Statistics

Bachelor Econometrics and Operations Research
Bachelor Econometrics and Data Science
School of Business and Economics

Exam:	Statistics
Code:	E_EOR1_STAT
Examinator:	Dr. E.J.J. Wijler
Co-reader:	N. Berhane
Date:	Wednesday, March 29, 2023
Time:	12:15
Duration:	2 hours
Calculator:	Allowed
Graphical calculator:	Allowed
Scrap paper:	Allowed
Open Book:	No
Number of questions:	2
Type of questions:	Open
Answer in:	English
Credit score:	100 credits counts for a 10
Grades:	made public within 10 working days
Inspection:	TBA
Number of pages:	3, including front page

- Give justifications for your answers unless stated otherwise.
- Be complete and explicit, but also clear and concise in your statements.
- If you think that further information is needed to answer a question, or that the question is ill-posed, then explain your reasoning
- The questions should be handed back at the end of the exam. Do not take it home.

Good luck!

Question 1 [25 points] - True/False Questions

For each of the following questions, argue whether you agree or disagree. Motivate your answer (points are only awarded for motivated answers). In all of the following, $X_1, \ldots, X_n$ is a random sample from some distribution $f(\boldsymbol{x} \mid \theta)$.

(a) **[5 points]** Let $\hat{\theta}$ be a best unbiased estimator (UMVUE) of θ and let τ be some arbitrary function. Then, $\hat{\eta} = \tau\left(\hat{\theta}\right)$ is a best unbiased estimator of $\eta = \tau(\theta)$

(b) **[5 points]** Let $\hat{\theta}_1$ be a consistent estimator of θ and $\hat{\theta}_2$ be an unbiased estimator of θ. Then, $\hat{\theta}_1$ is also unbiased, but $\hat{\theta}_2$ is not necessarily consistent.

(c) **[5 points]** The family of distributions $\left\{\frac{1}{2\sigma}e^{-\frac{x-\mu}{2\sigma}} \mid \mu \in \mathbb{R}, \sigma > 0\right\}$ forms a location-scale family.

(d) **[5 points]** Let $T(\boldsymbol{X}) = \prod_{i=1}^n X_i$ be a sufficient statistic for θ. The estimator $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n \log(X_i)$ can be a best unbiased estimator.

(e) **[5 points]** We have two estimators, $W_1(\boldsymbol{X})$ with $\mathbb{E}_\theta(W(\boldsymbol{X})) = \tau(\theta)$ and $W_2(\boldsymbol{X})$ with $\mathbb{E}_\theta(W_2(\boldsymbol{X})) = 0$ for all $\theta \in \Theta$. In addition, $\mathbb{E}_\theta(W_1(\boldsymbol{X})W_2(\boldsymbol{X})) = 0$ for all $\theta \in \Theta$. Then, $W_1(\boldsymbol{X})$ is the best unbiased estimator of $\tau(\theta)$.

Solution. (a) Disagree, there is no guarantee that $\hat{\eta}$ is an unbiased estimator of η, let alone a best unbiased estimator. We have seen repeatedly by application of Jensen's inequality that many transformations result in biased estimators.

(b) Disagree, because $\hat{\theta}_1$ may be biased. Consistent estimators may be biased, as long as the bias vanishes in the limit.

(c) Agree. Take the distribution

$$g(x \mid 0, 1) = \frac{1}{2}e^{-\frac{x}{2}}.$$

This generates the location-scale family

$$g(x \mid \mu, \sigma) = \left\{\frac{1}{\sigma}g\left(\frac{x-\mu}{\sigma}\right) \mid \mu \in \mathbb{R}, \sigma > 0\right\} = \left\{\frac{1}{2\sigma}e^{-\frac{x-\mu}{\sigma}}\mu \in \mathbb{R}, \sigma > 0\right\}.$$

(d) Agree. Sufficient statistics are not unique, as any invertible transformation of a sufficient statistics is again a sufficient statistic. In this case, $\tilde{T}(\boldsymbol{X}) = \log(T(\boldsymbol{X})) = \log\left(\prod_{i=1}^n X_i\right) = \sum_{i=1}^n \log(X_i)$. Since the estimator $\hat{\theta}$ is based on a sufficient statistic, it can be a best unbiased estimator (not necessarily though).

(e) Disagree. For $W_1(\boldsymbol{X})$ to be a best unbiased estimator it would have to be uncorrelated with *all* estimators U such that $\mathbb{E}_\theta(U) = 0$ for all $\theta \in \Theta$, of which $W_2(\boldsymbol{X})$ is just one example.

$\square$

Question 2 [75 points] - Carnaval in Maastricht

Your course lecturer is looking forward to celebrate carnaval in Maastricht. He spent weeks to create his colourful home-made costume. While he is happy with the result, he is afraid the costume will not survive heavy rainfall. Therefore, he only wants to wear his costume when the probability of rainfall is low. Since you don't trust the weather forecasts, you decide to help your lecturer predict the amount of rainfall (in mm) in Maastricht around carnaval.

After doing some research into weather modelling, you learn that the amount of daily rainfall can be modelled by the Gamma distribution with pdf

$$g(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 \leq x < \infty, \alpha, \beta > 0,$$

with $\mathbb{E}_{\theta}(X) = \alpha\beta$ and $\mathbb{V}\mathrm{ar}_{\theta}(X) = \alpha\beta^2$, where $\theta = (\alpha, \beta)$. Let $X_1, \ldots, X_n$ be a random sample from $g(x \mid \alpha, \beta)$ where X_i represent the amount of rainfall on day i. You have collected historical rainfalls for 15 historical carnaval days, say $x_1, \ldots, x_{15}$, and recorded the following information:

$$n = 15, \qquad \sum_{i=1}^{n} x_i = 120, \qquad \sum_{i=1}^{n} x_i^2 = 960.$$

(a) **(10pts)** Find sufficient statistics for (α, β). What would be the sufficient statistic if α was known?

For questions (b)-(e), assume that α is a known constant!

(b) **(10pts)** Show that the maximum likelihood estimator is equal to $\hat{\beta}_{ML} = \frac{\bar{X}}{\alpha}$.

(c) **(10pts)** Show that the method of moments estimators based on the first and second moment are given by

$$\hat{\beta}_{MM1} = \frac{\bar{X}}{\alpha} \quad \text{and} \quad \hat{\beta}_{MM2} = \sqrt{\frac{1}{n\alpha(1+\alpha)} \sum_{i=1}^{n} X_i^2},$$

respectively.

(d) **(15pts)** Are the estimators found in (b) and (c) unbiased estimators of β? Is any of them a UMVUE of β? Motivate your answer.

(e) **(10pts)** Show that $\hat{\beta}_{MM2}$ is a consistent estimator of β.

(f) **(10pts)** If the rainfall is more than 20mm, your lecturer's costume is unlikely to survive. Therefore, if the probability of more than 20mm of rainfall is larger than 5%, you will advise your lecturer to wear a different costume. Assume that $\alpha = 1$ and compute the probability of more than 20mm of rainfall *by maximum likelihood.* State your advice to your lecturer.

Hint: $\mathbb{P}_{\beta}(X > x)$ is just a function of β. Also, $\Gamma(1) = 1$.

(g) **(10pts)** Derive *either* the MME *or* the MLE of α in the case where both α and β are unknown. **Note**: it may not be possible to derive both of these estimators analytically, so choose wisely which of these two estimators you are going to try to derive.

Proof. (a) We can use both the factorization theorem or the exponential family.

Factorization Theorem

The joint distribution is given by

$$\begin{aligned} f(\boldsymbol{x} \mid \alpha, \beta) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} = \Gamma(\alpha)^{-n} \beta^{-n\alpha} \left(\prod_{i=1}^{n} x_i \right)^{\alpha-1} \exp\left(-\frac{\sum_{i=1}^{n} x_i}{\beta} \right) \\ &= \frac{1}{\prod_{i=1}^{n} x_i} \Gamma(\alpha)^{-n} \beta^{-n\alpha} \left(\prod_{i=1}^{n} x_i \right)^{\alpha} \exp\left(-\frac{\sum_{i=1}^{n} x_i}{\beta} \right). \end{aligned}$$

Then, we define $h(\boldsymbol{x}) = \frac{1}{\prod_{i=1}^{n} x_i}$ and

$$g(\boldsymbol{x} \mid T(\boldsymbol{x})) = \Gamma(\alpha)^{-n} \beta^{-n\alpha} T_1(\boldsymbol{x})^{\alpha} \exp\left(-\frac{T_2(\boldsymbol{x})}{\beta} \right),$$

with the sufficient statistics given by $\boldsymbol{T}(\boldsymbol{X}) = (T_1(\boldsymbol{X}), T_2(\boldsymbol{X})) = (\prod_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i)$.

If α were known, we would instead define $h(\boldsymbol{x}) = \frac{1}{\Gamma(\alpha)^n} (\prod_{i=1}^{n} x_i)^{\alpha-1}$ and

$$g(\boldsymbol{x} \mid T(\boldsymbol{x})) = \beta^{-n\alpha} \exp\left(-\frac{T(\boldsymbol{x})}{\beta} \right),$$

with $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ being the sufficient statistic.

Exponential family

Using the exponential family, we can rewrite the pdf as

$$g(x \mid \alpha, \beta) = \frac{1}{x\Gamma(\alpha)\beta^{\alpha}} \exp\left(\alpha \log(x) - \frac{x}{\beta} \right).$$

This is clearly part of the exponential family with decomposition $h(x) = x$, $c(\boldsymbol{\theta}) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}$, $w_1(\boldsymbol{\theta}) = \alpha$, $w_2(\boldsymbol{\theta}) = \frac{-1}{\beta}$, $t_1(x) = \log(x)$, $t_2(x) = x$. Hence, the sufficient statistics are given by $\boldsymbol{T}(\boldsymbol{X}) = (T_1(\boldsymbol{X}), T_2(\boldsymbol{X})) = (\sum_{i=1}^{n} \log(X_i), \sum_{i=1}^{n} X_i)$.

If α were known, then we would define $h(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}$, $c(\beta) = \frac{1}{\beta^{\alpha}}$, $w(\beta) = -\frac{1}{\beta}$, $t(\beta) = x$. Hence, the sufficient statistic is given by $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$.

(b) The log-likelihood is given by

$$\ell(\beta \mid \boldsymbol{x}) = -n \log(\Gamma(\alpha)) - n\alpha \log(\beta) + (\alpha - 1) \sum_{i=1}^{n} \log(x_i) - \frac{\sum_{i=1}^{n} x_i}{\beta}.$$

Taking the first and solving gives the stationary point

$$\frac{d\ell(\beta \mid \boldsymbol{x})}{d\beta} = -\frac{n\alpha}{\beta} + \frac{\sum_{i=1}^{n} x_i}{\beta^2} \overset{\text{s}}{=} 0 \Rightarrow \tilde{\beta} = \frac{\bar{X}}{\alpha}.$$

To check whether this is a global maximum, we check the second derivative at the stationary point:

$$\left.\frac{d^2\ell(\beta \mid \boldsymbol{x})}{d\beta^2}\right|_{\beta=\tilde{\beta}} = \frac{n\alpha}{\tilde{\beta}^2} - \frac{2\sum_{i=1}^n x_i}{\tilde{\beta}^3} = \frac{n\alpha}{\tilde{\beta}^2} - \frac{2n\alpha\tilde{\beta}}{\tilde{\beta}^3} = \frac{-n\alpha}{\tilde{\beta}^2} < 0.$$

Hence, we indeed have $\hat{\beta}_{ML} = \frac{\bar{X}}{\alpha}$.

(c) For the MME based on the first moment, we have

$$\bar{X} \overset{\text{s}}{=} \mathbb{E}_\beta(X) = \alpha\beta \Rightarrow \hat{\beta}_{MM1} = \frac{\bar{X}}{\alpha}.$$

For the second moment, note that

$$\mathbb{E}_\beta(X^2) = \mathbb{V}\mathrm{ar}_\theta(X) + (\mathbb{E}_\beta(X))^2 = \alpha\beta^2 + \alpha^2\beta^2 = \alpha(1+\alpha)\beta^2.$$

Hence,

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \overset{\text{s}}{=} \mathbb{E}_\theta(X^2) = \alpha(1+\alpha)\beta^2 \Rightarrow \hat{\beta}_{MM2} = \sqrt{\frac{1}{n\alpha(1+\alpha)}\sum_{i=1}^n X_i^2}.$$

(d) First, note that $\hat{\beta}_{ML} = \hat{\beta}_{MM1}$, so we verify just one of them. Then,

$$\mathbb{E}_\beta\left(\hat{\beta}_{ML}\right) = \frac{\mathbb{E}_\beta\left(\bar{X}\right)}{\alpha} = \frac{\mathbb{E}_\beta\left(X_1\right)}{\alpha} = \beta.$$

Hence, $\hat{\beta}_{ML}$ and $\hat{\beta}_{MM1}$ are unbiased estimators of β. For $\hat{\beta}_{MM2}$, we can show that

$$\mathbb{E}_\beta\left(\hat{\beta}_{MM2}\right) = \mathbb{E}_\beta\left(\sqrt{\frac{1}{n\alpha(1+\alpha)}\sum_{i=1}^n X_i^2}\right) < \sqrt{\frac{1}{n\alpha(1+\alpha)}\sum_{i=1}^n \mathbb{E}_\beta\left(X_i^2\right)} = \beta.$$

Hence, $\hat{\beta}_{MM2}$ is a biased estimator.

To check if $\hat{\beta}_{ML}$ is UMVUE, we can argue in two ways. First, $g(x \mid \beta)$ is a member of an exponential family. The parameter space for β is given by $(0, \infty)$, which certainly contains an open subset. Therefore, a complete and sufficient statistic is given by $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$. Since $\hat{\beta}_{ML}$ is unbiased and a function of $T(\boldsymbol{X})$, it is UMVUE by Lehmann-Scheffé.

We can also verify the variance and check if it equals the Cramèr-Rao lower bound. First the variance of $\hat{\beta}_{ML}$ is given by

$$\begin{aligned}\mathbb{V}\mathrm{ar}_\beta\left(\hat{\beta}_{ML}\right) &= \mathbb{V}\mathrm{ar}_\beta\left(\frac{\bar{X}}{\alpha}\right) = \frac{1}{n^2\alpha^2}\mathbb{V}\mathrm{ar}_\beta\left(\sum_{i=1}^n X_i\right) \overset{(*)}{=} \frac{1}{n^2\alpha^2}\sum_{i=1}^n \mathbb{V}\mathrm{ar}_\beta(X_i) \\ &\overset{(**)}{=} \frac{1}{n\alpha^2}\mathbb{V}\mathrm{ar}_\beta(X_1) = \frac{\beta^2}{n\alpha}.\end{aligned}$$

Next, the Fisher information number is given by

$$I_\beta = -\mathbb{E}_\beta\left(\frac{n\alpha}{\beta^2} - \frac{2\sum_{i=1}^n X_i}{\beta^2}\right) = -\frac{n\alpha}{\beta^2} + \frac{2\sum_{i=1}^n \mathbb{E}_\beta(X_i)}{\beta^3} = -\frac{n\alpha}{\beta^2} + \frac{2n\alpha}{\beta^2} = \frac{n\alpha}{\beta^2},$$

such that the Cramèr-Rao lower bound is given by

$$B(\beta) = I_\beta^{-1} = \frac{\beta^2}{n\alpha} = \mathbb{V}\mathrm{ar}_\beta\left(\hat{\beta}_{ML}\right).$$

Hence, $\hat{\beta}_{ML}$ and $\hat{\beta}_{MM1}$ both attain the lower bound and, therefore, are UMVUE.

(e) First, note that by the LLN, it holds that

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} \mathbb{E}_\beta(X_1^2) = \alpha(1+\alpha)\beta^2,$$

as $n \to \infty$. In addition, the function $h(x) = \sqrt{\frac{x}{\alpha(1+\alpha)}}$ for $x > 0$ is clearly continuous in x. Then, by the CMT, it holds that

$$\begin{aligned}\hat{\beta}_{MM2} &= \sqrt{\frac{1}{n\alpha(1+\alpha)}\sum_{i=1}^{n} X_i^2} = h\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) \xrightarrow{p} h\left(\mathbb{E}_\beta(X_1^2)\right) \\ &= \sqrt{\frac{1}{n\alpha(1+\alpha)}\sum_{i=1}^{n} \mathbb{E}_\beta(X_i^2)} = \sqrt{\frac{\alpha(1+\alpha)\beta^2}{\alpha(1+\alpha)}} = \beta,\end{aligned}$$

as $n \to \infty$. Hence, $\hat{\beta}_{MM2}$ is a consistent estimator of β.

(f) First, we derive an analytic expression for the probability as follows:

$$\mathbb{P}_\beta(X > x) = \int_x^\infty \frac{1}{\beta}e^{-y/\beta}dy = \frac{1}{\beta}\left[-\beta e^{-y/\beta}\right]_{y=x}^{\infty} = e^{-x/\beta} = \tau(\beta).$$

By the invariance principle, the maximum likelihood estimator of $\tau(\beta)$ is given by

$$\tau\left(\hat{\beta}_{ML}\right) = \mathbb{P}_{\hat{\beta}_{ML}}(X > x) = e^{-x/\hat{\beta}_{ML}}.$$

To calculate the probability of more than 20mm of rainfall, we note that in our sample $\hat{\beta}_{ML} = 120/15 = 8$. Hence,

$$\mathbb{P}_8(X > 20) = e^{-20/8} \approx 0.082.$$

Since this probability is larger than 5%, you advise the lecturer to stay at home or wear a different costume.

(g) The MLE would be very difficult to derive due to the Gamma function that appears in there. Therefore, we proceed with the derivation of the MME. Since both parameters are unknown, we will need two moments. Let $\boldsymbol{\theta} = (\alpha, \beta)$. Then the first moment is given by $\mathbb{E}_\theta(X_1) = \alpha\beta)$ and the second moment was derived to be $\mathbb{E}_\theta(X_1^2) = \alpha(1+\alpha)\beta^2$. Hence we need to solve the system

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} X_i &\stackrel{s}{=} \mathbb{E}_\theta(X_1) = \alpha\beta \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 &\stackrel{s}{=} \mathbb{E}_\theta(X_1^2) = \alpha(1+\alpha)\beta^2\end{aligned}$$

The first equation gives $\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}}$. Plugging this into the second equation, we obtain

$$\overline{X^2} = \left(\frac{\bar{X}}{\hat{\beta}} + \left(\frac{\bar{X}}{\hat{\beta}}\right)^2\right)\hat{\beta}^2 = \hat{\beta}\bar{X} + \bar{X}^2 \Rightarrow \hat{\beta} = \frac{\overline{X^2} - \bar{X}^2}{\bar{X}}.$$

Hence, plugging this back into the solution for α, we obtain

$$\hat{\alpha} = \frac{\bar{X}}{\frac{\overline{X^2} - \bar{X}^2}{\bar{X}}} = \frac{\bar{X}^2}{\overline{X^2} - \bar{X}^2}.$$

□