Exam - Statistics (WBMA009-05) 2023/2024

Date and time: November 8, 2023, 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. **Asymptotic confidence intervals and tests.** **30**
   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

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

   (a) **5** Show that the ML estimator of θ is: $\hat{\theta}_{ML} = \bar{X}/(r+\bar{X})$, 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) **5** Show that the expected Fisher information (for sample size 1) is

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

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

   (c) **10** Make use of the asymptotic efficiency of the ML estimator and give a one-sided asymptotic 95% confidence interval $(-\infty, U]$ for θ.

   (d) **10** Check whether a score-test to the level $\alpha = 0.02$ would reject the null hypothesis $H_0 : \theta = 0.9$ in favour of the alternative $H_1 : \theta \neq 0.9$.

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

   <u>HINT 2</u>: You can use the approximate quantiles provided in Table 1.

2. **Linear regression.** 30

Assume that we have n two-dimensional data points of the form:

$$(Y_1, x_1), ..., (Y_n, x_n)$$

where $Y_1, \ldots, Y_n$ are n independently (but not identically) Gaussian distributed random variables. Each random variable Y_i is accompanied by a non-random 'covariate' value $x_i \in \mathbb{R}$ that has an effect on the expectation of Y_i in the following way:

$$Y_i \sim \mathcal{N}(\theta \cdot x_i,\ 1), \quad (i = 1, \ldots, n)$$

where $\theta \in \mathbb{R}$ is an unknown parameter.

(a) 10 Show that the Maximum Likelihood estimator $\hat{\theta}_{ML}$ of θ is given by

$$\hat{\theta}_{ML} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$

It suffices to show that the 1st derivative of the log-likelihood is equal to zero.

(b) 5 Show that the ML estimator from (a) is unbiased estimator of θ and that its variance is given by:

$$V(\hat{\theta}_{ML}) = \left(\sum_{i=1}^{n} x_i^2\right)^{-1}$$

(c) 5 We have that:

$$\hat{\theta}_{ML} \sim \mathcal{N}\left(\theta,\ \left(\sum_{i=1}^{n} x_i^2\right)^{-1}\right)$$

Use this to specify a statistical level α test for the two-sided test problem:

$$H_0 : \theta = \theta_0 \quad \underline{\text{versus}} : \quad H_1 : \theta \neq \theta_0$$

Give the test statistic, the rejection region and the decision rule.

(d) 10 Now assume that the data fulfil:

$$\sum_{i=1}^{n} x_i y_i = 27$$

$$\sum_{i=1}^{n} x_i^2 = 81$$

Give an exact 95% confidence interval for θ.

<u>HINT</u>: You can use the quantiles provided in Table 1.

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.

3. **Sufficient statistic and UMP test.** 15
   Consider a random sample
   $$X_1, \dots, X_n \sim \mathcal{F}(\theta)$$
   from a distribution with sample space $S = \mathbb{R}^+$ that depends on a parameter $\theta > 0$ and whose density is:
   $$f_\theta(x) = \frac{1}{2} \cdot \theta^3 \cdot x^2 \exp\{-x\theta\} \qquad (x > 0)$$

   (a) 5 Give a sufficient statistic for θ.

   (b) 10 Consider the simple test problem
   $$H_0 : \theta = 4 \ \textbf{vs.} \ H_1 : \theta = 2$$
   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 UMP test for the above test problem.

4. **Test level, power and p-value of a statistical test.** 15
   Consider a random sample of size n from a Gaussian distribution with known variance parameter $\sigma^2 = 4$:
   $$X_1, \dots, X_n \sim \mathcal{N}(\mu, 4)$$
   and the simple test problem
   $$H_0 : \mu = 0 \ \textbf{vs.} \ H_1 : \mu = -0.54$$
   The null hypothesis is rejected when $\bar{x}_n \leq -0.4$.

   (a) 5 Given sample size $n = 100$, what is the test level?

   (b) 5 Given sample size $n = 100$, what is the power of the test?

   (c) 5 How large must n be to obtain the power 0.9?

   HINT: For solving this exercise use the approximate quantiles provided in Table 1.