

2 Weak law of large numbers and central limit theorem

Usually, the data processed by statisticians are taken from a *population*. The portion drawn is called a *sample*. The value of any population parameter is always constant, whereas sample parameters are random variables varying from sample to sample.

If the sample is large enough to embrace the entire population, there is no difference between the sample parameter and population parameter. In general, the larger the sample size, the more probable it is that the sample parameter comes arbitrarily close to the population parameter. This fact is often called *the law of large numbers*. It can be specified by Chebyshev theorem or Bernoulli's theorem.

The central limit theorem states, that the random variable X , designed as a sum of large number of mutually independent variables X_1, X_2, \dots, X_n , approaches (on general assumptions) the standard normal distribution. The Moivre-Laplace theorem is the simplest specification of CLT. Generalizing M-L theorem we obtain Lindberg-Levy theorem. The most generalized CLT was formulated by Ljapunov; this will not be shown.

Before we list the mentioned theorems, we have to define the term *convergence of the sequence of random variables*. In probability theory there exist several notions of convergence of random variables, three of them we provide.

Definition 2.1

The random sequence $(X_1, X_2, \dots, X_n, \dots)$ is said to *converges* toward random variable X

(i.) *surely* if:

$$\forall \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

[This is a "common" convergence of the sequence of numbers]

(ii.) *in probability*, if:

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(|X_n - X| \leq \varepsilon) = 1$$

[As n increases, the larger differences between X_n and X are extremely unlikely. It is also often denoted $\text{plim } X_n = X$]

(iii.) *in distribution*, if:

for the distribution functions $F_1(x_1) \sim X_1, \dots, F_n(x_n) \sim X_n, \dots$, eventually $F(x) \sim X$:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \text{ for any } x \text{ such that } F \text{ is continuous function}$$

[This is a weakest form of convergence, it is defined by means of distribution functions]

Remark 2.2

The random sequence may converge to the constant c , what is included in previous definition. Consider a random variable that attains the deterministic value c with probability one.

Theorem 2.3

1. Sure convergence of a random variable implies convergence in probability; convergence in probability implies convergence in distribution. The converse implications do not hold in general.]

2. The random sequence $(X_1, X_2, \dots, X_n, \dots)$ converges in probability toward constant μ if two following conditions hold:

$$\lim_{n \rightarrow \infty} E(X_n) = \mu \quad \wedge \quad \lim_{n \rightarrow \infty} D(X_n) = 0$$

Theorem 2.4 Weak law of large numbers (Chebyshev theorem)

Let us consider the random sequence $(X_1, X_2, \dots, X_n, \dots)$ of independent and identically-distributed random variables with constant expected value μ and constant variance σ^2 . Then the random sequence of sample means $(X_1, \frac{1}{2} \sum_{i=1}^2 X_i, \dots, \frac{1}{n} \sum_{i=1}^n X_i, \dots)$ converges in probability towards the expected value μ . Thus for any $\varepsilon > 0$ it holds:

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

or

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right) = 1$$

[Simply: $\text{plim} \frac{1}{n} \sum_{i=1}^n X_i = \mu$. Interpreting this result, the weak law essentially states that for sufficiently large sample there will be extremely high probability, that the sample mean will be arbitrarily close to the expected value (population mean). Thus for sufficiently large n the expected value μ may be estimated by sample mean.]

Theorem 2.5 Bernoulli (corollary of Chebyshev theorem)

Let the random variable Y_n gives the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability ϑ , $0 < \vartheta < 1$. Then the random sequence of relative frequencies $(Y_1, \frac{Y_2}{2}, \dots, \frac{Y_n}{n}, \dots)$ converges in probability towards the probability of success ϑ . Thus for any $\varepsilon > 0$ it holds:

$$P\left(\left|\frac{Y_n}{n} - \vartheta\right| < \varepsilon\right) \geq 1 - \frac{\vartheta(1 - \vartheta)}{n\varepsilon^2}$$

or

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{Y_n}{n} - \vartheta\right| < \varepsilon\right) = 1$$

The collocation between Chebyshev theorem and Bernoulli theorem will be clear from the following example.

Example 2.6

The probability, that the manufactured product is a waster, is $\frac{12}{3000}$. The production reviewing tested 3000 products. What is the probability that the difference between the relative frequency of wasters and the probability of wasters is 0.01 at most?

Solution

Let Y_{3000} stands for a random variable giving the number of wasters (successes) in 3000 independent experiments.

Then $Y_{3000} \sim Bi(3000, \frac{12}{3000})$

As n increases, the relative frequency of successes should be getting close to the probability of the success. We have to find out the probability, that for $n = 3000$ the relative frequency of successes

will deviate from the probability of the success not more than 0.01. Thus in Bernoulli theorem we will set up 0.01 as ε .

For any $\varepsilon > 0$ it holds: $P(|\frac{Y_n}{n} - \vartheta| < \varepsilon) \geq 1 - \frac{\vartheta(1-\vartheta)}{n\varepsilon^2}$.

Thus $P(|\frac{Y_{3000}}{3000} - \frac{12}{3000}| < 0,01) \geq 1 - \frac{\frac{12}{3000}(1-\frac{12}{3000})}{3000 \cdot 0,01^2} \doteq 0,872$

Using Chebyshev theorem X_i stands for random variable with Bernoulli distribution (zero-one or alternative distribution) which takes value 1 in case of waster (success) and value 0 in case of good-quality product (failure).

Thus $X_i \sim A(\frac{12}{3000})$, $i = 1, \dots, 3000$ $E(X_i) = \frac{12}{3000}$ $D(X_i) = \frac{12}{3000}(1 - \frac{12}{3000})$; X_1, \dots, X_{3000} are independent. Then set up 0.01 as ε and use Chebyshev theorem. (Consider that binomial random variable is equal to the sum of independent, identically distributed random variables, all Bernoulli distributed with success probability ϑ .)

Theorem2.7 central limit theorem (Lindberg-Lévy)

Let (X_1, \dots, X_n, \dots) be a random sequence of n independent and identically distributed random variables each having constant expected value μ and constant variance σ^2 . Let us consider $X = \sum_{i=1}^n X_i$ and let us derive the expected value and the variance of the random variable X .

$$E(X) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \mu = n\mu$$

$$D(X) = D(\sum_{i=1}^n X_i) = \sum_{i=1}^n D(X_i) = \sum_{i=1}^n \sigma^2 = n\sigma^2$$

Here let us consider standardized sum $U_n = \frac{X - E(X)}{\sqrt{D(X)}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$ [n can arbitrarily increase]

Then the random sequence of standardized sums $(U_1, U_2, \dots, U_n, \dots)$ converges in distribution toward random variable $U \sim N(0, 1)$. Thus

$$\forall u \in \mathbf{R} : \lim_{n \rightarrow \infty} P(U_n \leq u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(u),$$

where $\Phi(u)$ is the cumulative distribution function of $N(0, 1)$. We write for short $U_n \approx N(0, 1)$ and U_n is said to follow asymptotic standard normal distribution.

[Notice that $U_n = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\frac{\sigma}{\sqrt{n}}}$. So the central limit theorem states that as the sample size n increases, the distribution of the sample average of random variables X_1, \dots, X_n approaches the normal distribution with a mean μ and variance $\frac{\sigma^2}{n}$ irrespective of the shape of the original distribution.]

Theorem2.8 Moivre-Laplace (corollary of Lindberg-Lévy theorem)

Let $Y_n \sim Bi(n, \vartheta)$, $n = 1, 2, \dots$. Then $E(Y_n) = n\vartheta$ $D(Y_n) = n\vartheta(1 - \vartheta)$.

and $U_n = \frac{Y_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \approx N(0, 1)$

[Interpreting this result, Moivre-Laplace theorem essentially states that if n is large enough, then an excellent approximation to $Bi(n, \vartheta)$ is given by the normal distribution $N(n\vartheta, n\vartheta(1 - \vartheta))$. (Using convergence in distribution.)]

Remark 2.9

The Moivre-Laplace theorem allows us to apply the approximate formula, which substitutes onerous calculation of binomial distribution function by simple searching in tabs of standard normal distribution. Compare

Exact calculation:

$$P(Y_n \leq y) = \sum_{t=0}^y \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t} \dots \text{time-consuming onerous summation}$$

Approximation by normal distribution:

$$P(Y_n \leq y) = P\left(\frac{Y_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \leq \frac{y - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}\right) \approx \Phi\left(\frac{y - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}\right) \sim N(0, 1),$$

where $\Phi(u)$ is tabulated standard normal distribution function.

The approximation may be used if following assumptions hold:

$$n\vartheta(1-\vartheta) > 9 \quad \wedge \quad \frac{1}{n+1} < \vartheta < \frac{n}{n+1}.$$

Example 2.10

Consider 100 independent tosses of a die. Find the probability that the number "six" does appear at least 20 times.

Solution

Let Y_{100} stands for a random variable giving a number of obtained "sixes" in 100 tosses, $Y_{100} \sim Bi(100, \frac{1}{6})$.

First we have to validate the assumptions of using normal approximation.:

$$n\vartheta(1-\vartheta) = 100 \cdot \frac{1}{6} \left(1 - \frac{1}{6}\right) = \frac{500}{36} > 9 \quad \wedge \quad \frac{1}{101} < \frac{1}{6} < \frac{100}{101}, \text{ hence both assumptions hold.}$$

So that to estimate the wanted probability we can use the Moivre-Laplace theorem.

$$P(Y_{100} \geq 20) = 1 - P(Y_{100} < 20) = P(Y_{100} \leq 19) = 1 - P\left(\frac{Y_{100} - 100/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}} \leq \frac{19 - 100/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}}\right) = 1 - P(U_n \leq 0,626) \approx 1 - \Phi(0,626) = 1 - 0,73565 = 0,2635.$$

(The exact software calculation is 0,2198.)

There are some cases, in which the approximation to binomial distribution by normal distribution is not convenient. Especially in case of extremely small probability of success ϑ the approximation by Poisson distribution fits better.

Theorem 2.11 Poisson

Let Y_1, Y_2, \dots be a random sequence of independent random variables, $Y_n \sim Bi(n, \vartheta_n)$, $n = 1, 2, \dots$ and $\lim_{n \rightarrow \infty} n\vartheta_n = \lambda$. Then the sequence $Y_1, Y_2, \dots, Y_n, \dots$ converges in distribution towards random variable $Y \sim Po(\lambda)$, thus $Y_n \approx Po(\lambda)$.

[The random variable Y follows the Poisson distribution with parameter $\lambda = n\vartheta$, which can be used as an approximation to the binomial distribution.]

Remark 2.12

The Poisson theorem allows us to apply the approximate formula, which substitutes onerous calculation of binomial distribution (or probability) function by simple searching in tabs of Poisson distribution (or probability) function.

• $P(Y_n \leq y) = \sum_{t=0}^y \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t} \approx F_{n\vartheta}(y) \sim Po(n\vartheta)$, where $F_{n\vartheta}(y)$ is a Poisson distribution function with parameter $\lambda = n\vartheta$

• $P(Y_n = y) = \binom{n}{y} \vartheta^y (1-\vartheta)^{n-y} \approx \frac{(n\vartheta)^y}{y!} e^{-n\vartheta}$ (Compare with Poisson probability function.)

The approximation by Poisson distribution may be used if following assumptions hold:

$$n \geq 30 \wedge \vartheta \leq 0,1.$$

Example 2.13

Testing a reliability of a particular device, this gets out of order with the probability 0.05. Find the probability that testing 100 identical devices, there will be just 5, which get out of order.

Solution

Let Y_{100} stands for a random variable giving a number of defected devices in a sample of 100 inde-

pendent tests, $Y_{100} \sim Bi(100; 0, 05)$.

First we have to validate the assumptions of using Poisson approximation.:

$$100 \geq 30 \quad \wedge \quad 0,05 \leq 0,1.$$

Poisson approximation of wanted probability:

$P(Y_{100} = 5) \approx \frac{(100 \cdot 0,05)^5}{5!} e^{-100 \cdot 0,05}$, which is a Poisson probability function with parameter $\lambda = 100 \cdot 0,05$ in a point 5 and it is tabulated. Thus $p_5(5) = 0,17547$.

Exact calculation of wanted probability:

$$P(Y_{100} = 5) = \binom{100}{5} 0,05^5 (1 - 0,05)^{95} = \dots = 0,18$$