Probabilities and Statistics - Lecture 8

Olariu E. Florentin

April 22, 2024

Table of contents

Random Events

- When $|\Omega| \geqslant |\mathbb{R}|$ (i.e., Ω has, at least, a continuous cardinal), random events are defined in a different manner.
- The most notably difference in definition is that it is possible to have subsets $A \subseteq \Omega$ that are not random events: the random events family forms a σ-algeba $\mathcal{A} \subseteq \mathcal{P}(\Omega)$:
  - $\varnothing, \Omega \in \mathcal{A}$;
  - if $A_1, A_2 \in \mathcal{A}$, then $A_1 \cap A_2 \in \mathcal{A}$;
  - if $(A_n)_{n \geqslant 1} \subseteq \mathcal{A}$, then $\bigcup_{n \geqslant 1} A_n \in \mathcal{A}$.
- The probability function is defined only on $\mathcal{A}$ (with known axioms):

$$P : \mathcal{A} \to [0, 1].$$

Continuous Random Variables

- A function $X : \Omega \to \mathbb{R}$ is called random variable if

$$\text{for every interval } J \subseteq \overline{\mathbb{R}},\ X^{-1}(J) \in \mathcal{A}.$$

- A random variable $X : \Omega \to \mathbb{R}$ is called continuous if its distribution function is a continuous one (*Sometimes, this definition addresses the situation when $X(\Omega)$ has a continuous cardinal*).
- The distribution of such a variable is given by its distribution function:

$$F : \mathbb{R} \to [0, 1], F(a) = P(X \leqslant a),$$

- or by its probability density function, $f : \mathbb{R} \to [0, +\infty)$, such that F can be described like follows

$$F(a) = P(X \leqslant a) = \int_{-\infty}^{a} f(t)\, dt.$$

Continuous Random Variables

- Any function $f : \mathbb{R} \to [0, +\infty)$, such that $\int_{-\infty}^{\infty} f(t)\, dt = 1$, is the density function for a certain continuous random variable (or simply a continuous distribution).
- Using the probability density function we can compute (if the integrals exist) the expectation and the variance:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} tf(t)\, dt \text{ and } Var[X] = \int_{-\infty}^{+\infty} (t - \mathbb{E}[X])^2 f(t)\, dt.$$

- If $h : \mathbb{R} \to \mathbb{R}$ is a real (say, continuous) function, and X is a random variable with the density f, then $h(X)$ is a random variable having the following expected value

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(t) f(t)\, dt.$$

Continuous Random Variables

- The associated probabilities are computed like this

$$P(a < X \leqslant b) = F(b) - F(a) = \int_a^b f(t)\,dt$$

which is the area under the graph of f between $t = a$ and $t = b$.

- If F is continuous, $P(X = a) = 0$ and $P(a \leqslant X < b) = P(a < X \leqslant b) = P(a < X < b)$.
- For a given random variable $X : \Omega \to \mathbb{R}$, standardization consists in the following transformation of X:

$$Y = \frac{X - \mathbb{E}[X]}{StDev[X]}.$$

- The new variable is "standard", that is,

$$\mathbb{E}[Y] = 0 \text{ and } Var[Y] = 1.$$

Continuous Random Variables - Examples

Example 1. The life time in years, of some electronic component is a continuous random variable with the density

$$f(x) = \begin{cases} \dfrac{k}{x^4}, & x \geqslant 1 \\ 0, & x < 1 \end{cases}$$

Find k, its distribution function, and the probability for life-time to exceed 2 years.

Solution. We must have $f(t) \geqslant 0, \forall t \in \mathbb{R}$ and $\int\limits_{-\infty}^{\infty} f(t)\, dt = 1$, therefore

$k \geqslant 0$ and $1 = \int\limits_{1}^{\infty} \frac{k}{t^4}\, dt = \left[-\frac{k}{3t^3} \right]_1^{\infty} = \frac{k}{3}$ which gives $k = 3$.

Continuous Random Variables - Examples

The distribution function is $F(x) = \int_{-\infty}^{x} f(t)\,dt$, therefore

$$F(x) = \begin{cases} 1 - \frac{1}{x^3}, & x \geqslant 1 \\ 0, & x < 1 \end{cases}$$

Let X be the life time of this electronic component, the probability that the life-time exceeds 2 years is $P(X \geqslant 2) = 1 - P(X < 2) = 1 - F(2) = 1/8$ (because F is continuous).

Example 2. Let X be a continuous random variable with the following density function

$$f(x) = \begin{cases} \alpha x, & 0 \leqslant x \leqslant 2 \\ 0, & \text{otherwise} \end{cases}$$

Find α, its distribution function, the expectation and the variance of X.

Continuous Random Variables - Examples

Example 3. The time in minutes, it takes to reboot a certain system is a continuous variable with the density

$$f(t) = \begin{cases} C(10-x)^2, & 0 < x < 10 \\ 0, & \text{otherwise} \end{cases}$$

Compute C and the probability that it takes between 1 and 2 minutes to reboot.

Example 4. Lifetime in years, of a certain HD is a continuous random variable with density

$$f(t) = \begin{cases} K - \dfrac{x}{50}, & 0 < x < 10 \\ 0, & \text{otherwise} \end{cases}$$

Find K, the probability of a failure within first 5 years, and the expectation of the lifetime.

Remarkable Continuous Distributions

Uniform distribution. It is denoted by $U(a, b)$ and have the density function

$$f(t) = \begin{cases} 0, & x < a \\ \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & x > b \end{cases}$$

If $X : U(a, b)$, then $\mathbb{E}[X] = \dfrac{a+b}{2}$ and $Var[X] = \dfrac{(b-a)^2}{12}$.

$U(0, 1)$ is called *the standard uniform distribution.*

Remarkable Continuous Distributions

Exponential distribution. It is abbreviated by $Exp(\lambda)$ and have the density function ($\lambda > 0$ is the rate parameter)

$$f(t) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geqslant 0 \end{cases}$$

$X : Exp(\lambda)$, $\mathbb{E}[X] = \frac{1}{\lambda}$, $Var[X] = \frac{1}{\lambda^2}$.

Exponential distribution is used to model waiting time, interarrival time, hardware lifetime, failure time; in a sequence of rare events the time between events is exponentially distributed.

The Exponential distribution is memoryless (having waited for x_0 minutes get forgotten): regardless of the event $X > x$, when the total waiting time exceeds x, the remaining waiting time still has Exponential distribution: $P(X > x + \Delta x | X > x) = P(X > \Delta x)$ (why?).

Remarkable Continuous Distributions

Gaussian (normal) distribution. It is denoted by $N(\mu, \sigma^2)$ with the density function

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

If $X : N(\mu, \sigma^2)$, then $\mathbb{E}[X] = \mu$ and $Var[X] = \sigma^2$.
The distribution $N(0, 1)$ is called *the standard normal distribution*.
The values of a normal distributed variable have the following spreading (symmetrical around the mean): %68 belongs to a interval $[\mu - \sigma, \mu + \sigma]$, %95 belongs to $[\mu - 2\sigma, \mu + 2\sigma]$, and %99.7 belongs to $[\mu - 3\sigma, \mu + 3\sigma]$.

Remarkable Continuous Distributions

- Normal distribution has a proeminent role in Probability and Statistics for at least two reasons.
- As a consequence of the Central Limit Theorem (CLT - see below) sums and/or averages of identical distributed independent random variables have approximatively a Normal distribution.
- Normal distribution was found to be a good model for variables like temperature, weight, height or even student grades.
- The Normal distribution was tacitly used by de Moivre as an approximation to the binomial distribution and was later used by Laplace and Gauss.

Remarkable Continuous Distributions

Student (or t) distribution. It is denoted by $t(r)$ with the density function

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\dfrac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}}, & x > 0 \\ 0, & x \leqslant 0 \end{cases},$$

where $\Gamma(y) = \int\limits_0^{+\infty} x^{y-1} e^{-x}\, dx$. For a random variable $X : t(r)$, we have

$\mathbb{E}[X] = 0$ and $Var[X] = \dfrac{r}{r-2}$.

The larger the number of degrees of freedom, the more the distribution looks like the standard normal distribution.

Remarkable Continuous Distributions

Gamma distribution. It is denoted by $\Gamma(\alpha, \lambda)$ with the density function

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leqslant 0 \end{cases},$$

where $\Gamma(t) = \int\limits_0^{+\infty} x^{t-1} e^{-x}\ dx$. α is the shape parameter and λ is the rate (or the frequency) parameter. For a random variable $X : \Gamma(\alpha, \lambda)$, we have $\mathbb{E}[X] = \dfrac{\alpha}{\lambda}$ and $Var[X] = \dfrac{\alpha}{\lambda^2}$.

Suppose that we have a process that consists of α independent steps, and each step takes $Exp(\lambda)$ amount of time, then the total time follows a Gamma distribution.

That is, *Gamma* distribution is a sum of α independent Exponential variables.

Markov and Tchebychev inequalities

Proposition 1

Let $X \geqslant 0$ be a non-negative random variable. If $a > 0$, then

$$P(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}.$$

proof:

$$\mathbb{E}[X] = \int_0^{+\infty} tf(t)dt = \int_0^{a} tf(t)dt + \int_a^{+\infty} tf(t)dt \geqslant$$

$$\int_a^{+\infty} tf(t)dt \geqslant a\int_a^{+\infty} f(t)dt = aP(X \geqslant a).$$

■

Tchebychev's inequality

Proposition 2

Let X be a random variable having expectation μ and variance σ^2. Then

$$P(|X-\mu| \geqslant k) \leqslant \frac{\sigma^2}{k^2}.$$

proof: Consider the variable $Y = (X-\mu)^2$ and $a = k^2$ in Markov inequality

$$P(|X-\mu| \geqslant k) = P\left[(X-\mu)^2 \geqslant k^2\right] \leqslant \frac{\mathbb{E}\left[(X-\mu)^2\right]}{k^2} = \frac{\sigma^2}{k^2}.$$

■

Tchebychev's theorem

Theorem 1.1

Let $(X_n)_{n \geqslant 1}$ be a sequence of independent random variables having finite variances, uniformly bounded, that is $Var[X_n] \leqslant c$, for every $n \geqslant 1$. Then

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]\right| < \epsilon\right) = 1.$$

proof: We know that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] \text{ and}$$

$$Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] < \frac{c}{n}.$$

Tchebychev's theorem

Applying the Tchebychev's inequality for the variable $\frac{1}{n}\sum_{i=1}^{n} X_i$ we get

$$1 \geqslant P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]\right| < \epsilon\right) \geqslant 1 - \frac{Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]}{\epsilon^2} \geqslant 1 - \frac{c}{n\epsilon^2}.$$

Taking to the limit we obtain

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]\right| < \epsilon\right) = 1. \blacksquare$$

The weak law of large numbers

- The laws of large numbers say that as the number of identically distributed, random variables increases, their sample mean approaches their theoretical common mean (expectation).

Theorem 1.2

(The weak law of large numbers, Khintchine's law) Let $(X_n)_{n \geqslant 1}$ be a sequence of identically distributed independent random variables having mean μ and variance σ^2. Then

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu]\right| < \epsilon\right) = 1 \ \textit{or}$$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu]\right| \geqslant \epsilon\right) = 0.$$

The strong law of large numbers

proof: It is a consequnce of the previous theorem, since

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]=\mu.$$

∎

Theorem 1.3

(The strong law of large numbers) Let $(X_n)_{n\geqslant 1}$ be a sequence of identical distributed independent random variables having mean μ and variance σ^2. Then

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i=\mu\right)=1.$$

proof: It is more complex and will be omitted.

An example - frequencies

- Bernoulli which is credited with the first proof of the weak law of large numbers, formulated a version that addresses only the Bernoulli distribution.
- Suppose that we have a random experience and a related random event A with $P(A) = p$.
- We independently perform the experience, and consider the following sequence of random variables: $X_i = 1$ if A occurs at the ith performing, and 0 otherwise.
- The variables are independent and identically Bernoulli distributed with parameter p.

An example - frequencies

- The law of large numbers says that, with probability 1,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to p.$$

- $\sum_{i=1}^{n} X_i$ is the number of occurrences of A after n performings of the experience.
- In other words the law of large numbers says that the A occurs with frequency p.

Some History

- James Bernoulli proved the weak law of large numbers in 1700; Poisson generalized his result around 1800.
- Tchebychev discovered his inequality in 1866, and Markov extended Bernoulli's theorem to dependent random variables.
- In 1909 the Émile Borel proved what today is known as the strong law of large numbers that further generalizes Bernoulli's theorem.
- In 1926 Kolmogorov derived a more general condition that was sufficient for a set of mutually independent random variables to obey the law of large numbers. This condition is

$$\sum_{n \geqslant 1} \frac{Var[X_n]}{n^2} < +\infty.$$

Bernoulli's theorem

Proposition 3

Let α_n be a the number of occurrences of an event A in n independent performings of a random experience. If $f_n = \dfrac{\alpha_n}{n}$ is the relative frequency of occurrence of A, then the sequence $(f_n)_{n\geqslant 1}$ converges in probability to p, the probability of A.

proof: $\alpha_n = nf_n$ is a binomial distributed variable, hence $\mathbb{E}[\alpha_n] = np$ and $Var[\alpha_n] = np(1-p)$. Moreover

$$P(|f_n - p| < \epsilon) = P(|\alpha_n - np| < n\epsilon) = P(|\alpha_n - \mathbb{E}[\alpha_n]| < n\epsilon) \geqslant$$

$$\geqslant 1 - \frac{p(1-p)}{n\epsilon^2}.$$

Obviously, $\lim\limits_{n\to\infty} P(|f_n - p| < \epsilon) = 1$, for every $\epsilon > 0$. ■

The central limit theorem

Theorem 2.1

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

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

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

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 2.2

(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.
- What is $P(X = 20)$?

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

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

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 CDF, the probability density function - PDF, and (a value of a) a random variable having the specified distribution respectively.
- For generating discrete uniform random number 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 lenght 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

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

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

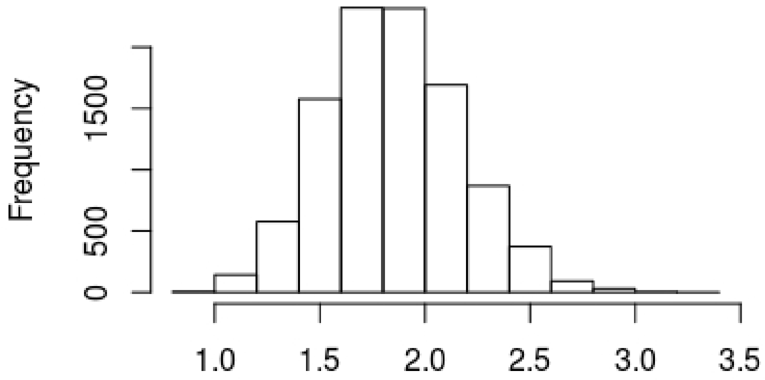
$$= pnorm(0.888) - pnorm(-0.888) = 2 \cdot pnorm(0.888) - 1 = 0.625$$

Illustrations of LLN and CLT

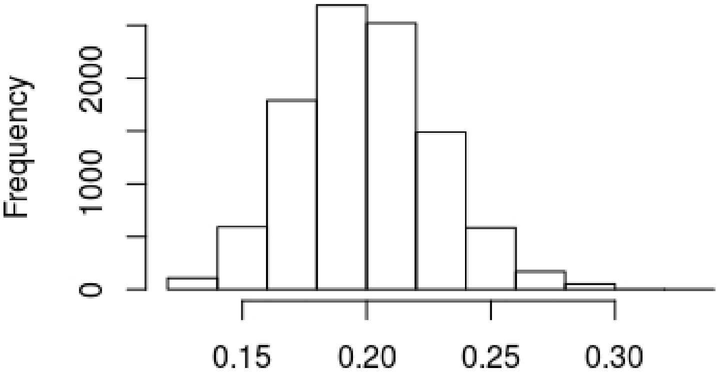
Example 5. (*Verifying CLT*) Consider a given probability distribution, X, with mean μ and variance σ^2, and a sequence of n independent identically distributed (with X) random variables X_i, $i = \overline{1,n}$. According to CLT, for large n, the sample mean, $\overline{x}_n$, has a normal distribution,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

We want to verify 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
$Abs.err$	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 absolute 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

$\mu \leftarrow 1/\lambda;$

$\sigma \leftarrow 1/\lambda^2;$ // 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.