

# Tutoriat PS 7: Central limit theorem and maximum likelihood estimation

Mihai Duzi, Lucan Cristian, Luminaru Ionut

December 4, 2025

## 1 Central limit theorem

### 1.1 The Drunkard's Walk

$$\dots -3, -2, -1, 0, 1, 2, 3, 4, \dots$$

$$X_i = \begin{cases} +1, & \text{if he steps to the right,} \\ -1, & \text{if he steps to the left,} \end{cases} \quad i \geq 1$$

The variables  $X_i$  are i.i.d. Rademacher:

$$X_i \sim \begin{cases} -1 & \text{with probability } \frac{1}{2}, \\ +1 & \text{with probability } \frac{1}{2}. \end{cases}$$

Define the sum

$$S_n := X_1 + X_2 + \dots + X_n,$$

which models the drunkard's position after  $n$  steps.

## Properties

Let  $S_m = X_1 + X_2 + \dots + X_m$ , where the variables  $X_i$  are i.i.d. with  $\mathbb{E}[X_i] = 0$  and  $\mathbb{E}[X_i^2] < \infty$ .

(i)  $\mathbb{E}[S_m] = m \cdot \mathbb{E}[X_1] = 0$ .

(On average, the position returns to the origin.)

(ii)  $\text{Var}(S_m) = m \cdot \text{Var}(X_1) = m$ .

(Since  $\text{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = 1 - 0^2 = 1$ .)

Since  $(X_i)_{i \geq 1}$  are i.i.d. with  $\mathbb{E}[X_i^2] < \infty$ , we may apply the \*\*Law of Large Numbers\*\*:

$$\bar{S}_m = \frac{S_m}{m} \xrightarrow{P} \mathbb{E}[X_1] = 0.$$

(The empirical mean of the position.)

Thus,

$$S_m \approx 0 \cdot m = 0 \quad \text{as } m \rightarrow \infty$$

(this is not deterministic, only the mean stabilizes).

## Central Limit Theorem

Let  $(X_m)_{m \geq 1}$  be i.i.d. random variables with

$$\mathbb{E}[X_1] = \mu, \quad \text{Var}(X_1) = \sigma^2 < \infty.$$

Then:

$$\frac{\sqrt{m}}{\sigma}(\bar{S}_m - \mu) \xrightarrow{D} Z, \quad \text{where } Z \sim \mathcal{N}(0, 1).$$

Equivalently,

$$\frac{\sqrt{m}}{\sigma}(\bar{S}_m - \mu) \approx \mathcal{N}(0, 1).$$

Thus,

$$\bar{S}_m \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right),$$

and therefore

$$S_m \approx \mathcal{N}(\mu m, m\sigma^2).$$

**Intuition:** Taking a lot of independent samples from a distribution will create a normal distribution

## 2 Maximum Likelihood Estimation

This is used to calculate the parameters of distributions when the probability/density is maximized, using samples. Below are examples for different distributions:

### 2.1 Binomial Distribution Example

Let's say I toss throw a coin 100 times and I get 55 heads. What is the probability of getting heads on a random throw?

Well we can't know for sure, but we can calculate p when the probability of getting 55 heads in 100 throws is maximized:

$$P(55 \text{ tails} | p) = \binom{100}{55} p^{55} (1-p)^{45}.$$

We use the notation  $\hat{p}$  for the MLE. We use mathematical analysis to find it, taking the derivative of the likelihood function and setting it equal to 0.

$$\frac{d}{dp} P(\text{data} \mid p) = \binom{100}{55} (55p^{54}(1-p)^{45} - 45p^{55}(1-p)^{44}) = 0.$$

Solving the equation we obtain

$$55p^{54}(1-p)^{45} = 45p^{55}(1-p)^{44}$$

$$55(1-p) = 45p \quad (\text{because } p \in (0, 1))$$

$$55 = 100p$$

The MLE is  $\hat{p} = 0.55$ .

## 2.2 Exponential Distribution Example

Assume that the lifetime of a certain type of light bulb is modeled by an exponential distribution with unknown parameter  $\lambda$ . We test 5 bulbs and find that their lifetimes are 2, 3, 1, 3 and 4 years. What is the MLE for  $\lambda$ ?

**Answer:** We must be careful with our notation. With 5 different values, it is best to use indices. Let  $X_i$  be the lifetime of the  $i$ -th bulb and let  $x_i$  be the observed value taken by  $X_i$ . Then each  $X_i$  has pdf

$$f_{X_i}(x_i) = \lambda e^{-\lambda x_i}.$$

Assuming that bulb lifetimes are independent, the joint pdf is the product of individual densities:

$$f(x_1, x_2, x_3, x_4, x_5 \mid \lambda) = (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2})(\lambda e^{-\lambda x_3})(\lambda e^{-\lambda x_4})(\lambda e^{-\lambda x_5}) = \lambda^5 e^{-\lambda(x_1+x_2+x_3+x_4+x_5)}.$$

This is a conditional density, since it depends on  $\lambda$ . If we treat the data as fixed and  $\lambda$  as the variable, this function is the likelihood. The observed data are:

$$x_1 = 2, \quad x_2 = 3, \quad x_3 = 1, \quad x_4 = 3, \quad x_5 = 4.$$

Thus the likelihood and log-likelihood are:

$$f(2, 3, 1, 3, 4 \mid \lambda) = \lambda^5 e^{-13\lambda},$$

$$\ln f(2, 3, 1, 3, 4 \mid \lambda) = 5 \ln(\lambda) - 13\lambda.$$

Finally, using calculus to find the MLE:

$$\frac{d}{d\lambda} (\text{log likelihood}) = \frac{5}{\lambda} - 13 = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{5}{13}.$$

### Observations:

1. We use uppercase letters for random variables and lowercase letters for the observed values they take.
2. The MLE for  $\lambda$  turns out to be the inverse of the sample mean. Indeed, if  $X \sim \exp(\lambda)$  then  $E(X) = 1/\lambda$ .

## 2.3 Normal Distributions

Assume the data  $x_1, x_2, \dots, x_n$  come from a normal distribution  $N(\mu, \sigma^2)$ , where  $\mu$  and  $\sigma$  are unknown. Find the maximum likelihood estimators for  $(\mu, \sigma^2)$ .

**Answer:** Let  $X_1, \dots, X_n$  be i.i.d.  $N(\mu, \sigma^2)$  and let  $x_i$  be the observed value of  $X_i$ . The density of each  $X_i$  is

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Since the  $X_i$  are independent, the joint pdf is

$$f(x_1, \dots, x_n | \mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Thus the likelihood and log-likelihood are:

$$f(x_1, \dots, x_n | \mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right),$$

$$\ln f(x_1, \dots, x_n | \mu, \sigma) = -n \ln(\sqrt{2\pi}) - n \ln(\sigma) - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Since  $\ln f$  is a function of two variables, we use partial derivatives. Differentiating with respect to  $\mu$ :

$$\frac{\partial}{\partial \mu} \ln f(x_1, \dots, x_n | \mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0,$$

$$\sum_{i=1}^n x_i = n\mu \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Differentiating with respect to  $\sigma$ :

$$\frac{\partial}{\partial \sigma} \ln f(x_1, \dots, x_n | \mu, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0,$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2.$$

## 2.4 Uniform distributions

Suppose that our data  $x_1, \dots, x_n$  are independent from a uniform distribution  $U(a, b)$ . Find the MLE for  $a$  and  $b$ .

**Answer:** This example is different from the previous ones because we need to use mathematical analysis to find the MLE. The density for  $U(a, b)$  is  $\frac{1}{b-a}$  on  $[a, b]$ . Therefore our likelihood function is

$$f(x_1, \dots, x_n | a, b) = \begin{cases} \left(\frac{1}{b-a}\right)^n, & \text{if all } x_i \text{ are in the interval } [a, b], \\ 0, & \text{otherwise.} \end{cases}$$

This is maximized by making  $b - a$  as small as possible. The only restriction is that the interval  $[a, b]$  must contain all the data. Thus, the MLE for the pair  $(a, b)$  is

$$\hat{a} = \min(x_1, \dots, x_n), \quad \hat{b} = \max(x_1, \dots, x_n).$$

## 2.5 Extra Example: Capture–recapture method

The capture–recapture method is a way to estimate the size of a wildlife population. The method assumes that each animal in the population is equally likely to be caught in a trap.

Suppose that 10 animals are caught, tagged, and released. A few months later, 20 animals are caught, examined, and released. Of these 20, 4 are found to be tagged. Estimate the size of the wildlife population using MLE for the probability that a wild animal is tagged.

**Answer:** Our unknown parameter  $n$  is the number of wild animals. Our data are that 4 of the 20 animals captured the second time are tagged (and that 10 animals were tagged the first time). The likelihood function is

$$P(\text{data} | n \text{ animals}) = \frac{\binom{n-10}{16} \binom{10}{4}}{\binom{n}{20}}.$$

(The numerator is the number of ways to choose 16 untagged animals from the  $n - 10$  untagged ones, times the number of ways to choose 4 from the 10 tagged animals. The denominator is the number of ways to choose 20 animals from the whole population of  $n$ .)

We can use R (with the command `dhyper(16, n-10, 10, 20)`) for various values of  $n$  to compute that the likelihood function is maximized when  $n = 50$ . This should make sense: the best estimate is when the fraction of tagged animals in the entire population is 10/50, which matches the fraction of animals captured the second time that are tagged (4/20).

## 2.6 Extra example: Hardy–Weinberg

Suppose that a particular gene has two alleles ( $A$  and  $a$ ), with allele  $A$  having frequency  $\theta$  in the population. That is, a random copy of the gene is  $A$  with probability  $\theta$  and  $a$  with probability  $1 - \theta$ . Since a diploid genotype consists of 2 genes, the probability of each genotype is:

genotype	probability
$AA$	$\theta^2$
$Aa$	$2\theta(1 - \theta)$
$aa$	$(1 - \theta)^2$

Suppose we test a random sample of people and find that  $k_1$  are  $AA$ ,  $k_2$  are  $Aa$ , and  $k_3$  are  $aa$ . Find the MLE for  $\theta$ .

**Answer:** The likelihood function is

$$P(k_1, k_2, k_3 | \theta) = \binom{k_1}{k_1 + k_2 + k_3} \binom{k_2}{k_2 + k_3} \binom{k_3}{k_3} \theta^{2k_1} (2\theta(1-\theta))^{k_2} (1-\theta)^{2k_3}.$$

The log-likelihood is

$$\text{constant} + 2k_1 \ln(\theta) + k_2 \ln(\theta) + k_2 \ln(1-\theta) + 2k_3 \ln(1-\theta).$$

Setting the derivative equal to 0:

$$\frac{2k_1 + k_2}{\theta} - \frac{k_2 + 2k_3}{1-\theta} = 0.$$

Solving for  $\theta$ , we find the MLE is

$$\hat{\theta} = \frac{2k_1 + k_2}{2k_1 + 2k_2 + 2k_3},$$

which is simply the fraction of allele  $A$  among all gene copies in the sampled population.