Probabilities and Statistics - Lecture 7

Olariu E. Florentin

March 30 , 2026

Table of contents

The central limit theorem

Theorem 1.1

(The central limit theorem, Lindeberg-Lévy) Let $(X_n)_{n\geqslant 1}$ be a sequence of identically distributed and independent random variables having mean μ and variance σ^2. Then

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1) \text{ or}$$

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leqslant a\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \exp\left(-t^2/2\right) dt, \forall a \in \mathbb{R}.$$

The central limit theorem

- The central limit theorem allows to estimate probabilities for sum of independent random variables.
- On the other hand, the theorem explains why so many processes (from social sciences, biology, psichology etc) follow the normal law.
- Essentially the central limit theorem says that, for large samples ($n \geqslant 30$), the variable

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

behaves like a standard normal law, $N(0,1)$.
- The central limit theorem holds even for dependent variables, if their correlation is very small.

Normal approximation to the binomial distribution

- Let X_n be a sequence of Bernoulli(p) independent variables.
- $X = \sum_{i=1}^{n} X_i$ is a binomial distributed variable, $B(n,p)$.
- Using the central limit theorem we get the de Moivre-Laplace theorem which says that for large values of n the variable

$$Y = \frac{X - \mathbb{E}[X]}{\sqrt{Var[X]}} = \frac{X - np}{\sqrt{np(1-p)}}$$

  is a standard normal variable ($N(0,1)$).
- The estimation is good for $np(1-p) \geqslant 10$.

Normal approximation to the binomial distribution

Theorem 2.1

(de Moivre-Laplace theorem) When k is around np, as n grows large we have

$$\binom{n}{k} p^k (1-p)^{n-k} \sim \frac{\exp -\frac{(k-np)^2}{2np(1-p)}}{\sqrt{2\pi np(1-p)}}.$$

- Consider the following example: let X be the number of tail occurrences in 40 flippings of a fair coin.
- Compute $P(X = 20)$.

$$P(X = 20) = P(19.5 \leqslant X \leqslant 20.5) =$$

$$= P\left(\frac{19.5 - 20}{\sqrt{10}} \leqslant \frac{X - 20}{\sqrt{10}} \leqslant \frac{20.5 - 20}{\sqrt{10}}\right) =$$

The continuity correction

$$P\left(-0.16 \leqslant \frac{X-20}{\sqrt{10}} \leqslant 0.16\right) \sim$$

$$\sim \Phi(0.16) - \Phi(-0.16) = 0.1272,$$

where $\Phi(\cdot)$ is the distribution function of $N(0,1)$.

- *Continuity correction* is an adjustment that is made whenever a discrete distribution is approximated by a continuous one.
- $P(X = 10) = P(9.5 \leqslant X \leqslant 10.5)$, $P(X > 15) = P(X \geqslant 15.5)$, $P(X < 13) = P(X \leqslant 12.5)$.

Generate uniform random numbers

- When we talk about random numbers we often understand values of an uniform random variable.
- There are two types of uniform random variables: discrete and continuous.
- For example in order to choose uniformly at random an integer between 1 and n (sometimes between 0 and $(n-1)$) we have to generate a value of a discrete random variable U_n.
- On the other hand for choosing uniformly at random a number in $[0,1]$ we have to generate a value of a continuous random variable $U_{[0,1]}$.
- Generally speaking, *to simulate a certain random variable means to generate values that follow its distribution*.

Generate random numbers

- Almost every programming language has random number generators of both types; we will use the random number generators from R.
- We review the R commands for commonly employed discrete and continuous distributions.
- Functions that start with p, q, d and r give the (cumulative) distribution function - CDF, the inverse of CDF, the probability density function - PDF, and (a value of a) a random variable having the specified distribution, respectively.
- For generating discrete uniform random numbers one can use the $sample()$ function.

Generate random numbers

Distribution	Commands			
Binomial	$pbinom()$	$qbinom()$	$dbinom()$	$rbinom()$
Geometric	$pgeom()$	$qgeom()$	$dgeom()$	$rgeom()$
Poisson	$ppois()$	$qpois()$	$dpois()$	$rpois()$
Uniform	$punif()$	$qunif()$	$dunif()$	$runif()$
Exponential	$pexp()$	$qexp()$	$dexp()$	$rexp()$
Normal	$pnorm()$	$qnorm()$	$dnorm()$	$rnorm()$
Student	$pt()$	$qt()$	$dt()$	$rt()$
Gamma	$pgamma()$	$qgamma()$	$dgamma()$	$rgamma()$

- You can find details about all these function using $help(name)$ in R or RStudio.

Generate random numbers

- In order to simulate a discrete random variable all we need to know is its distribution.

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}$$

- We simulate X like follows: we generate an uniform random number U and return x_i if $\sum_{j=1}^{i-1} p_j \leqslant U < \sum_{j=1}^{i} p_j$.

Illustrations of LLN and CLT

Example 1. (*LLN - Buffon's needle problem*) The problem (stated in 1733 and first solved in 1777 by french naturalist and mathematician Comte de Buffon) asks to find the probability that a needle of length l will cross a line, given a straight surface with equally spaced parallel lines at distance $2d$.

Suppose that the needle length is less than the distance between the lines (the easiest situation to analyse); there are two variables that determine the relative position of the needle to the closest line: the angle, x, at which the needle falls and the distance from the middle of the needle to this (closest) line, y.

The needle will cross the closest line if and only if $y \leqslant l/2 \sin x$, for every $x \in [0, \pi]$.

Illustrations of LLN and CLT

All the cases are completely described by the pairs $(x, y) \in [0, \pi] \times [0, d]$, and the favorable cases are the pairs belonging to the area under the graph of the function $f : [0, \pi] \to \mathbb{R}$, $f(x) = l/2 \sin x$.
Thus, the probability is

$$\frac{\int_0^\pi f(x)\,dx}{\pi \cdot d} = \frac{1}{\pi d} \int_0^\pi \frac{l}{2} \sin x \; dx = \frac{l}{2\pi d} \left[-\cos x \right]_0^\pi = \frac{l}{\pi d}.$$

For $l = d = 1$, that is the needle length is half of the distance between the lines, the probability is $1/\pi$.

Illustrations of LLN and CLT

Introduce the random experiment of launching the needle and define a Bernoulli variable, X, with value 1 if and only if the needle cross a line; the probability of success and the expectation of X is $1/\pi$.
If we independently repeat n times this experiment we will get an n size sample $(X_i)_{i=\overline{1,n}}$. Because of the Law of Large Numbers $\overline{x}_n \to 1/\pi$, thus, for large enough values of n,

$$\overline{x}_n = \frac{\text{number of successes}}{n} \approx \frac{1}{\pi}.$$

This kind of relation could be used to obtain an experimental approximation of π. Several needle casters already performed this experiment.

Illustrations of LLN and CLT

Example 2. (*Verifying LLN*) Consider a given probability distribution, X, having mean μ and variance σ^2, and a sequence of n independent identical distributed random variables X_i, $i = \overline{1, n}$. The Law of Large Numbers says that in a certain probabilistic sense, the sample mean converges to the known mean:

$$\overline{x}_n \to \mu \text{ as } n \to \infty$$

Let us verify this law using the Poisson distribution with different λ parameters (for a *Poisson* distribution $\mu = \lambda$).

λ	2	3	4	6	8	12	15
$\overline{x}_n$	1.955	2.977	4.003	6.027	8.018	12.093	14.925

We observe that the resulted statistics ($n = 5000$) are very close to the corresponding known expectations. (The samples are obtained using $rpois(n, \lambda)$.)

Illustrations of LLN and CLT

If we repeat the former test with Gamma distribution for different pairs of (α, λ) parameters (the expectation is $\mu = \alpha/\lambda$) we get

α	2	2	3	4	6	6	6	12
λ	1.5	2	2	3	5	4	8	4
$\overline{x}_n$	1.361	1.009	1.489	1.345	1.204	1.501	0.752	2.973
μ	1.333	1.000	1.500	1.333	1.200	1.500	0.750	3.000

Again, the resulted sample means ($n = 5000$) are very close to the corresponding expectations. (The samples are obtained using $rgamma(n, \alpha, \lambda)$.)

Illustrations of LLN and CLT

Example 3. (*CLT - de Moivre-Laplace*) The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that, on the average, only 30% of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that at least 151 first-year students attend this college.

Solution. Let X be the number of students that attend; we can assume that each accepted applicant will independently attend. Thus $X : B(450, 0.3)$ and

$$P(X > 150) = P(X \geqslant 150.5) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \geqslant \frac{150.5 - np}{\sqrt{np(1-p)}}\right) =$$

$$= P\left(\frac{X - 135}{\sqrt{81}} \geqslant \frac{15.5}{\sqrt{81}}\right) \approx P(Z \geqslant 1.722)$$

where $Z : N(0,1)$. Hence $P(X \geqslant 150) \approx 1 - pnorm(1.722) = 0.0425$.

Illustrations of LLN and CLT

Example 4. (*CLT*) The weights of a population of workers have mean 167 and standard deviation 27. If a sample of 36 workers is chosen, what is the probability that the sample mean of their weights lies between 163 and 170?

Solution. Let us denote by $\overline{x}_n$ the sample mean, from CLT, $\dfrac{\overline{x}_n - \mu}{\sigma/\sqrt{n}}$ approximately follows a standard normal distribution, therefore

$$P(163 \leqslant \overline{x}_n \leqslant 170) = P\left(\frac{163-167}{4.5} \leqslant \frac{\overline{x}_n - 167}{4.5} \leqslant \frac{170-167}{4.5}\right) =$$

$$= P\left(-0.888 \leqslant \frac{\overline{x}_n - 167}{4.5} \leqslant 0.888\right) \approx P(-0.888 \leqslant Z \leqslant 0.888) =$$

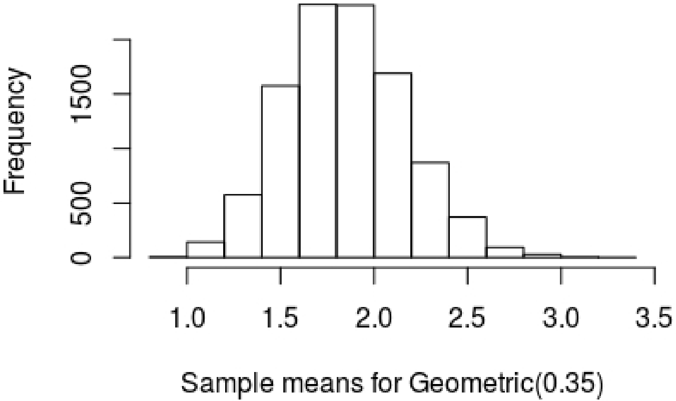
$$= pnorm(0.888) - pnorm(-0.888) = 2 \cdot pnorm(0.888) - 1 = 0.625$$

Illustrations of LLN and CLT

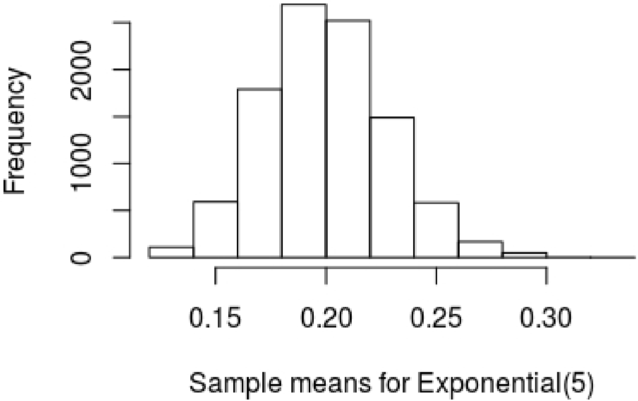
Example 5. (*Verifying CLT*) Consider a given probability distribution, X, with mean μ and variance σ^2, and a sequence of n independent identically distributed (with X) random variables X_i, $i = \overline{1, n}$. According to CLT, for large n, the sample mean, $\overline{x}_n$, has a normal distribution, $N(\mu, \sigma^2/n)$.

We want to check this assertion and take N such sample means and build a histogram. For our examples we used the geometric distribution $G(0.35)$ and the Exponential distribution $Exp(5)$ ($n = 50$, $N = 10000$).

Illustrations of LLN and CLT



Illustrations of LLN and CLT



Illustrations of LLN and CLT

Example 6. (*Verifying CLT*) Consider a given probability distribution, X, having mean μ and variance σ^2, and a sequence of n independent identical distributed random variables X_i, $i = \overline{1,n}$. This sequence can always be viewed as a sample; if $\overline{x}_n$ is the sample mean, CLT says that

$$\lim_{n\to\infty} P\left[\frac{\overline{x}_n - \mu}{\sigma/\sqrt{n}} \leqslant z\right] = P(Z \leqslant z),$$

where $Z : N(0,1)$. Usually, for large values of n we can make the following approximation

$$P_n(z) = P\left[\frac{\overline{x}_n - \mu}{\sigma/\sqrt{n}} \leqslant z\right] \approx P(Z \leqslant z).$$

A method to verify if this approximation is a good one: choose independently N samples (sequences) $\left(X_i^k\right)_{i=\overline{1,n}}^{k=\overline{1,N}}$, and compute

$$P^N = \frac{\left|\{k : \overline{x}_n^k \leqslant z\sigma/\sqrt{n} + \mu\}\right|}{N}.$$

Illustrations of LLN and CLT

That is, P^N is the number of samples (from those N) satisfying the inequality $\frac{\overline{x}_n - \mu}{\sigma/\sqrt{n}} \leqslant z$ over the number of samples. This statistic should approximate $P[Z \leqslant z]$. For Exponential distribution with $\lambda = 2$, $n = 50$, and $N = 2000$ the results are tabulated below (a sample of size n can be obtained with $rexp(n, \lambda)$).

z	-1.5	-1.0	-0.5	0	0.5	1.0	1.5
$P^N(z)$	0.055	0.154	0.313	0.509	0.723	0.831	0.931
$Rel.err$	16%	2.5%	1.6%	1.8%	4.6%	1.8%	0.2%
$pnorm(z)$	0.066	0.158	0.308	0.5	0.691	0.847	0.933

The relative error is equal with $\frac{|P(Z \leqslant z) - P^N(z)|}{P(Z \leqslant z)}$.

Illustrations of LLN and CLT

For computing $P^N(z)$ we used the following algorithm

$\mu \leftarrow 1/\lambda;$

$\sigma \leftarrow 1/\lambda;$ // why?

$c \leftarrow z * \sigma/\sqrt{n} + \mu;$

$j \leftarrow 1;$

for$(i = \overline{1, N})$

    if$(mean(rexp(n, \lambda)) \leqslant c)$

        j++;

return $j/N;$

End

Bibliography I

- Baron, M., *Probability and Statistics for Computer Scientist*, Chapman & Hall/CRC Press, 2013 or the electronic edition https://ww2.ii.uj.edu.pl/~z1099839/naukowe/RP/rps-michael-byron.pdf
- Johnosn, J. L., *Probability and Statistics for Computer Science*, Wiley Interscience, 2008.
- Lipschutz, S., *Theory and Problems of Probability*, Schaum's Outline Series, McGraw-Hill, 1965.
- Ross, S. M., *A First Course in Probability* , Prentice Hall, 5th edition, 1998.
- Shao, J., *Mathematical Statistics*, Springer Verlag, 1998.

Bibliography II

Stone, C. J., *A Course in Probability and Statistics*, Duxbury Press, 1996.