Re-Exam - Statistics (WBMA009-05) 2025/2026

Date and time: January 30, 2026, 15.00-17.00h
Place: Exam Hall 1, Blauwborgje 4

Rules to follow:

- This is a closed book exam. Consultation of books and notes is **not** permitted. You may use a simple (non-programmable) calculator.
- Write your name and student number on each sheet of paper.
- There are four exercises worth a total of 36 points.
- Your exam grade will be computed as follows:

$$\text{grade} = (\text{number of points} + 4) \cdot 0.25$$

  rounded to the nearest 0.5, with the usual exception that 5.5 is not possible.
- ALWAYS include the relevant equation(s) and/or short derivations.
- **We wish you success in completing the exam!**

START OF RE-EXAM
ON NEXT PAGE

α	0.025	0.05	0.1	0.5	0.9	0.95	0.975
$\mathcal{N}(0,1)$	-1.96	-1.64	-1.28	0	1.28	1.64	1.96

Table 1: Quantiles q_α of the $\mathcal{N}(0,1)$ distribution.

1. **Geometric distribution.** **10**
Consider a random sample of size n from the geometric distribution

$$X_1, \ldots, X_n \sim \text{Geo}(\theta)$$

with parameter $0 < \theta < 1$ and density:

$$f_\theta(x) = (1-\theta)^{x-1} \cdot \theta \qquad (x = 1, 2, 3, \ldots)$$

It can be shown that this implies the expectation $E[X_1] = \frac{1}{\theta}$.

(a) Compute the ML estimator of θ. 2
HINT: Check via the 2nd derivative whether you have a maximum.

(b) Compute the Fisher information $I(\theta)$ for a sample of size $n = 1$. 2

For the following two exercise parts (c-d) consider the test problem

$$H_0 : \theta = 0.6 \ \textbf{vs.} \ H_1 : \theta = 0.4$$

(c) Show that a statistical test that rejects H_0 if $\sum_{i=1}^{n} X_i > k_0$, where $k_0 > 0$ is a constant, is the Neyman-Pearson UMP test for the test problem. 2

(d) Consider the Neyman-Pearson UMP test from (c) with $n = 1$ and $k_0 = 3$. Compute the test level α. 2

For the last exercise part (e) assume $n = 5$ and the five realisations:
$X_1 = 1$, $X_2 = 3$, $X_3 = 1$, $X_4 = 3$, and $X_5 = 2$.

(e) Make use of the asymptotic efficiency (normality) of the ML estimator to construct an asymptotic two-sided 80% confidence interval for θ. 2
HINT: See Table 1 for a list of quantiles.

2. **Rao–Blackwell theorem.** **6**
Let $X_1, \ldots, X_n$ be a random sample from the uniform distribution on $(0, \theta)$ with $\theta > 0$ and density

$$f(x|\theta) = \frac{1}{\theta} \mathbf{1}_{(0,\theta)}(x), \qquad (x \in \mathbb{R}).$$

Define $M = \max\{X_1, \ldots, X_n\}$ and consider the estimator $T_0 = 2X_1$.

(a) Show that T_0 is an unbiased estimator of θ. 1

(b) Show that M is a sufficient statistic for θ. 2

(c) Compute the Rao–Blackwell estimator $W = \mathbb{E}[T_0|M]$ and simplify it. 3

3. **Lindley distribution.** **10**

Consider a random sample of size n from the Lindley distribution with density

$$f(x|\theta) = \frac{\theta^2}{1+\theta}(1+x)e^{-\theta x} \qquad (x > 0)$$

where $\theta > 0$, and denote $S := \sum_{i=1}^{n} X_i$.

(a) Show that S is a sufficient statistic for θ. 2

(b) Derive the maximum likelihood estimator of θ. 2
HINT: Check via the 2nd derivative whether you have a maximum.

(c) Consider the test problem

$$H_0 : \theta \leq \theta_0 \ \textbf{vs.} \ H_1 : \theta > \theta_0.$$

Show that the uniformly most powerful (UMP) test rejects H_0 if $S \leq c$. You do not need to compute the constant c. 2

(d) Compute the Fisher information $I(\theta)$ for a sample of size $n = 1$. 2

For the last exercise part (e) assume $n = 10$ and $\widehat{\theta}_{ML} = 2$.

(e) Make use of the asymptotic efficiency (normality) of the ML estimator to construct an asymptotic one-sided 95% confidence interval of the form $(-\infty, U]$ for θ. 2
HINT: See Table 1 for a list of quantiles.

4. **P-values and power.** **10**

Consider a random sample of size n from the exponential distribution with density

$$f(x|\theta) = \theta e^{-\theta x} \qquad (x > 0)$$

where $\theta > 0$, and the test problem

$$H_0 : \theta = \theta_0 \ \textbf{vs.} \ H_1 : \theta \neq \theta_0.$$

Define $S := \sum_{i=1}^{n} X_i$ and perform a likelihood ratio test (LRT).

(a) Derive the maximum likelihood (ML) estimator of θ. 2

(b) Derive the LRT statistic, and write it as a function of S. 2

(c) Show that LRT statistic is maximised at $S = n/\theta_0$ and decreases as S moves away from n/θ_0, so that the LRT rejects H_0 for

$$S \leq c_1 \quad \text{or} \quad S \geq c_2$$

for suitable constants $c_1 < c_2$. 4

(d) For $\theta \neq \theta_0$, write down in one line the power function of the LRT, in terms of S and the critical values c_1, c_2. 2

HINT: Recall the LRT statistic: $\dfrac{\sup_{\theta \in \Theta_0} L(\theta|x_1, \ldots, x_n)}{\sup_{\theta \in \Theta} L(\theta|x_1, \ldots, x_n)}$.

EXERCISE 1 - SOLUTIONS

(a) We have the likelihood $L(\theta) = \prod_{i=1}^{n} f_\theta(X_i) = (1-\theta)^{\sum_{i=1}^{n} X_i - n} \cdot \theta^n$, and the log likelihood:

$$l(\theta) = (\sum_{i=1}^{n} X_i - n) \cdot \log(1-\theta) + n \cdot \log(\theta)$$

Taking the 1st and the 2nd derivative of $l(\theta)$ yields

$$l'(\theta) = -\frac{\sum_{i=1}^{n} X_i - n}{1-\theta} + \frac{n}{\theta} \quad \text{and} \quad l''(\theta) = -\frac{\sum_{i=1}^{n} X_i - n}{(1-\theta)^2} - \frac{n}{\theta^2}$$

We have: $l'(\theta) = 0 \Leftrightarrow \ldots \Leftrightarrow \theta = \frac{n}{\sum_{i=1}^{n} X_i}$, and $l''(\theta) < 0$ for all $\theta \in (0,1)$. It follows

$$\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^{n} X_i}$$

(b) Use $l''(\theta)$ with $n = 1$ and the relation:

$$I(\theta) = -E[l''(\theta)] = E\left[\frac{X_1 - 1}{(1-\theta)^2} + \frac{1}{\theta^2}\right] = \frac{\frac{1}{\theta} - 1}{(1-\theta)^2} + \frac{1}{\theta^2} = \frac{1}{\theta^2(1-\theta)}$$

(c) Build the joint density (or likelihood) ratio:

$$\lambda(X) := \frac{L_X(0.6)}{L_X(0.4)} = \frac{(1-0.6)^{\sum_{i=1}^{n} X_i - n} \cdot 0.6^n}{(1-0.4)^{\sum_{i=1}^{n} X_i - n} \cdot 0.4^n} = \left(\frac{2}{3}\right)^{\sum_{i=1}^{n} X_i - n} \cdot \left(\frac{3}{2}\right)^n$$

The NP lemma states that the UMP test rejects H_0 if $\lambda(X) < k$.
As $\lambda(X)$ is monotone decreasing in $Y := \sum_{i=1}^{n} X_i$, we have $\lambda(X) < k \quad \Leftrightarrow \quad \sum_{i=1}^{n} X_i > k_0$.

(d) For $n = 1$ and $k_0 = 3$ we reject H_0 if $X_1 > 3$. Thus the UMP test is to the level:

$$\alpha = P(X_1 > 3) = 1 - P(X_1 \leq 3) = 1 - (0.6 + 0.4 \cdot 0.6 + 0.4^2 \cdot 0.6) = 0.064$$

(e) For large n we have: $\sqrt{n} \cdot (\hat{\theta}_{ML} - \theta) \sim \mathcal{N}(0, I(\theta)^{-1}) \Leftrightarrow \sqrt{I(\theta)}\sqrt{n} \cdot (\hat{\theta}_{ML} - \theta) \sim \mathcal{N}(0,1)$.

It follows: $P(-q_{0.9} \leq \sqrt{I(\theta)}\sqrt{n} \cdot (\hat{\theta}_{ML} - \theta) \leq q_{0.9}) = 0.8$.

$$\Leftrightarrow P(\hat{\theta}_{ML} - q_{0.9}\frac{1}{\sqrt{I(\theta)}\sqrt{n}} \leq \theta \leq \hat{\theta}_{ML} + q_{0.9}\frac{1}{\sqrt{I(\theta)}\sqrt{n}}) = 0.8$$

Here we have: $\hat{\theta}_{ML} = \frac{5}{10} = 0.5$ and the observed Fisher information

$$I(\hat{\theta}_{ML}) = \frac{1}{\hat{\theta}^2_{ML}(1-\hat{\theta}_{ML})} = \frac{1}{0.125} = 8$$

It follows that an approximate two-sided 80% CI for θ is:

$$\left[0.5 - 1.28 \cdot \frac{1}{\sqrt{40}}, 0.5 + 1.28 \cdot \frac{1}{\sqrt{40}}\right] = [0.298, 0.702]$$

EXERCISE 2 - SOLUTIONS

(a) For $X_1 \sim \text{Unif}(0,\theta)$ we have $\mathbb{E}[X_1] = \frac{\theta}{2}$. Hence $\mathbb{E}[T_0] = \mathbb{E}_\theta[2X_1] = 2\cdot\frac{\theta}{2} = \theta$. Therefore, T_0 is an unbiased estimator of θ.

(b) The likelihood is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\mathbf{1}_{(0,\theta)}(X_i) = \theta^{-n}\,\mathbf{1}_{(0,\theta)}(M).$$

This factorises as $L(\theta) = g(\theta, M)\,h(x_1,\ldots,x_n)$ with $g(\theta,M) = \theta^{-n}\mathbf{1}_{(0,\theta)}(M)$ and $h \equiv 1$. By the factorisation theorem, M is a sufficient statistic for θ.

(c) The Rao–Blackwell estimator is

$$W = \mathbb{E}(T_0|M) = 2\,\mathbb{E}(X_1|M).$$

By symmetry: $\mathbb{E}(X_1|M) = \cdots = \mathbb{E}(X_n|M)$.
Taking conditional expectations on both sides of $\sum_{i=1}^n X_i$ yields

$$\mathbb{E}\Big(\sum_{i=1}^{n} X_i \,\Big|\, M\Big) = \sum_{i=1}^{n} \mathbb{E}(X_i|M) = n\,\mathbb{E}(X_1|M).$$

Now condition on $M = m$. With probability 1 exactly one observation equals m and the remaining $n-1$ observations are iid $\text{Unif}(0,m)$. Therefore

$$\mathbb{E}\Big(\sum_{i=1}^{n} X_i \,\Big|\, M = m\Big) = m + (n-1)\cdot\frac{m}{2} = \frac{n+1}{2}\,m.$$

Hence

$$n\,\mathbb{E}(X_1|M=m) = \frac{n+1}{2}\,m \quad\Rightarrow\quad \mathbb{E}(X_1|M=m) = \frac{n+1}{2n}\,m.$$

Thus

$$W = 2\,\mathbb{E}(X_1|M) = 2\cdot\frac{n+1}{2n}\,M = \frac{n+1}{n}\,M.$$

Therefore the Rao–Blackwellised estimator is

$$W \;=\; \frac{n+1}{n}\,M \;=\; \frac{n+1}{n}\,\max\{X_1,\ldots,X_n\}.$$

EXERCISE 3 - SOLUTIONS

(a) Likelihood: $L(\theta) = \prod_{i=1}^n \frac{\theta^2}{1+\theta}(1+X_i)e^{-\theta X_i} = \left(\frac{\theta^2}{1+\theta}\right)^n \exp(-\theta S)\cdot\prod_{i=1}^n(1+X_i)$.
This factorises into a term $g(\theta, S)$ depending on the data only through S and a term $h(x_1,\ldots,x_n)$ free of θ. By the factorisation theorem, S is sufficient for θ.

(b) The log-likelihood is $\ell(\theta) = 2n\log\theta - n\log(1+\theta) - \theta S + c$.
First order derivative:

$$\ell'(\theta) = \frac{2n}{\theta} - \frac{n}{1+\theta} - S.$$

Setting $\ell'(\theta) = 0$ gives $\frac{2n}{\theta} - \frac{n}{1+\theta} = S \Leftrightarrow S\theta^2 + (S-n)\theta - 2n = 0$.

$$\Leftrightarrow \widehat{\theta}_{ML} = \frac{-(S-n) + \sqrt{(S-n)^2 + 8nS}}{2S}.$$

Note: As $S > 0$, the other root is negative, and so not a member of the parameter space. Check $\ell''(\theta) = -\frac{2n}{\theta^2} + \frac{n}{(1+\theta)^2} < 0$ at any θ, so it is indeed a maximum.

(c) For $\theta_1 > \theta_0$ the likelihood ratio is

$$\frac{L(\theta_1)}{L(\theta_0)} = \left(\frac{\theta_1}{\theta_0}\right)^{2n}\left(\frac{1+\theta_0}{1+\theta_1}\right)^n \exp\{-(\theta_1-\theta_0)S\}$$

All factors except the exponential are constant in S, and since $\theta_1 - \theta_0 > 0$, the ratio is strictly decreasing in S. Hence the family has a monotone likelihood ratio in S. By the Karlin–Rubin theorem, the UMP test for

$$H_0 : \theta \le \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0$$

rejects H_0 if $S \le c$. Note: c needs to be chosen such that $P_{\theta_0}(S \le c) = \alpha$ (not required).

(d) For $n = 1$ the log-likelihood is (compare part (b))

$$\ell''(\theta) = -\frac{2}{\theta^2} + \frac{1}{(1+\theta)^2}$$

The expected Fisher information is $I(\theta) = -\mathbb{E}_\theta[\ell''(\theta)] = \frac{2}{\theta^2} - \frac{1}{(1+\theta)^2}$.

(e) For large n we have:

$$\sqrt{n}\cdot(\hat{\theta}_{ML} - \theta) \sim \mathcal{N}(0, I(\theta)^{-1}) \Leftrightarrow \sqrt{I(\theta)}\sqrt{n}\cdot(\hat{\theta}_{ML} - \theta) \sim \mathcal{N}(0,1)$$

It follows:

$$P(q_{0.05} \le \sqrt{I(\theta)}\sqrt{n}\cdot(\hat{\theta}_{ML} - \theta)) = 0.95$$

$$\Leftrightarrow P(\theta \le \hat{\theta}_{ML} + q_{0.95}\frac{1}{\sqrt{I(\theta)}\sqrt{n}}) = 0.95$$

Here $q_{0.95} \approx 1.64$ and

$$I(2) = \frac{2}{2^2} - \frac{1}{(1+2)^2} = \frac{1}{2} - \frac{1}{9} = \frac{7}{18}.$$

Hence

$$U = 2 + 1.64\cdot\frac{1}{\sqrt{7/18}}\cdot\frac{1}{\sqrt{10}} \approx 2.83.$$

Thus, the asymptotic one-sided 95% CI is $(-\infty, 2.83]$.

EXERCISE 4 - SOLUTIONS

(a) The likelihood is

$$L(\theta) = \prod_{i=1}^{n} \theta e^{-\theta X_i} = \theta^n e^{-\theta S}, \qquad \theta > 0.$$

The log-likelihood is $\ell(\theta) = n \log \theta - \theta S$. Differentiate:

$$\ell'(\theta) = \frac{n}{\theta} - S, \qquad \ell''(\theta) = -\frac{n}{\theta^2} < 0 \quad (\theta > 0).$$

Hence the ML estimator is

$$\hat{\theta}_{ML} = \frac{n}{S}.$$

(b) The likelihood ratio statistic is

$$\frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\theta_0)}{L(\hat{\theta}_{ML})} = \frac{\theta_0^n e^{-\theta_0 S}}{\left(\frac{n}{S}\right)^n e^{-n}} = \left(\frac{\theta_0 S}{n}\right)^n \exp\{\, n - \theta_0 S \,\} =: \Lambda(S).$$

Hence Λ is a function of the random sample only through S.

(c) Consider $\log \Lambda(S)$:

$$\log \Lambda(S) = n \log\left(\frac{\theta_0 S}{n}\right) + n - \theta_0 S.$$

Differentiate w.r.t. S:

$$\frac{d}{dS} \log \Lambda(S) = \frac{n}{S} - \theta_0, \qquad \frac{d^2}{dS^2} \log \Lambda(S) = -\frac{n}{S^2} < 0.$$

Thus $\log \Lambda(S)$ (and hence $\Lambda(S)$) is strictly concave and has a unique maximum at

$$\frac{n}{S} - \theta_0 = 0 \quad \Rightarrow \quad S = \frac{n}{\theta_0}.$$

Therefore $\Lambda(S)$ decreases as S moves away from n/θ_0, i.e. the LRT rejects for *small* values of $\Lambda(S)$, which correspond to S being either too small or too large:

$$\text{reject } H_0 \quad \text{if} \quad S \le c_1 \text{ or } S \ge c_2,$$

for suitable constants $0 < c_1 < n/\theta_0 < c_2$ chosen to obtain size α.

(d) The power function is

$$\beta(\theta) = P_\theta(\text{reject } H_0) = P_\theta(S \le c_1) + P_\theta(S \ge c_2) \qquad (\theta > 0).$$