Exam - Statistics (WBMA009-05) 2025/2026

Date and time: November 5, 2025, 11.45-13.45h
Place: Exam Hall 4, 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 5 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 EXAM
ON NEXT PAGE

1. **Random sample.** $\boxed{8}$
Consider a distribution $\mathcal{D}(\theta_1, \theta_2)$ with density

$$f_{\theta_1,\theta_2}(x) = \frac{1-\theta_1}{\theta_2} \cdot \exp\left\{\frac{x}{\theta_2}\right\} \cdot I_{(-\infty,0)}(x) + \frac{\theta_1}{\theta_2} \cdot \exp\left\{-\frac{x}{\theta_2}\right\} \cdot I_{[0,\infty)}(x)$$

where $0 < \theta_1 < 1$ is **unknown**, and $\theta_2 > 0$ is **known**, and a random sample

$$X_1, \ldots, X_n \sim \mathcal{D}(\theta_1, \theta_2)$$

The log likelihood is then given by:

$$l(\theta_1) = K \cdot \log(\theta_1) + (n-K) \cdot \log(1-\theta_1) - n \cdot \log(\theta_2) - \frac{S}{\theta_2}$$

where $S := \sum_{i=1}^{n} |X_i|$, and K is the number of random variables with non-negative values. This implies for the order statistics:

$$X_{(1)} \leq \ldots \leq X_{(n-K)} < 0 \leq X_{(n-K+1)} \leq \ldots < X_{(n)}$$

$$\text{and we have } \sum_{i=1}^{n-K} X_{(i)} + \sum_{i=n-K+1}^{n} -X_{(i)} = -\sum_{i=1}^{n} |X_{(i)}| = -S.$$

   (a) $\boxed{2}$ Show that K is a sufficient statistic for θ_1.
   (b) $\boxed{2}$ Show that $P(X_1 > 0) = \theta_1$ and compute $E[K]$.
   (c) $\boxed{2}$ Compute the ML estimator of θ_1 and show that it is unbiased.
   (d) $\boxed{2}$ Show that the Fisher information for sample size $n = 1$ is $I(\theta_1) = \frac{1}{\theta_1(1-\theta_1)}$.

2. **UMP test.** $\boxed{6}$
Consider a random sample of size $n = 9$: $X_1, \ldots, X_9 \sim N(\mu, \sigma^2)$, where $\sigma^2 = 1$ is known and μ is **unknown**.

   (a) $\boxed{3}$ Derive the uniformly most powerful test (UMP) for

$$H_0 : \mu = 0 \text{ vs. } H_1 : \mu = 1$$

   (b) $\boxed{3}$ Give the critical value **for the density ratio test statistic** of this UMP test at the level $\alpha = 0.1$. You may use the quantiles from Table 1.

<u>HINT:</u> The $\mathcal{N}(\mu, \sigma^2)$ density is: $f(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right\}$.

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.

3. **Asymptotic confidence intervals and tests.** [10]
Consider a random sample $X_1, \dots, X_n$ from a Negative Binomial distribution with known parameter $r \in \mathbb{N}$ and **unknown** probability parameter $\theta \in (0,1)$. Recall that the density and the expectation of a Negative Binomial are

$$f(x) = \binom{x+r-1}{x} \cdot (1-\theta)^x \cdot \theta^r \quad (x \in \mathbb{N}_0), \quad \text{and} \quad E[X] = \frac{(1-\theta)r}{\theta}$$

(a) [2] Show that the ML estimator of θ is: $\hat{\theta}_{ML} = r/(r+\bar{X})$, where $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$. It suffices to show that the 1st derivative of the log-likelihood is equal to zero.

(b) [2] Show that the expected Fisher information (for sample size 1) is

$$I(\theta) = \frac{r}{\theta^2 \cdot (1-\theta)}$$

From now on we assume that $r = 2$, $n = 20$ and that $\bar{X} = 8$ has been observed.

(c) [3] Make use of the asymptotic efficiency of the ML estimator and give a one-sided asymptotic 95% confidence interval $(-\infty, U]$ for θ.

(d) [3] Check whether a score-test to the level $\alpha = 0.05$ would reject the null hypothesis $H_0 : \theta = 0.25$ in favor of the alternative $H_1 : \theta \neq 0.25$.

<u>**HINTS:**</u> You can use the quantiles from Table 1.

Score test: $\frac{d}{d\theta} l_X(\theta)/\sqrt{n \cdot I(\theta)}$ is asymptotically $N(0,1)$ distributed.

4. **Multivariate Gaussian distribution.** [4]
Consider a random sample from a Gaussian distribution

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

Show that the random vector $\mathbf{X} := (X_1, \dots, X_n)^\top$ has an n-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu} = \mu\mathbf{1}$ and covariance matrix $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a vector of ones.

<u>**HINT:**</u> Densities of the $\mathcal{N}(\mu, \sigma^2)$ and the n-variate $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ Gaussian distributions:

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right\}$$

$$f(\mathbf{x}) = (2\pi)^{-n/2} \cdot \det(\boldsymbol{\Sigma})^{-1/2} \cdot \exp\left\{-\frac{1}{2} \cdot (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\}$$

5. **Properties of Maximum Likelihood Estimators.** $\boxed{8}$
Consider a random sample from a distribution $\mathcal{F}_\theta$ that depends on one unknown parameter $\theta \in \mathbb{R}$:

$$X_1, \dots, X_n \sim \mathcal{F}_\theta$$

Let $\hat{\theta}_{\mathrm{ML},n}$ denote the ML estimator of θ, and $l'_X(\theta)$ and $l''_X(\theta)$ be the 1st and 2nd derivatives of the log-likelihood $l_X(\theta)$, respectively.

(a) $\boxed{4}$ Show that for large sample sizes n:

$$\sqrt{n}\,(\hat{\theta}_{\mathrm{ML},n} - \theta) \approx -\sqrt{n} \cdot \frac{l'_X(\theta)}{l''_X(\theta)}$$

(b) $\boxed{4}$ Show that for large sample sizes n:

$$\frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{I(\theta)}} \cdot l'_X(\theta) \sim \mathcal{N}(0,1)$$

where $I(\theta)$ is the expected Fisher information (of a sample of size 1).

HINT: You may assume that all required regulatory conditions are satisfied. In part (b) you can use that

$$I(\theta) = E\left[\left(\frac{d}{d\theta}\, l_{X_i}(\theta)\right)^2\right] \quad \text{and} \quad E\left[\frac{d}{d\theta}\, l_{X_i}(\theta)\right] = 0.$$

EXERCISE 1 - SOLUTIONS

(a) Likelihood:

$$L(\theta_1) = \left(\frac{1-\theta_1}{\theta_2}\right)^{n-K} \cdot \exp\left\{\frac{\sum_{i=1}^{n-K} x_{(i)}}{\theta_2}\right\} \cdot \left(\frac{\theta_1}{\theta_2}\right)^{K} \cdot \exp\left\{\frac{-\sum_{i=n-K+1}^{n} x_{(i)}}{\theta_2}\right\}$$

The factorization theorem with

$$\begin{aligned} h(x) &= \exp\left\{\frac{-\sum_{i=1}^{n} |x_{(i)}|}{\theta_2}\right\} = \exp\left\{\frac{-S}{\theta_2}\right\} \\ g(K,\theta_1) &= \left(\frac{1-\theta_1}{\theta_2}\right)^{n-K} \cdot \left(\frac{\theta_1}{\theta_2}\right)^{K} \end{aligned}$$

implies that K is a sufficient statistic for θ_1

(b)

$$P(X_i > 0) = \int_0^\infty \frac{\theta_1}{\theta_2} \cdot \exp\left\{-\frac{x}{\theta_2}\right\}\, dx = \frac{\theta_1}{\theta_2} \cdot \left(-\theta_2 \exp\left\{-\frac{\infty}{\theta_2}\right\} + \theta_2 \exp\left\{-\frac{0}{\theta_2}\right\}\right) = \theta_1$$

As we have a random sample of size n, it follows: $E[K] = \sum_{i=1}^{n} \theta_1 = n \cdot \theta_1$.

(c) Take the derivatives w.r.t. θ_1:

$$\begin{aligned} l'(\theta_1) &= \frac{K}{\theta_1} - (n-K) \cdot \frac{1}{1-\theta_1} \\ l''(\theta_1) &= -\frac{K}{\theta_1^2} - (n-K) \cdot \frac{1}{(1-\theta_1)^2} \end{aligned}$$

Setting the first derivative equal to zero, yields:

$$l'(\theta_1) = 0 \Leftrightarrow K(1-\theta_1) - (n-K)\cdot\theta_1 = 0 \Leftrightarrow \theta_1 = \frac{K}{n}$$

And as $l''(\theta_1) < 0$ for all θ_1, we indeed have a maximum, so that $\hat{\theta}_{1,ML} = \frac{K}{n}$.

To check whether the ML estimator is unbiased, we compute:

$$E\left[\hat{\theta}_{1,ML}\right] = E\left[\frac{K}{n}\right] = \frac{E[K]}{n} = \frac{n\cdot\theta_1}{n} = \theta_1$$

where we used the result from exercise part **(b)**.

(d) We re-use the 2nd derivative of the log-likelihood from part **(c)**, and we recall from part **(b)** that $E[K] = n \cdot \theta_1$. For $n = 1$ we get:

$$I(\theta_1) = -E\left[-\frac{K}{\theta_1^2} - (1-K)\frac{1}{(1-\theta_1)^2}\right] = \frac{E[K]}{\theta_1^2} + \frac{1-E[K]}{(1-\theta_1)^2} = \frac{1}{\theta_1} + \frac{1}{1-\theta_1} = \frac{1}{\theta_1(1-\theta_1)}$$

EXERCISE 2 - SOLUTIONS:
(a) Density ratio:

$$\begin{aligned}
W(X) &= \frac{f_{0,1}(X_1,\ldots,X_9)}{f_{1,1}(X_1,\ldots,X_9)} = \frac{\prod\limits_{i=1}^{9} f_{0,1}(X_i)}{\prod\limits_{i=1}^{9} f_{1,1}(X_i)} = \frac{\prod\limits_{i=1}^{9} \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{1}\cdot\exp\{-\frac{1}{2}\frac{(X_i-0)^2}{1^2}\}}{\prod\limits_{i=1}^{9} \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{1}\cdot\exp\{-\frac{1}{2}\frac{(X_i-1)^2}{1^2}\}} \\
&= \frac{\exp\{-\frac{1}{2}\sum\limits_{i=1}^{9} X_i^2\}}{\exp\{-\frac{1}{2}\sum\limits_{i=1}^{9}(X_i-1)^2\}} = \exp\{-\frac{1}{2}\sum_{i=1}^{9} X_i^2 + \frac{1}{2}\sum_{i=1}^{9}(X_i^2 - 2X_i + 1)\} \\
&= \exp\{4.5 - \sum_{i=1}^{9} X_i\}
\end{aligned}$$

$W(X)$ is monotone decreasing in $\sum\limits_{i=1}^{9} X_i$. Hence the UMP rejects if $\sum\limits_{i=1}^{9} X_i > k$

(b) Under H_0 we have: $\sum\limits_{i=1}^{9} X_i \sim N(0,9)$, so that $\frac{1}{3}\sum\limits_{i=1}^{9} X_i \sim N(0,1)$.

$$\begin{aligned}
P_{H_0}(W(X_1,\ldots,X_9) < k) &\Leftrightarrow P_{H_0}(\exp\{4.5 - \sum_{i=1}^{9} X_i\} < k) \\
\Leftrightarrow P_{H_0}(\sum_{i=1}^{9} X_i > 4.5 - \log(k)) &\Leftrightarrow P_{H_0}(\frac{1}{3}\sum_{i=1}^{9} X_i > 1.5 - \log(k)/3)
\end{aligned}$$

So $(1.5 - \log(k)/3)$ must be the 0.9-quantile $q_{0.9} = 1.28$ of the $N(0,1)$ distribution.

$$1.5 - \log(k)/3 = 1.28 \Leftrightarrow k = \exp(0.66) \approx 1.93$$

Hence, the UMP rejects H_0 if: $W(X) < 1.93$.

EXERCISE 3 - SOLUTIONS

(a): Log-likelihood:

$$
\begin{aligned}
l_X(\theta) &= \log\left(\prod_{i=1}^{n}\binom{x_i+r-1}{x_i}\cdot(1-\theta)^{x_i}\cdot\theta^r\right)\\
&= \log\left((\prod_{i=1}^{n}\binom{x_i+r-1}{x_i})\cdot(1-\theta)^{\sum_{i=1}^n x_i}\cdot\theta^{n\,r}\right)\\
&= \log\left(\prod_{i=1}^{n}\binom{x_i+r-1}{x_i}\right)+(\sum_{i=1}^{n}x_i)\log(1-\theta)+n\,r\log(\theta)
\end{aligned}
$$

Take the derivative w.r.t. θ and set it to 0:

$$
\frac{-\sum_{i=1}^n x_i}{1-\theta}+\frac{n\,r}{\theta}=0 \;\Leftrightarrow\; -(\sum_{i=1}^{n}x_i)\theta+n\,r(1-\theta)=0 \Leftrightarrow -(nr+\sum_{i=1}^{n}x_i)\theta+n\,r=0
$$

$$
\Leftrightarrow\; \theta=\frac{n\,r}{n\,r+\sum_{i=1}^n x_i}\Rightarrow\hat{\theta}_{ML}=\frac{r}{r+\bar{X}}
$$

(b): For $n=1$ we have: $\frac{d^2}{d\theta^2}l_{X_1}(\theta)=\frac{-x_1}{(1-\theta)^2}-\frac{r}{\theta^2}$, and the Fisher information is:

$$
\begin{aligned}
I(\theta) &= -E_\theta\left[\frac{d^2}{d\theta^2}l_{X_1}(\theta)\right]=E_\theta\left[\frac{X_1}{(1-\theta)^2}+\frac{r}{\theta^2}\right]=\frac{E[X_1]}{(1-\theta)^2}+\frac{r}{\theta^2}\\
&= \frac{\frac{r(1-\theta)}{\theta}}{(1-\theta)^2}+\frac{r}{\theta^2}=\frac{r}{(1-\theta)\theta}+\frac{r}{\theta^2}=\frac{r\theta+r(1-\theta)}{(1-\theta)\theta^2}=\frac{r}{\theta^2(1-\theta)}
\end{aligned}
$$

(c): Asymptotically $\sqrt{I(\theta)}\sqrt{n}\cdot(\hat{\theta}_{ML}-\theta)\sim\mathcal{N}(0,1)$, hence:

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

With $q_{0.05}=-1.64$ the one-sided 95% CI for θ is: $(-\infty,\hat{\theta}_{ML}+\frac{1.64}{\sqrt{I(\hat{\theta}_{ML})}\cdot\sqrt{n}}\,]$

Here we have $\hat{\theta}_{ML}=0.2$ and $\frac{1.64}{\sqrt{I(\hat{\theta}_{ML})}\sqrt{n}}=\frac{1.64}{\sqrt{\frac{2}{0.2^2\cdot 0.8}}\sqrt{20}}\approx 0.047.$

So the one-sided CI is: $(-\infty,0.247]$.

(d): Asymptotically: $\frac{\frac{d}{d\theta}l_X(\theta)}{\sqrt{n\cdot I(\theta)}}\sim N(0,1)$ where $\frac{d}{d\theta}l_X(\theta)=\frac{-\sum_{i=1}^n x_i}{1-\theta}+\frac{n\,r}{\theta}$.

Given $r=2$, $\bar{X}=8$ and $n=20$ and $\theta_0=0.25$ we get:
$\frac{-\sum_{i=1}^n x_i}{1-\theta}+\frac{n\,r}{\theta}=\frac{-160}{1-0.25}+\frac{40}{0.25}\approx-53.33$ and $\sqrt{n\cdot I(\theta)}=\sqrt{20\cdot\frac{2}{0.25^2\cdot 0.75}}\approx 29.21$

Therefore the score test statistic takes the value: $\frac{\frac{d}{d\theta}l_X(\theta)}{\sqrt{n\cdot I(\theta)}}=\frac{-53.33}{29.21}\approx-1.83.$

As the value is **not** lower than the $q_{0.025}$ quantile -1.96 of the $N(0,1)$, the score test would **NOT** reject the null hypothesis to the level 0.05.

EXERCISE 4 - SOLUTIONS

Show that the joint density and the multivariate density are identical:

$$\begin{aligned} f(\mathbf{x}) &= (2\pi)^{-n/2} \cdot \det(\sigma^2 \cdot \mathbf{I})^{-1/2} \cdot \exp\left\{-\frac{1}{2} \cdot (\mathbf{x} - \mu \cdot \mathbf{1})^T (\sigma^2 \cdot \mathbf{I})^{-1} (\mathbf{x} - \mu \cdot \mathbf{1})\right\} \\ &= (2\pi)^{-n/2} \cdot \frac{1}{\sigma^n} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right\} \\ &= (2\pi)^{-n/2} \cdot \frac{1}{\sigma^n} \cdot \prod_{i=1}^{n} \exp\left\{-\frac{1}{2} \cdot \frac{1}{\sigma^2} (x_i - \mu)^2\right\} \\ &= \prod_{i=1}^{n} (2\pi)^{-1/2} \cdot \frac{1}{\sigma} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(x_i - \mu)^2}{\sigma^2}\right\} = \prod_{i=1}^{n} f(x_i) = f(x_1, \ldots, x_n) \end{aligned}$$

EXERCISE 5 - SOLUTIONS

(a) 1st order Taylor series expansion of $l'_X(.)$ around the true parameter θ yields:

$$l'_X(\theta^\star) \approx l'_X(\theta) + (\theta^\star - \theta) \cdot l''_X(\theta)$$

Plug in $\theta^\star = \hat{\theta}_{ML,n}$ (which for large n will be close to the true θ):

$$l'_X(\hat{\theta}_{ML,n}) \approx l'_X(\theta) + (\hat{\theta}_{ML,n} - \theta) \cdot l''_X(\theta)$$

As $l'_X(\hat{\theta}_{ML,n}) = 0$, we get:

$$0 \approx l'_X(\theta) + (\hat{\theta}_{ML,n} - \theta) \cdot l''_X(\theta)$$

what can be transformed into:

$$-\frac{l'_X(\theta)}{l''_X(\theta)} \approx \hat{\theta}_{ML,n} - \theta \quad \Leftrightarrow \quad -\sqrt{n}\ \frac{l'_X(\theta)}{l''_X(\theta)} \approx \sqrt{n}\ (\hat{\theta}_{ML,n} - \theta)$$

(b) We have:

$$l'_X(\theta) = \frac{d}{d\theta}\ l_X(\theta) = \frac{d}{d\theta} \log\left(\prod_{i=1}^{n} f_\theta(X_i)\right) = \sum_{i=1}^{n} \frac{d}{d\theta}\ l_{X_i}(\theta)$$

This is the sum of the iid sample: $\frac{d}{d\theta}\ l_{X_1}(\theta), \ldots, \frac{d}{d\theta}\ l_{X_n}(\theta)$ with:

$$E\left[\frac{d}{d\theta} l_{X_i}(\theta)\right] = 0 \quad \text{and} \quad V\left(\frac{d}{d\theta}\ l_{X_i}(\theta)\right) = E\left[\left(\frac{d}{d\theta}\ l_{X_i}(\theta)\right)^2\right] - 0^2 = I(\theta)$$

Therefore, it follows from the **central limit theorem** for the true θ:

$$\sqrt{n} \cdot \frac{\left(\frac{1}{n} \cdot \sum\limits_{i=1}^{n} \frac{d}{d\theta}\ l_{X_i}(\theta)\right) - 0}{\sqrt{I(\theta)}} \xrightarrow{D} \mathcal{N}(0, 1)$$

Henceforth, we have for large n:

$$\frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{I(\theta)}} \cdot l'_X(\theta) \sim \mathcal{N}(0, 1)$$