Universiteit Leiden
Mathematisch Instituut
Dr. A. Dürre

20.12.2022

Exam

Introduction Mathematical Statistics

Semster I 2022-2023

Family name: First name:

Student number:

Remarks:

- The exam consists of **6** tasks.
- You have **180** minutes to complete the exam.
- You can only use the provided short version of our lecture notes.
- The solution must be documented well with calculations and if necessary references to theorems of our lecture. Giving just the solution is not sufficient.
- Whenever you need a certain quantile or probability try to express it using quantiles respectively the distribution function of standard distributions, like standard normal, t or F distribution.

Task	1	2	3	4	5	6	$\sum$
Points possible	10	14	6	5	5	9	49
Points achieved							

Grade exam: Grade homework: Final grade:

Task 1 (3 + 3 + 4) Let $X_1, \dots, X_n$ be independent and identically distributed with density $f_\lambda, \lambda > 0$ given by

$$f_\lambda(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} I_{\{x \geq 0\}}.$$

a) Compute a moment estimator for λ.
Solution:

$$\begin{aligned}
\mathbb{E}(X_1) &= \int_0^\infty x \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} dx \\
&= x \frac{1}{-\frac{1}{\lambda}} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} |_{x=0}^{\infty} + \int_0^\infty \frac{1}{-\frac{1}{\lambda}} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} dx && (1) \\
&= 0 + \lambda \underbrace{\int_0^\infty \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} \, dx}_{=1} = \lambda && (1)
\end{aligned}$$

So we can choose $\hat{\lambda} = \overline{X}$ (1)

b) Compute the Cramer-Rao bound (see Theorem 7.2). You can assume in the following that $\{\prod_{i=1}^n f_\lambda(x_i), \lambda > 0\}$ is regular for $n \in \mathbb{N}$.
Solution:
By Lemma 6.3 we have

$$i_\lambda = -\mathbb{E}(\ddot{l}_\lambda(X_1))$$

where

$$\begin{aligned}
\ddot{l}_\lambda(x) = \frac{\partial^2 \log(\frac{1}{\lambda} e^{-\frac{1}{\lambda}x})}{\partial^2 \lambda} &= \frac{\partial^2(-\log(\lambda) - \frac{1}{\lambda}x)}{\partial^2 \lambda} \\
&= \frac{\partial(-\frac{1}{\lambda} + \frac{1}{\lambda^2}x)}{\partial \lambda} \\
&= +\frac{1}{\lambda^2} - \frac{2}{\lambda^3}x && (1)
\end{aligned}$$

and thus

$$i_\lambda = -\mathbb{E}\left(\frac{1}{\lambda^2} - \frac{2}{\lambda^3}X_1\right) = \frac{2}{\lambda^3}\lambda - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}x \qquad (1)$$

By Theorem 7.2 $\text{Var}(\hat{\lambda}) \geq \frac{1}{n\frac{1}{\lambda^2}} = \frac{\lambda^2}{n}$ (1)
They do not need to use Lemma 6.3 and can calculate Var($\dot{l}(X_1)$) instead (using the hint in c))

c) Show that $\overline{X}$ is uniformly minimum variance unbiased (UMVU) estimator for λ. You can (but do not need to) use that $\mathbb{E}(X_1^2) = 2\lambda^2$.

Solution:

We can use Theorem 7.2 (but of course also 7.14+7.15) and have

$$\begin{aligned}
\mathrm{Var}(\overline{X}) &= \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} n \mathrm{Var}(X_1) & \textcircled{1} \\
&= \frac{1}{n}\left(\mathbb{E}(X_1^2) - [\mathbb{E}(X_1)]^2\right) = \frac{1}{n}(2\lambda^2 - \lambda^2) = \frac{\lambda^2}{n} & \textcircled{1}
\end{aligned}$$

which equals the Cramer-Rao bound. ①

Since the estimator is unbiased ($\mathbb{E}(\overline{X}) = \mathbb{E}(X_1) = \lambda$) ①

it is thus UMVU estimator.

Task 2 (7 + 5 + 2) Let $X_1, \ldots, X_n$ be independent and identically distributed with probability mass function (discrete density)

$$f_\theta(x) = \begin{cases} \frac{1}{2}(1-\theta) & |x| = 1 \\ \theta \quad \text{for} & x = 0 \\ 0 & x \notin \{-1, 0, 1\} \end{cases} \qquad \text{for } \theta \in (0,1).$$

a) Compute the maximum likelihood estimator of θ.

Solution:

Since we have independence:

$$\begin{aligned} L_{X_1,\ldots,X_n}(\theta) &= \prod_{i=1}^n f(X_i) \\ &= \prod_{i=1}^n (\frac{1}{2}(1-\theta))^{|X_i|} \cdot \theta^{1-|X_i|} \\ &= (\frac{1}{2}(1-\theta))^{\sum_{i=1}^n |X_i|} \theta^{\sum_{i=1}^n (1-|X_i|)} \end{aligned} \quad (1)$$

and therefore

$$\begin{aligned} \log(L_{X_1,\ldots,X_n}(\theta)) &= \sum_{i=1}^n |X_i| \log(\frac{1}{2}[1-\theta]) + (n - \sum_{i=1}^n |X_i|) \log(\theta) \\ &= \sum_{i=1}^n |X_i| (\log(\frac{1}{2}) + \log(1-\theta)) + (n - \sum_{i=1}^n |X_i|) \log(\theta). \end{aligned} \quad (1)$$

Deriving with respect to θ yields:

$$\begin{aligned} \frac{\partial \log(L_{X_1,\ldots,X_n}(\theta))}{\partial \theta} &= -\frac{\sum_{i=1}^n |X_i|}{1-\theta} + \frac{n - \sum_{i=1}^n |X_i|}{\theta} \overset{!}{=} 0 \\ &\Leftrightarrow \frac{-\theta \sum_{i=1}^n |X_i| + (1-\theta)(n - \sum_{i=1}^n |X_i|)}{\theta(1-\theta)} \overset{!}{=} 0 \end{aligned} \quad (1)$$

So we have

$$n - \sum_{i=1}^n |X_i| - \theta n = 0 \Leftrightarrow \theta = 1 - \frac{1}{n} \sum_{i=1}^n |X_i|. \quad (1)$$

The second derivative yields

$$\begin{aligned} \frac{\partial \log(L_{X_1,\ldots,X_n}(\theta))}{\partial \theta} &= -\frac{\sum_{i=1}^n |X_i|}{(1-\theta)^2} - \frac{n - \sum_{i=1}^n |X_i|}{\theta^2} \quad (1) \\ &< 0 \quad (1) \end{aligned}$$

(inequality since at least one of $\sum_{i=1}^n |X_i|$ and $n - \sum_{i=1}^n |X_i|$ has to be positive (note that $0 \leq |X_i| \leq 1$). So we have indeed found a maximum and the ML estimator equals $1 - \overline{|X|}$. (1)

b) We choose the Beta distribution with parameters $\alpha, \beta > 0$:

$$p_{\overline{\theta}}(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\gamma(\alpha)\gamma(\beta)}\gamma(\alpha+\beta)I_{\theta\in(0,1)}$$

as prior distribution. Compute the related Bayes estimator of θ.
Hint: We have $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$ for a Beta distributed random variable X with parameters $\alpha, \beta > 0$ (see 3.22).
Solution:
We have

$$\begin{aligned} p_{\overline{\theta}|X=x}(\theta) &= \frac{p_{X|\overline{\theta}=\theta}(x)p_{\overline{\theta}}(\theta)}{p_X(x)} && \text{(1)} \\ &= \frac{\left(\frac{1}{2}\right)^{\sum_{i=1}^n |x_i|}(1-\theta)^{\sum_{i=1}^n |x_i|}\theta^{\sum_{i=1}^n (1-|x_i|)}\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\gamma(\alpha)\gamma(\beta)}\gamma(\alpha+\beta)I_{\theta\in(0,1)}}{p_{X_1,\dots,X_n}(x_1,\dots,x_n)} \\ &= C(x_1,\dots,x_n)(1-\theta)^{\sum_{i=1}^n |x_i|+\beta-1}\theta^{n-\sum_{i=1}^n |x_i|+\alpha-1} && \text{(1)} \end{aligned}$$

thus $\overline{\theta}|X = x$ is Beta distributed with parameters $\tilde{\alpha} = n - \sum_{i=1}^n |x_i| + \alpha$ and $\tilde{\beta} = \sum_{i=1}^n |x_i| + \beta$ (1)
so that by 3.22 (or hint):

$$\mathbb{E}(\overline{\theta}|X = x) = \frac{n - \sum_{i=1}^n |x_i| + \alpha}{n - \sum_{i=1}^n |x_i| + \alpha + \sum_{i=1}^n |x_i| + \beta} = \frac{n - \sum_{i=1}^n |x_i| + \alpha}{n+\alpha+\beta} \quad \text{(1)}$$

and $\hat{\theta}_{\text{Bayes}} = \frac{n-\sum_{i=1}^n |x_i|+\alpha}{n+\alpha+\beta}$. (1)

c) How should one choose α and β to add as few prior information as possible?
Solution:
If we compare the ML estimator with the Bayes estimator we see that α and β are pseudo-observations (α corresponds to $|X_i| = 0$ and β to $|X_i| = 1$. (1)
Thus choosing α and β as small as possible adds as few information as possible . (1)

or

One can also argue that by choosing $\alpha = \beta = 1$ we have the uniform distribution on (0,1) as prior distribution. (1)
Thus the prior distribution treats every value of θ equally and thus does not seem to contain information about θ. (1)

Task 3 $(1+2+3)$ Let $X_1, \dots, X_n$ be independent and identically distributed with density $f_\theta = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\theta}}$ where $\mu \in \mathbb{R}$ and $\theta > 0$. We assume that μ is known.

a) Show that $\sum_{i=1}^n \frac{(X_i-\mu)^2}{\theta}$ is a pivot.
Solution:
Since $(X_i - \mu)/\sqrt{\theta_0} \sim N(0,1)$ is standard normal $\sum_{i=1}^n \frac{(X_i-\mu)^2}{\theta_0} \sim \chi_n^2$. Thus it does not depend on θ (or μ) (1)

b) Show that $\{g_\theta(x_1, \dots, x_n), \theta > 0\}$ where g_θ denotes the joint density of $X_1, \dots, X_n$ has a monotone likelihood ratio.
Solution:
We apply Lemma 7.32 and show that we have an exponential family (*of course you can also show the original definition*). Because of independence we have

$$\begin{aligned} g_\theta(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\theta}} \\ &= \underbrace{(2\pi)^{-\frac{n}{2}}\theta^{-\frac{n}{2}}}_{c(\theta)} exp(\underbrace{-\frac{1}{2\theta}}_{Q(\theta)} \underbrace{\sum_{i=1}^n (x_i-\mu)^2}_{V(x_1,\dots,x_n)}). \end{aligned} \quad (1)$$

Since $Q'(\theta) = \frac{1}{\theta^2} > 0$ $Q(\theta)$ is strictly monotone increasing (1)
$\{g_\theta(x_1, \dots, x_n), \theta > 0\}$ has a monotone likelihood ratio.

c) Construct the uniformly most powerful level α test for $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_0$.
Solution:
Because of b) we have a monotone likelihood ratio in $V(X_1, \dots, X_n) = \sum_{i=1}^n (X_i-\mu)^2$ thus using Karlin-Rubin Theorem (7.29) there exist c_α, γ_α such that

$$\psi(X_1, \dots, X_n) = \begin{cases} 1 & > \\ \gamma_\alpha \text{ if } \sum_{i=1}^n (X_i-\mu)^2 & = c_\alpha \\ 0 & < \end{cases} \quad (1)$$

such that $\psi(X_1, \dots, X_n)$ is optimal for $H_0 : \theta \le \theta_0$ vs $H_1 : \theta > \theta_0$. It remains to compute c_α and γ_α. Since the distribution of X_i and thus $\sum_{i=1}^n (X_i - \mu)^2$ is continuous $P(V(X_1, \dots, X_n) = c_\alpha) = 0$ and we can choose γ_α arbitrary (and thus 0). (1)
For c_α we have:

$$P_{\theta_0}(\sum_{i=1}^{n}(X_i - \mu^2) > c_\alpha) = \alpha \Leftrightarrow P(\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\theta_0} > \frac{c_\alpha}{\theta_0}) = \alpha.$$

Because of a) $\frac{c_\alpha}{\theta_0} = F^{-1}_{\chi^2_n}(0.95)$ so that $c_\alpha = F^{-1}_{\chi^2_n}(0.95)\theta_0$ where $F^{-1}_{\chi^2_n}(0.95)$ is the 0.95 quantile of the χ^2 distribution with n degrees of freedom.(**1**)

Task 4 Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be two samples. Show the following identity for the sample correlation coefficient

$$r_{XY} = \frac{S^2_{X+Y} - S^2_{X-Y}}{4S_X S_Y}$$

where S^2_{X+Y} and S^2_{X-Y} denote the sample variances based on $X_1+Y_1, \ldots, X_n+Y_n$ respectively $X_1 - Y_1, \ldots, X_n - Y_n$.

Solution:

We have

$$\begin{aligned}
S^2_{X+Y} &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i + Y_i - \frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i))^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n-1}([X_i - \overline{X}] + [Y_i - \overline{Y}])^2 && \text{①} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}([X_i - \overline{X}]^2 + 2[X_i - \overline{X}][Y_i - \overline{Y}] + [Y_i - \overline{Y}]^2) \\
&= S^2_X + S^2_Y + 2S_{XY} && \text{①}
\end{aligned}$$

and analogously

$$\begin{aligned}
S^2_{X+Y} &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - Y_i - \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i))^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n-1}([X_i - \overline{X}] - [Y_i - \overline{Y}])^2 && \text{①} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}([X_i - \overline{X}]^2 - 2[X_i - \overline{X}][Y_i - \overline{Y}] + [Y_i - \overline{Y}]^2) \\
&= S^2_X + S^2_Y - 2S_{XY} && \text{①}
\end{aligned}$$

so that

$$\frac{S^2_{X+Y} - S^2_{X-Y}}{4S_X S_Y} = \frac{S^2_X + S^2_Y + 2S_{XY} - (S^2_X + S^2_Y - 2S_{XY})}{4S_X S_Y} = \frac{4S_{X,Y}}{4S_X S_Y} = r_{XY} \quad \text{①}$$

Task 5 (3 + 2) The following R-function computes a test based on two data-vectors `x` and `y` of equal length.

```
f <- function(x,y,alpha) {
    A <- length(x)
    B <- cor(x,y)*sd(y)/sd(x)
    C <- mean(y)-B*mean(x)
    D <- y-x*B-C
    E <- sum(D^2)/(A-2)
    F <- matrix(c(rep(1,A),x),ncol=2)
    F <- t(F)%*%F
    G <- solve(F)[2,2]
    H <- B/sqrt(G*E)
    I <- qt(1-alpha,A-2)
    J <- as.numeric(H>I)
    return(J)
}
```

a) Which values are defined by B, C and I? **Solution:**

   - B: estimated slope of simple linear regression $\hat{\beta}$ ①
   - C: estimated intercept of simple linear regression $\hat{\alpha}$ ①
   - I: $1-\alpha$ quantile of a t-distribution with A-2 degrees of freedom ①

b) Which null-hypothesis is tested by the R-function?

- Which null-hypothesis is tested by the R-function?
  Solution:
  We test whether the slope of a simple linear regression is: ①

$$H_0 : \beta \leq 0 \text{ vs. } H_1 : \beta > 0. \quad ①$$

Task 6 (2 + 4 + 3) Consider the model

$$Y_i = \beta_0 + x_{i1}\beta_1 + \ldots + x_{ik}\beta_k + \epsilon_i \text{ for } i = 1, \ldots, n,$$

where $\epsilon_1, \ldots, \epsilon_n$ are iid with $\epsilon_1 \sim N(0, \sigma^2)$ for some $\sigma > 0$ and $\det(X'X) \neq 0$ where

$$X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}.$$

We want to predict the value of the dependent random variable Y for a combination of independent explanatory variables $z := (1, x_1, \ldots, x_k)$. A possible predictor is $\hat{Y} := z\hat{\beta}$ where $\hat{\beta}$ is the maximum likelihood estimator of $\beta = (\beta_0, \ldots, \beta_k)'$.

a) Derive the distribution of $\hat{Y}$. **Solution:**
By Theorem 8.17 we have $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$ ①
and by Theorem 8.7 with $A = z$ and $b = 0$ ①

$$z\hat{\beta} \sim N(z\beta, \sigma^2 z(X'X)^{-1}z')$$

b) Show that

$$\frac{\hat{Y} - z\beta}{\sqrt{z(X'X)^{-1}z'\hat{\sigma}^2 \frac{n}{n-k-1}}} \sim t_{n-k-1}$$

where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - (1, x_{i1}, \ldots, x_{ik})\hat{\beta})^2$.
Solution:
By Theorem 8.20 we know that $\hat{\beta}$ and $\hat{\sigma}^2$ are independent. ①
By a) we know that $\frac{z\hat{\beta} - z\beta}{\sqrt{\sigma^2 z(X'X)^{-1}z'}} \sim N(0, 1)$ ①
and by Theorem 8.19 $\frac{\hat{\sigma}^2}{\sigma^2}\frac{n}{n-k-1} \sim \chi^2_{n-k-1}$. ①
It follows that

$$\frac{\hat{Y} - z\beta}{\sqrt{z(X'X)z'\hat{\sigma}^2 \frac{n}{n-k-1}}} = \frac{\frac{\hat{Y} - z\beta}{\sqrt{z(X'X)^{-1}z'\sigma^2}}}{\sqrt{\frac{\hat{\sigma}^2 \frac{n}{n-k-1}}{\sigma^2}}} \quad ①$$

is t distributed with $n - k - 1$ degrees of freedom.

c) Derive a one-sided $(1 - \alpha)$ confidence interval of the form (d, ∞) for $z\beta$. Try to find a small interval (you will get no points for the trivial interval $(-\infty, \infty)$).
Solution:
We have

$$
\begin{aligned}
1-\alpha &= P\Big(\frac{\hat{Y}-z\beta}{\sqrt{z(X'X)^{-1}z'\hat{\sigma}^2\frac{n}{n-k-1}}} \le F^{-1}_{t_{n-k-1}}(1-\alpha)\Big) \quad \textcircled{1}\\
&= P\Big(\hat{Y}-z\beta \le F^{-1}_{t_{n-k-1}}(1-\alpha)\sqrt{z(X'X)^{-1}z'\hat{\sigma}^2\frac{n}{n-k-1}}\Big)\\
&= P\Big(\hat{Y}-F^{-1}_{t_{n-k-1}}(1-\alpha)\sqrt{z(X'X)^{-1}z'\hat{\sigma}^2\frac{n}{n-k-1}} \le z\beta\Big) \quad \textcircled{1}
\end{aligned}
$$

and thus

$$
\left(\hat{Y}-F^{-1}_{t_{n-k-1}}(1-\alpha)\sqrt{z(X'X)^{-1}z'\hat{\sigma}^2\frac{n}{n-k-1}}, \infty\right) \quad \textcircled{1}
$$

is a $1-\alpha$ confidence interval for $z\beta$.