Review Session 8 – Solutions

1 Solutions

1.1 Previous Core Competency Problems

Problem 1 (2018 Summer Practice, # 2)**.** Suppose that $X_1, \ldots, X_n$ are i.i.d. $\exp(1/\mu)$, where $\mathbb{E}(X_1) = \mu > 0$.

(i) Find the mean and variance of $\overline{X}_n = \sum_{i=1}^n X_i/n$. Hence, find the asymptotic distribution of $\overline{X}_n$ (properly standardized).

(ii) Let $T = \log \overline{X}_n$. Find the corresponding asymptotic distribution of T (properly standardized).

(iii) How can the asymptotic distribution of T be used to construct an approximate $(1-\alpha)$ confidence interval (CI) for μ? Explain your answer and give the desired CI.

Solution

We have $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathrm{Var}(\overline{X}_n) = \frac{\mu^2}{n}$. Standardized, this is

$$\frac{\overline{X}_n - \mu}{\mu/\sqrt{n}} = \sqrt{n}\left(\frac{\overline{X}_n}{\mu} - 1\right) \xrightarrow{\text{d}} N(0,1)$$

by CLT so that $\frac{\log(\overline{X}_n) - \log(\mu)}{\mu/\sqrt{n}} \xrightarrow{\text{d}} N(0, 1/\mu^2)$ by delta method. This gives approximate $(1-\alpha)$ CI for $\log \mu$

$$\log(\overline{X}_n) - z_{1-\alpha/2}/\sqrt{n} < \log \mu < \log(\overline{X}_n) + z_{1-\alpha/2}/\sqrt{n}$$

whence we can construct a $(1-\alpha)$ CI for μ by taking $\exp(\cdot)$ of both sides.

Problem 2 (2018 Summer Practice, # 5)**.** Suppose that $Y_1, \ldots, Y_n$ are i.i.d Poisson(λ), $\lambda > 0$ unknown. Assume that n is even, i.e., $n = 2k$ for some integer k. Consider

$$\hat{\lambda} = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

(a) Is $\hat{\lambda}$ an unbiased estimator of λ (show your steps)?

(b) Is $\hat{\lambda}$ a consistent estimator of λ, as $k \to \infty$ (show your steps)?

Solution

$\hat{\lambda}$ is unbiased since

$$\mathbb{E}[\hat{\lambda}] = \frac{1}{2k}\sum_{i=1}^{k}\mathbb{E}[Y_{2i}^2] - 2\mathbb{E}[Y_{2i}Y_{2i-1}] + \mathbb{E}[Y_{2i-1}^2] = \frac{1}{2k}\sum_{i=1}^{k}2(\lambda + \lambda^2) - 2\lambda^2 = \lambda$$

It is also consistent since for some function $f(\lambda)$,

$$\mathrm{Var}(\hat{\lambda}) = \frac{1}{4k^2}\sum_{i=1}^{k}\mathrm{Var}(Y_{2i}^2) + 4\,\mathrm{Var}(Y_{2i}Y_{2i-1}) + \mathrm{Var}(Y_{2i-1}^2) = \frac{1}{4k}f(\lambda) \xrightarrow{k\to\infty} 0$$

Problem 3 (2018 September, # 6)**.** Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\theta, 1)$, where $\theta \in \mathbb{R}$ is unknown. Let $\psi = \mathbb{P}_\theta(X_1 > 0)$.

(a) Find the maximum likelihood estimator $\hat{\psi}$ of ψ.

(b) Find an approximate 95% confidence interval for ψ.

(c) Let $Y_i = \mathbf{1}\{X_i > 0\}$, for $i = 1, \dots, n$. Define $\tilde{\psi} = (1/n)\sum_{i=1}^n Y_i$. Show that $\tilde{\psi}$ is a consistent estimator of ψ.

(d) Find the asymptotic distribution of both the estimators. Which estimator of ψ, $\hat{\psi}$ or $\tilde{\psi}$, is more preferable in this model and why?

Solution

(a) Let $\Phi(\cdot)$ be the standard normal cdf. Then, $\psi = 1 - \Phi(-\theta)$ so that the MLE of ψ is $1 - \Phi(-\hat{\theta})$ by functional invariance of the MLE.

(b) We have $\overline{X} \sim \mathcal{N}(\theta, 1/n)$ so that w.p. at least $1-\alpha$:

$$\frac{z_{\alpha/2}}{\sqrt{n}} - \overline{X} < -\theta < \frac{z_{1-\alpha/2}}{\sqrt{n}} - \overline{X}$$

Since $\Phi(\cdot)$ is monotone, this implies

$$1 - \Phi\left(\frac{z_{\alpha/2}}{\sqrt{n}} - \overline{X}\right) > 1 - \Phi(-\hat{\theta}) > 1 - \Phi\left(\frac{z_{1-\alpha/2}}{\sqrt{n}} - \overline{X}\right)$$

which is a $1-\alpha$ C.I. for ψ.

(c) This is just Law of Large Numbers.

(d) From CLT, we have $\sqrt{n}(\tilde{\psi} - \psi) \xrightarrow{\text{d}} N(0, \sigma^2)$ where the asymptotic variance σ^2 of $\tilde{\psi}$ is

$$\mathbb{E}[\mathbf{1}\{X_i > 0\}] - \mathbb{E}[\mathbf{1}\{X_i > 0\}]^2 = (1 - \Phi(-\theta)) - (1 - \Phi(-\theta))^2 = \Phi(-\theta)(1 - \Phi(-\theta)).$$

The asymptotic variance of $1 - \Phi(-\hat{\theta})$ is $\phi(-\theta)^2$ by delta method where $\phi(\cdot)$ is the standard normal pdf. Observe at $\theta = 0$ that

$$\frac{1}{2\pi} = \phi(0)^2 < \frac{1}{4} = (1 - \Phi(0)) \cdot \Phi(0)$$

We claim this is true for general θ as well. By symmetry of the transformation $\theta \mapsto -\theta$ it suffices to show $\phi(\theta)^2 < (1 - \Phi(\theta)) \cdot \Phi(\theta)$ for positive θ. Define $G(\theta) := (1 - \Phi(\theta)) \cdot \Phi(\theta) - \phi(\theta)^2$, which is just the difference between the two asymptotic variances. We clearly have $\lim_{\theta \to \infty} G(\theta) = 0$. Then, it suffices to show G is strictly decreasing on $(0, \infty)$, which will imply $G(\theta) > 0$ for all $\theta > 0$. We have

$$\frac{\partial}{\partial \theta} G(\theta) = \phi(\theta) - 2\phi(\theta) \cdot \Phi(\theta) + 2\theta \cdot \phi(\theta)^2,$$

where we substituted $-\theta \cdot \phi(\theta)$ for $\phi'(\theta)$ (which is a well-known identity for the standard normal pdf, and can be verified by computation). Then,

$$G'(\theta) < 0 \iff \phi(\theta) - 2\phi(\theta) \cdot \Phi(\theta) + 2\theta \cdot \phi(\theta)^2 < 0 \iff 1/2 < -\theta \cdot \phi(\theta) + \Phi(\theta)$$

Note that the RHS of the last inequality is equal to $1/2$ when $\theta = 0$. Thus, we can show this last inequality by taking the derivative of the RHS $-\theta \cdot \phi(\theta) + \Phi(\theta)$ and showing that it is positive.

$$\frac{\partial}{\partial \theta} - \theta \cdot \phi(\theta) + \Phi(\theta) = -\phi(\theta) - \theta\phi'(\theta) + \phi(\theta) = -\theta\phi'(\theta) = \theta^2\phi(\theta) > 0,$$

where we again use the identity $\phi'(\theta) = -\theta \cdot \phi(\theta)$. Thus, $G'(\theta) < 0$ for all $\theta > 0$ meaning $G(\theta)$ is positive on all of $(0, \infty)$.

Problem 4 (2019 May, # 3). n_1 people are given treatment 1 and n_2 people are given treatment 2. Let X_1 be the number of people on treatment 1 who respond favorably to the treatment and let X_2 be the number of people on treatment 2 who respond favorably. Assume that $X_1 \sim$ Binomial(n_1, p_1), and $X_2 \sim$ Binomial(n_2, p_2). Let $\psi = p_1 - p_2$.

(i) Find the maximum likelihood estimator $\hat{\psi}$ of ψ.

(ii) Find the Fisher information matrix $I(p_1, p_2)$.

(iii) Use the delta method to find the asymptotic standard error of $\hat{\psi}$.

Solution

(i) The MLE of p_1 is X_1/n_1 so that by functional invariance of MLE, $\hat{\psi} = \frac{X_1}{n_1} - \frac{X_2}{n_2}$.

(ii) The log-likelihood is (up to terms not containing p_1, p_2):

$$X_1 \log(p_1) + (n_1 - X_1)\log(1-p_1) + X_2 \log(p_2) + (n_2 - X_2)\log(1-p_2).$$

Next, we consider the second-order partials of this expression. The mixed partials clearly vanish. And we have

$$-\mathbb{E}\left[\frac{\partial^2}{\partial^2 p_1^2} X_1 \log(p_1) + (n_1 - X_1)\log(1-p_1)\right] = \mathbb{E}\left[\frac{X_1}{p_1^2} + \frac{n_1 - X_1}{(1-p_1)^2}\right] = \frac{n_1}{p_1} + \frac{n_1}{1-p_1}.$$

By symmetry, the univariate Fisher information w.r.t. p_2 is $\frac{n_2}{p_2} + \frac{n_2}{1-p_2}$. Thus,

$$I(p_1, p_2) = \begin{bmatrix} \frac{n_1}{p_1} + \frac{n_1}{1-p_1} & 0 \\ 0 & \frac{n_2}{p_2} + \frac{n_2}{1-p_2} \end{bmatrix}.$$

(iii) By multivariate CLT we have for $n_1 = n_2 = n$,

$$\sqrt{n}\left(\begin{pmatrix} X_1/n \\ X_2/n \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}\right) \xrightarrow{\text{d}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} p_1(1-p_1) & 0 \\ 0 & p_2(1-p_2) \end{pmatrix}\right).$$

Next, The function $g(p_1, p_2) = p_1 - p_2$ has gradient $\nabla g = (1, -1)$. Then, by the delta method, the asymptotic standard error of $\hat{\psi}$ will be

$$(\nabla g)^T \begin{pmatrix} p_1(1-p_1) & 0 \\ 0 & p_2(1-p_2) \end{pmatrix} (\nabla g) = p_1(1-p_1) + p_2(1-p_2).$$

> **Remark 1.1.** We could have also appealed to Cramer's theorem instead of multivariate CLT as the asymptotic covariance matrix is the inverse of $I(p_1, p_2)$.

Problem 5 (2019 May, # 6)**.** Denote by $\hat{\zeta}_n$ the MLE of $\zeta = p(1-p)$ based on n i.i.d. samples from Binomial$(1, p)$. Denote by p_0 the true value of p.

(a) If $p_0 \neq 1/2$, find the limiting (non-degenerate) distribution of $\hat{\zeta}_n$, with proper normalization.

(b) Derive the asymptotic distribution of $\hat{\zeta}_n$, when $p_0 = 1/2$.

Solution

(i) Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim}$ Binomial$(1, p)$. By functional invariance of MLE, $\hat{\zeta}_n = \overline{X}_n \cdot (1 - \overline{X}_n)$. Then, since $g(p) = p(1-p)$ has derivative $g'(p) = 1 - 2p$ and $\sqrt{n}(\overline{X}_n - p) \xrightarrow{\text{d}} N(0,1)$, by the delta method,

$$\sqrt{n}(\hat{\zeta} - p(1-p)) \xrightarrow{\text{d}} N(0, p(1-p) \cdot (1-2p)^2).$$

(ii) Since $g''(p) = -2 \neq 0$, second-order delta method gives $n(\hat{\zeta}_n - 1/4) \xrightarrow{\text{d}} -\frac{1}{4} \cdot \chi_1^2$.

Problem 6 (2019 September, # 3)**.** Suppose that X_n and Y_n are independent random variables, where X_n is asymptotically normal with mean 4 and standard deviation $1/\sqrt{n}$ (i.e., $\sqrt{n}(X_n - 4) \xrightarrow{\text{d}} N(0,1)$) and Y_n is asymptotically normal with mean 3 and standard deviation $2/\sqrt{n}$. Use the delta method to get an approximate mean and standard deviation of Y_n/X_n.

Solution

Since X_n, Y_n are independent, we have

$$\sqrt{n}\left(\begin{pmatrix} X_n \\ Y_n \end{pmatrix} - \begin{pmatrix} 4 \\ 3 \end{pmatrix}\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\mathbf{0}_2, \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}\right).$$

Let $g(x, y) = y/x$ which has gradient $\nabla g(x, y) = (-y/x^2, 1/x)$. Then, the asymptotic variance of $g(X_n, Y_n)$ is

$$\nabla g(4,3)^T \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \nabla g(4,3) = (3/16)^2 + 1/4 = 73/256.$$

Thus $\sqrt{n}(Y_n/X_n - 3/4) \xrightarrow{\mathrm{d}} \mathcal{N}(0, 73/256)$ by delta method.

Problem 7 (2019 September, # 5)**.** Let $X_1, \ldots, X_n$ be the number of accidents at an important intersection in the past n years. We are interested in estimating the probability of zero accidents next year. We model the X_i's as independent random variables distributed according to a Poisson distribution with mean λ.

(i) Let $q(\lambda)$ be the probability that there will be no accidents next year. Find an unbiased and consistent estimator of $q(\lambda)$.

(ii) Compute the maximum likelihood estimator of $q(\lambda)$ and derive its asymptotic distribution. Compare this estimator with the one obtained in (i).

Solution

(i) $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i = 0\}$ is clearly unbiased and consistent by WLLN.

(ii) $q(\lambda) = e^{-\lambda}$ so by functional invariance of MLE the MLE of $q(\lambda)$ is $\exp(-\overline{X}_n)$. By delta method,

$$\sqrt{n}(\exp(-\overline{X}_n) - \exp(-\lambda)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \lambda e^{-2\lambda}).$$

The asymptotic variance of the estimator from (i) is λ by CLT which is always larger than $\lambda e^{-2\lambda}$. Thus, from the perspective of comparing asymptotic variances, the MLE is better.

Problem 8 (2020 May, # 8)**.** Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p) random sample, i.e. $P(X_i = 1) = 1 - P(X_i = 0) = p$, $p \in (0, 1)$. Further, let $\theta = \mathrm{Var}(X_i)$.

(i) Find $\hat{\theta}$, the maximum likelihood estimator of θ.

(ii) Show that $\hat{\theta}$ is asymptotically normal when $p \neq 1/2$ and give the asymptotic variance.

(iii) When $p = 1/2$, derive a non-degenerated asymptotic distribution of $\hat{\theta}$ with an appropriate normalization. *Hint: try relating $\hat{\theta}$ to the statistic $(\overline{X}_n - 1/2)^2$.*

Solution

(i) By functional invariance of MLE, $\hat{\theta} = \overline{X}_n \cdot (1 - \overline{X}_n)$.

(ii) By delta method, $\hat{\theta}$ is asymptotically normal with asymptotic variance $p(1-p) \cdot (1-2p)^2$.

(iii) Second-order delta method gives $n(\overline{X}_n \cdot (1 - \overline{X}_n) - 1/4) \xrightarrow{\mathrm{d}} -\frac{1}{4} \cdot \chi_1^2$, which is derived by taking Taylor expansion of the function $g(\overline{X}_n) := \overline{X}_n \cdot (1 - \overline{X}_n)$ and then taking limits giving us (as suggested in the hint) the limit of $n(\overline{X}_n - 1/2)^2$ which is $\frac{1}{4} \cdot \chi_1^2$ by CLT.

Problem 9 (2020 May, # 9)**.** Let $X_1, \ldots, X_{2n}$ be an i.i.d. random sample with common pdf $f(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda} x}$ for $x > 0$. Consider the three estimators $\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\hat{\lambda}_2 = \frac{1}{n}\sum_{i=n+1}^{2n} X_i$, and $\hat{\lambda} = \frac{1}{2n}\sum_{i=1}^{2n} X_i$.

(i) Show that $T_1 = \hat{\lambda}_1 \hat{\lambda}_2$ is an unbiased and consistent estimator of λ^2.

(ii) Show that $T_2 = \hat{\lambda}^2$ is consistent for λ^2, but not unbiased.

(iii) Derive the asymptotic distribution of the estimators T_1 and T_2. Which one is more efficient asymptotically?

Solution

(i) The common pdf is exponential with mean λ. Thus, $T_1 = \hat{\lambda}_1 \hat{\lambda}_2$ is an unbiased (by the independence of $\hat{\lambda}_1$ and $\hat{\lambda}_2$) and consistent (by WLLN) estimator of λ^2.

(ii) By WLLN $T_2 = \hat{\lambda}^2$ is consistent. But, it's biased as

$$\mathbb{E}[\hat{\lambda}^2] = \frac{1}{4n^2} \sum_{i,j=1}^{(2n)^2} \mathbb{E}[X_i X_j] = \frac{1}{4n^2} \left((4n^2 - 2n) \cdot \lambda^2 + 2n \cdot (\lambda^2 + \lambda) \right) = \lambda^2 + \frac{1}{2n} \cdot \lambda.$$

(iii) Multiviariate delta method gives

$$\sqrt{2n}(T_1 - \lambda^2) \xrightarrow{\text{d}} \mathcal{N}(0, 2\lambda^4).$$

On the other hand, ordinary delta method gives for T_2,

$$\sqrt{2n}(T_2 - \lambda^2) \xrightarrow{\text{d}} \mathcal{N}(0, 4\lambda^4).$$

Thus, purely in terms of asymptotic variance, T_1 is more efficient.

Problem 10 (2020 September, # 6). Suppose $(X_1, \ldots, X_n)$ are i.i.d. from a Normal distribution with $\mathbb{E}X_i = \mathrm{Var}(X_i) = \theta$, where $\theta > 0$ is unknown.

(a) Find the MLE for θ explicitly.

(b) Find the asymptotic distribution of your MLE.

Solution

(a) The log-likelihood is

$$L(\theta) = -\frac{n}{2} \log(2\pi\theta) - \frac{1}{2\theta} \sum_{i=1}^{n} (X_i - \theta)^2,$$

which has derivative

$$L'(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{n} X_i^2 - \frac{n}{2}.$$

Setting $L'(\theta) = 0$, we get the quadratic equation $\theta^2 + \theta = T_n$ with $T_n := n^{-1} \sum_{i=1}^{n} X_i^2$. This has nonnegative root

$$\hat{\theta} = \frac{-1 + \sqrt{1 + 4T_n}}{2}.$$

It's straightforward to further verify that $L''(\hat{\theta}) < 0$.

(b) We proceed by delta method. We first have $\mathbb{E}[X_1^2] = \theta^2 + \theta$ and $\mathrm{Var}(X_1^2) = 4\theta^3 + 2\theta^2$. Thus, by CLT

$$\sqrt{n}(T_n - \theta^2 - \theta) \xrightarrow{\text{d}} \mathcal{N}(0, 4\theta^3 + 2\theta^2).$$

Let $g(t) := \frac{-1+\sqrt{1+4t}}{2}$ for $t > 0$, which satisfies $g'(\theta^2 + \theta) = \frac{1}{2\theta+1}$. Thus, delta method gives

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{4\theta^3 + 3\theta^2}{(1 + 2\theta)^2}\right) = \mathcal{N}\left(0, \frac{2\theta^2}{1 + 2\theta}\right).$$

> **Remark 1.2.** Since in part (a), you had to find the second derivative $L''(\theta)$, it could be faster to compute the Fisher information which is
> $$I_n(\theta) = \frac{n(2\theta+1)}{2\theta^2}.$$
> Then, the asymptotic normality of MLE's (noting all standard regularity conditions hold here) directly gives us the same convergence as above since $I_1(\theta)^{-1} = 2\theta^2/(2\theta+1)$.

Problem 11 (2021 May, # 3)**.** A random sample $X_1, \dots, X_n$ is drawn from a population with p.d.f.

$$f_\theta(x) = \frac{1}{2}(1+\theta x), x \in [-1,1],$$

and $f_\theta(x) = 0$ if $x \notin [-1,1]$, where $\theta \in [-1,1]$ is the unknown parameter.

1. Find an unbiased estimator of θ.
2. Is the estimator in (i) consistent? Provide a justification for your answer.

Solution

(i) $\mathbb{E}[X] = \int_{-1}^{1} \frac{1}{2}(1+\theta x) \cdot x \, dx = \frac{\theta}{3}$. Thus, $\hat{\theta} := 3 \cdot \overline{X}_n$ is an unbiased estimator of θ.

(ii) Yes, $\hat{\theta}$ is consistent by LLN.

Problem 12 (2021 May, # 5)**.** Let X and Y be a pair of random variables with the following distributional specification. $P(Y = 1) = 1 - P(Y = 0) = \alpha$ where $\alpha \in (0,1)$ and $X|Y = 0 \sim N(0, \sigma^2)$ and $X|Y = 1 \sim N(\mu, \sigma^2)$.

1. Find the conditional distribution of Y given X, i.e. $P(Y = 1|X = x)$.
2. Suppose that we have an i.i.d. random sample from this population, i.e. we observe i.i.d. copies (X_i, Y_i), $i = 1, \dots, n$. Write down the likelihood function and find maximum likelihood estimators $\hat{\alpha}_n, \hat{\mu}_n$ and $\hat{\sigma}_n^2$ of α, μ, and σ^2.
3. What are the asymptotic distributions of $\hat{\alpha}_n, \hat{\mu}_n$, and $\hat{\sigma}_n^2$ (properly standardized)?

Solution

(i) We have

$$\begin{aligned}\mathbb{P}(Y=1|X\in[x-h,x+h]) &= \frac{\mathbb{P}(X\in[x-h,x+h]\cap Y=1)}{\mathbb{P}(X\in[x-h,x+h])} \\ &= \frac{\mathbb{P}(X\in[x-h,x+h]|Y=1)\cdot\mathbb{P}(Y=1)}{\mathbb{P}(X\in[x-h,x+h]|Y=1)\cdot\mathbb{P}(Y=1)+\mathbb{P}(X\in[x-h,x+h]|Y=0)\cdot\mathbb{P}(Y=0)}.\end{aligned}$$

Dividing the numerator and denominator on the RHS above by $2h$ and taking the limit as $h \to 0$, we obtain

$$\mathbb{P}(Y=1|X=x) = \frac{\alpha\cdot\exp(-(x-\mu)^2/(2\sigma^2))}{\alpha\cdot\exp(-(x-\mu)^2/(2\sigma^2))+(1-\alpha)\cdot\exp(-x^2/(2\sigma^2))} = \frac{\alpha}{\alpha+(1-\alpha)\exp(x\mu/\sigma^2-\mu^2/(2\sigma^2))}.$$

Thus, $Y|X = x$ is a Bernoulli with the above parameter.

(ii) We have likelihood

$$\begin{aligned}L(\alpha,\mu,\sigma^2|X,Y) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu\cdot Y_i)^2}{2\sigma^2}} \cdot \alpha^{Y_i}(1-\alpha)^{1-Y_i} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu Y_i)^2\right)\cdot\alpha^{\sum_{i=1}^n Y_i}(1-\alpha)^{n-\sum_{i=1}^n Y_i} \implies \\ \log(L) &\propto -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu Y_i)^2 + \left(\sum_{i=1}^n Y_i\right)\log(\alpha) + \left(n-\sum_{i=1}^n Y_i\right)\log(1-\alpha).\end{aligned}$$

We see that for any value of μ, σ^2, the log-likelihood is maximized in terms of α at $\hat{\alpha} := \frac{1}{n}\sum_{i=1}^n Y_i$. Similarly, for any values of σ^2, α, the log-likelihood is maximized in terms of μ at $\hat{\mu} = \frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n Y_i^2}$. Then, the log-likelihood is further maximized in terms of σ^2 at $\hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu}\cdot Y_i)^2$.

(iii) We have

(a) $\frac{\sqrt{n}(\hat{\alpha}_n - \alpha)}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{d} \mathcal{N}(0,1)$ by CLT.

(b) By WLLN, we have $\hat{\alpha}_n \xrightarrow{d} \alpha$. Next, by Slutsky and another WLLN, we have since $\frac{1}{n}\sum_{i=1}^n Y_i^2 = \frac{1}{n}\sum_{i=1}^n Y_i = \hat{\alpha}_n$,

$$\hat{\mu}_n = \frac{\frac{1}{n}\sum_{i=1}^n Y_i \cdot X_i}{\frac{1}{n}\sum_{i=1}^n Y_i^2} \xrightarrow{d} \frac{1}{\alpha}\cdot \mathbb{E}[Y_1 \cdot X_1] = \mu.$$

Next, we have

$$\mathrm{Var}(Y_1 \cdot X_1) = \mathbb{E}[Y_1^2 X_1^2] - (\mathbb{E}[Y_1 X_1])^2 = \alpha(\sigma^2 + \mu^2) - \alpha^2\mu^2 = \alpha\cdot\sigma^2 + \alpha\mu^2 - \alpha^2\mu^2.$$

So, CLT on the i.i.d. sum $\sum_{i=1}^n Y_i X_i$ gives us

$$\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i\cdot X_i - \alpha\cdot\mu\right)}{\sqrt{\mathrm{Var}(Y_1\cdot X_1)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Then, by Slutsky, we have

$$\frac{\sqrt{n}\left(\hat{\mu}_n - \mu\right)}{\sqrt{\mathrm{Var}(Y_1\cdot X_1)/\alpha}} \xrightarrow{d} \mathcal{N}(0,1).$$

(c) First, it is clear that the value of $\hat{\sigma}_n^2$ does not depend on μ since it is location-invariant. Thus, let us assume WLOG that $\mu = 0$. Now, we have

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n}\sum_{i=1}^n \hat{\mu}_n \cdot X_i\cdot Y_i + \frac{\hat{\mu}_n^2}{n}\sum_{i=1}^n Y_i^2 \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{1}{n\sum_{i=1}^n Y_i^2}\left(\sum_{i=1}^n X_i\cdot Y_i\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{\left(\frac{1}{n}\sum_{i=1}^n X_i\cdot Y_i\right)^2}{\hat{\alpha}_n}
\end{aligned}$$

We will use multivariate delta method to proceed. The above RHS is a function of the vector $\left(\frac{1}{n}\sum_{i=1}^n X_i^2, \frac{1}{n}\sum_{i=1}^n X_i Y_i, \frac{1}{n}\sum_{i=1}^n Y_i\right)$. This vector has expectation:

$$\begin{aligned}
\mathbb{E}[X^2] &= (1-\alpha)\cdot\sigma^2 + \alpha\cdot(\sigma^2+\mu^2) = \sigma^2 + \alpha\mu^2 = \sigma^2 \\
\mathbb{E}[XY] &= \alpha\cdot\mu = 0 \\
\mathbb{E}[Y] &= \alpha.
\end{aligned}$$

Note that we simplified the first two expectations above by assuming $\mu = 0$. Next, let the function g be defined by

$$g(a,b,c) = a - \frac{b^2}{c}.$$

Then, g has first-order partials at the vector $(a,b,c) = (\sigma^2, 0, \alpha)$:

$$\begin{aligned}
\frac{\partial}{\partial a} g(a,b,c) &= 1 \\
\frac{\partial}{\partial b} g(a,b,c)|_{a=\sigma^2, b=0, c=\alpha} &= 0 \\
\frac{\partial}{\partial c} g(a,b,c)|_{a=\sigma^2, b=0, c=\alpha} &= 0.
\end{aligned}$$

Then, we will have by delta method

$$\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2, \frac{1}{n}\sum_{i=1}^{n}X_iY_i, \frac{1}{n}\sum_{i=1}^{n}Y_i\right) - g(\sigma^2, 0, \alpha)\right) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \tau^2),$$

where we note $g(\sigma^2, 0, \alpha) = \sigma^2$ and where (because we already established that two of the first-order partials of g vanish):

$$\tau^2 := \mathrm{Var}(X^2) \cdot \left(\frac{\partial}{\partial a} g(a, b, c)|_{(a,b,c)=(\sigma^2,0,\alpha)}\right)^2 = \mathrm{Var}(X^2).$$

Thus, it remains to compute $\mathrm{Var}(X^2)$. We have, again using the fact that $\mu = 0$ so that $X \sim \mathcal{N}(0, \sigma^2)$ unconditional of Y,

$$\mathrm{Var}(X^2) = \mathbb{E}[X^4] - \mathbb{E}[X^2]^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4.$$

Thus, $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathsf{d}} \mathcal{N}(0, 2\sigma^4)$.

Problem 13 (2021 May, # 6)**.** Suppose $X_1, \ldots, X_n$ are independent, with $X_i \sim N\left(\frac{\theta}{i}, 1\right)$. Here, $\theta \in \mathbb{R}$ is an unknown parameter.

(i) Find an unbiased estimator $\hat{\theta}_n$ for θ which depends on the entire data.

(ii) Find asymptotic non-degenerate distribution of your estimator, i.e. $d_n(\hat{\theta}_n - \theta)$ converges to a non-degenerate distribution.

(iii) Suppose that we impose a normal prior $\theta \sim N(0, \tau)$, where $\tau > 0$ is an known constant. Find the posterior distribution of θ given data $X_1, \ldots, X_n$.

Solution

(i) Since $\mathbb{E}[X_i] = \theta/i$, we have $\hat{\theta}_n := \sum_{i=1}^{n} i \cdot X_i/n$ is an unbiased estimator for θ.

(ii) We have $\hat{\theta}_n \sim \mathcal{N}(\theta, \sum_{i=1}^{n} i^2/n^2)$. Thus,

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\sum_{i=1}^{n} i^2/n^2}} \xrightarrow{\mathsf{d}} \mathcal{N}(0, 1).$$

(iii) We have

$$\begin{aligned}\pi(\theta|X_1, \ldots, X_n) &\propto \pi(\theta)L(X_1, \ldots, X_n|\theta) \\ &\propto e^{-\theta^2/(2\tau)} \cdot \exp\left(-\frac{\sum_{i=1}^{n}(X_i - \theta/i)^2}{2}\right) \\ &\propto \exp\left(-\frac{\theta^2}{2}\left(\frac{1}{\tau} + \sum_{i=1}^{n}\frac{1}{i^2}\right) + \theta\sum_{i=1}^{n} X_i/i\right).\end{aligned}$$

Completing the square in terms of θ on the RHS, we have

$$\theta|X_1, \ldots, X_n \sim \mathcal{N}\left(\frac{\sum_{i=1}^{n} X_i/i}{1/\tau + \sum_{i=1}^{n} 1/i^2}, \left(1/\tau + \sum_{i=1}^{n} 1/i^2\right)^{-1}\right).$$

Problem 14 (2021 Sept, # 4)**.** Suppose $\{U_i\}_{i\geq 1} \overset{\text{i.i.d.}}{\sim} U(0, \theta)$, for some $\theta > 0$.

1. Show that $T_n := \left(\prod_{i=1}^{n} U_i\right)^{1/n}$ converges in probability to a constant, and find this constant.
2. Find a function of T_n that is a consistent estimator for θ.
3. Find constants a_n and b_n such that $a_n(T_n - b_n)$ converges to a non-degenerate distribution.

Solution

(1) We have $\log(T_n) = \frac{1}{n}\sum_{i=1}^n \log(U_i)$ which goes to $\mathbb{E}[\log(U_i)]$ by law of large numbers. More explicitly, this expectation is

$$\int_0^\theta \frac{\log(u)}{\theta}\,du = \frac{u\log(u) - u}{\theta}\bigg|_0^\theta = \log(\theta) - 1.$$

Thus, $T_n \xrightarrow{\mathsf{P}} \exp(\log(\theta) - 1)$.

(2) By inverting the formula in (1), we have $\exp(\log(T_n) + 1)$ is a consistent estimator for θ.

(3) Since $\sqrt{n}(\log(T_n) - (\log(\theta) - 1))$ goes to a normal distribution by CLT, we have by Delta method that

$$\sqrt{n}(T_n - \exp(\log(\theta) - 1)),$$

goes to a non-degenerate distribution where we know the variance term is positive since $\frac{\partial}{\partial\theta}\exp(\log(\theta) - 1) > 0$ for all $\theta > 0$.