5.10 Central Limit Theorem

Unit 5: Averages

Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.

Let $X_1, X_2, X_3, \ldots$ be i.i.d. with mean μ and variance σ^2. The law of large numbers says that as $n \to \infty$, $\bar{X}_n$ converges to the constant μ (with probability 1). But what is its distribution along the way to becoming a constant? This is addressed by the central limit theorem (CLT), which, as its name suggests, is a limit theorem of central importance in statistics.

The CLT states that for large n, the distribution of $\bar{X}_n$ after standardization approaches a standard Normal distribution. By standardization, we mean that we subtract μ, the expected value of $\bar{X}_n$, and divide by $\sigma/\sqrt{n}$, the standard deviation of $\bar{X}_n$.

> THEOREM 5.10.1 (CENTRAL LIMIT THEOREM).
>
> As $n \to \infty$,
>
> $$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to \mathcal{N}(0,1) \text{ in distribution.}$$
>
> In words, the CDF of the left-hand side approaches Φ, the CDF of the standard Normal distribution.

The CLT is an *asymptotic* result, telling us about the limiting distribution of $\bar{X}_n$ as $n \to \infty$, but it also suggests an *approximation* for the distribution of $\bar{X}_n$ when n is a finite large number.

Central limit theorem, approximation form.

For large n, the distribution of $\bar{X}_n$ is approximately $\mathcal{N}(\mu, \sigma^2/n)$. Of course, we already knew from properties of expectation and variance that $\bar{X}_n$ has mean μ and variance σ^2/n; the central limit theorem gives us the additional information that $\bar{X}_n$ is approximately *Normal* with said mean and variance.

Let's take a moment to admire the generality of this result. The distribution of the individual X_j can be *anything in the world*, as long as the mean and variance are finite. We could have a discrete distribution like the Binomial, a bounded continuous distribution, or a distribution with multiple peaks and valleys. No matter what, the act of averaging will cause Normality to emerge. In the figure below we show histograms of the distribution of $\bar{X}_n$ for four different starting distributions and for $n = 1, 5, 30, 100$. We can see that as n increases, the distribution of $\bar{X}_n$ starts to look Normal, regardless of the distribution of the X_j.

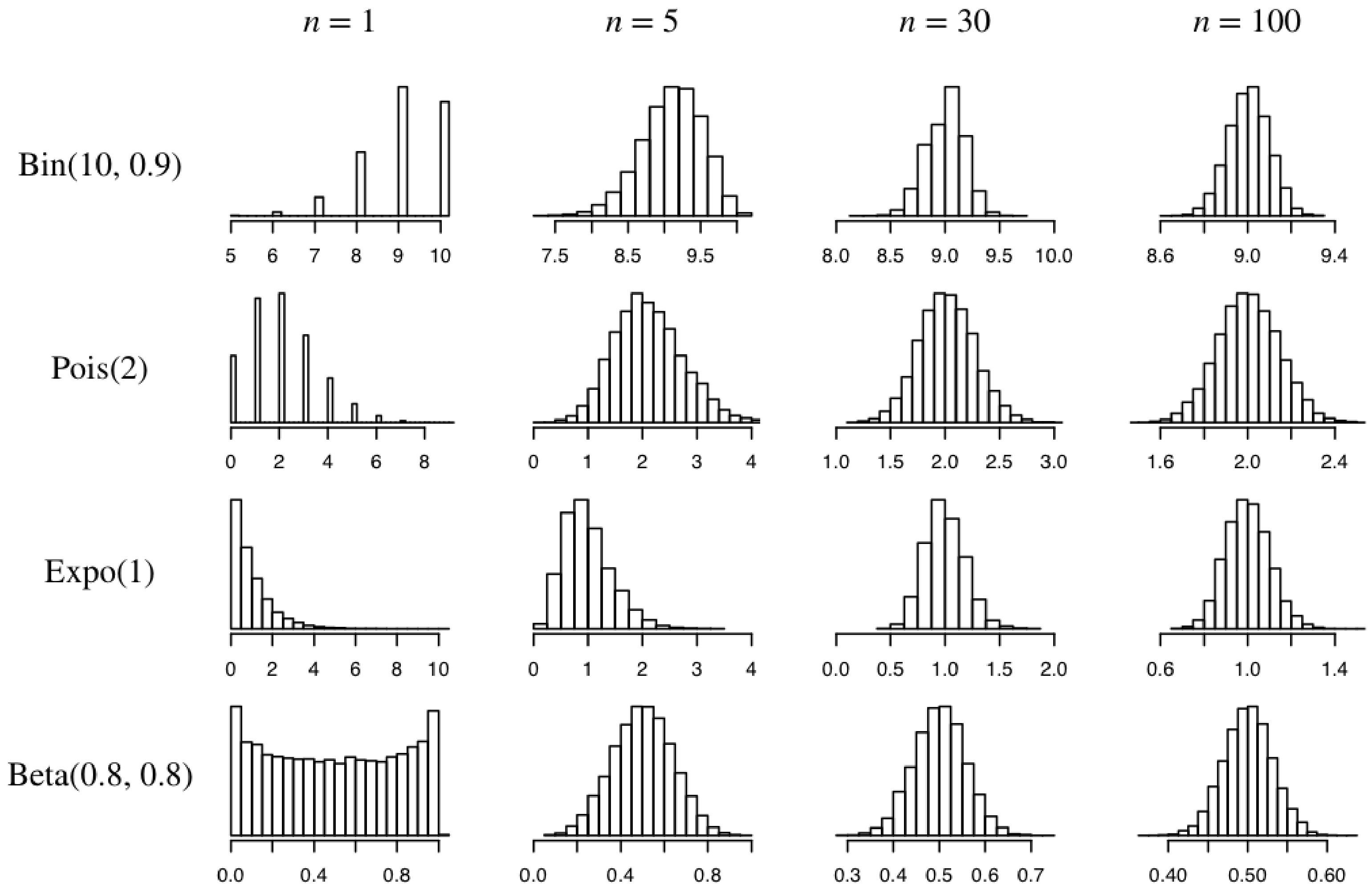

Figure 5.10.2: Central limit theorem. Histograms of the distribution of $\bar{X}_n$ for different starting distributions of the X_j (indicated by the rows) and increasing values of n (indicated by the columns). Each histogram is based on 10,000 simulated values of $\bar{X}_n$. Regardless of the starting distribution of the X_j, the distribution of $\bar{X}_n$ approaches a Normal distribution as n grows.

View Larger Image
Image Description

This does not mean that the distribution of the X_j is irrelevant, however. If the X_j have a highly skewed or multimodal distribution, we may need n to be very large before the Normal approximation becomes accurate; at the other extreme, if the X_j are already i.i.d. Normals, the distribution of $\bar{X}_n$ is exactly $\mathcal{N}(\mu, \sigma^2/n)$ for all n. Since there are no infinite datasets in the real world, the quality of the Normal approximation for finite n is an important consideration.

The CLT says that the sample mean $\bar{X}_n$ is approximately Normal, but since the sum $W_n = X_1 + \cdots + X_n = n\bar{X}_n$ is just a scaled version of $\bar{X}_n$, the CLT also implies W_n is approximately Normal. If the X_j have mean μ and variance σ^2, W_n has mean $n\mu$ and variance $n\sigma^2$. The CLT then states that for large n,

$$W_n \dot\sim \mathcal{N}(n\mu, n\sigma^2).$$

This is completely equivalent to the approximation for $\bar{X}_n$, but it can be useful to state it in this form because many of the named distributions we have studied can be considered as a sum of i.i.d. r.v.s.

Example 5.10.3 (Poisson convergence to Normal).

Let $Y \sim \mathrm{Pois}(n)$. By theorem 5.7.6, we can consider Y to be a sum of n i.i.d. $\mathrm{Pois}(1)$ r.v.s. Therefore, for large n,

$$Y \dot\sim \mathcal{N}(n, n).$$

Example 5.10.4 (Binomial convergence to Normal).

Let $Y \sim \mathrm{Bin}(n, p)$, we consider Y to be a sum of n i.i.d. $\mathrm{Bern}(p)$ r.v.s. Therefore, for large n,

$$Y \dot\sim \mathcal{N}\left(np, np(1-p)\right).$$

The Normal approximation to the Binomial distribution is complementary to the Poisson approximation discussed earlier in this unit. The Poisson approximation works best when n is large and p is small, while the Normal approximation works best when n is large and p is around $1/2$, so that the distribution of Y is symmetric or nearly so.

Example 5.10.5 (Volatile stock).

Each day, a very volatile stock rises 70% or drops 50% in price, with equal probabilities and with different days independent. Let Y_n be the stock price after n days, starting from an initial value of $Y_0 = 100$.

(a) Explain why $\log Y_n$ is approximately Normal for n large, and state its parameters.

(b) What happens to $E(Y_n)$ as $n \to \infty$?

(c) Use the law of large numbers to find out what happens to Y_n as $n \to \infty$.

Solution

(a) We can write $Y_n = Y_0(0.5)^{n-U_n}(1.7)^{U_n}$ where $U_n \sim \mathrm{Bin}(n, \frac{1}{2})$ is the number of times the stock rises in the first n days. This gives

$$\log Y_n = \log Y_0 - n\log 2 + U_n \log 3.4,$$

which is a location-scale transformation of U_n. By the CLT, U_n is approximately $\mathcal{N}(\frac{n}{2}, \frac{n}{4})$ for large n, so $\log Y_n$ is approximately Normal with mean

$$E(\log Y_n) = \log 100 - n\log 2 + (\log 3.4)E(U_n) \approx \log 100 - 0.081n$$

and variance

$$\mathrm{Var}(\log Y_n) = (\log 3.4)^2 \cdot \mathrm{Var}(U_n) \approx 0.374n.$$

(b) With notation as in (a), we can write

$$Y_n = Y_0(0.5)^n(3.4)^{U_n}.$$

Representing U_n as a sum of n i.i.d. $\mathrm{Bern}(1/2)$ r.v.s and using the fact (which we will show in Unit 6) that if X and Y are *independent* then $E(XY) = E(X)E(Y)$, we have

$$E(Y_n) = Y_0(0.5)^n E\left((3.4)^{U_n}\right) = Y_0(0.5)^n \left(E(3.4)^B\right)^n,$$

where $B \sim \mathrm{Bern}(1/2)$. This simplifies to

$$E(Y_n) = Y_0(0.5)^n \left(3.4 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}\right)^n = 1.1^n Y_0.$$

So $E(Y_n) \to \infty$ as $n \to \infty$.

Alternatively, we can use the notion of *conditional expectation given an r.v.*, which is covered in Unit 6 (so you may want to revisit this part of the solution after studying Unit 6). Conditioning on Y_n, we have

$$E(Y_{n+1}|Y_n) = \frac{1}{2}(1.7Y_n) + \frac{1}{2}(0.5Y_n) = 1.1Y_n,$$

so by Adam's law,

$$E(Y_{n+1}) = E(E(Y_{n+1}|Y_n)) = 1.1E(Y_n).$$

Thus, $E(Y_n) = 1.1^n Y_0$, which agrees with the result of the earlier approach. Again we have that $E(Y_n)$ goes to ∞ as $n \to \infty$.

(c) As in (a), let $U_n \sim \mathrm{Bin}(n, \frac{1}{2})$ be the number of times the stock rises in the first n days. Note that even though $E(Y_n) \to \infty$, if the stock goes up 70% one day and then drops 50% the next day, then overall it has dropped 15% since $1.7 \cdot 0.5 = 0.85$. So after many days, Y_n will be very small if about half the time the stock rose 70% and about half the time the stock dropped 50%---and the law of large numbers ensures that this *will* be the case! Writing Y_n in terms of U_n/n in order to apply LLN, we have

$$Y_n = Y_0(0.5)^{n-U_n}(1.7)^{U_n} = Y_0\left(\frac{(3.4)^{U_n/n}}{2}\right)^n.$$

Since $U_n/n \to 0.5$ with probability 1, $(3.4)^{U_n/n} \to \sqrt{3.4} < 2$ with probability 1, so $Y_n \to 0$ with probability 1.

Paradoxically, $E(Y_n) \to \infty$ but $Y_n \to 0$ with probability 1. To gain some intuition on this result, consider the extreme example where a gambler starts with \$100 and each day either quadruples his or her money or loses the entire fortune, with equal probabilities. Then on average the gambler's wealth doubles each day, which sounds good until one notices that eventually there will be a day when the gambler goes broke. The gambler's actual fortune goes to 0 with probability 1, whereas the expected value goes to infinity due to tiny probabilities of getting extremely large amounts of money.

WARNING 5.10.6 (THE EVIL CAUCHY).

The *Cauchy* distribution is the distribution of X/Y, where X and Y are i.i.d. $\mathcal{N}(0,1)$. The Cauchy turns out to have PDF

$$f(x) = \frac{1}{\pi(1+x^2)},$$

for all real x. The Cauchy PDF has much heavier tails than the standard Normal PDF, as shown in the figure below, which plots the Cauchy PDF (in black) compared with the $\mathcal{N}(0,1)$ PDF (in gray).

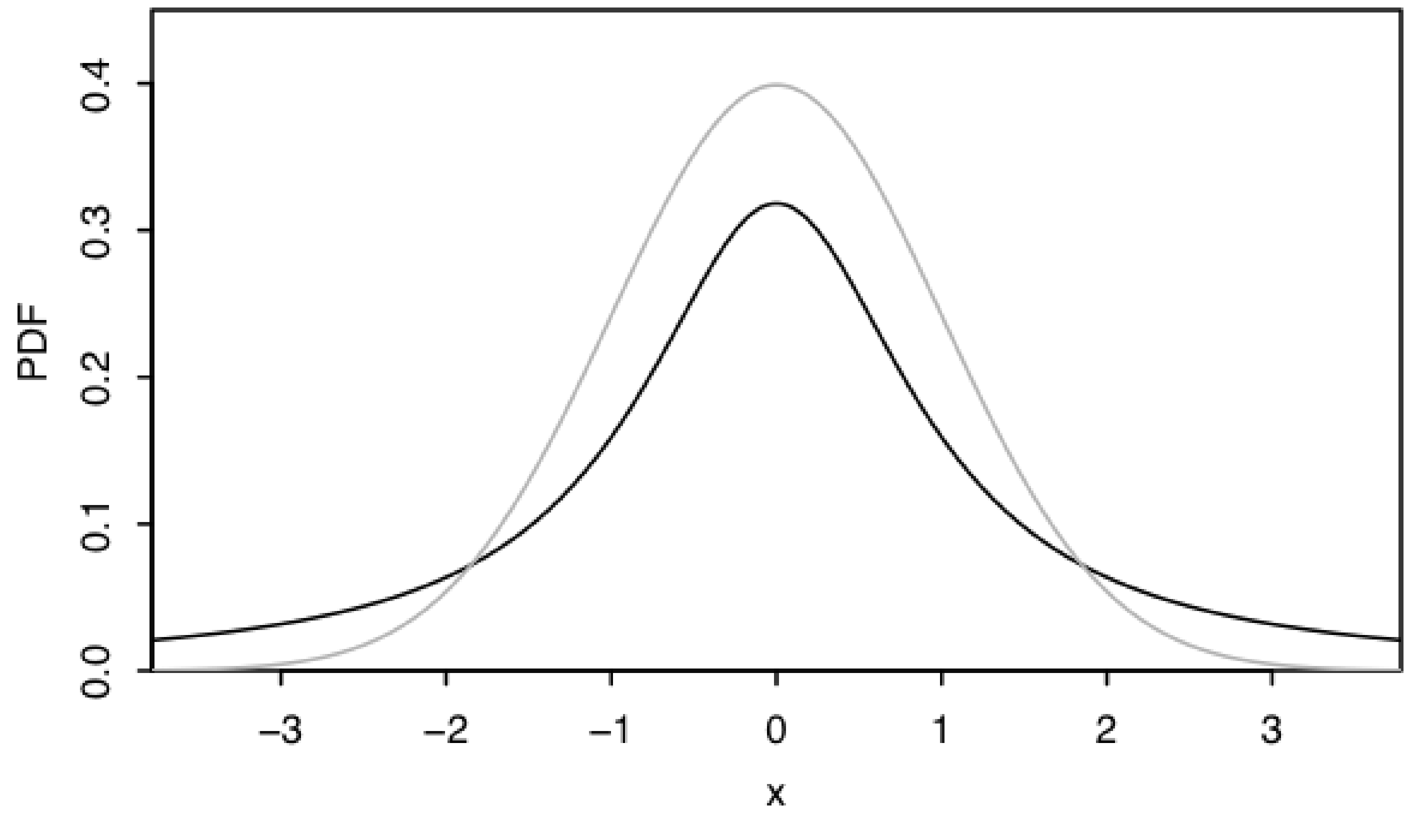

Figure 5.10.7: Cauchy PDF (black curve) compared with $\mathcal{N}(0,1)$ PDF (gray curve)

View Larger Image
Image Description

It turns out that the Cauchy does not have a finite mean or variance, so the law of large numbers and central limit theorem don't apply to the Cauchy. In fact, it turns out that the sample mean of n Cauchys is still Cauchy, no matter how large n gets! So the sample mean never approaches a Normal distribution, contrary to the behavior seen in the CLT. There is also no true mean for $\bar{X}_n$ to converge to, so LLN does not apply either.

Learn About Verified Certificates