frac{f_1}{f_0})) < 0, \text{and} \quad E_1(\log(\frac{f_1}{f_0})) > 0. \quad (**)$$

Third, for $\phi = 1$ we have:

$$\frac{1}{n}.\sum_{i=1}^{n}\log(\frac{f_1(x_i)}{f_0(x_i)}) = \frac{1}{n}.\log(\prod_{i=1}^{n}\frac{f_1(x_i)}{f_0(x_i)}) = \frac{1}{n}.\log(\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}) \geq \frac{\log(k)}{n},$$

and similarly for $\phi = 0$ we have:

$$\frac{1}{n}.\sum_{i=1}^{n}\log(\frac{f_1(x_i)}{f_0(x_i)}) \leq \frac{\log(k)}{n}.$$

Now, by Theorem 3.6 and two times application of $(*)$ we have:

$$\begin{aligned}
\lim_{n\to\infty}(\alpha.E_0[\phi_n(X)] + (1-\alpha)E_1[1-\phi_n(X)]) &= \lim_{n\to\infty}(\alpha.E_0[1_{f_1/f_0\geq k}|f = f_0] \\
&+ (1-\alpha)E_1[1_{f_1/f_0<k}|f = f_1]) \\
&= \lim_{n\to\infty}\alpha.P(\frac{1}{n}.\sum_{i=1}^{n}\log(\frac{f_1(x_i)}{f_0(x_i)}) \geq \frac{\log(k)}{n}|f = f_0) \\
&+ (1-\alpha).P(\frac{1}{n}.\sum_{i=1}^{n}\log(\frac{f_1(x_i)}{f_0(x_i)}) \leq \frac{\log(k)}{n}|f = f_1) \\
&= \alpha.P(E_0(\log(\frac{f_1}{f_0})) \geq 0) \\
&+ (1-\alpha).P(E_1(\log(\frac{f_1}{f_0})) \leq 0). \quad (***)
\end{aligned}$$

Finally, a comparison of $(**)$ and $(***)$ yields the desired result.
□

Problem 7.19. Consider a simple classification problem. An individual belongs to exactly one of k populations. Each population has a known density $f_i(x)(i = 1, \cdots, k)$ and it is known that a proportion p_i belong to population $i(p_1 + \cdots + p_k = 1)$. Given disjoint sets $R_1, \cdots, R_k$, a general classification rule is

classify as population i if $x \in R_i(i = 1, \cdots, k)$.

The total probability of correct classification is

$$C(R_1, \cdots, R_k) = \sum_{i=1}^{k} p_i \int_{R_i} f_i(x)dx.$$

We would like to find the classification rule (that is, the sets $R_1, \cdots, R_k$) that maximizes the total probability of correct classification.
(a) Suppose that $k = 2$. Show that the optional classification rule has

$$R_1 = \{x : \frac{f_1(x)}{f_2(x)} \geq \frac{p_2}{p_1}\} \quad R_2 = \{x : \frac{f_1(x)}{f_2(x)} < \frac{p_2}{p_1}\}.$$

(b) Suppose that f_1 and f_2 are Normal densities with different means but equal variances. find the optional classification rule using the result of part (a) (that is, find the regions R_1 and R_2.).
(c) Find the form of the optimal classification rule for general k.

Solution. (a) First, by $p_1 + p_2 = 1$ and $1_{R_1} + 1_{R_2} = 1$, it follows that:

$$\begin{aligned} C(R_1, R_2) &= \int (p_1.f_1(x).1_{R_1}(x))dx + \int (p_2.f_2(x).1_{R_2}(x))dx \\ &= \int ((p_1.f_1(x) - p_2.f_2(x)).1_{R_1}(x) + p_2.f_2(x))dx. \quad (*) \end{aligned}$$

Second, let R_1^*, R_2^* be two other disjoint sets with union $\mathbb{R}$. Then, for $k = \frac{p_2}{p_1}$, we have(Exercise !):

$$(1_{R_1}(x) - 1_{R_1^*}(x)).(f_1(x) - k.f_2(x)) \geq 0 \;\text{ for all }\; x.$$

Consequently:

$$\int (1_{R_1}(x) - 1_{R_1^*}(x)).(f_1(x) - (\frac{p_2}{p_1}).f_2(x))dx \geq 0 \;\Leftrightarrow$$

$$\int (1_{R_1}(x) - 1_{R_1^*}(x)).(p_1.f_1(x) - p_2.f_2(x))dx \geq 0 \;\Leftrightarrow$$

$$\int (1_{R_1}(x)).(p_1.f_1(x) - p_2.f_2(x))dx \geq \int (1_{R_1^*}(x)).(p_1.f_1(x) - p_2.f_2(x))dx \;\Leftrightarrow$$

$$\int ((p_1.f_1(x) - p_2.f_2(x)).1_{R_1}(x) + p_2.f_2(x))dx \geq \int ((p_1.f_1(x) - p_2.f_2(x)).1_{R_1^*}(x) + p_2.f_2(x))dx. \;\; (**)$$

Accordingly, by $(*)$ and $(**)$ it follows that:

$$C(R_1, R_2) \geq C(R_1^*, R_2^*).$$

(b)As $f_i(x) = \frac{1}{\sqrt{2.\pi\sigma}}.\exp(\frac{-(x-\mu_i)^2}{2.\sigma^2})$, $\;(\mu_1 \neq \mu_2)$, and $(x-\mu_1)^2 - (x-\mu_2)^2 = (2x - (\mu_1+\mu_2)).(-\mu_1+\mu_2)$, it follows that:

$$\begin{aligned} R_1 &= \{x|\frac{f_1(x)}{f_2(x)} \geq \frac{p_2}{p_1}\} \\ &= \{x|\exp(\frac{-1}{2.\sigma^2}((x-\mu_1)^2 - (x-\mu_2)^2)) \geq \frac{p_2}{p_1}\} \\ &= \{x|2.(\mu_1 - \mu_2).x \geq (\mu_1^2 - \mu_2^2) + 2.\sigma^2.\log(\frac{p_2}{p_1})\} \\ &= \{x|x \geq \frac{(\mu_1^2 - \mu_2^2) + 2.\sigma^2.\log(\frac{p_2}{p_1})}{2.(\mu_1 - \mu_2)}\}, \;\text{ if }\; \mu_1 > \mu_2, \\ &= \{x|x \leq \frac{(\mu_1^2 - \mu_2^2) + 2.\sigma^2.\log(\frac{p_2}{p_1})}{2.(\mu_1 - \mu_2)}\}, \;\text{ if }\; \mu_1 < \mu_2, \end{aligned}$$

and take $R_2 = \mathbb{R} - R_1$.

(c) As for $k = 2$ we have, $R_1 = \{x|p_1.f_1(x) - p_2.f_2(x) \geq 0\}$, taking $L_0(x) = 0$, $\;L_1(x) = p_1.f_1(x) - p_2.f_2(x)$ it follows that:

$$R_1 = \{x|L_1(x) \geq L_0(x)\}.$$

Consequently, for case $k > 2$ we may set (Floudas & Pardalos, 2009):

$$R_i = \{x|L_i(x) \geq L_j(x) \;\; (0 \leq j \leq i)\} \quad i = 1, 2, \cdots, k$$

in which for some $(q_{ij})_{j=1}^k$ we have:

$$\begin{aligned} L_0(x) &= 0, \\ L_i(x) &= p_i.f_i(x) - \sum_{1 \leq j \neq i \leq k} q_{ij}.f_j(x), \qquad i = 1, 2, \cdots, k. \end{aligned}$$

□

Problem 7.21. A heuristic (but almost rigorous) proof of Theorem 7.5. can be given by using the fact that the log-likelihood function is approximately quadratic in a neighbourhood of the true parameter value. Suppose that we have i.i.d. random variables $X_1, \cdots, X_n$ with density or frequency function $f(x;\theta)$ where $\theta = (\theta_1, \cdots, \theta_p)$, define

$$Z_n(u) = \ln(\mathcal{L}_n(\theta + u/\sqrt{n})/\mathcal{L}_n(\theta)) = u^T.V_n - \frac{1}{2}u^T I(\theta)u + R_n(u)$$

where $R_n(u) \to_p 0$ for each u and $V_n \to_d N_p(0, I(\theta))$.
(a) Suppose that we want to test the null hypothesis $H_0 : \theta_1 = \theta_{10}, \cdots, \theta_r = \theta_{r0}$. Show that, if H_0 is true, the LR statistic is $2\ln(\Lambda_n) = 2[Z_n(\widehat{U_n}) - Z_n(\widehat{U_{n0}})]$ where $\widehat{U_n}$ maximizes $Z_n(u)$ and $\widehat{U_{n0}}$ maximizes $Z_n(u)$ subject to the constraint that $u_1 = \cdots = u_1 = 0$.
(b) Suppose that $Z_n(u)$ is exactly quadratic (that is $R_n(u) = 0$). show that

$$\widehat{U_n} = I^{-1}(\theta)V_n \qquad \widehat{U_{n0}} = \begin{pmatrix} 0 \\ I_{22}^{-1}(\theta)V_{n2} \end{pmatrix}$$

where V_n and $I(\theta)$ are expressed as

$$V_n = \begin{pmatrix} V_{n1} \\ V_{n2} \end{pmatrix} \qquad I(\theta) = \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix}.$$

(c) Assuming that nothing is lost asymptotically in using the quadratic approximation, deduce Theorem 7.5. from parts (a) and (b).

Solution. (a) By definition and $\widehat{\theta_n} = \theta + \frac{\widehat{U_n}}{\sqrt{n}}$ for some $\widehat{U_n}$ and $\widehat{\theta_{n0}} = \theta + \frac{\widehat{U_{n0}}}{\sqrt{n}}$ for some $\widehat{U_{n0}}$ it follows that:

$$\begin{aligned} 2.\log(\Lambda_n) &= 2.\log(\frac{L_n(\widehat{\theta_n})}{L_n(\widehat{\theta_{n0}})}) = 2.\log(\frac{L_n(\theta + \frac{\widehat{U_n}}{\sqrt{n}})}{L_n(\theta + \frac{\widehat{U_{n0}}}{\sqrt{n}})}) = 2.\log(\frac{L_n(\theta + \frac{\widehat{U_n}}{\sqrt{n}})/L_n(\theta)}{L_n(\theta + \frac{\widehat{U_{n0}}}{\sqrt{n}})/L_n(\theta)}) \\ &= 2.\log(\frac{L_n(\theta + \frac{\widehat{U_n}}{\sqrt{n}})}{L_n(\theta)}) - 2.\log(\frac{L_n(\theta + \frac{\widehat{U_{n0}}}{\sqrt{n}})}{L_n(\theta)}) = 2[Z_n(\widehat{U_n}) - Z_n(\widehat{U_{n0}})]. \end{aligned}$$

(b) Let $Z_n(u) = U^T.V_n - \frac{1}{2}.U^T.I(\theta).U$. Then:

$$\frac{dZ_n(U)}{dU} = V_n^T - \frac{1}{2} * (2.U^T.I(\theta)) = V_n^T - U^T.I(\theta) = 0 \Rightarrow U^T.I(\theta) = V_n^T, \text{ or } U^T = V_n^T.I(\theta)^{-1}.$$

As $I(\theta)$ and $I^{-1}(\theta)$ are both symmetric we have:

$$\widehat{U_n} = (V_n^T.I(\theta)^{-1})^T = (I(\theta)^{-1})^T.V_n = I(\theta)^{-1}.V_n.$$

The second assertion follows similarly by consideration a projection P.

(c) Let $X \sim N_p(\mu_{p\times 1}, C_{p\times p})$. Then, a necessary and sufficient condition for the random variable $(X-\mu)^T.D.(X-\mu)$ to have a chi-square distribution with r degrees of freedom (in which $r = rank(DC)$) is that (Rao, 1973): $CDCDC = CDC$.

Now, using $V_n \rightarrow_d N_p(0, I(\theta))$ as $n \rightarrow \infty$ and the fact that:

$$\begin{aligned}
2.\log(\Lambda_n) &= 2[Z_n(\widehat{U_n}) - Z_n(\widehat{U_{n0}})] \\
&= 2[(\widehat{U_n}^T.V_n - \frac{1}{2}.\widehat{U_n}^T.I(\theta).\widehat{U_n}) - (\widehat{U_{n0}}^T.V_n - \frac{1}{2}.\widehat{U_{n0}}^T.I(\theta).\widehat{U_{n0}})] \\
&= 2[(V_n^T.I(\theta)^{-1}.V_n - \frac{1}{2}.V_n^T.I(\theta)^{-1}.I(\theta).I(\theta)^{-1}.V_n) - (V_n^T.P(\theta).V_n - \frac{1}{2}.V_n^T.P(\theta).I(\theta).P(\theta)^T.V_n)] \\
&= V_n^T.[I(\theta)^{-1} + P(\theta).(I(\theta).P(\theta) - 2.I)].V_n,
\end{aligned}$$

and taking $D = I(\theta)^{-1} + P(\theta).(I(\theta).P(\theta) - 2.I)$ and $C = I(\theta)$ in above mentioned Statement the assertion follows.

□

Problem 7.23. Suppose that $X_1, \cdots, X_n$ are independent Exponential random variables with $E(X_i) = \beta.t_i$ where $t_1, \cdots, t_n$ are known positive constants and β is unknown parameter.
(a) Show that the MLE of β is $\widehat{\beta_n} = \frac{1}{n}\sum_{i=1}^n X_i/t_i$.
(b) Show that $\sqrt{n}(\widehat{\beta} - \beta) \rightarrow_d N(0, \beta^2)$.
(c) Suppose we want to test $H_0 : \beta = 1$ versus $H_1 : \beta \neq 1$. show that the LR test of H_0 versus H_1 rejects H_0 for large values of

$$T_n = n(\widehat{\beta_n} - \ln(\widehat{\beta_n}) - 1)$$

where $\widehat{\beta_n}$ is defined as in part (a).
(d) Show that when H_0 is true, $2T_n \rightarrow_d \chi^2(1)$.

Solution. (a) As: $l(\beta; \mathbf{x}) = \sum_{i=1}^n(\log(\frac{1}{\beta.t_i}.e^{-\frac{x_i}{\beta.t_i}})) = \sum_{i=1}^n[-\log(\beta.t_i) - \frac{x_i}{\beta.t_i}]$, it follows that:

$$\frac{d}{d\beta}l(\beta; \mathbf{x}) = \frac{1}{\beta}.[-n + \frac{1}{\beta}.\sum_{i=1}^n \frac{x_i}{t_i}] = 0 \Rightarrow \widehat{\beta_n} = \frac{\sum_{i=1}^n \frac{x_i}{t_i}}{n}.$$

(b) Define $X_i^* = \frac{X_i}{t_i}, \quad i = 1, 2, \cdots$, then X_i^*'s are i.i.d. random variables with $E(X_i^*) = \beta$, and $Var(X_i^*) = \beta^2$. Hence, by Theorem 3.8 and Part (a):

$$\frac{\widehat{\beta_n} - \beta}{\beta/\sqrt{n}} \rightarrow_d N(0, 1), \quad (n \rightarrow \infty)$$

and the assertion follows.

(c) By Part (a) and:

$$\Lambda_n = \prod_{i=1}^n(\frac{f(X_i; \widehat{\beta_n})}{f(X_i; 1)}) = \prod_{i=1}^n \frac{e^{-x_i/t_i(1/\widehat{\beta_n}-1)}}{\widehat{\beta_n}} = \frac{e^{-\sum_{i=1}^n(x_i/t_i).(1/\widehat{\beta_n}-1)}}{\widehat{\beta_n}^n} = (\widehat{\beta_n})^{-n}.e^{-n+n.\widehat{\beta_n}},$$

it follows that:

$$T_n = \log(\Lambda_n) = -n.\log(\widehat{\beta_n}) - n + n.\widehat{\beta_n} = n.(\widehat{\beta_n} - \log(\widehat{\beta_n}) - 1).$$

(d) This is a direct consequence of Theorem 7.4. in which:

$$\begin{aligned}
l(\beta; x) &= -\log(\beta.t) - \frac{x}{\beta.t}, \quad l'(\beta; x) = -\frac{1}{\beta} + \frac{x}{\beta^2.t}, \quad l''(\beta; x) = \frac{1}{\beta^2} - \frac{2.x}{\beta^3.t} \\
E_\beta(l'(\beta; x)) &= 0, \quad E_\beta(l''(\beta; x)) = \frac{-1}{\beta^2}, \quad Var_\beta(l'(\beta; x)) = Var_\beta(-\frac{1}{\beta} + \frac{x}{\beta^2.t}) = \frac{1}{\beta^2},
\end{aligned}$$

and $I(\beta) = Var_\beta(l'(\beta; x)) = \frac{1}{\beta^2} = -E_\beta(l''(\beta; x)) = J(\beta)$.

□

Chapter 8

Linear and Generalized Linear Models

Problem 8.1. Suppose that $Y = X\beta + \epsilon$ where $\epsilon \sim N_n(0, \sigma^2.I)$ and X is $n \times (p+1)$. Let $\widehat{\beta}$ be the least squares estimator of β.
(a) Show that $S^2 = \frac{\|Y - X.\widehat{\beta}\|^2}{n-p-1}$ is an unbiased estimator of σ^2.
(b) Suppose that the random variable in ϵ are uncorrelated with common variance σ^2. show that S^2 is an unbiased estimator of σ^2.
(c) Suppose that $(X^T.X)^{-1}$ can be written as

$$(X^T.X)^{-1} = \begin{pmatrix} c_{00} & \cdots & c_{0p} \\ c_{11} & \cdots & c_{1p} \\ \cdots & \cdots & \cdots \\ c_{p0} & \cdots & c_{pp} \end{pmatrix}.$$

Show that $\frac{\widehat{\beta_j} - \beta_j}{S.\sqrt{c_{jj}}} \sim \mathcal{T}(n-p-1)$ for $j = 0, 1, \cdots, p$.

Solution. (a) First, let $\theta = X.\beta$ with $rank(X) = p+1$. Then, for $H = X.(X^T.X)^{-1}X^T$, we have $Y - \widehat{\theta} = (I_n - H).Y$. Thus:

$$(n-(p+1)).S^2 = Y^T.(I_n - H)^T.(I_n - H).Y = Y^T.(I_n - H)^2.Y = Y^T.(I_n - H).Y. \quad (*)$$

Second, by $H\theta = \theta$, it follows that $rank(I_n - H) = tr(I - H) = n - (p+1)$, and:

$$\begin{aligned} E(Y^T.(I_n - H).Y) &= tr((I_n - H).Var(Y)) + E(Y)^T(I_n - H).E(Y) \\ &= \sigma^2.tr(I_n - H) + \theta^T.(I_n - H).\theta = \sigma^2.(n-(p+1)). \quad (**) \end{aligned}$$

Thus, by $(*)$ and $(**)$ we have:

$$\begin{aligned} E(S^2) &= \frac{1}{n-(p+1)}(n-(p+1)).E(S^2) = \frac{1}{n-(p+1)}E((n-(p+1).S^2) \\ &= \frac{1}{n-(p+1)}E(Y^T.(I_n - H).Y) = \frac{1}{n-(p+1)}.\sigma^2.(n-(p+1)) = \sigma^2. \end{aligned}$$

(b) By Proposition 8.1(b) we have $n.\widehat{\sigma^2}/\sigma^2 \sim \chi^2(n-(p+1))$, and hence $E(n.\widehat{\sigma^2}/\sigma^2) = n-(p+1)$, implying: $E(S^2) = E(\frac{n}{n-(p+1)}\widehat{\sigma^2}) = \sigma^2$.

(c) By definition, for independent random variables Z, V with $Z \sim N(0,1)$ and $V \sim \chi^2(m)$, we have $T = \frac{T}{\sqrt{V/m}} \sim \mathcal{T}(m)$. Now, by Proposition 8.1. take independent random variables $\frac{\widehat{\beta_j}-\beta_j}{\sqrt{\sigma^2.c_{jj}}} \sim N(0,1)$ and $\frac{n.\widehat{\sigma^2}}{\sigma^2} = \frac{(n-(p+1))S^2}{\sigma^2} \sim \chi^2(n-(p+1))$ and $m = n-(p+1)$, it follows that:

$$\frac{\widehat{\beta_j}-\beta_j}{\sqrt{S^2.c_{jj}}} = \frac{\frac{\widehat{\beta_j}-\beta_j}{\sqrt{\sigma^2.c_{jj}}}}{\sqrt{\frac{[\frac{(n-(p+1))S^2}{\sigma^2}]}{n-(p+1)}}} \sim \mathcal{T}(n-(p+1)).$$

□

Problem 8.3. Consider the linear model $Y_i = \beta_0 + \beta_1.x_{i1} + \cdots + \beta_p.x_{ip} + \epsilon_i \quad (i = 1, \cdots, n)$ where for $j = 1, \cdots, p$ we have $\sum_{i=1}^n x_{ij} = 0$.
(a) Show that the least squares estimator of β_0 is $\widehat{\beta_0} = \overline{Y}$.
(b) Suppose that, in addition, we have $\sum_{i=1}^n x_{ij}.x_{ik} = 0$ for $1 \le j \ne k \le p$. Show that the least squares estimator of β_j is $\widehat{\beta_j} = \frac{\sum_{i=1}^n x_{ij}.Y_i}{\sum_{i=1}^n x_{ij}^2}$.

Solution. (a) As $\widehat{\beta} = (X^T.X)^{-1}.X^T.Y$ in which

$$\widehat{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_p \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix},$$

we may conclude that:

$$\begin{aligned}
\begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} &= \left(\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1j} & x_{2j} & \cdots & x_{nj} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix} \times \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} & \cdots x_{1p} \\ 1 & x_{21} & \cdots & x_{2k} & \cdots x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nk} & \cdots x_{np} \end{pmatrix} \right)^{-1} \\
&* \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1j} & x_{2j} & \cdots & x_{nj} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} \\
&= \left(\begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots \end{pmatrix} \right)^{-1} * \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}.y_i \\ \cdots \\ \sum_{i=1}^n x_{ip}.y_i \end{pmatrix} \\
&= \left(\begin{pmatrix} \frac{1}{n} & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots \end{pmatrix} \right) \times \begin{pmatrix} n.\overline{y} \\ \sum_{i=1}^n x_{i1}.y_i \\ \cdots \\ \sum_{i=1}^n x_{ip}.y_i \end{pmatrix},
\end{aligned}$$

implying $\widehat{\beta_0} = \frac{1}{n} * n.\overline{y} = \overline{y}$.

(b) In this case, we have:

$$\begin{pmatrix}\beta_0\\ \beta_1\\ \cdots\\ \beta_j\\ \cdots\\ \beta_p\end{pmatrix} = (\begin{pmatrix} n & 0 & 0 & \cdots & \cdots & 0\\ 0 & \sum_{i=1}^n x_{i1}^2 & 0 & \cdots & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & & \\ 0 & 0 & \cdots & \sum_{i=1}^n x_{ij}^2 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & & \\ 0 & 0 & \cdots & 0 & & \sum_{i=1}^n x_{ip}^2\end{pmatrix})^{-1} * \begin{pmatrix}\sum_{i=1}^n y_i\\ \sum_{i=1}^n x_{i1}.y_i\\ \cdots\\ \sum_{i=1}^n x_{ij}.y_i\\ \cdots\\ \sum_{i=1}^n x_{ip}.y_i\end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n} & 0 & 0 & \cdots & \cdots & 0\\ 0 & \frac{1}{\sum_{i=1}^n x_{i1}^2} & 0 & \cdots & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & & \\ 0 & 0 & \cdots & \frac{1}{\sum_{i=1}^n x_{ij}^2} & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & & \\ 0 & 0 & \cdots & 0 & & \frac{1}{\sum_{i=1}^n x_{ip}^2}\end{pmatrix}) * \begin{pmatrix}\sum_{i=1}^n y_i\\ \sum_{i=1}^n x_{i1}.y_i\\ \cdots\\ \sum_{i=1}^n x_{ij}.y_i\\ \cdots\\ \sum_{i=1}^n x_{ip}.y_i\end{pmatrix}$$

$$= \begin{pmatrix}\frac{\sum_{i=1}^n y_i}{n}\\ \frac{\sum_{i=1}^n x_{i1}.y_i}{\sum_{i=1}^n x_{i1}^2}\\ \cdots\\ \frac{\sum_{i=1}^n x_{ij}.y_i}{\sum_{i=1}^n x_{ij}^2}\\ \cdots\\ \frac{\sum_{i=1}^n x_{ip}.y_i}{\sum_{i=1}^n x_{ip}^2}\end{pmatrix},$$

yielding, $\widehat{\beta}_j = \frac{\sum_{i=1}^n x_{ij}.y_i}{\sum_{i=1}^n x_{ij}^2}$.

□

Problem 8.5. Suppose that $Y = \theta + \epsilon$ where θ satisfies $A\theta = 0$ for some known $q \times n$ matrix A having rank q. Define $\widehat{\theta}$ to minimize $\|Y - \theta\|^2$ subject to $A\theta = 0$. Show that $\widehat{\theta} = (I - A^T(A.A^T)^{-1}.A)Y$.

Solution. First, as $A.A^T$ is positive-definite and non-singular it is invertible, too. Second, using the method of Lagrange Multipliers, define $r(\theta) = \|Y - \theta\|^2 + \theta^T.A^T.\lambda$. Then:

$$\frac{d}{d\theta}r(\theta) = 2.\frac{d}{d\theta}(Y-\theta).(Y-\theta) + A^T.\lambda = 0 \Rightarrow \widehat{\theta_H} = Y - \frac{1}{2}.A^T.\lambda. \quad (*)$$

Next, estimating both sides of $(*)$ under A it follows that $0 = A\widehat{\theta_H} = A(Y - \frac{1}{2}.A^T.\lambda.) = AY - \frac{1}{2}A.A^T.\lambda$ or $A.Y = A.A^T.(\frac{\lambda}{2})$. Hence, by invertibility of $A.A^T$:

$$(A.A^T)^{-1}.A.Y = \frac{\lambda}{2}. \quad (**)$$

Accordingly, by $(*)$ and $(**)$ it follows that:

$$\widehat{\theta_H} = Y - A^T.(A.A^T)^{-1}.A.Y = [I - A^T.(A.A^T)^{-1}.A.] * Y.$$

□

Problem 8.7. (a) Suppose that $U \sim \chi^2(1, \theta_1^2)$ and $V \sim \chi^2(1, \theta_2^2)$ where $\theta_1^2 > \theta_2^2$. Show that U is stochastically greater than V.

(b) Suppose that $U_n \sim \chi^2(n, \theta_1^2)$ and $V_n \sim \chi^2(n, \theta_2^2)$ where $\theta_1^2 > \theta_2^2$. Show that U_n is stochastically greater than V_n.

Solution. (a) For standard normal distribution Z, in which $X - |\theta_i| \sim Z \quad (i = 1, 2)$, we have:

$$\begin{aligned} P(U > x) \geq P(V > x) \quad &\Leftrightarrow \quad P(X^2 > x|\theta_1^2) \geq P(X^2 > x|\theta_2^2) \\ &\Leftrightarrow \quad P(|X| > \sqrt{x}||\theta_1|) \geq P(|X| > \sqrt{x}||\theta_2|) \\ &\Leftrightarrow \quad P(|X| \leq \sqrt{x}||\theta_1|) \leq P(|X| \leq \sqrt{x}||\theta_2|) \\ &\Leftrightarrow \quad P(|Z + \theta_1| \leq \sqrt{x}) \leq P(|Z + \theta_2| \leq \sqrt{x}) \\ &\quad\quad\ \text{for all} \quad x > 0. \end{aligned}$$

Hence, considering standard normal C.D.F, Φ and p.d.f, f, it is enough to prove the function

$$g(\theta; x) = P(|Z + \theta| \leq \sqrt{x}) = \Phi(-\theta + \sqrt{x}) - \Phi(-\theta - \sqrt{x}), \quad \theta > 0$$

is decreasing. The proof is complete by considering the fact that

$$\frac{d}{d\theta} g(\theta; x) = -f(-\theta + \sqrt{x}) + f(-\theta - \sqrt{x}) \leq 0 \Leftrightarrow f(-(\theta + \sqrt{x})) = f(\theta + \sqrt{x}) \leq f(\sqrt{x} - \theta) : \quad \sqrt{x} - \theta \leq \sqrt{x} + \theta.$$

(b) Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ be two sets of independent random variables in which Y_i is stochastically smaller than X_i, with notation $Y_i \leq_{st} X_i \quad (1 \leq i \leq n)$. Then (Belzunce, Martineze & Mulero, 2016),

$$\sum_{i=1}^n Y_i \leq_{st} \sum_{i=1}^n X_i.$$

Now, by Part (a) and an application of above statement with $X_i^* = X_i^2 \sim \chi^2(1, \theta_1^2)$ and $Y_i^* = Y_i^2 \sim \chi^2(1, \theta_2^2) \quad (1 \leq i \leq n)$ it follows that:

$$V_n = \sum_{i=1}^n Y_i^* \leq_{st} \sum_{i=1}^n X_i^* = U_n.$$

□

Problem 8.9. Suppose that $Y = X\beta + \epsilon$ where $\epsilon \sim N_n(0, \sigma^2.I)$ and define the ridge estimator (Hoerl and Kennard, 1970) $\widehat{\beta_\lambda}$ to minimize $\|Y - X.\beta\|^2 + \lambda.\|\beta\|^2$ for some $\lambda > 0$. (Typically in practice, the columns of X are centred and scaled, and Y is centred.)

(a) Show that

$$\widehat{\beta_\lambda} = (X^T.X + \lambda.I)^{-1} X^T.Y = (I + \lambda.(X^T.X)^{-1})^{-1} \widehat{\beta}$$

where $\widehat{\beta}$ is the least squares estimator of β. Conclude that $\widehat{\beta_\lambda}$ is a biased estimator of β.

(b) Consider estimating $\theta = a^T.\beta$ for some known $a \neq 0$. Show that $MSE_\theta(a^T.\widehat{\beta_\lambda}) \leq MSE_\theta(a^T.\widehat{\beta})$ for some $\lambda > 0$.

Solution. (a) Let $G(\beta) \in \mathbb{R}^n$, then: $\frac{d}{d\beta}(\|G(\beta)\|^2) = 2(\frac{d}{d\beta}G(\beta)).G(\beta)$. Hence, by two times application of this rule it follows that:

$$\begin{aligned} \frac{d}{d\beta}(\|Y - X.\beta\|^2 + \lambda.\|\beta\|^2) \quad &= \quad 2.(\frac{d}{d\beta}(Y - X.\beta)).(Y - X.\beta) + \lambda.2.(\frac{d}{d\beta}\beta).\beta \\ &= \quad 2.(-X)^T.(Y - X.\beta) + 2.\lambda.I.\beta \\ &= \quad 2 * [-X^T.Y + (X^T.X + \lambda.I)\beta] = 0, \end{aligned}$$

and by comments on Page 407 it is implying :

$$\begin{aligned}\widehat{\beta_\lambda} &= (X^T.X+\lambda.I)^{-1}.X^T.Y \\ &= (X^T.X+\lambda.I)^{-1}.(X^T.X.\widehat{\beta}) \\ &= ((X^T.X)^{-1}.(X^T.X+\lambda.I))^{-1}.(\widehat{\beta}) \\ &= (I+\lambda.(X^T.X)^{-1})^{-1}.(\widehat{\beta}).\end{aligned}$$

(b)Take $U=(a^T.\widehat{\beta_\lambda}-a^T.\beta).(X^T.X)^{-1}.(a^T.\widehat{\beta_\lambda})$ and $V=(X^T.X)^{-1}.(a^T.\widehat{\beta_\lambda})$. Then, for $\lambda \geq \frac{-2.E_\theta(U)}{E_\theta(V)}$ it follows that:

$$0 \leq 2.\lambda.E_\theta(U)+\lambda^2.E_\theta(V). \quad (*)$$

Accordingly, by $(*)$ and considering the fact that $\widehat{\beta_\lambda}=(I+\lambda.(X^T.X)^{-1})^{-1}.(\widehat{\beta})$ implies $\widehat{\beta}=(I+\lambda.(X^T.X)^{-1}).(\widehat{\beta_\lambda})$, it follows that:

$$\begin{aligned}MSE_\theta(a^T.\widehat{\beta_\lambda}) &= E_\theta((a^T.\widehat{\beta_\lambda}-a^T.\beta)^2) \\ &\leq E_\theta((a^T.\widehat{\beta_\lambda}-a^T.\beta)^2)+2.\lambda.E_\theta((a^T.\widehat{\beta_\lambda}-a^T.\beta).(X^T.X)^{-1}.(a^T.\widehat{\beta_\lambda})) \\ &+ \lambda^2.E_\theta((X^T.X)^{-1}.(a^T.\widehat{\beta_\lambda})) \\ &= E_\theta((a^T.\widehat{\beta_\lambda}-a^T.\beta)^2)+2.E_\theta((a^T.\widehat{\beta_\lambda}-a^T.\beta).(\lambda.(X^T.X)^{-1}.a^T.\widehat{\beta_\lambda})) \\ &+ E_\theta(\lambda^2.((X^T.X)^{-1}.a^T.\widehat{\beta_\lambda})^2) \\ &= E_\theta(((a^T.\widehat{\beta_\lambda}-a^T.\beta)+(\lambda.(X^T.X)^{-1}.a^T.\widehat{\beta_\lambda}))^2) \\ &= E_\theta((a^T.(I+\lambda.(X^T.X)^{-1}).(\widehat{\beta_\lambda})-a^T.\beta)^2) \\ &= MSE_\theta(a^T.(I+\lambda.(X^T.X)^{-1}).(\widehat{\beta_\lambda})) \\ &= MSE_\theta(a^T.\widehat{\beta}).\end{aligned}$$

□

Problem 8.11. Suppose that $Y_i = x_i^T\beta + \epsilon_i (i=1,\cdots,n)$ where ϵ_i's are i.i.d. with mean 0 and finite variance. Consider the F statistic (call it F_n) for testing $H_0 : \beta_{r+1} = \cdots = \beta_p = 0$ where $\beta = (\beta_0,\cdots,\beta_p)^T$.
(a) Under H_0 and assuming the conditions of Theorem 8.5. on the x_i's, show that

$$(p-r).F_n \to_d \chi^2(p-r).$$

(b) If H_0 is not true, what happens to $(p-r).F_n$ as $n\to\infty$. ?

Solution. As $\frac{RSS}{\sigma^2} \sim \chi^2(n-p-1) =^d \sum_{i=1}^{n-p-1} X_i^* :\ X_i^* \sim^{i.i.d.} \chi^2(1)$, and $E(X_i^*)=1\ (1\leq i\leq n-p-1)$, an application of Theorem 3.6, implies that $(\frac{RSS}{\sigma^2})/(n-(p+1)) \to_p 1$. Then, an application of Theorem 3.2 with $g(x)=\frac{1}{x}$ yields

$$1/[(\frac{RSS}{\sigma^2})/(n-(p+1))] \to_p 1. \quad (*)$$

(a) Referring to Page 409:

$$(p-r).F_n = \frac{\frac{RSS_r-RSS}{\sigma^2}}{\frac{RSS}{(n-p-1)\sigma^2}} : \quad \frac{RSS_r-RSS}{\sigma^2} \sim \chi^2(p-r). \quad (**)$$

Now, by $(*)$ and $(**)$ and an application of Theorem 3.3.(b) yields:

$$(p-r).F_n = \frac{RSS_r - RSS}{\sigma^2} * \frac{1}{\frac{RSS}{(n-p-1)\sigma^2}} \to_d \chi^2(p-r).$$

(b) Referring to Pages 413-414:

$$(p-r).F_n = \frac{\frac{RSS_r - RSS}{\sigma^2}}{\frac{RSS}{(n-p-1)\sigma^2}} : \quad \frac{RSS_r - RSS}{\sigma^2} \sim \chi^2(p-r;\theta^2), \quad \theta^2 = \frac{\|X.\beta\|^2 - \|H_r.X.\beta\|^2}{\sigma^2}. \quad (***)$$

Accordingly, by $(*)$ and $(***)$ and another application of Theorem 3.3.(b) it follows that:

$$(p-r).F_n = \frac{RSS_r - RSS}{\sigma^2} * \frac{1}{\frac{RSS}{(n-p-1)\sigma^2}} \to_d \chi^2(p-r;\theta^2).$$

□

Problem 8.13. Consider the linear regression model $Y_i = x_i^T\beta + \epsilon_i$ $(i = 1, \cdots, n)$ where $\epsilon_1, \cdots, \epsilon_n$ are i.i.d. Exponential random variables with unknown parameter λ.
(a) Show that the density function of Y_i's is $f_i(y) = \lambda.\exp[-\lambda.(y - x_i^T\beta)]$ for $y \geq x_i^T\beta$.
(b) Show that the MLE of β for this model maximizes the function $g(u) = \sum_{i=1}^n x_i^T u$ subject to the constraints $Y_i \geq x_i^T u$ for $i = 1, \cdots, n$.
(c) Suppose that $Y_i = \beta.x_i + \epsilon_i (i = 1, \cdots, n)$ where $\epsilon_1, \cdots, \epsilon_n$ i.i.d. Exponential random variables with parameter λ and $x_i > 0$ for all i. If $\widehat{\beta_n}$ is the MLE of β, show that $\widehat{\beta_n} - \beta$ has an exponential distribution with parameter $\lambda.\sum_{i=1}^n x_i$.

Solution. (a) Let $y = h(\epsilon) = x^T.\beta + \epsilon$ with $f_\epsilon(y) = \lambda.e^{-\lambda.\epsilon}$ $\epsilon > 0$. By Theorem 2.3, for $\epsilon = h^{-1}(y) = y - x^T.\beta$ and $|J_{h^{-1}}(y)| = 1$ we have:

$$f_Y(y) = f_X(h^{-1}(y)).|J_{h^{-1}}(y)| = \lambda.e^{-\lambda.(y - x^T.\beta)} : \quad y - x^T.\beta \geq 0.$$

(b) As:

$$l(\beta;\mathbf{y}) = \sum_{i=1}^n \log(f_Y(y_i;\beta)) = \sum_{i=1}^n (\log(\lambda) - \lambda.(y_i - x_i^T.\beta)) = (n.\log(\lambda) - \lambda.\sum_{i=1}^n y_i) + \lambda.(\sum_{i=1}^n x_i^T.\beta) : \quad \lambda > 0,$$

it follows that:

$$arg(\max(l(\beta;\mathbf{y}))) = arg(\max(\sum_{i=1}^n x_i^T.\beta)) : \quad y_i - x_i^T.\beta \geq 0 \ \ (1 \leq i \leq n).$$

(c) By $y_i - \beta.x_i \geq 0$ $(1 \leq i \leq n)$ we have, $\frac{y_i}{x_i} \geq \beta$ $(1 \leq i \leq n)$, and by Part (b):

$$\widehat{\beta_n} = arg(\max(l(\beta;\mathbf{y}))) = arg(\max(\sum_{i=1}^n x_i^T.\beta)) = \min_{1\leq i\leq n} \frac{y_i}{x_i}.$$

Accordingly:

$$\begin{aligned} F_{\widehat{\beta_n}-\beta}(x) &= P(\widehat{\beta_n} - \beta \leq x) = P(\min_{1\leq i\leq n} \frac{Y_i}{x_i} \leq \beta + x) = 1 - P(\min_{1\leq i\leq n} \frac{Y_i}{x_i} \geq \beta + x) \\ &= 1 - \prod_{i=1}^n P(\frac{Y_i}{x_i} > \beta + x) = 1 - \prod_{i=1}^n P(\epsilon_i > x.x_i) = 1 - e^{-\lambda.\sum_{i=1}^n x_i.x}, \end{aligned}$$

and hence, $\widehat{\beta_n} - \beta \sim \exp(\lambda. \sum_{i=1}^{n} x_i)$.
□

Problem 8.15. Suppose that Y has a density or frequency function of the form

$$f(y;\theta,\phi) = \exp[\theta.y - \frac{b(\theta)}{\phi} + c(y,\phi)]$$

for $y \in A$, which is independent of the parameters θ and ϕ. This is an alternative to general family of distributions considered in Problem 8.14. and is particularly appropriate for discrete distributions.
(a) Show that the Negative Binomial distribution of Example 8.12 has this form.
(b) Show that $E_\theta(Y) = \phi^{-1}b'(\theta)$ and $Var_\theta(Y) = \phi^{-1}b''(\theta)$.

Solution. For $y = 0, 1, \cdots$, we can write:

$$\begin{aligned} f(y;\mu) &= \frac{\Gamma(y+\frac{1}{\alpha})}{y!.\Gamma(\frac{1}{\alpha})} * \frac{(\alpha.\mu)^y}{(1+\alpha.\mu)^{y+\frac{1}{\alpha}}} \\ &= C(y+\frac{1}{\alpha}-1,y) * (\frac{1}{1+\alpha.\mu})^{\frac{1}{\alpha}} * (\frac{\alpha.\mu}{1+\alpha.\mu})^y = C(y+\frac{1}{\alpha}-1,y) * (1-\frac{\alpha.\mu}{1+\alpha.\mu})^{\frac{1}{\alpha}} * (\frac{\alpha.\mu}{1+\alpha.\mu})^y \\ &= \exp[\log(C(y+\frac{1}{\alpha}-1,y)) + \frac{1}{\alpha}.\log(1-\frac{\alpha.\mu}{1+\alpha.\mu}) + y.\log(\frac{\alpha.\mu}{1+\alpha.\mu})], \end{aligned}$$

and by taking

$$\theta = \log(\frac{\alpha.\mu}{1+\alpha.\mu}), \quad \phi = \alpha, \quad c(y,\alpha) = \log(C(y+\frac{1}{\alpha}-1,y)), \quad b(\theta) = -\log(1-\frac{\alpha.\mu}{1+\alpha.\mu}) = -\log(1-e^\theta).$$

the assertion follows.

(b) First:

$$\begin{aligned} 0 &= \frac{d}{d\theta}(1) = \frac{d}{d\theta}\int_A f(y;\theta,\phi)dy = \frac{d}{d\theta}\int_A [\exp(\theta.y - \frac{b(\theta)}{\phi} + c(y,\phi))]dy \\ &= \int_A \frac{d}{d\theta}[\exp(\theta.y - \frac{b(\theta)}{\phi} + c(y,\phi))]dy = \int_A (y - \frac{b'(\theta)}{\phi}).f(y;\theta,\phi)dy \\ &= E_\theta(Y) - \frac{b'(\theta)}{\phi}, \end{aligned}$$

so, $E_\theta(Y) = \frac{b'(\theta)}{\phi}$.

Second, using first part it follows that:

$$\begin{aligned} \frac{b''(\theta)}{\phi} &= \frac{d}{d\theta}(\frac{b'(\theta)}{\phi}) = \frac{d}{d\theta}(E_\theta(Y)) = \frac{d}{d\theta}\int_A y.[\exp(\theta.y - \frac{b(\theta)}{\phi} + c(y,\phi))]dy \\ &= \int_A y.\frac{d}{d\theta}[\exp(\theta.y - \frac{b(\theta)}{\phi} + c(y,\phi))]dy = \int_A y.(y - \frac{b'(\theta)}{\phi}).f(y;\theta,\phi)dy \\ &= E_\theta(Y^2) - \frac{b'(\theta)}{\phi}.E_\theta(Y) = E_\theta(Y^2) - (E_\theta(Y))^2 = Var_\theta(Y). \end{aligned}$$

□

Problem 8.17. Lambert (1992) describes an approach to regression modelling of count data using a zero-inflated Poisson distribution. that is, the response variables $\{Y_i\}$ are nonnegative integer valued random variables with the frequency function of Y_i given by

$$P(Y_i = y) = \theta_i + (1-\theta_i)\exp(-\lambda_i) \text{ for } y = 0, \quad (1-\theta_i)\exp(-\lambda_i)\lambda_i^y/y! \text{ for } y = 1, 2, \cdots$$

where θ_i and λ_i depend on some covariates; in particular, it is assumed that

$$\ln(\frac{\theta_i}{1-\theta_i}) = x_i^T.\beta, \qquad \ln(\lambda_i) = x_i^T.\phi,$$

where $x_i (i = 1, \cdots, n)$ are covariates and β, ϕ are vectors of unknown parameters.
(a) the zero-inflated Poisson model can viewed as a mixture of a Poisson distribution and a distribution concentrated at 0. That is, let Z_i be a Bernoulli random variable with $P(Z_i = 0) = \theta_i$ such that $P(Y_i = 0|Z_i = 0) = 1$ and given $Z_i = 1, Y_i$ is Poisson distributed with mean λ_i. Show that

$$P(Z_i = 0|Y_i = y) = \theta_i/[\theta_i + (1-\theta_i).\exp(-\lambda_i)] \text{ for } y = 0, \quad 0 \text{ for } y \geq 1.$$

(b) Suppose that we could observe $(Y_1, Z_1), \cdots, (Y_n, Z_n)$ where the Z_i's are defined in part (a). Show that the MLE of β depends only on the Z_i's.
(c) Use the "Complete data" likelihood in part (b) to describe an EM algorithm for computing maximum likelihood estimates of β and ϕ.
(d) In the spirit of the zero-inflated Poisson model, consider the following simple zero-inflated Binomial model: for $i = 1, \cdots, n$, $Y_1, \cdots, Y_n$ are independent random variables with

$$P(Y_i = 0) = \lambda_i + (1-\lambda_i).(1-\theta_i)^m, \qquad P(Y_i = y) = (1-\lambda_i).C(m, y)\theta_i^y(1-\theta_i)^{m-y} \quad 1 \leq y \leq m$$

where $0 < \phi < 1$ and $\ln(\frac{\theta_i}{1-\theta_i}) = \beta_0 + \beta_1.x_i$, $\ln(\frac{\lambda_i}{1-\lambda_i}) = \phi_0 + \phi_1.x_i$ for some covariates $x_1, \cdots, x_n$. Derive an EM algorithm for estimating ϕ and β and use it to estimate the parameters for the data in Table 8.1.; for each observation, $m = 6$. with $m = 6$:

Table 8.1. Data for Problem 8.17; for each observation m=6.

x	y	x	y	x	y	x	y
0.3	0	0.6	0	1.0	0	1.1	0
2.2	1	2.2	0	2.4	0	2.5	0
3.0	4	3.2	0	3.4	4	5.8	5
6.2	0	6.5	5	7.1	4	7.6	6
7.7	4	8.2	4	8.6	4	9.8	0

.

(e) Carry out a likelihood ratio test for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$. (Assume that the standard χ^2 approximation can be applied.)

Solution. (a)

$$
\begin{aligned}
P(Z_i = 0|Y_i = 0) &= \frac{P(Y_i = y|Z_i = 0).P(Z_i = 0)}{P(Y_i = y)} \\
&= \frac{P(Y_i = y|Z_i = 0).P(Z_i = 0)}{[P(Y_i = y|Z_i = 0).P(Z_i = 0) + P(Y_i = y|Z_i = 1).P(Z_i = 1)]} \\
&= \frac{\theta}{\theta + e^{-\lambda}.(1-\theta)} \quad \text{if} \quad y = 0, \\
&= \frac{P(Y_i = y|Z_i = 0).P(Z_i = 1)}{P(Y_i = y)} \\
&\leq \frac{(1 - P(Y_i = 0|Z_i = 0)).P(Z_i = 1)}{P(Y_i = y)} \leq \frac{0.P(Z_i = 1)}{P(Y_i = y)} = 0 \quad \text{if} \quad y > 0.
\end{aligned}
$$

(b) As $f(y, z; \beta, \phi) = f(y; z, \beta, \phi) * f(z; \beta, \phi) = f(y; z, \phi) * f(z; \beta)$, the log-likelihood of the joint distribution can be written as:

$$
\begin{aligned}
l(\beta, \phi; y, z) &= \sum_{i=1}^{n} \log(f(z_i; \beta)) + \sum_{i=1}^{n} \log(f(y_i; z_i, \phi)) \\
&= [\sum_{i=1}^{n} (z_i.x_i^T.\beta - \log(1 + \exp(x_i^T.\beta)))] \\
&+ [\sum_{i=1}^{n} (1 - z_i).(y_i.x_i^T.\phi - \exp(x_i^T.\phi)) - \sum_{i=1}^{n} (1 - z_i).\log(y_i!)] \\
&= L_c(\beta; Y, Z) + L_c(\phi; Y, Z),
\end{aligned}
$$

in which the first term $L_c(\beta; Y, Z)$ is only dependent to z and so is $MLE(\beta)$.

(c) The $(k + 1)$th iteration of the EM algorithm requires three steps:

(i) E-Step: Estimate z_i via:

$$
\begin{aligned}
Z_i^{(k)} &= P(z_i = 0|y_i, \beta^{(k)}, \phi^{(k)}) \\
&= \frac{P(y_i|z_i = 0).P(z_i = 0)}{[P(y_i|z_i = 0).P(z_i = 0) + P(y_i|z_i = 1).P(z_i = 1)]} \\
&= \frac{1}{1 + \exp(-x_i^T.\beta^{(k)} - \exp(x_i^T.\phi^{(k)}))}.1_{y_i=0}.
\end{aligned}
$$

(ii) M-Step for ϕ: We find $\phi^{(k+1)}$ by maximizing $L_c(\phi; Y, Z^{(k)})$.

(iii) M-Step for β: We find $\beta^{(k+1)}$ by maximizing $L_c(\beta; Y, Z^{(k)})$ as a function of β given below:

$$
L_c(\beta; Y, Z^{(k)}) = \sum_{y_i=0} z_i^{(k)}.x_i^T.\beta - \sum_{y_i=0} z_i^{(k)} \log(1 + \exp(x_i^T.\beta)) - \sum_{i=1}^{n} (1 - z_i^{(k)}).\log(1 + \exp(x_i^T.\beta)).
$$

(d) First, the log-likelihood is given by (Hall, 2000):

$$
\begin{aligned}
l(\phi, \beta; \mathbf{y}) &= \sum_{i=1}^{n} [1_{y_i=0} * \log(e^{x_i^T.\phi} + (1 + e^{x_i^T.\beta})^{-m_i}) - \log(1 + e^{x_i^T.\phi}) \\
&+ 1_{y_i>0} * (y_i.x_i^T.\beta - m_i.\log(1 + e^{x_i^T.\beta}) + \log(C(m_i, y_i)))].
\end{aligned}
$$

Second, define

$$Z_i = 1 \text{ if } Y_i \text{ is generated from zero state}, \quad 0, \quad \text{if } Y_i \text{ is generated from binomial state.}$$

Then, given $Z = (z_1, \cdots, z_n)$ the complete data $\{(y_i, z_i)\}_{i=1}^n$ log-likelihood is of the form:

$$\begin{aligned} l_c(\phi, \beta; y, z) &= \log(\prod_{i=1}^n P(Y_i = y_i, Z_i = z_i)) \\ &= \sum_{i=1}^n [z_i.x_i^T.\phi - \log(1 + e^{x_i^T.\phi})] \\ &+ \sum_{i=1}^n [(1 - z_i) * (y_i.x_i^T.\beta - m_i.\log(1 + e^{x_i^T.\beta}) + \log(C(m_i, y_i)))] \\ &= l_c(\phi; y, z) + l_c(\beta; y, z). \end{aligned}$$

Finally, the EM algorithm by starting values $(\phi^{(0)}, \beta^{(0)})$ for the iteration $(k+1)$ has the following steps (Hall, 2000):

(i) E-Step: Estimate Z_i via:

$$Z_i^{(k)} = E(Z_i | y_i, \phi^{(k)}, \beta^{(k)}) = \frac{P(Z_i = 1 | y_i, \phi^{(k)}, \beta^{(k)}).P(Z_i = 1)}{\sum_{t=0}^1 P(Z_i = t | y_i, \phi^{(k)}, \beta^{(k)}).P(Z_i = t)} = \frac{1_{y_i=0}}{1 + \exp(-x_i^T.\phi^{(k)})(1 + e^{x_i^T.\beta^{(k)}})^{-m_i}}.$$

(ii) M-Step for ϕ : We find $\phi^{(k+1)}$ by maximizing $l_c(\phi; y, z^{(k)})$.

(iii) M-Step for β : We find $\beta^{(k+1)}$ by maximizing $l_c(\beta; y, z^{(k)})$.

(e) With the assumption $\phi_1 = 0$, the following SAS 9.4. Proc FMM output shows that with $p-value(\beta_1) = 0.0803 > 0.0500$, we cannot reject the null hypothesis $H_0 : \beta_1 = 0$ at 5% level.

Zero-Inflated Binomial Model

The FMM Procedure

Parameter Estimates for Binomial Model					
Component	**Effect**	**Estimate**	**Standard Error**	**z Value**	**Pr > |z|**
1	Intercept	-1.3727	1.3014	-1.05	0.2915
1	x	0.3481	0.1990	1.75	0.0803

Parameter Estimates for Mixing Probabilities					
Component		**Linked Scale**			
	Mixing Probability	**Logit(Prob)**	**Standard Error**	**z Value**	**Pr > |z|**
1	0.5492	0.1975	0.6067	0.33	0.7447
2	0.4508	-0.1975			

Figure 8.1. SAS 9.4. output for Table 8.1. data

□

Problem 8.19. Consider finding a quasi-likelihood function based on the variance function $V(\mu) = \mu^r$ for some specified $r > 0$.
(a) Find the function $\psi(\mu; y)$ for $V(\mu)$.
(b) Show that

$$\frac{d}{d\mu}\psi(\mu; y) = \frac{d}{d\mu}\ln f(y; \mu)$$

for some density or frequency function $f(y; \mu)$ when $r = 1, 2, 3$.

Solution. (a) Using $\frac{d}{d\mu}\psi(\mu, y) = \frac{y-\mu}{V(\mu)}$, it follows that:

$$\begin{aligned}\psi(\mu, y) &= \int (\frac{y-\mu}{V(\mu)})d\mu = \int (\frac{y-\mu}{\mu^r})d\mu = \\ &= [\frac{y}{1-r}.\mu^{1-r}.1_{r\neq 1} + y.\ln(\mu).1_{r=1}] - [\frac{y}{2-r}.\mu^{2-r}.1_{r\neq 2} + \ln(\mu).1_{r=2}] + c(y).\end{aligned}$$

In particular:

$$\begin{aligned} r=1 &: \psi(\mu, y) = y.\log(\mu) - \mu \quad \textit{Poisson} \\ r=2 &: \psi(\mu, y) = \frac{-y}{\mu} - \log(\mu) \quad \textit{Gamma} \\ r=3 &: \psi(\mu, y) = \frac{-y}{2.\mu^2} + \frac{1}{\mu} \quad \textit{InverseGaussian} \\ r=k &: \psi(\mu, y) = \mu^{-k} * (\frac{\mu.y}{1-k} - \frac{\mu^2}{2-k}) \quad k \neq 0, 1, 2. \end{aligned}$$

(b) Suppose that for some measure P on $\mathbb{R}$ to have:

$$dP_Y(y) = \exp(y.\theta - g(\theta)).dP(y) : \qquad \theta = \int \frac{d\mu}{V(\mu)}.$$

Then, $1 = \int dP_Y = \int \exp(y.\theta - g(\theta))dP(y) = e^{-g(\theta)}.\int e^{y\theta}dP(y)$ or $\int e^{y\theta}dP(y) = e^{g(\theta)}$. Consequently:

$$\begin{aligned} m_Y(t) &= E(e^{tY}) = \int e^{ty}.e^{y\theta - g(\theta)}dP(y) = \int e^{y(t+\theta)}dP(y).e^{-g(\theta)} \\ &= e^{g(t+\theta)}.e^{-g(\theta)} = e^{g(t+\theta)-g(\theta)}. \end{aligned}$$

Hence, $m_Y^{'}(0) = g^{'}(\theta) = \mu, \quad g^{''}(\theta) = V(\mu), \quad \frac{d\mu}{d\theta} = g^{''}(\theta) = V(\mu)$, implying:

$$\frac{d\ln(f(y;\mu))}{d\mu} = \frac{d\ln(f(y;\mu))}{d\theta}\frac{d\theta}{d\mu} = (y - g^{'}(\theta)).\frac{1}{V(\mu)} = \frac{y-\mu}{V(\mu)} = \frac{d\psi(\mu, y)}{d\mu}.$$

Finally, checking $\psi(\mu, y)$ for $V(\mu) = \mu^r \;\; (r = 1, 2, 3)$ in Part (a) we observe that:

$$\frac{d\ln(f(y;\mu))}{d\mu} = \frac{d\psi(\mu, y)}{d\mu} \quad r = 1, 2, 3.$$

□

Chapter 9

Goodness-of-Fit

Problem 9.1. The distribution of personal incomes is sometimes modelled by a distribution whose density function is

$$f(x;\alpha,\theta)=\frac{\alpha}{\theta}(1+\frac{x}{\theta})^{-(\alpha+1)}\quad\text{for } x\geq 0$$

for some unknown parameters $\alpha>0$ and $\theta>0$. The data given in Table 9.2 are a random sample of incomes (in 1000s of dollars) as declared on income tax forms. Thinking of these data as outcomes of i.i.d. random variables $X_1,\cdots,X_{40}$, define

$$Y_1=\sum_{i=1}^{40}I_{(X_i\leq 25)},\quad Y_2=\sum_{i=1}^{40}I_{(25<X_i\leq 40)},\quad Y_3=\sum_{i=1}^{40}I_{(40<X_i\leq 90)},\quad Y_4=\sum_{i=1}^{40}I_{(X_i>90)}.$$

Table 9.2. Data for Problem 9.1.

3.5	7.9	8.5	9.2	11.4	17.4	20.8	21.2
21.4	22.5	25.3	25.7	25.9	26.2	26.6	27.8
28.7	30.1	30.2	30.9	35.0	36.0	39.0	39.0
39.6	43.2	44.8	47.7	57.5	62.5	72.8	83.1
96.6	106.6	115.3	118.1	152.5	169.2	202.2	831.0

(a) What is the likelihood function for the parameters α and θ based on $(Y_1,\cdots,Y_4)$?
(b) Find the maximum likelihood estimates of α and θ based on the observed values of $(Y_1,\cdots,Y_4)$ in the sample.
(c) Test the null hypothesis that the density of the data is $f(x;\alpha,\theta)$ for some α and θ using both the LR statistics and Pearson χ^2 statistics. Compute approximate p-values for both test statistics.

Solution. (a)Take $S=[0,\infty)$ and define $A_1=[0,25], A_2=(25,40], A_3=(40,90], A_4=(90,\infty)$. Then, $S=\cup_{j=1}^4 A_j$, and furthermore:

$$\begin{aligned} p(a,b;\alpha,\theta) &= \int_a^b f(x;\alpha,\theta)dx=\int_a^b \frac{\alpha}{\theta}(1+\frac{x}{\theta})^{-(\alpha+1)}dx \\ &= \int_{a/\theta}^{b/\theta}\alpha.(1+x)^{-(\alpha+1)}dx=(1+\frac{a}{\theta})^{-\alpha}-(1+\frac{b}{\theta})^{-\alpha},\quad (a<b). \end{aligned}$$

So:

$$\begin{aligned} L(\alpha,\theta;y_1,y_2,y_3,y_4) &= \frac{40!}{y_1!y_2!y_3!y_4!} \\ &* [1-(1+\frac{25}{\theta})^{-\alpha}]^{y_1} * [(1+\frac{25}{\theta})^{-\alpha} - (1+\frac{40}{\theta})^{-\alpha}]^{y_2} \\ &* [(1+\frac{40}{\theta})^{-\alpha} - (1+\frac{90}{\theta})^{-\alpha}]^{y_3} * [(1+\frac{90}{\theta})^{-\alpha}]^{y_4}. \end{aligned}$$

(b) By data given in Table 9.2 $y_1 = 10, y_2 = 15, y_3 = 7, y_4 = 8$ and Part (a) the log-likelihood function is given by:

$$\begin{aligned} l(\alpha,\theta) &= \log(L(\alpha,\theta;10,15,7,8)) \\ &= \log(\frac{40!}{10!15!7!8!}) + 10*\log(1-(1+\frac{25}{\theta})^{-\alpha}) + 15*\log((1+\frac{25}{\theta})^{-\alpha} - (1+\frac{40}{\theta})^{-\alpha}) \\ &+ 7*\log((1+\frac{40}{\theta})^{-\alpha} - (1+\frac{90}{\theta})^{-\alpha}) + 8*\log((1+\frac{90}{\theta})^{-\alpha}). \end{aligned}$$

Next, define $g(\alpha,\theta) = -l(\alpha,\theta)$. Then, by Powell's Method for finding minimum values of g (Powell, 1964), it follows that: $\widehat{\alpha} = 0.0527$ and $\widehat{\theta} = 0.0917$.

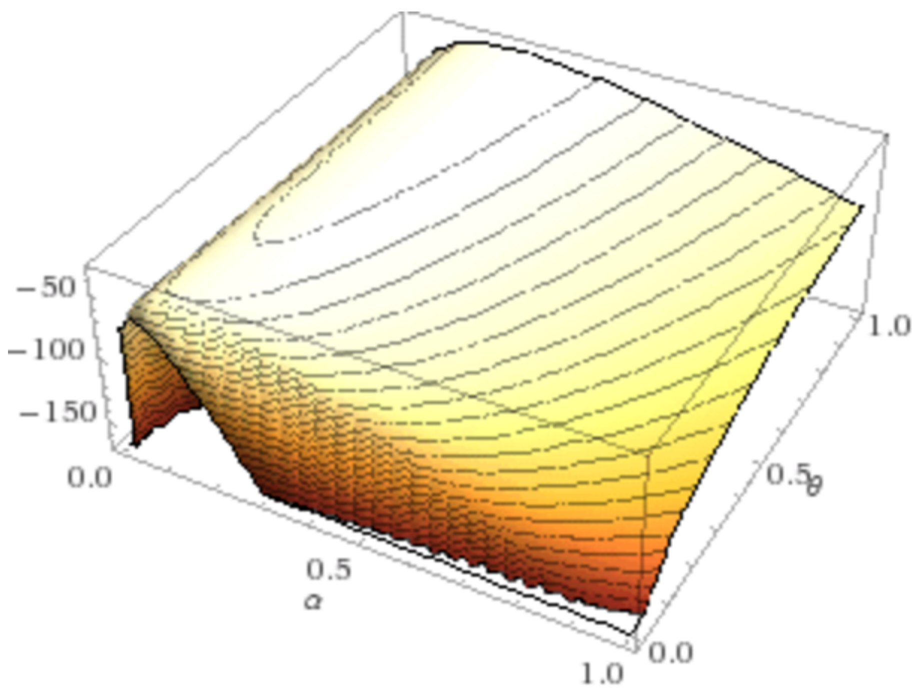

Figure 9.1 Plot of function $l(\alpha,\theta)$

(c) First, define:

$$p_1(\alpha,\theta) = 1-(1+\frac{25}{\theta})^{-\alpha}, \quad p_2(\alpha,\theta) = (1+\frac{25}{\theta})^{-\alpha} - (1+\frac{40}{\theta})^{-\alpha},$$

$$p_3(\alpha,\theta) = (1+\frac{40}{\theta})^{-\alpha} - (1+\frac{90}{\theta})^{-\alpha}, \quad p_4(\alpha,\theta) = (1+\frac{90}{\theta})^{-\alpha},$$

and evaluating at MLE values in Part (b) we get:

$$p_1(\widehat{\alpha},\widehat{\theta}) = 0.0256, \quad p_2(\widehat{\alpha},\widehat{\theta}) = 0.0181, \quad p_3(\widehat{\alpha},\widehat{\theta}) = 0.0303, \quad p_4(\widehat{\alpha},\widehat{\theta}) = 0.6955.$$

Second, by Theorem 9.1. and Theorem 9.2. for $k = 4$, $n = 40$ and $p = 2$ we have:

$$2.\ln(\Lambda_{40}) \sim \chi^2(1), \quad K^2_{40} \sim \chi^2(1).$$

To test the null hypothesis $H_0 : \phi_j = p_j(\alpha, \theta) \ \ (j = 1, 2, 3, 4)$ we have:

$$\begin{aligned}
2.\ln(\Lambda_{40}) &= 2.\sum_{j=1}^{4} y_j.\ln\Big(\frac{y_j}{40.p_j(\widehat{\alpha}, \widehat{\theta})}\Big) = 141.118 >> 3.841 \Rightarrow \text{Reject the null hypothesis at 5\% level} \\
K^2_{40} &= \sum_{j=1}^{4} \frac{(y_j - 40.p_j(\widehat{\alpha}, \widehat{\theta}))^2}{40.p_j(\widehat{\alpha}, \widehat{\theta})} = 401.939 >> 3.841 \Rightarrow \text{Reject the null hypothesis at 5\% level.}
\end{aligned}$$

The corresponding p-values for the above LR test statistics and Pearson χ^2 test statistics are both smaller than 0.00001.

□

Problem 9.3. Consider theorem 9.2. where now we assume that $\tilde{\theta_n}$ is some estimator (not necessarily the MLE from the Multinomial model) with

$$\sqrt{n}(\tilde{\theta_n} - \theta) \to_d N_p(0, C(\theta).)$$

(a) Show that $K^{*2}_n - 2\ln(\Lambda^*_n) \to_p 0$ (under the null hypothesis).
(b) What can be said about the limiting distribution of $2\ln(\Lambda^*_n)$ under this more general assumption on $\tilde{\theta_n}$?

Solution. (a) First, by null hypothesis $H_0 : p_j(\theta) = \phi_j \ (1 \le j \le k)$, $\widehat{\phi_j} = \frac{Y_{nj}}{n} \ \ (1 \le j \le k)$, and Example 3.12 for $X^*_i = 1_{X_i \in A_j} \sim Bernoulli(p_j(\theta)) \ \ (1 \le i \le n)$ and $\overline{X^*_n} = \frac{\sum_{i=1}^n X^*_i}{n} = \frac{Y_{nj}}{n}$ it follows that:

$$\sqrt{n}\Big(\frac{Y_{nj}}{n} - p_j(\theta)\Big) \to_d N(0, p_j(\theta) * (1 - p_j(\theta))) \ \ (1 \le j \le k). \ \ (*)$$

Second, it follows from assumption that:

$$\sqrt{n}.(p_j(\tilde{\theta_n}) - p_j(\theta)) \to_d N(0, p_j(\theta)^T.C(\theta).p_j(\theta)) \ \ (1 \le j \le k). \ \ (**)$$

Third, it follows from $(*)$ and $(**)$ that:

$$\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}) = \Big(\frac{Y_{nj}}{n} - p_j(\theta)\Big) + (p_j(\theta) - p_j(\tilde{\theta})) \simeq^d N\Big(0, \frac{p_j(\theta).(1 - p_j(\theta))}{n}\Big) + N\Big(0, \frac{p_j(\theta)^T.C(\theta).p_j(\theta)}{n}\Big),$$

and , consequently, for $|r_n| \le 1$:

$$\begin{aligned}
E\Big(\Big(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})\Big)^2\Big) &= Var\Big(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})\Big) \\
&\simeq \frac{p_j(\theta).(1 - p_j(\theta))}{n} + \frac{p_j(\theta)^T.C(\theta).p_j(\theta)}{n} \\
&+ 2.r_n.\frac{\sqrt{p_j(\theta).(1 - p_j(\theta))}.\sqrt{p_j(\theta)^T.C(\theta).p_j(\theta)}}{n} \\
&=^{\text{define}} \frac{k_j(\theta, r_n)}{n}, \quad (1 \le j \le k), \quad (n \ge 1). \quad (***)
\end{aligned}$$

Now, for $r_1, r_2, \epsilon > 0$ by $(***)$ and an application of Theorem 3.7 it follows that:

$$\begin{aligned} P(n^{r_1}.|\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})|^{r_2} > \epsilon) &= P(|\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})| > (\frac{\epsilon}{n^{r_1}})^{\frac{1}{r_2}}) \\ &\leq \frac{E((\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^2)}{(\frac{\epsilon}{n^{r_1}})^{\frac{1}{r_2}}} \\ &= \frac{k_j(\theta, r_n)}{\epsilon^{\frac{1}{r_2}}} * \frac{1}{n^{1-\frac{r_1}{r_2}}}, \quad (1 \leq j \leq k), \ (n \geq 1). \quad (\dagger) \end{aligned}$$

Accordingly, it follows from $(\dagger)$ that:

$$n^{r_1}.|\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})|^{r_2} \rightarrow_p 0, \quad (r_2 > r_1 > 0). \quad (\dagger\dagger)$$

Fourth, using $\sum_{j=1}^{k} \frac{Y_{nj}}{n} = 1 = \sum_{j=1}^{k} p_j(\tilde{\theta_n})$ it follows that:

$$\sum_{j=1}^{k} [\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})] = 0. \quad (\dagger\dagger\dagger)$$

Fifth, given $(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})) \rightarrow_p 0$, $(\dagger\dagger\dagger)$ and using Taylor expansion of $f(x) = \ln(x)$ around $a = p_j(\tilde{\theta_n})$ we have:

$$\begin{aligned} 2.\ln(\Lambda_n^*) &= 2n.\sum_{j=1}^{n} \frac{Y_{nj}}{n} * \ln(\frac{Y_{nj}}{n.p_j(\tilde{\theta_n})}) \\ &= 2n.\sum_{j=1}^{k} [((\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})) + p_j(\tilde{\theta_n})) * (\ln(\frac{Y_{nj}}{n}) - \ln(p_j(\tilde{\theta_n})))] \\ &= 2n.\sum_{j=1}^{k} [(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})) * (\ln(\frac{Y_{nj}}{n}) - \ln(p_j(\tilde{\theta_n})))] \\ &+ 2n.\sum_{j=1}^{k} [(p_j(\tilde{\theta_n}) * (\ln(\frac{Y_{nj}}{n}) - \ln(p_j(\tilde{\theta_n})))] \\ &= 2n.\sum_{j=1}^{k} [(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})) * (\frac{1}{p_j(\tilde{\theta_n})}.(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})) + O_p(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^2)] \\ &+ 2n.\sum_{j=1}^{k} [(p_j(\tilde{\theta_n})) * (\frac{1}{p_j(\tilde{\theta_n})}.(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})) - \frac{1}{2.(p_j(\tilde{\theta_n}))^2}.(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^2 + O_p(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^3] \\ &= 2n.\sum_{j=1}^{k} [(\frac{1}{p_j(\tilde{\theta_n})}.(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^2] + 2n.\sum_{j=1}^{k} O_p(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^3] \\ &+ 2n.\sum_{j=1}^{k} [(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n})] - n.\sum_{j=1}^{k} [(\frac{1}{p_j(\tilde{\theta_n})}.(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^2] + 2n.\sum_{j=1}^{k} O_p(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^3] \\ &= (K_n^*)^2 + 4n.\sum_{j=1}^{k} O_p(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^3. \quad (\ddagger) \end{aligned}$$

Now, by (†) for $r_1 = 1, r_2 \geq 3$ and (‡) it follows that:

$$2.\ln(\Lambda_n^*) - (K_n^*)^2 = 4n.\sum_{j=1}^{k} O_p(\frac{Y_{nj}}{n} - p_j(\tilde{\theta_n}))^3 \to_p 0.$$

(b) Let $(K_n)^2, 2.\ln(\Lambda_n)$ and $(K_n^*)^2, 2.\ln(\Lambda_n^*)$ be corresponding statistics to $p_j(\widehat{\theta_n})$ and $p_j(\tilde{\theta_n})$, respectively. Then, if $(K_n^*)^2 - (K_n)^2 \to_p 0$, then by Part (a) and using

$$2.\ln(\Lambda_n^*) = (2.\ln(\Lambda_n^*) - (K_n^*)^2) + ((K_n^*)^2 - (K_n)^2) + (K_n)^2$$

it follows that $2.\ln(\Lambda_n^*)$ has an asymptotic χ^2 distribution.
□

Problem 9.5. Suppose that $X_1, \cdots, X_n$ are i.i.d. continuous random variables whose range is the interval $(0,1)$. To test the null hypothesis that the X_i's are uniformly distributed, we can use the statistics:

$$V_n = (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sin(2\pi.X_i))^2 + (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\cos(2\pi.X_i))^2.$$

(a) Suppose that $X_i's$ are Uniform random variables on $[0,1]$. Show that as $n \to \infty$,

$$(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sin(2\pi.X_i), \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\cos(2\pi.X_i)) \to_d (Z_1, Z_2)$$

where Z_1 and Z_2 are independent $N(0,\sigma^2)$ random variables. Find the values of σ^2.
(b) Find the asymptotic distribution of V_n when the X_i's are uniformly distributed.
(c) Suppose that either $E[\sin(2\pi.X_i)]$ or $E[\cos(2\pi.X_i)]$ (or both) are non-zero. Show that $V_n \to_p \infty$ in the sense that $P(V_n \leq M) \to 0$ for any $M > 0$.
(d) Suppose that $\{v_{n,\alpha}\}$ is such that $P(V_n > v_{n,\alpha}) \geq \alpha$ when the X_i's are uniformly distributed. If the X_i's satisfy the condition given in part (c), show that

$$\lim_{n\to\infty} P(V_n > v_{n,\alpha}) = 1$$

for any $\alpha > 0$.

Solution. (a) We apply Example 3.11 with

$$\sqrt{n}.(\begin{pmatrix}\overline{X_n^*}\\ \overline{Y_n^*}\end{pmatrix} - \begin{pmatrix}\mu_{X^*}\\ \mu_{Y^*}\end{pmatrix}) \to_d N_2(0,C): \qquad C = \begin{pmatrix}\sigma_{X^*}^2 & \sigma_{X^*.Y^*}\\ \sigma_{X^*.Y^*} & \sigma_{Y^*}^2\end{pmatrix}$$

in which $X_i^* = \sin(2.\pi.X_i)$ and $Y_i^* = \cos(2.\pi.X_i)$, $(1 \leq i \leq n)$. To calculate the entries of the matrix C we first calculate the C.D.F. and p.d.f of X_i^* in which:

$$\begin{aligned}
F_{X_i^*}(x) &= P(\sin(2.\pi.X_i) \leq x) = P((0 \leq 2.\pi.X_i \leq \arcsin(x)) \cup (2.\pi \geq 2.\pi.X_i \geq \pi - \arcsin(x)))\\
&= P((0 \leq X_i \leq \frac{\arcsin(x)}{2.\pi}) \cup (1 \geq X_i \geq \frac{\pi - \arcsin(x)}{2.\pi})) = \frac{\pi + 2.\arcsin(x)}{2.\pi}.1_{[-1,1]}(x),\\
f_{X_i^*}(x) &= \frac{d}{dx}(F_{X_i^*}(x)) = \frac{1}{\pi.\sqrt{1-x^2}}.1_{[-1,1]}(x).
\end{aligned}$$

Consequently:

$$\begin{aligned}
\mu_{X^*} &= \int_{-1}^{1} (\frac{x}{\pi.\sqrt{1-x^2}})dx = 0, \\
\sigma^2_{X^*} &= \int_{-1}^{1} (\frac{x^2}{\pi.\sqrt{1-x^2}})dx = (\frac{2}{\pi}).(\frac{\arcsin(x) - x.(1-x^2)^{\frac{1}{2}}}{2}|_0^1) = \frac{1}{2}, \\
\sigma_{X^*.Y^*} &= E[X^*.Y^*] = \int_0^1 \sin(2.\pi.x).\cos(2.\pi.x)dx = \frac{1}{2}\int_0^{4\pi} \sin(x)\frac{dx}{4.\pi} = 0, \\
\mu_{Y^*} &= \int_0^1 \cos(2.\pi.x)dx = \int_0^{2.\pi} \cos(x)\frac{dx}{2.\pi} = 0, \\
\sigma^2_{Y^*} &= E((Y^*)^2) = E(1-(X^*)^2) = 1 - E((X^*)^2) = 1 - \frac{1}{2} = \frac{1}{2}.
\end{aligned}$$

(b) Let $Y \sim \chi^2(p)$ and $c > 0$. Then (Exercise !), $c.Y \sim Gamma(\alpha = \frac{p}{2}, \lambda = \frac{1}{2c})$. Now, consider the function g defined by $g(U_1, U_2) = U_1^2 + U_2^2$. Hence by Part (a) and Theorem 3.2.(b) for independent $Z_1, Z_2 \sim N(0, \frac{1}{2})$ and the mentioned note for $p = 2$, $c = 1/2$ and $Y = (\frac{Z_1}{\sqrt{1/2}})^2 + (\frac{Z_2}{\sqrt{1/2}})^2$ we have:

$$\begin{aligned}
V_n &= g(\sqrt{n}\overline{X^*_n}, \sqrt{n}\overline{Y^*_n}) \to_d g(Z_1, Z_2) = Z_1^2 + Z_2^2 \\
&= \frac{1}{2}((\frac{Z_1}{\sqrt{1/2}})^2 + (\frac{Z_2}{\sqrt{1/2}})^2) =^d \frac{1}{2}\chi^2(2) =^d Gamma(1,1) =^d \exp(1).
\end{aligned}$$

(c) Let $X^*_i = \sin(2.\pi.X_i)$ $(1 \le i \le n)$ with $E(X^*_i) = \mu^* \neq 0$, (the solution for other cases is analogous). Then by Theorem 3.6., $\overline{X^*_n} \to_p \mu^*$. Hence, by Theorem 3.2(a) for $g(x) = x^2$ it follows that $(\overline{X^*_n})^2 \to_p (\mu^*)^2$, implying:

$$n.(\overline{X^*_n})^2 \to_p \infty. \quad (*)$$

On the other hand, $V_n \ge n.(\overline{X^*_n})^2$, $(n \ge 1)$ and for $M > 0$ given $(V_n \le M) \subseteq (n.(\overline{X^*_n})^2 \le M)$, we have:

$$P(V_n \le M) \le P((\overline{X^*_n})^2 \le M), \quad (n \ge 1). \quad (**)$$

Accordingly, by $(*)$ and $(**)$ the assertion follows.

(d) Let $v_{n,\alpha} = O(n^r)$, $(0 < r < 1)$, so that $\sup_{n\in\mathbb{N}} |\frac{v_{n,\alpha}}{n^r}| \le M^*_r < \infty$. Then, by $(n^{1-r}.((\overline{X^*_n})^2 + (\overline{Y^*_n})^2) \ge M^*_r) \subseteq ((n^{1-r}.((\overline{X^*_n})^2 + (\overline{Y^*_n})^2) \ge \frac{v_{n,\alpha}}{n^r})$, $(n \ge 1)$ we have:

$$1 \ge P(V_n \ge v_{n,\alpha}) = P((n^{1-r}.((\overline{X^*_n})^2 + (\overline{Y^*_n})^2) \ge \frac{v_{n,\alpha}}{n^r})) \ge P(n^{1-r}.((\overline{X^*_n})^2 + (\overline{Y^*_n})^2) \ge M^*_r) \quad (n \ge 1). \quad (***)$$

But, by a small modification of proof in Part(c):

$$\lim_{n\to\infty} P(n^{1-r}.((\overline{X^*_n})^2 + (\overline{Y^*_n})^2) \ge M^*_r) = 1, \quad (****)$$

and; finally, by considering $(****)$ in $(***)$, the assertion follows.
□

Problem 9.7. A Brownian Bridge process can be represented by the infinite series

$$B(x) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{n} \frac{\sin(\pi.kx)}{k} Z_k$$

where $Z_1, Z_2, \cdots$ are i.i.d. Normal random variables with mean 0 and variance 1.
(a) Assuming that the expected values can be taken inside infinite summations, show that

$$E[B(x)B(y)] = \min(x, y) - xy$$

for $0 \leq x, y \leq 1$.
(b) Define

$$W^2 = \int_0^1 B^2(x)dx$$

using the the infinite series representation of $B(x)$. Show that the distribution of W^2 is simply the limiting distributions of the Cramer-von Mises statistics.

Solution. (a) Let K be a symmetric positive definite kernel on a $\sigma-$finite measure space $([0,1], M, \mu)$ with an orthonormal set $\{\phi_k\}_{k=1}^{\infty}$ of $L^2_{\mu}([0,1])$ such that its correspondent sequence of eigenvectors $\{\lambda_k\}_{k=1}^{\infty}$ with condition $\lambda_k.\phi_k(t) = \int_0^1 K(t,s)\phi_k(s)ds \quad (k \geq 1)$ is non-negative. Then, K has the representation

$$K(x,y) = \sum_{k=1}^{\infty} \lambda_k.\phi_k(x).\phi_k(y)$$

with convergence in L^2 norm, (Mercer, 1909).

Now, for the Mercer series representation of the kernel function $K(x,y) = \min(x,y) - x.y$ with $\lambda_k = \frac{1}{k^2.\pi^2}$ and $\phi_k(t) = \sqrt{2}.\sin(k.\pi.t)$ we have:

$$\begin{aligned}
E(B(x).E(y)) &= E((\frac{\sqrt{2}}{\pi}\sum_{k_1=1}^{\infty}\frac{\sin(\pi.k_1.x)}{k_1}Z_{k_1}).(\frac{\sqrt{2}}{\pi}\sum_{k_2=1}^{\infty}\frac{\sin(\pi.k_2x)}{k_2}Z_{k_2})) \\
&= \frac{2}{\pi^2}[\sum_{k_1,k_2=1}^{\infty}(\frac{\sin(\pi.k_1.x)}{k_1}.\frac{\sin(\pi.k_2.x)}{k_2}.E(Z_{k_1}.Z_{k_2}))] \\
&= \frac{2}{\pi^2}[\sum_{k_1=k_2=k=1}^{\infty}(\frac{\sin(\pi.k_1.x)}{k_1}.\frac{\sin(\pi.k_2.x)}{k_2}.E(Z_{k_1}.Z_{k_2}))] \\
&+ \frac{2}{\pi^2}[\sum_{k_1\neq k_2=1}^{\infty}(\frac{\sin(\pi.k_1.x)}{k_1}.\frac{\sin(\pi.k_2.x)}{k_2}.E(Z_{k_1}).E(Z_{k_2}))] \\
&= \sum_{k=1}^{\infty}(\frac{1}{k^2.\pi^2}).(\sqrt{2}.\sin(\pi.k.x).\sqrt{2}.\sin(\pi.k.y)) \\
&= \min(x,y) - x.y.
\end{aligned}$$

(b) Let $U_k = F(X_k) \sim Unif[0,1] \quad (1 \leq k \leq n)$ be independent with order statistics $U_{1,n} \leq U_{2,n} \leq \cdots \leq U_{n,n}$. Then,(Csorgo & Faraway, 1996):

$$\frac{1}{12.n} \leq W_n^2 = \frac{1}{12.n} + \sum_{k=1}^{n}(U_{k,n} - \frac{2k-1}{2n})^2 \leq \frac{n}{3}, \quad (n \geq 1) \quad (*),$$

where $W_n^2 = \frac{n}{3}$ if $U_{n,n} = 0$ or $U_{1,n} = 1$. Define $V_n(x) = F_{W_n^2}(x), \quad (-\infty < x < \infty)$. Then, $V_n(x) =$

$0, \quad (x \leq \frac{1}{12.n})$ and $1, \quad (x \geq \frac{n}{3})$. Now, by $(*)$ for the corresponding characteristics functions we have:

$$\begin{aligned}
\lim_{n\to\infty} \phi_{W_n^2}(t) &= \lim_{n\to\infty} E(e^{i.t.W_n^2}) = \lim_{n\to\infty} \int_{\frac{1}{12.n}}^{\frac{n}{3}} e^{i.t.x} dV_n(x) \\
&= (\frac{(-2.i.t)^{\frac{1}{2}}}{\sinh((-2.i.t)^{\frac{1}{2}})}) = \int_0^\infty e^{itx} dV(x) = \phi_V(t), \ (-\infty < t < \infty). \quad (**)
\end{aligned}$$

Accordingly, by $(**)$ and comments on Page 126, it follows that $W_n^2 \to_d W^2$.
□

Problem 9.9. Suppose that $X_1, \cdots, X_n$ are i.i.d. Exponential random variables with parameter λ. Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics and define the so-called normalized spacing (Pyke,1965)

$$\begin{aligned}
D_1 &= n.X_{(1)} \\
D_k &= (n-k+1).(X_{(k)} - X_{(k-1)}) \ \ (k = 2, \cdots, n).
\end{aligned}$$

According to Problem 2.26, $D_1, \cdots, D_n$ are also i.i.d. Exponential random variables with parameter λ.
(a) Let $\overline{X_n}$ be the sample mean of $X_1, \cdots, X_n$ and define

$$T_n = \frac{1}{n.(\overline{X_n})^2} \sum_{i=1}^n D_i^2.$$

Show that $\sqrt{n}(T_n - 2) \to_d N(0, 20)$.
(b) Why might T_n be a useful test statistic for testing the null hypothesis that the X_i's are Exponential?

Solution. (a) With the notation on Page 29, if $X \sim \exp(\lambda)$, then (Exercise !) $Y = X^{\frac{1}{\beta}} \sim Weibull(\lambda, \beta)$ with $E(Y) = (\frac{1}{\lambda})^{(\frac{1}{\beta})}.\Gamma(1+\frac{1}{\beta})$ and $Var(Y) = (\frac{1}{\lambda})^{(\frac{2}{\beta})}.[\Gamma(1+\frac{2}{\beta}) - \Gamma^2(1+\frac{1}{\beta})]$. Consequently, for $\beta = \frac{1}{2}$, we have $D_i^2 \sim^{i.i.d} Weibull(\lambda, \frac{1}{2})$ with $E(D_i^2) = \frac{2}{\lambda^2}$ and $Var(D_i^2) = \frac{20}{\lambda^4} \ \ (1 \leq i \leq n)$. Also, $\lambda.X_i \sim \exp(1)$ with $E(\lambda.X_i) = 1, \ \ (1 \leq i \leq n)$.Given these conclusions we have:

First, by Theorem 3.8 for $X_i^* = D_i^2, \mu_i^* = \frac{2}{\lambda^2}$ and $(\sigma^*)^2 = \frac{20}{\lambda^4}$, it follows that $\sqrt{n}.(\frac{\sum_{i=1}^n D_i^2}{n} - \frac{2}{\lambda^2}) \to_d N(0, \frac{20}{\lambda^4})$, and, by Theorem 3.4. for $g(x) = \lambda^2.x$:

$$\lambda^2.\sqrt{n}.(\frac{\sum_{i=1}^n D_i^2}{n} - \frac{2}{\lambda^2}) \to_d N(0, 20). \quad (*)$$

Second, by theorem 3.6. for $X_i^{**} = \lambda.X_i$, $\overline{\lambda.X_n} \to_p 1$, and by Theorem 3.2(a) for $g(x) = \frac{1}{x^2}$:

$$\frac{1}{(\overline{\lambda.X_n})^2} \to_p 1. \quad (**)$$

Third, given definition of T_n one may write:

$$\sqrt{n}(T_n - 2) = \lambda^2.\sqrt{n}.(\frac{1}{(\overline{\lambda.X_n})^2}.\frac{\sum_{i=1}^n D_i^2}{n} - \frac{2}{\lambda^2}). \quad (***)$$

Finally, an application of Theorem 3.3.(b) for $(*), (**)$, and $(***)$ proves the assertion.

(b) We apply Lagrange Multipliers Method for the function $f(a_1, \cdots, a_n) = a_1^2 + \cdots + a_n^2$ with constraint function $g(a_1, \cdots, a_n) = a_1 + \cdots + a_n - k$. Consider:

$$l(a_1, \cdots, a_n; \lambda) = f(a_1, \cdots, a_n) - \lambda . g(a_1, \cdots, a_n),$$

and, then, the system of equations

$$\begin{aligned} \frac{dl}{da_1} &= 2.a_1 - \lambda = 0, \\ &\cdots \\ \frac{dl}{da_n} &= 2.a_n - \lambda = 0, \\ \frac{dl}{\lambda} &= -(a_1 + \cdots + s_n - k) = 0, \end{aligned}$$

has the solution $a_1 = \cdots = a_n = \frac{\lambda}{2}$ by its first n equations and $\frac{\lambda}{2} = \frac{k}{n}$ by its last equation. Consequently, by the last two results, we have $a_i = \frac{k}{n}, \ \ (1 \leq i \leq n)$.

Finally, under the null hypothesis $D_i \sim \exp(\lambda), \ \ (1 \leq i \leq n)$ and considering $D_1 + \cdots + D_n = n.\overline{X_n} = k$, we notice the values of T_n and hence $\sqrt{n}(T_n - 2)$ are minimized allowing one not to potentially reject the null hypothesis.

□

Bibliography

[1] Knight, K. (2000) *Mathematical Statistics*, CRC Press, Boca Raton, ISBN: 1-58488-178-X

[2] Casella, G. & Berger, R.L. (2002) *Statistical Inference*, Brooks/Cole Cengage Learning, Second edition, ISBN-13: 978-0-495-39187-6

[3] David, H.A. (1981). *Order Statistics.* New york: John Wiley & Sons.

[4] Tate, R.F. & Klett, G.W. (1959). Optimal Confidence Intervals for a Variance of Normal Distribution. *J. Amer. Statis. Assoc.* 54, 674-682

[5] Floudas, C.A. & Pardalos, P.M (Eds). (2009). *Encyclopaedia of Optimization.* Second Edition, Springer. ISBN: 987-0-0387-74759-0

[6] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* Second edition. Hoboken, NJ: Wiley.

[7] Belzunce, F, Martinez-Riquelme. C, & Mulero, J. (2016). *An Introduction to Stochastic Orders.* Academic Press.

[8] Hall, D.B. (2000). Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometric.* 56, 1030-1039.

[9] Powell, M.J. (1964). An Efficient Method for Finding the Minimum of Functions of Several Variables without Calculating the Derivatives. *Computer Journal* 7(2): 155- 162.

[10] Mercer.J. (1909). Functions of Positive and Negative Type and Their Connection to the Theory of Integral Equations. *Philosophical Transactions of the Royal Society*, A 209, 415-446.

[11] Csorgo, S & Faraway, J.J. (1996). The Exact and Asymptotic Distributions of Cramer-von Mises Statistics. *J. R. Statistical. Society.* B. 58(1), 221-234.