Exam - Statistics (WBMA009-05) 2024/2025

Date and time: November 6, 2024, 18.15-20.15h
Place: Exam Hall 2, Blauwborgje 4

Rules to follow:

- This is a closed book exam. Consultation of books and notes is **not** permitted. You can use a simple (non-programmable) calculator.
- Write your name and student number onto each paper sheet. There are 4 exercises and you can reach 90 points. ALWAYS include the relevant equation(s) and/or short derivations.
- **We wish you success with the completion of the exam!**

START OF EXAM

1. **Cramer-Rao bound and asymptotic confidence interval.** **30**
   Consider a random sample

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

where $\mu = 0$ and the variance $\sigma^2 > 0$ is unknown. Recall the density of the $N(\mu, \sigma^2)$:

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\} \qquad (x \in \mathbb{R})$$

(a) **5** Show that $T(X_1, \dots, X_n) := \frac{1}{n}\sum_{i=1}^{n} X_i^2$ is a sufficient statistic.

(b) **5** Show that the estimator $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ is unbiased.

(c) **5** Show that the expected Fisher information (for sample size 1) is $(2\sigma^4)^{-1}$.

(d) **5** Check whether the estimator from (b) attains the Cramer-Rao bound.

(e) **10** Assume $n = 25$ and that the realization $\widehat{\sigma^2} = 2$ has been observed. Derive an asymptotic two-sided 80% confidence interval for σ^2.
The relevant quantiles can be found in Table 1.

HINT: Given a random sample $X_1, \dots, X_n$ from a $N(0,1)$ distribution, it follows: $Z := \left(\sum_{i=1}^{n} X_i^2\right) \sim \chi^2_n$, with $E[Z] = n$ and $Var(Z) = 2n$.

α	0.5	0.75	0.9	0.95	0.975	0.99	0.99997
q_α	0	0.7	1.3	1.6	2	2.3	4

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

2. **Random sample.** $\boxed{30}$
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 $0 < \theta_1 < 1$ and $\theta_2 > 0$. We consider a random sample of size n:

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

(a) $\boxed{10}$ Show that the log likelihood is given by:

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

where $S := \sum_{i=1}^{n} |X_i|$, and $K := \sum_{i=1}^{n} I_{[0,\infty)}(X_i)$.

HINTS: Use the order statistics and that:

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

$$\sum_{i=1}^{n-K} X_{(i)} + \sum_{i=n-K+1}^{n} -X_{(i)} = -S$$

For the following exercise parts (b-d), we assume that θ_2 is a **known** parameter.

(b) $\boxed{5}$ Show that $P(X_1 > 0) = \theta_1$.

(c) $\boxed{5+5}$ Compute the ML estimator of θ_1 and show that it is unbiased.

(d) $\boxed{5}$ Show that the Fisher information for sample size $n = 1$ is $I(\theta_1) = \frac{1}{\theta_1(1-\theta_1)}$.

3. **UMP test for Poisson sample.** $\boxed{10}$
Let $X_1, \ldots, X_n$ be a sample from a Poisson distribution with density:

$$p(x|\lambda) = e^{-\lambda} \cdot \frac{\lambda^x}{x!} \qquad (x \in \mathbb{N}_0)$$

where $\lambda > 0$ is an unknown parameter.
Recall that $T(X_1, \ldots, X_n) := \sum_{i=1}^{n} X_i$ has a Poisson distribution with parameter $n\lambda$.
Derive the uniform most powerful (UMP) test for the test problem

$$H_0 : \lambda = 2 \quad \text{vs.} \quad H_1 : \lambda = 3$$

to the significance level $\alpha = 0.05$. Use the symbol $q_{\lambda,\alpha}$ for the α quantile of a Poisson distribution with parameter λ.

4. **Likelihood Ratio Test Statistic.** $\boxed{20}$

Consider a random sample

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

where $\theta \in \mathbb{R}$. Let θ_0 and $\hat{\theta}_{ML}$ denote the true parameter and the ML estimator.

(a) $\boxed{10}$ Use a second order Taylor series expansion of the log-likelihood to show

$$-2\log\left(\frac{L(\theta_0)}{\max\limits_{\theta\in\mathbb{R}}\{L(\theta)\}}\right) \approx -(\hat{\theta}_{ML} - \theta_0)^2 \cdot l''(\hat{\theta}_{ML})$$

where $L(\theta)$ and $l(\theta)$ denote the likelihood and the log-likelihood, respectively, and $l''(.)$ is the 2nd derivative of the log-likelihood.

(b) $\boxed{5}$ Show that the law of the large number implies for $n \to \infty$:

$$\frac{1}{n} \cdot l''(\theta_0) \to -I(\theta_0)$$

where $I(.)$ is the expected Fisher information for a sample of size $n = 1$.

(c) $\boxed{5}$ Show that the central limit theorem implies for $n \to \infty$:

$$\frac{l'(\theta_0)}{\sqrt{n \cdot I(\theta_0)}} \to \mathcal{N}(0, 1)$$

HINTS: You can assume that all the required regulatory conditions are fulfilled. For example, that the sample space does not depend on θ and that the density $f(x|\theta)$ of $\mathcal{F}_\theta$ is twice continuously differentiable.

EXERCISE 1 - SOLUTIONS
Note that $X_i \sim N(0,\sigma^2)$ implies $\frac{1}{\sigma}X_i \sim N(0,1)$, so that (see hint):
$\sum_{i=1}^{n} \left(\frac{X_i}{\sigma}\right)^2$ is χ^2_n distributed.

(a) This follows from the factorization theorem, as:

$$f(x_1,\dots,x_n) = (2\pi)^{-n/2} \cdot \sigma^{-n} \exp\left\{ -0.5 \cdot \frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} \right\} = g(\sum_{i=1}^{n} x_i^2, \sigma^2) \cdot h(x_1,\dots,x_n)$$

with $h(x_1,\dots,x_n) = 1$

(b): We have: $E[\widehat{\sigma^2}] = \frac{1}{n} E\left[\sum_{i=1}^{n} \left(\frac{X_i}{\sigma}\right)^2\right] \cdot \sigma^2 = \frac{1}{n} \cdot n \cdot \sigma^2 = \sigma^2$

(c): Build the log-likelihood (for X_1 only):

$$l_{X_1}(\sigma^2) = \log\left(\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\{-\frac{1}{2} \cdot \frac{X_1^2}{\sigma^2}\}\right) = \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2} \cdot \log(\sigma^2) - \frac{1}{2} \cdot \frac{1}{\sigma^2} \cdot X_1^2$$

Take the first and the second derivative w.r.t. σ^2:

$$\begin{aligned}
\frac{d}{d\sigma^2}\, l_{X_1}(\sigma^2) &= -\frac{1}{2} \cdot \frac{1}{\sigma^2} \quad - \frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} \cdot (-1) \cdot X_1^2 \\
&= -\frac{1}{2} \cdot \frac{1}{\sigma^2} \quad + \frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} \cdot X_1^2 \\
\frac{d^2}{d\sigma^2\, d\sigma^2}\, l_{X_1}(\sigma^2) &= -\frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} \cdot (-1) \quad + \frac{1}{2} \cdot \frac{1}{(\sigma^2)^3} \cdot (-2) \cdot X_1^2 \\
&= +\frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} \quad - \frac{1}{(\sigma^2)^3} \cdot X_1^2
\end{aligned}$$

The expected Fisher information is:

$$\begin{aligned}
I(\sigma^2) &= -E\left[\frac{d^2}{d\sigma^2\, d\sigma^2}\, l_{X_1}(\sigma^2)\right] = -E\left[\frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \cdot X_1^2\right] \\
&= -\frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} \cdot E[X_1^2] = -\frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} \cdot \sigma^2 = \frac{1}{2\sigma^4}
\end{aligned}$$

(d) Yes, the estimator attains the Carmer-Rao bound $\frac{1}{n \cdot I(\sigma^2)} = \frac{1}{n \cdot \frac{1}{2\sigma^4}} = \frac{2\sigma^4}{n}$, because:

$$\begin{aligned}
Var(\widehat{\sigma^2}) &= \frac{1}{n^2} \cdot Var\left(\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n^2} \cdot Var\left(\sigma^2 \cdot \sum_{i=1}^{n} \left(\frac{X_i}{\sigma}\right)^2\right) \\
&= \frac{\sigma^4}{n^2} \cdot Var\left(\sum_{i=1}^{n} \left(\frac{X_i}{\sigma}\right)^2\right) = \frac{\sigma^4}{n^2} \cdot 2n = \frac{2\sigma^4}{n}
\end{aligned}$$

(e) For ML estimators we asymptotically have: $\sqrt{I(\theta)}\sqrt{n} \cdot (\hat{\theta}_{ML} - \theta) \sim \mathcal{N}(0,1)$, hence:

$$P(\hat{\theta}_{ML} - \frac{q_{1-\alpha/2}}{\sqrt{I(\theta)} \cdot \sqrt{n}} \leq \theta \leq \hat{\theta}_{ML} - \frac{q_{\alpha/2}}{\sqrt{I(\theta)} \cdot \sqrt{n}}) = 0.95$$

With $q_{0.90} = 1.3$ and $q_{0.10} = -1.3$, and $I(\theta)$ being replaced by $I(\widehat{\sigma^2})$, we get the CI:

$$\widehat{\sigma^2} \pm 1.3/(\sqrt{I(\widehat{\sigma^2})} \cdot \sqrt{n})$$

Here we have $\widehat{\sigma^2} = 2$ and $1.3/\left(\sqrt{I(\widehat{\sigma^2})} \cdot \sqrt{n}\right) = 1.3/\left(\sqrt{\frac{1}{2 \cdot 2^2}} \cdot \sqrt{25}\right) = 1.3/\sqrt{\frac{25}{8}} \approx 0.74$.

So the two-sided 80% CI is: $[1.26, 2.74]$.

EXERCISE 2 - SOLUTIONS

For $x < 0$ we have: $f_{\theta_1,\theta_2}(x) = \frac{1-\theta_1}{\theta_2} \cdot \exp\left\{\frac{x}{\theta_2}\right\}$
For $x \geq 0$ we have: $f_{\theta_1,\theta_2}(x) = \frac{\theta_1}{\theta_2} \cdot \exp\left\{-\frac{x}{\theta_2}\right\}$

(a) Assume that $X_{(1)} \leq \ldots \leq X_{(n-K)} < 0 \leq X_{(n-K+1)} \leq \ldots < X_{(n)}$. Then:

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} f_{\theta_1,\theta_2}(X_i) = \prod_{i=1}^{n} f_{\theta_1,\theta_2}(X_{(i)})$$

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

We obtain the log-likelihood:

$$\begin{aligned} l(\theta_1, \theta_2) &= (n-K)\log\left(\frac{1-\theta_1}{\theta_2}\right) + \frac{\sum\limits_{i=1}^{n-K} x_{(i)}}{\theta_2} + K\log\left(\frac{\theta_1}{\theta_2}\right) + \frac{\sum\limits_{i=n-K+1}^{n} -x_{(i)}}{\theta_2} \\ &= (n-K)\log\left(\frac{1-\theta_1}{\theta_2}\right) + K\log\left(\frac{\theta_1}{\theta_2}\right) - \frac{S}{\theta_2} \qquad \text{(see HINT)} \\ &= K \cdot \log(\theta_1) + (n-K)\cdot\log(1-\theta_1) - n\cdot\log(\theta_2) - \frac{S}{\theta_2} \end{aligned}$$

(b)

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

(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 $\hat{\theta}_{1,ML} = \frac{K}{n}$.

Check whether the ML estimator is unbiased:

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

(d) For $n = 1$:

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

EXERCISE 3 - SOLUTIONS
Compute the density ratio:

$$W(X_1, \ldots, X_n) = \frac{p(X_1, \ldots, X_n|\lambda = 3)}{p(X_1, \ldots, X_n|\lambda = 2)} = \frac{e^{-3n} \cdot \frac{3^{(\sum_{i=1}^{n} X_i)}}{\prod_{i=1}^{n} X_i!}}{e^{-2n} \cdot \frac{2^{(\sum_{i=1}^{n} X_i)}}{\prod_{i=1}^{n} X_i!}} = e^{-n} \cdot \left(\frac{3}{2}\right)^{\sum_{i=1}^{n} X_i}$$

The density ratio is a monotone increasing function in $\sum_{i=1}^{n} X_i$.

We reject H_0 if $e^{-n} \cdot \left(\frac{3}{2}\right)^{\sum_{i=1}^{n} X_i} > k \Leftrightarrow \sum_{i=1}^{n} X_i > k_0$.

Under H_0 the statistic $\sum_{i=1}^{n} X_i$ has a Poisson distribution with parameter $2n$.

Therefore, the decision rule is to reject H_0 if $\sum_{i=1}^{n} X_i > q_{2n,\,0.95}$.

EXERCISE 4 - SOLUTIONS:

(a) Second order Taylor series expansion:

$$\begin{aligned} l(\theta_0) &\approx l(\hat{\theta}_{ML}) + (\hat{\theta}_{ML} - \theta_0) \cdot l'(\hat{\theta}_{ML}) + \frac{1}{2}(\hat{\theta}_{ML} - \theta_0)^2 \cdot l''(\hat{\theta}_{ML}) \\ &\approx l(\hat{\theta}_{ML}) + \frac{1}{2}(\hat{\theta}_{ML} - \theta_0)^2 \cdot l''(\hat{\theta}_{ML}) \end{aligned}$$

as $l'(\hat{\theta}_{ML}) = 0$.

It follows

$$\begin{aligned} -2\log\left(\frac{L(\theta_0)}{\max\limits_{\theta\in\Theta}\{L(\theta)\}}\right) &= -2\log\left(\frac{L(\theta_0)}{L(\hat{\theta}_{ML})}\right) \\ &= -2 \cdot (l(\theta_0) - l(\hat{\theta}_{ML})) \\ &\approx -2 \cdot \left(l(\hat{\theta}_{ML}) + \frac{1}{2}(\hat{\theta}_{ML} - \theta_0)^2 \cdot l''(\hat{\theta}_{ML}) - l(\hat{\theta}_{ML})\right) \\ &\approx -(\hat{\theta}_{ML} - \theta_0)^2 \cdot l''(\hat{\theta}_{ML}) \end{aligned}$$

(b) We have:

$$\begin{aligned} \frac{1}{n} \cdot l''(\theta_0) &= \frac{1}{n} \cdot \left(\frac{d^2}{d\theta^2}\log(\prod_{i=1}^{n} f(X_i|\theta))\right)|_{\theta=\theta_0} \\ &= \frac{1}{n} \cdot \left(\frac{d^2}{d\theta^2}\sum_{i=1}^{n}\log(f(X_i|\theta))\right)|_{\theta=\theta_0} \\ &= \frac{1}{n} \cdot \sum_{i=1}^{n}\left(\frac{d^2}{d\theta^2}\log(f(X_i|\theta))\right)|_{\theta=\theta_0} \\ &= \frac{1}{n} \cdot \sum_{i=1}^{n} l''_{X_i}(\theta_0) \end{aligned}$$

The law of the Large numbers (LLN) implies that the mean converges in probability to

$$E\left[l''_{X_1}(\theta_0)\right] = E\left[\frac{d^2}{d\theta^2}\log(f(X_1|\theta))\right]|_{\theta=\theta_0} = -I(\theta_0)$$

(c) Exactly like in (b) we get: $\frac{1}{n} \cdot l'(\theta_0) = \ldots = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} l'_{X_i}(\theta_0)$.

Therefore it follows from the CLT:

$$\sqrt{n} \cdot \frac{\frac{1}{n} \cdot l'(\theta_0) - E[l'_{X_1}(\theta_0)]}{\sqrt{V\left(l'_{X_1}(\theta_0)\right)}} = \frac{\frac{1}{\sqrt{n}} \cdot l'(\theta_0)}{\sqrt{E\left[l'_{X_1}(\theta_0)^2\right]}} = \frac{l'(\theta_0)}{\sqrt{n \cdot I(\theta_0)}} \to \mathcal{N}(0,1)$$

because $E[l'_{X_1}(\theta_0)] = \int \frac{f'_{\theta_0}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) d_x = \frac{d}{d\theta}\int f_{\theta_0}(x) d_x = 0$ and $I(\theta_0) := E\left[l'_{X_1}(\theta_0)^2\right]$.