Statistics and Data Analysis - Lab-02

Barbara Żogała-Siudem

2025/2026

1 Cumulative distribution function and quantile function

1.0.1 Excercise (CDF)

Plot the cumulative distribution function for:

(a) $X \sim \exp(\lambda)$ for different values of $\lambda > 0$.
(b) $X \sim \mathcal{N}(0, 1)$, $X \sim \mathcal{N}(0, 0.1)$, $X \sim \mathcal{N}(1, 1)$
(c) $X \sim \mathrm{Bin}(20, 0.5)$, $X \sim \mathrm{Bin}(20, 0.1)$, $X \sim \mathrm{Bin}(20, 0.9)$

1.0.2 Exercise

When $X \sim \exp(\lambda)$, then $f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{(0,\infty)}(x)$.

Find:

- cumulative distribution function $F_X(t)$,
- quantile function $Q_X(p)$.

Find:

(a) $F_X(0)$
(b) $Q_X(0.5)$, how this quantile is called?
(c) $\mathbb{P}(X < \frac{\ln(5)}{\lambda})$

1.0.3 Exercise (CDF and quantiles function)

Assuming that $X \sim \mathcal{N}(0, \sigma^2)$, find

(a) $\mathbb{P}(X \in (-a, a))$, for $\sigma^2 \in \{1, 2\}$ and $a \in \{1, 3, 5\}$
(b) $F_X(0)$
(c) $\mathbb{P}(X = 0)$

Now, assume that $X \sim Cauchy()$ (standard Cauchy distribution). How results from point (a) would change? What does it mean?

2 Statistics

2.1 Pseudorandom number generation

Numpy package enables to generate samples from a variety of distributions. See `numpy.random` module and generate pseudo random samples from distributions. As always, check documentation to understand the parameters of given distributions.

- $\mathcal{N}(0, 1)$
- $\mathcal{U}(-3, 3)$
- Cauchy()

- Pareto(2)

2.2 Visualise empirical distribution

2.2.1 Exercise (visualize distributins)

For each generated sample plot:

- boxplot,
- violinplot,
- histogram with theoretical and estimated density,
- ECDF.

Use:

- `pandas.DataFrame` plotting functions
- `matplotlib.pyplot` functions
- `seaborn` package functions

2.2.2 Exercise (Central Limit Theorem)

Draw random samples from selected distribution (any distribution that satisfy the assumptions of CLT). Calculate appropriate sums and see if obtained histogram resembles the density of a normal distribution. What are the parameters of this normal distribution.

How can this theorem 'explain' why normal distribution is so commonly observed in reality?

2.2.3 Exercise (empirical quantiles)

Consider two samples:

(a) `x = [0, 0, 0, 0, 0, 1, 1, 8, 9, 9]`
(b) `x = np.random.normal(0, 1, 100)`

Calculate quantiles $q_{0.01}, q_{0.1}, q_{0.25}, q_{0.5}, q_{0.75}$. How results change between different methods for estimating quantiles?

2.2.4 Exercise (Inverse transform sampling)

One way to generate a random sample from given continuous distribution is the inverse transform sampling method. It bases on fact, that for a random variable $U \sim \mathcal{U}(0,1)$, and a continuously distributed $X \sim F_X$: $F_X^{-1}(U) \sim F_X$.

It means, that a sample $u_1, \ldots, u_n \sim \mathcal{U}(0,1)$ can be transformed to $F^{-1}(u_1), \ldots, F_X^{-1}(u_n) \sim F_X$

Knowing this fact, generate a sample from exponential distribution and plot its histogram along with theoretical density function.

2.2.5 Exercise (Q-Q plot)

To compare two empirical distributions or an epirical distribution againts a theoretical one we can also plot their quantiles against each other. This type of plot is called QQ-plot (quantile-quantile plot).

Plot QQ-plots for theoretical quantiles for standard normal distribution and:

(a) a sample from standard normal distribution,
(b) a sample from cauchy distribution
(c) a sample from exponential distribution