

Tutorial 7: Limit Theorems

Reminder

Weak Law of Large Numbers Let X_1, \dots, X_n be n random variables with the same distribution and uncorrelated. We assume that these random variables have a finite expectation m and finite variance σ^2 .

The random variable $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ converges in probability to m , i.e.:

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} m$$

Strong Law of Large Numbers Let X_1, \dots, X_n be n independent random variables with the same distribution. We assume that these random variables have a finite expectation m .

The random variable $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ converges almost surely to m , i.e. $\exists \Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0^c) = 0$ such that $\forall \omega \in \Omega_0$:

$$\lim_{n \rightarrow +\infty} \bar{X}_n(\omega) = m$$

Central Limit Theorem (CLT) Reminder:

- 1) *standardization*: if $Y \sim \mathcal{N}(m, \sigma^2)$, then $\frac{Y-m}{\sigma} \sim \mathcal{N}(0, 1)$
- 2) if X_1, \dots, X_n are n independent random variables with the same distribution $\mathcal{N}(m, \sigma^2)$, then $S_n = \sum_{i=1}^n X_i \sim \mathcal{N}(nm, n\sigma^2)$. We can also deduce that:

$$T_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

Central Limit Theorem:

Let X_1, \dots, X_n be n independent random variables with the same distribution, with mean m and variance σ^2 . The random variable $S_n = \sum_{i=1}^n X_i$ satisfies,

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y,$$

where $Y \sim \mathcal{N}(0, 1)$.

Chebyshev's Inequality **Chebyshev's inequality** (concentration inequality) shows a fraction of values of a random variable that are relatively far from its expectation.

Let X be a random variable with finite expectation $m = \mathbb{E}[X]$ and finite variance $\sigma^2 \neq 0$.

$\forall \alpha \in \mathbb{R}, \alpha > 0$:

$$\mathbb{P}(|X - m| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$$

Or another formulation:

$\forall k \in \mathbb{R}, k > 0$, i.e. $\alpha = k\sigma$:

$$\mathbb{P}(|X - m| \geq k\sigma) \leq \frac{1}{k^2}$$

Expectation

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx$$

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

Variance

$$Var(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \int_a^b (x - \mathbb{E}X)^2 f(x)dx$$

- $Var(X + Y) = Var[X] + Var[Y]$, (if X and Y are indep.)
- $Var(aX + b) = a^2Var(X)$

Exercise 1

During the second round of a presidential election, an “exit poll” survey is conducted on a sample of n people. We assume that the responses of respondents are independent, and that respondents do not lie. Moreover, we only consider expressed votes.

1. The polling institute announces the victory of the candidate among A and B who has the largest number of votes among the people polled. What are the probabilities of being wrong in the case where A has 51% of the votes and where n takes the values $n = 100$, $n = 500$, $n = 1000$? What is the minimum value of n for the probability of being wrong to be less than 5%?
2. Among the first 40 voters questioned, 26 voted for A . Is it worth continuing the poll?