PhD Prelim Exam
THEORY

(Majors & Co-Majors)

Summer 2013
(Given on 7/18/13)

PhD Prelim Exam Instructions

1. Put your **Student ID Number (last four digits)** at the top of EACH page.
2. Put the **Question Title** at the top of the page (e.g., Methods II).
3. Begin each **core question** on a new piece of paper.
4. Number pages consecutively beginning at 1 for the start of each new core question. (e.g. Theory I, p. 1; Methods II, p. 1)
5. At the end of the exam period, collect all pages and questions and place in the envelope provided.

Please **DO NOT WRITE** on the back side of the paper.

Part I

1. For a sequence of paired random variables (X_n, Y_n) defined on a probability space $(\Omega, \mathcal{F}, P)$, suppose X_n, Y_n are independent for each $n \geq 1$. If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ as $n \to \infty$, show that random variables X and Y are independent.

Part II

(**Questions 2-3**). Suppose X_n, $n \geq 1$, is a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ along with a random variable X_0. Let $A_k \in \mathcal{F}$, $k \geq 1$, be events such that, for each $k \geq 1$, $P(A_k) = 1$ and $A_k = \{\omega \in \Omega : |X_n(\omega) - X_0(\omega)| < \frac{1}{k} \text{ eventually}\}$.

2. For $k = 2$, express A_2 in terms of unions and/or intersections of events $B_m = \{\omega \in \Omega : |X_m(\omega) - X_0(\omega)| < \frac{1}{2}\}$, $m \geq 1$.

3. Show that $X_n \to X_0$ almost surely (P) as $n \to \infty$.

Part III

(**Questions 4-10**). Let $F : \mathbb{R} \cup \{-\infty\} \to \mathbb{R}$ be defined by $F(x) = 1 - e^{-x}$ if $x > 0$ and $F(x) = 0$ otherwise. Let $\mathcal{C} = \{(a, b] : -\infty \leq a, b < \infty\} \cup \{(b, \infty) : -\infty \leq b < \infty\}$ be a semialgebra on $\mathbb{R}$ with a measure μ defined by

$$\mu((a, b]) = F(b) - F(a) \quad \text{for } (a, b] \in \mathcal{C},\ -\infty \leq a, b < \infty,$$
$$\mu((b, \infty)) = 1 - F(b) \quad \text{for } (b, \infty) \in \mathcal{C},\ -\infty \leq b < \infty.$$

Questions 4-10 concern the construction of a unique extension of μ on $\mathcal{B}(\mathbb{R})$ (the Borel σ-algebra on $\mathbb{R}$).

4. Using sets $\{C_i \in \mathcal{C} : i \geq 1\}$ and μ, give a definition of an outer-measure μ^* on subsets of $\mathbb{R}$.

5. Provide a definition of a measurable set with respect to μ^*. Let $\mathcal{M}$ denote the collection of such sets.

6. Briefly explain the relationship between $(\mathbb{R}, \mathcal{M})$ and μ^*.

7. Briefly explain the relationship between $\mathcal{M}$ and $\mathcal{B}(\mathbb{R})$.

8. Briefly explain the relationship between μ^* and μ.

9. Explain why the extension of μ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is unique.

10. Give the Radon-Nikodym derivative of μ with respect to the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Part IV

(**Questions 11-12**). On a probability space $(\Omega, \mathcal{F}, P)$, suppose random variables $\{X_n : n \geq 1\}$ are uniformly integrable, i.e.,

$$\lim_{t\to\infty} \sup_{n\geq 1} \int_{|X_n|>t} |X_n| dP = 0$$

11. Show that $\sup_{n\geq 1} \mathrm{E}|X_n| < \infty$.

12. Show that, for any $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ (depending on ϵ) such that

$$\sup_{n\geq 1} \int_A |X_n| dP < \epsilon$$

for any $A \in \mathcal{F}$ with $P(A) < \delta_\epsilon$.

(**Question 13**). Suppose $\{X_n : n \geq 1\}$ are integrable random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that (i) $\sup_{n\geq 1} \mathrm{E}|X_n| < \infty$ and (ii) for any given $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that $\sup_{n\geq 1} \int_A |X_n| dP < \epsilon$ for any $A \in \mathcal{F}$ with $P(A) < \delta_\epsilon$.

13. Show that $\{X_n : n \geq 1\}$ is uniformly integrable.

(**Questions 14-16**). Suppose $\{X_n : n \geq 1\}$ are iid random variables on a probability space $(\Omega, \mathcal{F}, P)$ with $\mathrm{E}[|X_1|^p] < \infty$ for some real-valued $p \geq 1$.

(*Note:* In answering Questions 14-16, you may use results provided in Questions 11-13 above.)

14. Show that $\lim_{t\to\infty} \int_{|X_1|>t} |X_1|^p dP = 0$.

15. For $Y_i = |X_i|^p$, $i \geq 1$, show that $\{\sum_{i=1}^n Y_i/n : n \geq 1\}$ is uniformly integrable.

16. For $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, show that $\mathrm{E}\big[|S_n/n|^p\big] \to |\mathrm{E}X_1|^p$ as $n \to \infty$.

Part V

(**Questions 17-21**). Suppose, for $|\rho| < 1$ and a sequence $\{\varepsilon_i : \text{integer } i\}$ of iid random variables with mean $\mathrm{E}[\varepsilon_i] = 0$ and variance $\mathrm{E}[\varepsilon_i^2] = \sigma^2 \in (0, \infty)$, a process is defined as: $X_t = \rho X_{t-1} + \epsilon_t$, $t \geq 1$, where $X_0 = \sum_{i=0}^{\infty} \rho^i \varepsilon_{-i}$

17. Fix $t \geq 0$. Show that $S_{t,n} = \sum_{i=0}^{n} \rho^i \varepsilon_{t-i}$ converges with probability 1 as $n \to \infty$ (to a random variable $\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$).

18. For a given $t \geq 0$, show that $X_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$.

(**Additionally, for Questions 19-21**). Let $M_n = \sum_{i=1}^{n} X_i^2/n$ for $n \geq 1$. For a fixed integer $\ell > 1$, define $Z_{t,\ell} = \sum_{j=0}^{\ell} \rho^j \varepsilon_{t-j}$, $t \geq 1$, and let $M_{n,\ell} = \sum_{i=1}^{n} Z_{i,\ell}^2/n$, $n \geq 1$. You may assume that

$$M_{n,\ell} = \sum_{i=1}^{n} Z_{i,\ell}^2/n \to \mathrm{E}[Z_{1,\ell}^2] \quad \text{with probability 1 as } n \to \infty$$

for any *fixed* integer $\ell > 1$.

19. Show that

$$\lim_{\ell\to\infty} \sup_{n\geq 1} \mathrm{E}|M_n - M_{n,\ell}| \leq \lim_{\ell\to\infty} \left[\mathrm{E}\left(|X_1| \cdot |Z_{1,\ell} - X_1|\right) + \mathrm{E}\left(|Z_{1,\ell}| \cdot |Z_{1,\ell} - X_1|\right)\right] = 0.$$

(*Hint:* Consider adding/subtracting $\sum_{i=1}^{n} X_i Z_{i,\ell}/n$ and apply the Cauchy-Schwarz inequality.)

20. For

$$\begin{aligned} \Delta_{n,\ell}^{(1)} &= |M_{n,\ell} - \mathrm{E}[Z_{1,\ell}^2]|, \\ \Delta_{n,\ell}^{(2)} &= |M_n - M_{n,\ell}|, \\ \text{and} \quad \Delta_{n,\ell}^{(3)} &= |\mathrm{E}[Z_{1,\ell}^2] - \sigma^2(1-\rho^2)^{-1}|, \end{aligned}$$

show that

$$P(|M_n - \sigma^2(1-\rho^2)^{-1}| > \epsilon) \leq P(\Delta_{n,\ell}^{(1)} > \epsilon/3) + P(\Delta_{n,\ell}^{(2)} > \epsilon/3) + \mathbb{I}(\Delta_{n,\ell}^{(3)} > \epsilon/3)$$

holds for any given $\epsilon > 0$, $\ell > 1$, $n \geq 1$, where $\mathbb{I}(\cdot)$ denotes the indicator function.

21. Prove that $M_n = \sum_{i=1}^{n} X_i^2/n \xrightarrow{p} \sigma^2(1-\rho^2)^{-1}$ as $n \to \infty$. State any standard results that you use.

1. $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ implies $(X_n, Y_n) \xrightarrow{p} (X, Y)$, implying $(X_n, Y_n) \xrightarrow{d} (X, Y)$. From this, there are many other ways to show X, Y must be independent. For example, with characteristic functions of $s, t \in \mathbb{R}$,

$$\phi_X(t)\phi_Y(s) = \lim_{n\to\infty} \phi_{X_n}(t)\phi_{Y_n}(s) = \lim_{n\to\infty} \phi_{X_n,Y_n}(t,s) = \phi_{X,Y}(t,s),$$

implying independence.

Alternatively, fix any $x \in \mathbb{R}$ such that $P(X = x) = 0$ and $y \in \mathbb{R}$ such that $P(Y = y) = 0$. Then, $P(X = x, Y = y) = 0$ and (x, y) is a continuity point of the joint cdf of (X, Y) so that

$$P(X_n \le x, Y_n \le y) \to P(X \le x, Y \le y);$$

also,

$$P(X_n \le x, Y_n \le y) = P(X_n \le x) \cdot P(Y_n \le y) \to P(X \le x) \cdot P(Y \le y)$$

using independence with $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$, x is a continuity point of cdf of X and y is a continuity point of cdf of Y. Hence,

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

for any continuity point x of the cdf of X and any continuity point y of the cdf of Y; hence, $P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$ will hold for any $x, y \in \mathbb{R}$ (by choosing continuity points $x_m \downarrow x, y_m \downarrow y$ and using the right continuity (continuity from above) of the cdfs.

2. For $B_m = \{\omega \in \Omega : |X_m(\omega) - X_0(\omega)| < \frac{1}{2}\}$, $m \ge 1$, we have

$$A_2 = \{\omega \in \Omega : |X_n(\omega) - X_0(\omega)| < \frac{1}{2} \text{ eventually}\} = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} B_m$$

3. The events $A_k = \{\omega \in \Omega : |X_n(\omega) - X_0(\omega)| < \frac{1}{k} \text{ eventually}\}$ are decreasing ($A_{k+1} \subset A_k$), hence by monotone continuity from above

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} P(A_k) = \lim_{k\to\infty} 1 = 1;$$

alternatively, each $P(A_k) = 1$ so that $P\left(\bigcup_{k=1}^{\infty} A_k^c\right) \le \sum_{k=1}^{\infty} P\left(A_k^c\right) = 0$ implying that $P\left(\bigcap_{k=1}^{\infty} A_k\right) = 1 - P\left(\bigcup_{k=1}^{\infty} A_k^c\right) = 1$.

Pick $\omega \in \bigcap_{k=1}^{\infty} A_k$ and $\epsilon > 0$. Then $1/K < \epsilon$ for some integer K and, $\omega \in A_K$ implies, there exists some $N \equiv N(\omega)$ such that $|X_n(\omega) - X_0(\omega)| < 1/K < \epsilon$ for all $n \ge N$. Because $\epsilon > 0$ is arbitrary, then $\lim_{n\to\infty} X_n(\omega) = X_0(\omega)$ holds for any given $\omega \in \bigcap_{k=1}^{\infty} A_k$, where $P\left(\bigcap_{k=1}^{\infty} A_k\right) = 1$.

4. For any $A \subset \mathbb{R}$,

$$\mu^*(A) = \inf\left\{\sum_{i=1}^{\infty} \mu(C_i) : A \subset \bigcup_{i=1}^{\infty} C_i, \{C_i : i \geq 1\} \subset \mathcal{C}\right\}.$$

5. A set $E \subset \mathbb{R}$ is measurable if E acts like a cutting knife

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$$

for any $A \subset \mathbb{R}$. $\mathcal{M}$ is the collection of all measurable sets.

6. The collection $(\mathbb{R}, \mathcal{M}, \mu^*)$ is a measure space (i.e., $\mathcal{M}$ is a σ-algebra on $\mathbb{R}$ and μ^* is a measure on $\mathcal{M}$).

7. $\mathcal{C} \subset \mathcal{M}$ and $\mathcal{B}(\mathbb{R}) = \sigma\langle\mathcal{C}\rangle$ (the Borel sets are generated by C by definition) so that $\mathcal{B}(\mathbb{R}) = \sigma\langle\mathcal{C}\rangle \subset \mathcal{M}$.

8. For any $A \in \mathcal{C}$ (note again $\mathcal{C} \subset \mathcal{M}$), $\mu(A) = \mu^*(A)$.

9. For $\mathbb{R} = (-\infty, \infty) \in \mathcal{C}$, $\mu(\mathbb{R}) = F(\infty) - F(-\infty) = 1$ is a finite measure and $\mathcal{C}$ is a π-class (i.e., $\mathcal{C}$ is closed under intersection). Hence, if any two finite measures on $\sigma\langle\mathcal{C}\rangle$ agree for all sets in $\mathcal{C}$, then the measures must agree on all of $\sigma\langle\mathcal{C}\rangle = \mathcal{B}(\mathbb{R})$.

10. μ is the probability measure corresponding to the Exponential(1) distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. μ has a density with respect to the Lebesgue measure (almost everywhere) equal to $f(x) = e^{-x}$, $x > 0$ and $f(x) = 0$ otherwise for $x \in \mathbb{R}$.

11. Pick $M > 0$ such that $\sup_n \int_{|X_n|>M} |X_n| dP < 1$ then

$$\sup_n \mathrm{E}|X_n| = \sup_n \left(\int_{|X_n|>M} |X_n| dP + \int_{|X_n|\leq M} |X_n| dP \right) \leq 1 + M < \infty.$$

12. Given $\epsilon > 0$, pick $M > 0$ such that $\sup_n \int_{|X_n|>M} |X_n| dP < \epsilon/2$. Then, for $A \in \mathcal{F}$,

$$\sup_n \int_A |X_n| dP \leq \sup_n \left(\int_{|X_n|>M, A} |X_n| dP + \int_{|X_n|\leq M, A} |X_n| dP \right) < \epsilon/2 + MP(A) < \epsilon$$

if $P(A) < \epsilon/(2M)$.

13. Given ϵ, by assumption there exists some δ such that $\int_A |X_n| dP < \epsilon$ if $P(A) < \delta$ for $A \in \mathcal{F}$. Using the Markov inequality $P(|X_n| > M) \leq \mathrm{E}|X_n|/M$, pick large $M > 0$ such that
$$\sup_n P(|X_n| > M) \leq \frac{1}{M} \sup_n \mathrm{E}|X_n| < \delta$$
holds, since $\sup_n \mathrm{E}|X_n| < \infty$ by assumption. Then for each n, $\int_{|X_n|>M} |X_n| dP < \epsilon$ (i.e., take $A = \{|X_n| > M\}$ for each n). Hence, given $\epsilon > 0$, there exists some $M = M(\epsilon)$ such that
$$\sup_n \int_{|X_n|>M} |X_n| dP \leq \epsilon$$
i.e., X_n is uniformly integrable.

14. Let $Z_n = |X_1|^p \mathbb{I}(|X_1| > n)$ where $\mathbb{I}(\cdot)$ is the indicator function. As a function of a fixed $\omega \in \Omega$, $Z_n(\omega) = |X_1(\omega)|^p \mathbb{I}(|X_1(\omega)| > n) \to 0$ as $n \to \infty$ (i.e., as X_1 is a random variable, $X_1(\omega) \in \mathbb{R}$ so that $\mathbb{I}(|X_1(\omega)| > n) \to 0$). Hence, $Z_n \to 0$ almost surely (P). For all $\omega \in \Omega$, $|Z_n(\omega)| \leq |X_1(\omega)|^p$ where $\mathrm{E}[|X_1|^p] < \infty$. By the DCT, $\mathrm{E}Z_n \to \mathrm{E}0 = 0$ as $n \to \infty$, i.e,
$$\lim_{t\to\infty} \int_{|X_1|>t} |X_1|^p dP = \lim_{t\to\infty} \mathrm{E}[|X_1|^p \mathbb{I}(|X_1| > t)] = 0.$$

15. Define iid $Y_i = |X_i|^p$, $i \geq 1$, and set $T_n = \sum_{i=1}^n Y_i/n$, $n \geq 1$. By Problem 14, $\{Y_i\}$ is uniformly integrable:
$$\lim_{t\to\infty} \sup_i \int_{|Y_i|>t} |Y_i| dP = \lim_{t\to\infty} \int_{|X_1|>t^{1/p}} |X_1|^p dP = 0$$
(that is, uniform integrability follows because Y_1 is integrable and $\{Y_i\}$ are iid).

Fix $\epsilon > 0$. By Problem 12 ($\{Y_i\}$ is uniformly integrable), there exists a $\delta_\epsilon > 0$ such that $\sup_n \int_A Y_n dP < \epsilon$ for any $A \in \mathcal{F}$ with $P(A) < \delta_\epsilon$. Hence, if $P(A) < \delta_\epsilon$,
$$\sup_n \int_A T_n dP = \sup_n \frac{1}{n} \sum_{i=1}^n \int_A Y_i dP = \int_A Y_1 dP < \epsilon$$
while $\sup_n \mathrm{E}T_n = \mathrm{E}Y_1 < \infty$; by Problem 13, $\{T_n : n \geq 1\}$ is uniformly integrable.

16. By the SLLN, $S_n/n \to \mathrm{E}X_1$ w.p.1. so that $|S_n/n|^p \to |\mathrm{E}X_1|^p$ w.p.1. From this and because $|S_n/n|^p$ is uniformly integrable (by Jensen's inequality, $|S_n/n|^p \leq \sum_{i=1}^n Y_i/n$ holds and $\sum_{i=1}^n Y_i/n$ is uniformly integrable), it follows immediately that $\lim_{n\to\infty} \mathrm{E}[|S_n/n|^p] = \mathrm{E}\lim_{n\to\infty}[|S_n/n|^p] = \mathrm{E}[|\mathrm{E}X_1|^p] = |\mathrm{E}X_1|^p$ (i.e., uniform integrability on a probability space replaces the DCT). Alternatively, $|S_n/n|^p \leq \sum_{i=1}^n Y_i/n$ where $\sum_{i=1}^n Y_i/n \to \mathrm{E}Y_1 < \infty$ and $\mathrm{E}\sum_{i=1}^n Y_i/n = \mathrm{E}Y_1 \to \mathrm{E}Y_1$; by the extended DCT, the same conclusion holds.

17. Define independent $Z_i = \rho^i \varepsilon_{t-i}$, $i \geq 0$. As $\sum_{i=0}^{\infty} \mathrm{Var}(Z_i) = \sum_{i=0}^{\infty} \sigma^2 \rho^{2i} = \sigma^2(1-\rho^2)^{-1} < \infty$, it follows immediately that $\sum_{i=0}^{\infty} Z_i$ converges with probability 1 (w.p.1) (i.e., Levy's theorem).

    Alternatively, let $Y_i = Z_i \mathbb{I}(|Z_i| \leq 1)$. Then,

$$\sum_{i=0}^{\infty} \mathrm{Var}(Y_i) \leq \sum_{i=0}^{\infty} \mathrm{E}(Y_i^2) \leq \sum_{i=0}^{\infty} \mathrm{E}(Z_i^2) = \sum_{i=0}^{\infty} \rho^{2i}\sigma^2 = \sigma^2/(1-\rho^2) < \infty$$

$$\sum_{i=0}^{\infty} |\mathrm{E}(Y_i)| \leq \sum_{i=0}^{\infty} \mathrm{E}|Y_i| \leq \sum_{i=0}^{\infty} \mathrm{E}|Z_i| = \mathrm{E}|\varepsilon_1| \sum_{i=0}^{\infty} |\rho|^i = \mathrm{E}|\varepsilon_1|/(1-|\rho|) < \infty$$

$$\sum_{i=0}^{\infty} P(|Z_i| > 1) \leq \sum_{i=0}^{\infty} |\mathrm{E}(Z_i)| < \infty$$

    By the three-series theorem, $\sum_{i=1}^{n} Z_i$ converges w.p.1.

18. As an induction hypothesis, assume $X_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$ holds for some $t \geq 0$ (it holds for $t = 0$ by assumption). Then, $X_{t+1} = \rho X_t + \varepsilon_{t+1} = \rho \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} + \varepsilon_{t+1} = \sum_{i=0}^{\infty} \rho^{i+1} \varepsilon_{t-i} + \varepsilon_{t+1} = \sum_{j=1}^{\infty} \rho^j \varepsilon_{t+1-j} + \varepsilon_{t+1} = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t+1-j}$; that is, the form then holds for $t+1$.

19. For a fixed ℓ and $n \geq 1$,

$$\begin{aligned}
\mathrm{E}|M_n - M_{n,\ell}| &= \mathrm{E}\left| \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} X_i Z_{i,\ell} + \frac{1}{n}\sum_{i=1}^{n} X_i Z_{i,\ell} - \frac{1}{n}\sum_{i=1}^{n} Z_{i,\ell}^2 \right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(|X_i| + |Z_{i,\ell}|\right)|Z_{i,\ell} - X_i| \\
&= \mathrm{E}\left(|X_1| + |Z_{1,\ell}|\right)|Z_{1,\ell} - X_1| \\
&\leq \left[\mathrm{E}|X_1|^2 \mathrm{E}|Z_{1,\ell} - X_1|^2\right]^{1/2} + \left[\mathrm{E}|Z_{1,\ell}|^2 \mathrm{E}|Z_{1,\ell} - X_1|^2\right]^{1/2} \\
&\leq 2[C_0 C_{\ell+1}]^{1/2} = 2\frac{\sigma^2}{1-\rho^2}|\rho|^{\ell+1}
\end{aligned}$$

    using that $\{X_i\}_{i\geq 1}$ are identically distributed by Problem 18 (as are $\{Z_{i,\ell}\}_{i\geq 1}$) along with the Cauchy-Schwarz inequality and letting

$$C_j = \sum_{i=j}^{\infty} \mathrm{Var}(\rho^i \varepsilon_{t-i}) = \sum_{i=j}^{\infty} \sigma^2 \rho^{2i} = \frac{\sigma^2 \rho^{2j}}{1-\rho^2}, \quad j \geq 0.$$

    As $|\rho| < 1$,

$$\lim_{\ell\to\infty} \sup_n \mathrm{E}|M_n - M_{n,\ell}| \leq \lim_{\ell\to\infty} 2\frac{\sigma^2}{1-\rho^2}|\rho|^{\ell+1} = 0.$$

20. Recall that $\Delta^{(1)}_{n,\ell} = |M_{n,\ell} - \mathrm{E}Z^2_{1,\ell}|$, $\Delta^{(2)}_{n,\ell} = |M_n - M_{n,\ell}|$, $\Delta^{(3)}_{n,\ell} = |\mathrm{E}Z^2_{1,\ell} - \sigma^2(1-\rho^2)^{-1}|$.

    Note that the event $|M_n - \sigma^2(1-\rho^2)^{-1}| > \epsilon$ is a subset of $\cup_{i=1}^3 \{|\Delta^{(i)}_{n,\ell}| > \epsilon/3\}$. That is, $|M_n - \sigma^2(1-\rho^2)^{-1}| > \epsilon$ implies that $|\Delta^{(i)}_{n,\ell}| > \epsilon/3$ for some $i = 1,2,3$; if $|\Delta^{(i)}_{n,\ell}| \le \epsilon/3$ did hold for each $i = 1,2,3$, then $|M_n - \sigma^2(1-\rho^2)^{-1}| \le \sum_{i=1}^3 |\Delta^{(i)}_{n,\ell}| \le \epsilon$. Hence,

$$P(|M_n - \sigma^2(1-\rho^2)^{-1}| > \epsilon) \le P\left(\cup_{i=1}^3 \{|\Delta^{(i)}_{n,\ell}| > \epsilon/3\}\right) \le \sum_{i=1}^3 P\left(|\Delta^{(i)}_{n,\ell}| > \epsilon/3\right)$$

    where $P(|\Delta^{(3)}_{n,\ell}| > \epsilon/3) = \mathbb{I}(|\Delta^{(3)}_{n,\ell}| > \epsilon/3)$ as there is nothing random about $\Delta^{(3)}_{n,\ell}$.

21. Fix $\epsilon > 0$ and $\delta > 0$. By Problem 19, there exists some ℓ_1 such that

$$\sup_n P(\Delta^{(2)}_{n,\ell} > \epsilon/3) \le \sup_n \mathrm{E}|M_n - M_{n,\ell}|/(\epsilon/3) < \delta/2$$

    for all $\ell \ge \ell_1$. Also,

$$\lim_{\ell\to\infty} \mathrm{E}Z^2_{1,\ell} = \lim_{\ell\to\infty} \sigma^2 \sum_{i=0}^{\ell} \rho^{2i} = \sigma^2 \sum_{i=0}^{\infty} \rho^{2i} = \frac{\sigma^2}{1-\rho^2}$$

    so that there exists some ℓ_2 such that $|\mathrm{E}Z^2_{1,\ell} - \sigma^2(1-\rho^2)^{-1}| \le \epsilon/3$ for all $\ell \ge \ell_2$, implying

$$\mathbb{I}(|\Delta^{(3)}_{n,\ell}| > \epsilon/3) = \mathbb{I}(|\mathrm{E}Z^2_{1,\ell} - \sigma^2(1-\rho^2)^{-1}| > \epsilon/3) = 0$$

    for all $\ell \ge \ell_2$.

    Fix $\ell = \max\{\ell_1, \ell_2\}$; for this fixed ℓ, $M_{n,\ell} \to \mathrm{E}Z^2_{1,\ell}$ w.p.1, implying $M_{n,\ell} \xrightarrow{p} \mathrm{E}Z^2_{1,\ell}$, implying there exists an $N = N(\delta)$ such that $P(|M_{n,\ell} - \mathrm{E}Z^2_{1,\ell}| > \epsilon/3) < \delta/2$ for all $n \ge N$.

    Hence, for all $n \ge N = N(\delta)$ (and fixed $\ell = \max\{\ell_1, \ell_2\}$), by Problem 21,

$$P(|M_n - \sigma^2(1-\rho^2)^{-1}| > \epsilon) < \delta/2 + \delta/2 + 0 = \delta.$$

    As $\epsilon, \delta > 0$ were arbitrary, $M_n \xrightarrow{p} \sigma^2(1-\rho^2)^{-1}$ by definition.

Part I

Let $\underline{X} = (X_1, \ldots, X_n)$ (where $n \geq 2$) be a random vector such that X_1 is an exponential random variable with mean θ, and for $i = 2, \ldots, n$, conditional on $(X_1, \ldots, X_{i-1})$, X_i is an exponential random variable with mean θ / X_{i-1}. That is, the joint probability density function (pdf) of $\underline{X}$ is given by

$$f_\theta(\underline{x}) = \frac{x_1 \cdots x_{n-1}}{\theta^n} \exp\left\{ -\frac{x_1 + x_1 x_2 + \cdots + x_{n-1} x_n}{\theta} \right\}, \text{ for } \underline{x} \equiv (x_1, \ldots, x_n) \in (0, \infty)^n,$$

where $\theta \in (0, \infty)$ is an unknown parameter. Let $T(\underline{X}) = X_1 + \sum_{i=2}^n X_{i-1} X_i$.

1. Show that there exists a unique maximum likelihood estimator (MLE) of θ, $\widehat{\theta}_n$, and give an explicit expression for $\widehat{\theta}_n$.

2. Find a uniformly minimum variance unbiased estimator (UMVUE) of θ, $\widetilde{\theta}_n$.

3. Find the variance of $\widetilde{\theta}_n$, $\operatorname{Var}_\theta \widetilde{\theta}_n$, and show that $\operatorname{Var}_\theta \widetilde{\theta}_n$ attains the Cramér-Rao lower bound for the variance of any unbiased estimator of θ.

4. Consider estimating θ using the loss function
$$L(\theta, a) = \frac{(\theta - a)^2}{\theta^2}$$
in the class of estimators $\{\delta_c(\underline{X}) \in (0, \infty) | \delta_c(\underline{X}) = cT(\underline{X}),\ c \in (0, \infty)\}$. Find the best (minimum risk) estimator $\delta_{c_0}(\underline{X})$, and show that $\widetilde{\theta}_n$ in Question 3 is inadmissible.

5. Find the likelihood ratio test (LRT) statistic for testing $H_0\colon \theta = \theta_0$ versus $H_a\colon \theta \neq \theta_0$, where $\theta_0 \in (0, \infty)$ is a given number. Simplify this statistic (as much as possible) as a function of $T(\underline{X})$ and θ_0.

6. Show that the LRT of size $\alpha \in (0, 1)$ in Question 5 rejects H_0 if $T(\underline{X}) < c_1$ or $T(\underline{X}) > c_2$ for some constants c_1 and c_2 such that $0 < c_1 < c_2 < \infty$. Write down two equations from which c_1 and c_2 can be found numerically. You do not need to derive the pdf of $T(\underline{X})$.

Part II

Let $\{X_n\}_{n \geq 1}$ be a sequence of independent random variables and $c \in (0, \infty)$ be a given constant such that

$$\sum_{n=1}^\infty P(|X_n| > c) < \infty.$$

For $n \geq 1$, let

$$U_n = X_n I(|X_n| \leq c) \text{ and } V_n = X_n I(|X_n| > c);$$

let

$$S_n = \sum_{i=1}^{n} X_i, \quad S_n^{(1)} = \sum_{i=1}^{n} U_i, \quad \text{and} \quad S_n^{(2)} = \sum_{i=1}^{n} V_i;$$

let

$$\mu_n = ES_n^{(1)} = \sum_{i=1}^{n} EU_i \quad \text{and} \quad s_n = \left\{\text{Var}\, S_n^{(1)}\right\}^{1/2} = \left\{\sum_{i=1}^{n} \text{Var}\, U_i\right\}^{1/2}.$$

7. Show that $S_n^{(2)}$ converges with probability one as $n \to \infty$.

8. Suppose that $s_n \to \infty$ as $n \to \infty$. Show that there exist constants a_n and b_n (that may depend on μ_n and s_n) such that $(S_n - a_n)/b_n \xrightarrow{d} N(0,1)$.

9. Give a sequence of independent random variables $\{X_n\}_{n\geq 1}$ such that $\sum_{i=1}^n X_i/\sqrt{n} \xrightarrow{d} N(0,1)$, but none of the X_n's have finite means. Justify your conclusions.

Part III

10. Let X and Y be random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that $EX^2 < \infty$ and $EY^2 < \infty$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra. Show that

$$E(|XY|\,|\mathcal{G}) \leq \sqrt{E(X^2|\mathcal{G})E(Y^2|\mathcal{G})} \quad \text{a.s.}$$

Part IV

Let $Y_1, \ldots, Y_n$ be independent and identically distributed (iid) uniform $(0, \theta)$ random variables, where $\theta \in (1, \infty)$ is an unknown parameter. Suppose that we only observe

$$X_i = \begin{cases} Y_i & \text{if } Y_i > 1, \\ 1 & \text{if } Y_i \leq 1, \end{cases} \quad \text{for } i = 1, \ldots, n.$$

Let $X_{(n)} = \max_{1\leq i\leq n} X_i$ and $P_\theta^{X_{(n)}}$ be the probability distribution of $X_{(n)}$. Let μ be the Lebesgue measure on $[1, \infty)$, δ be the point mass at 1 (that is, $\delta(\{1\}) = 1$), and $\nu = \mu + \delta$.

11. Show that the Radon-Nikodym (R-N) derivative of $P_\theta^{X_{(n)}}$ with respect to ν is

$$f_\theta^{X_{(n)}}(x) = \theta^{-n} I_{\{1\}}(x) + n\theta^{-n}x^{n-1}I_{(1,\theta)}(x),$$

where $I_{\{1\}}(x) = 1$ if $x = 1$ and $I_{\{1\}}(x) = 0$ if $x \neq 1$.

12. Show that $X_{(n)}$ is a sufficient statistic for θ.

13. Show that $X_{(n)}$ is a complete statistic for θ.

Part V

14. Suppose that X is an Erlang $(k, 1)$ random variable with the pdf

$$f_k(x) = \frac{x^{k-1}e^{-x}}{(k-1)!}, \text{ for } x \in (0, \infty),$$

where k is an unknown positive integer. Consider estimating k with the loss function

$$L(k, a) = \frac{(k-a)^2}{k},$$

where $a \in \mathcal{A} \equiv \{1, 2, \ldots, \}$. Find a Bayes estimator of k with respect to the geometric (p) prior with the probability mass function (pmf)

$$g_p(k) = p(1-p)^{k-1}, \text{ for } k = 1, 2, \ldots,$$

where $p \in (0, 1)$ is known.

Part VI

15. Suppose that X is a geometric (p) random variable with the pmf

$$f_p(x) = p(1-p)^{x-1}, \text{ for } x = 1, 2, \ldots,$$

where $p \in \Theta \equiv (0, 1)$ is an unknown parameter. Suppose that one wishes to estimate p with the loss function

$$L(p, a) = \frac{(p-a)^2}{p(1-p)},$$

where $a \in \mathcal{A} \equiv [0, 1]$. Show that

$$\delta_0(X) = \begin{cases} 1, & \text{if } X = 1, \\ 0, & \text{if } X \geq 2, \end{cases}$$

is minimax in the class of estimators of p, $\{\delta | \delta(x) \in \mathcal{A} \text{ for } x = 1, 2, \ldots\}$.

1. The log-likelihood function is

$$L_n(\theta) = \log f_\theta(\underline{X}) = \sum_{i=1}^{n-1} \log X_i - n \log \theta - T(\underline{X})/\theta.$$

Thus, $L_n'(\theta) = -n/\theta + T(\underline{X})/\theta^2$. Solving the likelihood equation $L_n'(\theta) = 0$ for $\theta \in (0, \infty)$ gives $\theta = \widehat{\theta}_n = T(\underline{X})/n$. Since $L_n'(\theta) > 0$ for $\theta \in (0, \widehat{\theta}_n)$ and $L_n'(\theta) < 0$ for $\theta \in (\widehat{\theta}_n, \infty)$, $\widehat{\theta}_n$ is the unique MLE of θ.

2. Note that the joint pdf of $\underline{X}$ can be written as

$$f_\theta(\underline{x}) = \exp\left\{-n \log \theta + (-\theta^{-1}) T(\underline{x})\right\} \prod_{i=1}^{n-1} x_i,$$

which forms a one-dimensional exponential family for $\theta \in \Theta \equiv (0, \infty)$. Note that $\Gamma_\Theta = \{-\theta^{-1} \in \mathbb{R}^1 | \theta \in \Theta\} = (-\infty, 0)$ contains an open interval in $\mathbb{R}^1$. Thus $T(\underline{X})$ is a sufficient and complete statistic for θ.

Let $\eta = -\theta^{-1}$. Then $\theta = (-\eta)^{-1}$ and

$$f_\theta(\underline{x}) = (-\eta)^n \exp\{\eta T(\underline{x})\} \prod_{i=1}^{n-1} x_i.$$

Thus

$$E_\theta T(\underline{X}) = \frac{d}{d\eta}\left\{-\log(-\eta)^n\right\} = n(-\eta)^{-1} = n\theta.$$

So $\widetilde{\theta}_n \equiv T(\underline{X})/n = \widehat{\theta}_n$ is an unbiased estimator of θ. By the Lehmann-Scheffé theorem, $\widetilde{\theta}_n$ is a UMVUE of θ.

3. It follows from the solution to Question 2 that

$$\mathrm{Var}_\theta\, T(\underline{X}) = \frac{d^2}{d\eta^2}\left\{-\log(-\eta)^n\right\} = \frac{d}{d\eta}\{n(-\eta)^{-1}\} = n(-\eta)^{-2} = n\theta^2.$$

Thus $\mathrm{Var}_\theta\, \widetilde{\theta}_n = \theta^2/n$.

The Fisher information about θ contained in $\underline{X}$ at θ is

$$I(\theta) = E_\theta\left\{\frac{d}{d\theta} \log f_\theta(\underline{X})\right\}^2 = E_\theta\left\{-\frac{n}{\theta} + \frac{T(\underline{X})}{\theta^2}\right\}^2 = \frac{\mathrm{Var}_\theta\, T(\underline{X})}{\theta^4} = \frac{n}{\theta^2}.$$

Therefore, $\mathrm{Var}_\theta\, \widetilde{\theta}_n = 1/I(\theta)$, the Cramér-Rao lower bound for the variance of any unbiased estimator of θ.

4. The risk function of the estimator $\delta_c(\underline{X})$ is

$$R(\theta, \delta_c) = E_\theta\left\{\frac{(\theta - cT(\underline{X}))^2}{\theta^2}\right\} = \frac{1}{\theta^2}\left\{\mathrm{Var}_\theta[\theta - cT(\underline{X})] + [\theta - cE_\theta T(\underline{X})]^2\right\}$$

$$= \frac{c^2 n\theta^2 + (\theta - cn\theta)^2}{\theta^2} = (n^2+n)c^2 - 2cn + 1 = (n^2+n)\left(c - \frac{1}{n+1}\right)^2 + \frac{1}{n+1},$$

which is a quadratic function and is minimized at $c_0 \equiv 1/(n+1)$. So the best estimator is

$$\delta_{c_0}(\underline{X}) = \frac{T(\underline{X})}{n+1}.$$

Since $R(\theta, \widetilde{\theta}_n) = R(\theta, \delta_{1/n}) = 1/n > 1/(n+1) = R(\theta, \delta_{c_0})$ for all $\theta \in (0,\infty)$, $\widetilde{\theta}_n$ is inadmissible.

5. The LRT statistic is

$$\begin{aligned} \mathrm{LR}(\underline{X}) &= \frac{\sup_{\theta\in(0,\infty)} f_\theta(\underline{X})}{f_{\theta_0}(\underline{X})} = \frac{f_{\widehat{\theta}_n}(\underline{X})}{f_{\theta_0}(\underline{X})} = \frac{(T(\underline{X})/n)^{-n} e^{-n}}{\theta_0^{-n}\exp\{-T(\underline{X})/\theta_0\}} \\ &= n^n e^{-n}\{T(\underline{X})/\theta_0\}^{-n}\exp\{T(\underline{X})/\theta_0\}. \end{aligned}$$

6. The LRT of size α rejects H_0 if $\mathrm{LR}(\underline{X}) > k$ for some constant $k \in (0,\infty)$ such that $P_{\theta_0}(\mathrm{LR}(\underline{X}) > k) = \alpha$. That is,

$$g(T(\underline{X})) \equiv T(\underline{X}) - n\theta_0 \log T(\underline{X}) > c$$

for some constant $c \in \mathbb{R}^1$ such that $P_{\theta_0}(g(T(\underline{X})) > c) = \alpha$.

Note that $g'(t) = 1 - n\theta_0/t < 0$ for $t \in (0, n\theta_0)$ and $g'(t) > 0$ for $t \in (n\theta_0, \infty)$. Thus the LRT of size α rejects H_0 if $T(\underline{X}) < c_1$ or $T(\underline{X}) > c_2$ for some constants $c_1 \in (0, n\theta_0)$ and $c_2 \in (n\theta_0, \infty)$ such that

$$P_{\theta_0}(T(\underline{X}) < c_1) + P_{\theta_0}(T(\underline{X}) > c_2) = \alpha,$$

and $g(c_1) = g(c_2)$, that is,

$$c_1 - c_2 = n\theta_0 \log(c_1/c_2).$$

7. Clearly, $\{V_n\}_{n\ge1}$ is a sequence of independent random variables. For $n \ge 1$,

$$|V_n| = |X_n| I(|X_n| > c) = \begin{cases} 0, & \text{if } |X_n| \le c, \\ |X_n| > c, & \text{if } |X_n| > c. \end{cases}$$

Thus $W_n \equiv V_n I(|V_n| \le c) = 0$. So $\sum_{n=1}^\infty EW_n = 0$ converges, and $\sum_{n=1}^\infty \mathrm{Var}\, W_n = 0 < \infty$. Furthermore,

$$\sum_{n=1}^\infty P(|V_n| > c) = \sum_{n=1}^\infty P(|X_n| > c) < \infty.$$

By Kolmogorov's three-series theorem, $S_n^{(2)}$ converges with probability one.

8. Note that $\{U_n\}_{n\ge1}$ is a sequence of independent random variables with $|U_n| \le c < \infty$. Let $\widetilde{U}_n = U_n - EU_n$. Then $|\widetilde{U}_n| \le 2c < \infty$, $E\widetilde{U}_n = 0$, and $E\widetilde{U}_n^2 = \mathrm{Var}\, U_n$. Therefore,

$$\frac{1}{s_n^3}\sum_{j=1}^n E|\widetilde{U}_j|^3 \le \frac{2c}{s_n^3}\sum_{j=1}^n E\widetilde{U}_j^2 = \frac{2c}{s_n^3}\cdot s_n^2 = \frac{2c}{s_n} \to 0 \quad \text{as } n \to \infty.$$

This shows that the Lyapunov condition holds for $\{\widetilde{U}_n\}_{n\geq 1}$ (for $\delta = 1$) and thus

$$\frac{\sum_{j=1}^n \widetilde{U}_j}{s_n} = \frac{S_n^{(1)} - \mu_n}{s_n} \xrightarrow{d} N(0,1).$$

Let $a_n = \mu_n$ and $b_n = s_n$. Then

$$\frac{S_n - a_n}{b_n} = \frac{S_n^{(1)} - \mu_n}{s_n} + \frac{S_n^{(2)}}{s_n} \xrightarrow{d} N(0,1),$$

because $S_n^{(2)}/s_n \to 0$ with probability one, by the result from Question 7.

9. Consider a sequence of independent random variables $\{X_n\}_{n\geq 1}$ such that for $n \geq 1$,

$$P(X_n = 1) = P(X_n = -1) = (1 - 2^{-n})/2, \text{ and } P(X_n = 2^j) = 2^{-j} \text{ for } j > n.$$

Then for $c = 1$,

$$\sum_{n=1}^{\infty} P(|X_n| > c) = \sum_{n=1}^{\infty} P(|X_n| > 1) = \sum_{n=1}^{\infty} 2^{-n} = 1 < \infty,$$

$$\mu_n = \sum_{i=1}^n EU_i = \sum_{i=1}^n E\{X_i I(|X_i| \leq 1)\} = \sum_{i=1}^n \{1P(X_i = 1) + (-1)P(X_i = -1)\} = 0,$$

and

$$s_n^2 = \sum_{i=1}^n \operatorname{Var} U_i = \sum_{i=1}^n EU_i^2 = \sum_{i=1}^n (1 - 2^{-i}) = n - 1 + 2^{-n}.$$

It follows from the solution to Question 8 that $\sum_{i=1}^n X_i/s_n \xrightarrow{d} N(0,1)$. Thus

$$\sum_{i=1}^n X_i/\sqrt{n} = \left(\sum_{i=1}^n X_i/s_n\right)\left(1 - \frac{1}{n} + \frac{1}{n2^n}\right)^{1/2} \xrightarrow{d} N(0,1).$$

10. Let $U = \sqrt{E(X^2|\mathcal{G})}$ and $V = \sqrt{E(Y^2|\mathcal{G})}$. If $UV = 0$ a.s., then either $X = 0$ a.s. or $Y = 0$ a.s. Thus $XY = 0$ a.s. and $E(|XY|\,|\mathcal{G}) = 0$ a.s., so the inequality holds.

For $UV > 0$ a.s., noting that for a, $b \in \mathbb{R}^1$, $0 \leq (|a| - |b|)^2 = a^2 + b^2 - 2|a||b|$, we have $|ab| \leq a^2/2 + b^2/2$ and thus

$$\left|\frac{X}{U}\frac{Y}{V}\right| \leq \frac{X^2}{2U^2} + \frac{Y^2}{2V^2}.$$

Therefore, by the monotonicity and linearity of conditional expectations, we have

$$E\left(\frac{|XY|}{UV}\middle|\,\mathcal{G}\right) \leq \frac{1}{2}E\left(\frac{X^2}{U^2}\middle|\,\mathcal{G}\right) + \frac{1}{2}E\left(\frac{Y^2}{V^2}\middle|\,\mathcal{G}\right) \quad \text{a.s.}$$

Since U and V are both $\mathcal{G}$-measurable, so are $\frac{1}{UV}$, $\frac{1}{U^2}$, and $\frac{1}{V^2}$. Thus,

$$\frac{1}{UV}E\left(|XY|\,|\,\mathcal{G}\right) \leq \frac{1}{2U^2}E\left(X^2\middle|\,\mathcal{G}\right) + \frac{1}{2V^2}E\left(Y^2\middle|\,\mathcal{G}\right) = \frac{U^2}{2U^2} + \frac{V^2}{2V^2} = 1 \quad \text{a.s.}$$

Multiplying by $UV > 0$, this shows that

$$E(|XY|\,|\mathcal{G}) \leq UV = \sqrt{E(X^2|\mathcal{G})E(Y^2|\mathcal{G})} \quad \text{a.s.}$$

11. Note that $X_{(n)} = 1$ if and only if $Y_1 \leq 1, \ldots, Y_n \leq 1$. If $X_{(n)} > 1$, then $X_{(n)} = Y_{(n)} \equiv \max_{1\leq i\leq n} Y_i$. Thus, for any set $A \in \mathcal{B}([1, \theta))$,

$$\begin{aligned}
P_\theta^{X_{(n)}}(A) &= P_\theta(X_{(n)} \in A) = P_\theta(X_{(n)} \in A, X_{(n)} = 1) + P_\theta(X_{(n)} \in A, X_{(n)} > 1) \\
&= I_A(1)P_\theta(Y_1 \leq 1, \ldots, Y_n \leq 1) + P_\theta(Y_{(n)} \in A, Y_{(n)} > 1) \\
&= I_A(1)\theta^{-n} + \int_A n\theta^{-n}x^{n-1}I_{(1,\theta)}(x)d\mu(x) \\
&= \int_A \left(\theta^{-n}I_{\{1\}}(x) + n\theta^{-n}x^{n-1}I_{(1,\theta)}(x)\right) d\nu(x).
\end{aligned}$$

Thus the R-N derivative of $P_\theta^{X_{(n)}}$ with respect to ν is

$$f_\theta^{X_{(n)}}(x) = \theta^{-n}I_{\{1\}}(x) + n\theta^{-n}x^{n-1}I_{(1,\theta)}(x).$$

12. Note that for $n = 1$, $X_1 = X_{(1)}$. Thus from Question 11, the R-N derivative of the probability distribution of X_1 with respect to ν is

$$f_\theta^{X_1}(x) = \theta^{-1}I_{\{1\}}(x) + \theta^{-1}I_{(1,\theta)}(x) = \theta^{-1}I_{[1,\theta)}(x).$$

Thus the R-N derivative of the joint probability distribution of $\underline{X} = (X_1, \ldots, X_n)$, denoted by $P_\theta^{\underline{X}}$, with respect to ν^n is

$$f_\theta^{\underline{X}}(x_1, \ldots, x_n) = \prod_{i=1}^n \left\{\theta^{-1}I_{[1,\theta)}(x_i)\right\} = \theta^{-n}I_{[1,\theta)}(x_{(n)})I_{[1,\infty)}(x_{(1)}),$$

where $x_{(1)} = \min_{1\leq i\leq n} x_i$ and $x_{(n)} = \max_{1\leq i\leq n} x_i$. By the factorization theorem, $X_{(n)}$ is a sufficient statistic for θ.

13. For any measurable function of $X_{(n)}$, $h(X_{(n)})$, such that

$$E_\theta h(X_{(n)}) = 0 \text{ for all } \theta \in (1, \infty),$$

we need to show that $h(X_{(n)}) = 0$ a.s. $P_\theta^{\underline{X}}$ for all $\theta \in (1, \infty)$. It follows from Question 11 that

$$0 = E_\theta h(X_{(n)}) = \theta^{-n}h(1) + n\theta^{-n}\int_1^\theta h(x)x^{n-1}d\mu(x), \text{ for all } \theta \in (1, \infty).$$

Thus

$$\int_1^\theta h(x)x^{n-1}d\mu(x) = -\frac{h(1)}{n}, \text{ for all } \theta \in (1, \infty).$$

Letting $\theta \to 1$ we obtain that $h(1) = 0$. Differentiating both sides of the above equation with respect to θ we obtain that $h(x)x^{n-1} = 0$ and hence $h(x) = 0$ for all $x > 1$. Thus $h(x) = 0$ for all $x \geq 1$ and then $h(X_{(n)}) = 0$ a.s. $P_\theta^{\underline{X}}$ for all $\theta \in (1, \infty)$. Therefore, $X_{(n)}$ is a complete statistic for θ.

14. Note that

$$\sum_{j=1}^{\infty} f_j(x)g_p(j) = \sum_{j=1}^{\infty} \frac{x^{j-1}e^{-x}}{(j-1)!}p(1-p)^{j-1} = pe^{-x}\sum_{j=1}^{\infty}\frac{[(1-p)x]^{j-1}}{(j-1)!} = pe^{-x}e^{(1-p)x} = pe^{-px}.$$

Thus the posterior pmf of k given $X = x$ is

$$f^{k|X}(k|x) = \frac{f_k(x)g_p(k)}{\sum_{j=1}^{\infty} f_j(x)g_p(j)} = \frac{[(1-p)x]^{k-1}}{(k-1)!}e^{-(1-p)x} = \frac{u^{k-1}}{(k-1)!}e^{-u},$$

where $u = (1-p)x$. Furthermore, for $b \in [0,\infty)$,

$$E[L(k,b)|X=x] = E\left\{ \frac{(k-b)^2}{k} \middle| X = x \right\} = E(k|X=x) - 2b + E(k^{-1}|X=x)b^2$$

is a quadratic function of b and is minimized at

$$b^* = \left[E(k^{-1}|X=x)\right]^{-1} = \left[\sum_{k=1}^{\infty}\frac{1}{k}\frac{u^{k-1}}{(k-1)!}e^{-u}\right]^{-1} = \left[\frac{1}{u}\left(e^u - 1\right)e^{-u}\right]^{-1} = \frac{u}{1-e^{-u}}.$$

Thus, for $a \in \mathcal{A} = \{1, 2, \ldots\}$, $E[L(k,a)|X=x]$ is minimized at $a^* = \max(1, \lfloor b^* + 0.5\rfloor)$, where $\lfloor b^* + 0.5\rfloor$ denotes the largest integer less than or equal to $b^* + 0.5$. Therefore, a Bayes estimator of k is given by

$$\max\left(1, \left\lfloor \frac{(1-p)X}{1-e^{-(1-p)X}} + 0.5 \right\rfloor\right).$$

15. The risk function of δ_0 is

$$\begin{aligned} R(p,\delta_0) &= E_pL(p,\delta_0(X)) = \frac{(p-1)^2p}{p(1-p)} + \sum_{x=2}^{\infty}\frac{(p-0)^2p(1-p)^{x-1}}{p(1-p)} \\ &= 1 - p + p^2\sum_{x=2}^{\infty}(1-p)^{x-2} = 1 - p + p^2\cdot\frac{1}{p} = 1. \end{aligned}$$

The risk function of any estimator δ is

$$\begin{aligned} R(p,\delta) &= E_pL(p,\delta) = \sum_{x=1}^{\infty}[\delta(x)-p]^2(1-p)^{x-2} \\ &= \frac{[\delta(1)-p]^2}{1-p} + \sum_{x=2}^{\infty}[\delta(x)-p]^2(1-p)^{x-2}. \end{aligned}$$

If $\delta(1) \neq 1$, then $R(p,\delta) \to \infty$ as $p \to 1$ and thus $\sup_{0<p<1} R(p,\delta) = \infty$. If $\delta(1) = 1$, then

$$\sup_{0<p<1} R(p,\delta) \geq \lim_{p\to 0} R(p,\delta) = 1 + \sum_{x=2}^{\infty}[\delta(x)]^2 \geq 1.$$

Hence δ_0 is minimax.

Part I

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed standard uniform $U(0,1)$ random variables. For any $a \in \mathbb{R}$, denote the fractional part of a by a **mod** $1 \equiv a - \lfloor a \rfloor$, where $\lfloor a \rfloor$ is the largest integer smaller than or equal to a. Write $U_i = (\sum_{j=1}^{i} X_j)$ **mod** 1.

1. Show that the distribution of any U_i is also standard uniform $U(0,1)$. (**Hint:** Use induction.)

2. Show that U_i and U_k are independent for any $i < k$. (**Hint:** Show that U_1 and U_2 are independent. Thus, thus U_i and U_{i+1} are independent $\forall i = 1, 2, \ldots, n-1$ and proceed.)

Part II

Suppose that $X_1, X_2, \ldots, X_n$ be a simple random sample from the standard uniform $U(0,1)$ distribution. Let $Y_i = \frac{X_i}{\sum_{i=1}^{n} X_i}$ for $i = 1, 2, \ldots, n-1$.

3. Show that the joint density of $(Y_1, Y_2, \ldots, Y_{n-1})$ is given by

$$f(y_1, y_2, \ldots, y_{n-1}) = \frac{1}{n}\left(\min\left\{\frac{1}{y_1}, \frac{1}{y_2}, \ldots, \frac{1}{y_{n-1}}, \frac{1}{1 - \sum\limits_{i=1}^{n-1} y_i}\right\}\right)^n,$$

for $y_1, y_2, \ldots, y_{n-1} > 0$ and $\sum_{i=1}^{n-1} y_i < 1$.

Part III

The following theorems may be helpful in answering the questions in this part.

- **Sherman-Morrison Formula:** Let $\boldsymbol{A}$ be an invertible $n \times n$-matrix and $\boldsymbol{U}, \boldsymbol{V}$ be n-variate real vectors. Then, assuming that the inverse of $\boldsymbol{A} + \boldsymbol{U}\boldsymbol{V}'$ exists, it is given by

$$(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{V}')^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{U}\boldsymbol{V}'\boldsymbol{A}^{-1}}{1 + \boldsymbol{V}'\boldsymbol{A}^{-1}\boldsymbol{U}}.$$

- **Matrix Determinant Formula:** Let $\boldsymbol{A}$ be an invertible $n \times n$-matrix and $\boldsymbol{U}, \boldsymbol{V}$ be n-variate real vectors. Then, assuming that the determinant of $\boldsymbol{A} + \boldsymbol{U}\boldsymbol{V}'$ exists, it is given by

$$\mid \boldsymbol{A} + \boldsymbol{U}\boldsymbol{V}' \mid = (1 + \boldsymbol{V}'\boldsymbol{A}^{-1}\boldsymbol{U}) \mid \boldsymbol{A} \mid .$$

Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ be defined as follows:

$$Y_i = \alpha + \beta X_i + \epsilon_i, \qquad i = 1, 2, \ldots, n \tag{1}$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ are n non-stochastic variables and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ is a vector of zero-mean jointly multivariate normally-distributed random variables with a marginally exchangeable correlation structure. That is, $\boldsymbol{\epsilon} \sim MVN(\boldsymbol{0}, \sigma^2\boldsymbol{R})$ where $\boldsymbol{0} = (0, 0, \ldots, 0)$ and $\boldsymbol{R} = ((\rho_{ij}))$ is a correlation matrix with an exchangeable structure, *i.e.* $\rho_{ij} \equiv \rho \quad \forall \quad i \neq j \in \{1, 2, \ldots, n\}$. (Of course, $\rho_{ii} = 1 \quad \forall i = 1, 2, \ldots, n$.) An alternative way of writing (1) is

$$\boldsymbol{Y} = \alpha\boldsymbol{1} + \beta\boldsymbol{X} + \boldsymbol{\epsilon}.$$

Note that $\boldsymbol{1}$ here is an n-variate vector of 1's, *i.e.*, $\boldsymbol{1} \equiv (1, 1, \ldots, 1)'$, and that α, β, σ (and ρ) are all scalar quantities. For this problem, we assume that $\rho \geq 0$.

4. Show that the maximum likelihood estimates (MLEs) of (α, β) are the same as for simple linear regression (*i.e.*, in the case when $\rho = 0$).

5. Show that it is not possible to obtain MLEs of both ρ and σ when these (along with α and β) are unknown. (**Hint:** Consider the profile likelihood of (ρ, σ).)

6. Let the non-stochastic variables $X_1, X_2, \ldots, X_n$ have mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Suppose now that $\bar{X}_n$ is such that $\mid \bar{X}_n \mid \leq c \forall n$ and that $\sum_{i=1}^n (X_i - \bar{X}_n)^2 \to \infty$ as $n \to \infty$. Show the following with regard to the MLEs of $(\hat{\alpha}, \hat{\beta})$:

   a) The MLE $\hat{\beta}$ is unbiased.

   b) The MLE $\hat{\alpha}$ is unbiased.

   c) The MLE $\hat{\beta}$ is MSE-consistent.

   d) The MLE $\hat{\alpha}$ is not consistent as long as $\rho > 0$.

7. Find the MLE of $\sigma^2(1-\rho)$. Show that this is independent of $\hat{\beta}$. (**Hint:** Consider showing that each residual is independent of $\hat{\beta}$ and proceed accordingly.)

Part IV

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be n independent identically distributed random vectors from $f(\boldsymbol{x}; \boldsymbol{\varphi})$. Assume that there are redundancies in the specification of the p-variate vector $\boldsymbol{\varphi}$ and in particular, that $\boldsymbol{i}(\boldsymbol{\varphi})$, the Fisher information for $\boldsymbol{\varphi}$ based on observation X_i has rank $q < p$. Let $\hat{\boldsymbol{\varphi}}_n$ be the MLE of $\boldsymbol{\varphi}$. It is known that, as in the steps of the Multidimensional Central Limit Theorem (Breiman, 1982),

$$\sqrt{n}\boldsymbol{i}(\boldsymbol{\varphi})(\hat{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{i}(\boldsymbol{\varphi})). \tag{2}$$

Consider the spectral decomposition of $\boldsymbol{i}(\boldsymbol{\varphi})$ given by $\boldsymbol{i}(\boldsymbol{\varphi}) = \boldsymbol{\Gamma\Lambda\Gamma}'$, where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1 \vdots \boldsymbol{\gamma}_2 \vdots \ldots \vdots \boldsymbol{\gamma}_q)$ with $\boldsymbol{\gamma}_i, i = 1, 2, \ldots, q$ denoting the ith (p-variate) eigenvector corresponding to the ith eigenvalue λ_i, and $\boldsymbol{\Lambda} = diag\{\lambda, \ldots, \lambda_q\}$ being the diagonal matrix of positive eigenvalues of $\boldsymbol{i}(\boldsymbol{\varphi})$ of order $q \times q$. (**Note:** $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}_q$, but $\boldsymbol{\Gamma\Gamma}' \neq \boldsymbol{I}_p$). Further, let $\boldsymbol{i}^-(\boldsymbol{\varphi}) = \boldsymbol{\Gamma\Lambda}^{-1}\boldsymbol{\Gamma}'$ be a Moore-Penrose (generalized) inverse of $\boldsymbol{i}(\boldsymbol{\varphi})$. Consider a lower-dimensional representation $\boldsymbol{\zeta}$ of $\boldsymbol{\varphi}$, such that $\boldsymbol{\varphi} = \boldsymbol{\Gamma\zeta}$. Assume also that $\hat{\boldsymbol{\varphi}}_n = \boldsymbol{\Gamma}\hat{\boldsymbol{\zeta}}_n$.

8. Show that
$$\sqrt{n}(\hat{\boldsymbol{\zeta}}_n - \boldsymbol{\zeta}) \xrightarrow{d} N_q(\mathbf{0}, \boldsymbol{\Lambda}^{-1}).$$

9. Suppose that $\boldsymbol{C}$ is an $m \times p$-matrix ($m < q$) and that we are interested in testing the null hypothesis $\boldsymbol{C}\boldsymbol{\varphi} = \mathbf{0}$. Show that the Wald test statistic reduces to $n(\boldsymbol{C}\hat{\boldsymbol{\varphi}}_n)'(\boldsymbol{C}\boldsymbol{i}^-(\boldsymbol{\varphi})\boldsymbol{C}')^-(\boldsymbol{C}\hat{\boldsymbol{\varphi}}_n)$ with χ_r^2 null distribution where r is the rank of the matrix $\boldsymbol{C}\boldsymbol{i}^-(\boldsymbol{\varphi})\boldsymbol{C}'$ having Moore-Penrose inverse $(\boldsymbol{C}\boldsymbol{i}^-(\boldsymbol{\varphi})\boldsymbol{C}')^-$.

Part I:

1. Proof: It is obvious for $n=1$.

\$ now that $U_i \sim U(0,1)$ for $i = 1, 2, \ldots, n$.

Then $P(U_{n+1} \leq x) = P\left(\left(\sum_{i=1}^{n+1} X_i\right) \bmod 1 \leq x\right)$, $0 \leq x \leq 1$

$$= P\left(\left(\sum_{i=1}^{n} X_i + X_{n+1}\right) \bmod 1 \leq x\right)$$

$$= P\left(\left(\left(\sum_{i=1}^{n} X_i\right) \bmod 1 + X_{n+1}\right) \bmod 1 \leq x\right)$$

$$= P\left(\left(U_n + X_{n+1}\right) \bmod 1 \leq x\right)$$

From the above $U_n \sim U(0,1)$ and also note that $U_n \perp\!\!\!\perp X_{n+1}$.

So it is enough to show that for any two independent identically distributed r.v.s $X, Y \sim U(0,1)$, $(X+Y) \bmod 1 \sim U(0,1)$.

$$P\left((X+Y) \bmod 1 \leq x\right) = P(X+Y \leq x) + P(1 \leq X+Y \leq 1+x)$$

$$= \int_0^x y\,dy + \int_1^{1-x} (2-y)\,dy \text{ , using the density of } X+Y,\ X, Y \overset{iid}{\sim} U_{(0,1)}$$

$$= \frac{y^2}{2}\Big|_0^x + -\frac{(2-y)^2}{2}\Big|_1^{1-x}$$

$$= \frac{x^2}{2} + \frac{1}{2} - \frac{(1-x)^2}{2} = x.$$

QED.

2. We first show that $U_1 \perp\!\!\!\perp U_2$ are independent.

To see this, we have to show that

$$P(U_1 \leq x_1, U_2 \leq x_2) = x_1 x_2 \qquad 0 < x_1, x_2 < 1.$$

Case 1: \$ $x_1 > x_2$. Then

$$P(X_1 \leq x_1, (X_1+X_2) \bmod 1 \leq x_2)$$

$$= \int_0^{x_1} P((X_1+X_2) \bmod 1 \leq x_2 / X_1 = z)\,dz = \int_0^{x_1} P(1-z \leq x_2 \leq 1+x_2-z)\,dz$$

$$= x_1 x_2.$$

Case 2: \$ $x_1 < x_2$, Then

$$P(X_1 \leq x_1, (X_1+X_2) \bmod 1 \leq x_2) = \int_0^{x_1} P(0 \leq X_1+X_2 \leq x_2 / X_1 = z)\,dz$$

$$+ \int^{x_1} P(1 \leq X_1+X_2 \leq 1+x_2 / X_2 = z)\,dz$$

$$= \int_0^{x_1} P(X_2 \le x_2 - z)\, dz + \int_0^{x_1} P(1 - z \le X_2 \le 1)\, dz$$

$$= \int_0^{x_1} (x_2 - z)\, dz + \int_0^{x_1} z\, dz = x_1 x_2 .$$

Thus $U_1 \perp\!\!\!\perp U_2$.

Note that $\mathcal{L}(U_i, U_{i+1}) = \mathcal{L}(U_1, U_2)$ because

$U_{i+1} = (U_i + X_{i+1}) \bmod 1$ and from (1) we have shown that $U_i \sim \mathcal{U}(0,1)$. It follows that $\mathcal{L}(U_i, U_{i+1}) = \mathcal{L}(U_1, U_2)$.

Thus $U_i \perp\!\!\!\perp U_{i+1}$.

To show that $U_i \perp\!\!\!\perp U_j$ for $i < j$, note that

$$P(U_j \le x \mid U_k = u_k,\ k = i, i+1, \dots, j-1)$$

$$= P\big((U_{j-1} + X_j) \bmod 1 \le x \mid U_k = u_k,\ k = i, i+1, \dots, j-1\big)$$

$$= P(U_j \le x \mid U_{j-1} = u_{j-1}) .$$

But since $U_j \perp\!\!\!\perp U_{j-1}$ are independent $P(U_j \le x \mid U_{j-1} = u_{j-1}) = x$.

Therefore $U_j \perp\!\!\!\perp U_i, U_{i+1}, U_{i+2}, \dots U_{j-1}$ and in particular $U_j \perp\!\!\!\perp U_i$.

3. <u>Proof</u>: Let $Y_i = \dfrac{X_i}{\sum_{i=1}^{n} X_i}$, $i=1,2,\dots,n-1$.

$$Y_n = \sum_{i=1}^{n} X_i .$$

Then, $X_i = Y_n Y_i$, $i=1,2,\dots,n-1$.

$$X_n = Y_n - Y_n \sum_{i=1}^{n-1} Y_i = Y_n\left(1-\sum_{i=1}^{n-1} y_i\right)$$

$0 \leq y_n y_i \leq 1$ $\forall i=1,2,\dots,n-1$

& $0 \leq y_n\left(1-\sum_{i=1}^{n-1} y_i\right) \leq 1$

$\Leftrightarrow$ $0 \leq y_n \leq \dfrac{1}{y_i}$ $\forall i=1,2,\dots,n-1$

& $0 \leq y_n \leq \dfrac{1}{1-\sum_{i=1}^{n-1} y_i}$

The Jacobian of the transformation above is given by

$$|J| = \begin{vmatrix} y_n & & & & -y_n \\ & y_n & & 0 & -y_n \\ & & \ddots & & \vdots \\ & 0 & & y_n & -y_n \\ y_1 & y_2 & \cdots & y_{n-1} & \left(1-\sum_{i=1}^{n-1} y_i\right) \end{vmatrix}$$

$$= y_n^{n-1}\left[1-\sum_{i=1}^{n-1} y_i - (y_1, y_2, \dots, y_{n-1})\left(\frac{1}{y_n} I\right)\begin{pmatrix} -y_n \\ -y_n \\ \vdots \\ -y_n \end{pmatrix}\right]$$

$$= y_n^{n-1}\left(1-\sum_{i=1}^{n-1} y_i + \sum_{i=1}^{n-1} y_i\right) = y_n^{n-1}$$

The joint density of $Y_1, Y_2, \dots, Y_n$ is therefore given by

$$f(y_1, y_2, \dots, y_n) = y_n^{n-1}, \quad 0 \leq y_1, y_2, \dots, y_{n-1} \leq 1$$

Integrating out y_n yields: $\quad 0 \leq y_n \leq \min\left(\dfrac{1}{y_1}, \dfrac{1}{y_2}, \dots, \dfrac{1}{y_{n-1}}, \dfrac{1}{1-\sum_{i=1}^{n-1} y_i}\right)$.

$$f(y_1, y_2, \dots, y_{n-1}) = \int_0^{\min\left(\frac{1}{y_1}, \frac{1}{y_2}, \dots, \frac{1}{y_{n-1}}, \frac{1}{1-\sum_{i=1}^{n-1} y_i}\right)} y_n^{n-1}$$

$$= \left\{\min\left(\frac{1}{y_1}, \frac{1}{y_2}, \dots, \frac{1}{y_{n-1}}, \frac{1}{1-\sum_{i=1}^{n-1} y_i}\right)\right\}^n \Big/ n .$$

Part II:

$$Y = \alpha \underset{\sim}{1} + \beta \underset{\sim}{X} + \underset{\sim}{\varepsilon}, \quad \text{where } \underset{\sim}{\varepsilon} \sim N(0, \sigma^2 R),$$

with
$$R = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & & \vdots \\ \rho & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho & \cdots & \cdots & \cdots & 1 \end{pmatrix} = (1-\rho) I + \rho \underset{\sim}{1}\underset{\sim}{1}'$$

By the Sherman-Morrison formula,

$$R^{-1} = \left[(1-\rho) I + \rho \underset{\sim}{1}\underset{\sim}{1}'\right]^{-1} = \frac{1}{(1-\rho)} I - \frac{\rho}{(1-\rho)^2} \underset{\sim}{1}\underset{\sim}{1}' \Big/ \left(1 + \frac{\rho \underset{\sim}{1}'\underset{\sim}{1}}{(1-\rho)}\right)$$

$$= \frac{1}{(1-\rho)} \left[I - \frac{\rho}{1+(n-1)\rho} \underset{\sim}{1}\underset{\sim}{1}'\right].$$

$$|R| = \left|(1-\rho) I + \rho \underset{\sim}{1}\underset{\sim}{1}'\right| = \left(1 + \frac{\rho \underset{\sim}{1}'\underset{\sim}{1}}{(1-\rho)}\right) |(1-\rho) I|$$

$$= (1+(n-1)\rho)(1-\rho)^{n-1}.$$

4. The joint likelihood of $(\alpha, \beta, \sigma, \rho)$ given $\underset{\sim}{Y}$ is

$$\ell(\alpha, \beta, \sigma, \rho; \underset{\sim}{Y}) = \text{const} - n \ln \sigma - \frac{1}{2} \ln |R| - \frac{1}{2} (\underset{\sim}{Y} - \alpha \underset{\sim}{1} - \beta \underset{\sim}{X})' R^{-1} (\underset{\sim}{Y} - \alpha \underset{\sim}{1} - \beta \underset{\sim}{X}).$$

$$\left.\begin{array}{l} \dfrac{\partial \ell}{\partial \beta} = (\underset{\sim}{Y} - \alpha \underset{\sim}{1} - \beta \underset{\sim}{X})' R^{-1} \underset{\sim}{X} = 0 \\ \dfrac{\partial \ell}{\partial \alpha} = (\underset{\sim}{Y} - \alpha \underset{\sim}{1} - \beta \underset{\sim}{X})' R^{-1} \underset{\sim}{1} = 0 \end{array}\right\} \Rightarrow \begin{array}{l} (*)\ \alpha \underset{\sim}{1}' R^{-1} \underset{\sim}{X} + \beta \underset{\sim}{X}' R^{-1} \underset{\sim}{X} = \underset{\sim}{Y}' R^{-1} \underset{\sim}{X} \\ (**)\ \alpha \underset{\sim}{1}' R^{-1} \underset{\sim}{1} + \beta \underset{\sim}{X}' R^{-1} \underset{\sim}{1} = \underset{\sim}{Y}' R^{-1} \underset{\sim}{1} \end{array}$$

multiply both sides of the first equation (*) by $\underset{\sim}{1}' R^{-1} \underset{\sim}{1}$ and the second equation (**) by $\underset{\sim}{1}' R^{-1} \underset{\sim}{X}$ and subtract to get

$$\hat{\beta} = \frac{(\underset{\sim}{Y}' R^{-1} \underset{\sim}{X})(\underset{\sim}{1}' R^{-1} \underset{\sim}{1}) - (\underset{\sim}{Y}' R^{-1} \underset{\sim}{1})(\underset{\sim}{1}' R^{-1} \underset{\sim}{X})}{(\underset{\sim}{X}' R^{-1} \underset{\sim}{X})(\underset{\sim}{1}' R^{-1} \underset{\sim}{1}) - (\underset{\sim}{X}' R^{-1} \underset{\sim}{1})^2}$$

$$\underset{\sim}{y}'R^{-1}\underset{\sim}{x} = \frac{1}{(1-\rho)}\underset{\sim}{y}'\left[I - \frac{\rho\underset{\sim}{1}\underset{\sim}{1}'}{(1+(n-1)\rho)}\right]\underset{\sim}{x}$$

$$= \frac{1}{(1-\rho)}\left[\underset{\sim}{y}'\underset{\sim}{x} - \rho\frac{n^2\bar{y}\bar{x}}{1+(n-1)\rho}\right]$$

$$\underset{\sim}{1}'R^{-1}\underset{\sim}{1} = \frac{n}{(1-\rho)}\left[1 - \frac{n\rho}{1+(n-1)\rho}\right] = \frac{n}{1+(n-1)\rho}.$$

$$\underset{\sim}{x}'R^{-1}\underset{\sim}{x} = \frac{1}{(1-\rho)}\left[\underset{\sim}{x}'\underset{\sim}{x} - \frac{n^2\rho\bar{x}^2}{1+(n-1)\rho}\right].$$

$$\underset{\sim}{y}'R^{-1}\underset{\sim}{1} = \frac{1}{(1-\rho)}\underset{\sim}{y}'\left[I - \frac{\rho\underset{\sim}{1}\underset{\sim}{1}'}{1+(n-1)\rho}\right]\underset{\sim}{1} = \frac{1}{(1-\rho)}\left[n\bar{y} - \frac{n^2\rho\bar{y}}{1+(n-1)\rho}\right]$$

$$= \frac{n\bar{y}}{(1-\rho)}\left[1 - \frac{n\rho}{1+(n-1)\rho}\right] = \frac{n\bar{y}}{1+(n-1)\rho}$$

Similarly, $\underset{\sim}{x}'R^{-1}\underset{\sim}{1} = \dfrac{n\bar{x}}{1+(n-1)\rho}.$

Thus,
$$\hat{\beta} = \frac{\frac{1}{(1-\rho)}\left[\underset{\sim}{y}'\underset{\sim}{x} - \frac{\rho n^2\bar{y}\bar{x}}{1+(n-1)\rho}\right]\cdot\cancel{\frac{n}{1+(n-1)\rho}} - \frac{n^2\bar{x}\bar{y}}{[1+(n-1)\rho]^{\cancel{2}}}}{\frac{1}{(1-\rho)}\left[\underset{\sim}{x}'\underset{\sim}{x} - \frac{\rho n^2\bar{x}^2}{1+(n-1)\rho}\right]\cancel{\frac{n}{1+(n-1)\rho}} - \frac{n^2\bar{x}^2}{(1+(n-1)\rho)^{\cancel{2}}}}$$

$$= \frac{[1+(n-1)\rho]\underset{\sim}{y}'\underset{\sim}{x} - \rho n^2\bar{y}\bar{x} - n(1-\rho)\bar{x}\bar{y}}{[1+(n-1)\rho]\underset{\sim}{x}'\underset{\sim}{x} - \rho n^2\bar{x}^2 - n(1-\rho)\bar{x}^2}$$

$$= \frac{\cancel{[1+(n-1)\rho]}\underset{\sim}{y}'\underset{\sim}{x} - n\cancel{(1+(n-1)\rho)}\bar{x}\bar{y}}{\cancel{[1+(n-1)\rho]}\underset{\sim}{x}'\underset{\sim}{x} - n\cancel{(1+(n-1)\rho)}\bar{x}^2} = \frac{\underset{\sim}{y}'\underset{\sim}{x} - n\bar{x}\bar{y}}{\underset{\sim}{x}'\underset{\sim}{x} - n\bar{x}^2}.$$

From (**), $$\hat{\alpha} = \frac{\underset{\sim}{y}'R^{-1}\underset{\sim}{1} - \hat{\beta}\,\underset{\sim}{x}'R^{-1}\underset{\sim}{1}}{\underset{\sim}{1}'R^{-1}\underset{\sim}{1}} = \frac{n\bar{y} - \hat{\beta}n\bar{x}}{n} = \bar{y} - \hat{\beta}\bar{x}.$$

These are the same estimates as for the case when $\rho=0$, ie, the case of simple linear regression.

5. The joint likelihood of (ρ, σ) is (after maximizing for α, β):

$$\ell(\rho, \sigma ; \underset{\sim}{y}) = -n \ln \sigma - \frac{1}{2}(n-1)\ln(1-\rho) - \frac{1}{2}\ln(1+(n-1)\rho) - \frac{1}{2\sigma^2}\underset{\sim}{e}' R^{-1} \underset{\sim}{e}$$

where $\underset{\sim}{e} = \underset{\sim}{y} - \hat{\underset{\sim}{y}}$.

Note that $\underset{\sim}{e}' R^{-1} \underset{\sim}{e} = (1-\rho)^{-1}\left[\underset{\sim}{e}'\underset{\sim}{e} - \frac{\rho\, \underset{\sim}{e}'\underset{\sim}{1}\ \underset{\sim}{1}'\underset{\sim}{e}}{1+(n-1)\rho}\right] = (1-\rho)^{-1}\underset{\sim}{e}'\underset{\sim}{e}$.

$$= \frac{1}{(1-\rho)}\underset{\sim}{e}'\underset{\sim}{e}.$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \frac{1}{(1-\rho)} \underset{\sim}{e}'\underset{\sim}{e} = 0$$

$$\Rightarrow (1-\hat{\rho})\hat{\sigma}^2 = \frac{1}{n}\underset{\sim}{e}'\underset{\sim}{e} \qquad \text{—} \quad (***)$$

$$\frac{\partial \ell}{\partial \rho} = \frac{n-1}{2(1-\rho)} - \frac{(n-1)}{2(1+(n-1)\rho)} + \frac{1}{2\sigma^2(1-\rho)^2}\underset{\sim}{e}'\underset{\sim}{e} = 0. \quad (****)$$

Now if (***) were to be satisfied, then (****) would not have a solution because the result would not depend on data.

6. $\hat{\beta} = \dfrac{\underset{\sim}{x}'\underset{\sim}{y} - n\bar{x}\bar{y}}{S_{xx}}$ where $S_{xx} = \underset{\sim}{x}'\underset{\sim}{x} - n\bar{x}^2$

$= \dfrac{1}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'\underset{\sim}{y}$.

(a) $E(\hat{\beta}) = \dfrac{1}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'(\alpha\underset{\sim}{1} + \beta\underset{\sim}{x}) = \beta\dfrac{1}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'(\underset{\sim}{x})$

$= \beta\dfrac{1}{S_{xx}}(\underset{\sim}{x}'\underset{\sim}{x} - n\bar{x}^2) = \beta$.

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \dfrac{1}{n}\underset{\sim}{1}'\underset{\sim}{y} - \bar{x}\dfrac{1}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'\underset{\sim}{y}$

$= \left[\dfrac{1}{n}\underset{\sim}{1} - \dfrac{\bar{x}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})}{S_{xx}}\right]'\underset{\sim}{y}$.

(b) $E(\hat{\alpha}) = \left[\dfrac{1}{n}\underset{\sim}{1} - \dfrac{\bar{x}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})}{S_{xx}}\right]'\left[\alpha\underset{\sim}{1} + \beta\underset{\sim}{x}\right]$

$= \alpha\left[\dfrac{1}{n}\underset{\sim}{1}'\underset{\sim}{1} - \dfrac{\bar{x}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'\underset{\sim}{1}}{S_{xx}}\right] + \beta\left[\dfrac{1}{n}\underset{\sim}{1}'\underset{\sim}{x} - \dfrac{\bar{x}}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'\underset{\sim}{x}\right]$

$= \alpha[1 - 0] + \beta[\bar{x} - \bar{x}] = \alpha$.

(c) $Var(\hat{\beta}) = \dfrac{1}{S_{xx}^2}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})' Var(\underset{\sim}{y})(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})$

$= \dfrac{\sigma^2}{S_{xx}^2}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'((1-\rho)I + \rho\underset{\sim}{1}\underset{\sim}{1}')(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})$

$= \dfrac{\sigma^2(1-\rho)(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})}{S_{xx}^2} + \dfrac{\sigma^2}{S_{xx}^2}\underbrace{(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'\underset{\sim}{1}}_{0}\underbrace{\underset{\sim}{1}'(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})}_{0}$

$= \dfrac{\sigma^2(1-\rho)}{S_{xx}}$.

This implies that $Var(\hat{\beta}) \longrightarrow 0$ as $n \to \infty$.
Since $\hat{\beta}$ is unbiased, thus $\hat{\beta}$ is consistent for β.

(d) $\text{Var}(\hat{\alpha}) = \left[\frac{1}{n}\underset{\sim}{1} - \frac{\bar{x}}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})\right]' \text{Var}(\underset{\sim}{y}) \left[\frac{1}{n}\underset{\sim}{1} - \frac{\bar{x}}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})\right]$

$= \sigma^2\left[\frac{1}{n}\underset{\sim}{1} - \frac{\bar{x}}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})\right]' \left[(1-\rho)I + \rho \underset{\sim}{1}\underset{\sim}{1}'\right] \left[\frac{1}{n}\underset{\sim}{1} - \frac{\bar{x}}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})\right]$

$= \sigma^2\left[(1-\rho)\left\{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}^2}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})\right\} + \rho[1 - 0]\right]$

$= \sigma^2(1-\rho)\left[\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right] + \sigma^2\rho.$

Note $\text{Var}(\hat{\alpha}) \longrightarrow \sigma^2\rho > 0$ (unless $\rho \neq 0$).

So $\hat{\alpha}$ is not consistent for α.

7. The joint likelihood of (ρ, σ) is (after maximizing for α, β):

$\ell(\rho, \sigma; \underset{\sim}{y}) = -n \ln \sigma - \frac{1}{2}(n-1)\ln(1-\rho) - \frac{1}{2}\ln(1+(n-1)\rho) - \frac{1}{2\sigma^2(1-\rho)} \underset{\sim}{e}'\cancel{R^{-1}}\underset{\sim}{e}$

$= -\frac{n}{2}\ln[\sigma^2(1-\rho)] + \frac{1}{2}\ln(1-\rho) - \frac{1}{2}\ln(1+(n-1)\rho) - \frac{1}{2\sigma^2(1-\rho)}\underset{\sim}{e}'\underset{\sim}{e}$

$\frac{\partial \ell}{\partial[\sigma^2(1-\rho)]} = \frac{-n}{2[\sigma^2(1-\rho)]} + \frac{1}{2[\sigma^2(1-\rho)]^2} \cdot \underset{\sim}{e}'\underset{\sim}{e} = 0$

$\Rightarrow \widehat{\hat{\sigma}^2(1-\hat{\rho})} = \frac{\underset{\sim}{e}'\underset{\sim}{e}}{n}.$

$\underset{\sim}{e}^* = y - \hat{y} = y - \hat{\alpha}\underset{\sim}{1} - \hat{\beta}\underset{\sim}{x}$

$= y - \underset{\sim}{1}\left[\frac{1}{n}\underset{\sim}{1} - \frac{\bar{x}}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})\right]'y - \underset{\sim}{x}\frac{1}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'y$

$= y - \left[\frac{1}{n}\underset{\sim}{1}\underset{\sim}{1}' - \frac{1}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'\right]y$

$= \left[I - \frac{1}{n}\underset{\sim}{1}\underset{\sim}{1}' - \frac{1}{S_{xx}}(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})(\underset{\sim}{x} - \bar{x}\underset{\sim}{1})'\right]y.$

$= (I - H)\,y$ (where H is the usual hat matrix)

Now $\underset{\sim}{e}$ and $\hat{\beta}$ are both linear transformations of $\underset{\sim}{Y}$ so they are jointly multinormally distributed. (Note that $\underset{\sim}{e}'\underset{\sim}{1}=0$ so $\underset{\sim}{e}$ has a singular distribution.)

The ith residual $e_i = y_i - \hat{y}_i$

$$= j_i'(y - \hat{y}) \quad \text{where } j_i' = (00\ldots0,\underset{i}{1},00\ldots0)$$

$$= j_i'\underset{\sim}{e}$$

$$= j_i'(I-H)y.$$

It remains to show that

$$\mathrm{Cov}(e_i, \hat{\beta}) = 0$$

$$\mathrm{Cov}(e_i, \hat{\beta}) = \mathrm{Cov}\left(j_i'(I-H)y,\ \frac{1}{S_{xx}}(\underset{\sim}{X} - \bar{X}\underset{\sim}{1})'y\right)$$

$$= \frac{j_i'}{S_{xx}}(I-H)\left[\sigma^2(1-\rho)I + \rho\underset{\sim}{1}\underset{\sim}{1}'\right](\underset{\sim}{X} - \bar{X}\underset{\sim}{1})$$

Note that the term involving $\frac{j_i'}{S_{xx}}(I-H)\sigma^2(1-\rho)I(\underset{\sim}{X} - \bar{X}\underset{\sim}{1}) = 0$

since this is the same piece (without $(1-\rho)$) as in linear regression.

Also $\underset{\sim}{1}'(\underset{\sim}{X} - \bar{X}\underset{\sim}{1}) = 0$ so that $\mathrm{Cov}(e_i, \hat{\beta}) = 0$.

8. We have $\sqrt{n}\, i(\varphi)(\hat{\varphi}_n - \varphi) \xrightarrow{d} N(0, i(\varphi))$.

Write $\underset{p\times q}{i(\varphi)} = \underset{p\times q}{\Gamma}\,\underset{q\times q}{\Lambda}\,\underset{q\times p}{\Gamma'}$

Then $\sqrt{n}\, \Lambda^{-1/2}\Gamma' i(\varphi)(\hat{\varphi}_n - \varphi) \xrightarrow{d} N(0, \Lambda^{-1/2}\Gamma' i(\varphi)\Gamma\Lambda^{-1/2})$

$$\equiv N(0, I_q).$$

Let $\varsigma, \hat{\varsigma}_n$ be $\ni$ $\varphi = \Gamma\varsigma$, $\hat{\varphi}_n = \Gamma\hat{\varsigma}_n$. Thus $\varsigma = \Gamma'\varphi$, $\hat{\varsigma}_n = \Gamma'\hat{\varphi}_n$.

Then $\sqrt{n}\, \Lambda^{-1/2}\Gamma' i(\varphi)(\hat{\varphi}_n - \varphi) \equiv \sqrt{n}\, \Lambda^{1/2}(\hat{\varsigma}_n - \varsigma) \xrightarrow{d} N(0, I_q)$

Thus $\sqrt{n}\,(\hat{\varsigma}_n - \varsigma) \xrightarrow{d} N(0, \Lambda^{-1})$.

9. Let $g(\varphi) = C\varphi \equiv C\Gamma\varsigma = h(\varsigma)$. $H_0: C\varphi_0 = 0 \equiv H_0: C\Gamma\varsigma_0 = 0$

By the Taylor series expansion of $g(\hat{\varphi}_n) \equiv h(\hat{\varsigma}_n)$ around $h(\varsigma_0)$

$h(\hat{\varsigma}_n) = h(\varsigma_0) + \frac{\partial}{\partial\varsigma} h(\varsigma_0)(\hat{\varsigma}_n - \varsigma_0) + O_p(\frac{1}{n})$

$\Rightarrow \sqrt{n}\,(h(\hat{\varsigma}_n) - h(\varsigma_0)) = \frac{\partial}{\partial\varsigma} h(\varsigma_0)\sqrt{n}\,(\hat{\varsigma}_n - \varsigma_0)$

Now $\frac{\partial}{\partial\varsigma} h(\varsigma_0) \equiv C\Gamma$; also $\hat{\varphi}_n\ (\hat{\varsigma}_n)$ is consistent for $\varphi\ (\varsigma)$.

Thus $\sqrt{n}\,(C\Gamma\hat{\varsigma}_n) \xrightarrow{d} N(0, C i^{-}(\varphi) C')$

Write the spectral decomposition of $C i^{-}(\varphi) C' = PDP'$ where $r = \text{rank}(C i^{-}(\varphi) C')$

Then $\sqrt{n}\,(D^{-1/2}P'C\hat{\varphi}_n) \xrightarrow{d} N(0, I_r)$

from where it follows that

$$n(C\hat{\varphi}_n)'(C i^{-}(\varphi) C')^{-}(C\hat{\varphi}_n) \longrightarrow \chi^2_r \text{ under } H_0.$$

A large corporation lights its buildings with fluorescent bulbs that glow a random amount of time before burning out. These questions explore the properties and costs of "keeping the lights on."

Part I Problems 1 through 5 establish some facts we will need later.

1. Suppose the lifespan of a bulb in days follows an exponential distribution. If a bulb is expected to last $\frac{1}{\lambda}$ days, write down the pdf for the random lifetime X of a fluorescent bulb.

2. If one building has m iid bulbs, with lifetimes following the above exponential distribution, derive the pdf for the amount of time until the first bulb fails in this building.

3. If $Z_1 \sim \text{Poisson}(\mu)$ and $Z_2 \sim \text{Poisson}(\lambda)$ are independent, show that the distribution of $Z_1 + Z_2$ is $\text{Poisson}(\mu + \lambda)$.

4. Suppose bulbs that fail are immediately replaced with fresh bulbs. Let $Z(t)$ be the number of bulbs failed by time t in the building with m bulbs. It is a well-known fact that the number of bulbs failed in $(0, t]$ for a single socket follows a Poisson distribution with mean λt. Use the fact from Question 3 to conclude that the pmf for $Z(t)$ is

$$p_{Z(t)}(n) = \frac{(m\lambda t)^n e^{-m\lambda t}}{n!}, \qquad n = 0, 1, 2, \ldots. \tag{1}$$

5. Still assuming immediate replacement of failed bulbs, let W_k be the time of the kth bulb failure among the m in the building. It is well known that W_k follows a $\text{Gamma}(k, m\lambda)$ distribution with pdf

$$f_{W_k}(t) = \frac{(m\lambda)^k t^{k-1} e^{-m\lambda t}}{\Gamma(k)}, \qquad t > 0, \tag{2}$$

where $\Gamma(k) = (k-1)!$. Derive the following result for any integer $l < k$.

$$\mathbb{E}\left(\frac{e^{sW_k}}{W_k^l}\right) = \frac{(m\lambda)^k \Gamma(k-l)}{(m\lambda - s)^{k-l}\Gamma(k)}, \quad s < m\lambda.$$

Note that, this equation delivers the moment generating function of a $\text{Gamma}(k, m\lambda)$ distribution if $l = 0$.

Part II To save money, a company decides to wait until the kth bulb failure before replacing burned out bulbs in a building with total m bulbs. Technically, the bulb failure rate decreases from an initial rate $m\lambda$ to $(m-l)\lambda$ after the lth failure, but for answering this question you may assume the failure rate remains constant at $m\lambda$ even as the first k bulbs fail. This assumption is reasonable if the number of bulbs $m >> k$ much exceeds the number of failed bulbs that are tolerated. If it costs c to replace k bulbs, then $\frac{c}{W_k}$ is a measure of the cost rate. Answer the following questions.

6. Without explicitly computing it, provide a lower bound on $\mathbb{E}\left(\frac{c}{W_k}\right)$ as a function of the mean $\mathbb{E}(W_k) = \frac{k}{m\lambda}$.

7. Find the exact expectation $\mathbb{E}\left(\frac{c}{W_k}\right)$ assuming $k > 1$.

8. Suppose that failing to replace a burned out bulb results in an accumulated lost productivity given by $\frac{d}{\beta}\left(1 - e^{-\beta t}\right)$ per bulb t days after bulb burnout, where d and β are known constants. The cost decays over time because the lost productivity is gradually compensated by worker adjustments. Now, the total cost of burned out bulbs per cycle of length W_k is $c + C(W_1, W_2, \ldots, W_k)$, where c is the constant replacement cost and $C(W_1, W_2, \ldots, W_k)$ is the random cost of lost productivity. Find the expectation $\mathbb{E}\left[\frac{c+C(W_1,W_2,\ldots,W_k)}{W_k}\right]$ when $k > 2$.
Hint. If the kth bulb fails at time $W_k = t$, the production costs accumulated due to all k failed bulbs is

$$C(W_1, W_2, \ldots, W_{k-1}, t) \quad = \quad \frac{d}{\beta}\sum_{i=1}^{k}\left(1 - e^{-\beta(t-W_i)}\right).$$

Now, to simplify the expression, use the fact that given $W_k = t$ the failure times, $W_1, W_2, \ldots, W_{k-1}$, are distributed as the order statistics of $k-1$ iid Uniform$[0, t]$ random variables.

Part III The next questions focus on estimation of λ. You may once again assume bulbs are replaced instantaneously when they fail.

9. Show that the moment generating function of $Z(t)$ is given by $M_{Z(t)}(s) = e^{m\lambda t(e^s-1)}$, $s \in \Re$.

10. Show that, for any fixed $s \in \Re$,

$$\log M_{Z(t)}\left(\frac{s}{t}\right) \to m\lambda s \qquad\qquad \text{as } t \to \infty.$$

11. Prove that $\frac{Z(t)}{t} \xrightarrow{\text{p}} m\lambda$ as $t \to \infty$.

Part IV Suppose that the life of a bulb depends on environmental conditions, so that failures do not occur at a constant rate λ. The last questions begin building theory for the distribution of $Z(t)$ in this case. Let $X_i \sim$ Bernoulli(p_i) be independent and define $S_n = X_1 + X_2 + \cdots + X_n$. Similarly, let $Y_i \sim$ Poisson(p_i) be independent and $T_n = Y_1 + Y_2 + \cdots + Y_n$.

12. Consider any two vectors of discrete random variables, $\boldsymbol{X}$ and $\boldsymbol{Y}$ and set A. Prove

$$|P(\boldsymbol{X} \in A) - P(\boldsymbol{Y} \in A)| \leq P(\boldsymbol{X} \neq \boldsymbol{Y}).$$

13. Use the result in Problem 12 to show that

$$|P(S_n = k) - P(T_n = k)| \leq \sum_{i=1}^{n} P(X_i \neq Y_i). \tag{3}$$

14. Find a choice of joint distribution for (X_i, Y_i) such that $P(X_i \neq Y_i) \leq p_i^2$.
Hint. Assuming X_i and Y_i are independent will not work. Instead, try generating $U \sim \text{Unif}(0,1)$, then transforming this U into $X_i \sim \text{Bernoulli}(p_i)$ *and* $Y_i \sim \text{Poisson}(p_i)$. Also, note that $1 - p_i < e^{-p_i}$.

15. Use the work in this Part IV to prove that if $S_n \sim \text{Binomial}(n,p)$, then $S_n \xrightarrow{\text{d}} S \sim \text{Poisson}(\lambda)$ when $\lambda = np$ is held constant while $n \to \infty$.

16. Use the results of this Part IV to draw some conclusions about the distribution of $Z(t)$ when the failure rate λ is not constant in time t. Specifically, suppose $p_i = h\left(\frac{i}{n}t\right)\frac{t}{n}$, $i = 1, 2, \ldots, n$, for a non-negative function $h(\cdot)$, which is Riemann-integrable (*i.e.* $\sum_{i=1}^{n} p_i \to \int_0^t h(x)dx$ as $n \to \infty$). This question is not asking for formal proofs, but imagine approximating the count $Z(t)$ with $\sum_{i=1}^{n} X_i$ (with $X_i \overset{\text{iid}}{\sim} \text{Bernouilli}(p_i)$) as $n \to \infty$. Just sketch a rough argument justifying your conclusions about the distribution of $Z(t)$.

A large corporation lights its buildings with fluorescent bulbs that glow a random amount of time before burning out. These questions explore the properties and costs of "keeping the lights on."

Part I Problems 1 through 5 establish some facts we will need later.

1. Suppose the lifespan of a bulb in days follows an exponential distribution. If a bulb is expected to last $\frac{1}{\lambda}$ days, write down the pdf for the random lifetime X of a fluorescent bulb.

> **Solution:**
> $$f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

2. If one building has m iid bulbs, with lifetimes following the above exponential distribution, derive the pdf for the amount of time until the first bulb fails in this building.

> **Solution:** Let the failure times of the m bulbs be $X_1, \ldots, X_m$. Then, the failure of the first bulb is $Y = \min_j X_j$, and
> $$P(Y > x) = \prod_{j=1}^{m} e^{-\lambda x} = (e^{-\lambda x})^m = e^{-m\lambda x}.$$
> Take derivative of $P(Y \leq x) = 1 - e^{-m\lambda x}$ to get pdf
> $$f_Y(x) = m\lambda e^{-m\lambda x}.$$

3. If $Z_1 \sim \text{Poisson}(\mu)$ and $Z_2 \sim \text{Poisson}(\lambda)$ are independent, show that the distribution of $Z_1 + Z_2$ is Poisson$(\mu + \lambda)$.

> **Solution:** The probability generating function (pgf) for Z_1 is
> $$\begin{aligned} G_{Z_1}(z) &= \sum_{i=0}^{\infty} \frac{z^i \mu^i e^{-\mu}}{i!} = e^{-\mu + z\mu} \sum_{i=0}^{\infty} \frac{(z\mu)^i e^{-z\mu}}{i!} = e^{\mu(z-1)} \\ G_{Z_1+Z_2}(z) &= G_{Z_1}(z) G_{Z_2}(z) = e^{(\mu+\lambda)(z-1)}. \end{aligned}$$
> By uniqueness of pgfs, we conclude $Z_1 + Z_2 \sim \text{Poisson}(\mu + \lambda)$.

4. Suppose bulbs that fail are immediately replaced with fresh bulbs. Let $Z(t)$ be the number of bulbs failed by time t in the building with m bulbs. It is a well-known fact that the number of bulbs failed in $(0, t]$ for a single socket follows a Poisson distribution with mean λt. Use the fact from Question 3 to conclude that the pmf for $Z(t)$ is

$$p_{Z(t)}(n) = \frac{(m\lambda t)^n e^{-m\lambda t}}{n!}, \qquad n = 0, 1, 2, \ldots. \tag{1}$$

Solution: Let $Z_i(t)$ be the number of bulb failures in the ith socket. Combining the statement in Problem 2 that the bulbs are independent and the claim for a single socket given here, we have $Z_i(t) \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda t)$. Because $Z(t) = Z_1(t) + \cdots + Z_m(t)$ and the result of Problem 3 clearly extends to finite sums, we conclude

$$Z(t) \sim \text{Poisson}\,(m\lambda t)\,.$$

5. Still assuming immediate replacement of failed bulbs, let W_k be the time of the kth bulb failure among the m in the building. It is well known that W_k follows a Gamma$(k, m\lambda)$ distribution with pdf

$$f_{W_k}(t) = \frac{(m\lambda)^k t^{k-1} e^{-m\lambda t}}{\Gamma(k)}, \qquad t > 0, \tag{2}$$

where $\Gamma(k) = (k-1)!$. Derive the following result for any integer $l < k$.

$$\mathbb{E}\left(\frac{e^{sW_k}}{W_k^l}\right) = \frac{(m\lambda)^k \Gamma(k-l)}{(m\lambda - s)^{k-l}\Gamma(k)}, \quad s < m\lambda.$$

Note that, this equation delivers the moment generating function of a Gamma$(k, m\lambda)$ distribution if $l = 0$.

Solution: For $k > l$, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{e^{sW_k}}{t^l}\right) &= \int_0^\infty \frac{(m\lambda)^k}{\Gamma(k)} t^{k-l-1} e^{-(m\lambda - s)t} dt \\
&= \frac{(m\lambda)^k \Gamma(k-l)}{(m\lambda - s)^{k-l}\Gamma(k)} \int_0^\infty \frac{(m\lambda - s)^{k-l}}{\Gamma(k-l)} t^{k-l-1} e^{-(m\lambda - s)t} dt \\
&= \frac{(m\lambda)^k \Gamma(k-l)}{(m\lambda - s)^{k-l}\Gamma(k)}.
\end{aligned}$$

Part II To save money, a company decides to wait until the kth bulb failure before replacing burned out bulbs in a building with total m bulbs. Technically, the bulb failure rate decreases from an initial rate $m\lambda$ to $(m - l)\lambda$ after the lth failure, but for answering this question you may assume the failure rate remains constant at $m\lambda$ even as the first k

bulbs fail. This assumption is reasonable if the number of bulbs $m >> k$ much exceeds the number of failed bulbs that are tolerated. If it costs c to replace k bulbs, then $\frac{c}{W_k}$ is a measure of the cost rate. Answer the following questions.

6. Without explicitly computing it, provide a lower bound on $\mathbb{E}\left(\frac{c}{W_k}\right)$ as a function of the mean $\mathbb{E}(W_k) = \frac{k}{m\lambda}$.

Solution: Jensen's inequality on convex function $g(x) = 1/x$ yields expected cost

$$\mathbb{E}\left(\frac{c}{W_k}\right) \quad \geq \quad \frac{c}{\mathbb{E}(W_k)} = \frac{cm\lambda}{k},$$

using the mean of Gamma$(k, m\lambda)$ is $\frac{k}{m\lambda}$.

7. Find the exact expectation $\mathbb{E}\left(\frac{c}{W_k}\right)$ assuming $k > 1$.

Solution:

$$\begin{aligned}
\mathbb{E}\left(\frac{c}{W_k}\right) &= c\int_0^\infty \frac{1}{t}\frac{(m\lambda)^k t^{k-1} e^{-m\lambda t}}{\Gamma(k)} dw \\
&= c\int_0^\infty \frac{(m\lambda)^k t^{k-2} e^{-m\lambda t}}{\Gamma(k)} dt \\
&= \frac{cm\lambda}{k-1}.
\end{aligned}$$

8. Suppose that failing to replace a burned out bulb results in an accumulated lost productivity given by $\frac{d}{\beta}\left(1 - e^{-\beta t}\right)$ per bulb t days after bulb burnout, where d and β are known constants. The cost decays over time because the lost productivity is gradually compensated by worker adjustments. Now, the total cost of burned out bulbs per cycle of length W_k is $c + C(W_1, W_2, \ldots, W_k)$, where c is the constant replacement cost and $C(W_1, W_2, \ldots, W_k)$ is the random cost of lost productivity. Find the expectation $\mathbb{E}\left[\frac{c+C(W_1,W_2,\ldots,W_k)}{W_k}\right]$ when $k > 2$.

Hint. If the kth bulb fails at time $W_k = t$, the production costs accumulated due to all k failed bulbs is

$$C(W_1, W_2, \ldots, W_{k-1}, t) \quad = \quad \frac{d}{\beta}\sum_{i=1}^{k}\left(1 - e^{-\beta(t-W_i)}\right).$$

Now, to simplify the expression, use the fact that given $W_k = t$ the failure times, $W_1, W_2, \ldots, W_{k-1}$, are distributed as the order statistics of $k-1$ iid Uniform$[0, t]$ random variables.

Solution:

If $W_k = t$, then the production costs accumulated during time t due to all failed bulbs are (we'll write $C(t)$ for $C(W_1, W_2, \ldots, t)$)

$$\begin{aligned} C(t) &= \frac{d}{\beta}\sum_{i=1}^{k-1}\left(1 - e^{-\beta(t-W_i)}\right) \\ &= \frac{d(k-1)}{\beta}\left(1 - e^{-\beta U}\right), \end{aligned}$$

for $U \sim \text{Unif}(0,t)$. Since $\mathbb{E}\left(e^{-\beta U}\right) = \frac{e^{-\beta t}-1}{-\beta t}$, we condition on $W_k = t$ to obtain

$$\begin{aligned} \mathbb{E}\left[\frac{C(t)}{t} \mid W_k = t\right] &= \frac{d(k-1)}{\beta t}\left[1 - \frac{1}{\beta t}(1 - e^{-\beta t})\right] \\ &= d(k-1)\left[\frac{1}{\beta t} - \frac{1}{(\beta t)^2} + \frac{e^{-\beta t}}{(\beta t)^2}\right]. \end{aligned}$$

The desired unconditional expectation is obtained after a further expectation with respect to variable W_k and using the work of Question 5.

$$\mathbb{E}\left[\frac{c}{W_k} + \frac{C(W_k)}{W_k}\right] = \frac{cm\lambda}{k-1} + \frac{dm\lambda}{\beta} - \frac{d(m\lambda)^2}{\beta^2(k-2)} + \frac{d(m\lambda)^k}{(m\lambda+\beta)^{k-2}\beta^2(k-2)}.$$

Part III The next questions focus on estimation of λ. You may once again assume bulbs are replaced instantaneously when they fail.

9. Show that the moment generating function of $Z(t)$ is given by $M_{Z(t)}(s) = e^{m\lambda t(e^s-1)}$, $s \in \Re$.

Solution:

$$\begin{aligned} M_{Z(t)}(s) = \mathbb{E}\left[e^{sZ(t)}\right] &= \sum_{i=0}^{\infty} e^{si}\frac{e^{-m\lambda t}(m\lambda t)^i}{i!} \\ &= e^{-m\lambda t + m\lambda t e^s}\sum_{i=0}^{\infty}\frac{e^{-m\lambda t e^s}(m\lambda t e^s)^i}{i!} \\ &= e^{m\lambda t(e^s-1)}. \end{aligned}$$

10. Show that, for any fixed $s \in \Re$,

$$\log M_{Z(t)}\left(\frac{s}{t}\right) \to m\lambda s \qquad\qquad \text{as } t \to \infty.$$

Solution:

$$\begin{aligned}
\lim_{t\to\infty} \log M_{Z(t)}\left(\frac{s}{t}\right) &= m\lambda \lim_{t\to\infty} t(e^{s/t} - 1) \\
&= m\lambda \lim_{t\to\infty} \frac{e^{s/t} - 1}{1/t} \\
&= m\lambda \lim_{t\to\infty} \frac{-se^{s/t}/t^2}{-1/t^2} \\
&= m\lambda s.
\end{aligned}$$

We use L'Hôpital's rule to find the limit above.

11. Prove that $\frac{Z(t)}{t} \xrightarrow{\text{p}} m\lambda$ as $t \to \infty$.

Solution: By definition of the moment generating function,

$$M_{\frac{Z(t)}{t}}(s) = M_{Z(t)}\left(\frac{s}{t}\right),$$

which approaches $e^{m\lambda s}$ as $t \to \infty$ according to the previous problem. Convergence of moment generating functions implies convergence of cdfs (at points of continuity), so

$$\frac{Z(t)}{t} \xrightarrow{\text{d}} Y,$$

where Y is the random variable with mgf $e^{m\lambda s}$, but this is just the mgf of a point mass at $Y = m\lambda$. Thus,

$$\frac{Z(t)}{t} \xrightarrow{\text{d}} m\lambda,$$

but since $m\lambda$ is a constant, we also have

$$\frac{Z(t)}{t} \xrightarrow{\text{p}} m\lambda.$$

Part IV Suppose that the life of a bulb depends on environmental conditions, so that failures do not occur at a constant rate λ. The last questions begin building theory for the distribution of $Z(t)$ in this case. Let $X_i \sim \text{Bernoulli}(p_i)$ be independent and define $S_n = X_1 + X_2 + \cdots + X_n$. Similarly, let $Y_i \sim \text{Poisson}(p_i)$ be independent and $T_n = Y_1 + Y_2 + \cdots + Y_n$.

12. Consider any two vectors of discrete random variables, $\boldsymbol{X}$ and $\boldsymbol{Y}$ and set A. Prove

$$|P(\boldsymbol{X} \in A) - P(\boldsymbol{Y} \in A)| \leq P(\boldsymbol{X} \neq \boldsymbol{Y}).$$

Solution: Events $\{\boldsymbol{X} \in A\}$ and $\{\boldsymbol{Y} \in A\}$ can be partitioned into an unshared and shared parts.

$$\begin{aligned}
\{\boldsymbol{X} \in A\} &= \{\{\boldsymbol{X} \in A\} \cap \{\boldsymbol{X} \neq \boldsymbol{Y}\}\} \cup \{\{\boldsymbol{X} \in A\} \cap \{\boldsymbol{X} = \boldsymbol{Y}\}\} \\
\{\boldsymbol{Y} \in A\} &= \{\{\boldsymbol{Y} \in A\} \cap \{\boldsymbol{Y} \neq \boldsymbol{X}\}\} \cup \{\{\boldsymbol{Y} \in A\} \cap \{\boldsymbol{Y} = \boldsymbol{X}\}\} \\
P(\boldsymbol{X} \in A) - P(\boldsymbol{Y} \in A) &= P(\boldsymbol{X} \in A \cap \boldsymbol{X} \neq \boldsymbol{Y}) - P(\boldsymbol{Y} \in A \cap \boldsymbol{X} \neq \boldsymbol{Y}) \\
&\leq P(\boldsymbol{X} \neq \boldsymbol{Y}).
\end{aligned}$$

In the last step, we know, for any a, b, and c, that $b - a \leq c$ when $0 \leq a \leq c$ and $0 \leq b \leq c$. Similarly $a - b \leq c$. Thus the result is proved.

13. Use the result in Problem 12 to show that

$$|P(S_n = k) - P(T_n = k)| \leq \sum_{i=1}^{n} P(X_i \neq Y_i). \tag{3}$$

Solution: Let $\boldsymbol{X} = (X_1, \ldots, X_n)$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, and A be the event that the sum of the vector elements is k. Then, the result follows from Question 12 and the addition law for probability:

$$|P(S_n = k) - P(T_n = k)| \leq P(\boldsymbol{X} \neq \boldsymbol{Y}) \leq \sum_{i=1}^{n} P(X_i \neq Y_i).$$

14. Find a choice of joint distribution for (X_i, Y_i) such that $P(X_i \neq Y_i) \leq p_i^2$.
Hint. Assuming X_i and Y_i are independent will not work. Instead, try generating $U \sim \text{Unif}(0,1)$, then transforming this U into $X_i \sim \text{Bernoulli}(p_i)$ *and* $Y_i \sim \text{Poisson}(p_i)$. Also, note that $1 - p_i < e^{-p_i}$.

Solution: Let $U \sim \text{Unif}(0,1)$ and define $X_i = 1\{U \leq p_i\}$. Similarly, define $Y_i = F_{Y_i}^{-}(U)$, where $F_{Y_i}(y)$ is the cdf of a Poisson(p_i) and

$$F_{Y_i}^{-}(U) = \inf\{y : F_{Y_i}(y) \geq U\}.$$

The note is true because $1 - p_i < e^{-p_i} = 1 - p_i + p_i^2/2 - p_i^3/3! + \cdots$.

$$\begin{aligned}
P(X_i = Y_i = 0) &= P(Y_i = 0 \mid X_i = 0)P(X_i = 0) = P(Y_i = 0 \mid X_i = 0)(1 - p_i) \\
&= 1 - p_i \\
P(X_i = Y_i = 1) &= P(Y_i = 1 \mid X_i = 1)p_i = \frac{e^{-p_i} + p_i e^{-p_i} - e^{-p_i}}{p_i} p_i = e^{-p_i} \\
P(X_i = Y_i) &= 1 - p_i + e^{-p_1} \\
&> 1 - p_i + 1 - p_i \\
&= 2(1 - p_i) \\
&\geq (1 + p_i)(1 - p_i) = 1 - p_i^2.
\end{aligned}$$

15. Use the work in this Part IV to prove that if $S_n \sim \text{Binomial}(n, p)$, then $S_n \xrightarrow{\text{d}} S \sim \text{Poisson}(\lambda)$ when $\lambda = np$ is held constant while $n \to \infty$.

Solution: Let $p_i = p$ for all n. Then by Question 2, $T_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$ for all n and $T_n \xrightarrow{\text{d}} S$ trivially. Further, note that $p \to 0$ as $n \to \infty$. By constructing the joint distribution of S_n and T_n as in question 14, then we can bound

$$|P(S_n = k) - P(S = k)| = |P(S_n = k) - P(T_n = k)| \leq np^2 = \lambda p \to 0,$$

and conclude that S_n converges to $S \sim \text{Poisson}(\lambda)$ in distribution.

16. Use the results of this Part IV to draw some conclusions about the distribution of $Z(t)$ when the failure rate λ is not constant in time t. Specifically, suppose $p_i = h\left(\frac{i}{n}t\right)\frac{t}{n}, i = 1, 2, \ldots, n$, for a non-negative function $h(\cdot)$, which is Riemann-integrable (*i.e.* $\sum_{i=1}^n p_i \to \int_0^t h(x)dx$ as $n \to \infty$). This question is not asking for formal proofs, but imagine approximating the count $Z(t)$ with $\sum_{i=1}^n X_i$ (with $X_i \overset{\text{iid}}{\sim} \text{Bernouilli}(p_i)$) as $n \to \infty$. Just sketch a rough argument justifying your conclusions about the distribution of $Z(t)$.

Solution: By considering question 15, we see that the original process with constant failure rates can be thought of as the limit of a sum of independent Bernoulli random variables indicating arrivals in tiny discrete time intervals. Specifically, if the probability of failure in time interval of length Δt is $p = \lambda \Delta t$ (using analogous notation), then the number of failures in time t is Binomial with probability $\lambda \Delta t$ and mean $pn = \lambda \Delta t \frac{t}{\Delta t} = \lambda t$. As $\Delta t \to 0$, the Binomial distribution approaches the Poisson(λt).

However, the work of this Part IV shows the Poisson limit applies even when the Bernoulli probabilities are non-constant. Suppose the rate on the ith interval is

$p_i = \lambda_i \Delta t$. Then, the work of this Part shows that the number of failures before time t is still approximately Poisson, so long as the $p_i = \lambda_i \Delta t$ are not too big. Precisely, we have shown that

$$|P(S_n = k) - P(T_n = k)| \leq \sum_{i=1}^{n} p_i^2.$$

In fact, if the rate varies continuously, as $\lambda(t)$, then it appears this approximate Poisson will have mean $\sum_i \lambda_i \Delta t \to \int_0^t \lambda(u) du$ as $\Delta t \to 0$.

Part I

Let X_1 and X_2 be two independent observations from the Poisson distribution that has probability mass function (pmf) given by

$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, x = 0, 1, 2, \cdots,$$

where $\theta > 0$.

1. Let $\tau(\theta) = 1 - e^{-\theta} = P(X_1 > 0)$. Find an unbiased estimator of $\tau(\theta)$ based on X_1.
2. Show that $X_1 + X_2$ is a complete sufficient statistic for θ.
3. Find the uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$ based on (X_1, X_2) and prove that it is UMVU.

Part II

Note: If needed, you may use the following facts without proof:

- The pdf for the Gamma(a, b) distribution is $f(y|a, b) = \frac{1}{\Gamma(a)b^a} y^{a-1} \exp\left(-\frac{y}{b}\right), y > 0$.
- If $W \sim$ Gamma(a, b), then $mW \sim$ Gamma(a, mb).
- If $W \sim$ Gamma$(p/2, 2)$, then $W \sim \chi^2_p$ (chi-square with p degrees of freedom.)

Let $\mathbf{X} = (X_1, \cdots, X_n)$ be an i.i.d. sample with marginal pdf given by

$$f(x|\theta) = \left(\frac{c}{\theta}\right) x^{c-1} \exp\left(-\frac{x^c}{\theta}\right), x > 0,$$

where $c > 0$ is known, and $\theta > 0$ is unknown.

4. Show that $\sum_{i=1}^n X_i^c$ is a sufficient statistic for θ.
5. Show that the family of pdfs $\{f(x|\theta) = \left(\frac{c}{\theta}\right) x^{c-1} \exp\left(-\frac{x^c}{\theta}\right) : \theta > 0\}$ has monotone likelihood ratio.

6. Show that the uniformly most powerful (UMP) test for $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ is given by

$$\text{Reject } H_0 \text{ if } \sum_{i=1}^n X_i^c > k$$

where $0 < k < \infty$ is an appropriate constant. You don't need to give a specific form of k to answer this question.

7. Find the constant k in order to get the UMP test with size α. Express k in terms of a χ^2 quantile.

8. What is the power function of the test you derived in question **7** for $\theta > \theta_0$? Give your answer in terms of a probability statement about a χ^2 distribution.

9. Suppose that we are interested in constructing a $100(1-\alpha)\%$ confidence interval for θ of the form of $(L(\mathbf{X}), \infty)$. Find the lower confidence limit $L(\mathbf{X})$ based on $\sum_{i=1}^n X_i^c$.

Part III

Let $Y_1, \cdots, Y_n$ be an i.i.d. sample with marginal density or probability mass function given by $f(y|\mu)$, where parameter μ is unknown. It is desired to estimate μ with an interval estimator C using the loss function

$$L(\mu, C) = b \cdot (\text{Length}(C)) - \mathbb{I}_C(\mu),$$

where $b > 0$ is a given constant and indicator $\mathbb{I}_C(\mu) = 1$ if $\mu \in C$, and zero otherwise.

10. If μ has prior density function $\pi(\mu)$, assuming that $\pi(\mu|\mathbf{y})$ is unimodal, show that the Bayes rule for producing an interval is given by

$$C = \{\mu : \pi(\mu|\mathbf{y}) \geq b\},$$

where $\pi(\mu|\mathbf{y})$ is the posterior density of μ given $\mathbf{Y} = \mathbf{y}$ (where $\mathbf{Y} = (Y_1, \cdots, Y_n)$).

Note that you may write $\text{Length}(C) = \int_C 1 d\mu$ and use the fact that $P(\mu \in C|\mathbf{Y} = \mathbf{y}) = \int_C \pi(\mu|\mathbf{y}) d\mu$.

11. Show that the interval C given in question **10** depends on $\mathbf{Y}$ only through a sufficient statistic.

12. If you cannot complete questions **10** and **11**, accept the results and continue with this question. Let $Y_1, ..., Y_n$ be an i.i.d. sample from a distribution with pdf given by:

$$f(y) = e^{-(y-\mu)}, y > \mu$$

where $\mu > 0$ is the unknown parameter. Suppose also the prior density for μ is exponential with mean λ^{-1}, that is, $\pi(\mu) = \lambda \exp(-\lambda\mu)$ for $\mu > 0$ and $\pi(\mu) = 0$ otherwise. Suppose $n > \lambda$, and give an explicit description of the Bayes interval estimate in question **10**.

Part I

1. Using the indicator function $\mathbb{I}(\cdot)$, the estimator $Y = \mathbb{I}[X_1 > 0]$ (1 if $X_1 > 0$ and 0 if $X_1 = 0$) is an unbiased estimator of $\tau(\theta)$ as $\tau(\theta) = P(X_1 > 0) = \mathrm{E[Y]}$.

2. The Poisson(θ) model is a regular exponential family and, from this, it can be shown that $S = X_1 + X_2$ is a complete and sufficient statistic.

3. Because Y is unbiased for $\tau(\theta)$ and $S = X_1 + X_2$ is a complete and sufficient statistic, the UMVUE could be derived as

$$\mathrm{E[Y|S = s] = P(X_1 > 0|X_1 + X_2 = s)}$$

.

It can also be shown that $X_1|X_1 + X_2 = s \sim Binomial(s, 1/2)$. Hence

$$P(X_1 > 0|X_1 + X_2 = s) = 1 - P(X_1 = 0|X_1 + X_2 = s) = 1 - (1/2)^s.$$

So, $1 - (1/2)^s$ is UMVUE for $\tau(\theta) = 1 - e^{-\theta} = P(X_1 > 0)$.

Part II

4. The distribution with pdf $f(x|\theta) = \left(\frac{c}{\theta}\right) x^{c-1} \exp\left(-\frac{x^c}{\theta}\right)$ belongs to a one-parameter exponential family with $t(x) = x^c$ and $w(\theta) = -1/\theta$ (using notation for exponential family from the book of Statistical Inference by Casella and Berger). Hence, $\sum_{i=1}^n X_i^c$ is a sufficient statistic for θ.

5. Because $w'(\theta) = 1/(\theta^2) > 0$ for $\theta > 0$, $w(\theta)$ is an increasing function of θ. Then, this one-parameter exponential family has a monotone likelihood ratio (MLR) (See exercise 8.27 of Casella and Berger.) Students may also answer this question using the definition of MLR.

6. Now we know that $T(\mathbf{x}) = \sum_{i=1}^n X_i^c$ is sufficient and this family has a monotone likelihood ratio, we can apply the Karlin-Rubin Theorem for testing $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_0$ by

$$\text{rejecting } H_0 \text{ if } \sum_{i=1}^n X_i^c > k$$

where $0 < k < \infty$ is an appropriate constant.

7. Define $Y_i = X_i^c$. Since Y_i is a monotone transformation of X_i, we have

$$\begin{aligned} f_Y(y|\theta) &= f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| \\ &= \frac{c}{\theta}\left(y^{1/c}\right)^{c-1} \exp\left(-\left(y^{1/c}\right)^c/\theta\right)\left(\frac{1}{c}y^{(1/c-1)}\right) \\ &= \frac{1}{\theta}\exp\left(-\frac{y}{\theta}\right) \end{aligned}$$

Hence, $Y_i \sim$ Exponential(θ), and $T(\mathbf{x}) = \sum_{i=1}^n X_i^c \equiv \sum_{i=1}^n Y_i \sim$ Gamma(n, θ).
Using the hint, a further transformation gives us $\frac{2}{\theta}\sum_{i=1}^n X_i^c \sim$ Gamma$(n, 2)$ which is χ^2_{2n}.
Note that,

$$P\left(\sum_{i=1}^n X_i^c > k|\theta\right) = P\left(\frac{2}{\theta}\sum_{i=1}^n X_i^c > \frac{2}{\theta}k\right) = P\left(\chi^2_{2n} > \frac{2}{\theta}k\right).$$

As $\theta \uparrow$, $\frac{2}{\theta}k \downarrow$, and $P\left(\chi^2_{2n} > \frac{2}{\theta}k\right) \uparrow$.

Hence, the size of the test defined in question 6 is $\sup_{\theta \le \theta_0} P\left(\chi^2_{2n} > \frac{2}{\theta}k\right) = P\left(\chi^2_{2n} > \frac{2}{\theta_0}k\right)$.

To get a test with size α, we select k such that $P\left(\chi^2_{2n} > \frac{2}{\theta}k\right) = \alpha$, or, $\frac{2}{\theta}k = \chi^2_{2n,1-\alpha}$. That is, $k = \frac{\theta_0}{2}\chi^2_{2n,1-\alpha}$.

8. The power of the test is:

$$\beta(\theta) = P\left(\sum_{i=1}^n X_i^c > \frac{\theta_0}{2}\chi^2_{2n,1-\alpha}\right) = P\left(\frac{2}{\theta}\sum_{i=1}^n X_i^c > \frac{2}{\theta}\frac{\theta_0}{2}\chi^2_{2n,1-\alpha}\right) = P\left(\chi^2_{2n} > \frac{\theta_0}{\theta}\chi^2_{2n,1-\alpha}\right).$$

9. Note that $\frac{2}{\theta}\sum_{i=1}^n X_i^c$ is a pivot and $\frac{2}{\theta}\sum_{i=1}^n X_i^c \sim \chi^2_{2n}$. To construct a $100(1-\alpha)\%$ confidence interval of form $L(\mathbf{X}) \le \theta \le \inf$, note that for any constant $a > 0$,

$$\frac{2}{\theta}\sum_{i=1}^n X_i^c \le a \Leftrightarrow \frac{2}{a}\sum_{i=1}^n X_i^c \le \theta.$$

Hence, we choose a such that $P\left(\frac{2}{\theta}\sum_{i=1}^n X_i^c \le a\right) = 1-\alpha$. This suggests that $a = \chi^2_{2n,1-\alpha}$.
So the lower limit of the confidence interval is $L(\mathbf{X}) = \frac{2}{\chi^2_{2n,1-\alpha}}\sum_{i=1}^n X_i^c$.

Part III

10. The Bayes risk is $E\{E[L(\mu, C(\mathbf{Y}))|\mu]\} = E\{E[L(\mu, C(\mathbf{Y}))|\mathbf{Y}]\}$ by Fubini's Theorem.

Hence, the Bayes risk is minimized by choosing $C(\mathbf{Y})$ so that $E[L(\mu, C(\mathbf{Y}))|\mathbf{Y}\text{=}\mathbf{y}]$ is minimized for each $\mathbf{y}$ value. $E[L(\mu, C(\mathbf{Y}))|\mathbf{Y}\text{=}\mathbf{y}]$ is just the posterior expected loss, which is equal to

$$b \cdot (\text{Length of } C) - \text{P}(\mu \in C|\mathbf{Y}\text{=}\mathbf{y}) = b\int_C 1 d\mu - \int_C \pi(\mu|\mathbf{y})d\mu = \int_C (b - \pi(\mu|\mathbf{y}))d\mu,$$

and this is minimized for each y by taking $C = C(\mathbf{y}) = \{\mu : b - \pi(\mu|\mathbf{y}) \le 0\} = \{\mu : \pi(\mu|\mathbf{y}) \ge b\}$.

11. We have

$$\pi(\mu|\mathbf{y}) = \frac{f(\mathbf{y}|\mu)\pi(\mu)}{f(\mathbf{y})} = \frac{g(T(\mathbf{y})|\mu)h(\mathbf{y})\pi(\mu)}{f(\mathbf{y})} = g(T(\mathbf{y})|\mu)\pi(\mu)h(\mathbf{y})/f(\mathbf{y})$$

by the factorization theorem, where $T(\mathbf{y})$ is any sufficient statistic. Hence, the interval C given in question **10** depends on Y only through a sufficient statistic.

12. First note that the sufficient statistic is $T = Y_{(1)}$, the minimum of $Y_1, \cdots, Y_n$, and

$$\begin{aligned}\pi(\mu|\mathbf{y}) &= \frac{f(\mathbf{y}|\mu)\pi(\mu)}{f(\mathbf{y})} \\ &= \lambda \exp(-\lambda\mu) \exp\left(-\sum_{i=1}^{n}(y_i - \mu)\right) \mathbb{I}(0 < \mu \le T)/f(y),\end{aligned}$$

where

$$\begin{aligned}f(\mathrm{Y}) &= \lambda \exp\left(-\sum_{i=1}^{n} y_i\right) \int_0^T \exp\left((n-\lambda)\mu\right) d\mu \\ &= \frac{\lambda}{n-\lambda} \exp\left(-\sum_{i=1}^{n} y_i\right) \left(\exp((n-\lambda)T) - 1\right).\end{aligned}$$

Hence,

$$\pi(\mu|\mathbf{y}) = \pi(\mu|T) = \frac{(n-\lambda)\exp[(n-\lambda)\mu]}{\exp[(n-\lambda)T] - 1}$$

for $0 < \mu \le T$.

Using the Bayes rule given in question **10**, we have the following intervals:

When $n > \lambda$**:** This is the case when $\pi(\mu|\mathbf{y})$ is increasing in μ over $(0, T)$.

If $b < \pi(0|\mathbf{y})$, then $C = (0, T)$. If $b > \pi(T|\mathbf{y})$, then $C = \emptyset$.

If $\pi(0|\mathbf{y}) \le b \le \pi(T|\mathbf{y})$, $C = (L(T), T)$, where

$$L(T) = \frac{1}{n-\lambda} \log\left(\frac{b\left(\exp((n-\lambda)T) - 1\right)}{n-\lambda}\right)$$

is the solution to $\pi(\mu|T) = b$.