PhD Prelim Exam
THEORY
(Majors and Co-majors)

Summer 2009
(Given on 7/9/09)

Here, $I\!R$ denotes the set of all reals, $I\!N$ denotes the set of natural numbers and $\mathcal{B}(S)$ denotes the Borel sigma-field of a metric space S. m denotes the Lebesgue measure on $(I\!R, \mathcal{B}(I\!R))$.

Problem 1

(a) Define what is meant by a sequence of i.i.d. Normal(α, β^2) random variables $\{X_n, n \geq 1\}$, defined on $(\Omega, \mathcal{F}, P)$. *[Carefully define the terms "i.i.d.", "Normal" and "random variable"]*

(b) State the following inequalities for random variables defined on a probability space $(\Omega, \mathcal{F}, P)$: (i) Chebychev's inequality, (ii) Jensen's inequality, (iii) Hölder's inequality.

(c) Let $(\Omega, \mathcal{F})$ be a measurable space and $\Omega' \subset \Omega$. Define a new collection $\mathcal{F}_{\Omega'} = \{A \cap \Omega' : A \in \mathcal{F}\}$. Prove $\mathcal{F}_{\Omega'}$ is a σ-algebra of subsets of Ω' ($\mathcal{F}_{\Omega'}$ is known as the *trace $\sigma-$algebra of $\mathcal{F}$ on Ω'*).

(d) Suppose Z is a random variable defined on some probability space $(\Omega, \mathcal{F}, P)$. Prove that $Z \in L^q(\Omega, \mathcal{F}, P)$ implies $Z \in L^p(\Omega, \mathcal{F}, P)$, where $0 < p < q < \infty$.

Problem 2

(a) **(i)** State *Lebesgue's Dominated convergence Theorem (DCT)* for $f, f_1, f_2, \ldots$ real-valued measurable functions defined $(\Omega, \mathcal{F}, \mu)$, such that $f_n \to f$ *a.e.*(μ).
(ii) Prove that the conclusion of the DCT remains true even if "$f_n \to f$ a.e (μ) " is replaced by "$f_n \to f$ in measure" in the statement. *[If you are using real analysis facts in the proof, state them clearly.]*

(b) Let μ, ν be two measures on $(I\!R, \mathcal{B}(I\!R))$ and $f, g \geq 0$. Define the following
$$\mu * \nu, \ f * g, \ f * \mu$$
for $*$ indicating convolution.

(c) Let μ, ν be two measures on $(I\!R, \mathcal{B}(I\!R))$ and $\mu \ll m$, then show that
$$\frac{d(\mu * \nu)}{dm} = \frac{d\mu}{dm} * \nu$$

(d) Suppose X has distribution $\mu \equiv Exponential(1)$ and Y has distribution $\nu \equiv Binomial(3, 0.5)$ and assume X and Y are independent.
(i) Express the distribution of $(X + Y)$ using convolution notations.
(ii) Does $(X + Y)$ have a probability density function? If yes, write down the precise form of the density. If no, justify why it can not have a density.

Problem 3

(a) Let $\alpha \in I\!R, \beta \geq 0$ be constants. **(i)** Show that a $Normal(\alpha, \beta^2)$ random variable exists: In other words, using steps of the Caratheodory extension theorem, exhibit a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X : \Omega \to I\!R$, such that its c.d.f. is given by Φ where
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(u-\alpha)^2}{2\beta^2}} du, \ x \in I\!R.$$

(ii) Compute the following Lebesgue integral

$$\int_{\Omega} X(\omega) dP(\omega).$$

[Clearly state any result used in computing the above integral.]

(b) Let $\alpha \in \mathbb{R}$. Suppose X is a $Normal(0,1)$ random variable defined on some $(\Omega, \mathcal{F}, P)$. Define another measure $\tilde{P}$ on $(\Omega, \mathcal{F})$ as follows:

$$\tilde{P}(A) = \int_A e^{(\alpha X(\omega) - \frac{1}{2}\alpha^2)} dP(\omega), \quad A \in \mathcal{F}.$$

Show that the same measurable transformation X as a random variable on $(\Omega, \mathcal{F}, \tilde{P})$ is a $Normal(\alpha, 1)$ random variable. *[A similar "change of measure" result for Brownian motions is called the Girsanov's Theorem, used in many calculations (e.g. in Math. Finance).]*

(c) **(i)** For each $n \geq 1$, let μ^n be a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Under what conditions can we say that there is a probability measure P on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ such that $P(A \times \mathbb{R}^\infty) = \mu^n(A)$ for any $A \in \mathcal{B}(\mathbb{R}^n)$, for any $n \geq 1$? Describe a sequence of random variables $\{X_n : n \geq 1\}$ defined on $(\Omega = \mathbb{R}^\infty, \mathcal{F} = \mathcal{B}(\mathbb{R}^\infty), P)$ such that the joint distribution of $(X_1, \ldots, X_n)$ is given by μ^n, for all $n \geq 1$.
(ii) Using part (i) or otherwise, show that there exists $\{X_n : n \geq 1\}$ which is an i.i.d. sequence of $Normal(\alpha, \beta^2)$ random variables.

(d) Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d. $Normal(\alpha, \beta^2)$ random variables defined on $(\Omega, \mathcal{F}, P)$, each with density function $\phi_{(\alpha,\beta^2)}(\cdot)$. Let $x_1, \ldots, x_n \in \mathbb{R}$ and $L(\alpha, \beta^2; x_1, \ldots, x_n) = \Pi_{i=1}^n \phi_{(\alpha,\beta^2)}(x_i)$, then define

$$\Big(T_\alpha(x_1, \ldots, x_n), \ T_{\beta^2}(x_1, \ldots, x_n)\Big) = argmax_{(\alpha,\beta^2)} \, L(\alpha, \beta^2; x_1, \ldots, x_n).$$

Define *MLEs* as $T_1(\omega) = T_\alpha(X_1(\omega), \ldots, X_n(\omega))$, $T_2(\omega) = T_{\beta^2}(X_1(\omega), \ldots, X_n(\omega))$ for $\omega \in \Omega$.
(i) Briefly argue that T_1 and T_2 are random variables.
(ii) Stating clearly all results used, prove the *strong consistency of MLEs.* i.e.,

$$(T_1, T_2) \rightarrow (\alpha, \beta^2) \text{ with probability } 1 \ (P).$$

Problem 1 [6+9+7+8 = **30 points**]

(a) For a given probability space $(\Omega, \mathcal{F}, P)$, a function $X : \Omega \to I\!R$ is called a random variable, if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(I\!R)$. A sequence $\{X_n, n \geq 1\}$ of random variables defined on $(\Omega, \mathcal{F}, P)$ is called i.i.d. $Normal(\alpha, \beta^2)$ random variables if for each $n \geq 1$, the law (or the distribution or the induced measure) of the random variable on $(I\!R, \mathcal{B}(I\!R))$ denoted by P_{X_n} is :

$$P_{X_n}(A) = \int_A \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(x-\alpha)^2}{2\beta^2}} dm(x), \; A \in \mathcal{B}(I\!R),$$

and these sequence of random variables are independent w.r.t P, i.e. $P(X_1 \in B_1, \ldots, X_n \in B_n) = \Pi_{i=1}^n P(X_i \in B_i)$, for all $B_1, \ldots, B_n \in \mathcal{B}(I\!R)$ and for all $n \geq 1$.

(b) Suppose X is a random variable defined on $(\Omega, \mathcal{F}, P)$

(i) Chebychev Inequality: Let $E_P(X^2) < \infty$, and let $\mu = E_P(X), \sigma^2 = Var_P(X)$. Then

$$P\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}.$$

(ii) Jensen Inequality: If $P(X \in (a,b)) = 1$ for some interval (a,b), for some $\infty \leq a < b \leq \infty$ and let $\phi : (a,b) \to I\!R$ be convex. Then

$$\phi\left(\int X dP\right) \leq \int \phi(X) dP,$$

provided $\int X dP$ and $\int \phi(X) dP$ are finite.

(iii) Hölder Inequality: Let $1 < r < \infty$ and $X \in L^r(\Omega, \mathcal{F}, P)$ and Y is another random variable such that $Y \in L^s(\Omega, \mathcal{F}, P)$, with $s = r/(r-1)$. Then,

$$\int |XY| dP \leq \left(\int |X|^r dP\right)^{\frac{1}{r}} \left(\int |Y|^s dP\right)^{\frac{1}{s}}.$$

(c) To show that $\mathcal{F}_{\Omega'}$ is a $\sigma-$algebra of subsets of Ω', we verify the following:

(i) with $A = \Omega \in \mathcal{F}$, one gets $\Omega' \in \mathcal{F}_{\Omega'}$. Note that the whole space here is Ω'.

(ii) For the complement of a set E in Ω', we have to verify that if $E = A \cap \Omega' \in \mathcal{F}_{\Omega'}$, then $\Omega' \setminus E$ is in $\mathcal{F}_{\Omega'}$. But $\Omega' \setminus E = A^c \cap \Omega'$, and $A^c \in \mathcal{F}$, which implies that it belongs to $\mathcal{F}_{\Omega'}$ (note that here we used the notation $A^c = \Omega \setminus A$).

(iii) For the closure under countable union property, note that if $E_n = A_n \cap \Omega' \in \mathcal{F}_{\Omega'}, n \geq 1$, then $\cup_{n\geq 1} E_n = (\cup_{n \geq 1} A_n) \cap \Omega'$, and since $\{A_n\} \subset \mathcal{F}$ which is a $\sigma-$algebra, we know that $(\cup_{n\geq 1} A_n) \in \mathcal{F}$. Hence, from the representation above, $\cup_{n \geq 1} E_n \in \mathcal{F}_{\Omega'}$.

(d) Choose $X = Z^p$ and $Y \equiv 1$ as the constant random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Using the Hölder's inequality as stated in Problem 1(b) with $r = q/p > 1$, and $s = r/(r-1)$ that:

$$\int |Z|^p dP = \int |XY| dP \leq \left(\int |X|^r dP\right)^{\frac{1}{r}} \left(\int |Y|^s dP\right)^{\frac{1}{s}} = \left(\int |X|^{\frac{q}{p}} dP\right)^{\frac{p}{q}} .1 = \left(\int |Z|^q dP\right)^{\frac{p}{q}}$$

Which proves that $Z \in L^q(\Omega, \mathcal{F}, P)$ implies $Z \in L^p(\Omega, \mathcal{F}, P)$, where $0 < p < q < \infty$.

Problem 2 [(4+8)+9+7+(2+5) = **35 points**]

(a) (i) *Lebesgue's Dominated convergence Theorem (DCT):* Let $\{f_n\}, f, g$ be real-valued measurable functions defined $(\Omega, \mathcal{F}, \mu)$. If $|f_n| \le g$ *a.e.*(μ), and $g \in L^1(\Omega, \mathcal{F}, \mu)$, then $f_n \longrightarrow f$ *a.e.*(μ) implies,

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu, \quad \lim_{n\to\infty} \int |f_n - f| d\mu = 0.$$

(ii) Suppose $f_n \to f$ in measure, then for any subsequence $\{f_{n_k}\}$ of $\{f_n\}$, we have $f_{n_k} \to f$ in measure, as $k \to \infty$. For this subsequence, we know that there exists a *further* subsequence of functions $\{f_{n_{k_m}}\}$ such that $f_{n_{k_m}} \to f$ *a.e.*(μ), as $m \to \infty$ (Theorem 2.5.2 of the text). From the conditions of f_n, we have $|f_{n_{k_m}}| \le g$ and g satisfies all the properties of the DCT above. Hence, by applying DCT on the sub-subsequence $\{f_{n_{k_m}}\}$ we have

$$\lim_{m\to\infty} \int f_{n_{k_m}} d\mu = \int f d\mu, \quad \lim_{m\to\infty} \int |f_{n_{k_m}} - f| d\mu = 0.$$

Note that from the above argument, we have that for any subsequence of the real sequences $\{\int f_n d\mu\}$ and $\{\int |f_n - f| d\mu\}$, there exists a sub-subsequence along which the above convergence statements hold. Hence, the above convergence holds for the sequences $\{\int f_n d\mu\}$ and $\{\int |f_n - f| d\mu\}$, i.e.,

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu, \quad \lim_{n\to\infty} \int |f_n - f| d\mu = 0.$$

In other words, the conclusions of the DCT remains true in this case.

(b)

$$\begin{aligned}
(\mu * \nu)(A) &= \int\int I_A(x+y) d\mu(x) d\nu(y), \ A \in \mathcal{B}(I\!R), \\
(f * g)(x) &= \int f(x-y) g(y) dm(y), \ x \in I\!R, \\
(f * \mu)(x) &= \int f(x-y) d\mu(y), \ x \in I\!R.
\end{aligned}$$

(c) If $\mu \ll m$, then with $f = \frac{d\mu}{dm}$ we have by the definition of Radon-Nikodym derivatives, the following holds for $A \in \mathcal{B}(I\!R)$ (proving the claim):

$$\begin{aligned}
(\mu * \nu)(A) &= \int\int I_A(x+y) d\mu(x) d\nu(y) = \int \left(\int I_A(x+y) f(x) dm(x) \right) d\nu(y) \\
&= \int \left(\int I_A(u) f(u-y) dm(u) \right) d\nu(y) \quad \text{(by change of variable } u = x+y \\
&\qquad\qquad \text{and translation invariance of Lebesgue measure } m) \\
&= \int I_A(u) \left(\int f(u-y) d\nu(y) \right) dm(u) \quad \text{(by changing the order of} \\
&\qquad\qquad \text{integration, using Fubini's Theorem, with } f \ge 0) \\
&= \int_A (f * \nu)(u) dm(u),
\end{aligned}$$

(d) Let $X \sim \mu \equiv Exponential(1)$ and $Y \sim \nu \equiv Binomial(3, 0.5)$ and they are independent, then **(i)** the law of $X + Y$ is $\mu * \nu$. It follows from the fact that by independence, $P_{(X,Y)} = \mu \times \nu$ and with $h(x, y) = x + y$, we have $P(X + Y \in A) = (\mu \times \nu)h^{-1}(A) = (\mu * \nu)(A)$ for any $A \in \mathcal{B}(R)$.
(ii) Yes, $X + Y$ does have a density (w.r.t. Lebesgue measure) as we saw in part (c) above. It is given by

$$\begin{aligned}\left(\frac{d\mu}{dm} * \nu\right)(x) &= \int e^{(x-y)} I_{\{x-y>0\}} d\nu(y) \\ &= (0.5)^3 \left[e^{-x} I_{\{x>0\}} + 3e^{-x+1} I_{\{x>1\}} + 3e^{-x+2} I_{\{x>2\}} + e^{-x+3} I_{\{x>3\}}\right].\end{aligned}$$

Problem 3 [(5+5)+8+(3+4)+(3+7) = **35 points**]

(a) **(i)** Define $\mathcal{C} = \{(a, b], (b, \infty) : -\infty \leq a, b < \infty\}$, and define $\mu_\Phi((a, b]) = \Phi(b) - \Phi(a)$, $\mu_\Phi((b, \infty)) = 1 - \Phi(b)$. This defines a measure on the semi-algebra $\mathcal{C}$. Caratheodory extension theorem states that there exists an extension μ^* of μ_Φ to $(I\!R, \mathcal{M})$, and $\mathcal{C} \subset \mathcal{M}$, where

$$\mu^*(A) = \inf\left\{\sum_{n=1}^{\infty} \mu_\Phi(A_n) : \{A_n\} \subset \mathcal{C}, A \subset \cup_{n=1}^{\infty} A_n\right\}, \ A \subset I\!R,$$

and $\mathcal{M}$ is the set of μ^*-measurable sets, i.e. $\mathcal{M} = \{A \subset I\!R : \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c), \ \forall E \subset I\!R\}$. Since, $\mathcal{C} \subset \mathcal{M}$, and we know that $\mathcal{B}(I\!R) = \sigma\langle\mathcal{C}\rangle$, we have $\mathcal{B}(I\!R) \subset \mathcal{M}$. Hence, we get and extension of μ_Φ to $(I\!R, \mathcal{B}(I\!R))$ by restricting μ^* to $(I\!R, \mathcal{B}(I\!R))$. We will denote this measure as μ on $(I\!R, \mathcal{B}(I\!R))$. Define $(\Omega, \mathcal{F}, P) = (I\!R, \mathcal{B}(I\!R), \mu)$. Define $X : \Omega \to I\!R$ as $X(\omega) = \omega$. Clearly, it is measurable and μ is a probability (and hence X is a random variable). Also, the law of X is $P_X(A) = P(X^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}(I\!R)$. $F_X(x) = P(X \leq x) = \mu((-\infty, x]) = \Phi(x)$, for all $x \in I\!R$.
(ii) Here we use the following change of variable formula for Lebesgue integrals: If X is a measurable transformation from $(\Omega, \mathcal{F}, P)$ to $(I\!R, \mathcal{B}(I\!R))$, then $X \in L^1(\Omega, \mathcal{F}, P)$ iff $I \in L^1(I\!R, \mathcal{B}(I\!R), P_X)$, where $I(x) = x$, for $x \in I\!R$, and in that case,

$$\int X dP = \int_{I\!R} I(x) dP_X.$$

From the definition of P_X in the problem and relationship between Riemann and Lebesgue integrals, we know that the right side of the above equation is

$$\int_{I\!R} I(x) dP_X = \int_{I\!R} I(x) \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(x-\alpha)^2}{2\beta^2}} dm(x) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(x-\alpha)^2}{2\beta^2}} dx,$$

where is the last expression is a usual Riemann integral. We know from elementary calculus, that this integral is equal to α.

(b) Here, we need to verify that the law of X under $\tilde{P}$ is given by the $Normal(\alpha, 1)$ probability on $(I\!R, \mathcal{B}(I\!R))$. Note that, by definition $\frac{d\tilde{P}}{dP} = e^{(\alpha X - \frac{1}{2}\alpha^2)}$. Hence, using the change of variable theorem, we have for $A \in \mathcal{B}(I\!R)$

$$\begin{aligned}
\tilde{P}(X \in A) &= \int I_{\{X(\omega)\in A\}} d\tilde{P}(\omega) = \int I_{\{X(\omega)\in A\}} e^{(\alpha X(\omega) - \frac{1}{2}\alpha^2)} dP(\omega) \\
&= \int I_{\{x\in A\}} e^{(\alpha x - \frac{1}{2}\alpha^2)} dP_X(x) = \int_A e^{(\alpha x - \frac{1}{2}\alpha^2)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dm(x) \\
&= \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2}} dm(x).
\end{aligned}$$

(c) **(i)** From Kolmogorov's Consistency theorem, only condition we need here is that $\mu^{(n+1)}(A \times I\!R) = \mu^n(A)$ for all $A \in \mathcal{B}(I\!R^n)$ for all $n \geq 1$. For the sequence of random variables, define for $\omega = (\omega_1, \omega_2, \ldots) \in I\!R^\infty$, $X_n(\omega) = \omega_n$, $n \geq 1$. Then the joint law of $(X_1, \ldots, X_n)$ is given by

$$P_{(X_1,\ldots,X_n)}(A) = P((X_1, \ldots, X_n) \in A) = P(A \times I\!R^\infty) = \mu^n(A), \quad \text{for each } A \in \mathcal{B}(I\!R^n),\ n \geq 1.$$

(ii) In part **(i)**, define $\mu^{(1)} = \mu \equiv N(\alpha, \beta^2)$, and $\mu^{(n+1)} = \mu^{(n)} \times \mu$, for $n \geq 1$. That is, for each $n \geq 1$ μ^n is the product measure generated by μ on $(I\!R^n, \mathcal{B}(I\!R^n))$. Then for each $n \geq 1$, μ^n is a probability on $(I\!R^n, \mathcal{B}(I\!R^n))$ and $\mu^{(n+1)}(A \times I\!R) = \mu^n(A)$ for $A \in \mathcal{B}(I\!R^n)$. Hence, as in part **(i)**, we can construct a sequence of random variables $\{X_n : n \geq 1\}$ on some probability space $(\Omega, \mathcal{F}, P)$ such that joint distribution of $(X_1, \ldots, X_n)$ is given by μ^n, which is the product measure generated by μ. Hence $\{X_n : n \geq 1\}$ is an i.i.d. sequence.

(d) (i) By definition,

$$L(\alpha, \beta^2; x_1, \ldots, x_n) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(x_i-\alpha)^2}{2\beta^2}}\right] = C \exp\left(-\frac{n}{2}\ln\left(\beta^2\right) - \frac{1}{2\beta^2}\sum_{i=1}^n (x_i - \alpha)^2\right),$$

where C is a constant. This function is maximized at

$$\Big(T_\alpha(x_1, \ldots, x_n), T_{\beta^2}(x_1, \ldots, x_n)\Big) = \left(\frac{\sum_{i=1}^n x_i}{n},\ \frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2\right).$$

Since $T \equiv (T_1, T_2) : I\!R^n \to I\!R^2$ is a continuous function (and hence, measurable function), and $(X_1, \ldots, X_n) : \Omega \to I\!R^n$ is a measurable transformation, we get that T_1, T_2 are random variables.

(ii) Note that from $E_P(X_1) = \alpha$ and $E_P(X_1^2) = Var_P(X_1) + (E_P(X_1))^2 = \beta^2 + \alpha^2$ (can compute using calculations similar to Problem 1 (C)). Since $Y_n \equiv (Y_n^{(1)}, Y_n^{(2)}) = (X_n, X_n^2), n \geq 1$ are i.i.d random variables (vector valued), such that $E_P(|Y_1^{(1)}|) \leq (E_P(|Y_1^{(2)}|))^{\frac{1}{2}} < \infty$, we have from Etamadi's SLLN (vector valued version) that

$$\left(\frac{1}{n}\sum_{i=1}^n Y_i^{(1)},\ \frac{1}{n}\sum_{i=1}^n Y_i^{(2)}\right) \to \left(E_P(Y_1^{(1)}),\ E_P(Y_1^{(2)})\right) = \left(\alpha,\ \beta^2 + \alpha^2\right), \text{ a.e. } (P).$$

Now since, $h(x, y) = (x, y - x^2) : I\!R^2 \to I\!R^2$ is a continuous function, we get

$$(T_1, T_2) = h\left(\frac{1}{n}\sum_{i=1}^n Y_i^{(1)},\ \frac{1}{n}\sum_{i=1}^n Y_i^{(2)}\right) \to h\left(\alpha,\ \beta^2 + \alpha^2\right) = \left(\alpha,\ \beta^2\right), \text{ a.e. } (P).$$

1. Let $X_1, X_2, \ldots$ be i.i.d. real-valued random variables with unknown continuous distribution function $F(x) = P((-\infty, x])$, and let $F_n(x) = P_n((-\infty, x])$ be the empirical distribution function based on the sample $X_1, \ldots, X_n$ where $P_n(\cdot)$ is the empirical measure, i.e. for each Borel measurable set $A \in \mathcal{B}(\Re)$)

$$P_n(A) := \frac{1}{n}\sum_{i=1}^{n} 1_A(X_i)$$

where $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ if $x \notin A$.

   (a) Show that $P_n(A)$ is an unbiased estimator of $P(A)$ for each Borel measurable set $A \in \mathcal{B}(\Re)$.

   (b) For any fixed $x \in \Re$ such that $F(x) \in (0,1)$, derive the asymptotic distribution of $\sqrt{n}(F_n(x) - F(x))$.

   (c) Prove that $K_n := \sqrt{n}\sup_{x\in\Re}|F_n(x) - F(x)|$ has a distribution which does not depend on the true continuous distribution function F, i.e. K_n is a "distribution-free statistic."

   (d) Explain whether each of the following two statements is true:

$$\lim_{n\to+\infty}\sup_{x\in\Re}|P_n((-\infty,x]) - P((-\infty,x])| = 0 \qquad a.s. \tag{1}$$

$$\lim_{n\to+\infty}\sup_{A\in\mathcal{B}(\Re)}|P_n(A) - P(A)| = 0 \qquad a.s. \tag{2}$$

2. Let $(X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n)$ be independently normally distributed random variables, all with the same variance θ but with $EX_i = EY_i = \mu_i$, $i = 1, 2, \ldots, n$, all possibly different. The parameter θ is of interest for inference while the parameters μ_i, $i = 1, 2, \ldots, n$ are considered nuisance parameters here.

   (a) Obtain the maximum likelihood estimator (MLE) $\hat{\theta}$ for θ.

   (b) Show that the MLE $\hat{\theta}$ is not consistent for θ.

   (c) Show how the MLE $\hat{\theta}$ can be rescaled to give a consistent estimator for θ.

   (d) Show how confidence intervals for θ can be obtained.

   (e) What is the Fisher information for θ in (X_1, Y_1) when μ_1 is fixed? What property of the Fisher information suggests that the MLE $\hat{\theta}$ from (a) is biased (even asymptotically)?

Continue on the next page!

3. Suppose a potentially biased coin is independently tossed three times. For $i = 1, 2, 3$, let X_i be the number of heads observed in the ith toss. A natural model is $X_i \sim$ Bernoulli(θ), $\theta \in \Theta = (0, 1)$. We consider estimating θ^2 based on these i.i.d. observations.

   (a) What is the MLE for θ^2? Is it unbiased? (Explain.)

   (b) Derive the UMVUE for θ^2.

   (c) If there is exactly one head in these three tosses, what is the value of the UMVUE of θ^2? Explain whether you think the UMVUE is a better estimator than the MLE here.

   (d) Consider the following hierarchy:

$$X_i|\theta \sim \text{Bernoulli}(\theta),\ i = 1, 2, 3,$$

$$\theta \sim U(0, 1).$$

   Under this hierarchy, find the mean and variance of the random sum $S = X_1 + X_2 + X_3$.

4. Suppose $X_1, \ldots, X_n$ are i.i.d. random observations with a double exponential density $f(x|\theta_0)$, where for any θ

$$f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}, \quad \theta \in \Re$$

   Note that $f(x|\theta)$ is symmetric about θ and has variance $\sigma^2 = 2$.

   (a) Write down the log-likelihood function.

   (b) Derive the MLE $\hat{\theta}$ of θ_0 based on the sample $X_1, \ldots, X_n$.

   (c) Identify the limiting distribution of the MLE, i.e. the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$, and specify an asymptotic confidence interval for θ_0 with confidence $1 - \alpha$.

   (d) Show that the sample mean $\overline{X}$ is unbiased for θ_0 and identify the limiting distribution of $\sqrt{n}(\overline{X} - \theta_0)$. Is $\overline{X}$ asymptotically more efficient than the MLE $\hat{\theta}$?

   (e) Now suppose the observations $X_1, \ldots, X_n$ are actually an i.i.d. sample from a normal distribution with some unknown location paramter θ_0, i.e. $N(\theta_0, 1)$. Suppose the MLE in (b) obtained under the assumption of a double exponential density is used for inference on θ_0. Is the MLE $\hat{\theta}$ from part (b) a consistent estimator of θ_0? Will the asymptotic confidence interval specified in part (c) have coverage probability $1 - \alpha$ as $n \to \infty$?

1. Let $X_1, X_2, \ldots$ be i.i.d. real-valued random variables with unknown continuous distribution function $F(x) = P((-\infty, x])$, and let $F_n(x) = P_n((-\infty, x])$ be the empirical distribution function based on the sample $X_1, \ldots, X_n$ where $P_n(\cdot)$ is the empirical measure, i.e. for each Borel measurable set $A \in \mathcal{B}(\Re)$)

$$P_n(A) := \frac{1}{n}\sum_{i=1}^{n} 1_A(X_i)$$

   (a) Is $P_n(A)$ an unbiased estimator of $P(A)$ for each Borel measurable set $A \in \mathcal{B}(\Re)$?

   (b) For any fixed $x \in \Re$ such that $F(x) \in (0,1)$, derive the asymptotic distribution of $\sqrt{n}(F_n(x) - F(x))$.

   (c) Prove that $K_n := \sqrt{n}\sup_{x\in\Re}|F_n(x) - F(x)|$ has a distribution which does not depend on the true continuous distribution function F, i.e. K_n is a "distribution-free statistic".

   (d) Explain whether each of the following two statements are true:

$$\lim_{n\to+\infty}\sup_{x\in\Re}|P_n((-\infty,x]) - P((-\infty,x])| = 0 \qquad a.s. \tag{1.4.1}$$

$$\lim_{n\to+\infty}\sup_{A\in\mathcal{B}(\Re)}|P_n(A) - P(A)| = 0 \qquad a.s. \tag{1.4.2}$$

Solution

(a) Yes, $P_n(A)$ is an unbiased estimator of $P(A)$. In fact, $\forall A \in \mathcal{B}(\Re)$

$$\begin{aligned} \mathrm{E}P_n(A) &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}1_A(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}P(X_i \in A) \\ &= P(A) \end{aligned} \tag{1}$$

(b) Note that $\{1_{(-\infty,x]}(X_i)\}_{i=1}^{n}$ are i.i.d. Bernoulli RV's with

$$P(1_{(-\infty,x]}(X_i) = 1) = F(x)$$

$$\begin{aligned} \mathrm{E}(1_{(-\infty,x]}(X_i)) &= F(x) \\ \mathrm{Var}(1_{(-\infty,x]}(X_i)) &= F(x)[1-F(x)] \end{aligned}$$

When $F(x) \in (0,1)$, because $0 < \mathrm{Var}(1_{(-\infty,x]}(X_i)) < 1$, by the CLT

$$\sqrt{n}(F_n(x) - F(x)) \to^d N(0, F(x)[1-F(x)])$$

(c) Because when $F(x) = 0$ or $F(x) = 1$, $F_n(x) - F(x) = 0,\ a.s.$,

$$\sup_{x\in\Re} |F_n(x) - F(x)| = \sup_{F(x)\in(0,1)} |F_n(x) - F(x)|,\ a.s. \tag{2}$$

Let $p = F(x) \in (0,1)$. $F(x)$ is monotone and continuous, we can define $F^{-1}(p) = \inf\{x : F(x) \le p\} \in \Re$, then

$$1_{(-\infty,x]}(X_i) = 1_{(-\infty,F^{-1}(p)]}(X_i) \quad w.p.\ 1$$

so

$$\sup_{F(x)\in(0,1)} |F_n(x) - F(x)| = \sup_{p\in(0,1)} \left| \frac{1}{n} \sum 1_{(-\infty,F^{-1}(p)]}(X_i) - p \right| \quad a.s. \tag{3}$$

Because $\{1_{(-\infty,F^{-1}(p)]}(X_i)\}_{i=1}^n$ are i.i.d. bernoulli RV's with probability p to be 1, K_n has the same distribution, which doesn't depend on F.

There are other methods of proof including those using uniform representation.

(d) The Glivenko-Cantelli lemma about empirical CDF says

$$\lim_{n\to\infty} \sup_{x\in\Re} |P_n((-\infty,x]) - P((-\infty,x])| = 0 \quad a.s.$$

The Glivenko-Cantelli lemma can not be extended to all Borel-measurable sets. Let $A_n = \{X_1, \cdots, X_n\}$ which is a measurable set with measure 0. Then $P(A_n) = 0$ but $P_n(A_n) = 1$. Thus $\sup_{A\in\mathcal{B}(\Re)} |P_n(A) - P(A)| \ge 1$.

2. Let $(X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n)$ be independently normally distributed random variables, all with the same variance θ but with $EX_i = EY_i = \mu_i$, $i = 1, 2, \ldots, n$, all possibly different. The parameter θ is of major interest for inference and the parameters μ_i $i = 1, 2, \ldots, n$ are considered nuisance parameters here.

   (a) Obtain the maximum likelihood estimator (MLE) for θ.

   (b) Show that the MLE is inconsistent for θ.

   (c) Show how the MLE can be rescaled to give a consistent estimator for θ.

   (d) Show how confidence intervals for θ can be prescribed.

   (e) What is the Fisher information for θ in (X_1, Y_1) when μ_1 is fixed? What property of the Fisher information suggests that the MLE is biased (even asymptotically)?

Solution:

(a)

$$\begin{aligned}
f(X,Y;\mu,\theta) &= (2\pi\theta)^{-n} \exp\left(-\frac{\sum(X_i-\mu_i)^2 + \sum(Y_i-\mu_i)^2}{2\theta} \right) \\
\frac{\partial L(X,Y;\mu,\theta)}{\partial\theta} &= -\frac{n}{\theta} + \frac{\sum(X_i-\mu_i)^2 + \sum(Y_i-\mu_i)^2}{2\theta^2} \\
\frac{\partial L(X,Y;\mu,\theta)}{\partial\mu_i} &= \frac{(X_i-\mu_i) + (Y_i-\mu_i)}{\theta}
\end{aligned}$$

Solve these likelihood equations and it can be verified that the MLE's are

$$\begin{aligned}\hat{\mu}_i &= \frac{X_i + Y_i}{2} \\ \hat{\theta} &= \frac{1}{4n}\sum_{i=1}^{n}(X_i - Y_i)^2\end{aligned}$$

(b) Notice that $X_i - Y_i \sim N(0, 2\theta)$, hence $\mathrm{E}(X_i - Y_i)^2 = 2\theta, \ \forall i$

$$\mathrm{E}\hat{\theta} = \frac{\theta}{2}$$

So the MLE is biased, and by the law of large numbers, it is inconsistent for θ.

(c) $2\hat{\theta}$ gives a consistent estimator for θ.

(d) Because $\{X_i - Y_i\}$ are i.i.d. $N(0, 2\theta)$, then

$$\sum_{i=1}^{n}\frac{(X_i - Y_i)^2}{2\theta} \sim \chi_n^2$$

so the confidence interval can be obtained as

$$\left(\frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{2\chi^2_{n,\alpha/2}}, \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{2\chi^2_{n,1-\alpha/2}}\right)$$

i.e.

$$\left(\frac{\hat{\theta}}{8n\chi^2_{n,\alpha/2}}, \frac{\hat{\theta}}{8n\chi^2_{n,1-\alpha/2}}\right)$$

(e) The Fisher information for θ is

$$\begin{aligned}I(\theta) &= -\mathrm{E}_{\mu,\theta}\frac{\partial^2 L(X_1, Y_1; \mu, \theta)}{\partial\theta^2} \\ &= -\mathrm{E}_{\mu,\theta}\left(\frac{1}{\theta^2} - \frac{(X_1 - \mu_1)^2 + (Y_1 - \mu_1)^2}{\theta^3}\right) \\ &= \frac{1}{\theta^2}.\end{aligned}$$

The Cramér-Rao lower bound for θ is

$$CRLB = \frac{\theta^2}{n}$$

But for the MLE, we have

$$\mathrm{Var}\hat{\theta} = \frac{\theta^2}{2n}$$

The variance of the MLE is smaller than the Cramér-Rao lower bound which suggests the MLE is biased. Note that, the asymptotic variance of the MLE is smaller (even asymptotically) than the Cramér-Rao lower bound, which is possible due to the inconsistency of the MLE.

3. Suppose a biased coin is independently tossed three times. For $i = 1, 2, 3$, let X_i be the number of head observed in the ith toss, and a natural model is $X_i \sim \text{Bernoulli}(\theta)$, $\theta \in \Theta = [0, 1]$. We consider estimating θ^2 based on these i.i.d. observations.

   (a) What is the MLE for θ^2? Is it unbiased? (Explain why)

   (b) Derive the UMVUE for θ^2?

   (c) If there is exactly one head in these three tosses, what is the value of the UMVUE estimator of θ^2? Explain whther you think the UMVUE is a better estimator than the MLE in this case.

   (d) Consider the following hierarchy:

$$X_i|\theta \sim \text{Bernoulli}(\theta),\ i = 1, 2, 3,$$
$$\theta \sim U(0, 1).$$

   Under this hierarchy, find the mean and variance of the random sum $S = X_1 + X_2 + X_3$.

Solution:

3a). $f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $i = 1, 2, 3$.
The MLE of θ is $\hat{\theta} = \sum_{i=1}^{3} x_i/3$
By the invariance of MLE, the MLE of θ^2 is $\hat{\theta}^2 = [\sum_{i=1}^{3} x_i/3]^2$
The MLE of θ is $\hat{\theta} = \sum_{i=1}^{3} x_i/3$ which is unbiased. By Jensen's inequality the MLE of θ^2, $\hat{\theta}^2 = [\sum_{i=1}^{3} x_i/3]^2$, has a positive bias. In fact,

$$E[\sum_{i=1}^{3} x_i/3]^2 = (\theta + 2\theta^2)/3 > \theta^2$$

unless $\theta = 0$, or 1.

3b). Since Bernoulli family is an exponential family, thus $\sum X_i$ is complete and sufficient.

Note that
$E\sum_{i=1}^{3} X_i = 3\theta$
$E[\sum_{i=1}^{3} X_i]^2 = 3\theta + 6\theta^2$
$E(\frac{\sum_{i=1}^{3} X_i)^2 - \sum_{i=1}^{3} X_i}{6}) = \theta^2$ Then
$\frac{1}{6}(\sum_{i=1}^{3} X_i)^2 - \frac{1}{6}\sum_{i=1}^{3} X_i$ is the UMVUE for θ^2.

3c). If there is exactly one head in these three tosses,
$\tilde{\theta}^2_{UMVUE} = \frac{1}{6} - \frac{1}{6} \cdot 1 = 0$
$\hat{\theta}^2_{MLE} = \frac{1}{9}$

UMVUE often has good properties like unbiasedness and smallest variance, but, in this case the value of UMVUE for θ^2 is 0 which doesn't make much sense. Since, if θ^2 is zero, the sum of X_i's should be 0 all the time. Here, the MLE of θ^2 makes more sense even though it's not unbiased in this case.

3d). Under the assumption, we have the following hierarchy:

$$X_i|\theta \sim \text{Bernoulli}(\theta),\ i = 1, 2, 3,$$
$$\theta \sim U(0, 1).$$

Under this hierarchy, the random sum $S = X_1 + X_2 + X_3$ has a beta-binomial distribution.

$$ES = E(E(S|\theta)) = 3E(E(X_1|\theta)) = 3E\theta = 3/2.$$

$$\begin{aligned}\text{Var}(S) &= E(\text{Var}(S|\theta)) + \text{Var}(E(S|\theta)) \\ &= 3E\theta(1-\theta) + \text{Var}(3\theta) \\ &= 3E\theta - 3(E\theta)^2 + 6\text{Var}(\theta) \\ &= 3/4 + 6/12 = 5/4.\end{aligned}$$

4. Suppose $X_1, \ldots, X_n$ are i.i.d. random observations with a double exponential density $f(x|\theta_0)$, where for any θ

$$f(x|\theta) = \frac{1}{2}e^{-|x-\theta|}, \quad \theta \in \Re$$

Note that $f(x|\theta)$ is symmetric about θ and has variance $\sigma^2 = 2$.

(a) Write down the log-likelihood function.

(b) Derive the MLE $\hat{\theta}$ of θ_0 based on the sample $X_1, \ldots, X_n$.

(c) Identify the limiting distribution of the MLE, i.e. $\sqrt{n}(\hat{\theta} - \theta_0)$, and specify an asymptotic confidence interval for θ_0.

(d) Show that the sample mean $\overline{X}$ is unbiased for θ_0 and identify the limiting distribution of $\sqrt{n}(\overline{X} - \theta_0)$. Is $\overline{X}$ asymptotically more efficient than the MLE $\hat{\theta}$?

(e) Suppose the data are actually an i.i.d. sample from a normal distribution with some unknown location paramter θ_0, and suppose the MLE in (b) obtained under the assumption of a double exponential is used for inference for θ_0. Is the MLE $\hat{\theta}$ in (b) a consistent estimator of θ_0? Will the asymptotic confidence interval in part (c) have coverage probability $1 - \alpha$ as $n \to \infty$?

Solution:

(a) The log-likelihood function is

$$L_n(\theta, \mathbf{X}) = -\sum_{i=1}^{n} |X_i - \theta| - n\log(2).$$

(b) Maximizing $L_n(\theta, \mathbf{X})$ over θ is equivalent to minimizing $\sum_{i=1}^{n} |X_i - \theta|$, which yields the sample median $\hat{\theta}$.

(c) Note that the density has a continuous positive density with a unique median at θ_0 and the density value equals to 1/2 at θ_0. The limiting distribution of the sample median (or the MLE here) is normal:

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, (4f[(\theta_0|\theta_0)]^2)^{-1}) = N(0, 1). \qquad (\#)$$

Based on $(\#)$, a $1 - \alpha$ asymptotic confidence interval for θ_0 can be constrcted as follows:

$$[\hat{\theta} - z_{\alpha/2}/\sqrt{n},\ \hat{\theta} + z_{\alpha/2}/\sqrt{n}]$$

where $P(Z > z_{\alpha/2}) = \alpha/2$ and Z denotes the standard normal random variable.

(d) Note that $f(x|\theta_0)$ is symmetric about θ_0 and has variance $\sigma^2 = 2$. A distribution with a symmetric density has a mean (if it exists) equal to its median. Thus $EX = \theta_0$ and $\text{Var}(X) = 2$. Thus the sample mean $\overline{X}$ is a moment estimator. By the classical CLT,

$$\sqrt{n}(\overline{X} - \theta_0) \to_d N(0, 2).$$

Thus, compared with equation $(\#)$, the sample mean $\overline{X}$ is asymptotically less efficient than the MLE $\hat{\theta}$ which is the sample median.

(e) When the true density is normal with location parameter θ_0, i.e. $N(\theta_0, 1)$, the MLE obtained in (b) which is the sample median is still consistent as a point estimator. This is because the sample median converges to the true median of the distribution from which the data are sampled. However, the asymptotic confidence interval contructed in c is invalid because it does not have the desired coverage probability $1 - \alpha$. In this case, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, \pi/2).$$

This is different from (#) in part (c).

1. Suppose (U, V) are continuous random variables with joint probability density function:

$$f_{U,V}(u,v) = c(\theta u^2 + v^2),\ 0 < u < 1,\ 0 < v < 1,\ c > 0,\ \theta > 0.$$

   (a) Find c as a function of θ.

   (b) Prove that U and V are independent, or prove that U and V are not independent.

   (c) Derive the cumulative distribution function and probability density function for $W = V^2$.

2. Let X be a continuous random variable with cumulative distribution function:

$$F_X(x) = \tfrac{x}{\theta + x},\ x > 0,\ \theta > 0.$$

   and let X_i, $i = 1, 2, 3, ..., n$ be a random sample of size n from this distribution.

   (a) Derive an equation for which the solution is the maximum likelihood estimate of θ, $\hat{\theta}$.

   (b) Derive the asymptotic variance of $\hat{\theta}$.

   (c) Let

$$p_\Theta(\theta) = \begin{cases} \frac{1}{2} & \theta = 1, 2 \\ 0 & \text{other} \end{cases}$$

   be a prior probability mass function on the two values of θ indicated. Suppose a sample of size $n = 1$ is drawn. With respect to this prior, what is the posterior probability mass function for θ (in terms of $x = x_1$)?

3. Let:

   S_i, $i = 1, 2, 3, ..., n$ each have a uniform continuous distribution on the interval $[-\theta_1, \theta_1]$, $\theta_1 > 0$
   T_i, $i = 1, 2, 3, ..., n$ each have a uniform continuous distribution on the interval $[\theta_2, 2\theta_2]$, $\theta_2 > 0$

   such that all $2n$ random variables are independent.

   (a) Find the large sample (i.e. large n) distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n S_i$.

   (b) Find the large sample distribution of $\sqrt{n} \frac{\sum_{i=1}^n S_i}{\sum_{i=1}^n T_i}$.

   (c) Find the large sample distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{S_i}{T_i}$.

   (d) Find the (fixed n) maximum likelihood estimates of θ_1 and θ_2 ($\hat{\theta}_1$ and $\hat{\theta}_2$, respectively).

   (e) Derive the distribution of $\hat{\theta}_1$.

   (f) Derive the likelihood ratio statistic for testing $H_0 : \theta_1 = \theta_2$ versus $H_A : \theta_1 \neq \theta_2$.

1.a) $\int_0^1 \int_0^1 c(\theta u^2 + v^2)\, du\, dv = 1$

$= c\int_0^1 [\frac{1}{3}\theta u^3 + v^2 u]_0^1\, dv = c\int_0^1 \frac{1}{3}\theta + v^2\, dv$

$= c[\frac{1}{3}\theta v + \frac{1}{3}v^3]_0^1 = \frac{1}{3}c(\theta+1) \Rightarrow c = \frac{3}{\theta+1}$

1.b) $f_V(v) = \int_0^1 f(u,v)\, du = c(\frac{1}{3}\theta + v^2)$

$f_U(u|v) = \frac{c(\theta u^2 + v^2)}{c(\frac{1}{3}\theta + v^2)} = \frac{u^2\theta + v^2}{\frac{1}{3}\theta + v^2}$,

a function of $v \Rightarrow$ not independent

1.c) $W = V^2$, $V = \sqrt{W}$, $0 < W < 1$

$F_V(v') = c\int_0^{v'} \frac{1}{3}\theta + v^2\, dv = \frac{1}{3}c(\theta v + v^3)$

$\text{Prob}\{W \leq w'\} = \text{Prob}\{V \leq \sqrt{w'}\}$

$= \frac{1}{3}c(\theta\sqrt{w'} + w'^{3/2}) = F_W(w')$

2.a) $f_X(x) = \dfrac{\theta}{(\theta+x)^2}$

$$l = \theta^n \prod_{i=1}^{n} (\theta + x_i)^{-2}$$

$$\ln l = n \ln \theta - 2 \sum_{i=1}^{n} \ln(\theta + x_i)$$

$$\frac{\partial}{\partial \theta} \ln l = n\,\theta^{-1} - 2\sum_{i=1}^{n} (\theta + x_i)^{-1} \Rightarrow \sum_{i=1}^{n} \frac{\hat\theta - x_i}{\hat\theta(\hat\theta + x_i)} = 0$$

2b) $$\frac{\partial^2}{\partial \theta^2} \ln l = -n\,\theta^{-2} + 2\sum_{i=1}^{n} (\theta + x_i)^{-2}$$

$$E\left(\frac{\partial^2}{\partial \theta^2} \ln l\right) = -n\,\theta^{-2} + 2n\, \underbrace{E\left\{\frac{1}{(\theta+X)^2}\right\}}_{\int_0^\infty \frac{\theta}{(\theta+x)^4}\,dx}$$

$$= \theta \left[\frac{-\frac{1}{3}}{(\theta+x)^3}\right]_0^\infty = \frac{1}{3}\theta^{-2}$$

$$= -n\,\theta^{-2} + \frac{2}{3} n\,\theta^{-2} = -\frac{1}{3}\frac{n}{\theta^2}$$

expected information $= -E(\sim) = \frac{1}{3}\frac{n}{\theta^2}$

asymptotic $\mathrm{Var}(\hat\theta) = \left[\frac{1}{3}\frac{n}{\theta^2}\right]^{-1} = 3\,\frac{\theta^2}{n}$

2.c) $p_\Theta(\theta|x) \propto p_\Theta(\theta)\, f(x|\theta)$

$$= \frac{1}{2}\frac{\theta}{(\theta+x)^2} \qquad \theta = 1, 2$$

$$= 0 \qquad \text{other}$$

So,

$$\text{Prob}\{\Theta = 1 | x\} = \frac{\frac{1}{2}\frac{1}{(1+x)^2}}{\frac{1}{2}\frac{1}{(1+x)^2} + \frac{1}{2}\frac{2}{(2+x)^2}} = \frac{1}{1 + 2\frac{(1+x)^2}{(2+x)^2}}$$

$$= p_\Theta(1|x)$$

$$\text{Prob}\{\Theta = 2 | x\} = \frac{\frac{1}{2}\frac{2}{(2+x)^2}}{\frac{1}{2}\frac{1}{(1+x)^2} + \frac{1}{2}\frac{2}{(2+x)^2}} = \frac{2}{2 + \frac{(2+x)^2}{(1+x)^2}}$$

$$= p_\Theta(2|x)$$

$p_\Theta(\theta|x) = 0, \quad \theta \neq 1 \text{ or } 2$

3.a) $E(S) = 0 \qquad Var(S) = \frac{1}{3}\theta_1^2$

$E(T) = \frac{3}{2}\theta_2 \qquad Var(T) = \frac{1}{12}\theta_2^2$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i = \underbrace{\frac{\frac{1}{n}\sum S_i - 0}{\sqrt{\frac{1}{3}\theta_1^2}/\sqrt{n}}}_{\underset{CLT}{\Rightarrow} N(0,1)} \sqrt{\frac{1}{3}\theta_1^2} \Rightarrow N(0, \tfrac{1}{3}\theta_1^2)$$

3.b) $$\sqrt{n}\,\frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} T_i} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i \xrightarrow{CLT} N(0, \frac{1}{3}\theta_1^2)}{\frac{1}{n}\sum_{i=1}^{n} T_i \xrightarrow{LLN} \frac{3}{2}\theta_2} \Longrightarrow N\left(0, \frac{4}{27}\frac{\theta_1^2}{\theta_2^2}\right)$$

3.c) $\sqrt{n}\sum_{i=1}^{n}\frac{S_i}{T_i} \qquad E\left(\frac{S}{T}\right) = 0$

$$Var\left(\frac{S}{T}\right) = \int_{\theta_2}^{2\theta_2} \frac{1}{t^2}\,\frac{1}{2\theta_1\theta_2}\Big[\underbrace{\int_{-\theta_1}^{\theta_1} u^2\,du}_{\frac{2}{3}\theta_1^3}\Big]\,dt$$

$$= \frac{1}{3}\frac{\theta_1^2}{\theta_2}\int_{\theta_2}^{2\theta_2}\frac{1}{t^2}\,dt$$

$$= \frac{1}{3}\frac{\theta_1^2}{\theta_2}\underbrace{\Big[\frac{-1}{t}\Big]_{\theta_2}^{2\theta_2}}_{\left[\frac{-1}{2\theta_2} + \frac{1}{\theta_2}\right]}$$

$$= \frac{1}{6}\frac{\theta_1^2}{\theta_2^2}$$

$$\underset{CLT}{\Rightarrow} N\left(0, \frac{1}{6}\frac{\theta_1^2}{\theta_2^2}\right)$$

3.d) $f_S(s) = \frac{1}{2\theta_1} \quad -\theta_1 \le s \le \theta_1$

$= 0 \quad$ other

so $l(\theta_1) = \left(\frac{1}{2\theta_1}\right)^{\#(|s_i| \le \theta_1)} (0)^{\#(|s_i| > \theta_1)}$

$\max_{\theta_1} l(\theta_1) = \max_{\theta_1 \ge \text{all } |s_i|} \left(\frac{1}{2\theta_1}\right)^n \Rightarrow \underline{\hat{\theta}_1 = \max |s_i|}$

$f_T(t) = \frac{1}{\theta_2} \quad \theta_2 \le t \le 2\theta_2$

$= 0 \quad$ other

so $l(\theta_2) = \left(\frac{1}{\theta_2}\right)^{\#(\theta_2 \le t_i \le 2\theta_2)} (0)^{\#(t_i < \theta_2 \text{ or } t_i > 2\theta_2)}$

$\max_{\theta_2} l(\theta_2) = \max_{\theta_2 \le \text{all } t_i \le 2\theta_2} \left(\frac{1}{\theta_2}\right)^n$

$l(\theta_2)\Big|_{\theta_2 = \frac{1}{2}\max_i t_i} = \left(\frac{1}{\frac{1}{2}\max_i t_i}\right)^n$

$l(\theta_2)\Big|_{\theta_2 < \frac{1}{2}\max_i t_i} = 0$

$l(\theta_2)\Big|_{\theta_2 > \frac{1}{2}\max_i t_i} < l(\theta_2)\Big|_{\theta_2 = \frac{1}{2}\max_i t_i}$

so $\underline{\hat{\theta}_2 = \frac{1}{2}\max t_i}$

3.e)

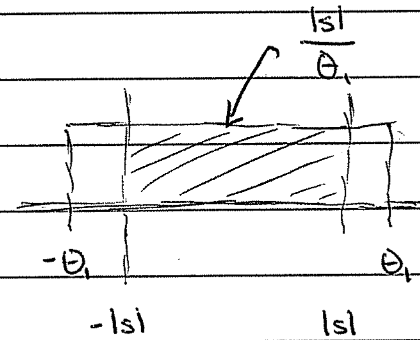
$$F_{\hat{\theta}_1}(\theta) = \text{Prob}\{\hat{\theta}_1 \le \theta\}$$

$$= \text{Prob}\{\max |s_i| \le \theta\}$$

$$= \text{Prob}\{\text{all } |s_i| \le \theta\}$$

$$= \text{Prob}\{|s| \le \theta\}^n$$

$$= \begin{cases} \left(\frac{\theta}{\theta_1}\right)^n & 0 \le \theta \le \theta_1 \\ 1 & \theta > \theta_1 \end{cases}$$



3F) From part (d),

Under H_A

$$\hat{\theta}_1 = \max |s_i|$$

$$\hat{\theta}_2 = \tfrac{1}{2} \max t_i$$

and the maximized likelihood is

$$L(\hat{\theta}_1, \hat{\theta}_2) = L(\hat{\theta}_1) L(\hat{\theta}_2) = \left(\frac{1}{2\hat{\theta}_1}\right)^n \left(\frac{1}{\hat{\theta}_2}\right)^n$$

Under H_0,

$$L(\theta) = \left(\frac{1}{2\theta}\right)^n \left(\frac{1}{\theta}\right)^n \quad \text{all } |s_i| \le \theta \text{ and } \theta \le \text{all } t_i \le 2\theta$$

$$= 0 \quad \text{other}$$

$$\Rightarrow \hat{\theta} = \max \{\max |s_i|, \tfrac{1}{2} \max t_i\}$$

and the maximized likelihood is

$$L(\hat{\theta}) = \left(\tfrac{1}{2}\right)^n \left(\frac{1}{\hat{\theta}}\right)^{2n}$$

The likelihood ratio is

$$\frac{\max L \text{ under } H_0}{\max L \text{ under } H_A} = \frac{\hat{\theta}^{-2n}}{\hat{\theta}_1^{-n} \hat{\theta}_2^{-n}} = \frac{\hat{\theta}_1^n \hat{\theta}_2^n}{\hat{\theta}^{2n}}$$

$$= \frac{\max |s_i|^n \left(\tfrac{1}{2} \max t_i\right)^n}{\max \{\max |s_i|, \tfrac{1}{2} \max t_i\}^{2n}}$$

1. Consider the following "randomized response" survey scheme:

   A pollster asks a potentially embarrassing question to a randomly chosen respondent. The respondent is told to flip a fair coin before answering; the respondent is instructed not to reveal the outcome of the toss to the pollster. The pollster then asks the question. If the truthful answer is "yes," the respondent is instructed to answer truthfully irrespective of the toss outcome. If the truthful answer is "no," the respondent is instructed to answer "yes" if "heads" appears and "no" otherwise.

   Note: This procedure essentially "anonymizes" response - the pollster cannot determine with certainty whether the respondent is being truthful or not because neither the truth nor toss outcome is known to the pollster.

   For the sake of concreteness, suppose the population is "Iowa high school students" and that the question is "Have you used drugs in the last 6 months?" Assume that the true proportion of Iowa high school students that actually have used drugs in the last 6 months is p. Suppose that this survey is to be administered to n randomly selected Iowa high school students, and that sampling is done with replacement. Finally, assume the students selected follow the instructions exactly.

   (a) Let Y be the number of "yes" responses obtained from administering this randomized response survey to n randomly selected students. Explain why

   $$Y \sim Binomial(n, \frac{1}{2}(1+p)).$$

   (b) Let $g(Y) = cY + d$ for constants c and d that do not depend on p. Find c, d such that $E[g(Y)] = p$.

   (c) Let $g(Y) = cY + d$ be defined as in part (b). How large must n be in order to ensure - irrespective of the true value of p - that the standard deviation of $g(Y)$ is less than 0.01? *(Note: if you were unable to do (b), leave your answer in terms of c and d.)*

2. Let the cumulative distribution function (cdf) of an absoluely continuous random variable X, call it $F_X(x)$, have the form

$$1 + \frac{\beta}{x^2},$$

for non-zero x in some interval of the form (c, ∞). Let $f_X(\cdot)$ denote the corresponding probability density funtion (pdf). (Neither β nor c have yet been specified.) Note that $f_X(x) > 0$ on the interval (c, ∞).

   (a) Show that of necessity, $\beta < 0$ and $c = -\sqrt{\beta}$.

   (b) Suppose $\beta = -1$, and let $Y = logX$. Find the moment generating function of Y. Use it to compute $E[Y]$.

   (c) Let $W = (Y - 1)^2$, where Y is as in part (b). Find the pdf of W.

   (d) Let Y be defined as in part (b), and set $Z = \lfloor Y \rfloor$; that is, Z is the greatest integer less than or equal to Y. (For example, $\lfloor 3.2 \rfloor = 3$.) Find the probability mass function of Z.

3. Let $Y_1, ..., Y_n$ be i.i.d. $N(0, 1)$ random variables, and $R_n^2 = \sum_{i=1}^n Y_i^2$,

   (a) R_n^2/n approaches a fixed limit L. What is L and what is the mode of convergence?

   (b) Show that $\sqrt{n}((R_n^2/n) - L)$ is asymptotically normal. Also identify the asymptotic mean and variance of R_n^2.

   (c) Show that $\sqrt{n}((R_n/\sqrt{n}) - \sqrt{L})$ is asymptotically normal. Also identify the asymptotic mean and variance of R_n.

Solution

Co-Major PhD, July 2009 (Theory I, 542)

1. (a) For a randomly selected student, define

D = the student has used drug in last 6 months.

$P(D) = p$ and $P(D^c) = (1-p)$ because the sample is selected with replacement

So $P(\text{answer yes}) = P(\text{answer yes} \mid D)\,P(D) + P(\text{answer yes} \mid D^c)\,P(D^c)$

$= 1 \times p + \frac{1}{2} \times (1-p) = \frac{1}{2}(1+p)$.

And the sample is selected randomly with size n,

so $Y \sim \text{Binomial}\left(n, \frac{1}{2}(1+p)\right)$.

(b) Since $Y \sim \text{Binomial}\left(n, \frac{1}{2}(1+p)\right)$.

$E[Y] = \frac{n}{2}(1+p)$

$E[cY+d] = c\frac{n}{2}(1+p) + d = \frac{cn}{2} + d + \frac{cn}{2}p$

We want $\frac{cn}{2} + d + \frac{cn}{2}p = p$

$$\Rightarrow \begin{cases} \frac{cn}{2} + d = 0 \\ \frac{cn}{2} = 1 \end{cases} \Longrightarrow \begin{cases} c = \frac{2}{n} \\ d = -1 \end{cases}$$

$g(Y) = \frac{2}{n}Y - 1$ has expectation $= p$.

(c) $Var[Y] = n\left(\frac{1}{2}(1+p)\right)\left(1-\frac{1}{2}(1+p)\right)$

$= \frac{n}{4}(1+p)(1-p)$

$Var[g(Y)] = c^2\, Var[Y] = \frac{c^2 n}{4}(1+p)(1-p)$

$= \frac{1-p^2}{n}$

$Std[g(Y)] = \sqrt{\frac{1-\rho^2}{n}}$

Because $1-\rho^2 \leq 1$, so $\sqrt{\frac{1-\rho^2}{n}} \leq \sqrt{\frac{1}{n}}$.

If we let $\sqrt{\frac{1}{n}} \leq 0.01$, then $Std[g(Y)] \leq 0.01$.

(i.e. $n \geq 10^4$).

#

2. (a) $F(x) = 1 + \frac{\beta}{x^2}$

We know $0 \leq F(x) \leq 1$, So $-1 \leq \frac{\beta}{x^2} \leq 0$.

$\frac{\beta}{x^2} \leq 0$ gives us that $\beta \leq 0$.

If $\beta = 0$, $F(x) = 1$ which can't be a legitimate CDF. So β must be < 0.

Also $-1 \leq \frac{\beta}{x^2}$ gives us that $-\beta \leq x^2$.

So x can either be $x \geq +\sqrt{-\beta}$ or $x \leq -\sqrt{-\beta}$.

If $x \leq -\sqrt{-\beta}$ (i.e. $-\sqrt{-\beta}$ is the upper limit of the support S_x), because $F(x)$ is continuous, we have $\lim_{x \to -\sqrt{-\beta}} F(x) = 1 + \frac{\beta}{-\beta} = 0 \neq 1$.

So x must be $x \geq \sqrt{-\beta}$.

Next we need to show that ∞ is the correct upper limit of S_x. If the upper limit of S_x is $c < \infty$, because $\beta < 0$ and $F(x)$ is continuous

$\lim_{x \to c} F_X(x) = 1 + \frac{\beta}{c^2} < 1$.

So the upper limie must be ∞.

In summary, $S_x = \{x : x \in [\sqrt{-\beta}, \infty)\}$ or

$S_x = \{x : x \in (\sqrt{-\beta}, \infty)\}$.

(b) $F_X(X) = 1 - \frac{1}{X^2}$, $X \geq 1$.

So the pdf of X is $f_X(x) = 2X^{-3}$, $X \geq 1$

The moment generating function of Y is

$M_Y(t) = E[e^{tY}] = E[X^t]$

$= \int_1^{\infty} X^t \, 2X^{-3} \, dX$

when $t \cancel{<} 2$, $M_Y(t)$ exists and $M_Y(t) = \frac{2}{2-t}$.

$E[Y] = M_Y'(0) = 2(2-t)^{-2}\Big|_{t=0} = \frac{1}{2}$.

(c) First, try to find the CDF of W.

$Pr(W \leq w) = P((Y-1)^2 \leq w) = P(1-\sqrt{w} \leq Y \leq 1+\sqrt{w})$

$= P(e^{1-\sqrt{w}} \leq X \leq e^{1+\sqrt{w}})$

Because X has the support region $X \geq 1$, we need to discuss this probability in two cases.

If $e^{1-\sqrt{w}} < 1$, (i.e. $w > 1$)

$P(W \leq w) = F_X(e^{1+\sqrt{w}}) - F_X(1) = 1 - e^{-2(1+\sqrt{w})}$

If $e^{1+\sqrt{w}} \geq 1$, (i.e. $w \leq 1$)

$P(W \leq w) = F_X(e^{1+\sqrt{w}}) - F_X(e^{1-\sqrt{w}})$

$= e^{-2(1-\sqrt{w})} - e^{-2(1+\sqrt{w})}$.

Therefore, the pdf of W is

$$f_W(w) = \begin{cases} e^{-2(1+\sqrt{w})} w^{-\frac{1}{2}} & w > 1 \\ e^{-2} w^{-\frac{1}{2}} (e^{2\sqrt{w}} + e^{-2\sqrt{w}}) & 0 \leq w \leq 1. \end{cases}$$

(d) If Z is a negative integer,
$P(Z=z) = 0$ because $X \geq 1$.

If z is a nonnegative integer,
$P(Z=z) = P(z \leq \log X < z+1)$
$= P(e^{z} \leq X < e^{z+1})$
$= 1 - e^{-2(z+1)} - (1-e^{-2z}) = e^{-2z}(1-e^{-2})$ #

3. (a) By the Law of Large number,
$R_n^2/n \xrightarrow{P} E[Y_i^2] = E[X_i^2] = 1 \triangleq L$.

(b) By the Central Limit Theorem,
because $E[Y_i^2]=1$ and $Var[Y_i^2] = 2$

$$\frac{\sqrt{n}\left(\frac{R_n^2}{n} - 1\right)}{\sqrt{2}} \xrightarrow{D} N(0,1), \text{ as } n \to \infty$$

So this implies that $R_n^2 \sim AN(n, 2n)$.
The asymptotic mean is n and the asymptotic Variance is $2n$.

(c) By the Delta method, let $f(x) = \sqrt{x}$, $f'(1) = \frac{1}{2}$

$$\frac{\sqrt{n}\left(\frac{R_n}{\sqrt{n}} - 1\right)}{\sqrt{2}} \xrightarrow{D} N(0, \tfrac{1}{4}), \text{ as } n \to \infty$$

So this implies that $R_n \sim AN(\sqrt{n}, \frac{1}{2})$.
The asymptotic mean is $\sqrt{n}$ and the asymptotic Variance is $\frac{1}{2}$.

#

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with the following probability density function (p.d.f.)

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1}, & \text{for } 0 < x < 1; \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$.

(a) Show that the random variable $Y = \sum_{i=1}^{n} \log(1/X_i)$ has a Gamma (n, θ) distribution with the p.d.f.

$$f(y|\theta) = \frac{1}{(n-1)!}\theta^n y^{n-1} \exp(-\theta y) \text{ for } y > 0.$$

(b) Find a uniformly minimum variance unbiased estimator (UMVUE) of $1/\theta$.

(c) Find a UMVUE of θ^r, where $r < n$ is a given real number.

(d) Show that the maximum likelihood estimator of θ based on $X_1, \ldots, X_n$ is

$$\widehat{\theta}_n = \frac{n}{\sum_{i=1}^{n} \log(1/X_i)}.$$

(e) Find the mean of $\widehat{\theta}_n$. Is $\widehat{\theta}_n$ unbiased for θ? Explain.

(f) Find the limiting distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ as $n \to \infty$.

(g) Find the limiting distribution of $\sqrt{n}(\delta_n - \theta)$ as $n \to \infty$, where $\delta_n = \overline{X}_n/(1 - \overline{X}_n)$ and $\overline{X}_n = \sum_{i=1}^{n} X_i/n$.

(h) Give the asymptotic relative efficiency of the estimator δ_n in part (g) with respect to the maximum likelihood estimator $\widehat{\theta}_n$.

(i) Determine the uniformly most powerful (UMP) level-α test of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 > 0$ is a given real number.

(j) Find the Bayes estimator of θ under squared error loss if the prior density of θ is $f(\theta) = \exp(-\theta)$ for $\theta > 0$.

(a) Let $Y_i = \log(1/X_i)$ for $i = 1, \ldots, n$. Then $Y = \sum_{i=1}^n Y_i$. Since $X_1 = \exp(-Y_1)$, the p.d.f. of Y_1 is given by

$$f(y_1|\theta) = \theta\{\exp(-y_1)\}^{(\theta-1)} \cdot \exp(-y_1) = \theta\exp(-\theta y_1), \quad \text{for } y_1 > 0;$$

that is, Y_1 has an exponential (θ) distribution. Thus, Y has a Gamma (n, θ) distribution with the given p.d.f.

(b) The joint p.d.f. of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) = \exp\left\{n\log\theta + (\theta - 1)\sum_{i=1}^n \log x_i\right\}\prod_{i=1}^n I_{(0,1)}(x_i),$$

where $I_{(0;1)}(x) = 1$ for $0 < x < 1$, and $I_{(0,1)}(x) = 0$ otherwise. Thus, $f(x_1, \ldots, x_n|\theta)$ forms a one-parameter exponential family and $Y = -\sum_{i=1}^n \log X_i$ is a complete and sufficient statistic for θ. From part (a), we have $E(Y) = n/\theta$. Thus, Y/n is a UMVUE of $1/\theta$.

(c) Note that θY has a Gamma $(n, 1)$ distribution. Thus, $E(\theta Y)^{-r} = \Gamma(n-r)/(n-1)!$ for $r < n$, where

$$\Gamma(n-r) = \int_0^\infty t^{n-r-1}\exp(-t)dt.$$

Thus, $Y^{-r}(n-1)!/\Gamma(n-r)$ is a UMVUE of θ^r for $r < n$.

(d) From part (b), we have

$$l'(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^n \log f(X_i|\theta) = \frac{n}{\theta} + \sum_{i=1}^n \log X_i.$$

Note that $l'(\theta) = 0$ has a unique solution at $\widehat{\theta}_n$ and $l''(\theta) < 0$ for $\theta > 0$. Thus, $\widehat{\theta}_n$ is the unique maximum likelihood estimator of θ.

(e) Note that $\widehat{\theta}_n = n/Y$. Then from part (c), we have

$$E(\widehat{\theta}_n) = nE(Y^{-1}) = \frac{n(n-2)!}{(n-1)!}\theta = \frac{n}{n-1}\theta, \quad \text{for } n > 1.$$

For $n = 1$, $E(\widehat{\theta}_n) = +\infty$. Thus, $\widehat{\theta}_n$ is a biased estimator of θ.

(f) The Fisher information number based on X_1 is

$$I_1(\theta) = \text{Var}\left(\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right) = \text{Var}\left(1/\theta + \log X_1\right) = \theta^{-2}.$$

Thus, the limiting distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ as $n \to \infty$ is $N(0, \theta^2)$.

(g) Note that $E(X_1) = \theta/(1+\theta)$ and $E(X_1^2) = \theta/(2+\theta)$. By the Central Limit Theorem, we have $\sqrt{n}(\overline{X}_n - E(X_1)) \to N(0, \text{Var}(X_1))$ as $n \to \infty$.

Let $g(x) = x/(1-x)$ for $0 < x < 1$. Then $g'(x) = (1-x)^{-2}$. Note that $g(\overline{X}_n) = \delta_n$ and $g(E(X_1)) = \theta$. Thus, $\sqrt{n}(\delta_n - \theta) \to N(0, \sigma^2)$ as $n \to \infty$, where

$$\sigma^2 = [g'(E(X_1))]^2 \text{Var}(X_1) = \frac{\theta(1+\theta)^2}{2+\theta}.$$

(h) The asymptotic relative efficiency of δ_n with respect to $\widehat{\theta}_n$ is

$$\frac{\theta^2}{\sigma^2} = \frac{\theta^2 + 2\theta}{(1+\theta)^2}.$$

(i) The UMP level-α test rejects H_0 if $\sum_{i=1}^n \log X_i \geq c(n, \alpha)$, where $c(n, \alpha)$ is the $(1-\alpha)$th quantile of $\sum_{i=1}^n \log X_i$ when $\theta = \theta_0$. From part (a), $c(n, \alpha) = -q(n, \alpha)$, where $q(n, \alpha)$ is the αth quantile of the Gamma (n, θ_0) distribution.

(j) The Bayes estimator of θ is given by

$$E(\theta|X_1, \ldots, X_n) = \frac{\int_0^\infty \theta f(X_1, \ldots, X_n|\theta) f(\theta) d\theta}{\int_0^\infty f(X_1, \ldots, X_n|\theta) f(\theta) d\theta}.$$

Note that, for $k = 0, 1$,

$$\int_0^\infty \theta^k f(X_1, \ldots, X_n|\theta) f(\theta) d\theta = \int_0^\infty \theta^{n+k} \left(\prod_{i=1}^n X_i\right)^{-1} \exp\left(-\theta\left(1 - \sum_{i=1}^n \log X_i\right)\right) d\theta$$

$$= \left(\prod_{i=1}^n X_i\right)^{-1} \left(1 - \sum_{i=1}^n \log X_i\right)^{-(n+k+1)} (n+k)!.$$

Thus, the Bayes estimator of θ is $(n+1)/(1 - \sum_{i=1}^n \log X_i)$.