Probability 1

Chapter 06 : Limit Theorems

Dr. Daniel Flores-Agreda,

(based on the notes of Prof. Davide La Vecchia)

Spring Semester 2021

Outline

Sequences of Random Variables

Definition

A sequence of random variables is an ordered list of random variables of the form

$$S_1, S_{\,2}, ..., S_n, ...$$

where, in an abstract sense, the sequence is infinitely long.

We would like to say something about how these random variables behave as n gets larger and larger (i.e. as n tends towards infinity, denoted by $n \to \infty$)

The study of such limiting behaviour is commonly called a study of ‘asymptotics’ — after the word asymptote used in standard calculus.

Example: Bernoulli Trials and their sum

Let $\tilde{Z}$ denote a dichotomous random variable with $\tilde{Z} \sim Bernoulli(p)$. A sequence of Bernoulli trials provides us with a sequence of values $\tilde{Z}_1, \tilde{Z}_2, ..., \tilde{Z}_n, ...$

$$P(''Success'') = P\left(\tilde{Z}_i = 1\right) = p \quad \text{and} \quad P(''Failure'') = P\left(\tilde{Z}_i = 0\right) = 1 - p$$

Now let

$$S_n = \sum_{s=1}^{n} \tilde{Z}_s,$$

the number of "Successes" in the first n Bernoulli trials. This yields a new sequence of random variables

$$\begin{array}{rcl} S_1 & = & \tilde{Z}_1 \\ S_2 & = & \left(\tilde{Z}_1 + \tilde{Z}_2\right) \\ & & \vdots \\ S_n & = & \left(\tilde{Z}_1 + \tilde{Z}_2 + \cdots + \tilde{Z}_n\right) = \sum_{i=1}^{n} \tilde{Z}_i \end{array}$$

This new sequence is such that $S_n \sim B(n, p)$ for each n.

Example: Bernoulli Trials and their sum

Now consider the sequence

$$P_n = S_n/n,$$

for $n = 1, 2, \ldots$, corresponds to the proportion of 'Successes'in the first n Bernoulli trials.

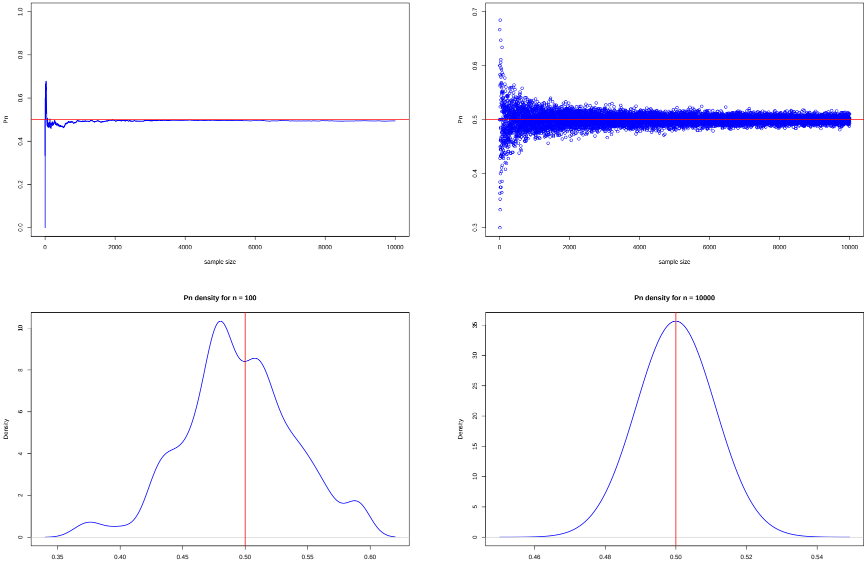
It is natural to ask how the behaviour of P_n is related to the true probability of a 'Success' (p).

Specifically, the open question at this point is:

"Do these results imply that P_n collapses onto the true p as n increases, and if so, in what way?"

To gain a clue, let us consider the simulated values of P_n.

Example: Bernoulli Trials and limit behaviour



Example: Bernoulli Trials and limit behaviour

Remark

This numerical illustration leads us to suspect that there is a sense in which P_n converges to p — notice that although the sequence is random, the 'limiting' value here is a constant (i.e. is non-random).

So, informally, we can claim that a sequence of random variables $X_1, X_2, ..., X_n, ...$ is thought to converge if the probability distribution of X_n gets more and more concentrated around a single point as n tends to infinity.

Convergence in Probability ($\xrightarrow{p}$)

More formally,

Definition

A sequence of random variables $X_1, X_2, ..., X_n, ...$ is said to **converge in probability** to a number α if for any arbitrary constant $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|X_n - \alpha| > \varepsilon\right) = 0$$

If this is the case, we write $X_n \xrightarrow{p} \alpha$ or $p\lim X_n = \alpha$.

A sequence of random variables $X_1, X_2, ..., X_n, ...$ is said to **converge in probability** to a random variable X if for any arbitrary constant $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0\,,$$

written $X_n \xrightarrow{p} X$ or $p\lim(X_n - X) = 0$.

Operational Rules for $\xrightarrow{p}$

Let us itemize some rules. To this end, let a be any (nonrandom) number so:

- If $X_n \xrightarrow{p} \alpha$ then
  - $aX_n \xrightarrow{p} a\alpha$ and
  - $a + X_n \xrightarrow{p} a + \alpha$,
- If $X_n \xrightarrow{p} X$ then
  - $aX_n \xrightarrow{p} aX$ and
  - $a + X_n \xrightarrow{p} a + X$
- If $X_n \xrightarrow{p} \alpha$ and $Y_n \xrightarrow{p} \gamma$ then
  - $X_nY_n \xrightarrow{p} \alpha\gamma$ and
  - $X_n + Y_n \xrightarrow{p} \alpha + \gamma$.

Operational Rules for $\xrightarrow{p}$

- If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ then
  - $X_n Y_n \xrightarrow{p} XY$ and
  - $X_n + Y_n \xrightarrow{p} X + Y$

- Let $g(x)$ be any (non-random) continuous function. If $X_n \xrightarrow{p} \alpha$ then

$$g(X_n) \xrightarrow{p} g(\alpha),$$

and if $X_n \xrightarrow{p} X$ then

$$g(X_n) \xrightarrow{p} g(X).$$

Convergence of Sample Moments as a motivation...

Suppose $X_1, X_2, ..., X_n, ...$ is a sequence of *i.i.d.* random variables with common distribution $F_X(x)$ and moments $\mu_r = E[X^r]$. At any given point along the sequence, $X_1, X_2, ..., X_n$ constitutes a simple random sample of size n.
For each fixed sample size n, the rth sample moment is (using an obvious notation)

$$M_{(r,n)} = \frac{1}{n}\left(X_1^r + X_2^r + \cdots + X_n^r\right) = \frac{1}{n}\sum_{s=1}^{n} X_s^r\,,$$

and we know that

$$E[M_{(r,n)}] = \mu_r \quad \text{and} \quad Var(M_{(r,n)}) = \frac{1}{n}(\mu_2 - \mu_1^2)\,.$$

Now consider the sequence of sample moments $M_{(r,1)}, M_{(r,2)}, ..., M_{(r,n)}, ...$ or, equivalently, $\{M_{(r,i)}\}_{i=1}^{n}$.

Convergence of Sample Moments as a motivation...

The distribution of $M_{(r,n)}$ (which is unknown because $F_X(x)$ has not been specified) is thus concentrated around μ_r for all n, with a variance which tends to zero as n increases.

So the distribution of $M_{(r,n)}$ becomes more and more concentrated around μ_r as n increases and therefore we might *anticipate* that

$$M_{(r,n)} \xrightarrow{p} \mu_r.$$

In fact, this result follows from what is known as the **Weak Law of Large Numbers** (WLLN).

The Weak Law of Large Numbers (WLLN)

Proposition

Let $X_1, X_2, ..., X_n, ...$ be a sequence of i.i.d. random variables with common probability distribution $F_X(x)$, and let $Y = h(X)$ be such that

$$\begin{aligned} E[Y] = E\left[h(X)\right] &= \mu_Y \\ Var(Y) = Var\left(h(X)\right) &= \sigma_Y^2 < \infty\,. \end{aligned}$$

Set

$$\overline{Y}_n = \frac{1}{n}\sum_{s=1}^{n} Y_s \quad \textit{where} \quad Y_s = h(X_s)\,, \quad s = 1, \dots, n\,.$$

Then for any two numbers ε and δ satisfying $\varepsilon > 0$ and $0 < \delta < 1$

$$P\left(\left|\overline{Y}_n - \mu_Y\right| < \varepsilon\right) \geq 1 - \delta$$

for all $n > \sigma_Y^2/(\varepsilon^2\delta)$. Choosing both ε and δ to be arbitrarily small implies that $p\lim_{n\to\infty}(\overline{Y}_n - \mu_Y) = 0$, or equivalently $\overline{Y}_n \overset{p}{\to} \mu_Y$.

The WLLN and Chebyshev's Inequality

- First note that $E[\overline{Y}_n] = \mu_Y$ and $Var(\overline{Y}_n) = \sigma_Y^2/n$.
- Now, according to **Chebyshev's inequality**

$$\begin{aligned} P\left(|\overline{Y}_n - \mu_Y| < \varepsilon\right) &\geq 1 - \frac{E\left[\left(\overline{Y}_n - \mu_Y\right)^2\right]}{\varepsilon^2} \\ &= 1 - \frac{\sigma_Y^2/n}{\varepsilon^2} \\ &= 1 - \frac{\sigma_Y^2}{n\varepsilon^2} \geq 1 - \delta \end{aligned}$$

for all $n > \sigma_Y^2/(\varepsilon^2\delta)$.

- Thus the WLLN is proven, provided we can verify **Chebyshev's inequality**.
- Note that by considering the limit as $n \to \infty$ we also have

$$\lim_{n\to\infty} P\left(\left|\overline{Y}_n - \mu_Y\right| < \varepsilon\right) \geq \lim_{n\to\infty}\left(1 - \frac{\sigma^2}{n\varepsilon^2}\right) = 1\,,$$

again implying that $\left(\overline{Y}_n - \mu_Y\right) \xrightarrow{p} 0$.

Chebyshev's (and Markov's) Inequality

- **Chebychev's Inequality**: For any random variable Z with mean μ_Z and variance $\sigma_Z^2 < \infty$

  $$P\left(|Z - \mu_Z| < r\sigma_Z\right) \geq 1 - \frac{1}{r^2}$$

  for all $r > 0$.

  - Note that an equivalent expression is given by

    $$P\left(|Z - \mu_Z| \geq r\sigma_Z\right) \leq \frac{1}{r^2} \tag{1}$$

  - This inequality says that the probability that a random variable lies more than r standard deviations away from its mean value is bounded above by $1/r^2$.
- Chebychev's inequality is, in turn, a special case of Markov's inequality.
- **Markov's inequality**: Let Z be random variable and $h(z)$ a non-negative valued function for all $z \in \mathbb{R}$. Then

  $$P(h(Z) \geq \zeta) \leq \frac{E[h(Z)]}{\zeta} \quad \text{for all}\, \zeta > 0. \tag{2}$$

Chebyshev's (and Markov's) Inequality

- To verify Markov's inequality, observe that

$$
\begin{aligned}
E[h(Z)] &= \int_{-\infty}^{\infty} h(z) f_Z(z)\, dz \\
&= \int_{\{z : h(z) \geq \zeta\}} h(z) f_Z(z)\, dz + \int_{\{z : h(z) < \zeta\}} h(z) f_Z(z)\, dz \\
&\geq \int_{\{z : h(z) \geq \zeta\}} h(z) f_Z(z)\, dz \\
&\geq \int_{\{z : h(z) \geq \zeta\}} \zeta f_Z(z)\, dz = \zeta P(h(Z) \geq \zeta)\,,
\end{aligned}
$$

  giving the desired result on division by ζ.

- Chebyshev's inequality now follows as a direct corollary of Markov's inequality on taking $h(z) = (z - \mu_Z)^2$ and $\zeta = r^2\sigma_Z^2$. It can be used to construct crude bounds on the probabilities associated with deviations of a random variable from its mean.

Example: Markov's Inequality

Example

Q. On the A2 highway (in the Luzern Canton), the speed limit is 80 Km/h. Most drivers are not driving so fast and the average speed on the high way is 70 Km/h. If Z denotes a randomly chosen driver's speed, what is the probability that such a person is driving faster than the speed limit?

A. Since we do not have the whole distribution of Z, but we have only limited info (i.e. we know $E[Z] = 70$ Km/h), we have to resort on Markov's inequality. So using (2) we obtain an upper bound to the probability:

$$P(Z \geq 80) \leq \frac{70}{80} = 0.875.$$

Example: Chebyshev's Inequality

Example

Q. On the A2 highway (in the Luzern Canton), the speed limit is 80 Km/h. Most drivers are not driving so fast and the average speed on the high way is 70 Km/h, **with variance** 9 $(Km/h)^2$. If Z denotes a randomly chosen driver's speed, what is the probability that such a person is driving faster than the speed limit?

A. Since we do not have the whole distribution of Z, but we have only limited info (i.e. we know $E[Z] = 70$ Km/h **AND** $V(Z) = 9\ (Km/h)^2$), we have to resort on Chebyshev's inequality and give an upper bound to the probability. Thus,

$$\begin{aligned} P(Z \geq 80) &= P(Z - E[Z] \geq 80 - 70) \\ &\leq P(|Z - E[Z]| \geq 10) \leq P\left(\frac{|Z - E[Z]|}{\sqrt{V(Z)}} \geq \frac{10}{\sqrt{9}}\right) \end{aligned}$$

Using (1), with $r = \frac{10}{3}$ and $\sigma_Z = 3$, we finally get

$$P(Z \geq 80) \leq P\left(\left|Z - E[Z]\right| \geq \left(\frac{10}{3}\right)3\right) \leq \frac{1}{\frac{10^2}{3^2}} \leq \frac{9}{100} \leq 0.09$$

A remark about Chebyshev's Inequality

Remark

Chebichev's inequality can be rewirtten in a different way.
Indeed, for any random variable Z with mean μ_Z and variance $\sigma_Z^2 < \infty$

$$P\left(|Z-\mu_Z| \geq \varepsilon\right) \leq \frac{E[Z-\mu_Z]^2}{\varepsilon^2} = \frac{\sigma_Z^2}{\varepsilon^2}. \tag{3}$$

It's easy to check that Eq. (3) coincides with Eq. (1), setting in Eq. (3)

$$\varepsilon = r\sigma_Z.$$

Do the check as an exercise!!

Convergence in Distribution

Definition

Consider, therefore, a sequence of random variables $X_1, X_2, ..., X_n, ...$ with corresponding CDFs $F_{X_1}(x), F_{X_2}(x), ..., F_{X_n}(x), ...$ We say that the sequence $X_1, X_2, ..., X_n, ...$ **converges in distribution** to the random variable X, having probability distribution $F_X(x)$, if and only if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous. In this case we write $X_n \xrightarrow{D} X$

Some Operational Rules for $\stackrel{D}{\rightarrow}$

- If $p\lim_{n\to\infty}(X_n - X) = 0$ then $X_n \stackrel{D}{\rightarrow} X$.
- Let a be any real number. If $X_n \stackrel{D}{\rightarrow} X$, then $aX_n \stackrel{D}{\rightarrow} aX$
- If $Y_n \stackrel{p}{\rightarrow} \phi$ and $X_n \stackrel{D}{\rightarrow} X$, then
  - $Y_n X_n \stackrel{D}{\rightarrow} \phi X$, and
  - $Y_n + X_n \stackrel{D}{\rightarrow} \phi + X$
- If $X_n \stackrel{D}{\rightarrow} X$ and $g\left(x\right)$ is any continuous function, then $g\left(X_n\right) \stackrel{D}{\rightarrow} g\left(X\right)$

Examples: Poisson and Normal Approximations to the Binomial Distribution

Example

Suppose $X_1, X_2, ..., X_n, ...$ is a sequence of independent random variables where $X_n \sim B(n, p)$ with probability of "Success" p.

- We already know that, if $p = \lambda/n$, where $\lambda > 0$ is fixed, then as n goes to infinity, $F_{X_n}(x)$ converges to the probability distribution of a *Poisson* (λ) random variable. So, $X_n \overset{D}{\to} X$, where $X \sim Poisson(\lambda)$

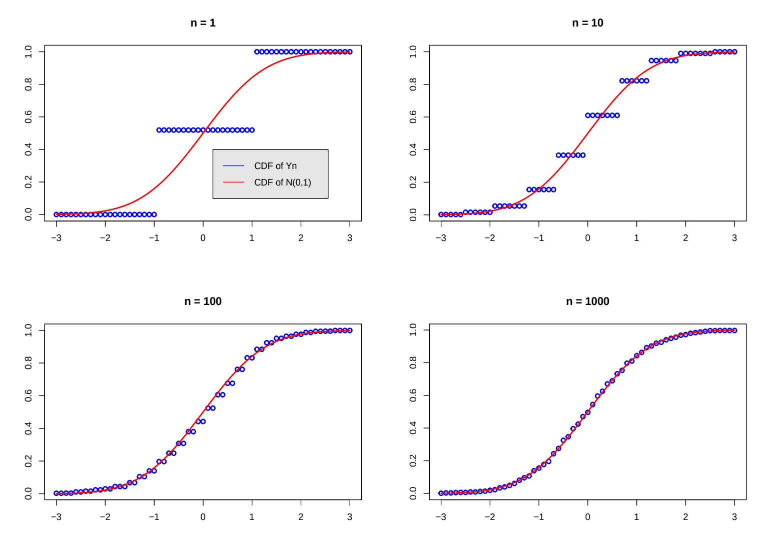
- Now consider another case. If p is fixed, the probability distribution of

$$Y_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

converges, as n goes to infinity, to that of a standard Normal random variable [Theorem of De Moivre-Laplace]. So, $Y_n \overset{D}{\to} Y$, where $Y \sim \mathcal{N}(0, 1)$.

Example cont'd (visualize Y_n)

Example



Example [convergence to an exp r.v.]

Example

Let us consider a sequence of continuous r.v.'s $X_1, X_2, ..., X_n, ...$, where X_n has range $(0, n]$, for $n > 0$ and CDF

$$F_{X_n}(x) = 1 - \left(1 - \frac{x}{n}\right)^n, \quad 0 < x \leq n.$$

Then, as $n \to \infty$, the limiting support is $(0, \infty)$, and $\forall x > 0$, we have

$$F_{X_n}(x) \to F_X(x) = 1 - e^{-x}$$

which is the CDF of an exponential r.v. (at all continuity points).

So, we conclude that X_n convergences in distribution to an exponential r.v., that is

$$X_n \overset{D}{\to} X, \quad X \sim \exp(1).$$

The Central Limit Theorem (CLT)

The following theorem is often said to be one of the most important results. Its significance lies in the fact that it allows accurate probability calculations to be made without knowledge of the underlying distributions!

Theorem

Let $X_1, X_2, ..., X_n, ...$ be a sequence of i.i.d. random variables and let $Y = h(X)$ be such that

$$\begin{aligned} E[Y] = E\left[h(X)\right] &= \mu_Y \\ Var(Y) = Var\left(h(X)\right) &= \sigma_Y^2 < \infty \,. \end{aligned}$$

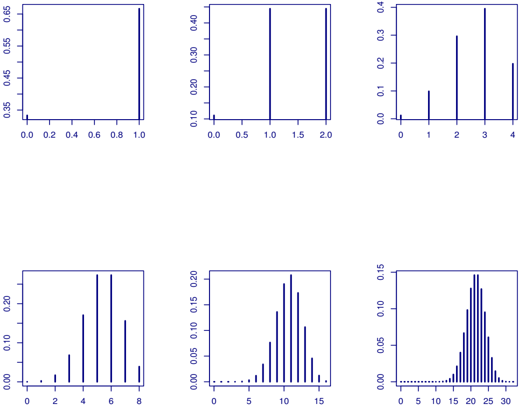
Set

$$\overline{Y}_n = \frac{1}{n}\sum_{s=1}^{n} Y_s \quad \text{where} \quad Y_s = h(X_s)\,, \quad s = 1, \ldots, n\,.$$

Then (under quite general regularity conditions)

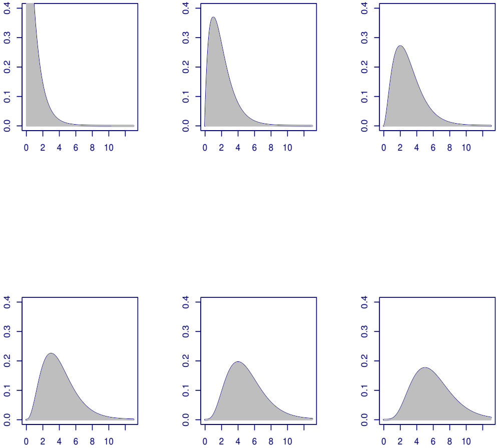
$$\frac{\sqrt{n}\left(\overline{Y}_n - \mu_Y\right)}{\sigma_Y} \overset{D}{\to} N\left(0, 1\right).$$

CLT: sum of Bernoulli rvs (from Kuonen)



Distributions de la somme de 1, 2, 4, 8, 16 *et* 32 *variables aléatoires indépendantes de Bernoulli avec* $p = 2/3$, *i.e.* $B(1, 2/3)$.

CLT: sum of Exp rvs (from Kuonen)



Distributions de la somme de 1, 2, 3, 4, 5 *et* 6 *variables aléatoires indépendantes exponentielles avec* $\lambda = 1$, *i.e.* $\mathrm{Exp}(1)$.

The Central Limit Theorem (CLT)

Remark

Several generalizations of this statement are available. For instance, one can state a CLT for data which are independent but NOT identically distributed. Another possibility is to define a CLT for data which are NOT independent, namely for dependent data — for this you need to attend my course about Time Series, in the Fall semester at the Master in Statistics at University of Geneva !!!!