Introduction to computational statistics

LMS January 2026

* Descriptive statistics

* Foundations of probability

* Hypothesis testing

* Bayesian statistics

LMS statistics and hypothesis testing

Chapter 1. Descriptive statistics

* Prediction and inference
* Population and sampling
* True parameters and statistical estimators

Chapter 2. Foundations of probability

* Probability and random events
* Discrete probability ; Bernoulli, Uniform, Binomial, Poisson
* Continuous probability ; Gaussian, exponential, Uniform.

Chapter 3. Hypothesis testing (I)

* the law of large numbers
* the central limit theorem
* Confidence intervals and critical regions.

Chapter 4. Hypothesis testing (II)

* The Fisher, Pearson, Neyman approach.
* Some examples : t-test, F-test, χ^2-test
* Parametric vs non parametric tests
* Error types in hypothesis testing (*) Bayesian statistics.

2.3 Continuous random variables

i) Discrete case : A random variable can have a countable / discrete number of outcomes x_i

Probability $P \in [0,1]$

Unitarity $\sum_{\forall x_i} P(x_i) = 1$

frequentist definition ✓

* Bernoulli events $\text{Bern}(x_i, p)$
* Uniform events $\mathcal{U}(x_i, n)$
* Binomial events $B(x_i; n, p)$
* Poisson events $P(x_i; \lambda)$

Probability "mass" distributions

ii) Continuous case : A random variable can take any value in a continuous range $x \in \mathbb{R}$

Density $f(x)$

Unitarity $\int_{-\infty}^{\infty} f(x)\,dx = 1$

frequentist definition X

* Gaussian distribution
* Exponential distribution
* Continuous uniform

Probability "density" distributions

i) Gaussian distribution

"Random variable x centered at some μ and spreaded σ"

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

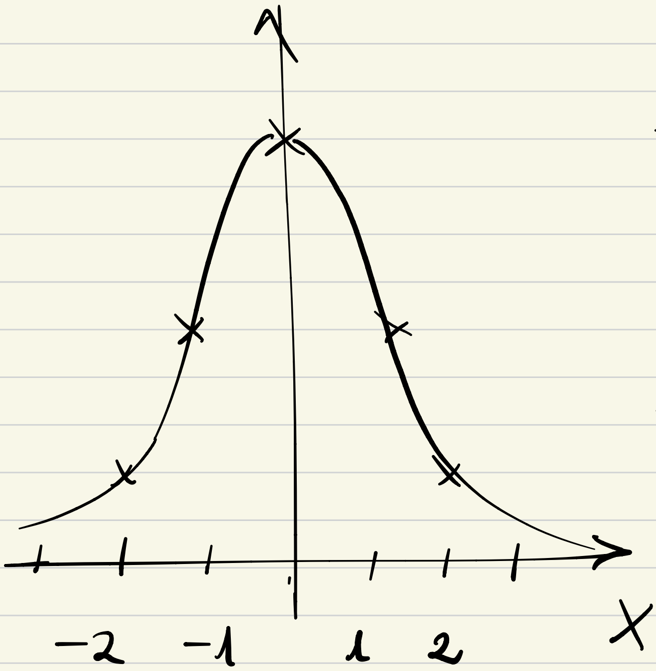
μ : mean value

σ : standard deviation

* Consider simplest case $\mu = 0$; $\sigma = 1$; ignore normalization factor

Visualize simplified $f(x) = e^{-x^2/2}$; plot some values

$$f(x) = e^{-x^2/2}$$



$$f(0) = e^{-0} = 1$$

$$f(1) = e^{-1/2} = \frac{1}{e^{1/2}} \simeq 0.607$$

$$f(-1) = e^{-(-1)^2/2} = \frac{1}{e^{1/2}} \simeq 0.607$$

$$f(\pm 2) = e^{-(\pm 2)^2/2} = \frac{1}{e^2} \simeq 0.135$$

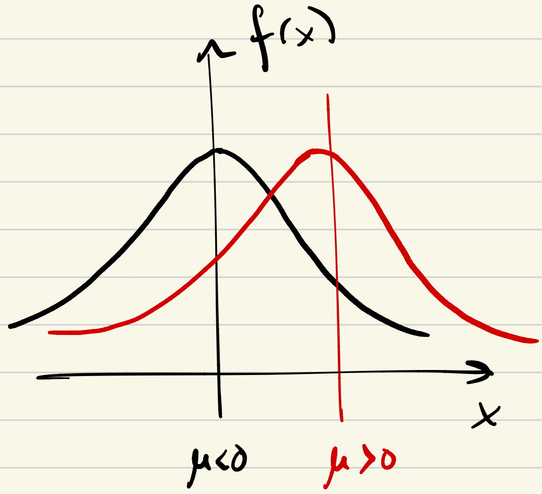
the exponential squared gives symmetric "bell" shape.

* Switch back on the mean value $\mu \neq 0$

for an arbitrary value of μ ;

$$f(x) = e^{-\frac{1}{2}(x-\mu)^2}$$

μ regulates central location

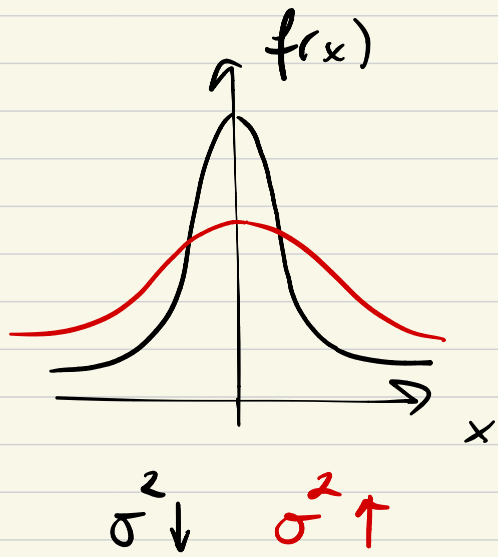


* Switch back on the variance $\sigma^2 \neq 1$

for an arbitrary value of σ^2 ;

$$f(x) = e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

σ^2 regulates shape / spread



* Add normalization factor

i) Remember discrete probability : Unitarity $\sum_{\forall x_i} P(x_i) = 1$

ii) Continuous case. Density distributions $\int_{-\infty}^{+\infty} dx\, f(x) = 1$

$$\int_{-\infty}^{\infty} dx \; e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \sigma\sqrt{2\pi}$$

(*) Exercise

* for $f(x)$ to behave as a probability, we add a normalization factor

$$\left\{ f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \right\}$$ Gaussian distribution ; $\int_{-\infty}^{\infty} dx\, f(x) = 1$

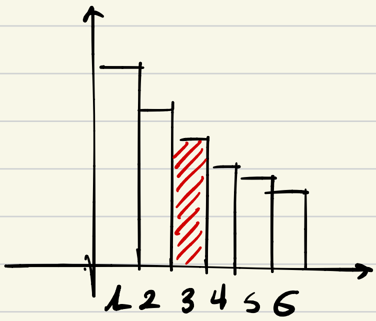
* Particular case $\mu = 0$; $\sigma = 1$

$$\left\{ f(x) = \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}x^2} \right\}$$ Standard Normal distribution $N(0,1)$

③ Hypothesis testing

* In discrete variables, the "mass" probability distributions represent exact probabilities.

$P(x)$: Bern(x), U(x), B(x), Pois(x), ...



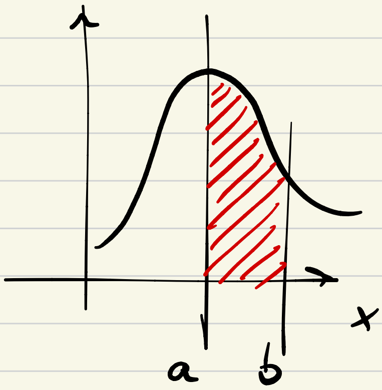
$$P(x=x_i) = B(x_i; n, p)$$ ✓

x : number of Hs in n tosses

* In continuous variables, the "density" probability distributions do not. We can only compute probabilities in a given range.

Never of a single value!

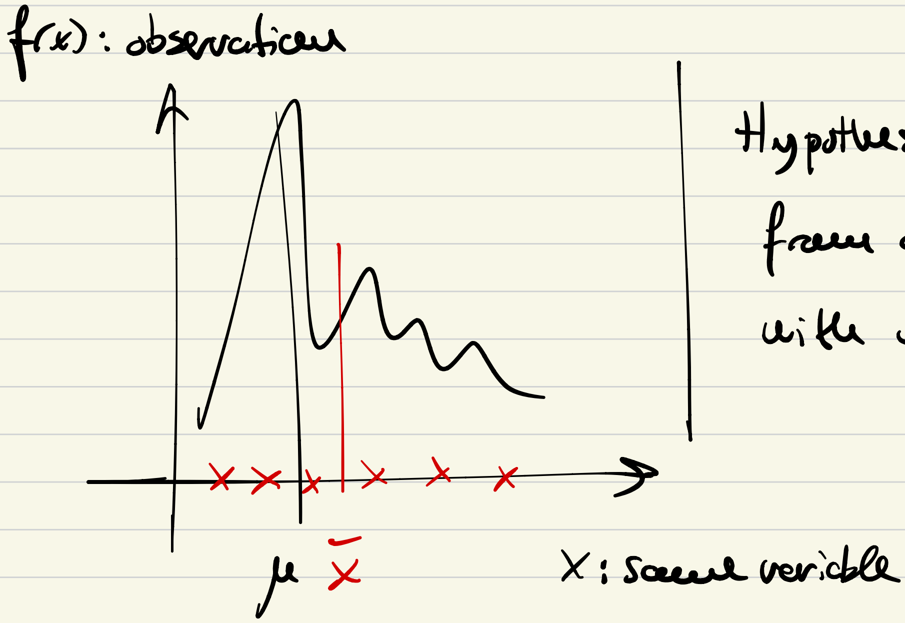
$f(x)$: Gauss(x), Exp(x), Unif(x), ...



$$P(a<x<b) = cdf(b) - cdf(a)$$

x : continuous variable

I want to compare sample mean $\bar{x}$ of a given sample (mutant) with an expected value of the true population μ (control)

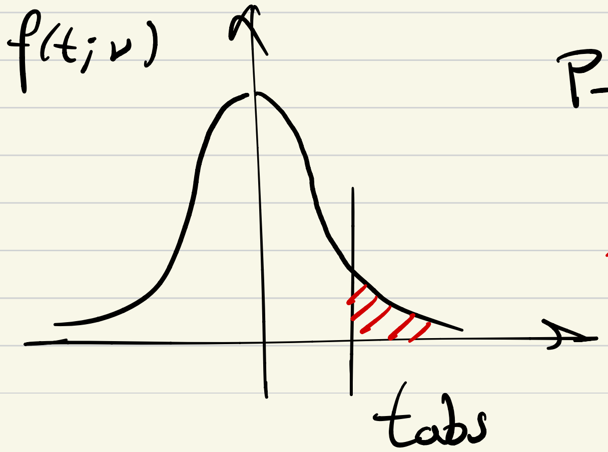


Hypothesis: my sample is drawn from a gaussian distribution with mean value μ

* Define a quantity that quantifies how different my sample mean is from the true / hypothesized μ.

$$\left\{ t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \right\}$$

* The t variable follows a "Student's t distribution"



P-value : $P(t \geq t_{obs}) = \int_{t_{obs}}^{\infty} f(t;\nu)\,dt$

3.2 The law of large numbers.

Bernoulli, Chebyshev, Borel (1713)

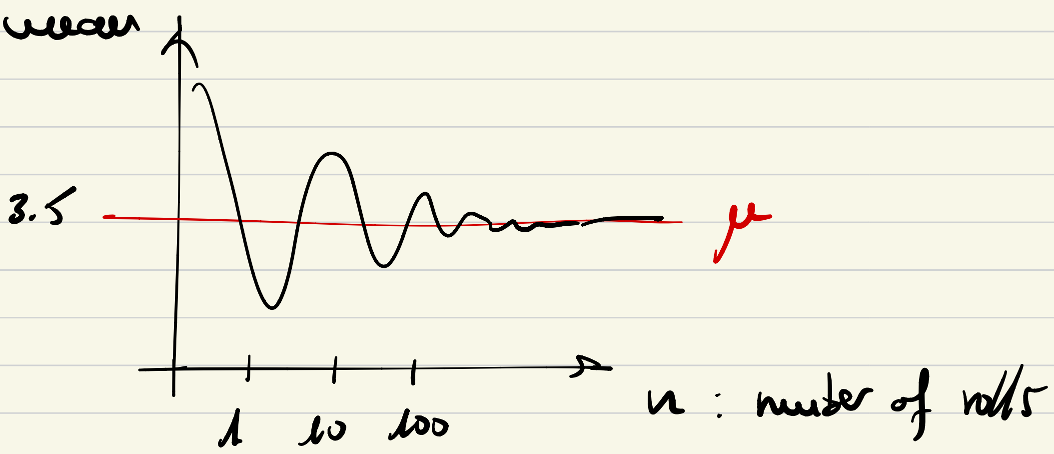
"Let $\{X_1, X_2, \dots, X_n\}$ independent and identically distributed (iid) random variables, with same true mean μ (exp. value), the sample mean $\bar{X}$ converges to μ as n increases"

$$\left\{ \lim_{n \to \infty} \bar{X} = \mu \ ; \text{ being } \bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \right\}$$

Example: rolling dice.

i) True / expected mean $\mu = \frac{1}{6}(1+2+3+4+5+6) = 3.5$

ii) As we increase n rolls, sample mean $\bar{X}$ will converge to the true / expected value μ.



3.3 The Central limit theorem

A. de Moivre, Laplace, Gauss (1733)

"For a set $\{x_1, \dots, x_n\}$ of n independent and identically distributed (iid) random variables, the distribution of sample means tends to gaussian as n increases, regardless of how are x_i distributed themselves"

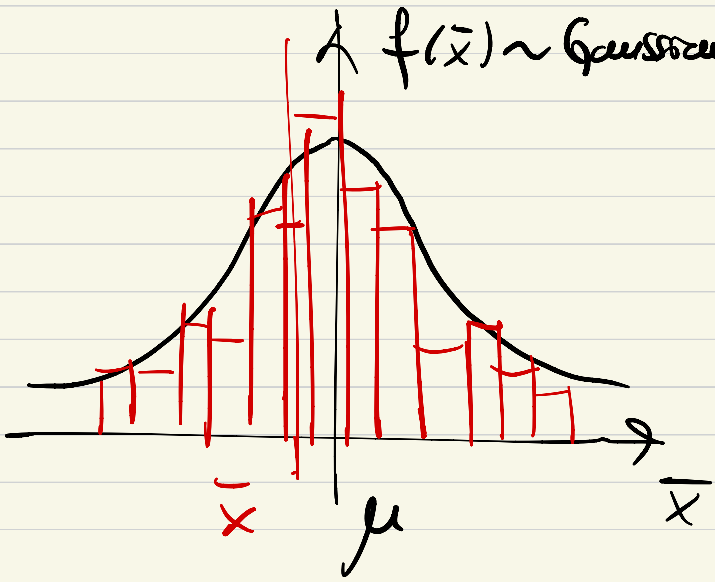
* Sample mean

$$\bar{x} \to \mathcal{N}(\mu, \sigma^2/n)$$

* Standard error (SE)

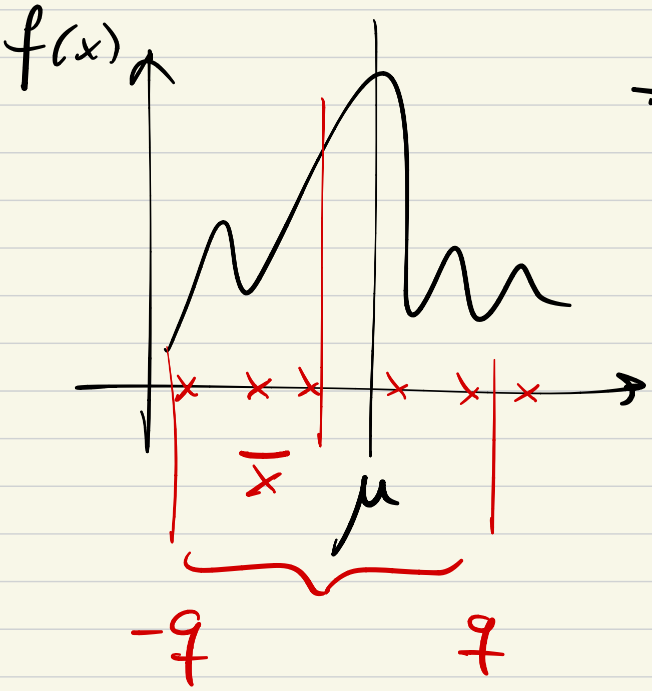
$$SE = \frac{\sigma_i}{\sqrt{n}} \; ; \text{ as I increase } n, \downarrow SE$$

* As we increase n rolls, sample mean $\bar{x}$ will tend to a Gauss(x)



3.4 Confidence intervals

"A p% confidence interval for μ is a random interval, centered at the sample mean $\bar{x}$, that contains μ with probability p"



* Assuming same distribution for x, with same true, unaccessible μ, I can just compute $\bar{x}, s^2$

x : random variable

* I can build an interval $(-q, q)$ centered at any sample mean $\bar{x}$, that contains μ with same "confidence"

$$P(-q \leq \bar{x} - \mu \leq q) = p$$

* Different distributions have well known intervals for certain values of p (10, 20, 30, ...)