Exercise Sheet 10

June 22nd 2023

Submission of the homework assignments until June 29th, 11:30 am online in TUM-moodle in groups of two. Please put the *full names* and *student IDs* of you *and* your partner on all parts of your submission. The solution will be discussed in the classes and published on moodle after some time.

Homework

Problem H 40 - Maximum Likelihood Estimators **[6 pts.]**

Let $\vec{X} = (X_1, X_2, \ldots, X_n)$ be independent sample variables of a random variable X. Find a maximum likelihood estimator of the respective parameter of the distribution under the assumption that

a) X is geometrically distributed with parameter $p \in]0,1[$,

b) X is exponentially distributed with parameter $\lambda > 0$.

For both cases, additionally analyze whether or not the estimator is unbiased.

Solution:

a) A single sample variable X_i which is geometrically distributed with parameter $p \in [0,1]$ has the density

$$f_{X_i}(x_i) = \begin{cases} p \cdot (1-p)^{x_i - 1} & \text{if } x_i \in \mathbb{N} \\ 0 & \text{else} \end{cases} .$$

Let $x_1, \ldots x_n \in \mathbb{N}$ be the actual samples (in contrast to the sample variables). Then, the likelihood function is defined by

$$L(x_1, \ldots, x_n; p) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} p \cdot (1-p)^{x_i - 1} = p^n \prod_{i=1}^{n} (1-p)^{x_i - 1} .$$

In the case that $x_i = 1$ for all $1 \leq i \leq n$ holds, the likelihood function equals to p^n and hence does not have a maximum in the open interval $]0,1[$. It follows that there exists no maximum likelihood estimator then.
Otherwise, we consider the natural logarithm - a strictly monotonously increasing function - of the likelihood function

$$\ln\left(L(x_1, \ldots, x_n; p)\right) = \ln\left(p^n \prod_{i=1}^{n} (1-p)^{x_i - 1}\right) = n \cdot \ln(p) + \ln(1-p) \cdot \sum_{i=1}^{n} (x_i - 1) . \checkmark$$

Define $g(p) = n \cdot \ln(p) + \ln(1-p) \cdot \sum_{i=1}^n (x_i - 1)$. To find the maximum of g, we consider the first derivative

$$g'(p) = \frac{n}{p} - \frac{\sum_{i=1}^n (x_i - 1)}{1-p}.$$

Some algebraic transformation lead to the zero of $g(p)$:

$$\begin{aligned}
\frac{n}{p} &= \frac{\sum_{i=1}^n (x_i - 1)}{1-p} \\
\Leftrightarrow (1-p)n &= p \cdot \sum_{i=1}^n (x_i - 1) \\
\Leftrightarrow n &= p \cdot \left(n + \sum_{i=1}^n (x_i - 1)\right) \\
\Leftrightarrow n &= p \sum_{i=1}^n x_i \\
\Leftrightarrow p &= \frac{n}{\sum_{i=1}^n x_i}\,.
\end{aligned}$$

To prove that this indeed is a maximum, we calculate the second derivative:

$$g''(p) = -\frac{n}{p^2} - \frac{\sum_{i=1}^n (x_i - 1)}{(1-p)^2}.$$

Obviously, this expression is negative for all $0 < p < 1$ which means that in fact $Y = \frac{n}{\sum_{i=1}^n X_i}$ is the maximum likelihood estimator of the geometric distribution. ✓ However, Y is not an unbiased estimator which can be seen already in the case of $n = 1$. Here, $Y = 1/X_1$ where X_i is geometrically distributed with parameter p. We find that

$$\mathbb{E}(Y) = \mathbb{E}(1/X_1) = \sum_{x=1}^{\infty} \frac{1}{x} \cdot p \cdot (1-p)^{x-1} = p + \sum_{x=2}^{\infty} \frac{1}{x} \cdot p \cdot (1-p)^{x-1} > p\,,$$

where the inequality holds since $p \in \,]0,1[$ implies that the sum is greater than 0. ✓

b) As the sample variables X_i are identically exponentially distributed with parameter λ, they have the density

$$f_{X_i}(x_i) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x_i} & \text{if } x_i \geq 0 \\ 0 & \text{else} \end{cases}.$$

Again, let $x_1, \ldots, x_n \geq 0$ be the actual samples. The likelihood function is given by

$$L(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \lambda \cdot e^{-\lambda \cdot x_i} = \lambda^n \cdot e^{-\lambda \cdot \sum_{i=1}^n x_i}\,.$$

This expression is to be optimized. If $x_i = 0$ holds for all $i \geq i \geq n$ the likelihood function is $L(x_1, \ldots, x_n; \lambda) = \lambda^n \cdot \exp -\lambda \cdot 0 = \lambda^n$. In this case, L does not have a maximum on the open interval $]0, \infty[$ and thus there cannot exist a maximum likelihood estimator by definition. However, this case occurs with probability 0 so

we may restrict our further considerations to the case that at least one $x_i > 0$. Then, it holds that $\sum_{i=1}^n x_i > 0$.
For convenience, we again use the natural logarithm of the likelihood function:

$$\ln\left(L(x_1,\ldots,x_n;\lambda)\right) = \ln\left(\lambda^n \cdot e^{-\lambda\cdot\sum_{i=1}^n x_i}\right) = n\cdot\ln(\lambda) - \lambda\sum_{i=1}^n x_i. \checkmark$$

Define the function $g(\lambda) = n\cdot\ln(\lambda) - \lambda\sum_{i=1}^n x_i$. Its first derivative is

$$g'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i\,.$$

Obviously, g has exactly one extremum, namely $\lambda = n/\sum_{i=1}^n x_i$. Further, the second derivative of g

$$g'{,}'(\lambda) = -\frac{n}{\lambda^2}$$

is always negative. We hence conclude that $n/\sum_{i=1}^n x_i$ in fact is a maximum. Since g is differentiable for all $\lambda > 0$ this maximum is global. Hence, $Y = n/\sum_{i=1}^n X_i$ is a maximum likelihood estimator for λ. ✓
To check whether or not the estimator Y is unbiased we again consider the case of a single sample, $n = 1$. We further make use of Jensen's inequality that tells that for an integrable real-valued random variable X and a convex function ϕ the relation

$$\mathbb{E}\left(\phi(X)\right) \geq \phi(\mathbb{E}(X))$$

holds where we (basically) have the equality sign only in the case that ϕ is a linear function. Considering now $\phi(x) = \frac{1}{x}$ being strictly convex over $\mathbb{R}^+$ we find for the expected value of the estimator Y

$$\mathbb{E}(Y) = \mathbb{E}\left(\frac{1}{X_1}\right) > \frac{1}{\mathbb{E}(X_1)} = \frac{1}{\frac{1}{\lambda}} = \lambda.$$

So, Y is not unbiased for λ. ✓
(At this point it might be helpful to recall that generally

$$\mathbb{E}\left(\frac{1}{X}\right) \neq \frac{1}{\mathbb{E}(X)},$$

i.e. the expected value of the reciprocal of a random variable is not equal to the reciprocal of the expected value of this random variable.)

Problem H 41 - Yarn Factories [3 pts.]

The strength of yarn produced in two different factories is to be tested. To do so, the two random samples of the maximum force (the force at tear) in kilonewtons (kN) were measured:

Factory A	150	149	152	146	146	149	148	152
Factory B	152	154	151	152	149	149	150	153

.

Find the confidence interval for the difference of the expected values $\mu_1 - \mu_2$ on the level of 90% under the assumption that the maximum force in the entire production batches is normally distributed with the same variance for both factories.

Solution: For our convention, we define the two random variables X_1 and X_2 for the strengths of a yarn from factory A and B, respectively, that are distributed $\mathcal{N}(\mu_1, \sigma)$ and $\mathcal{N}(\mu_2, \sigma)$ with equal but unknown σ. By simple calculations, we can determine the sample means and sample variances:

$$\bar{X}_1 = \frac{1}{8} \cdot \sum_{i=1}^{8} x_{1,i} = 149.0,\ \bar{X}_2 = 151.25,$$

$$S_1^2 = \frac{1}{8-1} \cdot \sum_{i=1}^{8} (x_{1,i} - \bar{X}_1)^2 = 5.429,\ S_2^2 = 3.357.$$

The confidence level is $1 - \alpha = 0.9$, so $1 - \frac{\alpha}{2} = 0.95$. From a table of the critical values of the t-distribution we find that the required quantile $t_{N,\alpha}$ of a t-distribution with $N = 2 \cdot (8-1) = 14$ degrees of freedom is $t_{14,0.95} = 1.716$. ✓

The test statistic of this scenario is

$$\begin{aligned} T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \sqrt{\frac{n_1+n_2-2}{\frac{1}{n_1}+\frac{1}{n_2}}} \\ = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{n-1} \cdot \sqrt{S_1^2+S_2^2}} \sqrt{\frac{2n-2}{\frac{2}{n}}} \\ = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2+S_2^2}} \sqrt{n} = \left(\bar{X}_1 - \bar{X}_2\right) \cdot \sqrt{\frac{n}{S_1^2+S_2^2}} \end{aligned}$$

where it was used that the two sample sizes n_1 and n_2 are equal to n. Hence, the confidence interval for the difference $\mu_1 - \mu_2$ on the level of 90% is given by

$$\begin{aligned} &\left[\mu_1 - \mu_2 - t_{N,1-\alpha/2} \cdot \sqrt{\frac{S_1^2+S_2^2}{n}}; \mu_1 - \mu_2 + t_{N,1-\alpha/2} \cdot \sqrt{\frac{S_1^2+S_2^2}{n}}\right] \\ \approx &\left[-2.25 - 1.716 \cdot \sqrt{\frac{5.429+3.357}{8}}; -2.25 + 1.716 \cdot \sqrt{\frac{5.429+3.357}{8}}\right] \\ &\approx [-4.1; -0.4] \end{aligned}$$ ✓

which means that $\mu_1 - \mu_2$ falls in $[-4.1; -0.4]$ with 90% probability. In particular, $\mu_1 - \mu_2 = 0$ is not in this interval, so the case $\mu_1 = \mu_2$ can be excluded on this level of confidence. Furthermore, the interval $[-4.1; -0.4]$ contains only negative numbers, so $\mu_1 - \mu_2 < 0$ and thus $\mu_1 < \mu_2$ holds by (at least) 90% probability. In conclusion, we can say that there is high probability that factory B produces better yarn than factory A. ✓

Problem H 42 - Estimating the Amount of Damage **[6 pts.]**

An insurance company tries to model the amount of damage of an insured accidence by means of the density

$$\tilde{\rho_\vartheta}(x) = \frac{1}{\vartheta^2} x e^{-x/\vartheta}, x \in \,]0, \infty[$$

for $x \in \mathbb{R}^+$. Here, $\vartheta > 0$ is an unknown parameter to be estimated based on the amount of damage of all reported accidents during the last year. For the sake of simplicity we may assume that the n reported values of the amount of damage $\vec{X} = (x_1, x_2, \ldots, x_n)$ are independent from each other.

a) Define an appropriate statistical model of the present situation in terms of a probability space for $\vec{X} \in \mathbb{R}^n_+$.

b) Show that the maximum likelihood estimator of ϑ is given by

$$\hat{\vartheta}(x_1, \ldots, x_n) = \frac{1}{2n} \sum_{i=1}^{n} x_i.$$

c) Prove or disprove that the estimator $\hat{\vartheta}$ is unbiased.

d) Suppose that $\hat{\vartheta}$ is unbiased. Describe how, in theory, to decide if $\hat{\vartheta}$ is the most efficient unbiased estimator for ϑ.

Solution:

a) The probability space can be defined by $(\mathbb{R}^n_+, \mathcal{B}(\mathbb{R}^n_+), Pr_\vartheta : \vartheta > 0)$ (i.e. the n-dimensional, positive real space, the corresponding Borel sets and a probability measure) where the probability measure Pr_ϑ has the density

$$\rho_\vartheta(x_1, \ldots, x_n) = \prod_{i=1}^{n} \tilde{\rho_\vartheta}(x_i) = \frac{1}{\vartheta^{2n}} \prod_{i=1}^{n} (x_i e^{-x_i/\vartheta}). \checkmark$$

b) To find the maximum of the likelihood function $L(x_1, \ldots, x_n; \vartheta) = \rho_\vartheta(\vec{X})$ we consider its logarithm:

$$\ln\left(L(x_1, \ldots, x_n; \vartheta)\right) = -2n \ln(\vartheta) + \sum_{i=1}^{n} \ln(x_i) - \sum_{i=1}^{n} \frac{x_i}{\vartheta}. \checkmark$$

At an extremum it has to hold

$$\frac{d}{d\vartheta} \ln\left(L(x_1, \ldots, x_n; \vartheta)\right) = -\frac{2n}{\vartheta} + \sum_{i=1}^{n} \frac{x_i}{\vartheta^2} = 0.$$

Since $\vartheta > 0$ it follows that the first derivative is zero at

$$\vartheta = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{\bar{X}}{2}. \checkmark$$

By means of the second derivative of the log-likelihood function

$$\frac{d^2}{d\vartheta^2} \ln\left(L(x_1, \ldots, x_n; \vartheta)\right) = \frac{2n}{\vartheta^2} - \sum_{i=1}^{n} \frac{2x_i}{\vartheta^3}$$

which, in this case, takes the value

$$\frac{8n}{\bar{X}^2} - \frac{2n \cdot \bar{X}}{\frac{\bar{X}^3}{8}} = -\frac{8n}{\bar{X}^2}$$

being negative since $x_i \in \mathbb{R}^+$ and thus also $\bar{X} > 0$, we see that $\hat{\vartheta} = \frac{\bar{X}}{2}$ is the desired maximum likelihood estimator. ✓
(An elegant alternative is

$$\frac{d^2}{d\vartheta^2} \ln\left(L(x_1, \ldots, x_n; \vartheta)\right) = -\frac{1}{\vartheta}\left(\frac{d}{d\vartheta} L(x_1, \ldots, x_n; \vartheta)\right) - \sum_{i=1}^{n} \frac{x_i}{\vartheta^3} = 0 - \sum_{i=1}^{n} \frac{x_i}{\vartheta^3} < 0$$

that makes use of the fact that the first derivative is zero for this choice of ϑ.)

c) To answer whether or not $\hat{\vartheta}$ is unbiased we consider its expected value

$$\mathbb{E}(\hat{\vartheta}) = \mathbb{E}\left(\frac{\sum_{i=1}^{n} X_i}{2n}\right) = \frac{1}{2n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{\mathbb{E}(X_i)}{2},$$

keeping in mind that all sample variables X_i are independent and identically distributed. The required expected value $\mathbb{E}(X_i)$ can be found by integration

$$\int_0^\infty x \tilde{\rho}_\vartheta(x) dx = \int_0^\infty \frac{1}{\vartheta^2} x^2 e^{-x/\vartheta} dx$$
$$= \frac{1}{\vartheta^2}\left[e^{-x/\vartheta}\left(-\vartheta x^2 - 2\vartheta x - 2\vartheta^3\right)\right]_0^\infty = \frac{1}{\vartheta^2}\left(0 + 2\vartheta^3\right) = 2\vartheta.$$

So, $\mathbb{E}(\hat{\vartheta}) = \vartheta$ which means that the estimator $\hat{\vartheta}$ is unbiased. ✓

d) Since we already know that the estimator $\hat{\vartheta}$ is unbiased, we know (by the lecture) that

$$MSE(\hat{\vartheta}) = Var(\hat{\vartheta}).$$

To prove that $\hat{\vartheta}$ is the best or, in other words, most efficient unbiased estimator we hence should determine its variance and prove that it is minimum among all other unbiased estimators. ✓
More specifically, here we may transform the likelihood function by

$$L(x_1, \ldots, x_n; \vartheta) = \frac{1}{\vartheta^{2n}} \prod_{i=1}^{n} (x_i e^{-x_i/\vartheta}) = \left(\prod_{i=1}^{n} x_i\right) e^{-\frac{1}{\vartheta}\sum_{i=1}^{n} x_i - 2n\ln(\vartheta)}$$
$$= \left(\prod_{i=1}^{n} x_i\right) e^{-\frac{2n}{\vartheta}\cdot\frac{1}{2n}\sum_{i=1}^{n} x_i - 2n\ln(\vartheta)} = \left(\prod_{i=1}^{n} x_i\right) e^{-\frac{2n}{\vartheta}T(x) - 2n\ln(\vartheta)}$$

which means that the likelihood function takes the form of a so-called exponential model with respect of $T(x) = \frac{1}{2n}\sum_{i=1}^{n} x_i = \hat{\vartheta}(x)$. Although not explicitly mentioned in the lecture this implies that $\hat{\vartheta} = T$ is a *best* unbiased estimator for $\tau(\vartheta) = \mathbb{E}(\hat{\vartheta}) = \vartheta$.

Problem H 43 - The Taxicab Fleet **[4 pts.]**

On a research trip to a foreign city Professor Günther five times has to take a taxicab. Each time he is assigned to one of the n caps of the taxi fleet of the city independently and by the same probability. Although Professor Günther does not know the exact parameter n he was told that all cabs are labeled with unique identification numbers between 1 and n. Let X be the largest identification number which Professor Günther sees on the five cabs of his drives. To estimate the size of the fleet n he uses the estimators $U_1 = X$ and $U_2 = \lceil c \cdot X \rceil$ where $c > 0$ is a real-valued constant.

a) Find a parameter c as small as possible such that $n \in [U_1, U_2]$ holds by a confidence level of 0.9.

b) Assume that 100 is the largest identification number Professor Günther sees on his cabs. Which value of n is to be expected according to the first part of the exercise?

Solution:

a) Denote be $X_1, \ldots, X_5$ the identification numbers Professor Günther observes during his trip and let X the maximum of those five numbers.
Trivially, $n \geq X$ in any case. Hence, it is sufficient to find a value for c in the upper bound $U_2 = \lceil c \cdot X \rceil$ that is as small as possible and, at the same time, satisfies $Pr(n \leq U_2) \geq 9/10$. First, simple transformations yield

$$Pr(n \leq U_2) = Pr(n \leq \lceil c \cdot X \rceil) \geq Pr(n \leq c \cdot X) = 1 - Pr(n > c \cdot X).$$

To have that the event "$n \leq U_2$" occurs by a probability of 9/10 or higher $Pr(n > c \cdot X)$ has to be equal to 1/10 at most. By further transformations we estimate this probability by

$$Pr(n > c \cdot X) \leq Pr(n \geq c \cdot X) = Pr\left(X \leq \frac{n}{c}\right) = Pr\left(X \leq \lfloor \frac{n}{c} \rfloor\right). \checkmark$$

The last equality is true because X is an integer number.
Let us now consider the cumulative distribution function of X. Since the random variables are independent and uniformly distributed, the probability that X takes a value smaller than some given natural number $m \leq n$ can be expressed as

$$Pr(X \leq m) = Pr(X_1 \leq m, \ldots, X_5 \leq m) = \prod_{i=1}^{5} Pr(X_i \leq m) = \prod_{i=1}^{5} \frac{m}{n} = \left(\frac{m}{n}\right)^5.$$

From this it follows that

$$Pr(c \cdot X < n) \leq \left(\frac{\lfloor n/c \rfloor}{n}\right)^5 \leq \left(\frac{n/c}{n}\right)^5 = \frac{1}{c^5}. \checkmark$$

In total, this means that we have to choose a value of c as small as possible that satisfies $(1/c^5) \leq 1/10$ or, equivalently, $c \geq 10^{1/5}$. Clearly, we set c to $10^{1/5} \approx 1.58$ and, by this, obtain the upper bound $U_2 = \left\lceil 10^{1/5} \cdot X \right\rceil$. $\checkmark$

b) In the case that the largest number seen by Professor Günther is 100, the upper bound takes the value $\lceil 10^{1/5} \cdot 100 \rceil = 159$. This means that he can be sure on the confidence level of 0.9 that there are at most 159 taxi caps in the city's fleet. ✓

Problem H 44 - Smokers' babies **[4 pts.]**

It is speculated that one may tell by the appearance of a baby whether or not its mother is a smoker. To test this hypothesis, the photos of 1600 babies, exactly half of them children of a smoking mother, are presented to an experienced midwife to be classified.
How many correct answers of the midwife do you expect if you like to exclude that the result was achieved by pure guessing by 99% confidence? You may use the normal approximation ($\Phi(2.33) \approx 0.99$).

Solution: Denote by $\vartheta \in [0,1[$ the hit ratio of the midwife. Since exactly half of the photos show babies of a smoking mother the midwife's detection power is better than pure guessing if $\vartheta > 0.5$. Thus, we test the null hypothesis H_0 that $\vartheta \in [0, 0.5]$, corresponding to pure guessing, versus H_1 that $\vartheta \in]0.5, 1]$. Further define the random variable $X \in \{0, \ldots, 1600\}$ the number of correct answers. By this the statistical model behind the task can be defined by the sample space $\Xi = \{0, \ldots, 1600\}$, the sigma algebra $\mathcal{P}(\Xi)$ and a probability measure Pr_ϑ that has the density of a binomial distribution $B_{n,\vartheta}$ with $n = 1600$. ✓
We can formulate the test by

$$\phi : \begin{cases} 0 & \text{if } X \leq c \quad (\text{if } H_0) \\ 1 & \text{if } X > c \quad (\text{if } H_1) \end{cases}$$

where c is the critical value of correct answers that is to be determined. It is required that H_0 can be excluded correctly by 99%, i.e. we would like to erroneously reject H_0 by 1%. This means that we aim at a confidence level of $\alpha = 0.01$, corresponding to the probability of the type-I error, so we require that

$$\sup_{\vartheta \in [0,0.5]} \mathbb{E}_\vartheta(\phi) \leq \alpha = 0.01. ✓$$

It is easy to see that the expected value of the test is

$$\mathbb{E}_\vartheta(\phi) = Pr(X > c) = B_{n,\vartheta}(X > c) = 1 - B_{n,\vartheta}(X \leq c - 1).$$

$B_{n,\vartheta}(X \leq c-1)$ is minimal over $[0, 0.5]$ at $\vartheta_0 = 0.5$ and thus $\mathbb{E}_\vartheta(\phi)$ has its maximum value in this case. If we additionally approximate the binomial distribution by the normal distribution we get

$$\alpha \geq \mathbb{E}_{\vartheta_0(\phi)} = 1 - B_{n,\vartheta_0}(X \leq c - 1) \approx 1 - \Phi\left(\frac{c - 1 - n\vartheta_0}{\sqrt{n\vartheta_0(1-\vartheta_0)}}\right). ✓$$

In the critical case we have

$$
\begin{aligned}
1-\alpha &= \Phi\left(\frac{c-1-n\vartheta_0}{\sqrt{n\vartheta_0(1-\vartheta_0)}}\right) \\
\Leftrightarrow \quad \frac{c-1-n\vartheta_0}{\sqrt{n\vartheta_0(1-\vartheta_0)}} &= \Phi^{-1}(1-\alpha)\,, \\
\text{i.e.} \quad \frac{c-1-1600\cdot 0.5}{\sqrt{1600\cdot 0.5\cdot 0.5}} &= \Phi^{-1}(0.99) \approx 2.33 \\
\Rightarrow \quad \frac{c-801}{\sqrt{400}} &\approx 2.33 \\
\Rightarrow c \approx 2.33\cdot 20 + 801 &= 847.6.
\end{aligned}
$$

So, the final answer is that the midwife has to give at least 848 correct answers such that we can exclude by more then 99% that this results from simple guessing. ✓