We consider the estimation of mean and variance of random variables. Let $X_1, \dots, X_n$ be i.i.d. random variables from a distribution with mean μ and variance σ^2. Let $\mu_h = \mathrm{E}(X_i - \mu)^h$ and $\alpha_h = \mathrm{E}(X_i^h)$ denote the hth centered moment and hth moment of X_i for $h = 3, 4$, respectively. Let

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

be the sample mean and sample variance respectively. Assume $\alpha_4 < \infty$.

We first consider the estimation of $\theta = \mu^2$.

1. Let $\hat{\theta}_1 = \bar{X}^2$. Show that $\hat{\theta}_1 \xrightarrow{p} \mu^2$.

2. Show that MSE($\hat{\theta}_1$) = $\begin{cases} 4n^{-1}\sigma^2\mu^2\{1 + o(1)\}, & \mu \neq 0; \\ 3n^{-2}\sigma^4\{1 + o(1)\}, & \mu = 0. \end{cases}$

3. As $\hat{\theta}_1$ is biased, we consider the estimator

$$\hat{\theta}_2 = \frac{1}{n(n-1)}\sum_{i \neq j} X_i X_j.$$

Show that $\hat{\theta}_2$ is an unbiased estimator for θ.

4. Find MSE($\hat{\theta}_2$) in terms of the moments of X_i.

5. Which estimator is better in terms of smaller MSE?

6. Show that $\hat{\theta}_2 \xrightarrow{p} \mu^2$.

7. Find the asymptotic distribution of $\sqrt{n}(\hat{\theta}_2 - \mu^2)$ for $\mu \neq 0$.

8. Find the asymptotic distribution of $1 + n\hat{\theta}_2/\sigma^2$ for $\mu = 0$.

Next, we estimate the coefficient of variation, $\omega = \sigma/\mu$. A natural estimator is $\hat{\omega} = S/\bar{X}$.

9. If $\mu \neq 0$, show that $\hat{\omega} \xrightarrow{p} \omega = \sigma/\mu$.

10. If $\mu \neq 0$, find the asymptotic distribution of $\sqrt{n}(\hat{\omega} - \omega)$.

11. If $\mu = 0$, show that $\hat{\omega}$ is not stochastically bounded.

12. If $\mu = 0$, does $\hat{\omega}/\sqrt{n}$ converge in distribution? If yes, find the asymptotic distribution; if not, show the reason.

13. If $\mu = 0$, does $E(\hat{\omega}/\sqrt{n})$ converge? If yes, find the limit; if not, show the reason.

Part I

1. Since $\mathrm{E}(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n \to 0$, we have $\bar{X} \to \mu$ in probability. Therefore, $\hat{\theta}_1 \to \mu^2$ in probability.

2. $\mathrm{E}(\hat{\theta}_1) = \mathrm{E}(\bar{X}^2) = \sigma^2/n + \mu^2$. $\mathrm{Var}(\hat{\theta}_1) = \mathrm{E}(\hat{\theta}_1^2) - (\sigma^2/n + \mu^2)^2 = \mathrm{E}(\bar{X}^4) - (\sigma^2/n + \mu^2)^2$. Note that

$$\begin{aligned}
\mathrm{E}(\bar{X}^4) &= \frac{1}{n^4} \sum_{i_1,i_2,i_3,i_4} \mathrm{E}(X_{i_1}X_{i_2}X_{i_3}X_{i_4}) \\
&= \frac{(n-1)(n-2)(n-3)}{n^3}\mu^4 + \frac{6(n-1)(n-2)}{n^3}\mu^2(\sigma^2+\mu^2) \\
&+ \frac{3(n-1)}{n^3}(\sigma^2+\mu^2)^2 + \frac{4(n-1)}{n^3}\alpha_3\mu + \frac{\alpha_4}{n^3} \\
&= \frac{(n-1)(n^2+n-3)}{n^3}\mu^4 + \frac{6(n-1)^2}{n^3}\sigma^2\mu^2 + \frac{3}{n^2}\sigma^4 + \frac{4\alpha_3\mu}{n^2} + \frac{\alpha_4 - 3\sigma^4 - 4\alpha_3\mu}{n^3}.
\end{aligned}$$

It follows that

$$\mathrm{Var}(\hat{\theta}_1) = 4n^{-1}\sigma^2\mu^2 + n^{-2}(2\sigma^4 + 4\alpha_3\mu - 12\sigma^2\mu^2 - 4\mu^4) + O(n^{-3}).$$

Therefore, $\mathrm{Var}(\hat{\theta}_1) = 4n^{-1}\sigma^2\mu^2\{1+o(1)\}$ if $\mu \neq 0$ and $\mathrm{Var}(\hat{\theta}_1) = 2n^{-2}\sigma^4\{1+o(1)\}$ if $\mu = 0$. Bias of $\hat{\theta}_1$ is σ^2/n. MSE of $\hat{\theta}_1$ is equal to $4n^{-1}\sigma^2\mu^2\{1+o(1)\}$ if $\mu \neq 0$, and it is equal to $3n^{-2}\sigma^4\{1+o(1)\}$ if $\mu = 0$.

3. $\mathrm{E}(\hat{\theta}_2) = \frac{1}{n(n-1)}\sum_{i\neq j} \mathrm{E}X_i\mathrm{E}X_j = \mu^2$.

4. For the variance of $\hat{\theta}_2$, we have

$$\mathrm{Var}(\hat{\theta}_2) = \frac{1}{n^2(n-1)^2}\sum_{i_1\neq i_2}\sum_{i_3\neq i_4} \mathrm{E}(X_{i_1}X_{i_2}X_{i_3}X_{i_4}) - \mu^4.$$

Since

$$\begin{aligned}
\sum_{i_1\neq i_2}\sum_{i_3\neq i_4} \mathrm{E}(X_{i_1}X_{i_2}X_{i_3}X_{i_4}) &= n(n-1)(n-2)(n-3)\mu^4 + 4n(n-1)(n-2)\mu^2(\sigma^2+\mu^2) \\
&+ 2n(n-1)(\sigma^2+\mu^2)^2,
\end{aligned}$$

we have

$$\mathrm{Var}(\hat{\theta}_2) = 4n^{-1}\sigma^2\mu^2 + 2\{n(n-1)\}^{-1}\sigma^4 = 4n^{-1}\sigma^2\mu^2 + 2n^{-2}\sigma^4\{1+o(1)\}.$$

5. From Questions 3 and 4, for $\mu = 0$, $\hat{\theta}_2$ has a smaller MSE. For $\mu \neq 0$,

$$\begin{aligned}
\mathrm{MSE}(\hat{\theta}_1) &= 4n^{-1}\sigma^2\mu^2 + n^{-2}(3\sigma^4 + 4\alpha_3\mu - 12\sigma^2\mu^2 - 4\mu^4) + O(n^{-3}) \\
\mathrm{MSE}(\hat{\theta}_2) &= 4n^{-1}\sigma^2\mu^2 + 2n^{-2}\sigma^4\{1+o(1)\}.
\end{aligned}$$

Therefore, $\hat{\theta}_1$ has a smaller variance if $\sigma^4 < 12\sigma^2\mu^2 + 4\mu^4 - 4\alpha_3\mu$.

6. Since $\mathrm{E}(\hat{\theta}_2) = \mu^2$ and $\mathrm{Var}(\hat{\theta}_2) \to 0$, we have $\hat{\theta}_2 \to \mu^2$ in probability.

7. Since $\hat{\theta}_2 = \bar{X}^2 - \sum X_i^2/n^2 + \hat{\theta}_2/n$, we have

$$\sqrt{n}(\hat{\theta}_2 - \mu^2) = \sqrt{n}(\bar{X}-\mu)(\bar{X}+\mu) - \frac{\sum X_i^2}{n^{3/2}} + \frac{\hat{\theta}_2}{\sqrt{n}}.$$

By the law of large numbers, $\sum X_i^2/n \to (\mu^2+\sigma^2)$ and $\bar{X}+\mu \to 2\mu$. Therefore, $\sum X_i^2/n^{3/2} \to 0$ and $\hat{\theta}_2/\sqrt{n} \to 0$ in probability. It follows that

$$\sqrt{n}(\hat{\theta}_2 - \mu^2) \to N(0, 4\sigma^2\mu^2).$$

8. If $\mu = 0$, we have $\sum X_i^2/n \to \sigma^2$ in probability. Then, $(n-1)\hat{\theta}_2/\sigma^2 = (\sqrt{n}\bar{X}/\sigma)^2 - \sum X_i^2/(n\sigma^2) \to \mathcal{X}_1^2 - 1$. Therefore, $n\hat{\theta}_2/\sigma^2 = \{n/(n-1)\}(n-1)\hat{\theta}_2/\sigma^2 \to \mathcal{X}_1^2 - 1$.

9. Since $\bar{X} \to \mu$ and $S^2 \to \sigma^2$ in probability, we have $\hat{\omega} \to \sigma/\mu$ in probability.

10. Note that
$$\hat{\omega} - \omega = \frac{S\mu - \sigma\bar{X}}{\mu\bar{X}} = \frac{(S/\sigma - 1)\sigma\mu - \sigma(\bar{X}-\mu)}{\mu\bar{X}}.$$
The denominator converges to μ^2 in probability. It suffices to consider the numerator. We can write $S/\sigma - 1$ as
$$\frac{S}{\sigma} - 1 = \frac{S^2-\sigma^2}{\sigma^2(S/\sigma+1)} = \frac{S^2-\sigma^2}{2\sigma^2} + \frac{S^2-\sigma^2}{2\sigma^2(s/\sigma+1)}(1 - s/\sigma).$$
Since $S^2 - \sigma^2 = O_p(n^{-1/2})$ and $S^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2 = n^{-1}\sum_{i=1}^n (X_i-\mu)^2 - (\bar{X}-\mu)^2 = n^{-1}\sum_{i=1}^n (X_i-\mu)^2 + O_p(n^{-1})$, it follows that $1 - S/\sigma = O_p(n^{-1/2})$ and
$$\frac{S}{\sigma} - 1 = \frac{1}{2\sigma^2 n}\sum_{i=1}^n \{(X_i-\mu)^2 - \sigma^2\} + O_p(n^{-1}).$$
This leads to
$$\sqrt{n}\{(S/\sigma-1)\sigma\mu - \sigma(\bar{X}-\mu)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\frac{\mu\{(X_i-\mu)^2-\sigma^2\}}{2\sigma} - \sigma(X_i-\mu)\right] + O_p(n^{-1/2}),$$
which is asymptotic normal with mean 0 and variance equal to
$$\mathrm{E}\left[\frac{\mu\{(X_i-\mu)^2-\sigma^2\}}{2\sigma} - \sigma(X_i-\mu)\right]^2 = \frac{\mu^2}{4\sigma^2}(\mu_4 - \sigma^4) + \sigma^4 - \mu\mu_3.$$
Therefore,
$$\sqrt{n}(\hat{\omega} - \omega) \to N(0, \sigma^4/\mu^4 - \mu_3/\mu^3 + (\mu_4 - \sigma^4)/(4\mu^2\sigma^2)).$$

11. As s is stochastically bounded, we only need to show $1/\bar{X}$ is not stochastically bounded. This is to show that, there exist a small $\epsilon > 0$, such that for any large M and $n_0 \in \mathbb{N}$, there exists $n > n_0$ such that $P(|1/\bar{X}| > M) > \epsilon$, which is true since $\bar{X} \to 0$ in probability.

12. Yes. But, can not use the continuous mapping theorem. Note that $\sqrt{n}\bar{X} \to N(0,\sigma^2)$. Let Z and $Z(\sigma)$ denote the random variable from the distribution $N(0,1)$ and $N(0,\sigma^2)$, respectively. By the slutsky's theorem, the convergence of $\hat{\omega}/\sqrt{n}$ is equivalent to the convergence of $1/(\sqrt{n}\bar{X})$. For $c > 0$, we have $P(1/(\sqrt{n}\bar{X}) < c) = P(\sqrt{n}\bar{X} < 0) + P(\sqrt{n}\bar{X} > 1/c) \to 0.5 + P(Z(\sigma) > 1/c) = P(1/Z(\sigma) < c)$. Similarly, for $c < 0$, $P(1/(\sqrt{n}\bar{X}) < c) = P(1/c < \sqrt{n}\bar{X} < 0) \to P(1/Z(\sigma) < c)$. Therefore, $1/(\sqrt{n}\bar{X}) \to 1/Z(\sigma)$, and $\hat{\omega}/\sqrt{n} \to 1/Z$.

13. No. $E|1/(\sqrt{n}\bar{X})|$ and $E|1/Z|$ do not exist.

Fact: You may use the following definition: The Gamma distribution with parameters (α, β) has density function

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\, x^{\alpha-1} e^{-x/\beta}.$$

Let $X_1, \ldots, X_n$ be independent discrete random variables with probability mass functions

$$\mathbb{P}(X_i = x) = \frac{e^{-\lambda i}\,(\lambda i)^x}{x!},$$

where $\lambda > 0, x \in \mathbb{N} \cup \{0\}$, and $i = 1, \ldots, n$. Answer the following questions.

1. Find $\mathbb{E}(X_i)$ and $\mathbb{E}(X_i^2)$.

2. Use the method of maximum likelihood to find the estimator for λ and call it $\hat{\lambda}_{\mathrm{MLE}}$.

3. Is $\hat{\lambda}_{\mathrm{MLE}}$ an unbiased estimator for λ? Why?

4. Can we use the method of moment to estimate λ? If yes, find the estimator and call it $\tilde{\lambda}$. If not, provide your reason(s).

5. Prove or disprove: $\hat{\lambda}_{\mathrm{MLE}}$ is the UMVUE for λ.

6. Prove or disprove: $\hat{\lambda}_{\mathrm{MLE}}$ is a consistent estimator for λ.

7. Prove or disprove: $\sqrt{n}(\hat{\lambda}_{\mathrm{MLE}} - \lambda)$ converges in distribution to normal distribution with mean zero and variance λ.

8. Find the Bayes estimator under the squared error loss function and call it $\hat{\lambda}_{\mathrm{B}}$.

9. Can we say the class of Gamma distribution priors is a conjugate family for the class of $\{f_{X_i}(x_i|\lambda)\}$? Explain your answer.

10. Find the relative efficiency of $\hat{\lambda}_{\mathrm{B}}$ with respect to $\hat{\lambda}_{\mathrm{MLE}}$? Can we say $\hat{\lambda}_{\mathrm{MLE}}$ is more efficient than $\hat{\lambda}_{\mathrm{B}}$? Why?

11. Prove or disprove: $\hat{\lambda}_{\mathrm{B}}$ is asymptotically unbiased for λ.

12. Prove or disprove: $\hat{\lambda}_{\mathrm{B}}$ is a consistent estimator for λ.

13. Prove or disprove: $T(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i$ is sufficient for λ.

14. Prove or disprove: $T(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i$ is complete.

15. Prove or disprove: $T(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i$ is minimal sufficient for λ.

For problems **16** to **20**, assume that the X_i's have common probability mass function

$$\mathbb{P}(X_i = x) = \frac{e^{-\lambda}\,\lambda^x}{x!},$$

where $\lambda > 0, x \in \mathbb{N} \cup \{0\}$, and $i = 1, \cdots, n$. Answer the following questions.

16. Find a size α LRT for the hypothesis test $H_0 : \lambda = \lambda_0$ vs $H_a : \lambda \neq \lambda_0$, and call it $\phi_1(\underset{\sim}{X})$ where $\underset{\sim}{X} = (X_1, \cdots, X_n)$. If you are using the MLE for λ, you don't need to verify it. Just use it.

17. Apply the Mood-Graybill-Boes (MGB) method to obtain a CI with significance level α for λ based on $T = \sum_{i=1}^{n} X_i$. [Hint: In your calculation, you may use the fact that if $Y \sim$ Gamma(a, b) then $\mathbb{P}(Y \leq y) = \mathbb{P}(X \geq a)$, where $X \sim$ Poisson (y/b) and a is an integer number.]

18. Find a variance stabilizing transformation based on $\hat{\lambda}_n = \bar{X}$, and find a corresponding large-sample CI for θ with approximate confidence level $1 - \alpha$. [Hint: You may start with the limiting distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda)$.]

For **Problem 19** and **Problem 20**, consider the hypothesis test $H_0 : \lambda \leq \lambda_0$ vs $H_a : \lambda > \lambda_0$, and let λ have a Gamma(a, b) distribution.

19. Find the posterior distribution of λ given $\underset{\sim}{x}$.

20. Find a $(1 - \alpha)$ credible interval for λ.

1. Each X_i has Poisson distributions with parameter λi for $i = 1, 2, \dots, n$. Therefore, $\mathbb{E}(X_i) =$Var$(X_i) = \lambda i$. This implies that $\mathbb{E}(X_i^2) = \lambda i + (\lambda i)^2$.

2. From the likelihood function, we see that

$$l(\lambda, x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda i}(\lambda i)^{x_i}}{x_i!} = \frac{e^{-\lambda \sum_{i=1}^{n} i}\ (\lambda)^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} i^{x_i}}{\prod_{i=1}^{n} x_i!}$$

and then

$$\log l(\lambda, x_1, \dots, x_n) = -\lambda \frac{n(n+1)}{2} + \Big(\sum_{i=1}^{n} x_i\Big) \log \lambda + \sum_{i=1}^{n} x_i \log i - \sum_{i=1}^{n} \log(x_i!).$$

Taking the derivative of $\log l(\lambda, x_1, \dots, x_n)$ with respect to λ and solving $\frac{d}{d\lambda} \log l(\lambda, x_1, \dots, x_n)|_{\lambda=\hat{\lambda}} = 0$ yields $\hat{\lambda} = \frac{2\sum_{i=1}^{n} x_i}{n(n+1)}$. Since $\frac{d2}{d\lambda^2} \log l(\lambda, x_1, \dots, x_n)|_{\lambda=\hat{\lambda}} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$, therefore $\hat{\lambda}$ maximizes the likelihood function, and we call it $\hat{\lambda}_{\text{MLE}}$.

3. $\hat{\lambda}_{\text{MLE}}$ is an unbiased estimator for λ since

$$\mathbb{E}(\hat{\lambda}_{\text{MLE}}) = \mathbb{E}\Big(\frac{2\sum_{i=1}^{n} x_i}{n(n+1)}\Big) = \Big(\frac{2\sum_{i=1}^{n} \lambda i}{n(n+1)}\Big) = \Big(\frac{2\lambda}{n(n+1)}\Big) \sum_{i=1}^{n} i = \lambda$$

4. The method of moments cannot be applied here since each random variables X_i has a different distribution ($X_i \sim$ Poisson(λi)).

5. We prove $\hat{\lambda}_{\text{MLE}}$ is UMVUE by showing that the variance of the MLE is the same as the CRLB. To see this, we write

$$\begin{aligned} I_n(\lambda) &= \mathbb{E}\Big[\Big(\frac{\partial}{\partial \lambda} \log f(x_1, \dots, x_n; \lambda)\Big)^2\Big] \\ &= \mathbb{E}\Big[\Big(\frac{\sum_{i=1}^{n} X_i}{\lambda}\Big)^2 + \frac{n^2(n+1)^2}{4} - 2\frac{n(n+1)}{2}\frac{\sum_{i=1}^{n} X_i}{\lambda}\Big] \\ &= \frac{1}{\lambda^2}\Big\{\frac{\lambda n(n+1)}{2} + \frac{\lambda^2 n^2(n+1)^2}{4}\Big\} + \frac{n^2(n+1)^2}{4} - \frac{n^2(n+1)^2}{2} \\ &= \frac{n(n+1)}{2\lambda}. \end{aligned}$$

On the other side,

$$\text{Var}\Big(\hat{\lambda}_{\text{MLE}}\Big) = \frac{4}{n^2(n+1)^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{2\lambda}{n(n+1)}.$$

Therefore, the variance of the $\hat{\lambda}_{\text{MLE}}$ is the same as the CRLB which implies that $\hat{\lambda}_{\text{MLE}}$ is the UMVUE.

6. We prove that $\hat{\lambda}_{\mathrm{MLE}}$ is consistent for λ as follows: For any fixed $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbb{P}\Big(|\hat{\lambda}_{\mathrm{MLE}} - \lambda| \geq \varepsilon\Big) &= \mathbb{P}\Big(\Big(\hat{\lambda}_{\mathrm{MLE}} - \lambda\Big)^2 \geq \varepsilon^2\Big) \\
&\leq \frac{\mathbb{E}\Big(\Big(\hat{\lambda}_{\mathrm{MLE}} - \lambda\Big)^2\Big)}{\varepsilon^2} \\
&= \frac{\mathrm{Var}\Big(\hat{\lambda}_{\mathrm{MLE}}\Big)}{\varepsilon^2} = \Big(\frac{2}{n(n+1)}\Big)^2 \frac{1}{\varepsilon^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) \\
&= \frac{2\lambda}{n(n+1)\varepsilon^2} \to 0
\end{aligned}$$

as $n \to \infty$. Therefore, $\hat{\lambda}_{\mathrm{MLE}}$ is consistent for λ.

7. For any fixed $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbb{P}\Big(\sqrt{n}|\hat{\lambda}_{\mathrm{MLE}} - \lambda| \geq \varepsilon\Big) &= \mathbb{P}\Big(n\Big(\hat{\lambda}_{\mathrm{MLE}} - \lambda\Big)^2 \geq \varepsilon^2\Big) \\
&\leq \frac{\mathbb{E}\Big(\Big(\hat{\lambda}_{\mathrm{MLE}} - \lambda\Big)^2\Big)}{\varepsilon^2/n} \\
&= \frac{n\mathrm{Var}\Big(\hat{\lambda}_{\mathrm{MLE}}\Big)}{\varepsilon^2} = \frac{n}{\varepsilon^2}\Big(\frac{2}{n(n+1)}\Big)^2 \frac{n(n+1)\lambda}{2} \\
&= \frac{2\lambda}{(n+1)\varepsilon^2} \to 0
\end{aligned}$$

as $n \to \infty$. Therefore, $\sqrt{n}(\hat{\lambda}_{\mathrm{MLE}} - \lambda)$ converges to zero in probability and hence $\sqrt{n}(\hat{\lambda}_{\mathrm{MLE}} - \lambda)$ converges to zero in distribution. Therefore, $\sqrt{n}(\hat{\lambda}_{\mathrm{MLE}} - \lambda)$ does not converge to a normal distribution.

8. We know that the Bayes estimator under the squared error loss is the posterior mean. To find the posterior distribution, write

$$\begin{aligned}
f(\lambda|\underset{\sim}{x}) \propto f(\underset{\sim}{x}|\lambda)f(\lambda) &\propto \prod_{ki=1}^{n} \Big[e^{-\lambda i}(\lambda i)^{x_i}\Big] \lambda^{\alpha} e^{-\lambda/\beta} \\
&\propto \lambda^{\alpha + \sum_{i=1}^{n} x_i - 1} e^{-\left(\frac{n(n+1)}{2} + \frac{1}{\beta}\right)\lambda}.
\end{aligned}$$

Therefore, $\lambda|\underset{\sim}{x}$ ~Gamma($\alpha + \sum_{i=1}^{n} x_i$, $\frac{2\beta}{n(n+1)\beta+2}$). Now, $\hat{\lambda}_{\mathrm{B}} = \mathbb{E}(\lambda|\underset{\sim}{x}) = \frac{2\beta(\alpha+\sum_{i=1}^{n} x_i)}{n(n+1)\beta+2}$.

9. Yes. The posterior distribution is also a gamma distribution. Therefore, the class of gamma distribution priors is a conjugate family for the class of $\{f_{X_i}(x_i|\lambda)\}$.

10. By the definition, the relative efficiency of $\hat{\lambda}_{\mathrm{B}}$ with respect $\hat{\lambda}_{\mathrm{MLE}}$, r.e. $(\hat{\lambda}_{\mathrm{B}}, \hat{\lambda}_{\mathrm{MLE}}, \lambda)$, is $\frac{\mathrm{Var}(\hat{\lambda}_{\mathrm{MLE}})}{\mathrm{Var}(\hat{\lambda}_{\mathrm{B}})}$. Next, we compute the variance of each estimator:

$$\mathrm{Var}(\hat{\lambda}_{\mathrm{MLE}}) = \frac{2\lambda}{n(n+1)}$$

and

$$\mathrm{Var}(\hat{\lambda}_{\mathrm{B}}) = \frac{2\beta^2 n(n+1)\lambda}{(n(n+1)\beta+2)^2}.$$

Therefore,

$$\mathrm{r.e.}(\hat{\lambda}_{\mathrm{B}}, \hat{\lambda}_{\mathrm{MLE}}, \lambda) = \big(1 + \frac{2}{n(n+1)\beta}\big)^2 > 1.$$

Although the variance of $\hat{\lambda}_{\mathrm{B}}$ is less than the variance of $\hat{\lambda}_{\mathrm{MLE}}$, however we cannot say which one is more efficient since $\hat{\lambda}_{\mathrm{B}}$ is biased.

11. We prove that $\hat{\lambda}_{\mathrm{B}}$ is asymptotically unbiased for λ. To see this, write

$$\mathbb{E}(\hat{\lambda}_{\mathrm{B}}) = \mathbb{E}\Big(\frac{2\beta(\alpha + \sum_{i=1}^n X_i)}{n(n+1)\beta+2}\Big) = \frac{2\beta\Big(\frac{\alpha}{n(n+1)} + \frac{\lambda}{2}\Big)}{\beta + \frac{2}{n(n+1)}} \to \frac{\beta\lambda}{\beta} = \lambda$$

as $n \to \infty$. Therefore, $\hat{\lambda}_{\mathrm{B}}$ is asymptotically unbiased for λ.

12. We prove that $\hat{\lambda}_{\mathrm{B}}$ is consistent for λ. Note that $\hat{\lambda}_{\mathrm{B}}$ can be decomposed as

$$\hat{\lambda}_{\mathrm{B}} = \frac{2\beta(\alpha + \sum_{i=1}^n X_i)}{n(n+1)\beta+2} = \Big(\underbrace{\frac{2\alpha\beta}{n(n+1)\beta+2}}_{:=a_n}\Big) + \frac{\underbrace{\Big(\frac{2}{n(n+1)}\sum_{i=1}^n X_i\Big)}_{=\hat{\lambda}_{\mathrm{MLE}}}}{\underbrace{\Big(1 + \frac{2}{n(n+1)\beta}\Big)}_{:=b_n}}.$$

We see that $a_n \to 0$ in probability and $b_n \to 1$ in probability. Now, apply the Slutsky's theorem along with the fact that $\hat{\lambda}_{\mathrm{MLE}}$ is consistent to see that $\hat{\lambda}_{\mathrm{B}} \to \lambda$ is probability. This completes the proof.

13. We prove that $T(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$ is sufficient for λ. By writing the joint density function, we have

$$f(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^n \frac{e^{-\lambda i}(\lambda i)^{x_i}}{x_i!} = \underbrace{\Big(\prod_{i=1}^n i_i^x\Big)}_{:=h(x_1,\ldots,n)} \underbrace{e^{-\frac{n(n+1)}{2}\lambda}\, e^{\log\lambda \sum_{i=1}^n x_i}}_{:=g(x_1,\ldots,x_n,\lambda)}.$$

The function h does not depend on λ, and the function g depends on both x_i's and λ. Therefore, factorization theorem implies that $T(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$ is sufficient for λ.

14. $T(X_1,\ldots,X_n)=\sum_{i=1}^n X_i$ is complete for λ since the joint density function can be written in terms of exponential family.

15. There are two ways to show that $T(X_1,\ldots,X_n)=\sum_{i=1}^n X_i$ is minimal sufficient for λ. One way is to apply the definition of minimal sufficiency. The other way is to use the fact that $T(X_1,\ldots,X_n)$ is sufficient and complete and this gives the desired result.

16. Recall that $\hat{\lambda}_{\text{MLE}}=\hat{\lambda}$ is the MLE for λ over the entire parameter space. Next, we apply the LRT to see that

$$\lambda(\underset{\sim}{x})=\frac{f(\underset{\sim}{x};\lambda_0)}{f(\underset{\sim}{x},\hat{\lambda})}=\frac{e^{-n\lambda_0}\ \lambda_0^{\sum_{i=1}^n x_i}}{e^{-n\hat{\lambda}}\ \hat{\lambda}^{\sum_{i=1}^n x_i}}$$

and

$$\phi_1(\underset{\sim}{x})=\begin{cases}1 & \text{if } \ \lambda(\underset{\sim}{x})<k\\ \gamma & \text{if } \ \lambda(\underset{\sim}{x})=k\\ 0 & \text{if } \ \lambda(\underset{\sim}{x})>k\end{cases}$$

where $\gamma\in[0,1]$, $0\le k\le 1$ are constants determined by $\mathbb{E}_{\lambda_0}(\phi_1\underset{\sim}{X})=\alpha$. Define $T=\sum_{i=1}^n X_i$ ~Possion($n\lambda$). We can write

$$\begin{aligned}\lambda(\underset{\sim}{x})<k &\Leftrightarrow \ln(\lambda(\underset{\sim}{x}))<\ln k\\ &\Leftrightarrow -n\lambda(\lambda_0-\hat{\lambda})+\sum_{i=1}^n x_i\ln\Big(\frac{\lambda_0}{\hat{\lambda}}\Big)<\ln k\\ &= T+T\ln\lambda_0-T\ln(\frac{T}{n})<\ln k+n\lambda_0=k_1.\end{aligned}$$

Next, define $g(t)=t+t\ln\lambda_0-t\ln t+t\ln n=(1+\ln n\lambda_0)t-t\ln t$. Note that

$$g^{'}(t)=1+\ln(n\lambda_0)-(\ln t+1)=\begin{cases}>0 & \text{if } \ t<n\lambda_0\\ =0 & \text{if } \ t=n\lambda_0\\ <0 & \text{if } \ t>n\lambda_0\end{cases}$$

Therefore,

$$\lambda(\underset{\sim}{x})<k\Leftrightarrow T<c_1(\le c_1+1)\text{ or }T>c_2(\ge c_2+1)\text{ s.t. }g(c_1)=g(c_2)=k$$

and

$$\phi_1(\underset{\sim}{X})=\begin{cases}1 & \text{if } \ T\le c_1-1 \textit{ or } T\ge c_2+1\\ \gamma & \text{if } \quad T=c_1 \textit{ or } T=c_2\\ 0 & \text{if } \quad c_1+1\le T\le c_2-1\end{cases},$$

where

$$\alpha=\mathbb{P}\Big(T\le c_1-1\Big)+\mathbb{P}\Big(T\ge c_2+1\Big)+\gamma\Big(\mathbb{P}\Big(T=c_1\Big)+\mathbb{P}\Big(T=c_2\Big)\Big).$$

17. Define $Q(t,\lambda) = \mathbb{P}\Big(T \le t|\lambda\Big)$ where $T = \sum_{i=1}^n X_i$ has a Poission distribution with parameter $n\lambda$. Also, let $\chi^2_{\nu,\alpha}$ be a point such that $\mathbb{P}(\chi^2_\nu \ge \chi^2_{\nu,\alpha}) = \alpha$. Note that

$$\begin{aligned}\mathbb{P}\Big(T \ge t+1|\lambda\Big) &= \mathbb{P}(Y \le n) \quad Y \sim \text{Gamma}(t+1, 1/\lambda)\\ &= \mathbb{P}(2\lambda Y \le 2\lambda n) \quad 2\lambda Y \sim \chi^2_{2(t+1)}\\ &= \mathbb{P}(\chi^2_{2(t+1)} \le 2\lambda n)\end{aligned}$$

is increasing in λ. Therefore, $Q(t,\lambda)$ is decreasing in λ. Now, based on the MGB method, we find the lower bound λ_L and upper bound λ_U such that

$$\mathbb{P}\Big(T \ge t|\lambda_L\Big) = \alpha/2 \quad \& \quad \mathbb{P}\Big(T \le t|\lambda_U\Big) = \alpha/2.$$

Note that

$$\mathbb{P}\Big(T \ge t|\lambda_L\Big) = \alpha/2 \Rightarrow \mathbb{P}\Big(\chi^2_{2t} \ge 2\lambda_L n\Big) = \alpha/2 \Rightarrow \lambda_L = \frac{1}{2n}\chi^2_{2t,1-\alpha/2}$$

and

$$\mathbb{P}\Big(T \le t|\lambda_U\Big) = \alpha/2 \Rightarrow \mathbb{P}\Big(\chi^2_{2(t+1)} \le 2\lambda_U n\Big) = 1-\alpha/2 \Rightarrow \lambda_U = \frac{1}{2n}\chi^2_{2(t+1),\alpha/2}.$$

Hence $[\frac{1}{2n}\chi^2_{2T,1-\alpha/2}, \frac{1}{2n}\chi^2_{2(T+1),\alpha/2}]$ is a CI for λ with a confidence coefficient greater than or equals to $1-\alpha$.

18. By CLT, $\sqrt{n}(\hat{\lambda}_n - \lambda)$ $\to$N$(0,\lambda)$ in distribution as $n \to \infty$. Let $g^{'}(\lambda) = \frac{1}{\sqrt{\lambda}}$. Then $g(\lambda) = 2\sqrt{\lambda}$ and $\sqrt{n}(2\sqrt{\hat{\lambda}_n} - 2\sqrt{\lambda})$ $\to$N$(0,1)$ in distribution as $n \to \infty$. This implies that Therefore, a corresponding large-sample CI for λ with confidence coefficient $1-\alpha$ has the form

$$\begin{aligned}C_{\underset{\sim}{X}} &= \Big\{\lambda > 0 : -z_{\alpha/2} \le 2\sqrt{n}(\sqrt{\hat{\lambda}_n} - \sqrt{\lambda}) \le z_{\alpha/2}\Big\}\\ &= \Big\{\lambda > 0 : \sqrt{\hat{\lambda}_n} - \frac{z_{\alpha/2}}{2\sqrt{n}} \le \sqrt{\lambda} \le \sqrt{\hat{\lambda}_n} + \frac{z_{\alpha/2}}{2\sqrt{n}}\Big\},\end{aligned}$$

where $\mathbb{P}(Z \ge z_\alpha) = \alpha$.

19. We first find the posterior distribution as follows:

$$\begin{aligned}f(\lambda|\underset{\sim}{x}) &\propto f(\underset{\sim}{x}|\lambda)\pi(\lambda) \propto e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}\lambda^{a-1}e^{-\lambda/\boldsymbol{\beta}}\\ &\propto \lambda^{a+\sum_{i=1}^n x_i - 1}e^{-(n+1/b)\lambda}.\end{aligned}$$

Therefore, $\lambda|\underset{\sim}{x}$ $\propto$Gamma$\Big(a + \sum_{i=1}^n x_i, \frac{b}{nb+1}\Big)$.

20. The Bayes interval with confidence coefficient $1-\alpha$ has the form

$$C_{\underset{\sim}{X}} = \left\{ \lambda > 0 : a \leq \lambda \leq b \right\}$$

such that

$$\mathbb{P}_{\lambda|\underset{\sim}{X}}(a \leq \lambda \leq b) = 1 - \alpha.$$

Using the fact that $\frac{2(nb+1)}{b} \sim \chi^2_{\chi}$, we see that

$$C_{\underset{\sim}{X}} = \left\{ \lambda > 0 : \chi^2_{2(a+\sum_{i=1}^n x_i),1-\alpha/2} \leq \frac{2(nb+1)}{b}\lambda \leq \chi^2_{2(a+\sum_{i=1}^n x_i),\alpha/2} \right\},$$

where $\mathbb{P}(\chi^2_\nu \geq \chi^2_{\nu,\alpha}) = \alpha$. Finally,

$$\left[\frac{b\chi^2_{2(a+\sum_{i=1}^n x_i),1-\alpha/2}}{2(nb+1)}, \frac{b\chi^2_{2(a+\sum_{i=1}^n x_i),\alpha/2}}{2(nb+1)} \right]$$

is a $(1-\alpha)$ credible interval for λ.

Some useful notations

- For two subsets A and $B \subseteq \Omega$, $A\Delta B = (A \cap B^c) \cup (B \cap A^c)$, where A^c and B^c denote, respectively, the complements of A and B in Ω.
- The conditional expectation of a function f given a sub-σ-field $\mathcal{B}$ under a probability measure P is denoted by $E_P(f|\mathcal{B})$.
- For two measures μ and λ, we say $\mu << \lambda$ if μ is dominated byλ.

Part I

Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space and $f : \Omega \to \mathbb{R}$ is measurable and integrable function.

1. Show that for each $\epsilon > 0$ there is an integrable simple function g such that $\int_\Omega |f - g| d\mu < \epsilon$.
2. Suppose that $\mathcal{F}_0$ is a field generating $\mathcal{F}$. Show that if μ is σ-finite on $\mathcal{F}_0$, then the function g in **Problem 1** can be taken to be of the form $\sum_i x_i I_{A_i}$ where $x_i \in \mathbb{R}$ and $A_i \in \mathcal{F}_0$.
3. Show by example that the conclusion to **Problem 2** may be false if μ is not σ-finite on $\mathcal{F}_0$.

Part II

Suppose $\mathcal{A}$ is the field of finite and cofinite subsets of an infinite set Ω.

$$\nu(A) = \begin{cases} 0 & \text{if } A \text{ is finite} \\ 1 & \text{if } A^c \text{ is finite} \end{cases}$$

4. Show that ν is countably additive on $\mathcal{A}$ if and only if Ω is uncountable.

Part III

5. State Tonelli's theorem. *Hint*: It is the simpler version of the Fubini's theorem for non-negative functions.
6. Prove for a cdf F that $\int_{-\infty}^{\infty} \big(F(x + c) - F(x)\big) dx = c$ for all c.

Part IV

Let Ω be a non-empty set and $(\Omega, \mathcal{F}, P)$ be a probability space.

7. Suppose $A \in \mathcal{F}$ and there exist sequences $\{B_n\}$ and $\{C_n\}$ of events such that B_n and C_n are independent for each n, and
$$\lim_{n\to\infty} P(A \,\Delta\, B_n) = \lim_{n\to\infty} P(A \,\Delta\, C_n) = 0.$$
Show that $P(A)$ is either 0 or 1.

8. Suppose $\{Y_i\}_{i \in I}$ is a family of random variables on an index set I.

a) Show that if $\int_0^\infty \sup_{i \in I} P(|Y_i| \geq t) dt < \infty$ then Y_i's are uniformly integrable.

b) Show by a counterexample that the converse to **Problem 8a** is not necessarily true.

Part V

Suppose $X_1, X_2, \ldots, X_n$ are iid with pdf

$$f(x;\theta) = \exp\big(-(x-\theta)\big) I(x > \theta)$$

where $I(\cdot)$ is the indicator function and $\theta > 0$ is an unknown parameter. Let $M_n = \min\{X_1, \ldots, X_n\}$ denote the smallest order statistic.

9. Show that the cdf of M_n is given by

$$G(x) = \begin{cases} 0 & \text{if } x < \theta \\ 1 - \exp(-n(x-\theta)) & \text{if } x > \theta \end{cases}$$

10. Show that $M_n \xrightarrow{P} \theta$.

11. Suppose $c > \theta$ is fixed. Consider the random variable:

$$U_n = \begin{cases} \exp(M_n - c) & \text{if } M_n < c \\ 1 & \text{if } M_n \geq c. \end{cases}$$

Find sequences (a_n) and (b_n) of real numbers (possibly also depending on c and θ), such that $L_n := a_n(U_n - b_n)$ converges in distribution to a random variable L having Exponential distribution with mean $\exp(\theta - c)$.

12. Show that $E(L_n^k) \to E(L^k)$ as $n \to \infty$ for all positive integers k, where L_n and L are as defined in **Problem 11**.

Part VI

13. Suppose $\{Y_n\}$ is a sequence of independent random variables with distribution functions F_n. Show that

$$Y_n \to 0 \text{ a.s. if and only if } \sum_{n=1}^{\infty} \{1 - F_n(\epsilon) + F_n(-\epsilon)\} < \infty, \ \forall \epsilon > 0.$$

14. Use the CLT to prove that

$$\lim_{n \to \infty} \sum_{k=n}^{\infty} e^{-n} \frac{n^k}{k!} = \frac{1}{2}.$$

Part I

Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space and $f : \Omega \to \mathbb{R}$ is measurable and integrable function.

1. Show that for each $\epsilon > 0$ there is an integrable simple function g such that $\int_\Omega |f - g| d\mu < \epsilon$.

Solution: Let $f^+ = f I_{f \geq 0}$ and $f^- = -f I_{f<0}$. Then both f^+ and f^- are integrable because f is integrable. By definition of integral, there exists sequences of $\mathcal{F}$-measurable non-negative simple functions g_n and h_n such that

$$g_n \uparrow f^+, \quad \int_\Omega g_n d\mu \uparrow \int_\Omega f^+ d\mu, \quad h_n \uparrow f^-, \quad \int_\Omega h_n d\mu \uparrow \int_\Omega f^- d\mu.$$

Thus, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$, such that

$$0 \leq \int_\Omega f^+ d\mu - \int_\Omega g_N d\mu < \epsilon/2 \quad \text{and} \quad 0 \leq \int_\Omega f^- d\mu - \int_\Omega h_N d\mu < \epsilon/2.$$

Let $g = g_N - h_N$. Then g is an integrable $\mathcal{F}$-measurable simple function since both g_N and h_N are so, and,

$$\int_\Omega |f - g| d\mu = \int_\Omega |f^+ - f^- - g_N + h_N| d\mu \leq \int_\Omega |f^+ - g_N| d\mu + \int_\Omega |f^- - h_N| d\mu < \epsilon/2 + \epsilon/2 = \epsilon.$$

□

2. Suppose that $\mathcal{F}_0$ is a field generating $\mathcal{F}$. Show that if μ is σ-finite on $\mathcal{F}_0$, then the function g in **Problem 1** can be taken to be of the form $\sum_i x_i I_{A_i}$ where $x_i \in \mathbb{R}$ and $A_i \in \mathcal{F}_0$.

Solution: From Problem **1**, there exists $h = \sum_{i=1}^K x_i I_{B_i}$ where $K \in \mathbb{N}$, $x_i \in \mathbb{R}$, $B_i \in \mathcal{F}$ for all i, with $\mu(B_i) < \infty$ for each i and $B_i \cap B_j = \emptyset$ whenever $i \neq j$, and

$$\int_\Omega |f - h| d\mu < \epsilon/2$$

Since μ is σ-finite on $\mathcal{F}_0$, and $\mu(B_i) < \infty$, for each i, there exists $A_i \in \mathcal{F}_0$ such that

$$\mu(B_i \Delta A_i) < \frac{\epsilon}{2K(x^* + 1)}$$

where $x^* = \max_{i \leq K} |x_i| < \infty$. (See Theorem 1.3.4 of A&L). Define $g = \sum_{i=1}^K x_i I_{A_i}$. Then g is an $\mathcal{F}_0$ measurable and integrable simple function since $x_i I_{A_i}$ is a $\mathcal{F}_0$-measurable and integrable simple function for each i. Then

$$\begin{aligned}
\int_\Omega |f - g| d\mu \leq & \int_\Omega |f - h| d\mu + \int_\Omega |h - g| d\mu < \epsilon/2 + \int_\Omega \left| \sum_{i=1}^K x_i (I_{B_i} - I_{A_i}) \right| d\mu \\
< & \epsilon/2 + \sum_{i=1}^K |x_i| \int_\Omega |I_{B_i} - I_{A_i}| d\mu = \epsilon/2 + \sum_{i=1}^K |x_i| \mu(B_i \Delta A_i) \\
< & \frac{\epsilon}{2} + \sum_{i=1}^K |x_i| \frac{\epsilon}{2K(x^* + 1)} < \epsilon.
\end{aligned}$$

□

3. Show by example that the conclusion to **Problem 2** may be false if μ is not σ-finite on $\mathcal{F}_0$.

Solution: Take $\Omega = (0, 1]$, $\mathcal{F} =$ Borel $\sigma-$field, $\mathcal{F}_0$ to be the field of all intervals contained in Ω, and μ to be the counting measure: $\mu(A) =$ cardinality of A. Let $f(\omega) = 1$ if $\omega = 0.5$, and 0 otherwise. Then f is Ω measurable and integrable with $\int_\Omega f d\mu = 1$. The function f itself is a simple function. But for any $\mathcal{F}_0$ measurable simple function g the only finite value $\int_\Omega g d\mu$ can take is 0. Thus a $\mathcal{F}_0$ measurable function satisfying Problem **2** cannot exist. □

Part II

Suppose $\mathcal{A}$ is the field of finite and cofinite subsets of an <u>infinite</u> set Ω. set

$$\nu(A) = \begin{cases} 0 & \text{if } A \text{ is finite} \\ 1 & \text{if } A^c \text{ is finite} \end{cases}$$

4. Show that ν is countably additive on $\mathcal{A}$ if and only if Ω is uncountable.

Solution:
<u>Only if part:</u> Suppose Ω is countably infinite: $\Omega = \{x_1, x_2, x_3, \ldots\}$. Let $A_i = \{x_i\}$. Thus $A_i \in \mathcal{A}$ for all i and are pairwise disjoint. Also, $\cup A_i = \Omega \in \mathcal{A}$. However, $\nu(\cup A_i) = 1$ but $\sum_i \nu(A_i) = \sum_i 0 = 0$ so that ν is not countably additive on $\mathcal{A}$.
<u>If part:</u> Suppose Ω is uncountable and let $\{A_i\}$ be a collection of pairwise disjoint sets in $\mathcal{A}$ such that $\cup A_i \in \mathcal{A}$.

If $\cup A_i$ is finite, then so are each of A_i. In this case, $\nu(\cup A_i) = 0 = \sum_i 0 = \sum_i \nu(A_i)$.

Otherwise $(\cup A_i)^c$ is finite so that $\nu(\cup A_i) = 1$. But then $\cup A_i$ must be uncountable since Ω is uncountable. Since countable union of finite sets cannot be uncountable, at least one A_i must be cofinite. We claim that only one A_i must be cofinite. To prove that, assume on the contrary that A_1 and A_2 are both cofinite (and also disjoint). Then $A_1^c \cup A_2^c$ is also finite. Then $A_1 \cap A_2$ is non-empty - a contradiction. Thus $\sum_i \nu(A_i) = 1$ since all but one term is zero and only one term is 1.

Thus ν is countably additive on $\mathcal{A}$. □

Part III

5. State Tonelli's theorem. *Hint*: It is the simpler version of Fubini's theorem for non-negative functions.

Solution: See Athreya and Lahiri, Theorem 5.2.1 □

6. Prove for a cdf F that $\int_{-\infty}^{\infty} \big(F(x + c) - F(x)\big) dx = c$ for all c.

Solution: First assume $c > 0$. Then by Tonelli's theorem,

$$\int_{-\infty}^{\infty} \big(F(x+c) - F(x)\big)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x < z \le x+c)dF(z)dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(z-c \le x < z)dxdF(z) = \int_{-\infty}^{\infty} cdF(z) = c(F(\infty) - F(-\infty)) = c.$$

Now if $c < 0$, then

$$\int_{-\infty}^{\infty} \big(F(x+c) - F(x)\big)dx = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x+c < z \le x)dF(z)dx$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(z \le x < z-c)dxdF(z) = -\int_{-\infty}^{\infty} (-c)dF(z) = c(F(\infty) - F(-\infty)) = c.$$

$\square$

Part IV

Let Ω be a non-empty set and $(\Omega, \mathcal{F}, P)$ be a probability space.

7. Suppose P is a probability measure on $(\Omega, \mathcal{F})$, $A \in \mathcal{F}$ and there exist sequences $\{B_n\}$ and $\{C_n\}$ of events such that B_n and C_n are independent for each n, and

$$\lim_{n\to\infty} P(A \,\Delta\, B_n) = \lim_{n\to\infty} P(A \,\Delta\, C_n) = 0.$$

Show that $P(A)$ is either 0 or 1.

Solution: Let $P(A) = p$. First, note that for any event U

$$P(U) = P(A) - P(A-U) + P(U-A).$$

Next note that, $P(B_n \Delta A) \to 0 \Rightarrow P(B_n - A) \to 0$ and $P(A - B_n) \to 0$. Hence,

$$P(B_n) = P(A) - P(A - B_n) + P(B_n - A) \to P(A) = p.$$

Similarly, $P(C_n - A) \to 0$ and $P(A - C_n) \to 0$, and hence

$$P(C_n) = P(A) - P(A - C_n) + P(C_n - A) \to P(A) = p.$$

Now

$$\begin{aligned} P(B_n \cap C_n) &= P(A) - P(A - (B_n \cap C_n)) + P((B_n \cap C_n) - A) \\ &= p - P((A \cap B_n^c) \cup (A \cap C_n^c)) + P((B_n \cap C_n) - A) \end{aligned}$$

However,

$$P((A \cap B_n^c) \cup (A \cap C_n^c)) \le P(A - B_n) + P(A - C_n) \to 0$$

and $P((B_n \cap C_n) - A) \le P(B_n - A) \to 0$. Thus, $P(B_n \cap C_n) \to p$. But since, B_n and C_n are independent for each n, $P(B_n \cap C_n) = P(B_n)P(C_n) \to p^2$. Hence $p = p^2$ so that $p = 0$ or 1. $\square$

8. Suppose $\{Y_i\}_{i\in I}$ is a family of random variables on an index set I.

a) Show that if $\int_0^\infty \sup_{i\in I} P(|Y_i| \ge t)dt < \infty$ then Y_is are uniformly integrable.

b) Show by a counterexample that the converse to **Problem 8a** is not necessarily true.

Solution:
a. We need to show

$$\lim_{x\to\infty}\sup_{i\in I}\int_{|Y_i|\ge x}|Y_i|dP = 0.$$

To that end, by Fubini's theorem,

$$\int_{|Y_i|\ge x}|Y_i|dP = \int_{|Y_i|\ge x}\int_0^{|Y_i|}dtdP = \int_0^\infty\int_{|Y_i|\ge t\vee x}dPdt = \int_0^\infty P(|Y_i|\ge t\vee x)dt.$$

So that

$$\sup_{i\in I}\int_{|Y_i|\ge x}|Y_i|dP = \sup_{i\in I}\int_0^\infty P(|Y_i|\ge t\vee x)dt \le \int_0^\infty \sup_{i\in I}P(|Y_i|\ge t\vee x)dt.$$

Now, since $\sup_{i\in I}P(|Y_i|\ge t)$ is a non-increasing function of t, and since $\int_0^\infty \sup_{i\in I}P(|Y_i|\ge t)dt<\infty$, it must be that $\sup_{i\in I}P(|Y_i|\ge t)\to 0$ as $t\to\infty$. Thus for each t, $\sup_{i\in I}P(|Y_i|\ge t\vee x)\to 0$ as $x\to\infty$.. Hence the result follows from the dominated convergence theorem.

b. Consider $I=[3,\infty)$ and Y_i take the value i with probability $1/(i\log i)$ and 0 with probability $1-1/(i\log i)$. Then, Y_i's are U.I. since

$$\lim_{x\to\infty}\sup_{i\in I}\int_{|Y_i|>x}|Y_i|dP = \lim_{x\to\infty}x/(x\log x) = 0$$

but

$$\int_0^\infty \sup_{i\in I}P(|Y_i|>t)dt \ge \int_3^\infty dt/(t\log t) = \infty.$$

□

Part V

Suppose $X_1, X_2, \ldots, X_n$ are iid with pdf

$$f(x;\theta) = \exp\big(-(x-\theta)\big)I(x>\theta)$$

9. Show that the cdf of M_n is given by

$$G(x) = \begin{cases} 0 & \text{if } x<\theta \\ 1-\exp(-n(x-\theta)) & \text{if } x>\theta \end{cases}$$

Solution: Note that the cdf of X_is is given by

$$F(x;\theta) = \begin{cases} 0 & \text{if } x<\theta \\ 1-\exp(-(x-\theta) & \text{if } x\ge\theta. \end{cases}$$

The result is obvious for $x<\theta$. For $x\ge\theta$, Since $X_1, X_2, \ldots, X_n$ are iid

$$P(M_n>x) = P(X_1>x)^n = \exp(-n(x-\theta)),$$

so that the result follows. This says that $M_n-\theta\sim$ Exponential distribution with rate n, □

10. Show that $M_n \xrightarrow{P} \theta$.

Solution: For any $\epsilon > 0$, note that using the cdf from Problem **10**, we have

$$P(|M_n - \theta| > \epsilon) = P(M_n > \theta + \epsilon) = \exp(-(n\epsilon)) \to 0$$

as $n \to \infty$. $\square$

11. Suppose $c > \theta$ is fixed. Consider the random variable:

$$U_n = \begin{cases} \exp(M_n - c) & \text{if } M_n < c \\ 1 & \text{if } M_n \geq c. \end{cases}$$

Find sequences (a_n) and (b_n) of real numbers (possibly also depending on c and θ), such that $L_n := a_n(U_n - b_n)$ converges in distribution to a random variable L having Exponential distribution with mean $\exp(\theta - c)$.

Solution: Since $c > \theta$, by continuous mapping theorem, $U_n \downarrow \exp(\theta - c)$ in probability. Thus $b_n = \exp(\theta - c)$. We will show that,

$$P(n(U_n - b_n) \leq x) \to 1 - \exp\left(-xe^{c-\theta}\right)$$

for $x \in (0, \infty)$. To that end, first note that,

$$P(n(U_n - b_n) \leq x) = P(n(U_n - b_n) \leq x, M_n > c) + P(n(U_n - b_n) \leq x, M_n < c).$$

Since the first term on the right side at most $P(M_n > c) \to 0$ as $M_n \to \theta$, we focus on the second term. First consider n sufficiently large so that, $\log(1 + xe^{c-\theta}/n) < c - \theta$ which is possible since the left side goes to zero as $n \to \infty$. For all such n,

$$\begin{aligned} P(n(U_n - b_n) \leq x, M_n < c) &= P(M_n \leq \theta + \log(1 + xe^{c-\theta}/n),\ M_n < c) \\ &= P(M_n \leq \theta + \log(1 + xe^{c-\theta}/n)) \\ &= 1 - \exp\left(-n\log(1 + xe^{c-\theta}/n)\right) \\ &= 1 - \left(1 + \frac{xe^{c-\theta}}{n}\right)^{-n} \to 1 - \exp\left(-xe^{c-\theta}\right) \end{aligned}$$

Hence, $L_n := n\left(U_n - \exp(\theta - c)\right) \to L \sim$ Exponential distribution with mean $\exp(\theta - c)$. $\square$

12. Show that $E(L_n^k) \to E(L^k)$ as $n \to \infty$ for all positive integers k, where L_n and L are as defined in **Problem 11**.

Solution: Let $\lambda = \exp(c - \theta)$. Then,

$$EL^k = \int_0^\infty x^k \lambda e^{-\lambda x} dx = \lambda\Gamma(k+1)/\lambda^{k+1} = k!/\lambda^k.$$

Now note that, $M_n - \theta \sim$ Exponential distribution with rate n. Thus for all $n \geq k$

$$\begin{aligned}
EL_n^k &= En^k(e^{M_n - c} - e^{c-\theta})^k = n^k e^{k(c-\theta)} E(e^{M_n-\theta} - 1)^k \\
&= n^k \lambda^{-k} \int_0^\infty (e^z - 1)^k n e^{-nz} dz \\
&= n^{k+1} \lambda^{-k} \int_0^1 (1/u - 1)^k u^n (1/u) du \qquad \text{with } u = e^{-z} \\
&= n^{k+1} \lambda^{-k} \int_0^1 (1-u)^k u^{n-k-1} du \\
&= n^{k+1} \lambda^{-k} \frac{\Gamma(k+1)\Gamma(n-k)}{\Gamma(n+1)} = n^{k+1} \lambda^{-k} \frac{k!(n-k-1)!}{n!} \\
&= k! \lambda^{-k} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k}{n}\right) \to k!/\lambda^k,
\end{aligned}$$

since $k \in \mathbb{N}$ is fixed. □

Part VI

13. Suppose $\{Y_n\}$ is a sequence of independent random variables with distribution functions F_n. Show that

$$Y_n \to 0 \text{ a.s. if and only if } \sum_{n=1}^\infty \{1 - F_n(\epsilon) + F_n(-\epsilon)\} < \infty, \ \forall \epsilon > 0.$$

Solution: Note that $1 - F_n(\epsilon) + F_n(-\epsilon) = P(\{Y_n \leq -\epsilon\} \cup \{Y_n > \epsilon\})$. This problem is a simple application of the two Borel-Cantelli lemmas.

If part: For each $j \in \mathbb{N}$, define $A_{n,j} = P(\{Y_n \leq -1/j\} \cup \{Y_n > 1/j\})$, $A_j = \limsup A_{n,j}$ and $A = \cup_{j \in \mathbb{N}} A_j$. Thus from the given condition, for each j, (with $\epsilon = 1/j$), we have $\sum_n P(A_{n,j}) < \infty$, so that By Borel-Cantelli lemma, $P(A_j) = 0$. Hence $P(A) \leq \sum_j P(A_j) = 0$.

If $w \notin A$, then $w \notin A_j$ for any j, hence $Y_n(w) < 1/j$ for sufficiently large n. Since $j \in \mathbb{N}$ is arbitrary, this is equivalent to saying $Y_n \to 0$ on A^c.

Only if part: Assume there exists $\epsilon > 0$ such that $\sum_{n=1}^\infty \{1 - F_n(\epsilon) + F_n(-\epsilon)\} = \infty$, i.e., $\sum_n P(\{Y_n \leq -\epsilon\} \cup \{Y_n > \epsilon\}) = \infty$. Since Y_ns are independent, we have from the second Borel-Cantelli lemma, that, $P(\limsup_n(\{Y_n \leq -\epsilon\} \cup \{Y_n > \epsilon\})) = 1$. In particular, $|Y_n| \geq \epsilon$ infinitely often almost surely. Hence $Y_n \not\to 0$ a.s. □

14. Use the CLT to prove that

$$\lim_{n\to\infty} \sum_{k=n}^\infty e^{-n} \frac{n^k}{k!} = \frac{1}{2}.$$

Solution: The sum in the left hand side is the probability that a Poisson random variable with mean n is bigger than or equal to n. So let us assume $X_1, X_2, \ldots, X_n$ are iid Poisson(1) random variables and let $S_n = \sum_{i=1}^n X_i$. Then $S_n \sim$ Poisson(n) and hence

$$\lim_{n\to\infty} \sum_{k=n}^{\infty} e^{-n} \frac{n^k}{k!} = P(S_n \geq n) = P\left(\frac{S_n - n}{\sqrt{n}} \geq 0\right) \to 1 - \Phi(0) = 1/2$$

by the CLT. □