## Part I

Suppose $\{X_i\}_{i=1}^n$ is a random sample from an unknown probability distribution with finite mean, variance, and central moments denoted by

$$\mu = \mathrm{E}(X_i), \quad \sigma^2 = \mathrm{E}\{(X_i - \mu)^2\}, \quad \text{and} \quad \mu_3 = \mathrm{E}\{(X_i - \mu)^3\}.$$

Denote the sample mean and variance, respectively, by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2.$$

It is of interest to study the relationship between $\overline{X}$ and $S^2$ for different distributions. It is well known that $\overline{X}$ and $S^2$ are independent if and only if $X_i$'s are normally distributed, which then implies that $\mathrm{Cov}(\overline{X}, S^2) = 0$. Could $\mathrm{Cov}(\overline{X}, S^2)$ be zero for some distributions even though $\overline{X}$ and $S^2$ are NOT independent? We will explore such issues in **Problems 1-4**.

**1.** Let $Y_i = X_i - \mu$, for $i = 1, \ldots, n$, and $S_Y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$. Show that

$$S^2 = S_Y^2, \quad \text{and} \quad \mathrm{Cov}(\overline{X}, S^2) = \mathrm{E}(\overline{Y} \cdot S_Y^2).$$

**2. (General Expression):** Use the identity $\sum_{i=1}^n (Y_i - \overline{Y})^2 = \sum_{i=1}^n Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^n Y_i\right)^2$ and the last identity in **Problem 1** to show that

$$\mathrm{Cov}(\overline{X}, S^2) = \frac{\mu_3}{n}.$$

**3. (Special Case I):** Suppose $X_1, \ldots, X_n$ are iid Bernoulli$(p)$ random variables, where $0 < p < 1$.

(a) Calculate $\mathrm{Cov}(\overline{X}, S^2)$ explicitly, and

- determine the values of $p$ for which $\mathrm{Cov}(\overline{X}, S^2) < 0$,
- determine the values of $p$ for which $\mathrm{Cov}(\overline{X}, S^2) = 0$,
- determine the values of $p$ for which $\mathrm{Cov}(\overline{X}, S^2) > 0$.

(b) Compute $P(S^2 = 0 \mid \overline{X} = 1)$ and $P(S^2 = 0)$.

(c) Show that $\overline{X}$ and $S^2$ are NOT independent (even though the covariance between the two could be zero).

**4. (Special Case II):** Suppose $X_i \overset{iid}{\sim} \mathrm{Poisson}(\theta)$ for $i = 1, \ldots, n$, where $\theta > 0$. Then, show that

$$\mathrm{Cov}(\overline{X}, S^2) = \mathrm{Var}(\overline{X}).$$

## Part II

Suppose $\{X_1, \ldots, X_n\}$ is a random sample from a Uniform$(0, \sigma)$ distribution, for some unknown parameter $\sigma > 0$. Let $\left\{X_{(1)}, \ldots, X_{(n)}\right\}$ denote the corresponding order statistics and define the range $R = X_{(n)} - X_{(1)}$, where $X_{(1)}$ is the smallest order statistic and $X_{(n)}$ is the largest order statistic.

**5.** For fixed $1 \leq i \leq n-1$ and $0 < x_{(1)} < \ldots < x_{(i)} < x_{(i+1)} < \sigma$, show that the conditional pdf of $X_{(i+1)}$ given $X_{(1)} = x_{(1)}, \ldots, X_{(i)} = x_{(i)}$ is the same as the conditional pdf of $X_{(i+1)}$ given $X_{(i)} = x_{(i)}$.

**6.** Find the (marginal) pdf of the range $R$ beginning from the joint pdf of $R$ and $X_{(n)}$.

**7.** Define $\left\{W_n = 2n\left(1 - \frac{R}{\sigma}\right), n \geq 1\right\}$, show that the pdf of $W_n$ is given by

$$h_n(x) = \begin{cases} \frac{n-1}{4n} x \left(1 - \frac{x}{2n}\right)^{n-2} & \text{for } 0 < x < 2n, \\ 0 & \text{otherwise} \end{cases}.$$

**8.** For $h_n(x)$ defined in **Problem 7**, show that there is a pdf $h$ such that

$$\lim_{n \to \infty} h_n(x) = h(x), \quad \text{for } 0 < x < \infty.$$

Let $W$ be a random variable with pdf $h$. Name the distribution of $W$ explicitly.

## Part III

First, consider the case of positive random variables $X_1$ and $X_2$, and define $Z = X_1 - X_2$ and $W = X_1/X_2$. Then, note that the two sets $\{Z = 0\}$ and $\{W = 1\}$ are identical. So, the question arises "Is the conditional pdf of $X_1$ given $Z = 0$ equal to the conditional pdf of $X_1$ given $W = 1$?" In **Problems 9-13** you will consider this issue.

Let $\{X_1, X_2, \ldots, X_n\}$ be iid with pdf

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

**9.** Find the joint pdf of $X_1$ and $Z$. Be sure to specify the region where the joint density is positive.

**10.** Show that the conditional pdf of $X_1$ given $Z = 0$ is

$$f_{X_1|Z=0}(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}.$$

**11.** Find the joint pdf of $X_1$ and $W$ and then show that the marginal pdf of $W$ is

$$f_W(w) = \begin{cases} \frac{1}{(1+w)^2} & \text{for } w > 0, \\ 0 & \text{otherwise.} \end{cases}.$$

**12.** Show that the conditional pdf of $X_1$ given $W = 1$ is

$$f_{X_1|W=1}(x) = \begin{cases} 4xe^{-2x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

## Part IV

Suppose $X_1, X_2, \ldots, X_n$ are iid with an exponential marginal pdf defined in **Part III**.

**14.** Consider the entire sample $\{X_1, X_2, \ldots, X_n\}$. Assume that $n$ is an even integer. Find the limiting distribution of the random variable $Z_n = \sqrt{2/n}\sum_{i=1}^{n}(-1)^{i-1}X_i$.

Some distributional facts that you may use without proof are:

**Fact 1:** If $X$ is an exponential random variable with mean $\lambda$, that is $X \sim$ Exp $(\lambda)$, then the pdf of $X$ is

$$f(x|\lambda) = \begin{cases} \frac{1}{\lambda}\exp\Big\{-\frac{x}{\lambda}\Big\} & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda > 0$. For such $X$, Var$(X) = \lambda^2$.

**Fact 2:** If $X_1$ and $X_2$ are independent random variables with $X_i \sim$ Gamma $(\alpha_i, 1)$, for $\alpha_i > 0$, $i = 1, 2$, then $X_1/[X_1 + X_2] \sim$ Beta $(\alpha_1, \alpha_2)$.

**Fact 3:** If $X$ is an inverse gamma random variable with parameters $(\alpha, \beta)$, that is $X \sim$ IG $(\alpha, \beta)$, then the pdf of $X$ is

$$f(x|\alpha,\beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\exp\Big\{-\frac{\beta}{x}\Big\} & \text{if } x > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha, \beta > 0$. Further, for such $X$, if $\alpha > 1$, E$(X) = \beta/(\alpha - 1)$.

Suppose that $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are iid random vectors where $X_1$ and $Y_1$ are independently distributed with $X_1 \sim$ Exp $(\lambda)$ and $Y_1 \sim$ Exp $(\mu)$. Let $\boldsymbol{\theta} = (\lambda, \mu)$, and $\rho = \lambda/\mu$.

1. Find a two-dimensional sufficient statistic for $\boldsymbol{\theta}$ based on $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.
2. Argue that there is a unique maximizer of the likelihood function of $\boldsymbol{\theta}$ based on these $n$ $(X, Y)$ pairs, call it $\hat{\boldsymbol{\theta}}_n$. Find $\hat{\boldsymbol{\theta}}_n$.
3. Find the asymptotic (bivariate) normal distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ as $n \to \infty$.
4. Show that $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$.
5. Find the MLE of $\rho$, call it $\hat{\rho}_n$.
6. Using the result of **Problem 3**, find the limiting distribution of $\sqrt{n}(\hat{\rho}_n - \rho)$ as $n \to \infty$.
7. Show that $Q((X_1, \ldots, X_n, Y_1, \ldots, Y_n), \boldsymbol{\theta}) = (\sum_{i=1}^n X_i/\lambda) + (\sum_{i=1}^n Y_i/\mu)$ is a pivotal quantity.
8. Construct a confidence set for $\boldsymbol{\theta}$ with confidence coefficient $(1-\alpha)$ using the pivotal quantity from **Problem 7**.

**9**. Let $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $\bar{Y} = \sum_{i=1}^{n} Y_i/n$. Show that the likelihood ratio test (LRT) statistic for testing $H_0 : \lambda = \mu$ against $H_1 : \lambda \neq \mu$ can be expressed in terms of the statistic

$$R = \frac{\bar{X}}{\bar{X} + \bar{Y}} \, .$$

**10**. Find the distribution of $R$ when $H_0$ is true, where $H_0$ and $R$ are defined in **Problem 9**.

**11**. Derive the size $\alpha$ LRT for testing $H_0 : \lambda = \mu$ against $H_1 : \lambda \neq \mu$. (Hint: Using the result of **Problem 10**, you can express this size $\alpha$ LRT in terms of the $(\alpha/2)$th quantile of the distribution of $R$ under $H_0$.)

For **Problems 12-14**, assume that the prior density of $\boldsymbol{\theta}$ is $\pi(\boldsymbol{\theta}) = \pi_1(\lambda)\pi_2(\mu)$, where $\pi_1(\lambda)$ is the IG $(\alpha_\lambda, \beta_\lambda)$ density, and $\pi_2(\mu)$ is the IG $(\alpha_\mu, \beta_\mu)$ density for known positive values of $\alpha_\lambda, \beta_\lambda, \alpha_\mu$ and $\beta_\mu$.

**12**. Derive the posterior density of $\boldsymbol{\theta}$.

**13**. Is the family of prior densities $\{\pi(\boldsymbol{\theta}) : \text{all } \alpha_\lambda, \beta_\lambda, \alpha_\mu, \beta_\mu > 0\}$ conjugate in this model? Explain.

**14**. Let $\mathbf{t} = (t_1, t_2)$. Derive the Bayes estimator of $\boldsymbol{\theta}$ under the loss function

$$L(\boldsymbol{\theta}, \mathbf{t}) = \frac{(t_1 - \lambda)^2}{\lambda} + \frac{(t_2 - \mu)^2}{\mu}.$$

For **Problems 15-18**, assume that we observe only $\{(Z_i, \Delta_i), i = 1, \ldots, n\}$ where $Z_i = \min(X_i, Y_i)$, $\Delta_i = 1$ if $Z_i = X_i$ and $\Delta_i = 0$ if $Z_i = Y_i$, for $i = 1, \ldots, n$.

**15**. Write down the likelihood function of $\boldsymbol{\theta} = (\lambda, \mu)$ based on the observations $\{(Z_i, \Delta_i), i = 1, \ldots, n\}$.

**16**. If $0 < \sum_{i=1}^{n} \Delta_i < n$, find the MLE of $\boldsymbol{\theta} = (\lambda, \mu)$, call it $\tilde{\boldsymbol{\theta}}_n = (\tilde{\lambda}_n, \tilde{\mu}_n)$.

**17**. Show that $\tilde{\boldsymbol{\theta}}_n = (\tilde{\lambda}_n, \tilde{\mu}_n)$ is a consistent estimator of $\boldsymbol{\theta} = (\lambda, \mu)$.

**18**. Does the MLE of $(\lambda, \mu)$ exist

(a) when $\sum_{i=1}^{n} \Delta_i = 0$?

(b) when $\sum_{i=1}^{n} \Delta_i = n$?

Explain.

**This question set is grouped into four parts:**

**Part I** concerns properties of measures/distributions.

**Part II** involves strong/weak laws of large numbers and CLT.

**Parts III & IV** often address modes of convergence for random variables (e.g., almost surely or in distribution) and convergence of expectations/integrals (e.g., DCT or uniform integrability).

**One question in Part III requires use of the following:**

**Berry-Esseen theorem** for iid bounded variables: If $W_1, \cdots, W_m$ are iid bounded random variables with mean $\mu_W$ and variance $\sigma_W^2 > 0$ such that $P(|W_1| \le C) = 1$ for some $C > 0$, then

$$\sup_{z\in\mathbb{R}} \left| P\left( \frac{\sum_{i=1}^m (W_i - \mu_W)}{\sqrt{n\sigma_W^2}} \le z \right) - \Phi(z) \right| \le 5.5 \frac{1}{\sqrt{m}} \frac{C}{\sqrt{\sigma_W^2}} \text{ holds for any integer } m \ge 1,$$

where $\Phi(\cdot)$ denotes the standard normal cdf. *You will be prompted on when to consider this.*

## Part I

Let $X_1$ be a positive Exponential($\theta$) random variable with cdf

$$F_\theta(t) \equiv P(X_1 \le t) = 1 - e^{-t/\theta}, \quad t > 0,$$

depending on a parameter $\theta > 0$. Suppose further that $X_1$ is observable only up to a given point $t_c > 0$ (a right censoring time) so that the available observation based on $X_1$ is given by $A_1$, where

$$A_1 \equiv \begin{cases} X_1 & \text{if } X_1 \le t_c, \\ t_c & \text{if } X_1 > t_c. \end{cases}$$

**1**. Find the cdf $F_{A_1}(t)$ of $A_1$ for $t > 0$.

**2**. Noting that $A_1$ only assumes values in the continuous range $(0, t_c]$, should we classify the random variable $A_1$ as continuous? Briefly explain.

**3**. Show that $P(A_1 \in D) = \mu(D)$ holds for any $D \in \mathcal{B}(\mathbb{R})$ in the Borel $\sigma$-field $\mathcal{B}(\mathbb{R})$ of $\mathbb{R}$, where $\mu$ is a probability measure defined as

$$\mu(D) \equiv \int_0^{t_c} I(x \in D)\theta^{-1} e^{-x/\theta} dx + e^{-t_c/\theta} I(t_c \in D), \quad D \in \mathcal{B}(\mathbb{R}),$$

based on the indicator function $I(\cdot)$.

*Hint:* Show $F_{A_1}(t) = \mu((-\infty, t])$ for $t \in \mathbb{R}$ and explain why this fact suffices.

## Part II

Based on the set-up in **Part I**, let $X_1, X_2, \ldots$ denote an iid sequence of positive Exponential($\theta$) random variables on a common probability space $(\Omega, \mathcal{F}, P)$, having marginal cdf

$$F_\theta(t) \equiv P(X_1 \le t) = 1 - e^{-t/\theta}, \quad t > 0.$$

Suppose $X_1, \ldots, X_n$ represent the failure times of $n$ products where, due to practical constraints, these failure times are only directly observed up to a fixed time point $t_c > 0$. Hence, for any $n \ge 1$, the available observations based on $X_1, \ldots, X_n$ are given by

$$A_i \equiv \begin{cases} X_i & \text{if } X_i \le t_c, \\ t_c & \text{if } X_i > t_c, \end{cases} \qquad i = 1, \ldots, n.$$

In addition to observations $A_1, \ldots, A_n$, further define an observable count

$$R_n = \sum_{i=1}^{n} I(X_i \le t_c)$$

of the number of failures among $X_1, \ldots, X_n$ occurring up to time $t_c$ (where $I(\cdot)$ is the indicator function). Note, among $X_1, \ldots, X_n$, that $n - R_n$ variables exceed $t_c$, though their exact values are unobserved.

**4**. Show that

$$\frac{n - R_n}{n} \to e^{-t_c/\theta} \qquad \text{almost surely (a.s.)} \;\; \text{as } n \to \infty,$$

stating any standard results used.

**5**. Using the fact that $\mathrm{E}(A_1) = \theta F_\theta(t_c)$, show that

$$\hat{\theta}_n \equiv \frac{n}{R_n} \cdot \frac{1}{n} \sum_{i=1}^{n} A_i \to \theta \qquad \text{(a.s.)} \quad \text{as } n \to \infty$$

(the estimator $\hat{\theta}_n$ of $\theta$ is strongly consistent), stating any standard results used.

Based on a further fixed time $t_w > t_c$, next define a parametric function

$$p \equiv p(\theta) = 1 - e^{-(t_w - t_c)/\theta}$$

and an associated estimator $\hat{p}_n \equiv p(\hat{\theta}_n) \equiv 1 - e^{-(t_w - t_c)/\hat{\theta}_n}$ based on $\hat{\theta}_n$ from **Problem 5**.

**6**. Using the fact that $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in distribution to a standard normal $Z_0 \sim N(0,1)$ variable scaled by $\theta[F_\theta(t_c)]^{-1/2}$ (i.e., $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \theta[F_\theta(t_c)]^{-1/2} Z_0$ as $n \to \infty$), prove that

$$\sqrt{\frac{n - R_n}{n}} \cdot \frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} \sigma Z_0 \quad \text{as } n \to \infty, \quad \textbf{and} \quad \text{give an expression for } \sigma > 0,$$

stating any standard results used. (*No algebraic simplifications of $\sigma$ are needed.*)

## Part III

Based on the available observations $A_1, \dots, A_n$ (for $n > 1$) and the number $R_n \equiv \sum_{i=1}^n I(X_i \le t_c)$ of failures (up to time $t_c$) from **Part II**, consider predicting the number of failures, denoted as

$$Y_n \equiv \sum_{i=1}^n I(t_c < X_i \le t_w),$$

among $X_1, \dots, X_n$ occuring in a future interval $(t_c, t_w]$ for some fixed $t_w > t_c$.

**7**. Briefly explain why (without formal proof) the joint distribution of variables $(R_n, Y_n)$, from $X_1, \dots, X_n$, has mass function:

$$P(Y_n = y, R_n = r) = \frac{n!}{y!r!(n-y-r)!}[F_\theta(t_c)]^r[F_\theta(t_w) - F_\theta(t_c)]^y[1 - F_\theta(t_w)]^{n-r-y},$$

for $0 \le y, r \le n$ and $y + r \le n$.

**8**. Argue that the conditional distribution of $Y_n$ given $R_n = r$, for fixed integer $0 \le r < n$, is the same as the distribution of a sum $\sum_{i=1}^{n-r} B_i$ of iid Bernoulli$(p)$ random variables $B_1, \dots, B_{n-r}$ for $p \equiv 1 - e^{-(t_w - t_c)/\theta}$, stating any standard results used.

Using $R_n$ and the (conditional) probability $p \equiv 1 - e^{-(t_w - t_c)/\theta}$ from **Problem 8**, define an "upper prediction bound" $\widetilde{Y}_{\alpha,n}(p)$ for $Y_n$ as

$$\widetilde{Y}_{\alpha,n}(p) \equiv \Phi^{-1}(\alpha)\sqrt{(n - R_n)p(1-p)} + (n - R_n)p,$$

based on a fixed $\alpha \in (0,1)$ (a confidence level) and associated standard normal percentile given by $\Phi^{-1}(\alpha)$ (i.e., $\Phi(\Phi^{-1}(\alpha)) = \alpha$). (Technically, $\widetilde{Y}_{\alpha,n}(p)$ is not a "statistic" as this quantity involves $p$.)

**9**. By applying the **Berry-Esseen theorem (page 1)** and **Problem 8**, show that

$$\left|P\left(Y_n \le \widetilde{Y}_{\alpha,n}(p)\middle|R_n = r\right) - \alpha\right| \le 5.5\frac{1}{n-r}\frac{1}{\sqrt{p(1-p)}} \quad \text{holds for any integer } 0 \le r < n.$$

*Hint:* $P(Y_n \le \widetilde{Y}_{\alpha,n}(p)\Big|R_n = r)$ equals $P\big(\sum_{i=1}^{n-r} B_i \le \Phi^{-1}(\alpha)\sqrt{(n-r)p(1-p)} + (n-r)p\big)$.

**10**. Based on **Problem 9**, show that for any possible value of $R_n \in \{0, \dots, n\}$:

$$\left|P\left(Y_n \le \widetilde{Y}_{\alpha,n}(p)\middle|R_n\right) - \alpha\right| \le 2I(R_n = n) + 5.5\frac{1}{n - R_n}\frac{1}{\sqrt{p(1-p)}}I(0 \le R_n < n),$$

where $I(\cdot)$ denotes the indicator function.

*Hint:* The bound in **Problem 9** does not apply when $R_n = n$, where then $Y_n = 0$ degenerately.

**11**. Based on **Problems 4** and **10**, show that

$$P\left(Y_n \le \widetilde{Y}_{\alpha,n}(p)\middle|R_n\right) \to \alpha \quad \text{(a.s.)} \quad \text{as } n \to \infty.$$

**12**. Stating any standard results used, prove that

$$P\big(Y_n \le \widetilde{Y}_{\alpha,n}(p)\big) = \mathrm{E}\left[P\Big(Y_n \le \widetilde{Y}_{\alpha,n}(p)\Big|R_n\Big)\right] \to \alpha \quad \text{as } n \to \infty.$$

**Part IV**

While **Part III** showed that $\widetilde{Y}_{\alpha,n}(p) \equiv \Phi^{-1}(\alpha)\sqrt{(n-R_n)p(1-p)} + (n-R_n)p$ from **Problems 9-12** provides an $100\alpha\%$ upper prediction bound for $Y_n$ with asymptotically correct coverage of $\alpha$, note that $\widetilde{Y}_{\alpha,n}(p)$ depends on an unknown parameter $p \equiv 1 - e^{-(t_w - t_c)/\theta}$. Based on the available data $A_1, \ldots, A_n$ and count $R_n \equiv \sum_{i=1}^n I(X_i \leq t_c)$, a natural approach to producing a useful prediction bound may be to plug the estimator (actually MLE) $\hat{p}_n = 1 - e^{-(t_w-t_c)/\hat{\theta}_n}$ of $p = 1 - e^{-(t_w-t_c)/\theta}$ from **Problem 6** into the form $\widetilde{Y}_{\alpha,n}(p)$. So we re-define an approximate $100\alpha\%$ upper prediction bound for $Y_n$ as

$$\widetilde{Y}_{\alpha,n}(\hat{p}_n) \equiv \Phi^{-1}(\alpha)\sqrt{(n-R_n)\hat{p}_n(1-\hat{p}_n)} + (n-R_n)\hat{p}_n.$$

**Fact:** For this prediction bound $\widetilde{Y}_{\alpha,n}(\hat{p}_n)$ for $Y_n$, it turns out that the conditional probability

$$CP_n \equiv P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(\hat{p}_n) \middle| R_n, A_1, \ldots, A_n\right)$$

satisfies

$$|CP_n - \Phi_n| \xrightarrow{p} 0 \quad \text{as } n \to \infty,$$

for the quantity

$$\Phi_n \equiv \Phi\left(\Phi^{-1}(\alpha) \cdot \frac{\sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{p(1-p)}} + \sqrt{\frac{n-R_n}{n}}\,\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}}\right).$$

*You may use this fact may in the following, where the notation $CP_n$ and $\Phi_n$, $n \geq 1$, also appears.*

**13**. Based on **Problem 6** and the normal limit $\sigma Z_0$ appearing there, carefully explain why

$$\Phi_n \xrightarrow{d} \Phi(\Phi^{-1}(\alpha) + \sigma Z_0) \quad \text{as } n \to \infty,$$

for the quantity $\Phi_n$ from the **Fact** above, stating any standard results used.

**14**. Show that $CP_n \xrightarrow{d} \Phi(\Phi^{-1}(\alpha) + \sigma Z_0)$ as $n \to \infty$, stating any standard results used.

**15**. Explain why the conditional probability variables $CP_n$, $n \geq 1$, are uniformly integrable.

**16**. Prove that

$$P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(\hat{p}_n)\right) = \mathrm{E}\left[CP_n\right] \to \int_{-\infty}^{\infty} \Phi(\Phi^{-1}(\alpha) + \sigma z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \quad \text{as } n \to \infty,$$

stating any standard results used.

**17**. When substituting an estimator $\hat{p}_n$ of $p$ into the bound form $\widetilde{Y}_{\alpha,n}(p)$ from **Problems 9-12**, the resulting approximate $100\alpha\%$ upper prediction bound $\widetilde{Y}_{\alpha,n}(\hat{p}_n)$ for $Y_n$ can be seen to exhibit a troubling issue not shared by $\widetilde{Y}_{\alpha,n}(p)$. Briefly explain this issue.

## Part I

Suppose $\{X_i\}_{i=1}^n$ is a random sample from an unknown probability distribution with finite mean, variance, and their central moments denoted by

$$\mu = \mathrm{E}(X_i), \quad \sigma^2 = \mathrm{E}\{(X_i-\mu)^2\}, \quad \text{and} \quad \mu_3 = \mathrm{E}\{(X_i-\mu)^3\}.$$

Denote the sample mean and variance, respectively, by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2.$$

It is of interest to study the relationship between $\overline{X}$ and $S^2$ for different distributions. It is well known that $\overline{X}$ and $S^2$ are independent if and only if $X_i$'s are normally distributed, which then implies that $\mathrm{Cov}(\overline{X}, S^2) = 0$. In general, what is the covariance between $\overline{X}$ and $S^2$, $\mathrm{Cov}(\overline{X}, S^2)$? Could $\mathrm{Cov}(\overline{X}, S^2)$ be zero for some distributions even though $\overline{X}$ and $S^2$ are NOT independent? We will explore answers to some of these questions below.

**1.** Let $Y_i = X_i - \mu$, for $i = 1, \ldots, n$, and $S_Y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$, show that

$$S^2 = S_Y^2, \quad \text{and} \quad \mathrm{Cov}(\overline{X}, S^2) = \mathrm{E}(\overline{Y}\cdot S_Y^2).$$

(1) Proof of $S^2 = S_Y^2$:

$$\begin{aligned} S_Y^2 &= \frac{1}{n-1}\sum_{i=1}^n (Y_i-\overline{Y})^2 = \frac{1}{n-1}\sum_{i=1}^n \left\{X_i - \mu - \frac{\sum_{i=1}^n (X_i-\mu)}{n}\right\}^2 \\ &= \frac{1}{n-1}\sum_{i=1}^n \left\{X_i - \mu - \frac{\sum_{i=1}^n X_i - n\mu}{n}\right\}^2 = \frac{1}{n-1}\sum_{i=1}^n \{X_i - \mu - (\overline{X}-\mu)\}^2 \\ &= \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = S^2. \end{aligned}$$

(2) Proof of $\mathrm{Cov}(\overline{X}, S^2) = \mathrm{E}(\overline{Y}\cdot S_Y^2)$:
Notice that $\mathrm{E}(\overline{X}) = \mu,\ \mathrm{E}(S^2) = \mathrm{E}(S_Y^2) = \sigma^2$, one has

$$\begin{aligned} \mathrm{Cov}\left(\overline{X}, S^2\right) &= \mathrm{Cov}\left(\overline{X}, S_Y^2\right) = \mathrm{E}\left[\{\overline{X} - \mathrm{E}(\overline{X})\}\{S_Y^2 - \mathrm{E}(S_Y^2)\}\right] \\ &= \mathrm{E}\{(\overline{X}-\mu)(S_Y^2-\sigma^2)\} = \mathrm{E}(\overline{X}S_Y^2 - \mu S_Y^2 - \sigma^2\overline{X} + \mu\sigma^2) \\ &= \mathrm{E}(\overline{X}S_Y^2) - \mu\mathrm{E}(S_Y^2) - \sigma^2\mathrm{E}(\overline{X}) + \mu\sigma^2 = \mathrm{E}(\overline{X}S_Y^2) - \mu\sigma^2 \\ &= \mathrm{E}\{(\overline{Y}+\mu)S_Y^2\} - \mu\sigma^2 = \mathrm{E}(\overline{Y}S_Y^2) + \mu\mathrm{E}(S_Y^2) - \mu\sigma^2 = \mathrm{E}(\overline{Y}S_Y^2). \end{aligned}$$

2. **General Expression:** Use the identity $\sum_{i=1}^n (Y_i - \overline{Y})^2 = \sum_{i=1}^n Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^n Y_i\right)^2$ and the last identity in **Problem 1** to show that

$$\mathrm{Cov}(\overline{X}, S^2) = \frac{\mu_3}{n}.$$

*Proof.*

$$\begin{aligned}
\mathrm{Cov}(\overline{X}, S^2) &= \mathrm{E}(\overline{Y} S_Y^2) = \frac{1}{n(n-1)}\mathrm{E}\left[\sum_{i=1}^n Y_i \left\{\sum_{i=1}^n Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^n Y_i\right)^2\right\}\right] \\
&= \frac{1}{n(n-1)}\left[\mathrm{E}\left(\sum_{i=1}^n Y_i \sum_{j=1}^n Y_j^2\right) - \frac{1}{n}\mathrm{E}\left\{\sum_{i=1}^n Y_i \left(\sum_{j=1}^n Y_j\right)^2\right\}\right] \\
&= \frac{1}{n(n-1)}(I_1 - I_2), \qquad (1)
\end{aligned}$$

where

$$I_1 = \mathrm{E}\left(\sum_{i=1}^n Y_i \sum_{j=1}^n Y_j^2\right) = \mathrm{E}\left(\sum_{i=1}^n Y_i^3\right) = n\mu_3,$$

and

$$\begin{aligned}
I_2 &= \frac{1}{n}\left\{\sum_{i=1}^n Y_i \left(\sum_{j=1}^n Y_j\right)^2\right\} = \frac{1}{n}\mathrm{E}\left\{\sum_{i=1}^n Y_i \left(\sum_{j=1} Y_j^2 + 2\sum_{j=1}^n \sum_{k=j+1}^n Y_j Y_k\right)\right\} \\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^n Y_i^3\right) = \mu_3.
\end{aligned}$$

Substituting $I_1$ and $I_2$ into Equation (1) completes the proof. □

3. **(Special Case I):** Suppose $X_1, \ldots, X_n$ are iid Bernoulli($p$) random variables, where $0 < p < 1$.

   (a) Calculate $\mathrm{Cov}(\overline{X}, S^2)$ explicitly, and
   - determine the values of $p$ for which $\mathrm{Cov}(\overline{X}, S^2) < 0$;
   - determine the values of $p$ for which $\mathrm{Cov}(\overline{X}, S^2) = 0$;
   - determine the values of $p$ for which $\mathrm{Cov}(\overline{X}, S^2) > 0$.

   *Proof.* Since $X_i \overset{iid}{\sim}$ Bernoulli($p$), $\mu_3 = p(1-p)(1-2p)$, so that one has

$$\mathrm{Cov}(\overline{X}, S^2) = \frac{p(1-p)(1-2p)}{n}. \qquad (2)$$

   Thus, one can see

- $\text{Cov}(\overline{X}, S^2) > 0$, if $p < \frac{1}{2}$;
- $\text{Cov}(\overline{X}, S^2) = 0$, if $p = \frac{1}{2}$;
- $\text{Cov}(\overline{X}, S^2) < 0$, if $p > \frac{1}{2}$.

$\square$

(b) Compute $P(S^2 = 0 \mid \overline{X} = 1)$ and $P(S^2 = 0)$ explicitly.

*Proof.*

$$\begin{aligned} \overline{X} = 1 &\Rightarrow X_1 = X_2 = \cdots = X_n = 1 \Rightarrow S^2 = 0 \\ &\Rightarrow P(S^2 = 0 \mid \overline{X} = 1) = 1, \end{aligned}$$

and

$$\begin{aligned} S^2 = 0 &\Rightarrow X_1 = X_2 = \cdots = X_n = 0 \text{ OR } X_1 = X_2 = \cdots = X_n = 1 \\ &\Rightarrow P(X_1 = \cdots = X_n = 0) = (1-p)^n,\ P(X_1 = \cdots = X_n = 1) = p^n \\ &\Rightarrow P(S^2 = 0) = p^n + (1-p)^n. \end{aligned}$$

$\square$

(c) Show that $\overline{X}$ and $S^2$ are NOT independent (even though the covariance between the two could be zero).

*Proof.* It has been shown in part (a) that when $p = \frac{1}{2}$, $\text{Cov}(\overline{X}, S^2) = 0$. However, when $p = \frac{1}{2}$, $P(S^2 = 0) = \left(\frac{1}{2}\right)^{n-1} \neq P(S^2 = 0 \mid \overline{X} = 1)$. So $\overline{X}$ and $S^2$ are NOT independent even though the covariance between the two could be zero. $\square$

**4. (Special Case II):** Suppose $X_i \overset{iid}{\sim} \text{Poisson}(\theta)$ for $i = 1, \ldots, n$, where $\theta > 0$. Then, show that

$$\text{Cov}(\overline{X}, S^2) = \text{Var}(\overline{X}).$$

*Proof.* $X_i \overset{iid}{\sim} \text{Poisson}(\theta) \Rightarrow \text{E}(X_i) = \theta,\ \text{E}(X_i^2) = \theta^2 + \theta,\ \text{E}(X_i^3) = \theta^3 + 3\theta^2 + \theta,\ \text{Var}(X_i) = \text{E}(X_i^2) - \{\text{E}(X_i)\}^2 = \theta,\ \mu_3 = \text{E}(X_i^3 - 3\theta X_i^2 + 3\theta^2 X_i - \theta^3) = \theta$. Thus, $\text{Cov}(\overline{X}, S^2) = n^{-1}\mu_3 = n^{-1}\theta$. On the other hand, $\text{Var}(\overline{X}) = \frac{1}{n}\text{Var}(X_i) = n^{-1}\theta = \text{Cov}(\overline{X}, S^2)$. $\square$

## Part II

Suppose $\{X_1, \ldots, X_n\}$ is a random sample from Uniform$(0, \sigma)$, for some unknown parameter $\sigma > 0$. Let $\left\{X_{(1)}, \ldots, X_{(n)}\right\}$ denote the order statistics and define the range $R = X_{(n)} - X_{(1)}$, where $X_{(1)}$ is the smallest order statistic and $X_{(n)}$ is the largest order statistic.

**5.** For fixed $1 \le i \le n-1$ and $0 < x_{(1)} < \ldots < x_{(i)} < x_{(i+1)} < \sigma$, show that the conditional probability density function (p.d.f) of $X_{(i+1)}$ given $X_{(1)} = x_{(1)}, \ldots, X_{(i)} = x_{(i)}$ is the same as the conditional pdf of $X_{(i+1)}$ given $X_{(i)} = x_{(i)}$.

*Proof.* Since $X_i \overset{iid}{\sim} \text{Uniform}(\sigma)$, one has

$$f(x) = f_{X_i}(x) = \begin{cases} \frac{1}{\sigma}, & 0 < x < \sigma \\ 0, & \text{otherwise} \end{cases}, \quad \text{and} \quad F(x) = F_{X_i}(x) = \begin{cases} 0, & x \le 0 \\ \frac{x}{\sigma}, & 0 < x < \sigma \\ 1, & x \ge \sigma \end{cases}.$$

The joint density of $X_{(i)}$ and $X_{(i+1)}$ is

$$f_{X_{(i)},X_{(i+1)}}(x_{(i)}, x_{(i+1)}) = \frac{n!}{(i-1)!(n-i-1)!}\{F(x_{(i)})\}^{i-1}\{1-F(x_{(i+1)})\}^{n-i-1}f(x_{(i)})f(x_{(i+1)})$$

with $0 < x_{(i)} < x_{(i+1)} < \sigma$. The density of $X_{(i)}$ is

$$f_{X_{(i)}}(x_{(i)}) = \frac{n!}{(i-1)!(n-i)!}\{F(x_{(i)})\}^{i-1}\{1-F(x_{(i)})\}^{n-i}f(x_{(i)}),\ 0 < x_{(i)} < \sigma.$$

So the conditional density is

$$\begin{aligned} f_{X_{(i+1)} \mid X_{(i)}}(x_{(i+1)} \mid x_{(i)}) &= \frac{f_{X_{(i)},X_{(i+1)}}(x_{(i)}, x_{(i+1)})}{f_{X_{(i)}}(x_{(i)})} \\ &= (n-i)\left\{\frac{1-F(x_{(i+1)})}{1-F(x_{(i)})}\right\}^{n-i-1}\left\{\frac{f(x_{(i+1)})}{1-F(x_{(i)})}\right\}. \end{aligned}$$

The desire result follows by realizing that $\frac{1-F(x_{(i+1)})}{1-F(x_{(i)})}$ and $\frac{f(x_{(i+1)})}{1-F(x_{(i)})}$ are the cdf and pdf of the population whose distribution is obtained by truncating the distribution $F(x)$ on the left at $x_{(i)}$. □

**6.** Find the marginal p.d.f of the range $R$ based on the joint pdf of $R$ and $X_{(n)}$.

*Proof.* The joint density of $X_{(1)}$ and $X_{(n)}$ is

$$\begin{aligned} f_{X_{(1)},X_{(n)}}(x_{(1)}, x_{(n)}) &= n(n-1)\{F(x_{(n)}) - F(x_{(1)})\}^{n-2}f(x_{(1)})f(x_{(n)}) \\ &= n(n-1)\left\{\frac{x_{(n)} - x_{(1)}}{\sigma}\right\}^{n-2}\frac{1}{\sigma^2} \\ &= \frac{n(n-1)}{\sigma^n}(x_{(n)} - x_{(1)})^{n-2}. \end{aligned}$$

Let $R = X_{(n)} - X_{(1)},\ X_{(n)} = X_{(n)}$ then $X_{(1)} = X_{(n)} - R,\ X_{(n)} = X_{(n)},\ |J| = 1,$

$$f_{R,X_{(n)}}(r, x_{(n)}) = \frac{n(n-1)}{\sigma^n}r^{n-2}, \quad 0 < r < x_{(n)} < \sigma.$$

Thus, the marginal p.d.f of $R$ is

$$f_R(r) = \int_r^\sigma f_{R,X_{(n)}}(r, x_{(n)})dx_{(n)} = \int_r^\sigma \frac{n(n-1)}{\sigma^n} r^{n-2} dx_{(n)} = \frac{n(n-1)}{\sigma^{n-1}} r^{n-2}\left(1 - \frac{r}{\sigma}\right),$$

with $0 < r < \sigma$. ☐

**7.** Define $\left\{W_n = 2n\left(1 - \frac{R}{\sigma}\right), n \geq 1\right\}$. For each $n$ show that the pdf of $W_n$ is given by

$$h_n(x) = \begin{cases} \frac{n-1}{4n} x\left(1 - \frac{x}{2n}\right)^{n-2}, & \text{for } 0 < x < 2n, \\ 0, & \text{otherwise} \end{cases}.$$

*Proof.* Since $W_n = 2n\left(1 - \frac{R}{\sigma}\right)$, $R = \sigma\left(1 - \frac{W_n}{2n}\right)$, $|J| = \frac{\sigma}{2n}$, then

$$h_n(x) = \frac{n-1}{\sigma^{n-1}}\left(1 - \frac{x}{2n}\right)^{n-2} \sigma^{n-2} \frac{x}{2} \frac{\sigma}{2n} = \frac{n-1}{4n} x\left(1 - \frac{x}{2n}\right)^{n-2},\ 0 < x < 2n.$$

☐

**8.** For $h_n(x)$ defined in **Problem 7**, show that there is a pdf $h$ such that

$$\lim_{n\to\infty} h_n(x) = h(x), \quad \text{for } 0 < x < \infty.$$

Let $W$ be a random variable with pdf $h$. Name the distribution of $W$ explicitly.

*Proof.* From $h_n(x)$ defined in **Problem 7**, one has

$$h_n(x) = \frac{1}{2n} \frac{\Gamma(n+1)}{\Gamma(2)\Gamma(n-1)} \left(\frac{x}{2n}\right)^{2-1} \left(1 - \frac{x}{2n}\right)^{n-1-1},\ 0 < x < 2n.$$

Thus, $W_n/(2n) \sim \text{Beta}(2, n-1)$ with $\text{E}(W_n) = \frac{4n}{n+1}$ and $\text{Var}(W_n) = \frac{8n^2(n-1)}{(n+1)^2(n+2)}$.
Since $n \to \infty$, then $\lim_{n\to\infty} \text{E}(W_n) = 4$, and $\lim_{n\to\infty} \text{Var}(W_n) = 8$ stay constant. When $n \to \infty$, one has that $I_{0<x<2n} \to I_{x>0}$, $\text{B}(2, n-1) \sim (n-1)^2$, and $\left(1 - \frac{x}{2n}\right)^{n-2} \sim \exp(-\frac{(n-1)x}{2n}) \sim \exp(-\frac{x}{2})$. Hence,

$$\lim_{n\to\infty} h_n(x) = h(x) = \frac{1}{\Gamma(2)2^2} x^{2-1} e^{-\frac{x}{2}},\ x > 0.$$

$$W \sim \text{Gamma}(2, 2) \text{ OR } \chi^2(4).$$

☐

## Part III

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample with common probability density function (pdf)

$$f(x) = \begin{cases} e^{-x}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

For Problems 9–13, consider the case where $X_1 > 0$ and $X_2 > 0$, and define the following two quantities: $Z = X_1 - X_2$ and $W = X_1/X_2$. Then, note that the two sets $\{Z = 0\}$ and $\{W = 1\}$ are identical. So, the following question arises: Is the conditional pdf of $X_1$ given $Z = 0$ equal to the conditional pdf of $X_1$ given $W = 1$? In the subsequent parts you will answer this question.

**9.** Find the joint pdf of $X_1$ and $Z$. Be sure to specify the range of $X_1$ and $Z$ in the joint density.

*Proof.*

$$f_{X_1,X_2}(x_1, x_2) = e^{-(x_1+x_2)},\ x_1, x_2 > 0.$$

Since $\begin{cases} Z = X_1 - X_2 \\ X_1 = X_1 \end{cases}$, one has $\begin{cases} X_2 = X_1 - Z \\ X_1 = X_1 \end{cases}$, and $|J| = 1$. So

$$f_{X_1,Z}(x_1, z) = e^{-(2x_1 - z)},\ -\infty < z < x_1,\ \max(0, z) < x_1 < \infty.$$

□

**10.** Show that the conditional pdf of $X_1$ given $Z = 0$ is

$$f_{X_1|Z=0}(x) = \begin{cases} 2e^{-2x}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

*Proof.* Based on (12), the marginal density of $Z$ is

$$\begin{aligned} \text{When } z > 0: \ f_Z(z) &= \int_z^\infty f_{X_1,Z}(x_1, z)dx_1 = \int_z^\infty e^{-(2x_1 - z)}dx_1 \\ &= -\frac{1}{2}e^z\ e^{-2x_1}\Big|_z^\infty = \frac{1}{2}e^z e^{-2z} = \frac{1}{2}e^{-z} \\ \text{When } z \le 0: \ f_Z(z) &= \int_0^\infty f_{X_1,Z}(x_1, z)dx_1 = \int_0^\infty e^{-(2x_1 - z)}dx_1 \\ &= -\frac{1}{2}e^z\ e^{-2x_1}\Big|_0^\infty = \frac{1}{2}e^z. \end{aligned}$$

In summary,

$$f_Z(z) = \begin{cases} \frac{1}{2}e^z, & \text{for } z \le 0 \\ \frac{1}{2}e^{-z}, & \text{for } z > 0 \end{cases}.$$

Hence,

$$f_{X_1|Z=0}(x) = 2e^{-2x},\ \text{for } x > 0.$$

□

**11.** Find the joint pdf of $X_1$ and $W$ and show that the marginal pdf of $W$ is

$$f_W(w) = \begin{cases} \frac{1}{(1+w)^2}, & \text{for } w > 0 \\ 0, & \text{otherwise} \end{cases}.$$

*Proof.*

$$f_{X_1,X_2}(x_1, x_2) = e^{-(x_1+x_2)},\ x_1, x_2 > 0.$$

Since $\begin{array}{l} W = X_1/X_2 \\ X_1 = X_1 \end{array}$, one has $\begin{array}{l} X_2 = X_1/W \\ X_1 = X_1 \end{array}$, and $|J| = 1$, the joint density of $Z_1$ and $W$ is

$$f_{X_1,W}(x_1, w) = \frac{x_1}{w^2} e^{-(1+\frac{1}{w})x_1},\ \text{for } x_1, w > 0.$$

Then the marginal density of $W$ is

$$\begin{aligned} f_W(w) &= \int_0^\infty \frac{x_1}{w^2} e^{-(1+\frac{1}{w})x_1} dx_1 = \frac{1}{w^2} \frac{w}{1+w} \int_0^\infty \frac{w+1}{w} e^{-(1+\frac{1}{w})x_1} dx_1 = \frac{1}{w^2} \frac{w^2}{(1+w)^2} \\ &= \frac{1}{(1+w)^2},\ w > 0. \end{aligned}$$

□

**12.** Show that the conditional pdf of $X_1$ given $W = 1$ is

$$f_{X_1|W=1}(x) = \begin{cases} 4xe^{-2x}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

*Proof.* Based on (14), one has

$$f_{X_1|W=1}(x) = 4xe^{-2x},\ \text{for } x > 0.$$

□

**14.** Consider the entire sample $\{X_1, X_2, \ldots, X_n\}$. Assume that $n$ is an even integer. Find the limiting distribution of the random variable $Z_n = \sqrt{2/n} \sum_{i=1}^n (-1)^{i-1} X_i$.

*Proof.* Notice that $f_{X_i}(x) = e^{-x},\ x > 0$, the MGF of $X_i$ is

$$M_{X_i}(t) = \mathrm{E}\left(e^{tX_i}\right) = \int_0^\infty e^{tx} e^{-x} dx = \frac{1}{1-t},\ t < 1.$$

$$\begin{aligned} M_{Z_n}(t) &= \mathrm{E}\left(e^{tZ_n}\right) = \mathrm{E}\left(e^{\sqrt{2/n}tX_1} e^{-\sqrt{2/n}tX_2} \cdots e^{\sqrt{2/n}tX_{n-1}} e^{-\sqrt{2/n}tX_n}\right) \\ &= \left(\frac{1}{1-\sqrt{2}t/\sqrt{n}}\right)^{\frac{n}{2}} \left(\frac{1}{1+\sqrt{2}t/\sqrt{n}}\right)^{\frac{n}{2}} = \left(\frac{1}{1-\frac{t^2}{n/2}}\right)^{\frac{n}{2}} \end{aligned}$$

Hence, $\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2}$, and $Z_n \sim N(0, 2)$. □

**1**. Since the joint pdf of $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is $(\lambda\mu)^{-n} \exp\{-(\sum_{i=1}^n x_i/\lambda + \sum_{i=1}^n y_i/\mu)\}$, by factorization theorem $(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i)$ is sufficient for $\theta$.

**2**. Let $\bar{x} = \sum_{i=1}^n x_i/n$ and $\bar{y} = \sum_{i=1}^n y_i/n$. The loglikelihood function for $(\lambda, \mu)$ is

$$\log \ell(\lambda, \mu) = -n\Big[\log\lambda + \log\mu + \frac{\bar{x}}{\lambda} + \frac{\bar{y}}{\mu}\Big].$$

Thus

$$\frac{\partial \log \ell(\lambda, \mu)}{\partial \lambda} = -n\Big[\frac{1}{\lambda} - \frac{\bar{x}}{\lambda^2}\Big] \text{ and } \frac{\partial \log \ell(\lambda, \mu)}{\partial \mu} = -n\Big[\frac{1}{\mu} - \frac{\bar{y}}{\mu^2}\Big].$$

Hence, the solution of the likelihood equations is $\hat{\theta}_n \equiv (\bar{x}, \bar{y})$. Note that, irrespective of the value of $\mu$, $\frac{\partial \log \ell(\lambda,\mu)}{\partial \lambda} \gtreqqless 0$ if $\lambda \lesseqqgtr \bar{x}$. Similarly, irrespective of the value of $\lambda$, $\frac{\partial \log \ell(\lambda,\mu)}{\partial \mu} \gtreqqless 0$ if $\mu \lesseqqgtr \bar{y}$. Thus, the likelihood function has a unique maximizer at $\hat{\theta}_n$.

**3**. By CLT $\sqrt{n}(\bar{X} - \lambda) \xrightarrow{d} N(0, \lambda^2)$ and $\sqrt{n}(\bar{Y} - \mu) \xrightarrow{d} N(0, \mu^2)$. Since $X_i$'s and $Y_i$'s are independent, $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N_2(0, V(\theta))$ where $V_{11}(\theta) = \lambda^2$, $V_{22}(\theta) = \mu^2$, and $V_{12}(\theta) = 0 = V_{21}(\theta)$.

**4**. Since

$$(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}}\sqrt{n}(\hat{\theta}_n - \theta),$$

and $\frac{1}{\sqrt{n}} \to 0$, by Slutsky's theorem $(\hat{\theta}_n - \theta) \xrightarrow{d} 0$, which is equivalent to $(\hat{\theta}_n - \theta) \xrightarrow{P} 0$.

**5**. Since $(\bar{X}, \bar{Y})$ is the MLE of $\theta$, by the invariance property of MLE, the MLE of $\rho = \rho(\theta)$ is $\hat{\rho}_n = \rho(\bar{X}, \bar{Y}) = \frac{\bar{X}}{\bar{Y}}$.

**6**. Note that $\partial\rho(\theta)/\partial\theta = (1/\mu, -\lambda/\mu^2)^T$. By Delta method,

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{d} N(0, (1/\mu, -\lambda/\mu^2)V(\theta)(1/\mu, -\lambda/\mu^2)^T).$$

That is $\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{d} N(0, 2\lambda^2/\mu^2)$.

**7**. Since $X_i$'s are independent Exp $(\lambda)$ random variables, $\sum_{i=1}^n X_i \sim$ Gamma $(n, \lambda)$, implying $\sum_{i=1}^n X_i/\lambda \sim$ Gamma $(n, 1)$. Similarly, since $Y_i$'s are independent Exp $(\mu)$ random variables, $\sum_{i=1}^n Y_i \sim$ Gamma $(n, \mu)$, implying $\sum_{i=1}^n Y_i/\mu \sim$ Gamma $(n, 1)$. Also, since $X_i$'s and $Y_i$'s are independent, $Q((X_1, \ldots, X_n, Y_1, \ldots, Y_n), \theta) = \sum_{i=1}^n X_i/\lambda + \sum_{i=1}^n Y_i/\mu \sim$ Gamma $(2n, 1)$. Thus, $Q((X_1, \ldots, X_n, Y_1, \ldots, Y_n), \theta) = \sum_{i=1}^n X_i/\lambda + \sum_{i=1}^n Y_i/\mu$ is a pivotal quantity.

**8**. We know that $2Q((X_1, \ldots, X_n, Y_1, \ldots, Y_n), \theta) \sim$ Gamma $(2n, 2) \equiv \chi^2_{4n}$, a $(1 - \alpha)$ confidence set for $\theta$ can be found from $\{\theta : 2Q((X_1, \ldots, X_n, Y_1, \ldots, Y_n), \theta) \le \chi^2_{4n,1-\alpha}\}$, where $\chi^2_{4n,\alpha}$ is the $\alpha$th quantile of $\chi^2_{4n}$.

**9**. The LRT statistic is

$$\lambda(\mathbf{x},\mathbf{y}) = \frac{\sup_{\theta\in\Theta_0}\ell(\theta)}{\sup_{\theta\in\Theta}\ell(\theta)} = \frac{\sup_\lambda \frac{1}{\lambda^{2n}}\exp\Big[-\frac{n\bar{x}+n\bar{y}}{\lambda}\Big]}{\sup_{\theta\in\Theta}\frac{1}{\lambda^n\mu^n}\exp\Big[-\frac{n\bar{x}}{\lambda}-\frac{n\bar{y}}{\mu}\Big]}$$

As before, differentiating the log likelihood function under $H_0$, it can be shown that the maximum of the numerator is attained at $\hat{\lambda}_0 = [n\bar{x}+n\bar{y}]/2n = [\bar{x}+\bar{y}]/2$. Thus,

$$\lambda(\mathbf{x},\mathbf{y}) = 2^{2n}\frac{\bar{x}^n\bar{y}^n}{(\bar{x}+\bar{y})^{2n}} = 2^{2n}r^n(1-r)^n.$$

**10**. Since

$$R = \frac{\bar{X}}{\bar{X}+\bar{Y}} = \frac{\sum_{i=1}^n X_i/\lambda}{\sum_{i=1}^n X_i/\lambda + \sum_{i=1}^n Y_i/\lambda},$$

and under $H_0$, $\sum_{i=1}^n X_i/\lambda$ and $\sum_{i=1}^n Y_i/\lambda$ are independent Gamma $(n,1)$ random variables, by the Fact 2, we have $R \sim$ Beta $(n,n)$.

**11**. Since $g(r) = 2^{2n}r^n(1-r)^n$ is symmetric about its maximizer $0.5$, $\lambda(\mathbf{x},\mathbf{y}) < c$ is equivalent to $r < 0.5-a$ or $r > 0.5+a$ for some $a$. Thus, $P_{H_0}(\lambda(\mathbf{X},\mathbf{Y}) < c) = \alpha$ is equivalent to $P_{H_0}(R < 0.5-a) + P_{H_0}(R > 0.5+a) = \alpha$. Under $H_0$, $R \sim$ Beta $(n,n)$, which is a symmetric distribution about its mode $0.5$. Thus, $P_{H_0}(R < 0.5-a) = P_{H_0}(R > 0.5+a) = \alpha/2$. So, $0.5 - a =$ Beta$(\alpha/2,n,n)$, the $\alpha/2$th quantile of Beta$(n,n)$. Thus, the size $\alpha$ LRT for testing $H_0 : \lambda = \mu$ against $H_1 : \lambda \neq \mu$ is $R <$ Beta$(\alpha/2,n,n)$ or $R > 1-$ Beta$(\alpha/2,n,n)$.

**12**. The posterior density of $\theta$ is

$$\begin{aligned}\pi(\theta|\mathbf{x},\mathbf{y}) &\propto \frac{1}{\lambda^n\mu^n}\exp\Big[-\frac{n\bar{x}}{\lambda}-\frac{n\bar{y}}{\mu}\Big]\lambda^{-\alpha_\lambda-1}\exp\Big\{-\frac{\beta_\lambda}{\lambda}\Big\}\mu^{-\alpha_\mu-1}\exp\Big\{-\frac{\beta_\mu}{\mu}\Big\}\\ &= \lambda^{-(\alpha_\lambda+n)-1}\exp\Big\{-\frac{n\bar{x}+\beta_\lambda}{\lambda}\Big\}\mu^{-(\alpha_\mu+n)-1}\exp\Big\{-\frac{n\bar{y}+\beta_\mu}{\mu}\Big\}.\end{aligned}$$

Thus $\pi(\theta|\mathbf{x},\mathbf{y}) = \pi(\lambda|\mathbf{x})\pi(\mu|\mathbf{y})$ , where $\pi(\lambda|\mathbf{x})$ is the density of IG $(\alpha_\lambda+n, n\bar{x}+\beta_\lambda)$ and $\pi(\mu|\mathbf{y})$ is the density of IG $(\alpha_\mu+n, n\bar{y}+\beta_\mu)$.

**13**. Since the posterior density of $\theta$ is also a product of two inverse gamma densities of $\lambda$ and $\mu$, respectively, the prior family is conjugate.

**14**. Note that

$$E[L(\theta,t)|\mathbf{x},\mathbf{y}] = \int_0^\infty \frac{(t_1-\lambda)^2}{\lambda}\pi(\lambda|\mathbf{x})d\lambda + \int_0^\infty \frac{(t_2-\mu)^2}{\mu}\pi(\mu|\mathbf{y})d\mu.$$

Now,

$$\int_0^\infty \frac{(t_1-\lambda)^2}{\lambda}\pi(\lambda|\mathbf{x})d\lambda \propto \int_0^\infty (t_1-\lambda)^2\lambda^{-(\alpha_\lambda+n+1)-1}\exp\Big\{-\frac{n\bar{x}+\beta_\lambda}{\lambda}\Big\}d\lambda,$$

which is minimized at the mean of IG $(\alpha_\lambda + n + 1, n\bar{x} + \beta_\lambda)$. Thus the Bayes estimator of $\theta$ under the loss function $L(\theta, t)$ is

$$\Big(\frac{n\bar{X} + \beta_\lambda}{\alpha_\lambda + n}, \frac{n\bar{Y} + \beta_\mu}{\alpha_\mu + n}\Big).$$

**15**. Note that

$$\begin{aligned} P(Z_i \le z, \Delta_i = 1) &= P(X_i \le z, X_i \le Y_i) \\ &= \int_0^z \int_x^\infty \frac{1}{\lambda\mu} \exp\Big\{-\frac{x}{\lambda}\Big\} \exp\Big\{-\frac{y}{\mu}\Big\} dy dx \\ &= \int_0^z \frac{1}{\lambda} \exp\Big\{-x\Big(\frac{1}{\lambda} + \frac{1}{\mu}\Big)\Big\} dx \\ &= \frac{\mu}{\lambda + \mu}\Big[1 - \exp\Big\{-z\Big(\frac{1}{\lambda} + \frac{1}{\mu}\Big)\Big\}\Big]. \end{aligned}$$

Similarly,

$$P(Z_i \le z, \Delta_i = 0) = P(Y_i \le z, Y_i \le X_i) = \frac{\lambda}{\lambda + \mu}\Big[1 - \exp\Big\{-z\Big(\frac{1}{\lambda} + \frac{1}{\mu}\Big)\Big\}\Big].$$

Let $D = \sum_{i=1}^n \Delta_i$ and $\bar{z} = \sum_{i=1}^n z_i/n$. Thus, the likelihood function of $(\lambda, \mu)$ is

$$\ell_1(\lambda, \mu) = \frac{1}{\lambda^D} \frac{1}{\mu^{n-D}} \exp\Big\{-n\bar{z}\Big(\frac{1}{\lambda} + \frac{1}{\mu}\Big)\Big\}.$$

**16**. If $0 < D < n$,

$$\frac{\partial \log \ell_1(\lambda, \mu)}{\partial \lambda} = -\Big[\frac{D}{\lambda} - \frac{n\bar{z}}{\lambda^2}\Big] \text{ and } \frac{\partial \log \ell_1(\lambda, \mu)}{\partial \mu} = -\Big[\frac{n - D}{\mu} - \frac{n\bar{z}}{\mu^2}\Big].$$

The likelihood equations have unique solutions at $\tilde{\lambda}_n = \sum_{i=1}^n z_i/D$ and $\tilde{\mu}_n = \sum_{i=1}^n z_i/(n - D)$, respectively. Thus the MLE of $(\lambda, \mu)$ is $(\sum_{i=1}^n Z_i/D, \sum_{i=1}^n Z_i/(n - D))$.

**17**. Note that $\Delta_i$'s are independent with

$$P(\Delta_i = 1) = P(Z_i < \infty, \Delta_i = 1) = \frac{\mu}{\lambda + \mu} = 1 - P(\Delta_i = 0).$$

Thus, $\Delta_i$'s are iid Bernoulli random variables with $P(\Delta_i = 1) = \mu/(\lambda + \mu)$. Next,

$$\begin{aligned} P(Z_i \le z) &= P(Z_i \le z, \Delta_i = 1) + P(Z_i \le z, \Delta_i = 0) \\ &= 1 - \exp\Big\{-z\Big(\frac{1}{\lambda} + \frac{1}{\mu}\Big)\Big\}. \end{aligned}$$

Thus, $Z_i$'s are iid exponential random variables with mean $\lambda\mu/(\lambda+\mu)$. Hence by WLLN,

$$\frac{\sum_{i=1}^n Z_i}{D} = \frac{\sum_{i=1}^n Z_i/n}{D/n} \xrightarrow{P} \frac{\lambda\mu/(\lambda+\mu)}{\mu/(\lambda+\mu)} = \lambda.$$

Similarly,

$$\frac{\sum_{i=1}^n Z_i}{n-D} = \frac{\sum_{i=1}^n Z_i/n}{1-D/n} \xrightarrow{P} \mu.$$

Thus, $(\tilde{\lambda}_n, \tilde{\mu}_n)$ is a consistent estimator of $(\lambda, \mu)$.

**18**. If $D = 0$,

$$\ell_1(\lambda,\mu) = \frac{1}{\mu^n} \exp\left\{ -n\bar{z}\left(\frac{1}{\lambda} + \frac{1}{\mu}\right)\right\},$$

which is increasing in $\lambda$ for any fixed $\mu$. Thus, there does not exist an MLE of $\lambda$. Similarly, if $D = n$,

$$\ell_1(\lambda,\mu) = \frac{1}{\lambda^n} \exp\left\{ -n\bar{z}\left(\frac{1}{\lambda} + \frac{1}{\mu}\right)\right\},$$

which is increasing in $\mu$ for any fixed $\lambda$. Thus, there does not exist an MLE of $\mu$.

**1**. We have $P(A_1 \le t) = 1$ for $t \ge t_c$ (i.e, $A_1 \le t_c$ with probability 1), and also

$$P(A_1 \le t) = P(X_1 \le t) = F_\theta(t) = 1 - e^{-t/\theta}$$

for $0 < t < t_c$. Hence, for $t > 0$, the cdf of $A_1$ is

$$F_{A_1}(t) = \begin{cases} 1 - e^{-t/\theta} & 0 < t < t_c, \\ 1 & t \ge t_c. \end{cases}$$

**2**. The range of $A_1$ is continuous but the cdf of $A_1$ is not: it jumps at $t_c$ by

$$e^{-t_c/\theta} = P(A_1 = t_c) = F_\theta(t_c) - \lim_{t \uparrow t_c} F_\theta(t) = 1 - [1 - e^{-t_c/\theta}].$$

The random variable $A_1$ is neither continuous nor discrete (but rather "mixed").

**3**. Fix $t > 0$ and plug $(-\infty, t]$ into the measure $\mu(\cdot)$ giving

$$\mu((-\infty, t]) = \int_0^{t_c} I(x \le t)\theta^{-1} e^{-x/\theta} dx + e^{-t_c/\theta} I(t_c \le t) = \begin{cases} 1 - e^{-t/\theta} & t < t_c \\ 1 & t \ge t_c, \end{cases}$$

which is the cdf of $A_1$. As their cdfs match (both cdfs are also 0 for $t \le 0$), $\mu(\cdot)$ and $P(A_1 \in \cdot)$ prescribe the same distributions/probability measures on the Borel sets $\mathcal{B}(\mathbb{R})$.

**4**. Note $n^{-1} \sum_{i=1}^n R_n/n = \sum_{i=1}^n I(X_i \le t_c)/n$ is a sample mean of iid variables $I(X_i \le t_c)$ with mean $\mathrm{E}I(X_i \le t_c) = P(X_i \le t_c) = F_\theta(t_c)$. By the SLLN, $n^{-1} \sum_{i=1}^n R_n \to F_\theta(t_c)$ holds as $n \to \infty$ (a.s.), and consequently, $(n - R_n)/n = 1 - n^{-1} R_n \to 1 - F_\theta(t_c) = e^{-t_c/\theta}$.

**5**. Due to a sample mean of iid terms with finite mean $\mathrm{E}A_1 = \theta F_\theta(t_c)$, we have $\sum_{i=1}^n A_i/n \to \mathrm{E}A_1 = \theta F_\theta(t_c)$ (a.s.). By **Problem 4**, $R_n/n \to 1 - e^{-t_c/\theta} = F_\theta(t_c) > 0$ as $n \to \infty$ (a.s.). Because the function $h(x, y) = x/y$, for $x, y \ge 0$ (say), is continuous at $(\theta F_\theta(t_c), F_\theta(t_c))$, it follows that

$$\hat{\theta}_n = h\left(\sum_{i=1}^n A_i/n, R_n/n\right) \to h(\theta F_\theta(t_c), F_\theta(t_c)) = \frac{\theta F_\theta(t_c)}{F_\theta(t_c)} = \theta$$

as $n \to \infty$ (a.s.) from $(\sum_{i=1}^n A_i/n, R_n/n) \to (\theta F_\theta(t_c), F_\theta(t_c))$ (a.s.).

**6**. Using the delta method, where $p'(\theta) \equiv dp(\theta)/d\theta = e^{-(t_w - t_c)/\theta}(t_w - t_c)/\theta^2 \ne 0$, we have

$$\sqrt{n}(\hat{p}_n - p) = \sqrt{n}[p(\hat{\theta}_n) - p(\theta)] \xrightarrow{d} N(0, [p'(\theta)]^2 \theta^2 [F_\theta(t_c)]^{-1}) \sim p'(\theta)\theta[F_\theta(t_c)]^{-1/2} Z_0$$

from the assumption given that $\sqrt{n}[\hat{\theta}_n - \theta] \xrightarrow{d} \theta[F_\theta(t_c)]^{-1/2} Z_0 \sim N(0, \theta^2 [F_\theta(t_c)]^{-1})$, involving a standard normal $Z_0$. Since $(n - R_n)/n \to e^{-t_c/\theta}$ as $n \to \infty$ (a.s.) by **Problem 4**, we have $(n - R_n)/n \xrightarrow{p} e^{-t_c/\theta}$ as $n \to \infty$ and then Slutsky's theorem gives

$$\sqrt{\frac{n - R_n}{n}} \cdot \frac{1}{\sqrt{p(1-p)}} \cdot \sqrt{n}(\hat{p}_n - p) \xrightarrow{d} \sqrt{e^{-t_c/\theta}} \cdot \frac{1}{\sqrt{p(1-p)}} p'(\theta)\theta[F_\theta(t_c)]^{-1/2} Z_0 \sim \sigma Z_0$$

for $\sigma = \sqrt{e^{-t_c/\theta}}\theta[F_\theta(t_c)]^{-1/2} p'(\theta)/\sqrt{p(1-p)}$.

**7**. Each observation $X_i$ has three disjoint possibilities: "$X_i \le t_c$," "$X_i \in (t_c, t_w]$" or "$X_i > t_w$," with corresponding probabilities $F_\theta(t_c)$, $F_\theta(t_w) - F_\theta(t_c)$ and $1 - F_\theta(t_w)$ that sum to 1. The random variables $(R_n, Y_n)$ represent cell counts for the number of times (among $n$ iid trials) that $X_i \le t_c$ occurs (i.e., $R_n$) and the number of times that $X_i \in (t_c, t_w]$ occurs (i.e., $Y_n$), so that $(R_n, Y_n)$ is multinomial$(n, F_\theta(t_c), F_\theta(t_w) - F_\theta(t_c))$-distributed.

**8.** The count $R_n \equiv \sum_{i=1}^n I(X_i \le t_c)$ is Binomial$(n, F_\theta(t_c))$-distributed. Conditional on $R_n = r$ (integer $0 \le r < n$), the support/range of $Y_n \equiv \sum_{i=1}^n I(X_i \in (t_c, t_w])$ is $\{0, \dots, n-r\}$ and the distribution of $Y_n$ is

$$
\begin{aligned}
P(Y_n = y | R_n = r) &= \frac{P(R_n = r, Y_n = y)}{P(R_n = r)} \\
= \frac{n!}{r!y!(n-r-y)!}&[F_\theta(t_c)]^r [F_\theta(t_w) - F_\theta(t_c)]^y [1 - F_\theta(t_w)]^{n-r-y} \times \frac{r!(n-r)!}{n!}[F_\theta(t_c)]^{-r}[1 - F_\theta(t_c)]^{-(n-r)} \\
= \frac{(n-r)!}{y!(n-r-y)!}&p^y(1-p)^{(n-r-y)}
\end{aligned}
$$

for

$$p = \frac{F_\theta(t_w) - F_\theta(t_c)}{1 - F_\theta(t_c)} = \frac{e^{-t_c/\theta} - e^{-\theta/t_w}}{e^{-t_c/\theta}} = 1 - e^{-(t_w - t_c)/\theta}$$

That is, $Y_n | R_n = r$ is Binominal$(n-r, p)$ distributed, if $(R_n, Y_n)$ is multinomial$(n, F_\theta(t_c), F_\theta(t_w) - F_\theta(t_c))$-distributed, and so has the same distribution as $\sum_{i=1}^{n-r} B_i$ for iid Bernoulli$(p)$ variables $B_1, \dots, B_{n-r}$.

**9.** Conditional on $R_n = r < n$, by **Problem 8**, we have

$$
\begin{aligned}
P\left(Y_n \le \widetilde{Y}_{\alpha,n}(p) \middle| R_n = r\right) &= P\left(\sum_{i=1}^{n-r} B_i \le \Phi^{-1}(\alpha)\sqrt{(n-r)p(1-p)} + (n-r)p\right) \\
= P&\left(\frac{\sum_{i=1}^{n-r}(B_i - p)}{\sqrt{(n-r)p(1-p)}} \le \frac{\Phi^{-1}(\alpha)\sqrt{(n-r)p(1-p)} + (n-r)p - (n-r)p}{\sqrt{(n-r)p(1-p)}}\right) \\
= P&\left(\frac{\sum_{i=1}^{n-r}(B_i - p)}{\sqrt{(n-r)p(1-p)}} \le \Phi^{-1}(\alpha)\right)
\end{aligned}
$$

for iid Bernoulli$(p)$ variables $B_1, \dots, B_{n-r}$ with mean $p$ and variance $p(1-p)$. By the Berry-Esseen theorem (since $|B_i| \le 1$), we have

$$\left| P\left(\frac{\sum_{i=1}^{n-r}(B_i - p)}{\sqrt{(n-r)p(1-p)}} \le \Phi^{-1}(\alpha)\right) - \Phi(\Phi^{-1}(\alpha)) \right| \le 5.5 \frac{1}{n-r} \frac{1}{\sqrt{p(1-p)}}.$$

**10.** If $R_n = n$, then $Y_n = 0$ and $\widetilde{Y}_{\alpha,n}(p) \equiv \Phi^{-1}(\alpha)\sqrt{(n-R_n)p(1-p)} + (n-R_n)p = 0$ and so that $P\left(Y_n \le \widetilde{Y}_{\alpha,n}(p) \middle| R_n = n\right) = 1$; hence,

$$\left| P\left(Y_n \le \widetilde{Y}_{\alpha,n}(p) \middle| R_n = n\right) - \alpha \right| = 1 - \alpha \le 2$$

when $R_n = n$. The bound in question (for any value of $R_n$) now follows from this (when $R_n = n$) combined from the bound in **Problem 9** (when $0 \le R_n < n$).

**11.** By **Problem 4**, $(n - R_n)/n \to e^{-t_c/\theta} > 0$ as $n \to \infty$ (a.s.). This implies that $(n - R_n) = n \cdot (n - R_n)/n \to \infty$ and $R_n/n \to e^{-t_c/\theta}$ as $n \to \infty$ (a.s.), so that the bound in **Problem 10** behaves as

$$2I(R_n = n) + 5.5\frac{1}{\sqrt{n - R_n}}\frac{1}{\sqrt{p(1-p)}}I(0 \le R_n < n) \to 0$$

as $n \to \infty$ (a.s.); that is, $I(R_n = n) = I(R/n = 1) \to 0$ (a.s.), while $I(0 \le R_n < n)$ is bounded by 1 (it converges to 1 too) with $1/\sqrt{n - R_n} \to 0$ (a.s.).

**12**. By **Problem 11**, $P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(p)\Big| R_n\right) \to \alpha$ as $n \to \infty$ (a.s.) and the conditional probability $P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(p)\Big| R_n\right)$ is bounded by 1 for all $n \geq 1$ (a.s.). Hence, as $n \to \infty$, the standard DCT gives that

$$P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(p)\right) = \mathrm{E}\left[P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(p)\Big| R_n\right)\right] = \int P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(p)\Big| R_n\right) dP \to \int \alpha dP = \alpha.$$

**13**. We have $\hat{p}_n \xrightarrow{p} p$ as $n \to \infty$ along with

$$\sqrt{\frac{n-R_n}{n}}\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} \sigma Z_0$$

by **Problem 6** (for some $\sigma > 0$ and $Z_0 \sim N(0,1)$). By Slutsky's theorem,

$$\Phi^{-1}(\alpha)\cdot\frac{\sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{p(1-p)}} + \sqrt{\frac{n-R_n}{n}}\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} \Phi^{-1}(\alpha) + \sigma Z_0$$

follows. By this and the fact that the standard normal cdf $\Phi(\cdot)$ is continuous, the continuous mapping theorem gives

$$\Phi\left(\Phi^{-1}(\alpha)\cdot\frac{\sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{p(1-p)}} + \sqrt{\frac{n-R_n}{n}}\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}}\right) \xrightarrow{d} \Phi(\Phi^{-1}(\alpha) + \sigma Z_0).$$

**14**. This follows by Slutsky's theorem from **Problem 13** and the **Fact**.

**15**. The conditional probability variables $CP_n \equiv P\left(Y_N \leq \widetilde{Y}_{\alpha,n}(\hat{p}_n)\Big| R_n, A_1, \ldots, A_n\right)$, $n \geq 1$, are bounded by 1 and therefore uniformly integrable; i.e., "tail expectations"

$$\sup_{n\geq 1} \mathrm{E}|CP_n| I(|CP_n| > t)$$

can (uniformly) be made arbitrarily small for large $t > 0$ (in fact, equaling 0 for $t > 1$).

**16**. Note "uniformly integrability + convergence in distribution" imply "convergence in expectation." Consequently, by **Problems 14-15**, it follows that

$$P\left(Y_n \leq \widetilde{Y}_{\alpha,n}(\hat{p}_n)\right) = \mathrm{E}\big[CP_n\big] \to \mathrm{E}\Big[\Phi(\Phi^{-1}(\alpha)+\sigma Z_0)\Big] = \int_{-\infty}^{\infty}\Phi(\Phi^{-1}(\alpha)+\sigma z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz,$$

since $Z_0 \sim N(0,1)$.

**17**. The prediction bound $\widetilde{Y}_{\alpha,n}(\hat{p}_n)$ fails to have asymptotically correct coverage $\alpha$, unlike $\widetilde{Y}_{\alpha,n}(p)$.