Delft University of Technology
Faculty EEMCS, Department DIAM

van Mourik Broekmanweg 6
2628 XE Delft

Practice Final Exam AM2080

This written final exam contains 5 questions, each question counts for 20% of the final grade of the written final exam. Within each question the different parts have equal weights. You are only allowed to use the sheets with information on probability distributions, R-commands, and tables. You are not allowed to use any books or notes.

1. Let $X_1, \dots, X_n$ be independent identically distributed random variables with probability density $p_\theta(x) = \mathrm{e}^{\theta - x} 1_{\{x \geq \theta\}}$, where $\theta > 0$ is unknown.

   You may use that $X_i - \theta$ has an exponential distribution with parameter 1, for $i = 1, \dots, n$, and that $X_{(1)} - \theta$ has an exponential distribution with parameter n, where $X_{(1)} = \min_{1 \leq i \leq n} X_i$.

   (a) Determine the method of moment estimator for θ.

   (b) For which $c \in \mathbb{R}$ does the estimator $X_{(1)} + c$ have the smallest mean squared error for estimating θ?

2. Let $X_1, \dots, X_n$ be independent identically distributed random variables with a uniform distribution on the interval $[-\theta, \theta]$, where $\theta > 0$ is unknown. We test $H_0 : \theta \leq 1$ against $H_1 : \theta > 1$ with test statistic $T = \max_{1 \leq i \leq n} |X_i|$ at significance level α_0. The critical region is of the form $K_T = [c_{\alpha_0}, \infty)$, for some $c_{\alpha_0} > 0$.

   (a) Show that

   $$\mathrm{P}_\theta \left(\max_{1 \leq i \leq n} |X_i| \leq t \right) = \begin{cases} 0 & t < 0; \\ (t/\theta)^n & t \in [0, \theta]; \\ 1 & t > \theta. \end{cases}$$

   (b) Give the definition of the size of the test based on T and determine c_{α_0}.

   (c) Give the definition of the power function of the test based on T and determine an expression in terms of n, α_0, and θ.

3. Let $X_1, \dots, X_n$ be a sample from the distribution with probability density

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

   and 0 for $x < 0$. The parameter $\theta > 0$ is unknown. We can prove that $X_1^2, \dots, X_n^2$ are exponentially distributed with parameter θ and that if $\theta = 1$, the random variable $2 \sum_{i=1}^n X_i^2$ has a chi-square distribution with $2n$ degrees of freedom.

(a) Show that $2\theta \sum_{i=1}^n X_i^2$ is a pivot.

(b) Determine a 95% confidence interval for θ by means of the pivot in part (a).

4. Let $X_1, \ldots, X_n$ be independent identically distributed random variables with probability density

$$p_\theta(x) = \theta x^{\theta-1} \quad \text{for } x \in (0,1)$$

and 0 otherwise, where $\theta > 0$ is unknown. We test $H_0 : \theta = 2$ against $H_1 : \theta \neq 2$. You may use that the maximum likelihood estimator for θ is given by $\widehat{\theta} = -n / \sum_{i=1}^n \log X_i$.

(a) Give the definition of the likelihood ratio statistic and show that it is a strictly increasing function of $\widehat{\theta}$.

(b) Compute the Fisher information and give an approximated 95% confidence interval for θ based on the maximum likelihood estimator $\widehat{\theta}$.

5. Consider the linear regression model

$$Y_i = \theta + \frac{\theta}{2} x_i^2 + e_i, \quad \text{for } i = 1, \ldots, n,$$

where $e_1, \ldots, e_n$ are independent identically distributed random variables with a $N(0,1)$ distribution and $\theta \in \mathbb{R}$ is unknown.

(a) Compute the least squares estimator $\widehat{\theta}$ for θ.

(b) Show that $\widehat{\theta}$ is unbiased for θ.

Antwoorden:

1a First determine the expectation. Since $X_1 - \theta$ has an exponential distribution with parameter 1, we have

$$\mathrm{E}_\theta X_1 = \mathrm{E}_\theta(X_1 - \theta) + \theta = 1 + \theta.$$

The method of moments estimator is the solution of

$$\overline{X} = 1 + \theta$$

which gives method of moments estimator $\widehat{\theta} = \overline{X} - 1$.

1b Since $X_{(1)} - \theta$ has an exponential distribution with parameter n, we have

$$\begin{aligned}
\mathrm{MSE}(X_{(1)} + c) &= \mathrm{var}_\theta(X_{(1)} + c) - \big(\mathrm{E}_\theta(X_{(1)} + c) - \theta)\big)^2 \\
&= \mathrm{var}_\theta(X_{(1)} - \theta) - \big(\mathrm{E}_\theta(X_{(1)} - \theta) + c\big)^2 \\
&= \frac{1}{n^2} + \left(\frac{1}{n} + c\right)^2
\end{aligned}$$

It follows that the mean squared error is minimal for $c = -1/n$.

2a Because of independence of $X_1, \ldots, X_n$, we find

$$\begin{aligned}
\mathrm{P}_\theta\left(\max_{1\le i\le n} |X_i| \le t\right) &= \mathrm{P}_\theta\left(|X_1| \le t, \ldots, |X_n| \le t\right) \\
&= \mathrm{P}_\theta\left(|X_1| \le t\right) \cdots \mathrm{P}_\theta\left(|X_n| \le t\right).
\end{aligned}$$

Furthermore, for each $i = 1, \ldots, n$, it holds that $\mathrm{P}_\theta\left(|X_i| \le t\right) = 0$, for $t < 0$, and $\mathrm{P}_\theta\left(|X_i| \le t\right) = 1$, for $t > \theta$. Furthermore, for $0 < t\theta$ we find

$$\mathrm{P}_\theta\left(|X_i| \le t\right) = \mathrm{P}_\theta\left(-t \le X_i \le t\right) = \frac{2t}{2\theta} = \frac{t}{\theta}.$$

This means that

$$\mathrm{P}_\theta\left(\max_{1\le i\le n} |X_i| \le t\right) = \begin{cases} 0 & t < 0 \\ (t/\theta)^n & 0 < t < \theta \\ 1 & t > \theta \end{cases}$$

2b The size of the test is defined as

$$\sup_{\theta \le 1} \mathrm{P}_\theta\left(\max_{1\le i\le n} |X_i| > c_{\alpha_0}\right)$$

From part (a) we know that

$$\mathrm{P}_\theta\left(\max_{1\le i\le n} |X_i| > c_{\alpha_0}\right) = 1 - \mathrm{P}_\theta\left(\max_{1\le i\le n} |X_i| \le c_{\alpha_0}\right) = 1 - \left(\frac{c_{\alpha_0}}{\theta}\right)^n.$$

Because the expression on the right hand side is increasing in θ, it follows that the size of the test is equal to

$$\sup_{\theta \leq 1} \mathrm{P}_\theta \left(\max_{1 \leq i \leq n} |X_i| > c_{\alpha_0} \right) = \mathrm{P}_{\theta=1} \left(\max_{1 \leq i \leq n} |X_i| > c_{\alpha_0} \right) = 1 - c_{\alpha_0}^n.$$

The critical value must satisfy

$$1 - c_{\alpha_0}^n \leq \alpha_0 \quad \Leftrightarrow \quad c_{\alpha_0} \geq (1 - \alpha_0)^{1/n}$$

Because the critical region must be chosen as large as possible, we conclude that $c_{\alpha_0} = (1 - \alpha_0)^{1/n}$.

2c The definition of the power function is

$$\pi(\theta) = \mathrm{P}_\theta \left(\max_{1 \leq i \leq n} |X_i| > c_{\alpha_0} \right).$$

From part (a), we have

$$\mathrm{P}_\theta \left(\max_{1 \leq i \leq n} |X_i| > c_{\alpha_0} \right) = 1 - \mathrm{P}_\theta \left(\max_{1 \leq i \leq n} |X_i| \leq c_{\alpha_0} \right) = 1 - \left(\frac{c_{\alpha_0}}{\theta} \right)^n.$$

Together with part (b), this means that

$$\pi(\theta) = 1 - \frac{1 - \alpha_0}{\theta^n}.$$

3a We must show that the distribution of $2\theta \sum_{i=1}^n X_i^2$ does not depend on θ. It suffices to show that the distribution of each θX_i^2 does not depend on θ.
To this end, we determine the distribution function of θX_i^2. We find that

$$\mathrm{P}_\theta(\theta X_i^2 \leq t) = \mathrm{P}_\theta(X_i \leq \sqrt{t}/\sqrt{\theta}) = \int_0^{\sqrt{t}/\sqrt{\theta}} 2x\theta \mathrm{e}^{-\theta x^2} \, \mathrm{d}x = \int_0^t \mathrm{e}^{-y} \, \mathrm{d}y$$

where we use a change of variables $y = \theta x^2$.
The right hand side no longer depends on θ, so that the distribution of θX_i^2 does not depend on θ. This means that the distribution of $2\theta \sum_{i=1}^n X_i^2$ also does not depend on θ.

3b We first determine values c and d such that

$$\mathrm{P}_\theta \left(c < 2\theta \sum_{i=1}^n X_i^2 < d \right) = 0.95.$$

Because $2\theta \sum_{i=1}^n X_i^2$ is pivot, it has the same distribution as $2\theta \sum_{i=1}^n X_i^2$ with $\theta = 1$. It is given that this distribution is a chi-square distribution with $2n$ degrees of freedom. Therefore, we can take

$$c = \chi^2_{2n,0.025} \quad \text{and} \quad d = \chi^2_{2n,0.975}.$$

We find that

$$0.95 = \mathrm{P}_\theta\left(c < 2\theta\sum_{i=1}^n X_i^2 < d\right) = \mathrm{P}_\theta\left(\frac{c}{2\sum_{i=1}^n X_i^2} < \theta < \frac{d}{2\sum_{i=1}^n X_i^2}\right).$$

This means that

$$\left(\frac{c}{2\sum_{i=1}^n X_i^2}, \frac{d}{2\sum_{i=1}^n X_i^2}\right) = \left(\frac{\chi^2_{2n,0.025}}{2\sum_{i=1}^n X_i^2}, \frac{\chi^2_{2n,0.975}}{2\sum_{i=1}^n X_i^2}\right)$$

is a 95% confidence interval for θ.

4a [1 point]
The likelihood ratio statistic is defined as

$$\lambda_n = \frac{\sup_{\theta>0} p_\theta(X_1,\dots,X_n)}{\sup_{\theta=2} p_\theta(X_1,\dots,X_n)}$$

This leads to the following expression

$$\begin{aligned}\lambda_n &= \frac{\sup_{\theta>0} p_\theta(X_1,\dots,X_n)}{\sup_{\theta=2} p_\theta(X_1,\dots,X_n)}\\ &= \frac{p_{\widehat{\theta}}(X_1,\dots,X_n)}{p_{\theta=2}(X_1,\dots,X_n)}\\ &= \frac{\widehat{\theta}^n\left(\prod_{i=1}^n X_i\right)^{\widehat{\theta}-1}}{2^n\left(\prod_{i=1}^n X_i\right)}\\ &= \frac{\widehat{\theta}^n}{2^n}\left(\prod_{i=1}^n X_i\right)^{\widehat{\theta}-2}\end{aligned}$$

where $\widehat{\theta}$ is the maximum likelihood estimator.

NB: the second part of this question is not taken into account.
Note that

$$\prod_{i=1}^n X_i = \exp\left(\sum_{i=1}^n \log X_i\right) = \mathrm{e}^{-n/\widehat{\theta}}$$

This means that

$$\lambda_n = \frac{\widehat{\theta}^n}{2^n}\left(\prod_{i=1}^n X_i\right)^{\widehat{\theta}-2} = \frac{\widehat{\theta}^n}{2^n}\mathrm{e}^{-n(\widehat{\theta}-2)/\widehat{\theta}} = \frac{\widehat{\theta}^n}{2^n}\mathrm{e}^{-n(1-2/\widehat{\theta})}$$

This is a function of $\widehat{\theta}$, but it is **not** strictly increasing in $\widehat{\theta}$.

4b The Fisher information can be obtained by

$$i_\theta = -\mathrm{E}_\theta\ddot{\ell}_\theta(X_1).$$

We have that

$$\ell_\theta(x) = \log p_\theta(x) = \log\theta + (\theta - 1)\log x$$

so that

$$\dot{\ell}_\theta(x) = \frac{1}{\theta} + \log x \quad \text{and} \quad \ddot{\ell}_\theta(x) = -\frac{1}{\theta^2}.$$

We conclude that the Fisher information is equal to $i_\theta = 1/\theta^2$.
To compute an approximated 95% confidence interval for θ on the basis of the maximum likelihood estimator, there are two possibilities.

1. This method is based on the result

$$\sqrt{n\widehat{i_\theta}}(\widehat{\theta} - \theta) \approx N(0,1)$$

   where $\widehat{i_\theta}$ is an estimator for the Fisher information $i_\theta = 1/\theta^2$.

   Both plug-in and observed information lead to the traditional symmetric large sample interval

$$\theta = \widehat{\theta} \pm \frac{1}{\sqrt{n i_{\widehat{\theta}}}}\xi_{0.975} = \widehat{\theta} \pm \frac{1.96}{\sqrt{n}}\widehat{\theta}$$

   This gives the interval

$$\left[\frac{n}{\sum_{i=1}^n \log(1/X_i)}\left(1 - \frac{1.96}{\sqrt{n}}\right), \frac{n}{\sum_{i=1}^n \log(1/X_i)}\left(1 + \frac{1.96}{\sqrt{n}}\right)\right]$$

2. The other method is based on the result that

$$\sqrt{n i_\theta}(\widehat{\theta} - \theta) = \frac{\sqrt{n}}{\theta}(\widehat{\theta} - \theta) = \sqrt{n}\left(\frac{\widehat{\theta}}{\theta} - 1\right)$$

   is asymptotically standard normally distributed.

   This means

$$\mathrm{P}_\theta\left(-\xi_{1-\alpha/2} < \sqrt{n}\left(\frac{\widehat{\theta}}{\theta} - 1\right) < \xi_{1-\alpha/2}\right) \approx 1 - \alpha$$

   Or equivalently

$$\mathrm{P}_\theta\left(1 - \frac{\xi_{1-\alpha/2}}{\sqrt{n}} < \frac{\widehat{\theta}}{\theta} < 1 + \frac{\xi_{1-\alpha/2}}{\sqrt{n}}\right) \approx 1 - \alpha$$

   which leads to the interval

$$\left[\frac{\widehat{\theta}}{1 + \frac{\xi_{1-\alpha/2}}{\sqrt{n}}}, \frac{\widehat{\theta}}{1 - \frac{\xi_{1-\alpha/2}}{\sqrt{n}}}\right] = \left[\frac{\frac{n}{\sum_{i=1}^n \log(1/X_i)}}{1 + \frac{1.96}{\sqrt{n}}}, \frac{\frac{n}{\sum_{i=1}^n \log(1/X_i)}}{1 - \frac{1.96}{\sqrt{n}}}\right]$$

   which is not symmetric around $\widehat{\theta}$.

5a The least squares estimator is the minimizer of

$$\sum_{i=1}^{n}\left(Y_i-\theta-\frac{\theta}{2}x_i^2\right)^2$$

Putting the derivative with respect to θ equal to zero, gives

$$\begin{aligned}
&\sum_{i=1}^{n}\left(Y_i-\theta-\frac{\theta}{2}x_i^2\right)\left(1+\frac{x_i^2}{2}\right)=0\\
\Leftrightarrow\quad &\sum_{i=1}^{n}Y_i\left(1+\frac{x_i^2}{2}\right)=\theta\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2\\
\Leftrightarrow\quad &\theta=\frac{\sum_{i=1}^{n}Y_i\left(1+\frac{x_i^2}{2}\right)}{\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2}.
\end{aligned}$$

To check whether this is a minimum we compute the second derivative, which is equal to

$$\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2>0.$$

This means that the previous solution θ above is indeed a minimum, so that the least squared estimator is equal to

$$\widehat{\theta}=\frac{\sum_{i=1}^{n}Y_i\left(1+\frac{x_i^2}{2}\right)}{\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2}$$

5b We must show that $\mathrm{E}_\theta\widehat{\theta}=\theta$.
By the linearity property of expectations, we get

$$\mathrm{E}_\theta\widehat{\theta}=\frac{\sum_{i=1}^{n}(\mathrm{E}_\theta Y_i)\left(1+\frac{x_i^2}{2}\right)}{\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2}$$

and since

$$\mathrm{E}_\theta Y_i=\mathrm{E}_\theta\left(\theta+\frac{\theta}{2}x_i^2+e_i\right)=\theta+\frac{\theta}{2}x_i^2+\mathrm{E}_\theta e_i=\theta+\frac{\theta}{2}x_i^2=\theta\left(1+\frac{x_i^2}{2}\right).$$

It follows that

$$\mathrm{E}_\theta\widehat{\theta}=\frac{\sum_{i=1}^{n}\theta\left(1+\frac{x_i^2}{2}\right)\left(1+\frac{x_i^2}{2}\right)}{\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2}=\theta\frac{\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2}{\sum_{i=1}^{n}\left(1+\frac{x_i^2}{2}\right)^2}=\theta.$$

This show that $\widehat{\theta}$ is unbiased for θ.