DISCUSSION WORKSHEET 06

PSTAT 120B: Mathematical Statistics, I
Summer Session A, 2024 with Instructor: Ethan P. Marzban

Conceptual Review

(a) What is a **likelihood**? What about a **log-likelihood**?

(b) How do we obtain a **maximum likelihood estimator** for a parameter? What do we do if the likelihood is nondifferentiable in the parameter of interest?

(c) What is a **sufficient statistic**? How can the **factorization theorem** help us find a statistic that is sufficient for a given parameter?

Problem 1

Let $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where both $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown parameters.

(a) Derive an expression for $\mathcal{L}_{\vec{\mathbf{Y}}}(\mu, \sigma^2)$, the likelihood of the sample $\vec{\boldsymbol{Y}}$. Recall that, since our sample is assumed to be i.i.d.,

$$\mathcal{L}_{\vec{\mathbf{Y}}}(\mu, \sigma^2) = \prod_{i=1}^{n} f_{Y_i}(y_i; \mu, \sigma^2)$$

Solution:

$$\begin{aligned} \mathcal{L}_{\vec{\mathbf{Y}}}(\mu, \sigma^2) &= \prod_{i=1}^{n} f(Y_i; \mu, \sigma^2) \\ &= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(Y_i - \mu)^2 \right\} \right] \\ &= \left(\frac{1}{2\pi} \right)^{n/2} \cdot (\sigma^2)^{-n/2} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \mu)^2 \right\} \end{aligned}$$

(b) Derive an expression for $\ell_{\vec{\mathbf{Y}}}(\mu, \sigma^2)$, the log-likelihood of the sample $\vec{\boldsymbol{Y}}$. Also compute

$$\frac{\partial}{\partial \mu} \ell_{\vec{\mathbf{Y}}}(\mu, \sigma^2) \qquad \text{and} \qquad \frac{\partial}{\partial \sigma^2} \ell_{\vec{\mathbf{Y}}}(\mu, \sigma^2)$$

Solution: Taking the natural logarithm of our answer to part (a), we find

$$\ell_{\vec{\mathbf{Y}}}(\mu, \sigma^2) = \ln \mathcal{L}_{\vec{\mathbf{Y}}}(\mu, \sigma)^2 = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \mu)^2$$

We now take the required partial derivatives:

$$\frac{\partial}{\partial \mu}\ell_{\vec{\mathbf{Y}}}(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)$$
$$\frac{\partial}{\partial \sigma^2}\ell_{\vec{\mathbf{Y}}}(\mu, \sigma^2) = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(Y_i - \mu)^2$$

One thing to note: we take derivatives with respect to σ^2, not σ. This is because we're really thinking of σ^2 as a parameter in itself (i.e. the population variance) - so, whenever we take the derivative with respect to σ^2, we treat σ^2 as the whole variable, not σ.

(c) In the two derivatives you found in part (b), replace all instances of μ with $\widehat{\mu}_{\text{MLE}}$, all instances of σ^2 with $\widehat{\sigma^2}_{\text{MLE}}$. Set the resulting expressions equal to zero and solve for $\widehat{\mu}_{\text{MLE}}$ and $\widehat{\sigma^2}_{\text{MLE}}$.

Solution: The system of equations we wish to solve is:

$$\frac{1}{\widehat{\sigma^2}_{\text{MLE}}}\sum_{i=1}^{n}(Y_i - \widehat{\mu}_{\text{MLE}}) = 0$$
$$-\frac{n}{2}\cdot\frac{1}{\widehat{\sigma^2}_{\text{MLE}}} + \frac{1}{2(\widehat{\sigma^2}_{\text{MLE}})^2}\sum_{i=1}^{n}(Y_i - \widehat{\mu}_{\text{MLE}})^2 = 0$$

Let's focus on the first equation. Multiplying both sides by $\widehat{\sigma^2}_{\text{MLE}}$, we find:

$$\sum_{i=1}^{n} Y_i - n\cdot\widehat{\mu}_{\text{MLE}} = 0 \implies \widehat{\mu}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i =: \overline{Y}_n$$

Substituting this into the second equation and simplifying, we find

$$\frac{1}{(\widehat{\sigma^2}_{\text{MLE}})}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 = n \implies \widehat{\sigma^2}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$$

(d) The **equivariance** property of maximum likelihood estimators is as follows: given the MLE $\widehat{\theta}_{\text{MLE}}$ for a parameter θ, the MLE of $\tau(\theta)$ [where $\tau(\cdot)$ is an appropriately-behaved function] is $\tau(\widehat{\theta}_{\text{MLE}})$. For example, the MLE of θ^3 is $(\widehat{\theta}_{\text{MLE}})^3$.

Use the equivariance property and your answer to part (c) to derive an expression for the maximum likelihood estimator of σ, the population *standard deviation*.

Solution: Since $\sigma = \sqrt{\sigma^2}$,

$$\widehat{\sigma}_{\mathsf{MLE}} = \sqrt{(\widehat{\sigma^2}_{\mathsf{MLE}})} = \boxed{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2}}$$

Problem 2

Something's gone awry with *GauchoPop*'s newest bottling machine! Specifically, the new soda dispenser doesn't fill each bottle entirely - rather, the proportion Y_i of a randomly-selected bottle that is full of soda follows the distribution with density

$$f(y;\theta) = \theta y^{\theta-1} \cdot \mathbb{1}_{\{0 \leq y \leq 1\}}$$

where $\theta > 0$ is an unknown parameter. Let $Y_1, \cdots, Y_n$ denote the proportion of fill contained in n randomly-selected *GauchoPop* bottles.

(a) Find $\widehat{\theta}_{\text{MLE}}$, the maximum likelihood estimator for θ.

Solution:

$$\mathcal{L}_{\vec{\mathbf{Y}}}(\theta) = \prod_{i=1}^{n} f(Y_i;\theta)] = \prod_{i=1}^{n} \left[\theta Y_i^{\theta-1} \cdot \mathbb{1}_{\{0 \leq Y_i \leq 1\}}\right] = \theta^n \left(\prod_{i=1}^{n} Y_i\right)^{\theta-1} \cdot \prod_{i=1}^{n} \mathbb{1}_{\{0 \leq Y_i \leq 1\}}$$

$$\ell_{\vec{\mathbf{Y}}}(\theta) = \log \mathcal{L}_{\vec{\mathbf{Y}}}(\theta) = n \ln(\theta) + (\theta - 1)\sum_{i=1}^{n} \ln(Y_i) + \sum_{i=1}^{n} \ln \mathbb{1}_{\{0 \leq Y_i \leq 1\}}$$

$$\frac{\partial}{\partial \theta}\ell_{\vec{\mathbf{Y}}}(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln(Y_i)$$

Therefore, $\widehat{\theta}_{\text{MLE}}$ must satisfy

$$\frac{n}{\widehat{\theta}_{\text{MLE}}} + \sum_{i=1}^{n} \ln(Y_i) \implies \widehat{\theta}_{\text{MLE}} = -\frac{n}{\sum_{i=1}^{n} \ln(Y_i)} = \boxed{\left[\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{Y_i}\right)\right]^{-1}}$$

(b) Company regulations requires that any bottles with fewer than 0.8 fill be labeled as "not fit for sale." Let τ denote the true proportion of bottles that end up labeled as "not fit for sale" - find $\widehat{\tau}_{\text{MLE}}$, the maximum likelihood estimator for τ.

Hint: Use the equivariance property

Solution: Note that

$$\tau := \mathbb{P}(Y_i < 0.8) = \int_{0.8}^{1} \theta y^{\theta-1}\, \mathrm{d}y = (0.8)^\theta$$

By the **equivariance property** of the maximum likelihood estimator,

$$\widehat{\tau}_{\text{MLE}} = \widehat{(0.8)^\theta}_{\text{MLE}} = (0.8)^{(\widehat{\theta}_{\text{MLE}})} = \boxed{\left(\frac{4}{5}\right)^{\frac{n}{\sum_{i=1}^{n} \ln\left(\frac{1}{Y_i}\right)}}}$$

Problem 3

Let $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[\theta, 1]$ where $\theta < 1$ is an unknown parameter.

(a) Find $\widehat{\theta}_{\text{MM}}$, the method of moments estimator for θ.

Solution:

$$\mathbb{E}[Y_i] = \frac{\theta + 1}{2} \implies \frac{\widehat{\theta}_{\text{MM}} + 1}{2} = \overline{Y}_n \implies \widehat{\theta}_{\text{MM}} = 2\overline{Y}_n - 1$$

(b) Show that the likelihood of the sample $\vec{\boldsymbol{Y}}$ is given by

$$\mathcal{L}_{\vec{\mathbf{Y}}}(\theta) = \left(\frac{1}{1-\theta} \right)^n \cdot \mathbb{1}_{\{\theta \leq Y_{(1)}\}} \cdot \mathbb{1}_{\{Y_{(n)} \leq 1\}}$$

where $Y_{(1)}$ denotes the first order statistic (sample minimum) and $Y_{(n)}$ denotes the n^{th} order statistic (sample maximum). Justify your logic.

Solution: We begin by finding the likelihood of the sample $\vec{\boldsymbol{Y}}$:

$$\begin{aligned} \mathcal{L}_{\vec{\mathbf{Y}}}(\theta) &= \prod_{i=1}^{n} f_Y(Y_i; \theta) = \prod_{i=1}^{n} \left[\frac{1}{1-\theta} \cdot \mathbb{1}_{\{\theta \leq Y_i \leq 1\}} \right] \\ &= \left(\frac{1}{1-\theta} \right)^n \cdot \prod_{i=1}^{n} \mathbb{1}_{\{\theta \leq Y_i\}} \cdot \prod_{i=1}^{n} \mathbb{1}_{\{Y_i \leq 1\}} \end{aligned}$$

Let's focus on simplifying the two products of indicators. Both products are nonzero only when each of the constituent indicators is nonzero. Hence, the first product of indicators is only nonzero when all of the Y_i's are greater than θ, which occurs when $Y_{(1)} \geq \theta$. Similarly, the second product of indicators is nonzero only when all of the Y_i's are less than 1, which occurs when $Y_{(n)} \leq 1$. Hence,

$$\mathcal{L}_{\vec{\mathbf{Y}}}(\theta) = \left(\frac{1}{1-\theta} \right)^n \cdot \mathbb{1}_{\{\theta \leq Y_{(1)}\}} \cdot \mathbb{1}_{\{Y_{(n)} \leq 1\}}$$

(c) Find $\widehat{\theta}_{\text{MLE}}$, the maximum likelihood estimator for θ.

Solution: By definition,

$$\widehat{\theta}_{\text{MM}} = \operatorname*{arg\ max}_{\theta} \left\{ \mathcal{L}_{\vec{\mathbf{Y}}}(\theta) \right\}$$

Note that $\mathcal{L}_{\vec{\mathbf{Y}}}(\theta)$ is nondifferentiable in θ - hence, we cannot simply take the derivative wrt. θ and set equal to zero. Instead, we must maximize $\mathcal{L}_{\vec{\mathbf{Y}}}(\theta)$ analytically.

First note that, for $\theta < 1$, the function $\left(\frac{1}{1-\theta} \right)^n$ is an increasing function in θ; hence, it is maximized by setting θ to be as large as possible. The indicator $\mathbb{1}_{\{\theta \leq Y_{(1)}\}}$ essentially restricts θ to be no larger than $Y_{(1)}$. Hence, putting these two facts together, we see that the likelihood is maximized at $Y_{(1)}$; i.e.

$$\widehat{\theta}_{\text{MLE}} = Y_{(1)}$$

(d) Find the exact sampling distribution of $\widehat{\theta}_{\text{MLE}}$, and use this to determine whether or not $\widehat{\theta}_{\text{MLE}}$ is an unbiased estimator for θ.

Solution: By our formula for the density of the first order statistic,

$$f_{Y_{(1)}}(y) = n\left[\overline{F_Y}(y)\right]^{n-1} \cdot f_Y(y)$$

Since $Y_i \overset{\text{i.i.d.}}{\sim} \text{Unif}[\theta, 1]$, we have

$$\overline{F_Y}(y) = \begin{cases} 1 & \text{if } y < \theta \\ \frac{1-y}{1-\theta} & \text{if } \theta \leq y \leq 1 \\ 0 & \text{if } y \geq 1 \end{cases}$$

Therefore:

$$\begin{aligned} f_{Y_{(1)}} &= n\left[\overline{F_Y}(y)\right]^{n-1} \cdot f_Y(y) \\ &= n\left(\frac{1-y}{1-\theta}\right)^{n-1} \cdot \frac{1}{1-\theta} \cdot \mathbb{1}_{\{\theta \leq y \leq 1\}} = \frac{n(1-y)^{n-1}}{(1-\theta)^n} \cdot \mathbb{1}_{\{\theta \leq y \leq 1\}} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}[Y_{(1)}] &= \int_{-\infty}^{\infty} y f_{Y_{(1)}}(y)\, \mathrm{d}y \\ &= \frac{n}{(1-\theta)^n} \cdot \int_{\theta}^{1} y(1-y)^{n-1}\, \mathrm{d}y \\ &= \frac{n}{(1-\theta)^n} \cdot \int_0^{1-\theta} (1-u)u^{n-1} \\ &= \frac{n}{(1-\theta)^n} \cdot \left[\frac{1}{n}(1-\theta)^n - \frac{1}{n+1}(1-\theta)^{n+1}\right] \\ &= 1 - \frac{n}{n+1}(1-\theta) = \frac{n+1-n+n\theta}{n+1} = \frac{n\theta+1}{n+1} \end{aligned}$$

As this does not equal θ for any finite sample size n, we can conclude that $\widehat{\theta}_{\text{MLE}}$ is a biased estimator for θ.

(e) Show that $U := Y_{(1)}$, the first order statistic, is a sufficient sufficient statistic for θ.

Solution: Going back to our likelihood from part (b), we can rearrange terms to see

$$\mathcal{L}_{\vec{\boldsymbol{Y}}}(\theta) = \underbrace{\left[\left(\frac{1}{1-\theta}\right)^n \cdot \mathbb{1}_{\{\theta \leq Y_{(1)}\}}\right]}_{:=g(Y_{(1)}, \theta)} \times \underbrace{\left[\mathbb{1}_{\{Y_{(n)} \leq 1\}}\right]}_{:=h(\vec{\boldsymbol{Y}})}$$

Since the likelihood factors into the product of two functions, one involving only θ and $Y_{(1)}$ and another involving only $\vec{\boldsymbol{Y}}$, we can use the **factorization theorem** to conclude that $U := Y_{(1)}$ is a sufficient statistic for θ.