Law of Large Numbers & Central Limit Theorem

EE-209 - Eléments de Statistiques pour les Data Sciences

Convergence of random variables

Convergence in Probability (convergence en probabilités)

We say that a sequence of r.v.s X_n converges in probability to X and write $X_n \xrightarrow{\mathbb{P}} X$ if

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0, \qquad \text{for any } \varepsilon > 0.$$

Almost sure convergence (convergence presque sûre)

We say that a sequence of r.v.s X_n converges almost surely to X and write $X_n \xrightarrow{\text{a.s.}} X$ if

$$\mathbb{P}\Big(\lim_{n\to\infty} X_n = X\Big) = 1.$$

Theorem

$$\left(X_n \xrightarrow{\text{a.s.}} X\right) \qquad \underset{\not\Leftarrow}{\Rightarrow} \qquad \left(X_n \xrightarrow{\mathbb{P}} X\right)$$

Strong Law of Large Numbers (SLLN)

If $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}[|f(X_1)|] < \infty$, then, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[f(X_1)].$$

- In particular,
  the **sample mean** $\bar{X}$ converges to the **population mean** or **expectation** $\mathbb{E}[X_1]$.
- If for some identically distributed r.v.s we have $\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow{\mathbb{P}} \mathbb{E}[f(X_1)]$, then we say that there is
  a *weak* law of large numbers.
- The strong law of large numbers implies the weak law of large numbers.

LLN for the empirical mean of i.i.d. Bernoullis $Ber(p)$

We consider

- a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$.
- $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ the fraction of the throws where heads was observed.

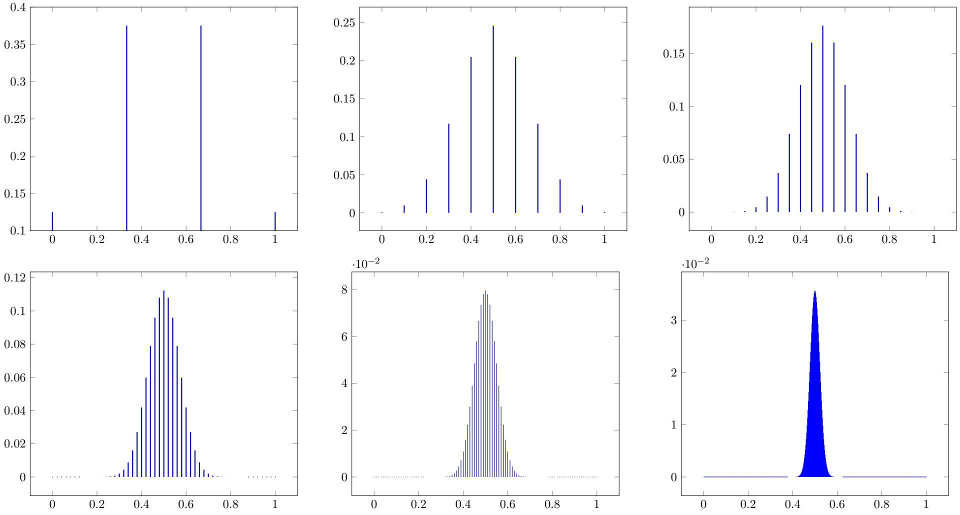
Since $\mathbb{E}[|X_1|] = p < \infty$, by the SLLN, we have

$$\bar{X} \xrightarrow{\text{a.s.}} \mathbb{E}[X_1] = p.$$

Remark:

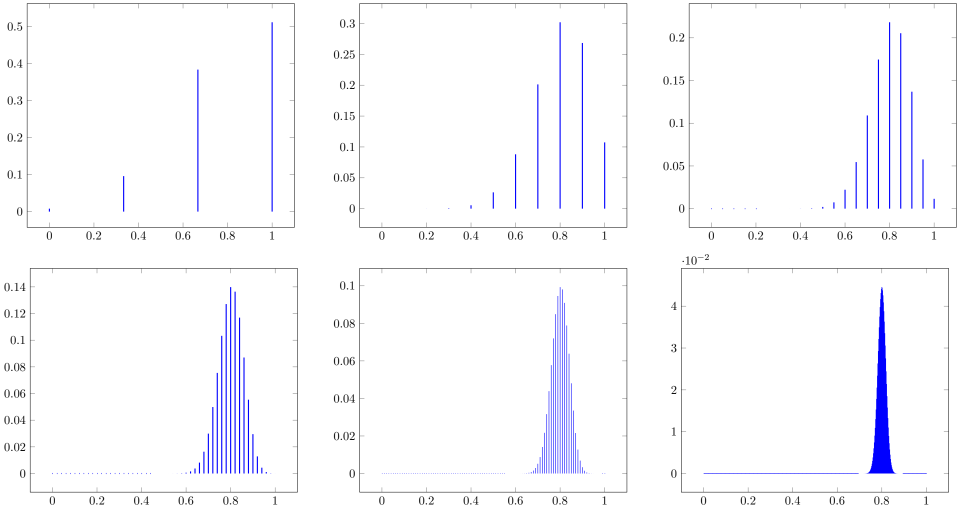
- $N := \sum_{i=1}^n X_i = n\bar{X}_n \sim \text{Bin}(n, p)$
- so $\bar{X}_n = \frac{N}{n}$ is a scaled Binomial r.v.

LLN for the Ber(0.5): pmf of $\bar{X}_n$ for $n \in \{3, 10, 20, 50, 100, 500\}$



We see that the distribution of $\bar{X}_n$ concentrates around $p = 0.5$.

LLN for the Ber(0.8): pmf of $\bar{X}_n$ for $n \in \{3, 10, 20, 50, 100, 500\}$



We see that the distribution of $\bar{X}_n$ concentrates around $p = 0.8$.

Convergence in distribution

Definition

Let $(X_n)_{n\geq 0}$ be a sequence of random variables,

- we say that $(X_n)_{n\geq 0}$ converges in distribution to X
- and we write $X_n \xrightarrow{(d)} X$

if, for each point $x \in \mathbb{R}$ where F_X is continuous, $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_X(x), \qquad \forall x \in \mathbb{R}.$

Equivalent definition

For any finite partition $a_0 = -\infty < a_1 < \ldots < a_K = \infty$,

"the histograms of X_n converge to the histograms of X"

in the sense that for any a_{k-1}, a_k where F_X is continuous,

$$\mathbb{P}(X_n \in [a_{k-1}, a_k]) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X \in [a_{k-1}, a_k]).$$

Central Limit Theorem (CLT)

Theorem

If $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}[f(X_1)^2] < \infty$, then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(f(X_i) - \mu_f\right) \xrightarrow{(d)} \mathcal{N}(0, \sigma_f^2) \quad \text{with} \quad \mu_f = \mathbb{E}[f(X_1)],\ \sigma_f^2 = \mathsf{Var}(f(X_1)).$$

where $\xrightarrow{(d)}$ is the *convergence in distribution.*

In particular,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{(d)} \mathcal{N}(0, 1) \quad \text{with} \quad \mu = \mathbb{E}[X_1],\ \sigma^2 = \mathsf{Var}(X_1).$$

CLT for the empirical mean of i.i.d. Bernoullis $Ber(p)$

We consider again

- a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$.
- $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ the fraction of the throws where heads was observed.

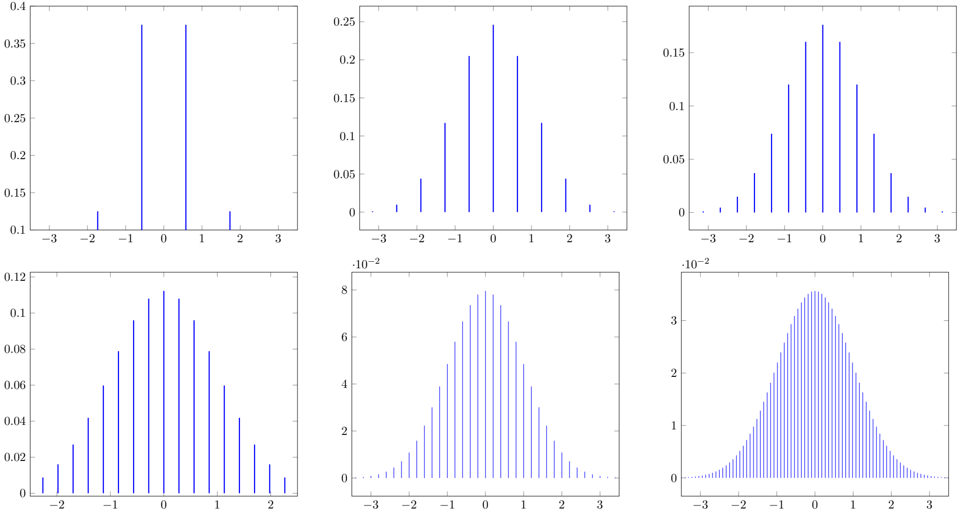
Since $\mathbb{E}[X_1^2] = p < \infty$, by the CLT, we have

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{(d)} \mathcal{N}(0,1) \quad \text{with} \quad \mu = \mathbb{E}[X_1] = p, \quad \sigma^2 = \text{Var}(X_1) = p(1-p).$$

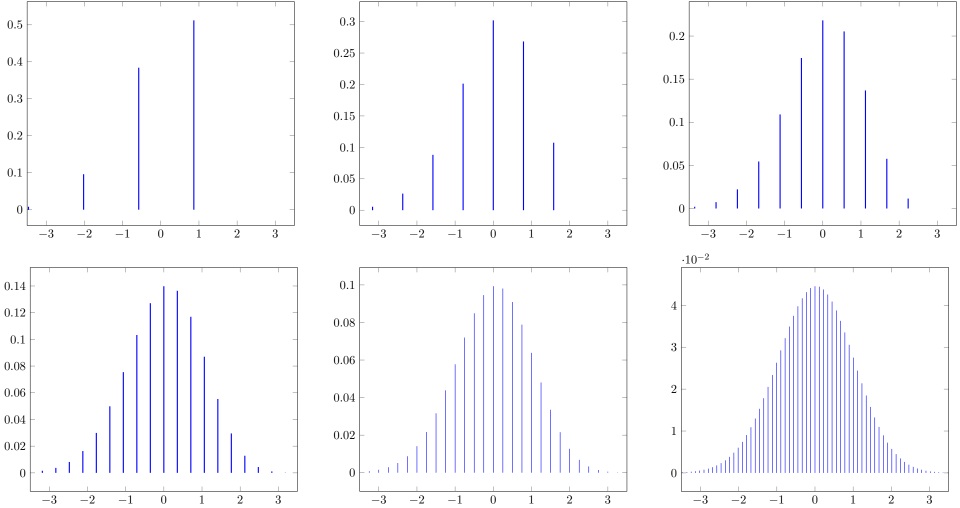
in other words

$$\frac{\sqrt{n}\,(\bar{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{(d)} \mathcal{N}(0,1)$$

CLT for the Ber(0.5): pmf of $\sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p)$ for $N \in \{3, 10, 20, 50, 100, 500\}$



CLT for the Ber(0.8): pmf of $\sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p)$ for $n \in \{3, 10, 20, 50, 100, 500\}$



Example 2: LLN and CLT for the empirical mean of i.i.d. $\mathcal{U}[0,1]$ r.v.s.

We consider

- a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0,1]$.
- $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ the empirical mean.
- we have $\mathbb{E}[|X_1|] = \mathbb{E}[X_1] = \frac{1}{2}$ and $\text{Var}(X_1) = \frac{1}{12} < \infty$.

Since $\mathbb{E}[|X_1|] < \infty$, by the SLLN, we have

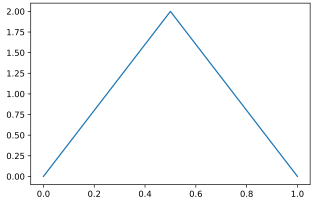
$$\bar{X} \xrightarrow{\text{a.s.}} \mathbb{E}[X_1] = \frac{1}{2}.$$

Since the mean and variance are finite, we have $\mathbb{E}[X_1^2] < \infty$, and by the CLT, we have

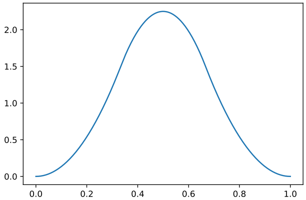
$$\frac{\sqrt{n}\,(\bar{X}_n - 0.5)}{\sqrt{1/12}} \xrightarrow{(d)} \mathcal{N}(0,1).$$

LLN for means of Uniforms $\bar{X}_n$ with $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0,1]$

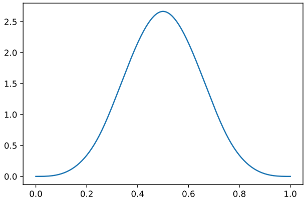
Probability density functions $p_{\bar{X}_n}$ of $\bar{X}_n$:



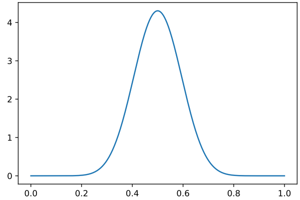
$n = 2$



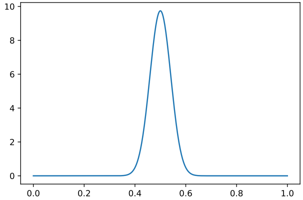
$n = 3$



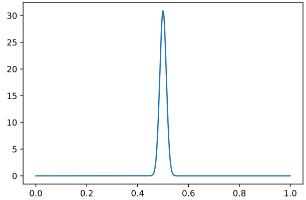
$n = 4$



$n = 10$



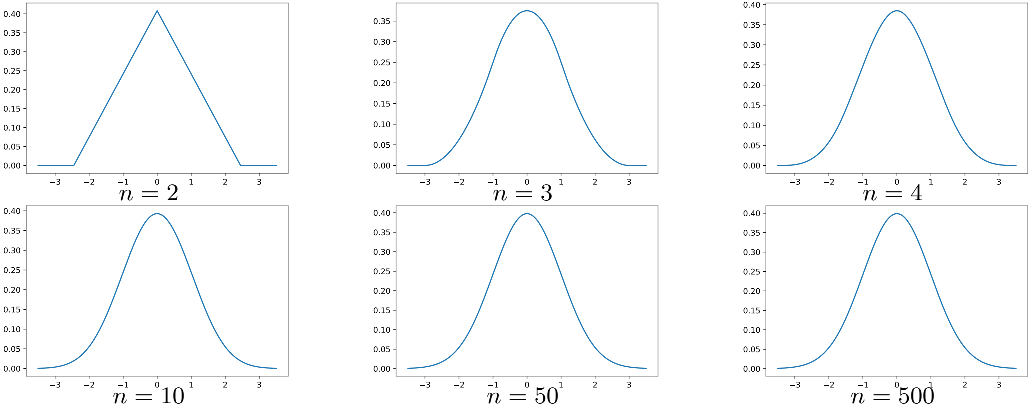
$n = 50$



$n = 500$

CLT$^+$ for *standardized* means $\sqrt{12n}(\bar{X}_n - 0.5)$ with $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0,1]$

Probability density functions $p_{\sqrt{12n}(\bar{X}_n - 0.5)}$ of $\sqrt{12n}(\bar{X}_n - 0.5)$:



Actually, the result seen here is stronger than the CLT because the *pdfs* of $\sqrt{12n}(\bar{X}_n - 0.5)$ become Gaussian (and not only the cdfs).

CLT combined with Slutsky's lemma for the case $\hat{\sigma} \xrightarrow{\mathbb{P}} \sigma$.

We will often use the CLT to know how close $\bar{X}_n$ is from $\mu := \mathbb{E}[X_1]$, but this depends on σ which is typically unknown...

Fortunately, the CLT is still valid if we have an estimate : $\hat{\sigma}$ of σ which converges to it.

Theorem

If $\hat{\sigma} \xrightarrow{\mathbb{P}} \sigma$, then $$\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \xrightarrow{(d)} \mathcal{N}(0,1) \quad \text{with} \quad \mu = \mathbb{E}[X_1],\ \sigma^2 = \mathrm{Var}(X_1).$$

This is guaranteed by a theoretical result called Slutsky's lemma beyond the scope of the course.

Relationship between the TCL and the LLN

- We always have $\mathbb{E}[X^2] \geq \mathbb{E}[|X|]^2$ so that if $\mathbb{E}[X^2] < \infty$ then $\mathbb{E}[|X|] < \infty$ as well.
- So if the conditions to apply the TCL are met then the SLLN applies as well.

Central Limit Theorem: multivariate version

We consider now r.v. $X_i = (X_{i1}, \dots, X_{id})^\top$ taking values in $\mathbb{R}^d$.

Theorem

If $X_1, \dots, X_n$ are i.i.d. with $\mathbb{E}[\|X_1\|^2] < \infty$, then,

$$\sqrt{n}(\bar{X} - \boldsymbol{\mu}) \xrightarrow{(d)} \mathcal{N}(0, \boldsymbol{\Sigma}),$$

with $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)^\top$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ the covariance matrix of X_1 with entry

$$\boldsymbol{\Sigma}_{jk} = \mathrm{cov}(X_{1j}, X_{1k}) = \mathbb{E}[(X_{1j} - \mu_j)(X_{1k} - \mu_k).]$$

Continuous mapping theorem

Theorem

Let $f : \mathbb{R}^d \to \mathbb{R}^k$ is a *continuous function*

$$\text{if} \quad Y_n \xrightarrow{\text{a.s.}} Y \quad \text{then} \quad f(Y_n) \xrightarrow{\text{a.s.}} f(Y)$$

$$\text{if} \quad Y_n \xrightarrow{\mathbb{P}} Y \quad \text{then} \quad f(Y_n) \xrightarrow{\mathbb{P}} f(Y)$$

$$\text{if} \quad Y_n \xrightarrow{(d)} Y \quad \text{then} \quad f(Y_n) \xrightarrow{(d)} f(Y).$$