Vrije Universiteit Amsterdam
Department of Mathematics

Exam Mathematical Statistics (XB_0049)
December 20, 2022, 18.45–21.00

You are not allowed to use calculators, phones, laptops or other tools. Clearly put your name and student number on all sheets that you submit. Your answer may contain quantiles from a distribution. Expressions for densities and quantiles are given on page 3 of the exam.
Unless stated differently, always add an explanation to your answer.

1. Let $X_1, ..., X_n$ be independent random variables with density

$$p_\theta(x) = \begin{cases} 0 & x < 1 \\ \theta x^{-(\theta+1)} & x \geq 1, \end{cases}$$

where $\theta > 1$ is unknown.

   (a) [2 pt]. Determine the method of moments estimator for θ.
   (b) [2 pt]. Find a one-dimensional sufficient statistic for θ.
   (c) [2 pt]. Determine the maximum likelihood estimator for θ.
   (d) [3 pt]. Determine the Bayes estimator for θ with respect to the prior $\pi(\theta) = \mathrm{e}^{1-\theta}$, for $\theta > 1$ and 0 elsewhere.

2. Let $X_1, \ldots, X_n$ be independent random variables with $N(0, \sigma^2)$ distribution with $\sigma^2 > 0$ unknown.

   (a) [2 pt]. Show that $\left(\sum_{i=1}^n X_i^2\right)/\sigma^2$ is a pivot for σ^2.
   (b) [2 pt]. Determine a confidence interval for σ^2 of confidence level $1 - \alpha$, based on the distribution of the pivot in part (a).

3. Let $X_1, \ldots, X_n$ be independent random variables with probability density

$$p_\theta(x) = \theta^2 x \mathrm{e}^{-\theta x}, \quad x \geq 0$$

and $p_\theta(x) = 0$, for $x < 0$, where $\theta > 0$ is unknown. This density corresponds to that of the Gamma distribution with parameters 2 and θ. It can be shown that the maximum likelihood estimator for θ is given by $\widehat{\theta} = 2/\overline{X}$ (in the following you don't need to show this yourself).

   (a) [3 pt]. Determine the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ and show that the approximate confidence interval for θ is given by the set

$$\left\{\theta > 0 : \log(\theta\overline{X}/2) - \theta\overline{X}/2 \geq -1 - \chi^2_{1,1-\alpha}/4n\right\}.$$

   (b) [3 pt]. Compute the Fisher information and give an approximate confidence interval for θ with confidence level $1 - \alpha$ based on the asymptotic distribution of the maximum likelihood estimator.

4. Consider the linear regression model

$$Y_i = \beta x_i + e_i, \quad \text{for } i = 1, \ldots, n,$$

where $e_1, \ldots, e_n$ are independent random variables with the $N(0, \sigma^2)$ distribution.

(a) [2 pt]. Show that the least squares estimator for β is given by

$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

(b) [1 pt]. Another estimator for β is given by

$$\widetilde{\beta} = \frac{\overline{Y}}{\overline{x}} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}.$$

Show that $\widetilde{\beta}$ is an unbiased estimator for β.

(c) [3 pt]. The mean squared error for $\widehat{\beta}$ is given by

$$\text{MSE}(\widehat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

Determine the mean squared error for $\widetilde{\beta}$ and show that it is larger than $\text{MSE}(\widehat{\beta})$.

You may use

- that the least squares estimator is unbiased for β;
- the inequality of Cauchy-Schwarz:

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right).$$

5. Let $X_1, \ldots, X_n$ be independent random variables with distribution function

$$F_\theta(x) = \begin{cases} 1 - \mathrm{e}^{\theta - x} & x \geq \theta \\ 0 & x < \theta, \end{cases}$$

for some unknown $\theta > 0$. We want to test $H_0 : \theta \leq 1$ against $H_1 : \theta > 1$ with test statistic

$$T = X_{(1)} = \min\{X_1, \ldots, X_n\}$$

at significance level $0 < \alpha_0 < 1$. We reject $H_0 : \theta \leq 1$ for large values of $X_{(1)}$.

(a) [2 pt]. Show that

$$\mathrm{P}_\theta(X_{(1)} \geq t) = \begin{cases} \mathrm{e}^{n(\theta - t)} & t \geq \theta \\ 1 & t < \theta. \end{cases}$$

(b) [3 pt]. Give the definition of the size of the test and show that the critical region is given by

$$K = \left\{(x_1, \ldots, x_n) : x_{(1)} \geq 1 - \frac{\log \alpha_0}{n}\right\}.$$

Distributions and notation

- If $X \sim N(\mu, \sigma^2)$, then its density is given by $p(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2)}{2\sigma^2}\right)$ and $\Phi(x) := \mathbb{P}(X \leq x)$. The (lower) quantile ξ_α is defined by $\Phi(\xi_\alpha) = \alpha$, for $\alpha \in (0, 1)$.
- If $X \sim \chi^2_k$, then $\chi^2_{k,\alpha}$ is defined by $\mathbb{P}(X \leq \chi^2_{k,\alpha}) = \alpha$.
- If $X \sim t_k$, then $t_{k,\alpha}$ is defined by $\mathbb{P}(X \leq t_{k,\alpha}) = \alpha$.
- If $X \sim Ga(\alpha, \beta)$, then its density is given by $p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{(0,\infty)}(x)$. $\mathbb{E}X = \alpha/\beta$, $\mathrm{Var}(X) = \alpha/\beta^2$. The special case of $\alpha = 1$ corresponds to the exponential distribution with parameter β.
- If $X \sim Bin(n, p)$, then $\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, for $x \in \{0, 1, \ldots, n\}$. $\mathbb{E}X = np$, $\mathrm{Var}(X) = np(1-p)$.
- If $X \sim Beta(\alpha, \beta)$, then its density is given by $p(x) = B(\alpha, \beta)^{-1} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{[0,1]}(x)$. $\mathbb{E}X = \alpha/(\alpha+\beta)$, $\mathrm{Var}(X) = \alpha\beta/\left((\alpha+\beta)^2(\alpha+\beta+1)\right)$

1. (a) The expectation of X_1 is

$$\mathrm{E}_\theta X_1 = \int_1^\infty x p_\theta(x)\,\mathrm{d}x = \int_1^\infty x\theta x^{-(\theta+1)}\,\mathrm{d}x$$
$$= \theta\int_1^\infty x^{-\theta}\,\mathrm{d}x = \theta\left[-\frac{x^{-\theta+1}}{\theta-1}\right]_1^\infty = \frac{\theta}{\theta-1}.$$

The method of moments estimator is the solution of

$$\overline{X} = \frac{\theta}{\theta-1} \Leftrightarrow (\theta-1)\overline{X} = \theta \Leftrightarrow \theta(\overline{X}-1) = \overline{X} \Leftrightarrow \theta = \frac{\overline{X}}{\overline{X}-1}$$

(b)

$$L(\theta; X_1,\ldots,X_n) = \prod_{i=1}^n \theta X_i^{-(\theta+1)} = \theta^n\left(\prod_{i=1}^n X_i\right)^{-\theta-1} = g_\theta(V(X_1,\ldots,X_n))h(X_1,\ldots,X_n)$$

We can hence take $h \equiv 1$ and $V(X_1,\ldots,X_n) = \prod_{i=1}^n X_i$ which, by the factorisation theorem, is sufficient for θ.

(c) To determine the ML estimator we compute the loglikelihood

$$\log L(\theta; X_1,\ldots,X_n) = n\log\theta - (\theta+1)\sum_{i=1}^n \log X_i$$

Differentiation gives

$$\frac{\mathrm{d}L}{\mathrm{d}\theta} = \frac{n}{\theta} - \sum_{i=1}^n \log X_i = 0 \quad\Leftrightarrow\quad \theta = \frac{n}{\sum_{i=1}^n \log X_i}.$$

Since the second derivative is $\partial^2 \log L/\partial\theta^2 = -n/\theta^2 < 0$, this is a maximum. Hence, the ML estimator is given by $n/\sum_{i=1}^n \log X_i$.

(d) The posterior density is proportional to

$$p_\theta(X_1,\ldots,X_n)\pi(\theta) = \prod_{i=1}^n \theta X_i^{-(\theta+1)}\mathrm{e}^{1-\theta}$$
$$= \theta^n \mathrm{e}^{-(\theta+1)\sum_{i=1}^n \log X_i}\mathrm{e}^{1-\theta}$$
$$= \theta^n \mathrm{e}^{-\theta(1+\sum_{i=1}^n \log X_i)}\mathrm{e}^{1-\sum_{i=1}^n \log X_i)}$$
$$\propto \theta^n \mathrm{e}^{-\theta(1+\sum_{i=1}^n \log X_i}$$

The right hand side is proportional to the gamma density

$$\frac{\theta^{\alpha-1}\lambda^\alpha \mathrm{e}^{-\lambda\theta}}{\Gamma(\alpha)}$$

with $\alpha = n+1$ and $\lambda = 1 + \sum_{i=1}^n \log X_i$. Hence the posterior density must be a gamma distribution with these parameters and the Bayes estimator is the corresponding mean

$$\frac{\alpha}{\lambda} = \frac{n+1}{1+\sum_{i=1}^n \log X_i}$$

2. (a) First note that for each $i = 1, \ldots, n$, the random variable $Z_i = X_i/\sigma$ has a $N(0,1)$ distribution, which does not depend on σ^2. This means that every combination of $X_1/\sigma, \ldots, X_n/\sigma$ has a distribution that no longer depends on σ^2.
In particular, the distribution of

$$\frac{\sum_{i=1}^n X_i^2}{\sigma^2} = \sum_{i=1}^n (X_i/\sigma)^2$$

no longer depends on σ^2, which means that it is a pivot for σ^2.

(b) From part (a) we have that

$$\frac{\sum_{i=1}^n X_i^2}{\sigma^2} = \sum_{i=1}^n (X_i/\sigma)^2 = \sum_{i=1}^n Z_i^2$$

which is the sum of n independent squared $N(0,1)$ distributed random variables. By definition this has a χ^2-distribution with parameter n.
This means that

$$\mathrm{P}\left(\chi^2_{n,\alpha/2} \le \frac{\sum_{i=1}^n X_i^2}{\sigma^2} \le \chi^2_{n,1-\alpha/2}\right) = 1-\alpha$$

or equivalently

$$\mathrm{P}\left(\frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,1-\alpha/2}} \le \sigma^2 \le \frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,\alpha/2}}\right) = 1-\alpha.$$

The confidence interval of level $1-\alpha$ for σ^2 is then given by

$$\left[\frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,1-\alpha/2}}, \frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,\alpha/2}}\right].$$

3. (a) The likelihood ratio statistic is defined by

$$\begin{aligned}
\lambda_n &= \frac{\sup_{\theta\in\Theta} L(\theta)}{\sup_{\theta\in\Theta_0} L(\theta)} \\
&= \frac{L(\widehat{\theta})}{L(\theta_0)} \\
&= \frac{\prod_{i=l}^n \widehat{\theta}^2 X_i \mathrm{e}^{-\widehat{\theta} X_i}}{\prod_{i=l}^n \theta_0^2 X_i \mathrm{e}^{-\theta_0 X_i}} \\
&= \frac{\widehat{\theta}^{2n}\left(\prod_{i=1}^n X_i\right)\mathrm{e}^{-\widehat{\theta}\sum_{i=1}^n X_i}}{\theta_0^{2n}\left(\prod_{i=1}^n X_i\right)\mathrm{e}^{-\theta_0\sum_{i=1}^n X_i}} \\
&= \left(\frac{\widehat{\theta}}{\theta_0}\right)^{2n} \exp\left(-n(\widehat{\theta}-\theta_0)\overline{X}\right) \\
&= \left(\frac{2}{\theta_0\overline{X}}\right)^{2n} \exp\left(-n(2-\theta_0\overline{X})\right).
\end{aligned}$$

This means that

$$2\log\lambda_n = 4n(\log 2 - \log(\theta_0\overline{X})) - 2n(2-\theta_0\overline{X}).$$

Since the likelihood ratio test rejects for large values of λ_n and that the distribution of $2\log\lambda_n$ can be approximated by a χ^2_1 distribution, if follows that the likelihood ratio confidence region of confidence level $1-\alpha$ is then given by the set

$$\begin{aligned}
&\left\{\theta : 4n(\log 2 - \log(\theta\overline{X})) - 2n(2-\theta\overline{X}) \le \chi^2_{1,1-\alpha}\right\}\\
&= \left\{\theta : (\log 2 - \log(\theta\overline{X})) - (1-\theta\overline{X}/2) \le \chi^2_{1,1-\alpha}/4n\right\}\\
&= \left\{\theta : \log(\theta\overline{X}) - \log 2 + (1-\theta\overline{X}/2) \ge -\chi^2_{1,1-\alpha}/4n\right\}\\
&= \left\{\theta : \log(\theta\overline{X}/2) - \theta\overline{X}/2 \ge -1 - \chi^2_{1,1-\alpha}/4n\right\}.
\end{aligned}$$

(b) We have

$$\begin{aligned}
\ell_\theta(x) &= \log p_\theta(x) = 2\log\theta + \log x - \theta x\\
\dot{\ell}_\theta(x) &= \frac{\partial \ell_\theta(x)}{\partial\theta} = \frac{2}{\theta} - x\\
\ddot{\ell}_\theta(x) &= \frac{\partial^2 \ell_\theta(x)}{\partial\theta^2} = -\frac{2}{\theta^2}.
\end{aligned}$$

The Fisher information can be computed by means of the definition

$$i_\theta = \mathrm{var}_\theta \dot{\ell}_\theta(X_1) = \mathrm{var}_\theta\left(\frac{2}{\theta} - X_1\right) = \mathrm{var}_\theta X_1.$$

As the density corresponds to that of a Gamma distribution with parameters 2 and θ, the variance is given in the book: $2/\theta^2$.

It is easier to use directly

$$-i_\theta = \mathrm{E}_\theta\, \ddot{\ell}_\theta(X_1) = -\left(-\frac{2}{\theta^2}\right) = \frac{2}{\theta^2}.$$

The approximated confidence interval is given by

$$\widehat{\theta} \pm \frac{1}{\sqrt{n\widehat{i_\theta}}}\xi_{1-\alpha/2}$$

where $\widehat{\theta}$ is the maximum likelihood estimator and i_θ can be estimates, either by plug-in

$$\widehat{i_\theta} = i_{\widehat{\theta}} = \frac{2}{\widehat{\theta}^2}$$

or by means of the observed information

$$\widehat{i_\theta} = -\frac{1}{n}\sum_{i=1}^n \ddot{\ell}_{\widehat{\theta}}(X_i) = -\frac{1}{n}\sum_{i=1}^n\left(-\frac{2}{\widehat{\theta}^2}\right) = \frac{2}{\widehat{\theta}^2}.$$

This results in confidence interval

$$\widehat{\theta} \pm \frac{\widehat{\theta}}{\sqrt{2n}}\xi_{1-\alpha/2} = \frac{2}{\overline{X}} \pm \frac{2}{\overline{X}}\frac{\xi_{1-\alpha/2}}{\sqrt{2n}}$$

4. (a) The least squares estimator is the solution of

$$\min_{\beta} \sum_{i=1}^{n} (Y_i - \beta x_i)^2$$

We find that

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{n} (Y_i - \beta x_i)^2 = -2 \sum_{i=1}^{n} (Y_i - \beta x_i) x_i = -2 \sum_{i=1}^{n} x_i Y_i + 2\beta \sum_{i=1}^{n} x_i^2.$$

This is equal to zero if

$$\beta = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

Furthermore, the second derivative is $\partial^2/\partial\beta^2 = 2\sum_{i=1}^n x_i^2 > 0$, so that this is a minimum. This means that the least squares estimator is given by

$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

(b)

$$\mathrm{E}\,\widetilde{\beta} = \mathrm{E}\left[\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}\right] = \frac{\sum_{i=1}^n \mathrm{E}[Y_i]}{\sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n (\beta x_i + \mathrm{E}\, e_i)}{\sum_{i=1}^n x_i} = \frac{\beta \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i} = \beta.$$

(c) Because $\widehat{\beta}$ is unbiased the mean squared error is given by

$$\begin{aligned}
\mathrm{MSE}(\widetilde{\beta}) = \mathrm{var}(\widetilde{\beta}) &= \mathrm{var}\left(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}\right) \\
&= \frac{1}{\left(\sum_{i=1}^n x_i\right)^2} \sum_{i=1}^{n} \mathrm{var}(Y_i) = \frac{1}{\left(\sum_{i=1}^n x_i\right)^2} \sum_{i=1}^{n} \mathrm{var}(e_i) \\
&= \frac{1}{\left(\sum_{i=1}^n x_i\right)^2} \sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{\left(\sum_{i=1}^n x_i\right)^2}.
\end{aligned}$$

We must show that

$$\begin{aligned}
\mathrm{MSE}(\widehat{\beta}) \le \mathrm{MSE}(\widetilde{\beta}) &\Leftrightarrow \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \le \frac{n\sigma^2}{\left(\sum_{i=1}^n x_i\right)^2} \\
&\Leftrightarrow \left(\sum_{i=1}^{n} x_i\right)^2 \le n \sum_{i=1}^{n} x_i^2
\end{aligned}$$

This precisely Cauchy-Schwarz with $a_i = 1$ and $b_i = x_i$.

5. (a) Because $X_1, \ldots, X_n$ are independent with the same distribution function F_θ, we have that

$$\begin{aligned}
\mathrm{P}_\theta(X_{(1)} \ge c_{\alpha_0}) &= \mathrm{P}_\theta(X_1 \ge c_{\alpha_0}, \ldots, X_1 \ge c_{\alpha_0}) \\
&= \mathrm{P}_\theta(X_1 \ge c_{\alpha_0}) \cdots \mathrm{P}_\theta(X_1 \ge c_{\alpha_0}) \\
&= (1 - F_\theta(c_{\alpha_0}))^n \\
&= \begin{cases} \mathrm{e}^{n(\theta - c_{\alpha_0})} & \theta \le c_{\alpha_0} \\ 1 & \theta > c_{\alpha_0}. \end{cases}
\end{aligned}$$

(b) Since we reject for large values of T, the critical region is of the form $K_T = [c, \infty)$. This means that the size of the test is defined by

$$\alpha = \sup_{\theta \le 1} \mathrm{P}_\theta(X_{(1)} \ge c_{\alpha_0}) = \sup_{\theta \le 1} \mathrm{e}^{n(\theta - c_{\alpha_0})}.$$

where c_{α_0} should be chosen as small as possible such that $\alpha \le \alpha_0$.
The function $\theta \mapsto \mathrm{e}^{n(\theta - c_{\alpha_0})}$ increases on $\theta \in (0, c_{\alpha_0}]$ and then remains constant equal to 1. This means that the largest value of $\mathrm{P}_\theta(X_{(1)} \ge c_{\alpha_0})$ that is less than α_0, is attained at $\theta = 1$, so that

$$\alpha = \sup_{\theta \le 1} \mathrm{e}^{n(\theta - c_{\alpha_0})} = \mathrm{e}^{n(1 - c_{\alpha_0})}.$$

To find c_{α_0}, we must solve

$$\mathrm{e}^{n(1 - c_{\alpha_0})} \le \alpha_0 \Leftrightarrow n(1 - c_{\alpha_0}) \le \log \alpha_0 \Leftrightarrow c_{\alpha_0} \ge 1 - \frac{\log \alpha_0}{n}.$$

Since we must choose c_{α_0} as small as possible, we conclude that $c_{\alpha_0} = 1 - (\log \alpha_0)/n$, and

$$K = \left\{ (x_1, \ldots, x_n) : x_{(1)} \ge 1 - \frac{\log \alpha_0}{n} \right\}.$$