Statistical Inference 1 Winter Term 2024/2025

Lecture: Prof. Dr. Thomas Augustin, Prof. Dr. Christian Heumann
Exercises: Sergio Buttazzo, Jan Anders
Tutorial, Corrections: Sophie Hopp, Michael Speckbacher **Suggested Solution**

Exercise Sheet 3

Exercise 3.1 - Kullback Leibler Divergence

A large supermarket chain starts counting the time (in minutes) between customers arriving at one of their stores. After 6 people, they obtained the following sample: (0.8, 0.5, 1.2, 0.6, 1.6). The count is subsequently stopped for unknown reasons, but the manager still tasks you to come up with a reasonable model for the distribution of the time interval between customers arriving with the little data they gathered.

(a) Assuming the customers each arrive on their own and don't have any other relationship with each other, what is a reasonable modelling assumption? State the assumed distribution and estimate the relative parameter(s) via maximum likelihood. **Note:** No need to derive the formula of the ML estimator here.

We know: Customers arrive at a constant rate and independently of each other. Modelling assumption: The time between customers arriving is exponentially distributed (See also Statistics II) and all samples are independent and identically distributed. The exponential distribution only has one parameter λ, which describes the rate at which customers arrive. The inverse of the rate $\frac{1}{\lambda}$ is it's expected value. The MLE for λ is the inverse sample mean, or $\widehat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}$.

We get:

X: Time between each customer in minutes, $X \overset{i.i.d.}{\sim} Exp(\frac{1}{\bar{x}})$

Also thinkable: Weibull Distribution as a generalization of exponential (disadvantage: We have to estimate two parameters) or Gamma distribution (though gamma would rather be used in case we were modelling the time for serving each customer).

(b) Your manager tells you of measurements in other stores of the chain investigating the same task. One research team reported that the rate at which customers arrived was 1.5 per minute, another reported 1 customer per minute on average. Use the Kullback-Leibler divergence to find out which of these two stores is more similar to the initial store in terms of customer arrival times. To do so, start by deriving a function that allows for fast computation of the KL-divergence between two models with the same distributional assumption but different parameters.

$$
\begin{aligned}
\mathcal{KL}(f_\theta, f_\phi) = &= \int f_\theta(x)\log\frac{f_\theta(x)}{f_\phi(x)}dx \\
&= \int \theta\exp(-\theta x)\log\frac{\theta\exp(-\theta x)}{\phi\exp(-\phi x)}dx \\
&= \int \theta\exp(-\theta x)\left[\log(\theta)-\log(\phi)-\theta x+\phi x\right]dx \\
&= \int \theta\exp(-\theta x)\left[\log(\theta)-\log(\phi)\right]+\theta\exp(-\theta x)x\left[-\theta+\phi\right]dx \\
&= (\log\theta-\log\phi)\underbrace{\int\theta\exp(-\theta x)dx}_{1 \text{ by } \int f_\theta=1} + (\phi-\theta)\underbrace{\int_D \theta\exp(-\theta x)x\,dx}_{\mathbb{E}[x]=\theta^{-1} \text{ if } x\sim f_\theta} \\
&= \log\theta-\log\phi+\frac{\phi-\theta}{\theta} \\
&= \log\theta-\log\phi+\frac{\phi}{\theta}-1
\end{aligned}
$$

Our (assumed) true model: $X_\theta \sim Exp(\widehat{\theta})$ with $\widehat{\theta} = \frac{n}{\sum_{i=1}^n x_i} = 1.064$
Model 1: $X_{\phi 1} \sim Exp(1)$
Model 2: $X_{\phi 2} \sim Exp(1.5)$

$\mathcal{KL}(f_\theta, f_{\phi 1}) = log(\widehat{\theta}) - log(\phi_1) + \frac{\phi_1}{\widehat{\theta}} - 1 = log(1.064) - log(1) + \frac{1}{1.064} - 1 = 0.0019$
$\mathcal{KL}(f_\theta, f_{\phi 2}) = log(\widehat{\theta}) - log(\phi_2) + \frac{\phi_2}{\widehat{\theta}} - 1 = log(1.064) - log(1.5) + \frac{1.5}{1.064} - 1 = 0.066$
Calculating the divergence in the other order $\mathcal{KL}(f_\phi, f_\theta)$ also works.

Exercise 3.2 - MSE

Let $X_1, \dots, X_n$ be independently identically distributed with $X_i \sim Bin(1, \pi)$. Our goal is to estimate the parameter π given the sample. Four estimators for π are to be compared with each other:

$$
\begin{aligned}
T_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i, \\
T_2 &= \frac{2}{n}\sum_{i=1}^{n/2} X_i, \\
T_3 &= 0.5, \\
T_4 &= \left(\sum_{i=1}^{n} X_i + 1\right)/(n+2).
\end{aligned}
$$

(a) Determine the bias, variance and MSE for each estimator analytically.

- T1

$$
\mathbb{E}(T_1) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\cdot n\cdot\pi = \pi \quad \Rightarrow T_1 \text{ is unbiased, Bias}(T_1,\pi) = 0
$$

$$
\text{MSE}(T_1,\pi) \overset{T_1 \text{ unbiased}}{=} \text{Var}(T_1) \overset{X_i \text{ i.i.d.}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) = \frac{1}{n^2}\cdot n\cdot\pi\cdot(1-\pi) = \frac{\pi(1-\pi)}{n}
$$

- T2

$$\mathbb{E}(T_2) = \frac{2}{n}\sum_{i=1}^{n/2}\mathbb{E}(X_i) = \frac{2}{n}\cdot\frac{n}{2}\cdot\pi = \pi \quad \Rightarrow T_2 \text{ is unbiased, Bias}(T_2,\pi) = 0$$

$$\text{MSE}(T_2,\pi) \overset{T_2 \text{ unbiased}}{=} \text{Var}(T_2) \overset{X_i \text{ i.i.d.}}{=} \frac{4}{n^2}\sum_{i=1}^{n/2}\text{Var}(X_i) = \frac{4}{n^2}\cdot\frac{n}{2}\pi(1-\pi) = \frac{2\pi(1-\pi)}{n}$$

- T3

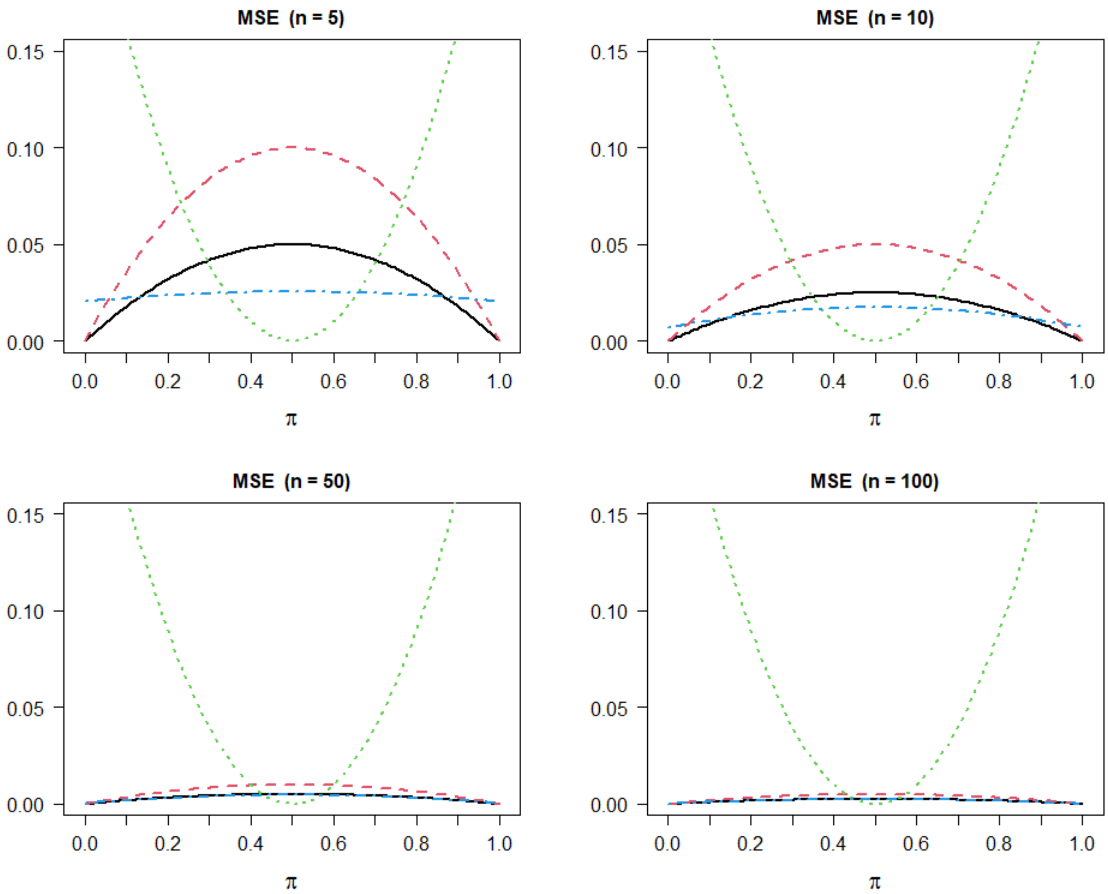
$$\begin{aligned}\mathbb{E}(T_3) &= \mathbb{E}(0.5) = 0.5 \neq \pi \quad \Rightarrow T_3 \text{ is biased, except for } \pi = 0.5.\\ \text{Bias}(T_3,\pi) &= 0.5 - \pi\\ \text{Var}(T_3) &= \text{Var}(0.5) = 0 \quad \Rightarrow MSE(T_3,\pi) = (\text{Bias}(T_3,\pi))^2 = (0.5-\pi)^2\end{aligned}$$

- T4

$$\begin{aligned}\mathbb{E}(T_4) &= \frac{1}{n+2}\left(\sum_{i=1}^{n}\mathbb{E}(X_i)+1\right) = \frac{n\pi+1}{n+2} \neq \pi \quad \Rightarrow T_4 \text{ is biased}\\ \text{Bias}(T_4,\pi) &= \frac{n\pi+1}{n+2} - \pi = \frac{1-2\pi}{n+2}\\ \text{Var}(T_4) &= \frac{\text{Var}\left(\sum_{i=1}^{n} X_i + 1\right)}{(n+2)^2} = \frac{\text{Var}\left(\sum_{i=1}^{n} X_i\right)}{(n+2)^2}\\ &\overset{X_i \text{ i.i.d.}}{=} \frac{\sum_{i=1}^{n}\text{Var}(X_i)}{(n+2)^2} = \frac{n\pi(1-\pi)}{(n+2)^2}\\ \Rightarrow \text{MSE}(T_4,\pi) &= \text{Var}(T_4) + (\text{Bias}(T_4,\pi))^2\\ &= \frac{n\pi(1-\pi)}{(n+2)^2} + \left[\frac{1-2\pi}{n+2}\right]^2\\ &= (n\pi - n\pi^2 + 1 - 4\pi + 4\pi^2)\cdot\frac{1}{(n+2)^2}\\ &= ((n-4)\pi(1-\pi)+1)\cdot\frac{1}{(n+2)^2}\end{aligned}$$

(b) In the figure below, the *MSE*s of the four estimators are drawn for different values of π and n. Which function belongs to which estimator and why?

(c) What do the results of the comparison of the different estimators mean for choosing an appropriate estimator?



$MSE(T_1)$: black (solid line)
$MSE(T_2)$: red (striped)
$MSE(T_3)$: green (dotted)
$MSE(T_4)$: blue (striped and dotted)

T_3 is generally unsuitable as an estimator. T_4 never reaches an MSE of zero, but has a small overall MSE compared to T_1 and T_2. If a prior range for π is known, for example, $\pi \in [0.4; 0.6]$, T_4 might be a reasonable choice. The MSE of T_2 behaves very similarly to that of T_1, but is always 2 times higher, since T_2 uses only the information from half the sample, so it is not a good choice. Overall, T1 is the best unbiased estimate to use if we have no prior information about π.

Exercise 3.3 - Asymptotic Behaviour of Estimates

Let $X_1, \ldots, X_n$ be i.i.d. normally distributed with $\mu = 0$ and unknown variance σ^2. Consider the following estimator for σ^2:

$$T_n = \frac{2}{n} X_1^2 + \frac{n-2}{n(n-1)} \sum_{i=2}^{n} X_i^2$$

(a) Is T_n an unbiased estimator for σ^2? *Reminder*: $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

$$\mathbb{E}(T_n) = \frac{2}{n} \mathbb{E}(X_1^2) + \frac{n-2}{n(n-1)} \sum_{i=2}^{n} \mathbb{E}\left(X_i^2\right)$$

We have

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \quad \Rightarrow \quad \mathbb{E}(X^2) = \text{Var}(X) + [\mathbb{E}(X)]^2$$

In this case, we have $\mathbb{E}(X) = \mu = 0$ and thus: $\mathbb{E}(X^2) = \text{Var}(X)$.

$$\begin{aligned} \Rightarrow \mathbb{E}(T_n) &= \frac{2}{n}\text{Var}(X_1) + \frac{n-2}{n(n-1)}\sum_{i=2}^{n}\text{Var}(X_i) \\ &= \frac{2}{n}\sigma^2 + \frac{n-2}{n(n-1)}(n-1)\sigma^2 \\ &= \left(\frac{2}{n} + \frac{n-2}{n}\right)\sigma^2 \\ &= \sigma^2 \end{aligned}$$

$\Rightarrow T_n$ is unbiased for σ^2.

(b) Determine the variance of T_n.

$$\begin{aligned} \text{Var}(T_n) &= \text{Var}\left(\frac{2}{n}X_1^2 + \frac{n-2}{n(n-1)}\sum_{i=2}^{n}X_i^2\right) \\ &= \frac{4}{n^2}\text{Var}(X_1^2) + \frac{(n-2)^2}{n^2(n-1)^2}\sum_{i=2}^{n}\text{Var}(X_i^2) \end{aligned}$$

To determine $Var(X_i^2)$, we know that $\frac{X_i-\mu}{\sigma} \sim N(0,1)$. Since $\mu = 0$, this simplifies our expression to $\frac{X_i}{\sigma} \sim N(0,1)$. When we square a standard normally distributed variable, the result follows a Chi-Square distribution with 1 degree of freedom. Therefore,

$$\left(\frac{X_i}{\sigma}\right)^2 = \frac{X_i^2}{\sigma^2} \sim \chi^2(1)$$

This tells us that $\frac{X_i^2}{\sigma^2}$ follows a Chi-Square distribution with 1 degree of freedom. Recall that if $Y \sim \chi^2(n)$, then $Var(Y) = 2n$. Thus, the variance of $\frac{X_i^2}{\sigma^2}$ is:

$$\text{Var}\left(\frac{X_i^2}{\sigma^2}\right) = 2 \Rightarrow \text{Var}(X_i^2) = 2\sigma^4$$

Plugging in:

$$\begin{aligned} \text{Var}(T_n) &= \frac{8}{n^2}\sigma^4 + \frac{(n-2)^2}{n^2(n-1)^2}\cdot(n-1)\cdot 2\cdot\sigma^4 \\ &= \left(\frac{8}{n^2} + \frac{2(n-2)^2}{n^2(n-1)}\right)\cdot\sigma^4 = \left(\frac{8n-8+2n^2-8n+8}{n^2(n-1)}\right)\cdot\sigma^4 \\ &= \frac{2}{n-1}\cdot\sigma^4 \end{aligned}$$

(c) Is T_n a consistent estimator for σ^2?

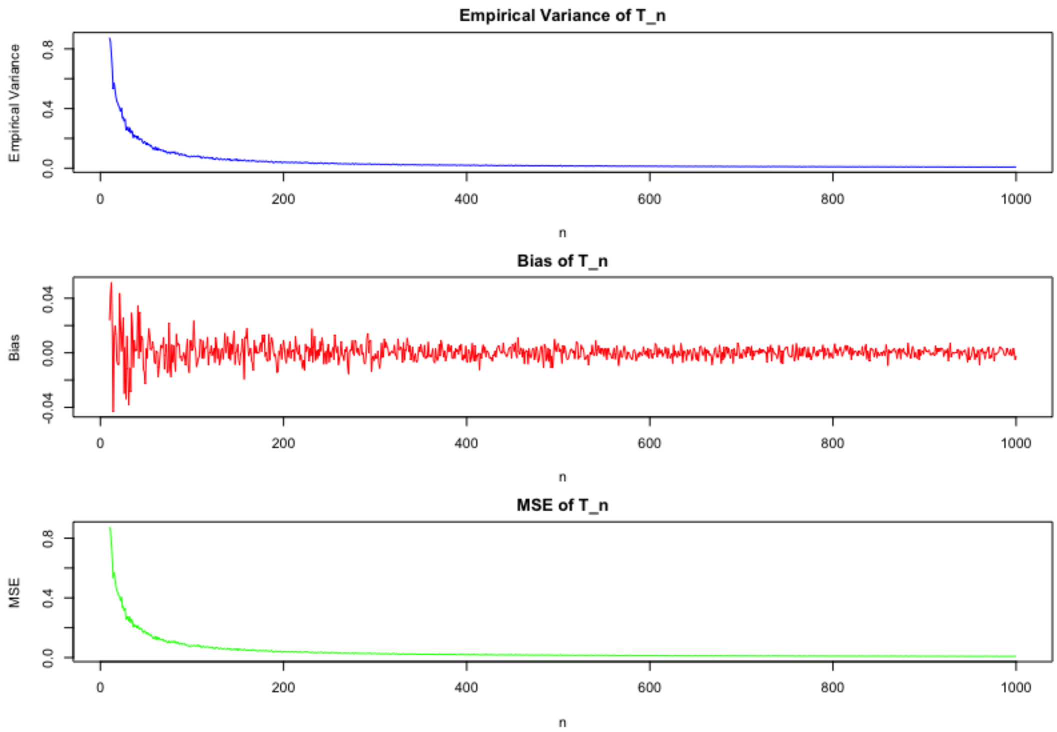
Since we already proved that T_n is unbiased we just need to acknowledge that $\text{Var}(T_n) = \frac{2}{n-1}\sigma^4$ converges to 0 for large n:

$$\lim_{n\to\infty}\text{Var}(T_n) = \lim_{n\to\infty}\frac{2}{n-1}\sigma^4 = 0$$

The variance of T_n shall now be calculated for different values of n, using the formula determined in (b) and a simulation study.

(d) In R, write a function that simulates a lot of random samples (e.g. 1000) of size n from a normal distribution with $\mu = 0$ and $\sigma^2 = 2$ and calculates the estimate $\hat{T}_n$ for each sample. Now, determine the empirical Variance, Bias and MSE of these estimators $\hat{T}_n$. Plot the resulting curves for $n \in [10, 1000]$.

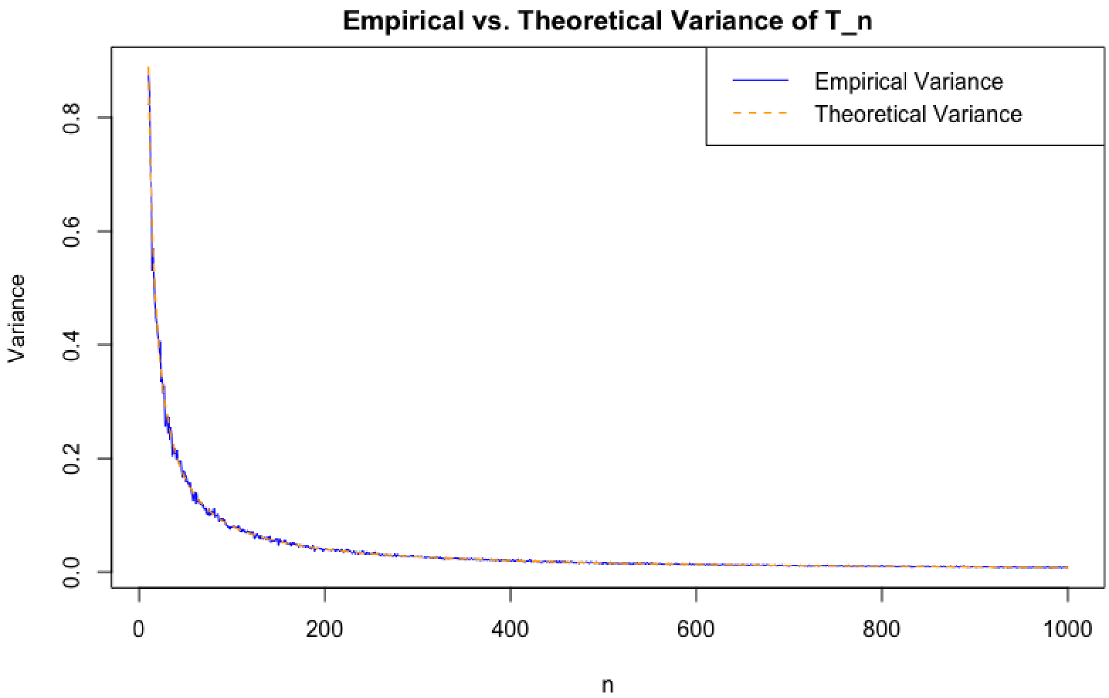
See `Ex_3_3.R`.



We can see that the variance of T_n is reduced as a function of n. The bias is distributed around 0, confirming our result that T_n is unbiased. However, the large variance has an influence on the empirical bias as well. Overall, the bias with it's low range in comparison to the variance has a negligible influence on the MSE.

(e) Compare the results obtained in the simulation study in R to the variance we would theoretically expect using the formula from (b) with $\sigma^2 = 2$.

See `Ex_3_3.R`.



We observe that the estimate for the variance of $\hat{T}_n$, using 1000 simulated samples to generate the variance estimate for each n, basically perfectly matches the theoretical variance of the estimate.