

THE CENTRAL LIMIT THEOREM

Throughout the discussion below, let X_1, X_2, \dots be i.i.d. rv's, each with finite expected value μ and finite nonzero standard deviation σ . Given n , define \bar{X}_n to be the average $(X_1 + \dots + X_n)/n$, and define S_n to be the sum $X_1 + \dots + X_n$. Then

$$\begin{aligned} E(S_n) &= n\mu \\ V(S_n) &= n\sigma^2 \\ \text{st.dev.}(S_n) &= \sigma\sqrt{n}. \end{aligned}$$

(The equation for the variance of S_n holds because the X_i are independent, so the variance of the sum of the X_i is the sum of the variances.) Now, $\bar{X}_n = (1/n)S_n$, so

$$\begin{aligned} E(\bar{X}_n) &= \mu &= (1/n) E(S_n) \\ V(\bar{X}_n) &= \frac{\sigma^2}{n} &= (1/n)^2 V(S_n) \\ \text{st.dev.}(\bar{X}_n) &= \frac{\sigma}{\sqrt{n}} &= (1/n) \text{st.dev.}(S_n) \end{aligned}$$

REMARKS

1. We can think of the i.i.d. condition as meaning that the X_i are repeated experiments, or alternately random samples, from some given probability distribution.
2. The expected value of the sample average is the same as the expected value of each X_i . This is common sense. We can think of \bar{X}_n as an estimate of the true population average μ .
3. As n gets bigger, the spread (standard deviation) of \bar{X}_n gets smaller. This is common sense: a bigger sample should give a more reliable estimate of the true population average.
4. $\mathcal{N}(\mu, \sigma)$ denotes the normal distribution with mean μ and standard deviation σ .

THEOREM (Central Limit Theorem) Suppose that X_1, X_2, \dots is a sequence of i.i.d. rv's, each with finite expected value μ and finite nonzero standard deviation σ . Let Z_n be the standardized version of \bar{X}_n , i.e.

$$Z_n = \frac{\bar{X}_n - \mu}{(\sigma/\sqrt{n})}.$$

Then as $n \rightarrow \infty$, $Z_n \rightarrow \mathcal{N}(0, 1)$.

REMARKS

1. Note the CLT has an extra assumption (finite variance) which the LLN does not have. The CLT gives more information when it is applicable.
2. The CLT is an incredible law of nature. Under modest assumptions, the process of independent repetition has a universal effect on the averaging process, depending only on the mean and standard deviation of the underlying population.
3. The expression “ $Z_n \longrightarrow \mathcal{N}(0, 1)$ ” abbreviates “ Z_n converges in distribution to $\mathcal{N}(0, 1)$ as $n \rightarrow \infty$ ”. Informally, this means that if n is large enough, then we have (for all numbers $a < b$)

$$\text{Prob}\left(a < \frac{(\bar{X}_n - \mu)}{(\sigma/\sqrt{n})} < b\right) \approx \text{Prob}(a < Z < b)$$

where \approx means “approximately equals”, and as n goes to infinity, the approximation gets as good as we want. If we want to be completely precise, we express this by saying that for every $\epsilon > 0$, there exists N such that whenever $n \geq N$ and $a < b$, we have

$$\left| \text{Prob}(a < Z_n < b) - |\Phi(b) - \Phi(a)| \right| < \epsilon$$

where Φ is the cumulative distribution function for $\mathcal{N}(0, 1)$ (which is given approximately by the tables in the back of our textbook).

4. There are other ways to express the approximation above:

$$\begin{aligned} \text{Prob}\left(a(\sigma/\sqrt{n}) < (\bar{X}_n - \mu) < b(\sigma/\sqrt{n}) \right) &\approx \text{Prob}(a < Z < b), \quad \text{or} \\ \text{Prob}\left(\mu + a(\sigma/\sqrt{n}) < \bar{X}_n < \mu + b(\sigma/\sqrt{n}) \right) &\approx \text{Prob}(a < Z < b) \end{aligned}$$

where Z is any r.v. which has the standard normal distribution $\mathcal{N}(0, 1)$. We could also use the notation $\bar{X}_n \longrightarrow \mathcal{N}(\mu, \sigma/\sqrt{n})$ to describe this situation.

RULE OF THUMB

How large should n be for the CLT approximation to be good enough? Really, that depends on the particular r.v. X and on how good “good enough” has to be. One rule of thumb (the rule used, for example, in the Devore text) is that, unless we have explicit information to the contrary, $n > 30$ is large enough for “good enough”.

EXAMPLE

Let us go through the approximations above in an example, with $a = -2$ and $b = 2$. We will take the r.v.'s X_i corresponding to flipping a fair coin.

So, each X_i equals 0 with probability $1/2$, and equals 1 with probability $1/2$.

For each X_i , $\mu = .5$ and $\sigma = \sqrt{(.5)(1 - .5)} = .5$. Let $n = 10,000$.

Then the standard deviation for \bar{X}_n is $(\sigma/\sqrt{n}) = (.5)/\sqrt{10,000} = .005$.

Here are the approximations above with these numbers put in.

$$\begin{aligned}\text{Prob}(-2 < \frac{(\bar{X}_n - .5)}{.005} < 2) &\approx \text{Prob}(-2 < Z < 2) \\ \text{Prob}(-2(.005) < (\bar{X}_n - .5) < 2(.005)) &\approx \text{Prob}(-2 < Z < 2) \\ \text{Prob}(.5 - 2(.005) < \bar{X}_n < .5 + 2(.005)) &\approx \text{Prob}(-2 < Z < 2) .\end{aligned}$$

If we compute out the last line, we get

$$\begin{aligned}\text{Prob}(.49 < \bar{X}_n < .51) &\approx \text{Prob}(-2 < Z < 2) = \text{Prob}(Z \leq 2) - \text{Prob}(Z \leq -2) \\ &= .9772 - .0228 = .9544.\end{aligned}$$

This means: if the experiment is to flip a fair coin 10,000 times:

then in about 95% of those experiments,

the percentage of the flips which equal heads will be between 49% and 51%.

SUMS AND THE CLT

Let us look again at one of the ways to express the CLT:

$$\text{Prob}\left(a \frac{\sigma}{\sqrt{n}} < (\bar{X}_n - \mu) < b \frac{\sigma}{\sqrt{n}}\right) \approx \text{Prob}(a < Z < b) .$$

Remember, $\bar{X}_n = (X_1 + \dots + X_n)/n$. If we multiply each element of the inequality on the left by n , we don't change the truth of the inequality, so we don't change its probability. So we get

$$\text{Prob}(a\sigma\sqrt{n} < (X_1 + \dots + X_n) - n\mu < b\sigma\sqrt{n}) \approx \text{Prob}(a < Z < b) .$$

So, we can also use a normal approximation for the probability that sums lie in some range. (For the special case of coin flipping, we already did this with the normal approximation to the binomial distribution. In fact, the description above explains how the normal approximation to the binomial distribution can be deduced as a consequence of the Central Limit Theorem.)

LAW OF LARGE NUMBERS

Let us see that the LLN is a consequence of the CLT, in the case that the CLT applies.

Suppose $\epsilon > 0$, and we have i.i.d. rv's as in the Central Limit Theorem. Then

$$\begin{aligned} & \text{Prob}\left(-\epsilon < \bar{X}_n - \mu < \epsilon\right) \\ &= \text{Prob}\left(\left(\frac{-\epsilon\sqrt{n}}{\sigma}\right) < \frac{\bar{X}_n - \mu}{(\sigma/\sqrt{n})} < \left(\frac{\epsilon\sqrt{n}}{\sigma}\right)\right) \\ &\approx \text{Prob}\left(\frac{-\epsilon\sqrt{n}}{\sigma} < Z < \frac{\epsilon\sqrt{n}}{\sigma}\right) \end{aligned}$$

where Z is any random variable with the standard normal distribution. Therefore for any given $\epsilon > 0$, no matter how small,

$$\lim_{n \rightarrow \infty} \text{Prob}\left(-\epsilon < \bar{X}_n - \mu < \epsilon\right) = 1 .$$

This last statement is one way to state the Law of Large Numbers.