Introduction to lab05

Barbara Żogała-Siudem

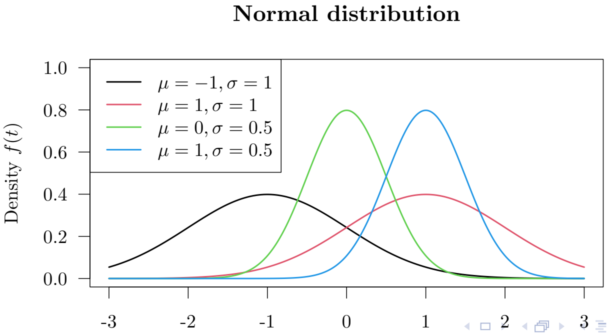
20 listopada 2025

Normal distribution

Normal distribution: $X \sim \mathcal{N}(\mu, \sigma)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad \text{for} \;\; \sigma > 0$$

$$\mathbb{E}X = \mu$$

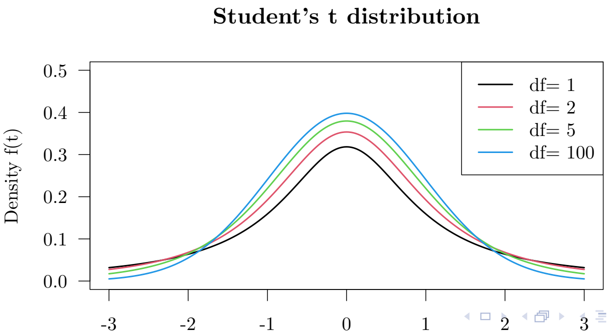


Student's t distribution

Student's t distribution: $X \sim t_\nu$

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad \text{for } \nu > 0$$

$$\mathbb{E}X = 0, \quad \text{for } \nu > 1, \text{otherwise undefined}$$



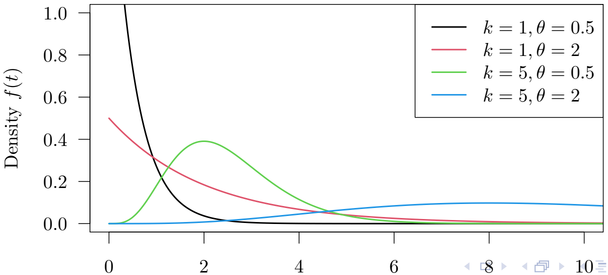
Gamma distribution

Gamma distribution: $X \sim \Gamma(k, \theta)$

$$f(x) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \quad \text{for } x > 0,\ k, \theta > 0.$$

$$\mathbb{E}X = k\theta$$

Gamma distribution



Python functions to random sample generation

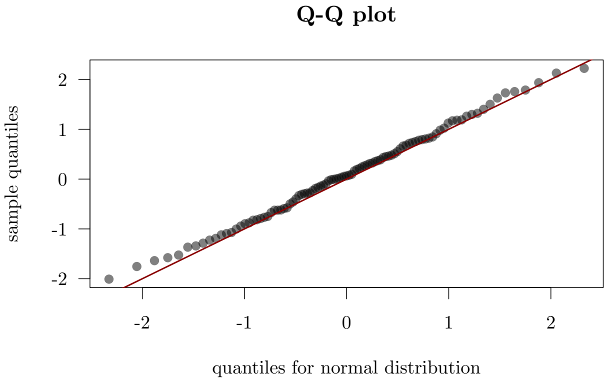
Useful functions:

- `numpy.random.normal()`
- `numpy.random.standard_t()`
- `numpy.random.gamma()`

See documentation for parameters.

Q-Q plot

- Q-Q plot: Q stands for quantile
- compare visually two distributions
- OX: quantiles for 1st distribution
- OY: quantiles for 2nd distribution



python - hints

Histogram

```
import matplotlib.pyplot as plt
plt.hist()
```

Q-Q plot

```
import statsmodels.api as sm
sm.qqplot()
```

Plot Q-Q plot also manually, to understand what it is

```
scipy.stats.norm.ppf()
np.quantile()
plt.scatter()
plt.axline()
```

Idea

Task

Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$
Question: How to check if $\mu = 0$?
Estimator of μ in normal distribution: $\bar{X}$.

Assume that:

- $\bar{X} = 0.5$,
- $\bar{X} = 0.01$

So can we assume that $\mu = 0$? Well, it depends:

- How 'sure' do we want bo be?
- How big is the sample?

Types of errors

Statistical test

H_0: $\mu = 0$ (we especially do not reject it by mistake)
H_1: $\mu \neq 0$

What kind of mistakes (errors) we can make:

decision \truth	$\mu = 0$	$\mu \neq 0$
$\mu = 0$	OK	II-type error (β)
$\mu \neq 0$	I-type error (α)	OK

Are errors equally bad?

Screening for diseases

H_0: disease
H_1: healthy
We especially do not want miss detecting illness!

Guilty or not guilty

H_0: not guilty
H_1: guilty
We especially do not want send to prison innocent people...

Is there a bomb in a luggage?

H_0: bomb
H_1: not bomb
We especially do not want miss detecting bomb!

Significance level and power of the test

Test

H_0: $\mu = 0$ (we especially do not reject it by mistake)
H_1: $\mu \neq 0$

Significance level α

$$\alpha = \mathbb{P}(\text{reject } H_0 | H_0 \text{ is true})$$

In other words:

$$\mathbb{P}(\text{I-type error}) = \alpha$$

Power of the test $1 - \beta$

$$\beta = \mathbb{P}(\text{do not reject } H_0 | H_1 \text{ is true})$$

In other words:

$$\mathbb{P}(\text{II-type error}) = \beta$$

How do actually statistical test work?

Let us get back to our sample:

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$$

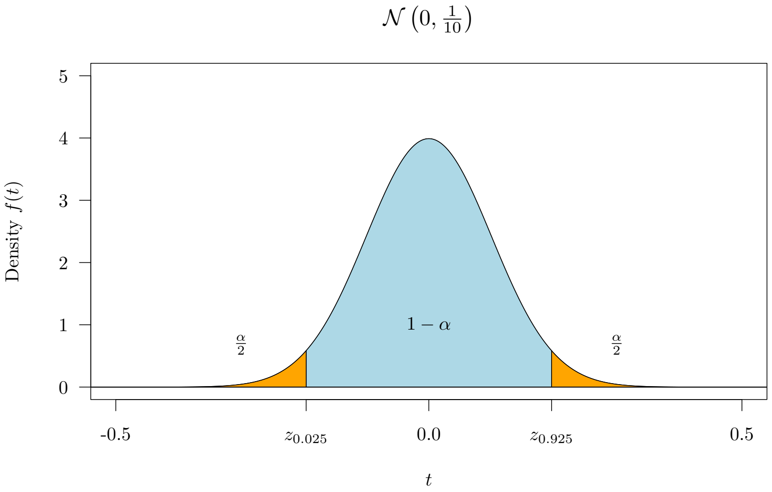
Let us assume that H_0 is true, then

$$\bar{X} \sim \mathcal{N}(0, \frac{1}{\sqrt{n}})$$

Which leads to conclusion, that

$$\mathbb{P}\left(\bar{X} \in (z_{\alpha/2}, z_{1-\alpha/2})\right) = 1 - \alpha$$

How do actually statistical test work?



p-value

Critical area

We check if the value of the test statistic lies in a critical area. If so, then we reject H_0.

$$T(X_1, \ldots, X_n) \in K_\alpha \implies \text{reject } H_0$$

K_α in our example: orange area.

p-value

- Probability that we obtain at least as 'bad' results of the test statistic as the one we actually observed, assuming that H_0 is true.
- Small values – we reject H_0.
- Tests in statistical packages return p-value.

Tests needed for this class

For more on testing see the following lecture

https://www.ibspan.waw.pl/~opara/statistics_and_data_analysis/05_hypothesis_testing.pdf

Some tests which can be used today:

- `scipy.stats.kstest()`
- `scipy.stats.shapiro()`
- `scipy.stats.ttest_ind()`
- `scipy.stats.ranksums()`