Review Session 6 – Solutions

1 Solutions

1.1 Previous Core Competency Problems

Problem 1 (May 2018, # 3)**.** Let $W_1, W_2, \ldots, W_k$ be unbiased estimators of a parameter θ with $\mathrm{Var}(W_i) = \sigma_i^2$ and $\mathrm{Cov}(W_i, W_j) = 0$ if $i \neq j$.

(a) Show that among all estimators of the form $\sum_{i=1}^k a_i W_i$, where a_i's are constants and $\mathbb{E}_\theta(\sum_i a_i W_i) = \theta$, the estimator $W^* = \frac{\sum_i W_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$ has minimum variance.

(b) Show that $\mathrm{Var}(W^*) = \frac{1}{\sum_i 1/\sigma_i^2}$.

Solution

It's straightforward to compute $\mathrm{Var}(W^*) = \frac{1}{\sum_i 1/\sigma_i^2}$. By Cauchy-Schwarz, we have

$$1 = \left|\sum_{i=1}^k a_i\right|^2 = \left|\sum_{i=1}^k (a_i\sigma_i)\left(\frac{1}{\sigma_i}\right)\right|^2 \leq \left|\sum_{i=1}^k a_i^2\sigma_i^2\right|\left|\sum_{i=1}^k \frac{1}{\sigma_i^2}\right| \implies \frac{1}{\sum_{i=1}^k \frac{1}{\sigma_i^2}} \leq \sum_{i=1}^k a_i^2\sigma_i^2 = \mathrm{Var}\left(\sum_{i=1}^k a_i W_i\right)$$

Thus, W^* minimizes the variance among all estimators of the form $\sum_{i=1}^k a_i W_i$.

Problem 2 (May 2018, # 6)**.** Consider observed response variables $Y_1, \ldots, Y_n \in \mathbb{R}$ that depend linearly on covariates $x_1, \ldots, x_n$ as follows:

$$Y_i = \beta x_i + \epsilon_i, \text{ for } i = 1, \ldots, n.$$

Here, the ϵ_i's are independent Gaussian noise variables, but we do not assume they have the same variance. Instead, they are distributed as $\epsilon_i \sim N(0, \sigma_i^2)$ for possibly different variances $\sigma_1^2, \ldots, \sigma_n^2$. The unknown parameter of interest is β.

(a) Suppose that the error variances $\sigma_1^2, \ldots, \sigma_n^2$ are all known. Find the MLE $\hat{\beta}$ for β in this case and derive an explicit formula for $\hat{\beta}$. Show that $\hat{\beta}$ minimizes a certain weighted least-squares criterion.

(b) Show that the estimate $\hat{\beta}$ in part (a) is unbiased, and derive a formula for the variance of $\hat{\beta}$ in terms of $\sigma_1^2, \ldots, \sigma_n^2$ and $x_1, \ldots, x_n$.

(c) Compute the Fisher information $I(\beta)$ in this model (still assuming $\sigma_1^2, \ldots, \sigma_n^2$ are known constants). Compare this value with the variance of $\hat{\beta}$ derived in part (b).

Solution

Let **Y** $= (Y_1, \ldots, Y_n)$. We have log-likelihood

$$L(\beta|\mathbf{Y}, \sigma_1^2, \ldots, \sigma_n^2) = \frac{1}{(2\pi)^{n/2}\prod \sigma_i} \exp\left(-\sum \frac{(y_i - \beta x_i)^2}{2\sigma_i^2}\right) \implies$$

$$\log L(\beta|\mathbf{Y}, \sigma_1^2, \ldots, \sigma_n^2) = \log C - \sum_{i=1}^n \frac{(Y_i - \beta x_i)^2}{2\sigma_i^2}$$

where C does not depend on β. Thus, the least-squares criterion to be minimized is then

$$\sum_{i=1}^{n} \frac{(Y_i - \beta x_i)^2}{2\sigma_i^2}$$

and this has solution at

$$\frac{\partial}{\partial \beta} \log L(\beta|\mathbf{Y}, \sigma_1^2, \ldots, \sigma_n^2) = \sum \frac{-2x_iY_i + 2\beta x_i^2}{2\sigma_i^2} = 0 \implies \hat{\beta} = \frac{\sum x_iY_i/\sigma_i^2}{\sum x_i^2/\sigma_i^2}.$$

Showing $\mathbb{E}[\hat{\beta}] = \beta$ is straightforward and we have

$$\text{Var}(\hat{\beta}) = \frac{1}{\sum x_i^2/\sigma_i^2}$$

We get $I(\beta) = \sum x_i^2/\sigma_i^2$. We have $I(\beta)^{-1} = \text{Var}(\hat{\beta})$ meaning $\hat{\beta}$ satisfies the Cramer-Rao lower bound.

Problem 3 (May 2018, # 7). Suppose that $X \sim$ Poisson(λ) and its parameter $\lambda > 0$ has a prior distribution Gamma(α, β) given by density

$$f(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-y\beta} y^{\alpha-1}, \text{ for } y \geq 0, \text{ (and 0 otherwise).}$$

(a) Find the posterior distribution of λ given the observation X, and identify the distribution with its parameters.

(b) Find the mean of this posterior distribution.

Solution

(a) It's straightforward to show $\pi(\lambda|X) \sim$ Gamma$(x + \alpha, \beta + 1)$.

(b) Since the posterior is a gamma, $\mathbb{E}[\lambda|X] = \frac{x+\alpha}{\beta+1}$. To prove this, we have

$$\begin{aligned}
\mathbb{E}[\lambda|X] &= \int_0^\infty \frac{(\beta+1)^{x+\alpha}}{\Gamma(x+\alpha)} e^{-\lambda(\beta+1)} \lambda^{x+\alpha} \, d\lambda \\
&= \int_0^\infty \frac{\Gamma(x+\alpha+1)}{\Gamma(x+\alpha)(\beta+1)} f_{\Gamma, x+\alpha+1, \beta+1}(\lambda) \, d\lambda \\
&= \frac{\Gamma(x+\alpha+1)}{(\beta+1)\Gamma(x+\alpha)} \\
&= \frac{x+\alpha}{\beta+1}
\end{aligned}$$

where we make use of the fact that $\Gamma(k+\alpha+1) = (k+\alpha)\Gamma(k+\alpha)$ and $f_{\Gamma,k+\alpha+1,\beta}$ denotes the pdf of a Gamma$(k+\alpha+1, \beta+1)$ distribution.

Problem 4 (May 2018, # 8). Suppose $X_1, X_2 \overset{i.i.d.}{\sim} Ber(p)$ for some unknown parameter $p \in (0, 1)$. Find an unbiased estimator for the following functions of p, if there exists one.

(a) $g(p) = 2p$.

(b) $g(p) = p(1-p)$.

(c) $g(p) = p^2$.

(d) $g(p) = p^3$.

Solution

(a) $X_1 + X_2$

(b) $S^2 := \frac{1}{n-1}\sum_{i=1}^{2}(X_i - \overline{X}_2)^2$

(c) $\overline{X}_2 - S^2$

(d) We claim no unbiased estimator exists. For contradiction, suppose $U(X_1, X_2)$ is an unbiased estimator of p^3. Then,

$$\mathbb{E}[U(X_1, X_2)] = p^3 \implies (1-p)p(U(1,0) + U(0,1)) + (1-p)^2 U(0,0) + p^2 U(1,1) = p^3.$$

However, the RHS is a polynomial of degree 3 in p, while the LHS is of degree 2, meaning the above cannot be satisfied for all $p \in [0,1]$ for any choice of $\{U(0,1), U(1,0), U(0,0), U(1,1)\}$.

Problem 5 (September 2019, # 7)**.** Suppose that $X_1, \ldots, X_n$ are i.i.d. uniform random variables on $[0, \theta]$ for some $\theta \in [1,2]$.

(i) What is the MLE of θ?

(ii) Suppose that, instead of X_i's, we only observe, for all $i = 1, \ldots, n$,

$$Y_i = \begin{cases} X_i & \text{if } X_i \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

What is the MLE of θ based on $\{Y_1, \ldots, Y_n\}$?

Solution

(i) We have the joint likelihood is $L(\theta | X_1, \ldots, X_n) = 1/\theta^n \cdot \mathbf{1}\{\max_i X_i \le \theta\}$. This is maximized over $\theta \in [1,2]$ at $\hat{\theta} := \max\left(1, \max_{i=1,\ldots,n} X_i\right)$.

(ii) We have the likelihood function of θ based on a single Y_1 is

$$L(\theta | Y_1) = \frac{1}{\theta} \cdot \mathbf{1}\{Y_i \in (0,1]\} + \frac{\theta - 1}{\theta} \cdot \mathbf{1}\{Y_i = 0\}.$$

Thus, the joint likelihood is

$$L(\theta | Y_1, \ldots, Y_n) = \left(\frac{1}{\theta}\right)^m \left(1 - \frac{1}{\theta}\right)^{n-m},$$

where $m := \sum_{i=1}^n \mathbf{1}\{Y_i > 0\}$ is the number of Y_i's which are positive. We claim the above is maximized at $\hat{\theta} = \min(\max(n/m, 1), 2)$ (i.e., round n/m to the nearest point in $[1,2]$).

To see this, we can first recognize the term $(1/\theta)^m (1 - 1/\theta)^{n-m}$ as a binomial pdf with parameter $1/\theta$. Thus, it must be maximized at $\theta^* = n/m$. In fact, θ^* is the only stationary point of the function

$$\theta \mapsto (1/\theta)^m (1 - 1/\theta)^{n-m},$$

and this function is increasing for $\theta < \theta^*$ and decreasing for $\theta > \theta^*$. Thus, the constrained maximum $\max_{\theta \in [1,2]} (1/\theta)^m (1 - 1/\theta)^{n-m}$ must be achieved at n/m if $n/m \in [1,2]$ or else the nearest boundary point among $\{1, 2\}$ to n/m.

Problem 6 (September 2019, # 8)**.** Suppose that a measurement Y is recorded with a $N(\theta, \sigma^2)$ sampling distribution, with σ known and θ known to lie in the interval $[0,1]$ (but otherwise unknown). Consider two point estimators of θ: (a) the posterior mean $\hat{\theta}_B$ based on the assumption of a uniform prior distribution on θ on $[0,1]$, and (b) the maximum likelihood estimate $\hat{\theta}_M$, restricted to the range $[0,1]$.

(i) Show that, as $\sigma \to \infty$, $\hat{\theta}_B$ converges in distribution (to Y_1, say). Identify the limit Y_1. [**Hint:** You may first find the distribution of $\Theta | Y = y$ and then take limits.]

(ii) Show that, as $\sigma \to \infty$, $\hat{\theta}_M$ converges in distribution (to Y_2, say). Identify the limit Y_2.

(iii) If σ is large enough, which estimator $\hat{\theta}_M$ or $\hat{\theta}_B$ has a higher mean squared error, for any value of θ in $[0,1]$. You may answer this question by comparing the mean squared errors of Y_1 and Y_2 for estimating θ.

Solution

(i) With uniform prior $\pi_\theta \sim$ Unif$([0,1])$, we have posterior

$$\pi(\theta|Y) \propto L(Y|\theta) \cdot \pi(\theta) = \exp\left(-\frac{(Y-\theta)^2}{2\sigma^2}\right) \cdot \mathbf{1}\{\theta \in [0,1]\} \cdot \frac{1}{\sqrt{2\pi\sigma^2}}.$$

To compute the limiting distribution, let $x \in [0,1]$ and let $Z \sim \mathcal{N}(0,1)$. Then, the CDF of $\theta|Y$ is

$$\begin{aligned}
\mathbb{P}(\theta \le x|Y) &= \mathbb{P}(0 \le \mathcal{N}(Y,\sigma^2) \le x | 0 \le \mathcal{N}(Y,\sigma^2) \le 1) \\
&= \mathbb{P}(0 \le \sigma \cdot Z + Y \le x | 0 \le \sigma \cdot Z + Y \le 1) \\
&= \mathbb{P}(-Y/\sigma \le Z \le (x-Y)/\sigma | -Y/\sigma \le Z \le (1-Y)/\sigma) \\
&= \frac{\mathbb{P}(-Y/\sigma \le Z \le (x-Y)/\sigma)}{\mathbb{P}(-Y/\sigma \le Z \le (1-Y)/\sigma)}.
\end{aligned}$$

Then, letting Φ, ϕ be the standard normal cdf and pdf, respectively, we have the above is

$$\frac{\Phi\left(\frac{x-Y}{\sigma}\right) - \Phi\left(-\frac{Y}{\sigma}\right)}{\Phi\left(\frac{1-Y}{\sigma}\right) - \Phi\left(-\frac{Y}{\sigma}\right)}.$$

To take limit as $\sigma \to \infty$, we use l'Hôpital's rule:

$$\begin{aligned}
\lim_{\sigma\to\infty} \frac{\Phi\left(\frac{x-Y}{\sigma}\right) - \Phi\left(-\frac{Y}{\sigma}\right)}{\Phi\left(\frac{1-Y}{\sigma}\right) - \Phi\left(-\frac{Y}{\sigma}\right)} &= \lim_{\sigma\to\infty} \frac{-\left(\frac{X-Y}{\sigma^2}\right)\phi\left(\frac{x-Y}{\sigma}\right) - \left(\frac{Y}{\sigma^2}\right)\phi\left(-\frac{Y}{\sigma}\right)}{-\left(\frac{1-Y}{\sigma^2}\right)\phi\left(\frac{1-Y}{\sigma}\right) - \frac{Y}{\sigma^2}\phi\left(-\frac{Y}{\sigma}\right)} \\
&= \lim_{\sigma\to\infty} \frac{-(x-Y)\phi\left(\frac{x-Y}{\sigma}\right) - Y\phi\left(-\frac{Y}{\sigma}\right)}{-(1-Y)\phi\left(\frac{1-Y}{\sigma}\right) - Y\phi\left(-\frac{Y}{\sigma}\right)} \\
&= \frac{-x\phi(0)}{-\phi(0)} \\
&= x.
\end{aligned}$$

Thus, $\theta|Y \xrightarrow{\text{d}}$ Unif$([0,1])$ meaning by bounded convergence theorem, $\hat{\theta}_B := \mathbb{E}[\theta|Y] \xrightarrow{\sigma\to\infty} 1/2$.

(ii) The MLE is seen to be $\hat{\theta}_M = \max(\min(Y,1),0)$ (i.e., round Y to the nearest point of $[0,1]$). Next, we compute the limit of the cdf $\mathbb{P}(\hat{\theta}_M \le x)$. If $x < 0$, this is clearly 0 and if $x \ge 1$, this is 1. For $x \in [0,1)$, we have

$$\mathbb{P}(\hat{\theta}_M \le x) = \mathbb{P}(Y \le x) = \mathbb{P}(\sigma \cdot Z + \theta \le x) = \mathbb{P}\left(Z \le \frac{x-\theta}{\sigma}\right) \xrightarrow{\sigma\to\infty} \mathbb{P}(Z \le 0) = 1/2.$$

Thus, we have $\hat{\theta}_M \xrightarrow{\sigma\to\infty}$ Ber$(1/2)$.

(iii) The MSE of $\mathbb{E}[Y_1|Y] = 1/2$ is $(1/2-\theta)^2$, while the MSE of $Y_2 \sim$ Ber$(1/2)$ is $\frac{1}{2}(1-\theta)^2 + \frac{1}{2}\theta^2$. Straightforward algebra shows Y_1 has lower MSE for all $\theta \in [0,1]$;

Problem 7 (May 2020, # 6)**.** Suppose that we have single observation from X from the exponential distribution with parameter λ. Define $T(X) = I(X > 1)$, where I is the indicator function. Set $\psi(\lambda) := e^{-\lambda}$.

(i) Show that $T(X)$ is unbiased for $\psi(\lambda)$.

(ii) Find the (Fisher) information bound for unbiased estimators of $\psi(\lambda)$.

(iii) Show that the variance of $T(X)$ is strictly larger than the information bound.

Solution

(i) We have $\mathbb{P}(X > 1) = 1 - (1 - e^{-\lambda}) = e^{-\lambda}$ so that $T(X)$ is unbiased for $\psi(\lambda)$.

(ii) The Fisher information $I(\lambda)$ w.r.t. λ can be computed as $1/\lambda^2$ and then the Fisher information w.r.t. $\psi(\lambda)$ is, by chain rule:

$$I(\psi) = I(\lambda) \cdot \left(\frac{\partial \lambda}{\partial \psi}\right)^2 = \frac{e^{2\lambda}}{\lambda^2}.$$

Thus, the CR lower bound is $\lambda^2/e^{2\lambda}$

(iii) $\mathrm{Var}(T(X)) = \psi(\lambda) - \psi(\lambda)^2 = e^{-\lambda} - e^{-2\lambda}$. Thus, we wish to show

$$e^{-\lambda} - e^{-2\lambda} > \frac{\lambda^2}{e^{2\lambda}} \iff e^{\lambda} > \lambda^2 + 1.$$

However, we claim the last inequality is always true for $\lambda > 0$. Taking a power series expansion of e^{λ}, we have

$$e^{\lambda} > 1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} > \lambda^2 + 1 \iff \frac{\lambda^2}{6} - \frac{\lambda}{2} + 1 > 0.$$

However, the quadratic $\lambda^2/6 - \lambda/2 + 1$ has a negative discriminant meaning it is is always positive. Thus, the variance of $T(X)$ is strictly larger than the information bound.

Problem 8 (September 2020, # 5)**.** Consider the following Bayesian model

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Uniform}([0, \theta]) \text{ and } \theta \sim \text{Pareto}(\beta, \lambda)$$

where the pdf of the Pareto distribution is given by

$$\pi(\theta; \beta, \lambda) = \frac{\beta \lambda^{\beta}}{\theta^{(\beta+1)}}, \quad \theta > \lambda, \quad \beta, \lambda > 0.$$

Moreover, for this exercise you may assume $\beta > 1$.

(a) Use the Bayes formula to derive the posterior density of θ as explicitly as possible.

(b) Compute the prior and posterior means of θ.

Solution

(a) The posterior of θ is given by

$$L(\theta | X_1, \ldots, X_n) \propto \mathbf{1}\{\theta > \max(\lambda, X_{(n)})\} \cdot \theta^{-(n+\beta+1)}.$$

Thus, $\theta | X_1, \ldots, X_n \sim$ Pareto$(n + \beta, \max(\lambda, X_{(n)}))$.

(b)

$$\mathbb{E}_{\pi}[\theta] = \int_{\lambda}^{\infty} \frac{\beta \lambda^{\beta}}{\theta^{\beta}}\, d\theta = \frac{\lambda \beta}{\beta - 1}$$

$$\mathbb{E}[\theta | X] = \max(\lambda, X_{(n)}) \cdot \frac{n + \beta}{n + \beta - 1},$$

where we use the first formula to conclude the second formula.

Problem 9 (May 2021, # 1)**.** Let $X_1, \ldots, X_n$ be an i.i.d. random sample with common density function

$$f(x) = \begin{cases} 3\theta^3 x^{-4} & \text{for } x \geq \theta \\ 0 & \text{otherwise} \end{cases},$$

where $\theta > 0$ is an unknown parameter.

(i) Apply the method of moments to obtain an unbiased estimator of θ.

(ii) Find the maximum likelihood estimator (MLE) of θ and show that it is biased.

(iii) Which of the above two estimators has a smaller mean squared error (MSE)?

Solution

(i) We have

$$\mathbb{E}[X] = \int_0^\infty 3\theta^3 x^{-3}\, dx = \frac{3}{2}\theta.$$

Thus, $\hat{\theta}_{\mathsf{MOM}} = \overline{X}_n \cdot \frac{2}{3}$.

(ii) The likelihood function is

$$L(\theta|X_1, \dots, X_n) = \frac{3^n \cdot \theta^{3n}}{\prod_{i=1}^n X_i^4} \mathbf{1}\{X_{(1)} \geq \theta\},$$

which is maximized at $\hat{\theta}_{\mathsf{MLE}} = X_{(1)}$. Next, to simplify the computation of the bias, we'll use the *tail probability formula for expectations*: for a nonnegative random variable $Y \geq 0$:

$$\mathbb{E}[Y] = \int_0^\infty \mathbb{P}(Y > t)\, dt.$$

Using this formula, we have

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_{\mathsf{MLE}}] &= \int_0^\infty \mathbb{P}(X_{(1)} > t)\, dt \\
&= \int_0^\theta 1\, dt + \int_\theta^\infty \mathbb{P}(X_{(1)} > t)\, dt \\
&= \theta + 3^n \theta^{3n} \int_\theta^\infty \left(\int_t^\infty \frac{1}{x^4}\, dx \right)^n dt \\
&= \theta + 3^n \theta^{3n} \int_\theta^\infty \frac{t^{-3n}}{3^n}\, dt \\
&= \theta + \theta^{3n} \left(\frac{\theta^{-3n+1}}{-3n+1} \right) \\
&= \theta \cdot \left(1 + \frac{1}{-3n+1} \right).
\end{aligned}$$

Thus, $\hat{\theta}_{\mathsf{MLE}}$ is biased.

(iii) We have

$$\mathbb{E}[X^2] = \int_\theta^\infty 3\theta^3 x^{-2}\, dx = 3\theta^2.$$

Thus, $\mathrm{Var}(X) = 3\theta^2 - \frac{9}{4}\theta^2 = \frac{3}{4}\theta^2$ and $\mathrm{Var}(\overline{X}_n) = \frac{3}{4}\theta^2/n$. Thus, by a bias-variance decomposition:

$$\mathbb{E}[(\hat{\theta}_{\mathsf{MOM}} - \theta)^2] = \mathrm{Var}(\hat{\theta}_{\mathsf{MOM}}) = \frac{1}{3}\theta^2/n.$$

Next, for the sake of using a bias-variance decomposition, we first compute $\mathbb{E}[\hat{\theta}^2_{\mathsf{MLE}}]$ and then $\mathrm{Var}(\hat{\theta}_{\mathsf{MLE}})$. We have, again using the tail probability formula for expectations,

$$\begin{aligned}
\mathbb{E}[\hat{\theta}^2_{\mathsf{MLE}}] &= \int_0^\infty \mathbb{P}(X_{(1)} > \sqrt{t})\,dt \\
&= \theta^2 + \int_{\theta^2}^\infty \mathbb{P}(X_{(1)} > \sqrt{t})\,dt \\
&= \theta^2 + \int_{\theta^2}^\infty 3^n\theta^{3n}\left(\int_{\sqrt{t}}^\infty \frac{1}{x^4}\,dx\right)^n dt \\
&= \theta^2 + 3^n\theta^{3n}\int_{\theta^2}^\infty \frac{t^{-3n/2}}{3^n}\,dt \\
&= \theta^2 + \theta^{3n}\frac{\theta^{(-3n/2+1)*2}}{-3n/2+1} \\
&= \theta^2\left(1 + \frac{1}{-3n/2+1}\right).
\end{aligned}$$

Thus,

$$\mathrm{Var}(\hat{\theta}_{\mathsf{MLE}}) = \mathbb{E}[\hat{\theta}^2_{\mathsf{MLE}}] - \mathbb{E}[\hat{\theta}_{\mathsf{MLE}}]^2 = \theta^2\left(1 + \frac{1}{-3n/2+1} - \left(1 + \frac{1}{-3n+1}\right)^2\right).$$

Thus, the MSE of $\hat{\theta}_{\mathsf{MLE}}$ is

$$\mathbb{E}[(\hat{\theta}_{\mathsf{MLE}} - \theta)^2] = \left(\frac{1}{-3n+1}\right)^2\theta^2 + \mathrm{Var}(\hat{\theta}_{\mathsf{MLE}}) = \theta^2\left(-\frac{2}{-3n+1} + \frac{1}{-3n/2+1}\right).$$

We claim $\hat{\theta}_{\mathsf{MLE}}$ has a smaller MSE. This follows from:

$$\frac{1}{3n} > -\frac{2}{-3n+1} + \frac{1}{-3n/2+1} \iff 9n^2 > 3n-2,$$

where the last inequality is always true.

Problem 10 (September 2021, # 1)**.** Let $X_1, \ldots, X_n$ be an i.i.d. sample with common density

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(i) Find a one dimensional sufficient statistic T_n.

(ii) Derive the cumulative distribution function F_n of T_n.

(iii) Give an exact $(1-\alpha)$-confidence interval for θ. (Hint: What is the distribution of $F_n(T_n)$?).

Solution

(i) $X_{(1)}$ is sufficient based on computing the joint pdf

$$\prod_{i=1}^n f(x_i;\theta) = e^{-\sum_i x_i + n\theta}\cdot\mathbf{1}\{x_{(1)} \geq \theta\}.$$

(ii) We have

$$\mathbb{P}(X_{(1)} > c) = \left(\int_c^\infty e^{-x+\theta}\cdot\mathbf{1}\{x \geq \theta\}\,dx\right)^n = e^{n\theta - n(c\vee\theta)},$$

so that $F_n(c) = 1 - e^{n\theta - n(c\vee\theta)}$.

(iii) We have $n(\theta - X_{(1)} \vee \theta)$ is a $\log(\mathrm{Unif}([0,1]))$ random variable so that if $[L_\alpha, U_\alpha]$ is an interval such that $n(\theta - X_{(1)} \vee \theta) \in [L_\alpha, U_\alpha]$ w.p. $1-\alpha$, then $\theta \in \left(X_{(1)} + \frac{\log(L_\alpha)}{n}, X_{(1)} + \frac{\log(U_\alpha)}{n}\right)$ (note we get rid of the maximum in $X_{(1)} \vee \theta$ since $X_{(1)} \geq \theta$ w.p. 1).

Problem 11 (September 2021, # 2). Let X and Y be two independent exponential random variables with parameters λ and μ, respectively, i.e. $\mathbb{P}(X \geq x, Y \geq y) = e^{-\lambda x - \mu y}$, $x \geq 0$, $y \geq 0$. Define random variables

$$T = \min(X, Y) \text{ and } \Delta = \begin{cases} 1 & X < Y \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the probability density function of T and the probability mass function of Δ.

(ii) Find the joint distribution function of (T, Δ).

(iii) Suppose we have a random sample (T_i, Δ_i), $i = 1, \ldots, n$, i.e. i.i.d. copies of (T, Δ). Write down the likelihood function and find the MLE of λ.

Solution

(i) We have

$$\mathbb{P}(T \geq x) = e^{-x(\lambda+\mu)} \implies f_T(x) = \frac{\partial}{\partial x} 1 - e^{-x(\lambda+\mu)} = (\lambda+\mu)e^{-x(\lambda+\mu)},$$

and

$$\mathbb{P}(\Delta = 1) = \mathbb{P}(X < Y) = \mathbb{E}_X[\mathbb{P}(X < Y|X)] = \int_0^\infty \mathbb{P}(Y > x) \cdot f_X(x)\, dx = \int_0^\infty e^{-\mu x} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda+\mu}.$$

(ii) We have

$$\mathbb{P}(T \leq t, \Delta \leq d) = \begin{cases} \mathbb{P}(T \leq t) = 1 - e^{-t(\lambda+\mu)} & d = 1 \\ \mathbb{P}(T \leq t, X \geq Y) = \mathbb{E}_Y[\mathbb{P}(X \geq Y|Y) \cdot \mathbf{1}\{Y \leq t\}] = \int_0^t \mathbb{P}(X > y) f_Y(y)\, dy & d = 0 \end{cases}.$$

This last integral is $\frac{\mu}{\lambda+\mu}\left(1 - e^{-t(\lambda+\mu)}\right)$.

(iii) First, we see from the part (ii) that

$$\mathbb{P}(T \leq t, X < Y) = \mathbb{P}(T \leq t) - \mathbb{P}(T \leq t, X \geq Y) = \frac{\lambda}{\lambda+\mu} - e^{-t(\lambda+\mu)} + \frac{\mu}{\lambda+\mu} e^{-t(\lambda+\mu)}.$$

Taking the derivative with respect to t then gives us "joint likelihood" $L(t, 1) = \lambda e^{-t(\lambda+\mu)}$. Note that "joint likelihood" here involves both continuous and discrete distributions. The derivation above (although not entirely rigorous) essentially first takes the discrete derivative with respect to Δ of the joint distribution function from (ii) and then the continuous derivative with respect to t.

Similarly, the joint likelihood $L(t, 0) = \mu e^{-(\lambda+\mu)t}$. Thus, the likelihood function is

$$L(\{(T_i, \Delta_i)\}_{i=1}^n) = \prod_{i=1}^n e^{-T_i(\lambda+\mu)} (\mu \cdot \mathbf{1}\{\Delta_i = 0\} + \lambda \cdot \mathbf{1}\{\Delta_i = 1\}) = \lambda^{\sum_{i=1}^n \Delta_i} \mu^{n - \sum_{i=1}^n \Delta_i} \exp\left(-(\lambda+\mu)\sum_{i=1}^n T_i\right).$$

Taking $\log$, the part that depends on λ in the above is

$$\log(\lambda) \sum_{i=1}^n \Delta_i - \lambda \sum_{i=1}^n T_i,$$

so that the MLE is $\lambda_{\text{MLE}} = \frac{\sum_i \Delta_i}{\sum_i T_i}$.

1.2 Additional Practice

Problem 12 (Casella & Berger, Exercise 7.12). Let $X_1, \dots, X_n$ be a random sample from a population with pmf

$$\mathbb{P}_\theta(X = x) = \theta^x (1-\theta)^{1-x}, x = 0 \text{ or } 1, 0 \le \theta \le \frac{1}{2}.$$

(i) Find the method of moments estimator and MLE of θ.

(ii) Find the mean squared errors of each of the estimators.

(iii) Which estimator is preferred? Justify your choice.

Solution

(i) The method of moments estimator is obtained from setting

$$\mathbb{E}[X] = \theta = \overline{X}_n \implies \hat{\theta}_{\text{MOM}} = \overline{X}_n.$$

For the MLE, the log-likelihood is

$$L(\theta | X_1, \dots, X_n) := \left(\sum_{i=1}^n X_i\right) \log(\theta) + \left(n - \sum_{i=1}^n X_i\right) \log(1-\theta).$$

The above is increasing in θ for $\theta \le \overline{X}_n$ and decreasing in θ for $\theta \ge \overline{X}_n$. If $\overline{X}_n > 1/2$, then $L(\theta|X_1, \dots, X_n)$ is increasing for $\theta \in [0, 1/2]$ and is thus maximized at $\theta = 1/2$. Otherwise, we have $L(\theta|X_1, \dots, X_n)$ is maximized at $\theta = \overline{X}_n$. Thus, $\hat{\theta}_{\text{MLE}} = \min\{\overline{X}_n, 1/2\}$.

(ii) The MSE of $\hat{\theta}_{\text{MOM}}$ is

$$\text{MSE}(\hat{\theta}_{\text{MOM}}) = \text{Var}(\hat{\theta}_{\text{MOM}}) + \text{Bias}(\hat{\theta}_{\text{MOM}})^2 = (\theta(1-\theta)/n) + 0^2 = \theta(1-\theta)/n.$$

The MSE of $\hat{\theta}_{\text{MLE}}$ is

$$\begin{aligned}\text{MSE}(\hat{\theta}_{\text{MLE}}) &= \sum_{y=0}^n (\hat{\theta}_{\text{MLE}} - \theta)^2 \binom{n}{y} \theta^t (1-\theta)^{n-y} \\ &= \sum_{y=0}^{\lfloor n/2 \rfloor} \left(\frac{y}{n} - \theta\right)^2 \binom{n}{y} \theta^y (1-\theta)^{n-y} + \sum_{y=\lfloor n/2 \rfloor + 1}^n \left(\frac{1}{2} - \theta\right)^2 \binom{n}{y} \theta^y (1-\theta)^{n-y}.\end{aligned}$$

(iii) We first note that

$$\text{MSE}(\hat{\theta}_{\text{MOM}}) = \text{Var}(\overline{X}_n) = \sum_{y=0}^n \left(\frac{y}{n} - \theta\right)^2 \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

Thus,

$$\begin{aligned}\text{MSE}(\hat{\theta}_{\text{MLE}}) - \text{MSE}(\hat{\theta}_{\text{MOM}}) &= \sum_{y=\lfloor n/2 \rfloor + 1}^n \left(\left(\frac{y}{n} - \theta\right)^2 - \left(\frac{1}{2} - \theta\right)^2\right) \binom{n}{y} \theta^t (1-\theta)^{n-y} \\ &= \sum_{\lfloor n/2 \rfloor + 1}^n \left(\frac{y}{n} + \frac{1}{2} - 2\theta\right) \left(\frac{y}{n} - \frac{1}{2}\right) \binom{n}{y} \theta^t (1-\theta)^{n-y}.\end{aligned}$$

Noting that $y/n > 1/2$ and $\theta \le 1/2$ inside the sum, we have that every summand of the above is positive for $\theta > 0$. Thus, $\text{MSE}(\hat{\theta}_{\text{MOM}}) \le \text{MSE}(\hat{\theta}_{\text{MLE}})$ where the inequality is strict for $\theta \in (0, 1/2]$.