Statistics

L3: Sampling Distributions

Katarzyna Filipiak

Institute of Mathematics
Poznań University of Technology

2025/2026

Random sample

Aim of statistics: to draw conclusions about population from a set of observed data.

Sample – a known and measurable entity representing a population having unknown characteristics

Reasons for sampling – examples:

- economic
- timeliness
- large populations
- destructive nature of the observations
- inaccessible populations

Random sample

Observation, before the experiment is performed, is modeled by a random variable X, having its own probability distribution $f(x)$ (distribution of the whole population)

Sample of size n – set of n independent random variables $X_1, X_2, \ldots, X_n$ with the same probability distribution $f(x)$ as X.

$X_1, X_2, \ldots, X_n$ – random variables representing unknown measurements, which after random sampling become 1st, 2nd, $\ldots$, nth observation

$x_1, x_2, \ldots, x_n$ – observations (realizations of random variables $X_1, X_2, \ldots, X_n$)

Example 1

The researcher would like to verify how many bicycles are owned by the families living in the city of size 50,000 families. Thus, 100 families were asked for the number of bicycles. Let X denote the random variable that counts the number of bicycles in a family. Define population, sample, random variable and give the example of observations.

Population:

Sample:

Random variable:

Observations (example):

Sample statistics

Sample statistic – arbitrary function of random variables (function of a sample) $X_1, X_2, \dots X_n$, which does not contain unknown parameters, e.g.

- sample mean,
- sample variance,
- sample standard deviation,
- sample proportion (for population having binomial distribution).

Statistic $=$ function of random variables $=$ RANDOM VARIABLE!!!

Statistic, as a random variable, has its own probability distribution!!!

Example 2

Let the data in a table below represent the midterm examination results for the 5 students attending a metallurgy seminar. The population consists of the exam grades. Although such a tiny sample is unusual and statistical investigations ordinarily involve much larger populations, the miniature scale of this example will simplify our presentation.

Name	Daniel	Robert	Ana	Ida	John
Grade	3	2	3	4	2

A simple random sample of 2 student will be selected.

Establish the distribution of the grade sample mean.

X – grade, X_1, X_2 – grades of two randomly chosen students

Population:

D	R	A	I	J
3	2	3	4	2

Sample mean distribution (1)

Observed feature in a population – random variable $X \sim N(\mu, \sigma)$,

σ - known

Sample: $X_1, X_2, \ldots, X_n,$ $X_i \sim N(\mu, \sigma),$ σ - known

Sample mean:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Sample total:

$$T = X_1 + X_2 + \ldots + X_n \sim N(n \cdot \mu, \sqrt{n} \cdot \sigma)$$

Sample mean distribution (2)

Observed feature in a population – random variable X (arbitrary distr.)

Large sample: $X_1, X_2, \ldots, X_n \quad (n > 30)$

Central Limit Theorem

Consider a population having mean μ and finite standard deviation σ. Let $\overline{X}$ represent the mean of n independent random observations from this population. The sampling distribution of $\overline{X}$ tends toward a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, i.e.

$$\overline{X} \underset{\text{app}}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Sample total: $T = X_1 + X_2 + \ldots + X_n \underset{\text{app}}{\sim} N(n \cdot \mu, \sqrt{n} \cdot \sigma)$

Sample mean distribution (3)

Observed feature in a population – random variable $X \sim N(\mu, \sigma)$,

σ - unknown

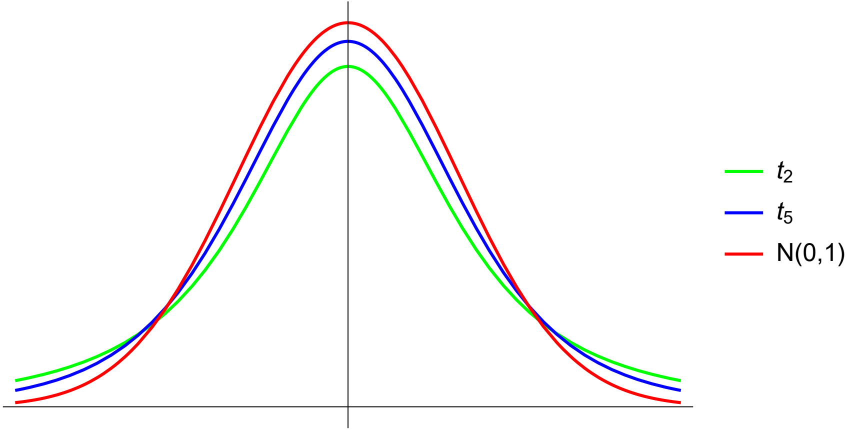
Sample: $X_1, X_2, \ldots, X_n$, $X_i \sim N(\mu, \sigma)$, σ - unknown

(Standardized) sample mean:

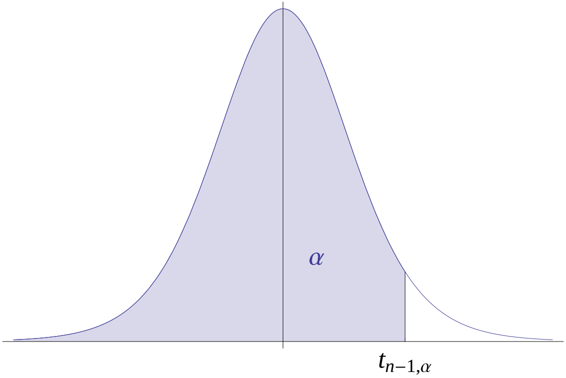
$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

$$S^2 = \tfrac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n \cdot \overline{X}^2\right)$$

t-Student distribution



t-Student distribution



$$P(t < t_{n-1,\alpha}) = \alpha$$

$t_{n-1,\alpha}$ – quantile of t_{n-1}: $\texttt{qt}(\alpha,\, n-1)$

Example 3

Assume that the weight of Poles has normal distribution with the mean $\mu = 70$ and standard deviation $\sigma = 10$.
9 students got into the elevator. What is the probability, that they exceeded the norm of 650 kg?

Example 4

Electronic components often have lifetimes (L) that empirically have been shown to fit an exponential distribution with $\lambda = 0.01$.
Suppose that a random sample of $n = 81$ items is chosen. Sample lifetimes might be used to establish μ from the computed mean time between failures, $\overline{L}$. What is the probability, that $\overline{L}$ falling within $\mu \pm 5$ days?

Sample mean distribution – summary

(1) $X_i \sim N(\mu, \sigma)$, μ, σ – known:

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \quad \Rightarrow \quad \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

$$T \sim N(n\,\mu, \sqrt{n}\,\sigma)$$

(2) X_i with μ, σ – known, distribution of X_i is arbitrary, sample is large:

$$\overline{X} \underset{\text{app}}{\sim} N(\mu, \frac{\sigma}{\sqrt{n}}) \quad \Rightarrow \quad \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \underset{\text{app}}{\sim} N(0, 1)$$

$$T \underset{\text{app}}{\sim} N(n\,\mu, \sqrt{n}\,\sigma)$$

(3) $X_i \sim N(\mu, \sigma)$, μ – known, σ – unknown: $\displaystyle \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$

Sample variance distribution

Observed feature in a population – random variable $X \sim N(\mu, \sigma)$,
σ - known

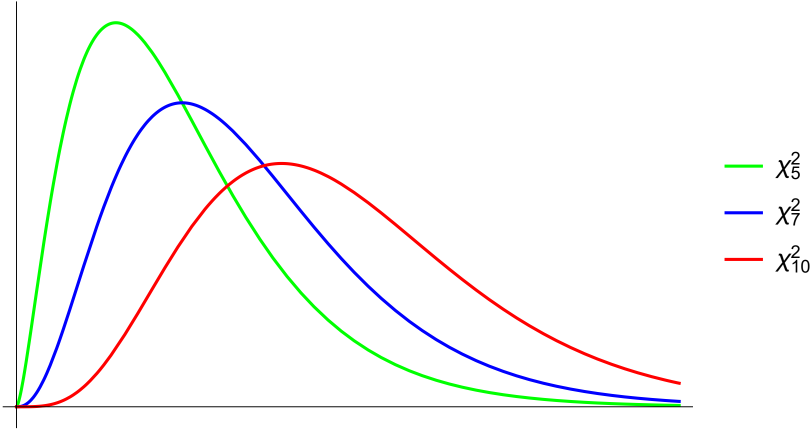
Sample: $X_1, X_2, \dots, X_n$, $X_i \sim N(\mu, \sigma)$, σ - known

Sample variance: $S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$

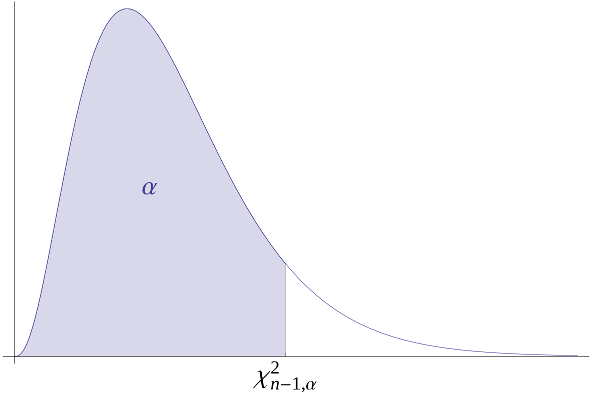
Sample variance distribution:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

χ^2 - distribution



χ^2 distribution



$P(\chi^2 < \chi^2_{n-1,\alpha}) = \alpha$

$\chi^2_{n-1,\alpha}$ – quantile of χ^2_{n-1}: $\texttt{qchisq}(\alpha, n-1)$

Sample proportion (probability of success)

Observed feature in a population – random variable $X \sim \text{bin}(1, p)$

Sample (large): $X_1, X_2, \ldots, X_n$ ($n > 100$), $X_i \sim \text{bin}(1, p)$

$$T = \sum_{i=1}^{n} X_i \text{ – number of "successes" in a sample}$$

Sample proportion: $\widehat{p} = \dfrac{T}{n}$

Sample proportion distribution:

$$\widehat{p} \underset{\text{app}}{\sim} N\left(p, \sqrt{pq/n}\right)$$

Example 5

An advertising agency ran a campaign to introduce a product. At the end of its campaign, it claimed that at least 25% of all consumers were now familiar with the product. If 25% of all consumers actually knew of the product, what is the probability that no more than 232 in a random sample of 1000 consumers were familiar with the product?