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 $p \in [0,1]$.

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 $\lambda > 0$.

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 $\sigma > 0$.

1.4 Exponential, with probability density function

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

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

1.5 Uniform, with probability density function

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

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

2 Exercise: MLE and Fisher Information

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

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

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

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

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

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

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).

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.

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 $N \in \mathbb{N}$.

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.