Exercise Sheet 2: Data Science Methods

Technische Universität München

Winter semester 2024-2025

Instructions

You can hand in your solutions during the next tutorial on 21/11/24 or submit them earlier at the group office 1269 in the Physics Department. For code, you can send it to `alan.zander@tum.de`.

1 Exercise: Expectation values and variances

Determine the expectation value and the variance of the following distributions:

1.1 Bernoulli, with probability mass function

$$f(k;p) = p^k\,(1-p)^{1-k}\,, \tag{1}$$

where $k \in \{0,1\}$ and $p \in [0,1]$.

Solution:

$$\mathbb{E}[X] = \sum_{k\in\{0,1\}} k f(k;p) = f(1;p) = p.$$

$$\mathbb{E}[X^2] = \sum_{k\in\{0,1\}} k^2 f(k;p) = f(1;p) = p,$$

$$\Rightarrow \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p\,(1-p)\,.$$

1.2 Poisson, with probability mass function

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \tag{2}$$

where $k \in \mathbb{N}_0$ and $\lambda > 0$.

Solution:

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k f(k;\lambda) = \sum_{k=1}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\lambda \underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{=e^{\lambda}} = \lambda.$$

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 f(k;\lambda) = \sum_{k=1}^{\infty} k^2\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{(k-1)!} =$$

$$e^{-\lambda}\lambda\left(\sum_{k=1}^{\infty}(k-1)\,\frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\right) = e^{-\lambda}\lambda\left(\underbrace{\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}}_{e^{\lambda}\mathbb{E}[X]} + \underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{e^{\lambda}}\right) = \lambda\,(\lambda+1)\,,$$

$$\Rightarrow \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda.$$

1.3 Normal (Gaussian), with probability density function

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3}$$

where $x, \mu \in \mathbb{R}$ and $\sigma > 0$.

Solution:

$$\mathbb{E}[X] = \int_{\mathbb{R}} \mathrm{d}x x f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \mathrm{d}x x e^{-\frac{(x-\mu)^2}{2\sigma^2}} \overset{y\equiv x-\mu}{=} \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \mathrm{d}y\,(y+\mu)\,e^{-\frac{y^2}{2\sigma^2}} =$$

$$\frac{1}{\sqrt{2\pi}\sigma}\underbrace{\int_{\mathbb{R}} \mathrm{d}y y e^{-\frac{y^2}{2\sigma^2}}}_{=0} + \mu \underbrace{\frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \mathrm{d}y e^{-\frac{y^2}{2\sigma^2}}}_{=1} = \mu.$$

The integral over symmetric intervals of continuous odd functions vanishes, so the first term is zero. Now, the variance can be also written as (this is the actual definition)

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \int_{\mathbb{R}} \mathrm{d}x\,(x-\mu)^2\, f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \mathrm{d}y y^2 e^{-\frac{y^2}{2\sigma^2}}.$$

This is of the form

$$\int_{\mathbb{R}} \mathrm{d}x x^2 e^{-ax^2} \overset{\text{int. by parts}}{=} -\frac{1}{2a}\underbrace{x e^{-ax^2}\Big|_{-\infty}^{\infty}}_{=0} + \frac{1}{2a}\underbrace{\int_{\mathbb{R}} \mathrm{d}x e^{-ax^2}}_{=\sqrt{\frac{\pi}{a}}},$$

where the last term is just a Gaussian integral. Hence,

$$\mathrm{Var}(X) = \sigma^2.$$

1.4 Exponential, with probability density function

$$f(x;\tau) = \frac{e^{-x/\tau}}{\tau}, \tag{4}$$

where $x \geq 0$ and $\tau > 0$.

Solution:

$$\mathbb{E}[X] = \int_0^\infty \mathrm{d}x x f(x;\tau) = \frac{1}{\tau}\int_0^\infty \mathrm{d}x x e^{-x/\tau} \overset{\text{int. by parts}}{=} \underbrace{-x e^{-x/\tau}\Big|_0^\infty}_{=0} + \tau\underbrace{\int_0^\infty \mathrm{d}x \frac{e^{-x/\tau}}{\tau}}_{=1} = \tau$$

$$\mathbb{E}[X^2] = \int_0^\infty \mathrm{d}x x^2 f(x;\tau) = \frac{1}{\tau}\int_0^\infty \mathrm{d}x x^2 e^{-x/\tau} \overset{\text{int. by parts}}{=} \underbrace{-x^2 e^{-x/\tau}\Big|_0^\infty}_{=0} + 2\tau\underbrace{\int_0^\infty \mathrm{d}x x \frac{e^{-x/\tau}}{\tau}}_{=\mathbb{E}[X]} = 2\tau^2,$$

$$\Rightarrow \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \tau^2.$$

1.5 Uniform, with probability density function

$$f(x;a,b) = \frac{1}{b-a}, \tag{5}$$

where $x \in [a,b]$ and $a \neq b$.

Solution:

$$\mathbb{E}[X] = \int_a^b \mathrm{d}x x f(x;a,b) = \int_a^b \mathrm{d}x x \frac{1}{b-a} = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{a+b}{2}.$$

$$\mathbb{E}[X^2] = \int_a^b \mathrm{d}x x^2 f(x;a,b) = \int_a^b \mathrm{d}x x^2 \frac{1}{b-a} = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{1}{3}\left(a^2+ab+b^2\right),$$

$$\Rightarrow \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{12}\left(b-a\right)^2.$$

2 Exercise: MLE and Fisher Information

Determine the Maximum Likelihood Estimator (MLE) for the following parameters:

2.1 Bernoulli: $\hat{p}_{\mathrm{MLE}}$.

Solution: Let $\vec{k} = (k_1, k_2, \ldots, k_n)$ be an experiment of n independent observations. Then, the log-likelihood function, which we want to maximize, reads

$$\log L(p;\vec{k}) = \log f(\vec{k}|p) = \sum_{i=1}^{n} \log f(k_i;p) = \sum_{i=1}^{n} \left[k_i \log p + (1-k_i)\log(1-p)\right].$$

$$\Rightarrow \frac{\partial \log L}{\partial p} = \sum_{i=1}^{n}\left[\frac{k_i}{p} - \frac{1-k_i}{1-p}\right] = \frac{1}{p\,(1-p)}\sum_{i=1}^{n}\left[k_i - p\right] = \frac{n}{p\,(1-p)}\left[\overline{k} - p\right].$$

Maximizing, $\left.\frac{\partial \log L}{\partial p}\right|_{\hat{p}_{\mathrm{MLE}}} \overset{!}{=} 0$, directly yields

$$\hat{p}_{\mathrm{MLE}} = \overline{k} = \frac{1}{n}\sum_{i=1}^{n} k_i.$$

2.2 Poisson: $\hat{\lambda}_{\mathrm{MLE}}$.

Solution: Let $\vec{k} = (k_1, k_2, \ldots, k_n)$ be an experiment of n independent observations. Then, the log-likelihood function, which we want to maximize, reads

$$\log L(\lambda;\vec{k}) = \log f(\vec{k}|\lambda) = \sum_{i=1}^{n} \log f(k_i;\lambda) = \sum_{i=1}^{n} \left[k_i \log\lambda - \lambda - \log k_i!\right].$$

$$\Rightarrow \frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^{n}\left[\frac{k_i}{\lambda} - 1\right] = \frac{n}{\lambda}\left[\overline{k} - \lambda\right].$$

Maximizing, $\left.\frac{\partial \log L}{\partial \lambda}\right|_{\hat{\lambda}_{\mathrm{MLE}}} \overset{!}{=} 0$, directly yields

$$\hat{\lambda}_{\mathrm{MLE}} = \overline{k} = \frac{1}{n}\sum_{i=1}^{n} k_i.$$

2.3 Normal (Gaussian): $\hat{\mu}_{\text{MLE}}$, $\hat{\sigma}^2_{\text{MLE}}$.

Solution: Let $\vec{x} = (x_1, x_2, \dots, x_n)$ be an experiment of n independent observations. Then, the log-likelihood function, which we want to maximize, reads

$$\log L(\mu, \sigma^2; \vec{x}) = \log f(\vec{x}|\mu, \sigma^2) = \sum_{i=1}^{n} \log f(x_i; \mu, \sigma^2) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log\left(2\pi\sigma^2\right) - \frac{(x_i - \mu)^2}{2\sigma^2} \right].$$

$$\Rightarrow \frac{\partial \log L}{\partial \mu} = -\sum_{i=1}^{n} \left[\frac{(x_i - \mu)}{\sigma^2} \right] = \frac{n}{\sigma^2} \left[\overline{x} - \mu \right],$$

$$\frac{\partial \log L}{\partial \sigma^2} = \sum_{i=1}^{n} \left[-\frac{1}{2}\frac{1}{\sigma^2} + \frac{(x_i - \mu)^2}{2\sigma^4} \right] = \frac{n}{2\sigma^4} \left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 - \sigma^2 \right].$$

Maximizing, $\left.\frac{\partial \log L}{\partial \mu}\right|_{\hat{\mu}_{\text{MLE}}} \overset{!}{=} 0$ and $\left.\frac{\partial \log L}{\partial \sigma^2}\right|_{\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}}} \overset{!}{=} 0$, directly yields, again

$$\hat{\mu}_{\text{MLE}} = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \hat{\mu}_{\text{MLE}}\right)^2.$$

2.4 Exponential: $\hat{\tau}_{\text{MLE}}$

Solution: Let $\vec{x} = (x_1, x_2, \dots, x_n)$ be an experiment of n independent observations. Then, the log-likelihood function, which we want to maximize, reads

$$\log L(\tau; \vec{x}) = \log f(\vec{x}|\tau) = \sum_{i=1}^{n} \log f(x_i; \tau) = \sum_{i=1}^{n} \left[-\log \tau - \frac{x_i}{\tau} \right].$$

$$\Rightarrow \frac{\partial \log L}{\partial \tau} = \sum_{i=1}^{n} \left[\frac{x_i}{\tau^2} - \frac{1}{\tau} \right] = \frac{n}{\tau^2} \left[\overline{x} - \tau \right].$$

Maximizing, $\left.\frac{\partial \log L}{\partial \tau}\right|_{\hat{\tau}_{\text{MLE}}} \overset{!}{=} 0$, directly yields, again

$$\hat{\tau}_{\text{MLE}} = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

2.5 Uniform: $\hat{b}_{\text{MLE}}$, assuming a is known.

Solution: Let $\vec{x} = (x_1, x_2, \dots, x_n)$ be an experiment of n independent observations. This time we will only maximize the likelihood function, since there is no advantage in taking the logarithm,

$$L(b; \vec{x}, a) = f(\vec{x}|a, b) = \prod_{i=1}^{n} f(x_i; a, b) \geq 0.$$

Note that this expression becomes zero if only one of the x_i lies outside of $[a, b]$. This means we need to choose $b \geq x_i$ for all $i = 1, \dots, n$, to maximize $L(b; \vec{x}, a)$. Remember that we assume "a" is known, so we keep it fixed. That said, the likelihood function takes the form.

$$L(b; \vec{x}, a) = \frac{1}{(b - a)^n}.$$

Its derivative is

$$\frac{\partial L}{\partial b} = -\frac{n}{(b-a)^{n+1}} < 0.$$

This means $L(b; \vec{x}, a)$ decreases for larger b. From these two considerations, we conclude that $L(b; \vec{x}, a)$ is maximized for

$$\hat{b}_{\mathrm{MLE}} = \max\{x_1, \ldots, x_n\}.$$

2.6 Choose from above two distributions and compute their Cremér-Rao Lower Bound (please **do not** choose the uniform distribution, since a regularity condition for the CRLB to hold is not met).

Solution: To compute the Cremér-Rao Lower Bound we compute first the Fisher information given by

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right].$$

Note that for all parameters (we will leave out σ^2, since its MLE estimator $\hat{\sigma}^2_{\mathrm{MLE}}$ is not unbiased and the CRLB as was presented in the lecture holds only for unbiased estimators), the *score* has the form

$$\frac{\partial \log L\left(\theta; \vec{x}\right)}{\partial \theta} = \frac{n}{\mathrm{Var}\left(X\right)}\left[\overline{x} - \mathbb{E}[X]\right].$$

Hence, the Fisher information is

$$I\left(\theta\right) = \frac{n^2}{\mathrm{Var}\left(X\right)^2} \underbrace{\mathbb{E}\left[\left(\overline{x} - \mathbb{E}[X]\right)^2\right]}_{=\mathrm{Var}(\overline{x}) = \frac{\mathrm{Var}(X)}{n}} = \frac{n}{\mathrm{Var}\left(X\right)},$$

where we used the fact that the sample mean $\overline{x}$ is an unbiased estimator of the population mean. That said, the CRLB reads,

$$\mathrm{Var}\left(\hat{\theta}\right) \geq \frac{\mathrm{Var}(X)}{n}.$$

This also agrees with the fact that MLE estimators are asymptotically efficient. In this case, they are even efficient for every sample size.

3 Exercise: Central Limit Theorem

Plot the distribution of the MLEs for `n = [1, 2, 5, 10, 100, 1000]` observations in each experiment, assuming the following distributions. The number N of experiments is up to you. Then fit a Gaussian to verify the agreement with the CLT.

Solution: See Jupyter Notebook.

3.1 Bernoulli: $p = 0.5$.

3.2 Exponential: $\tau = 1$.

3.3 Now try this assuming the so-called Cauchy distribution and explain why this is not a violation of the CLT.

4 Exercise: MC Integration

We want to estimate a definite integral of the form

$$I = \int_a^b \mathrm{d}x f(x). \tag{6}$$

We then introduce a probability density function $p(x) > 0$ in the range of the integration $[a, b]$:

$$I = \int_a^b \mathrm{d}x \underbrace{\frac{f(x)}{p(x)}}_{\equiv w(x)} p(x) = \mathbb{E}[w(x)]. \tag{7}$$

The natural choice for p is the uniform distribution $p(x) = \frac{1}{b-a}$ for the range of integration $[a, b]$, otherwise zero. In order to estimate $\mathbb{E}[w(x)]$ and therefore I, we can generate many observations $w(x_i)$ according to $p(x)$ and construct the sample average that is a consistent and unbiased estimator of $I = \mathbb{E}[w(x)]$. I.e.

$$\hat{I} \equiv \frac{1}{N}\sum_{i=1}^{N} w(x_i) = \frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i), \tag{8}$$

with $\hat{I} \to I$ for $N \to \infty$ and $\mathbb{E}[\hat{I}] = I$ for all $N \in \mathbb{N}$.

Solution: See Jupyter Notebook.

4.1 Estimate the integral I with $f(x) = e^x$ and $[a, b] = [-1, 1]$.

4.2 Plot the distribution of $\hat{I}$ for $N = 10, 100, 1000$.

4.3 Can you approximate the variance of $w(x)$ with the help of the distributions calculated in 4.2? Explain.

4.4 Calculate $\mathrm{Var}(w)$ analytically to confirm your results from 4.3.

4.5 This approach can be naturally extended to higher dimensions, transforming the interval $[a, b]$, into a multidimensional integration volume V. Think of a way of calculating the area of a circle of radius r and determine the value of π from that.