Exam Statistics

Bachelor Econometrics and Operations Research
Bachelor Econometrics and Data Science
School of Business and Economics
Tuesday, March 27, 2019

Exam:	Statistics
Code:	E_EOR1_STAT
Coordinator:	M.H.C. Nientker
Co-reader:	J.M. Sneek
Date:	March 27, 2019
Time:	08:45
Duration:	2 hours
Calculator:	Not allowed
Graphical calculator:	Not allowed
Number of questions:	3
Type of questions:	Open
Answer in:	English
Credit score:	100 credits counts for a 10, you get 1 free credit
Grades:	Made public within 10 working days
Number of pages:	3, 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 (12 points). During class we assumed that the heights of men can be modelled accurately by a normal distribution and thus we used the statistical model $\{g_{(\mu,\sigma^2)} \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$, where

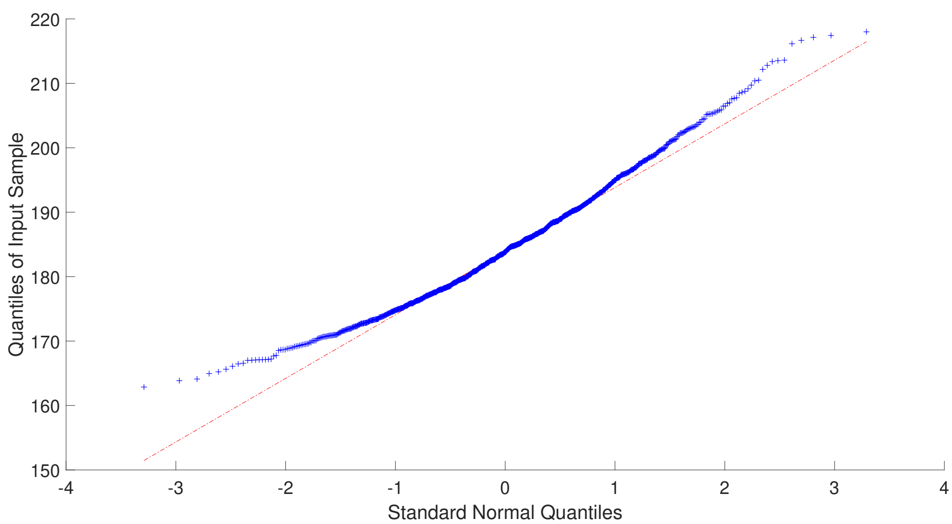
$$g_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \qquad x \in \mathbb{R}.$$

In practice we would like to check this assumption to make sure that our analysis is valid.

(6 points) a. Let $Z \sim \mathrm{N}(0, 1)$. Show that the set of normal distributions $\{g_{(\mu,\sigma^2)} \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$ is equal to the location-scale family of Z.

(6 points) b. We have measured one thousand random men and constructed a QQ-plot. Based on Figure 1, what do you conclude concerning the normality assumption we made? Why?

Figure 1: QQ-plot of data against the standard normal distribution



Solution.

a. Let $\mu \in \mathbb{R}$ and $\sigma > 0$ and define $Y = \mu + \sigma Z$. Then Y has cdf

$$G_{\mu,\sigma}(x) = P(Y \le x) = P(\mu + \sigma Z \le x) = P\left(Z \le \frac{z-\mu}{\sigma}\right) = G\left(\frac{z-\mu}{\sigma}\right).$$

The set of distributions $\{G_{\mu,\sigma} \mid \mu \in \mathbb{R}, \sigma > 0\}$ is called the location-scale family of Z. If $Z \sim N(0,1)$, then we know that $\mu + \sigma Z \sim N(\mu, \sigma^2)$ and conversely if $Y \sim N(\mu, \sigma^2)$, then $-\mu + \frac{1}{\sigma} Y \sim N(0,1)$ thus the location-scale family of Z is the set of all normal distributions.
3 points for understanding what the location scale family is, 3 points for showing the result.

b. The points in the QQ-plot seem to not lie on a straight line, therefore we conclude that the normality assumption probably does not hold.
6 points for a correct conclusion, based on interpretation of the graph.

Question 2 (42 points). Let $X_1, \ldots, X_n$ be an independent and identically distributed sequence of random variables from a population in $\{g_\theta \mid \theta > 0\}$, where

$$g_\theta(x) = \frac{1}{\theta}, \qquad 0 \le x \le \theta.$$

(6 points) a. Derive the first moment $\mathbb{E}_\theta X_1$ and show that the moment estimator $\hat{\theta}_{MOM}$ of θ_0 is equal to $2\overline{X}$.

(9 points) b. Describe a weakness of $\hat{\theta}_{MOM}$ in this case and give a suggestion on how to improve it.

(12 points) c. Show that the ML estimator $\hat{\theta}_{ML}$ of θ_0 is equal to $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

(6 points) d. Describe a weakness of $\hat{\theta}_{ML}$ in this case. Is this estimator biased for θ_0? Do not use any calculations to answer this question.

(9 points) e. Suppose that we know that $\theta_0 = 6$ and that we have simulated $x = (1, 1, 4, 4, 3, 4)$. Show that $\hat{\theta}_{MOM}$ is closer to θ_0 than $\hat{\theta}_{ML}$ and conclude that the moment estimator is closer for this realization. Why can't we conclude that $\hat{\theta}_{MOM}$ is better than $\hat{\theta}_{ML}$ in general? Propose a method to compare estimators and explain why it is better than comparing for a given realization.

Solution.

a. We have

$$\mathbb{E}(X_1) = \int_0^\theta x\frac{1}{\theta}dx = \frac{1}{\theta}\left[\frac{1}{2}x^2\right]_0^\theta = \frac{\theta^2}{2\theta} = \frac{\theta}{2}.$$

Next, we have to solve the equation $\overline{X} = \theta/2$, so $\hat{\theta} = 2\overline{X}$.
3 points for the expectation, 2 point for the right equation, 1 point for the moment estimator.

b. Note that θ_0 must guaranteed be at least as large as the maximum $x_{(n)}$ in the observed sample x, as the domain of the pdf depends on θ_0. If the moment estimator $\hat{\theta}_{MOM}$ is smaller than the maximum, i.e. $2\bar{x} < x_{(n)}$, then we are setting θ_0 too low as we are not using this information. We can improve by for example taking the maximum of the two random variable $\max\{2\bar{x}, x_{(n)}\}$.
6 points for correctly pointing out a flaw in $\hat{\theta}_{MOM}$, 3 points for suggesting an improvement.

c. The domain of our random variables depends on the unknown parameter θ_0. Therefore the likelihood is discontinuous and thus we cannot find the ML estimator by looking for stationary points. Instead we apply direct maximisation:

$$\begin{aligned}
L(\theta \mid x) &= \prod_{i=1}^n g_\theta(x)\mathbb{1}_{\{0 \le x_i \le \theta\}} = \prod_{i=1}^n \frac{1}{\theta}\mathbb{1}_{\{0 \le x_i \le \theta\}} \\
&= \left(\prod_{i=1}^n \frac{1}{\theta}\right)\left(\prod_{i=1}^n \mathbb{1}_{\{0 \le x_i\}}\right)\left(\prod_{i=1}^n \mathbb{1}_{\{x_i \le \theta\}}\right) \\
&= \frac{1}{\theta^n}\mathbb{1}_{\{0 \le x_{(1)}\}}\mathbb{1}_{\{x_{(n)} \le \theta\}}
\end{aligned}$$

The function $\frac{1}{\theta^n}$ is decreasing as a function of θ, so we would like to make θ as small as possible. However, if $\theta < x_{(n)}$, then the indicator function collapses to zero and thus the maximum is attained at $\tilde{\theta} = x_{(n)}$. We conclude that

$$\hat{\theta}_{ML} = X_{(n)}.$$

3 points for realising that we cannot differentiate, 2 points for writing the joint likelihood as a product of univariate densities, 2 points for splitting the $\frac{1}{\theta^n}$, 2 points for rewriting the product of indicator functions, 2 points for the correct understanding of the shape of the likelihood, 1 point for the correct conclusion.

d. The maximum will always underestimate θ_0, as θ_0 has to be at least as large as $x_{(n)}$, as the domain of the pdf depends on θ_0. (remember that we have seen that $\frac{n+1}{n}x_{(n)}$ is the UMVU estimator). We can immediately conclude that the estimator must therefore have a positive bias.
4 points for correctly pointing out a flaw in $\hat{\theta}_{ML}$, 2 points for the correct bias deduction.

e. We have $\theta_0 - \hat{\theta}_{MOM} = 6 - \frac{2\times 17}{6} = \frac{1}{3}$ and $\theta_0 - \hat{\theta}_{ML} = 6 - 4 = 2$. Therefore the moment estimator is closer. We cannot conclude that $\hat{\theta}_{MOM}$ is better than $\hat{\theta}_{ML}$ in general, because there are also many realisations for which $\hat{\theta}_{ML}$ is closer to θ_0 than $\hat{\theta}_{MOM}$. During class we have discussed to measure performance on average, that is, in expectation. A possible example of such a measure is the mean squared error $\mathbb{E}_\theta\left[(\hat{\theta} - \theta)^2\right]$.
3 points for correctly showing the moment estimator is better, 3 points for correctly mentioning that many other realisations can give the opposite result, 3 points for suggesting a performance measure that uses the expectation.

Question 3 (45 points). Let $X = (X_1, \dots, X_n)$ be an independent and identically distributed sequence of random variables from a population in $\{g_\theta \mid \theta > 0\}$, where

$$g_\theta(x) = \theta x^{\theta - 1} \qquad 0 \leq x \leq 1.$$

(12 points) a. Show that the maximum likelihood estimator of θ_0 is given by

$$\hat{\theta}_{ML} = -1/\overline{\log X} = -\left(\frac{1}{n}\sum_{i=1}^{n} \log X_i\right)^{-1}.$$

(9 points) b. Show that $T(X) = \sum_{i=1}^{n} \log X_i$ is a sufficient statistic for θ_0 using the factorization theorem.

(12 points) c. Calculate the Cramér-Rao lower bound for θ_0. Hint: use the theory on exponential families.

(12 points) d. Find the UMVU estimator for θ_0. Show that part c won't help here and explain why. You are allowed to use that

$$\mathbb{E}_\theta(1/T(X)) = -\frac{1}{n-1}\theta \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}_\theta(1/T(X)) = \frac{1}{(n-1)^2(n-2)}\theta^2.$$

SOLUTION.

a. For the likelihood we have:

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} \theta x_i^{\theta - 1}.$$

This gives us the log-likelihood:

$$\log L(\theta \mid x) = \sum_{i=1}^{n} \log\left(\theta x_i^{\theta-1}\right) = n \log \theta + (\theta - 1)\sum_{i=1}^{n} \log x_i.$$

We obtain derivatives

$$\frac{d}{d\theta} \log L(\theta \mid x) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i.$$

$$\frac{d^2}{d\theta^2} \log L(\theta \mid x) = -\frac{n}{\theta^2}.$$

We find the stationary point by setting the first derivative to zero and obtain

$$\tilde{\theta} = -\frac{n}{\sum_{i=1}^{n} \log x_i} = -1/\overline{\log x}.$$

When we evaluate the second derivative at this point we get

$$\frac{d^2}{d\theta^2} \log L(\tilde{\theta} \mid x) = -\frac{n}{\tilde{\theta}^2} < 0,$$

since $\overline{\log x} < 0$ and thus $\tilde{\theta} > 0$ with probability one. We have a unique stationary point, which is a local maximum. Therefore we have found the global maximum. We conclude that

$$\hat{\theta}_{ML} = -1/\overline{\log X}.$$

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

b. We have by independence

$$f_\theta(x) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \exp\left((\theta-1)\sum_{i=1}^{n} \log x_i\right).$$

We can define $k_\theta(T(x)) = \theta^n \exp\left((\theta-1)\sum_{i=1}^{n} \log x_i\right)$ and $h(x) = 1$. We conclude by the factorization theorem that T is a sufficient statistic.
3 points for splitting the joint density into a product of univariate densities, 3 points for writing it as a function of $T(x)$, 3 points for correctly applying the factorization theorem.

c. We have

$$g_\theta(x) = \theta x^{\theta-1} = \theta e^{(\theta-1)\log x}.$$

We therefore have a one dimensional exponential family with $h(x) = 1, c(\theta) = \theta, t_1(x) = \log x$ and $w_1(\theta) = \theta - 1$.

We calculate the Cramér-Rao lower bound:

$$\begin{aligned}
\log g_\theta(x) &= \log \theta x^{\theta-1} = \log\theta + (\theta-1)\log x. \\
\frac{d}{d\theta}\log g_\theta(x) &= \frac{1}{\theta} + \log x. \\
\frac{d^2}{d\theta^2}\log g_\theta(x) &= -\frac{1}{\theta^2}. \\
i_\theta &= -\mathbb{E}\left(\frac{d^2}{d\theta^2}\log g_\theta(X_1)\right) = \frac{1}{\theta^2}. \\
B(\theta) &= \frac{\tau'(\theta)^2}{n i_\theta} = \frac{1}{n\frac{1}{\theta^2}} = \frac{\theta^2}{n}.
\end{aligned}$$

3 points for correctly showing its an exponential family, 2 points for the log univariate pdf, 2 points for each derivative, 3 points for obtaining the correct bound.

d. We already know from b that $T(X)$ is sufficient. It follows from the exponential family proof in c, and the fact that $\Theta = (0,\infty) \subseteq \mathbb{R}$ clearly contains an open set, that

$$T(X) = \sum_{i=1}^{n} t_1(X_i) = \sum_{i=1}^{n} \log(X_i)$$

is complete. We know that $\mathbb{E}_\theta(1/T(X)) = -\frac{1}{n-1}\theta$, so we use

$$\phi(T) = -\frac{n-1}{T(X)} = \frac{n-1}{n}\hat{\theta}_{ML}.$$

Now $\phi(T)$ is unbiased by construction and a function of the sufficient and complete statistic T. It follows that $\phi(T)$ must be UMVU by the Lehmann-Schefé theorem.

Next, for the variance of the UMVU estimator we find

$$\begin{aligned}\mathbb{V}\mathrm{ar}_\theta\, \phi(T) &= \mathbb{V}\mathrm{ar}_\theta \left(-\frac{n-1}{T(X)}\right) = (n-1)^2\, \mathbb{V}\mathrm{ar}_\theta \left(\frac{1}{T(X)}\right) \\ &= (n-1)^2 \frac{1}{(n-1)^2(n-2)}\theta^2 = \frac{\theta^2}{n-2} > \frac{\theta^2}{n} = B(\theta).\end{aligned}$$

We see that the UMVU estimator has strictly larger variance than the Cramér-Rao lower bound. It follows that we can never find it via question b, as any unbiased estimator has variance larger than the UMVU estimator and thus than the Cramér-Rao lower bound, so no unbiased estimator attains the lower bound.
2 points for correctly finding the complete and sufficient statistic, 2 points for finding the function ϕ, 2 points for the correct conclusion that $\phi(T)$ must be UMVU by the Lehmann-Schefé theorem, 3 points for calculating the variance, 3 points for correctly concluding that the UMVU estimator doesn't attain the Cramér-Rao lower bound and thus that we cannot use this bound to find the UMVU estimator.