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 **5** 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 | $\sum$ |
|---|---|---|---|---|---|---|
| Points possible | 22 | 5 | 7 | 5 | 8 | 47 |
| Points achieved | | | | | | |

Grade exam: Grade homework: Final grade:

**Task 1** (7 + 6 + 2 + 5 + 2) Let $X_1, \dots, X_n$ be independent and identically Poisson distributed with probability mass function (discrete density) $f_\theta, \theta \geq 0$ given by

$$f_\theta(x) = \frac{\theta^x}{x!} e^{-\theta} I_{\{x \in \mathbb{N}_0\}}.$$

a) Compute the maximum likelihood estimator of $\theta$.

*Hint: Consider first the case where not all random variables attain the value 0:* $\{\omega \in \Omega : \sum_{i=1}^n X_i(\omega) > 0\}$. *Treat then the special case:* $\{\omega \in \Omega : \sum_{i=1}^n X_i(\omega) = 0\}$.

**Solution:**

Since $X_1, \dots, X_n$ are independent we have:

$$L_{X_1,\dots,X_n}(\theta) = \prod_{i=1}^n \frac{\theta^{X_i}}{X_i!} e^{-\theta} I_{\{X_i \in \mathbb{N}_0\}} \quad ①$$

$$= e^{-n\theta} \frac{\theta^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!} I_{\{X_1,\dots,X_n \in \mathbb{N}\}} \quad ①$$

and thus

$$\log(L_{X_1,\dots,X_n}(\theta)) = -n\theta + \sum_{i=1}^n X_i \log(\theta) - \sum_{i=1}^n \log(X_i!) + \log(I_{\{X_1,\dots,X_n \in \mathbb{N}\}}). \quad ①$$

For maximization we calculate the first derivative

$$\frac{\partial L_{X_1,\dots,X_n}(\theta)}{\partial \theta} = -n + \frac{\sum_{i=1}^n X_i}{\theta} \stackrel{!}{=} 0$$

$$\Leftrightarrow \theta = \frac{1}{n} \sum_{i=1}^n X_i = \overline{X}. \quad ①$$

The second derivative equals:

$$\frac{\partial^2 L_{X_1,\dots,X_n}(\theta)}{\partial^2 \theta} = -\frac{\sum_{i=1}^n X_i}{\theta^2}$$

which is negative if $\sum_{i=1}^n X_i > 0$ and thus a maximum. ①

If $\sum_{i=1}^n X_i = 0$ we have

$$L_{0,\dots,0}(\theta) = \prod_{i=1}^n \frac{\theta^0}{0!} e^{-\theta} = e^{-\theta n} \quad ①$$

which is maximal for $\theta = 0$ since $e^{-\theta n}$ is strictly monotonic decreasing. ①

Thus

$$\hat{\theta}_{ML} = \begin{cases} \overline{X} & \text{if } \sum_{i=1}^n X_i > 0 \\ 0 & \text{if } \sum_{i=1}^n X_i = 0 \end{cases} = \overline{X}.$$

b) Show that $\frac{1}{n}\sum_{i=1}^n X_i$ is uniformly minimum variance unbiased estimator for $\theta$.

*Hint: You can use here and in the following tasks that $\{\prod_{i=1}^n f_\theta(x_i), \theta \geq 0\}$ is regular and that for a Poisson distributed random variable $X$ : $\mathbb{E}(X) = \theta$ and $\mathbb{E}(X^2) = \theta + \theta^2$.*

**Solution:**

We have

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{\theta}{n}. \qquad ①$$

For the Cramer-Rao bound we compute $i_\theta$ :

$$\begin{aligned}
i_\theta &= \mathrm{Var}\left(\frac{\partial \log\left(\frac{\theta^X}{X!}e^{-\theta} I_{\{X\in\mathbb{N}_0\}}\right)}{\partial\theta}\right) \qquad ①\\
&= \mathrm{Var}\left\{\frac{\partial\left[X\log(\theta) - \log(X!) - \theta + \log(I_{\{X\in\mathbb{N}_0\}})\right]}{\partial\theta}\right\}\\
&= \mathrm{Var}\left(\frac{X}{\theta} - 1\right)\\
&= \frac{1}{\theta^2}\mathrm{Var}(X)\\
&= \frac{1}{\theta^2}(\mathbb{E}(X^2) - \mathbb{E}(X)^2)\\
&= \frac{1}{\theta^2}(\theta^2 + \theta - \theta^2) = \frac{1}{\theta} \qquad ①
\end{aligned}$$

and thus the Cramer Rao bound equals

$$\frac{1}{ni_\theta} = \frac{\theta}{n} \qquad ①$$

so that $\frac{1}{n}\sum -i = 1^n X_i$ reaches the Cramer-Rao bound ①

and is unbiased because of $\mathbb{E}(\overline{X} = \mathbb{E}(X_1) = \theta$. ①

Thus $\hat{\theta}$ is UMVU estimator for $\theta$.

c) Construct an asymptotic two sided confidence interval of confidence level $1-\alpha$ for $\theta$. You can assume that all conditions for Theorem 6.6 are fulfilled. You will not get points for the trivial interval $(-\infty, \infty)$.

**Solution:**

By Remark 6.7 we have

$$\left[\overline{X} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\hat{i_\theta}^{-\frac{1}{2}}}{\sqrt{n}}, \overline{X} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\hat{i_\theta}^{-\frac{1}{2}}}{\sqrt{n}}\right] \qquad ①$$

where we can choose $\hat{i_\theta}^{-\frac{1}{2}} = i_{\hat{\theta}}^{-\frac{1}{2}} = \left(\frac{1}{\overline{X}}\right)^{-\frac{1}{2}} = \sqrt{\overline{X}}$ ①

where the second last equation follows from the calculations in b).

d) We use the gamma distribution with parameters $\alpha, \beta > 0$ as prior:

$$f_{\overline{\theta}}(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \cdot I_{\{\theta \geq 0\}}.$$

Compute the Bayes estimator of $\theta$.

*Hint: Let $Y$ be a gamma distributed random variable with parameters $\alpha, \beta > 0$, then $\mathbb{E}(Y) = \frac{\alpha}{\beta}$.*

**Solution:**

By 3.20 we have

$$\hat{\theta}_B = \int \theta p_{\overline{\theta}|X=x}(\theta)d\theta \qquad (1)$$

where by Remark 3.19 and calculations in a)

$$\begin{aligned} p_{\overline{\theta}|X=x}(\theta) &= \frac{p_{X|\overline{\theta}=\theta}(x)p_{\overline{\theta}}(\theta)}{p_X(x)} \\ &= \frac{e^{-n\theta}\frac{\theta^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}I_{\{X_1,\dots,X_n\in\mathbb{N}_0\}}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}I_{\theta\geq 0}}{p_{X_1,\dots,X_n}(x_1,\dots,x_n)} \\ &= C(x_1,\dots,x_n)\theta^{\alpha-1+\sum_{i=1}^n X_i}e^{-(n+\beta)n}. \qquad (1) \end{aligned}$$

Thus $\overline{\theta}|X = x$ is gamma distributed (1)

with parameters $\tilde{\alpha} = \alpha + \sum_{i=1}^n X_i$ and $\tilde{\beta} = n + \beta$. (1)

Applying the hint then yields

$$\mathbb{E}(\overline{\theta}|X = x) = \frac{\alpha + \sum_{i=1}^n X_i}{n+\beta} = \hat{\theta}_B \qquad (1)$$

e) How should one choose $\alpha$ and $\beta$ to add as few prior information as possible?

**Solution:** If we compare the Bayes estimator with the ML estimator, we see that $\alpha$ and $\beta$ represent pseudo-observations. More in detail $\alpha$ is the sum of $\beta$ pseudo-observations. (1)

Choosing $\alpha, \beta$ as close to 0 as possible adds therefore as few as possible prior information. (1)

Alternatively one can see that the density of a gamma distributed random variable gets flat (does not depend on $\theta$) if $\beta \to 0$ and $\alpha \to 1$. $\beta = 0$ and $\alpha = 1$ as $\beta = \alpha = 0$ yield to so called improper prior distribution.

**Task 2** Let $X_1, \dots, X_n$ be independent and identically distributed with density

$$f_\theta(x) := \begin{cases} \theta - \theta^2 \cdot |x| & |x| \leq \frac{1}{\theta} \\ 0 & |x| > \frac{1}{\theta}. \end{cases}$$

Compute a moment estimator for $\theta$.

**Solution:** Since the density is symmetric $\mathbb{E}(X) = 0$ does not give us information about $\theta$, we have to look at the second moment. ①

By symmetry we have

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\frac{1}{\theta}}^{\frac{1}{\theta}} x^2(\theta - \theta^2|x|)dx & ① \\
&= 2\int_0^{\frac{1}{\theta}} (x^2\theta - \theta^2 x^3)dx \\
&= 2\left[\frac{1}{3}x^3\theta - \theta^2\frac{1}{4}x^4\right]_{x=0}^{\frac{1}{\theta}} \\
&= \frac{2}{3\theta^3}\theta - \frac{1}{2\theta^4}\theta^2 = \frac{1}{6\theta^2} & ①
\end{aligned}$$

which we solve for $\theta$ :

$$\mathbb{E}(X^2)\theta^2 = \frac{1}{6} \Leftrightarrow \theta^2 = \frac{1}{6\mathbb{E}(X^2)} \Leftrightarrow \theta = \left(\frac{1}{6\mathbb{E}(X^2)}\right)^{\frac{1}{2}}. \quad ①$$

It follows that $\hat{\theta} = \left(\frac{1}{6\frac{1}{n}\sum_{i=1}^n X_i^2}\right)^{\frac{1}{2}}$ is a moment estimator for $\theta$ ①
(there are a lot more possible).

**Task 3** (3 + 4) Let $X_1, \ldots, X_n$ be independent and identically distributed with density

$$f_\theta(x) = \theta e^{-\theta x} I_{\{x \geq 0\}}$$

for $\theta > 0$

a) Show that $\{g_\theta(x_1, \ldots, x_n), \theta > 0\}$ where $g_\theta$ denotes the joint density of $x_1, \ldots, x_n$ has a monotone likelihood ratio in $V(X_1, \ldots, X_n) = -\sum_{i=1}^n X_i$.

**Solution:**

We show that $\{g_\theta(x_1, \ldots, x_n), \theta > 0\}$ is an exponential family and then apply Lemma 7.32. Because of independence the joint density equals

$$\begin{aligned}
g_\theta(x_1, \ldots, x_n) &= \prod_{i=1}^n \left[\theta e^{-\theta x_i} I_{\{x_i \geq 0\}}\right] & ① \\
&= \underbrace{\theta^n}_{c(\theta)} \exp\left(\underbrace{\theta}_{Q(\theta)} \underbrace{\left[-\sum_{i=1}^n x_i\right]}_{V(X_1,\ldots,X_n)}\right) \underbrace{I_{\{x_1,\ldots,x_n \geq 0\}}}_{h(x_1,\ldots,x_n)} & ①
\end{aligned}$$

and since $Q(\theta) = \theta$ is strictly increasing $\{g_\theta(x_1, \ldots, x_n), \theta > 0\}$ has a monotone likelihood ration in $-\sum_{i=1}^n X_i$. ①

b) Construct the uniformly most powerful level $\alpha$ test for $H_0: \ \theta \leq \theta_0$ vs. $H_1: \ \theta > \theta_0$.

*Hint: For $X_1, \ldots, X_n$ independent and identically exponentially distribute with paramater $\lambda$* is $\sum_{i=1}^n X_i$ *Erlang distributed with parameters $n, \lambda$ and density* $h_{n,\lambda}(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} I_{\{x \geq 0\}}$.

**Solution:**

By the Karlin-Rubin (Theorem 7.29) and a) ①

the most powerful level $\alpha$ test for $H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$ is of the form

$$\psi_\alpha(X_1, \ldots, X_n) = \begin{cases} 1 & > \\ \gamma_\alpha \text{ if } & -\sum_{i=1}^n X_i = c_\alpha. \\ 0 & < \end{cases}$$ ①

or equivalently

$$\psi_\alpha(X_1, \ldots, X_n) = \begin{cases} 1 & < \\ \gamma_\alpha \text{ if } & \sum_{i=1}^n X_i = \tilde{c}_\alpha. \\ 0 & > \end{cases}$$

. By the hint $\sum_{i=1}^n X_i$ is Erlang distributed with parameters $n, \theta$. Since the distribution is continuous,we can choose $\gamma_\alpha = 0$. ①

Thus it remains to calculate $c_\alpha$. We need $P_{\theta_0}(\sum_{i=1}^n X_i < \tilde{c}_\alpha) = \alpha$ thus $\tilde{c}_\alpha = F_{n,\theta_0}^{-1}(\alpha)$ where $F_{n,\theta_0}^{-1}(\alpha)$ denotes the $\alpha$ quantile of the Erlang distribution with parameters $n, \theta_0$. ①

**Task 4** (3 + 2) The following R-function creates a plot based on a $n$-dimensional data vector `y` and a $n \times k$ matrix `x`.

```
f <- function(x,y) {
    A <- length(y)
    B <- length(x[1,])
    C <- cbind(1,x)
    D <- solve(t(C)%*%C)%*%t(C)%*%y
    E <- y-C%*%D
    F <- numeric(A)
    for (i in 1:A) {
     F[i] <- t(C[i,])%*%solve(t(C)%*%C)%*%C[i,]
     }
    plot(F,E)
}
```

a) Which values are defined by $D$, $E$ and $F$?
**Solution:**

- D: vector of estimated regression coefficients $\hat{\beta}$ (1)
- E: vector of estimated residuals $\hat{\epsilon}$ (1)
- F: vector of leverages $h$ (1)

b) What can one deduce from such a plot?
The scatterplot of residuals and leverages shows if there are influential observation within the linear regression. (1)
Observations with large leverage and large absolute (estimated) residual (this observations would be in the top right or top bottom corner of the plot) influence the estimated regression coefficients strongly. (1)

**Task 5** (4 + 4) 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

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

a) Show that $\sum_{i=1}^{n} \hat{\epsilon}_i = 0$ where $\hat{\epsilon}_i$ for $i = 1, \ldots, n$ denote the estimated residuals based on a maximum likelihood estimator of $\beta$ (situation of Theorem 8.16).
*Hint: Remember that $\hat{\beta}$ is minimizer of $f(\beta) := \sum_{i=1}^{n}(Y_i - X\beta)^2$.*
**Solution:**
This follows from the so called normal equations which are necessary to derive the maximum likelihood estimator of $\theta$. The likelihood equals

$$\begin{aligned} L_{Y_1,\ldots,Y_n}(\beta, \sigma^2) &= \prod_{i=1}^{n} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(Y_i - \beta_0 - X_{i1}\beta_1 - \ldots - X_{ik}\beta_k)^2} \right) \\ &= (2\pi)^{-\frac{n}{2}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} [Y_i - \beta_0 - X_{i1}\beta_1 - \ldots - X_{ik}\beta_k]^2). \end{aligned}$$

Since we want to maximize the likelihood we have to minimize $\sum_{i=1}^{n}[Y_i - \beta_0 - X_{i1}\beta_1 - \ldots - X_{ik}\beta_k]^2$ with respect to $\beta$. That is what the hint says. So these

steps are not necessary. The derivative with respect to $\beta_0$ ①
equals

$$\frac{\partial \sum_{i=1}^n [Y_i - \beta_0 - X_{i1}\beta_1 - \ldots - X_{ik}\beta_k]^2}{\partial \beta_0} = \sum_{i=1}^n [Y_i - \beta_0 - X_{i1}\beta_1 - \ldots - X_{ik}\beta_k](-1) \stackrel{!}{=} 0 \quad ①$$

Using the hint we know that $\hat{\beta}$ is minimizer and thus solves the above equation, ①
so that:

$$0 = \sum_{i=1}^n [Y_i - \hat{\beta}_0 - X_{i1}\hat{\beta}_1 - \ldots - X_{ik}\hat{\beta}_k](-1) = \sum_{i=1}^n \hat{\epsilon}_i. \quad ①$$

b) Denote by $F_m^{-1}(\alpha)$ the $\alpha$ quantile of a chi-square distributed random variable with $m$ degrees of freedom and by $\hat{\sigma}^2$ the maximum likelihood estimator of $\sigma^2$ as defined in Theorem 8.16. Show that

$$\left(\sqrt{\frac{n\hat{\sigma}^2}{F_{n-k-1}^{-1}(1-\alpha/2)}}, \sqrt{\frac{n\hat{\sigma}^2}{F_{n-k-1}^{-1}(\alpha/2)}}\right)$$

a $1-\alpha$ confidence interval for $\sigma$ is.

**Solution:**

By 8.19 we have that $n\frac{\hat{\sigma}^2}{\sigma^2}$ is chisquare distributed with $n-k-1$ degrees of freedom ①
and thus:

$$\begin{aligned}
1-\alpha &= P\left(F_{n-k-1}^{-1}(\alpha/2) < n\frac{\hat{\sigma}^2}{\sigma^2} < F_{n-k-1}^{-1}(1-\alpha/2)\right) && ① \\
&= P\left(\frac{F_{n-k-1}^{-1}(\alpha/2)}{n\hat{\sigma}^2} < \frac{1}{\sigma^2} < \frac{F_{n-k-1}^{-1}(1-\alpha/2)}{n\hat{\sigma}^2}\right) \\
&= P\left(\sqrt{\frac{F_{n-k-1}^{-1}(\alpha/2)}{n\hat{\sigma}^2}} < \frac{1}{\sigma} < \sqrt{\frac{F_{n-k-1}^{-1}(1-\alpha/2)}{n\hat{\sigma}^2}}\right) && ① \\
&= P\left(\frac{1}{\sqrt{\frac{F_{n-k-1}^{-1}(\alpha/2)}{n\hat{\sigma}^2}}} > \sigma > \frac{1}{\sqrt{\frac{F_{n-k-1}^{-1}(1-\alpha/2)}{n\hat{\sigma}^2}}}\right) \\
&= P\left(\sqrt{\frac{n\hat{\sigma}^2}{F_{n-k-1}^{-1}(\alpha/2)}} > \sigma > \sqrt{\frac{n\hat{\sigma}^2}{F_{n-k-1}^{-1}(1-\alpha/2)}}\right) && ①
\end{aligned}$$