Exam Statistics

Bachelor Econometrics and Operations Research
Bachelor Econometrics and Data Science
Faculty of Economics and Business Administration
Wednesday, March 30, 2022

Exam:	Statistics
Code:	E_EOR1_STAT
Coordinator:	M.H.C. Nientker
Date:	March 30, 2022
Time:	12:15
Duration:	2 hours
Calculator:	Not allowed
Graphical calculator:	Not allowed
Number of questions:	4
Type of questions:	Open
Answer in:	English
Credit score:	88 credits counts for a 10
Grades:	Made public within 10 working days
Number of pages:	2, including front page

- Read the entire exam carefully before you start answering the questions.
- Be clear and concise in your statements, but justify every step in your derivations.
- The questions should be handed back at the end of the exam. Do not take it home.

Good luck!

Question 1. Let $X_1, \ldots, X_n$ be an independent and identically distributed sequence of random variables from a population in $\{g_\mu \mid \mu \in \mathbb{R}\}$, where

$$g_\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \qquad x > 0.$$

Note that σ^2 is assumed to be known.

(8 points) a. Show that the moment estimator $\hat{\mu}_{MOM}$ of μ_0 is equal to the sample average $\overline{X}$.

(8 points) b. Calculate the mean squared error of $\hat{\mu}_{MOM}$.

(8 points) c. Find a sufficient and complete statistic for μ_0.

(8 points) d. Find an UMVU estimator of μ_0^2. Hint: start from $\overline{X}^2$.

Solution.

a. To find the moment estimator we have to solve for μ in the following equation

$$\overline{X} = \mathbb{E}_\mu X_1 = \mu.$$

This immediately delivers $\hat{\mu} = \overline{X}$.

4 for the equation on how to find a moments estimator, 4 for the result.

b. We have

$$\begin{aligned}
\mathbb{E}_\mu(\hat{\mu}) &= \mathbb{E}_\mu(\overline{X}) = \mathbb{E}_\mu\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\mu(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \mu \\
\text{Bias}_\mu(\hat{\mu}) &= \mathbb{E}_\mu(\hat{\mu}) - \mu = \mu - \mu = 0 \\
\mathbb{V}\text{ar}_\mu(\hat{\mu}) &= \mathbb{V}\text{ar}_\mu(\overline{X}) = \mathbb{V}\text{ar}_\mu\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\mathbb{V}\text{ar}_\mu\left(\sum_{i=1}^n X_i\right) \\
&\overset{ind}{=} \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\text{ar}_\mu(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n} \\
\text{MSE}(\mu, \hat{\mu}) &= \mathbb{V}\text{ar}_\mu(\hat{\mu}) + \text{Bias}_\mu(\hat{\mu})^2 = \frac{\sigma^2}{n} + 0^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

2 points for each derivation. Minus points for forgetting independence or incorrect subscripts.

c. We can rewrite the density as

$$g_\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} e^{\mu^2/2\sigma^2} e^{-\mu x/\sigma^2}.$$

We therefore obtain an exponential family with $h(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$, $c(\mu) = e^{\mu^2/2\sigma^2}$, $t_1(x) = x$, $w_1(\mu) = \frac{-\mu}{\sigma^2}$. We therefore conclude that our statistical model is an exponential family. Moreover the parameter space $\mathbb{R}$ contains an open subset, say $(0,1)$. Therefore, by the results on exponential families, we can conclude that $T(\vec{X}) = \sum_{i=1}^{\infty} t_1(X_i) = \sum_{i=1}^{\infty} X_i = n\overline{X}$ is sufficient and complete.

3 points for rewriting density, 2 points for identifying functions, 2 points for identifying the statistic, 1 points for open subset.

d. The Lehmann-Schefé theorem tells us that any function of $T(\vec{X})$ is UMVU for its mean. We use the hint to start with the function $\phi_1(t) = \frac{t^2}{n^2}$ to find

$$\mathbb{E}_\mu \phi_1(T(\vec{X})) = \mathbb{E}_\mu \overline{X}^2 = \mathbb{V}\mathrm{ar}_\mu\, \overline{X} + \mathbb{E}_\mu^2 \overline{X} = \sigma^2/n + \mu^2.$$

This expectation is not equal to μ yet, so we relocate and use $\phi_2(t) = \phi_1(t) - \sigma^2/n = \frac{t^2}{n^2} - \sigma^2/n$ to find

$$\mathbb{E}_\mu \phi_2(T(\vec{X})) = \mathbb{E}_\mu \phi_1(T(\vec{X})) - \sigma^2/n = \sigma^2/n + \mu^2 - \sigma^2/n.$$

We conclude that $\phi_2(T(\vec{X})) = \overline{X}^2 - \sigma^2/n$ is UMVU for μ_0^2.

1 point for finding ϕ_1, 2 points for finding the first expectation, 2 points for finding ϕ_2, 1 point for the second expectation, 2 points for finding the UMVU estimator with correct conclusion.

Question 2. Let $X_1, \ldots, X_n$ be an independent and identically distributed sequence of random variables from a population in $\{g_\theta \mid \theta \in \Theta\}$, let W be an unbiased estimator for $\tau(\theta_0)$ and let T be a sufficient statistic for θ_0.

(8 points) a. Give the formal definition of sufficiency. What is the intuitive interpretation based on summarizing the data?

(8 points) b. State the Rao-Blackwell theorem. Why do we need sufficiency for this result?

SOLUTION.

a. A statistic $T(\vec{X})$ is a *sufficient statistic* for θ_0 if the conditional distribution of the sample $\vec{X}$ given the value of $T(\vec{X})$ does not depend on θ_0. The intuitive interpretation is that a sufficient statistic is a summary of the data that still contains all the relevant information about θ_0.

4 points for correct definition, 4 points for correct intuitive interpretation.

b. The Rao-Blackwell theorem states that $\phi(T) = \mathbb{E}(W \mid T)$ is an unbiased estimator of $\tau(\theta_0)$ and that $\mathbb{V}\mathrm{ar}_\theta\, \phi(T) \leq \mathbb{V}\mathrm{ar}_\theta\, W$. That is, $\phi(T)$ is uniformly better than W. Sufficiency of T is needed in this theorem, because typically the distribution, and hence the expectation, of $W \mid T$ depends on θ_0 and thus $\phi(T)$ would not be well defined. Sufficiency ensures that the conditional distribution of the sample $\vec{X}$, and hence the conditional distribution of W, given the value of $T(\vec{X})$ does not depend on θ_0. Therefore $\phi(T)$ is only well defined if T is sufficient.

4 points for correct definition, 4 points for correct explanation.

Question 3. Let $X_1, \ldots, X_n$ be an independent and identically distributed sequence of random variables from a population in $\{g_\lambda \mid \lambda > 0\}$, where

$$g_\lambda(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

In this question you are allowed to use that $\mathbb{E}_\lambda X_1 = 1/\lambda$ and $\mathbb{V}\mathrm{ar}_\lambda\, X_1 = 1/\lambda^2$.

(8 points) a. Show that $\hat{\lambda}_{ML} = 1/\overline{X}$ is the maximum likelihood estimator of λ_0.

(8 points) b. Show that $\overline{X}$ is an UMVU estimator for $\tau(\lambda_0) = 1/\lambda_0$ using the Cramér-Rao lower bound.

(8 points) c. Find an asymptotic distribution for $\hat{\lambda}_{ML}$ given the general result on the asymptotic distribution of maximum likelihood estimators. Make sure the asymptotic variance does not depend on λ_0.

Solution.

a. To derive the ML estimator we write down the log likelihood

$$\begin{aligned}\log L(\lambda \mid \vec{x}) = \log f_\lambda(\vec{x}) = \log \prod_{i=1}^n g_\lambda(x_i) = \sum_{i=1}^n \log g_\lambda(x_i) = \sum_{i=1}^n \log\left(\lambda e^{-\lambda x_i}\right) \\ = \sum_{i=1}^n \log \lambda - \lambda x_i = n \log \lambda - \lambda \sum_{i=1}^n x_i.\end{aligned}$$

Taking derivatives we obtain

$$\begin{aligned}\frac{d}{d\lambda} \log L(\lambda \mid \vec{x}) &= \frac{n}{\lambda} - \sum_{i=1}^n x_i \\ \frac{d^2}{d\lambda^2} \log L(\lambda \mid \vec{x}) &= -\frac{n}{\lambda^2}\end{aligned}$$

Setting the first derivative equal to zero gives us the stationary point $\tilde{\lambda} = 1/\overline{x}$. As each $x_i > 0$ we have $\overline{x} > 0$ which means $\tilde{\lambda} > 0$ and thus the second derivative at λ evaluates to $-n/\tilde{\lambda}^2 < 0$. We conclude that the found stationary point is a local maxima. Since it is the only stationary point, we conclude by the lemma that we have found the global maximum and thus we have $\hat{\lambda} = 1/\overline{X}$.

2 for the log likelihood, 1 for each derivative, 1 for finding the stationary point, 1 for concluding it's a maximum, 1 for stating the ML estimator, 1 bonus point for remarking anything about boundary or limit points.

b. We start out by finding the information number. We have

$$\begin{aligned}\log g_\lambda(X) &= \log\left(\lambda e^{-\lambda x}\right) = \log \lambda - \lambda X \\ \frac{\partial}{\partial \lambda} \log g_\lambda(X) &= \frac{1}{\lambda} - X \\ i_\lambda = \mathbb{V}\mathrm{ar}_\lambda \frac{\partial}{\partial \lambda} \log g_\lambda(X) &= \mathbb{V}\mathrm{ar}\left(\frac{1}{\lambda} - X\right) = \mathbb{V}\mathrm{ar}_\lambda\, X = \frac{1}{\lambda^2}.\end{aligned}$$

We then get for the Cramér-Rao lower bound that

$$B(\lambda) = \frac{\tau'(\lambda)^2}{n i_\lambda} = \frac{\left(-1/\lambda^2\right)^2}{n/\lambda^2} = \frac{1}{n\lambda^2}.$$

For our estimator we have similarly as in exercise 1a

$$\begin{aligned}
\mathbb{E}_\lambda(\overline{X}) &= \mathbb{E}_\lambda\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\lambda(X_i) = \frac{1}{n}\sum_{i=1}^n \lambda = \frac{1}{\lambda} \\
\mathrm{Bias}_\lambda(\hat{\lambda}) &= \mathbb{E}_\mu(\hat{\lambda}) - \tau(\lambda) = \frac{1}{\lambda} - \frac{1}{\lambda} = 0 \\
\mathbb{V}\mathrm{ar}_\lambda(\overline{X}) &= \mathbb{V}\mathrm{ar}_\lambda\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\mathbb{V}\mathrm{ar}_\lambda\left(\sum_{i=1}^n X_i\right) \\
&\overset{ind}{=} \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\mathrm{ar}_{\frac{1}{\lambda}}(X_i) = \frac{1}{n^2}\sum_{i=1}^n \frac{1}{\lambda^2} = \frac{1}{n\lambda^2}
\end{aligned}$$

We conclude that $\overline{X}$ is unbiased for $\tau(\lambda_0)$ with variance equal to the Cramér-Rao lower bound. Therefore it must be UMVU.

3 points for information number, 2 points for Cramér-Rao lower bound, 1 point for unbiased, 1 point for variance, 1 point for conclusion.

c. We know from the general result of ML estimators that

$$\tau(\hat{\lambda}_{ML}) \approx N\left(\tau(\lambda_0), \frac{\tau'(\lambda_0)^2}{n i_{\lambda_0}}\right),$$

We know from part b. that $i_{\lambda_0} = \frac{1}{\lambda_0^2}$. Using the plug-in information number we find

$$i_{\hat{\lambda}} = \frac{1}{1/\overline{X}^2} = \overline{X}^2.$$

We conclude that

$$\hat{\lambda}_{ML} \approx N\left(\tau(\lambda_0), \frac{\tau'(\lambda_0)^2}{n i_{\lambda_0}}\right) = N\left(\lambda_0, \frac{1}{n\overline{X}^2}\right).$$

3 points for knowing asymptotic general result, 3 points for plug-in information number, 2 points for correct conclusion.

Question 4. Let X be Uniform$(0,1)$ distributed, that is, it has pdf

$$g(x) = 1, \qquad 0 \leq x \leq 1.$$

(8 points) a. Show that the location scale family of g is $\{g_{(\mu,\sigma^2)} \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$, where

$$g_{(\mu,\sigma^2)} = \frac{1}{\sigma}, \qquad \mu \leq x \leq \mu + \sigma.$$

(8 points) b. Use the factorization theorem to find a sufficient statistic T for (μ, σ^2). Note that the domain of $g_{(\mu,\sigma^2)}$ depends on the parameters μ and σ^2.

SOLUTION.

a. Let $Y = \mu + \sigma X$. Then the domain of Y is $\{y \in \mathbb{R} \mid \mu \leq y \leq \mu + \sigma\}$, since the domain of X is $\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$. Next,

$$G_Y(y) = P(Y \leq y) = P(\mu + \sigma X \leq y) = P\left(X \leq \frac{y-\mu}{\sigma}\right) = G_X\left(\frac{y-\mu}{\sigma}\right)$$
$$g_Y(y) = \frac{d}{dy}G_Y(y) = \frac{1}{\sigma}g_X\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma} \times 1 = \frac{1}{\sigma}.$$

Letting $\mu \in \mathbb{R}$ and $\sigma > 0$ gives us a whole set of distributions, which is exactly equal to the one stated in the question.

2 points for any attempt on getting a pdf for a transformed random variable, 2 points for getting the right cdf, 2 points for getting the right pdf, 1 point for finding the right domain, 1 point for final conclusion.

b. We have

$$\begin{aligned} f_{(\mu,\sigma)}(x) &= \prod_{i=1}^{n} g_{(\mu,\sigma)}(x_i) = \prod_{i=1}^{n} \frac{1}{\sigma}\mathbb{1}_{\{\mu \leq x_i \leq \mu+\sigma\}} \\ &= \frac{1}{\sigma^n}\prod_{i=1}^{n} \mathbb{1}_{\{x_i \geq \mu\}}\mathbb{1}_{\{x_i \leq \mu+\sigma\}} \\ &= \frac{1}{\sigma^n}\prod_{i=1}^{n} \mathbb{1}_{\{x_i \geq \mu\}}\prod_{i=1}^{n}\mathbb{1}_{\{x_i \leq \mu+\sigma\}} \\ &= \frac{1}{\sigma^n}\mathbb{1}_{\{x_{(1)} \geq \mu\}}\mathbb{1}_{\{x_{(n)} \leq \mu+\sigma\}}. \end{aligned}$$

Let $T(X) = \big(X_{(1)}, X_{(n)}\big)$, then $f_{(\mu,\sigma)}(x)$ depends on the data only through $T(x)$, so T is sufficient.

1 point for any attempt on rewriting the joint distribution of the data, 1 point for writing as a product of $g(x_i)$, 1 point for including the indicator function, 2 points for rewriting the product of indicator functions as one indicator function for the minimum, 2 points for rewriting the product of indicator functions as one indicator function for the maximum, 1 point for finding the right statistic and correct conclusion.