Imperial College London

Mathematics: Year 2

Joint Distributed Random Variables

Xin Wang

October 1, 2020

Abstract

In Statistics, the two most important but difficult to understand concepts are Law of Large Numbers (LLN) and Central Limit Theorem (CLT). These form the basis of the popular hypothesis testing framework. In the practical world, its impossible to analyse an entire population. Hence, mathematicians resort to sampling from the population, perform analysis and draw conclusions about the population based on the sample.

Contents

1 Introduction

It is understood intuitively that the average of many measurements of the same unknown quantity tends to give a better estimate than a single measurement. This is because the random error of each measurement cancels out in the average. In thi section, this intuition is precisely defined in two ways: the law of large numbers (LoLN) and the central limit theorem (CLT).

The random variables X_i evaluated are **independent and identically distributed**. This means that random variables $X_1, X_2, \ldots, X_n$ are independent and have the **same underlying distribution**.

The random variables that are independent and identically distributed satisfies the following properties:

1. Random variables $X_1, \ldots, X_n$ are all independent.

2. Random variables have the same expectation

$$E(X_i) = \mu$$

   $\forall i$ where μ is the mean.

3. Random variables have the same variance

$$\mathrm{Var}(X_i) = \sigma^2$$

   $\forall i$ where σ^2 is the variance and σ is the standard deviation.

As a breif the law of large numbers and central limit theorem are based on **the multiple independent samples from the same distribution**.

The LoLN shows two conclusions:

- The average of many independent samples is (with high probability) close to the mean of the underlying distribution.

- The density histogram of many independent samples is (with high probability) close to the graph of the density of the underlying distribution.

The central limit theorem says that the average of many independent copies of a random variable is **approximately a normal random variable**. The CLT goes on to give precise values for the mean and standard deviation of the normal variable.

It should be noted that mathematics cannot tell us if the experiment is producing data worth averaging. For example, if the measuring device is defective or poorly calibrated then the average of many measurements will be wrong even if the experiment is highly accurate. This is an example of **systematic error or sampling bias**, as opposed to the random error which can be controlled by the law of large numbers.

2 Law of large numbers

As mentioned previously,X_i are independent and identically-distributed thus X_i all have the same mean μ and standard deviation σ.

Let $\overline{X}_n$ be a random variable and the average of $X_1, \ldots, X_n$:

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

The law of large numbers and central limit theorem takes $\overline{X}_n$ and shows the value and distribution of $\overline{X}_n$ respectively.

- LoLN: As n grows, the probability that X_n is close to μ goes to 1.
- CLT: As n grows, the distribution of X_n converges to the normal distribution $N(\mu, \frac{\sigma^2}{n})$

Example 1: Averages of Bernoulli random variables

Bernoulli variable: A random variable that takes value 1 in case of success and 0 in case of failure.

Suppose each X_i is an independent flip of a fair coin

$$X_i \sim \text{Bernoulli}(0.5)$$
$$\mu = 0.5$$

$\overline{X}_n$ is the proportion of H in n flips, and it is expected that this proportion is close to 0.5 for large n since $\mu = 0.5$. However, note that due to the nature of randomness, this is not guaranteed e.g. it is possible to get 1000 H in 1000 flips, though the probability is very small.

A more accurate definition of this intuitive idea is: **There is with high probability** the sample average $\overline{X}_n$ is close to the mean 0.5 for large n.

Proof: Investigate the probability of th sample average $\overline{X}_n$ being within 0.1 of the mean μ 0.5.

$$P(|\overline{X}_n - 0.5| < 0.1) \text{ or equivalently } P(0.4 \leq \overline{X}_n \leq 0.6)$$

The law of large numbers says that this probability goes to 1 as the number of flips n gets large. This is seen through any program used to calculate the binomial probability distribution $P(0.4 \leq \overline{X}_n \leq 0.6)$:

$n = 10$:	`pbinom(6, 10, 0.5) - pbinom(3, 10, 0.5)`	$= 0.65625$
$n = 50$:	`pbinom(30, 50, 0.5) - pbinom(19, 50, 0.5)`	$= 0.8810795$
$n = 100$:	`pbinom(60, 100, 0.5) - pbinom(39, 100, 0.5)`	$= 0.9647998$
$n = 500$:	`pbinom(300, 500, 0.5) - pbinom(199, 500, 0.5)`	$= 0.9999941$
$n = 1000$:	`pbinom(600, 1000, 0.5) - pbinom(399, 1000, 0.5)`	$= 1$

Figure 1: R code producing the following values for $P(0.4 \leq \overline{X}_n \leq 0.6)$

The law of large numbers still apply with the probability of being within 0.01 of the mean. It will take larger values of n to raise the probability to near 1.

$n = 10$:	`pbinom(5, 10, 0.5) - pbinom(4, 10, 0.5)`	$= 0.2460937$
$n = 100$:	`pbinom(51, 100, 0.5) - pbinom(48, 100, 0.5)`	$= 0.2356466$
$n = 1000$:	`pbinom(510, 1000, 0.5) - pbinom(489, 1000, 0.5)`	$= 0.49334$
$n = 10000$:	`pbinom(5100, 10000, 0.5) - pbinom(4899, 10000, 0.5)`	$= 0.9555742$

Figure 2: R code producing the following values for $P(0.49 \leq \overline{X}_n \leq 0.51)$

This convergence of the probability to 1 is proof of the LoLN.

2.1 Formal statement of the law of large numbers

Given that random variables $X_1, X_2, \ldots, X_n$ are independent and share identical distribution with mean μ and variance σ^2, for each n, let X_n be the average of the first n variables.

For any $a > 0$

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < a) = 1$$

The equation precisely defines that as n increases the probability of being within a of the mean goes to 1 where a is a small tolerance of error from the true mean μ. If the tolerance a is decreased or the probability p is increased, then N will need to be larger. For example, in the previous example, if the probability that the proportion of heads X_n is within $a = 0.1$ of $\mu = 0.5$ is to be at least $p = 0.99999$, then $n > N = 500$ is large enough.

2.2 Binomial distribution and normal distribution

Suppose we have n random variables, X_i for $i = 1, 2, \ldots, n$, mutually independent and identical i.e. the same distribution, each having mean μ and variance σ^2:

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\mu$$
$$\operatorname{Var}\left(\sum_{i=1}^{n} E(X_i)\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) = n\sigma^2$$

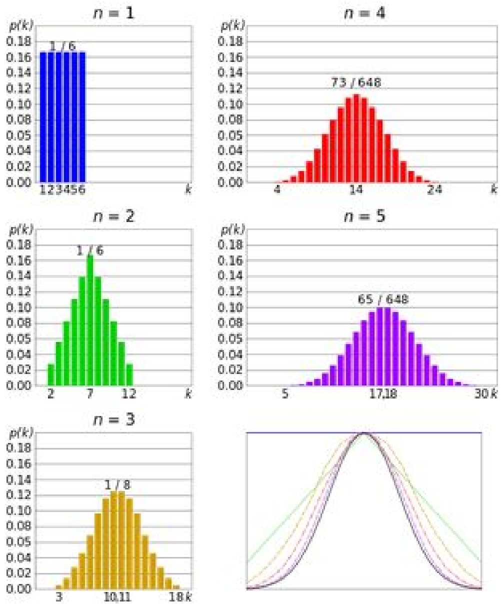
It can be shown that, as $n \to \infty$ the distribution of this sum converges to the normal distribution:

$$X_1 + X_2 + \cdots + X_n \to N(n\mu, n\sigma^2)$$

3 Central Limit Theorem

Central Limit Theorem: No matter what the shape of the population distribution is, the sampling distribution of the sample means approaches a normal distribution as the sample size n gets larger.

Graphically speaking, the CLT can be illustrated with rolling a fair die. The more times the die is rolled, the more likely the shape of the distribution of the means tends to look like a normal distribution graph.



An essential component of the Central Limit Theorem is that the average μ of the sample will be the population mean μ. Similarly, if by finding the average of all of the standard deviations in the sample, the actual standard deviation σ for your population is found.

3.1 Standardisation

Given a random variable X with mean μ and standard deviation σ, standardisation of X is defined as a new random variable:

$$Z = \frac{X - \mu}{\sigma}$$

Note:

- Z has mean of 0 and standard deviation of 1.
- if X has a normal distribution, then the standardization of X results in the standard normal distribution Z

3.2 Statement of the central limit theorem

Suppose $X_1, X_2, \ldots, X_n$, are independent and share the identical distribution, the random variables each having mean μ and standard deviation σ.

For each n let S_n denote the sum of $X_1, X_2, \ldots, X_n$:

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

and let X_n denote the average of $X_1, \ldots, X_n$:

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{S_n}{n}$$

The properties of mean and variance for both S_n and $\overline{X}_n$ are defined as:

$$\begin{aligned} E(S_n) &= n\mu, & \operatorname{Var}(S_n) &= n\sigma^2, & \sigma_{S_n} &= \sqrt{n}\,\sigma \\ E(\overline{X}_n) &= \mu, & \operatorname{Var}(\overline{X}_n) &= \frac{\sigma^2}{n}, & \sigma_{\overline{X}_n} &= \frac{\sigma}{\sqrt{n}}. \end{aligned}$$

Notice that the mean and variance for both S_n and $\overline{X}_n$ are multiples of each other thus both would have the same standardisation:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

This leads to the Central Limit Theorem:

$$\overline{X}_n \approx N(\mu, \frac{\sigma^2}{n}) \qquad S_n \approx N(n\mu, n\sigma^2) \qquad Z_n \approx N(0, 1)$$

These three equations show:

- $\overline{X}_n$ is approximately a normal distribution with the same mean as X but a smaller variance.
- S_n is approximately normal.
- Standardized $\overline{X}_n$ and S_n are approximately standard normal.

The central limit theorem is used to approximate a sum or average of i.i.d random variables by a normal random variable. This is extremely useful because it is usually easy to do computations with the normal distribution.

3.3 Moivre-Laplace Theorem

In probability theory, the de Moivre–Laplace theorem is a special case of the central limit theorem.

de Moivre–Laplace theorem: The normal distribution may be used as an approximation to the binomial distribution under certain conditions.

If $X \sim \text{Binomial}(n, p)$, then $X = Y_1 + \ldots, Y_n$ is the sum of independent Bernoulli with parameter p and $q = 1 - p$ if $n \to +\infty$:

$$P(X \leq a) = P\left(\frac{X - np}{\sqrt{npq} \leq \frac{a-np}{\sqrt{npq}}}\right) \to P\left(Z \leq \frac{a - np}{\sqrt{npq}}\right)$$