Lecture 20: The Central Limit Theorem

Anup Rao

February 23, 2018

We discuss the central limit theorem.

There is a good reason why we should expect to see the normal distribution show up in nature quite often. That reason is the central limit theorem. The theorem asserts that the average value of n samples must converge to the normal distribution in *probability*. We will get to what that means in a second.

The Central Limit Theorem

The central limit theorem expresses the fact that whenever you take the sum of many independent identically distributed random variables, you end up with something that looks like the normal distribution. To state the theorem in way that is true requires careful thought about the right definitions to use. Before we do that, let us look at the examples shown in Figure 1, 2 and 3.

Theorem 1. *Given any real-valued distribution with expectation μ and standard deviation σ, suppose $X_1, X_2, \ldots, X_n$ are sampled independently according to this distribution and*

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then the cdf of Y_n converges to the cdf of the standard normal, in the sense that for every α,

$$\lim_{n\to\infty} p(Y_n \leq \alpha) = \int_{-\infty}^{\alpha} e^{-x^2/2}\, dx.$$

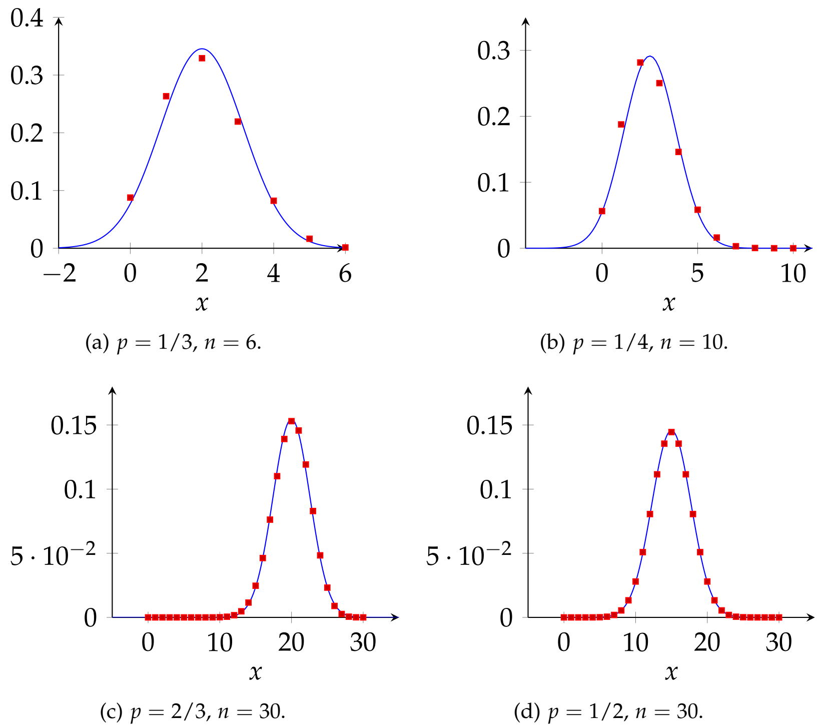
Note that $\int_{-\infty}^{\alpha} e^{-x^2/2}\, dx$ is just the cdf of the normal distribution with mean 0 and standard deviation 1. So, the theorem asserts that the cdf of the average convergers to the cdf of the standard normal.

The proof of the central limit theorem is pretty complicated, so we will not discuss it in this course.

Using the Normal Distribution as an Approximator

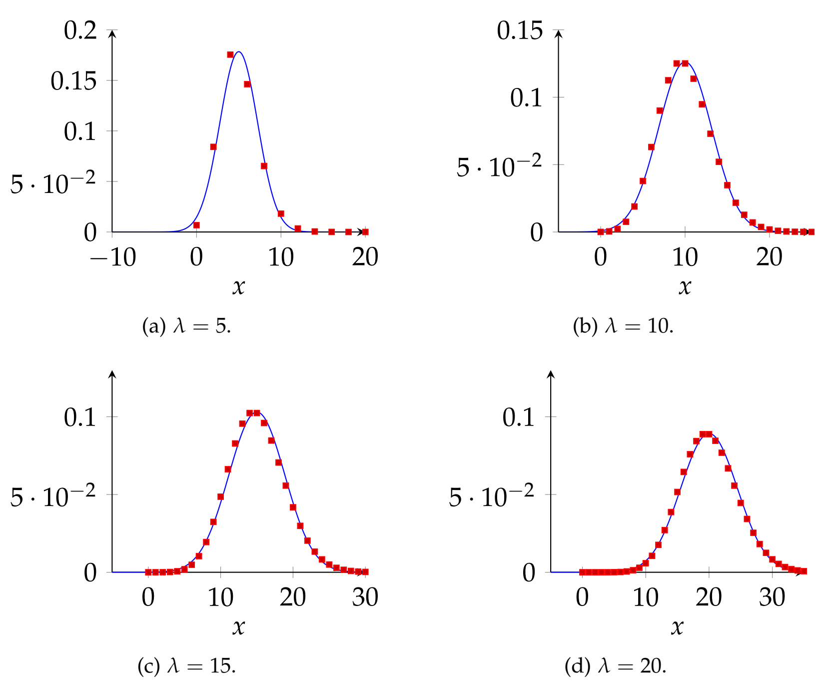
The central limit theorem suggests that when we are studying the average of many independent samples from the same distribution, then the normal distribution is a good approximator for the average.

Beware: the CLT does not say how fast the average converges to the normal, so you never really know if n is large enough for the approximation to be good.



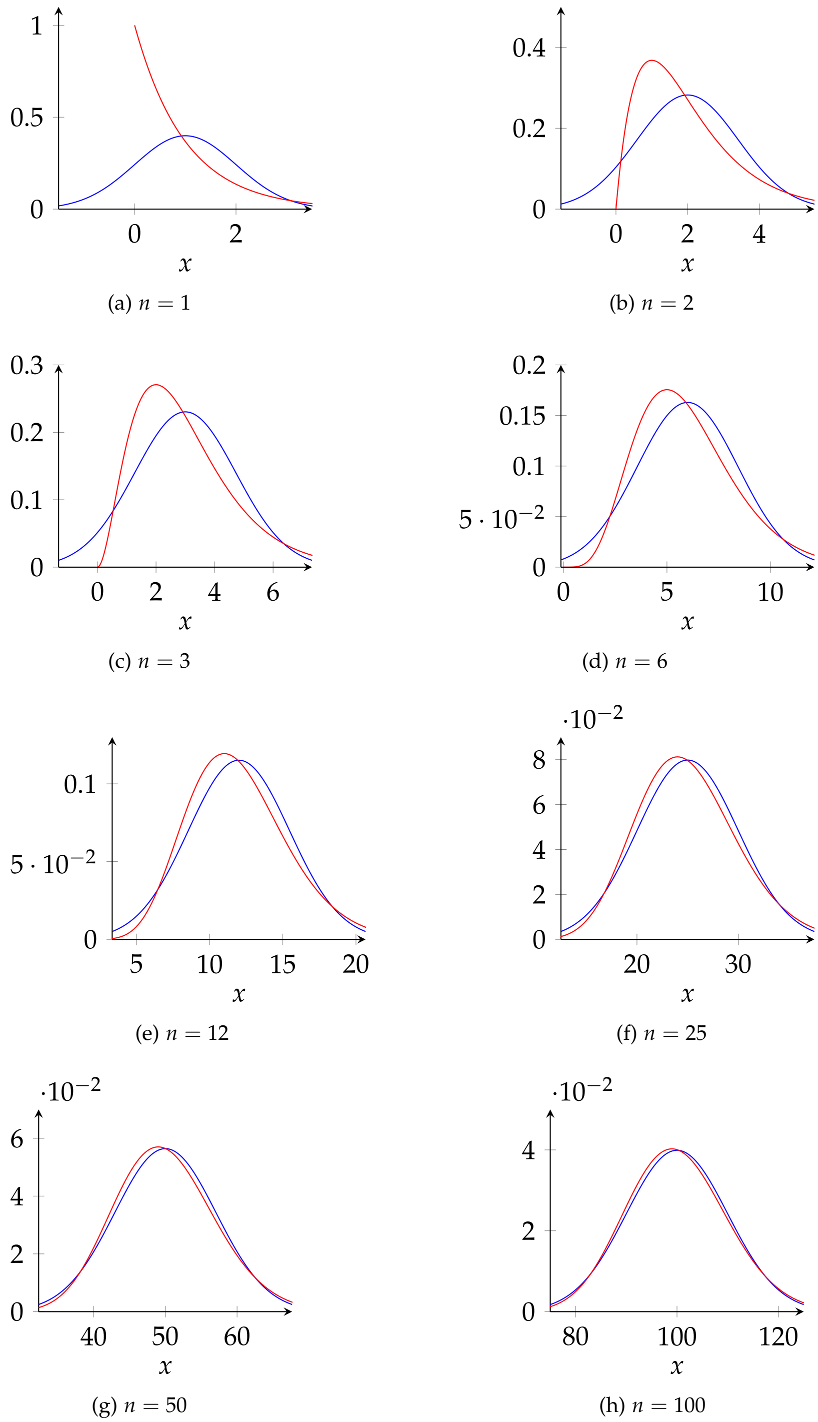
(a) $p = 1/3$, $n = 6$.
(b) $p = 1/4$, $n = 10$.
(c) $p = 2/3$, $n = 30$.
(d) $p = 1/2$, $n = 30$.

Figure 1: The CLT in pictures—several examples of binomial distributions, along with the pdf of the normal distribution with the same mean and variance are shown. In each case, the mean of the normal is set to np and the variance is set to $np(1-p)$.



(a) $\lambda = 5$.
(b) $\lambda = 10$.
(c) $\lambda = 15$.
(d) $\lambda = 20$.

Figure 2: The CLT in pictures—several examples of Poisson distributions, along with the pdf of the normal distribution with the same mean and variance are shown. Recall that the sum of two Poisson's with parameter λ is a Poisson with parameter 2λ, so as λ gets larger, the normal should approximate the Poisson, because the Poisson is the sum of many independent Poissons with smaller λ.



(a) $n = 1$

(b) $n = 2$

(c) $n = 3$

(d) $n = 6$

(e) $n = 12$

(f) $n = 25$

(g) $n = 50$

(h) $n = 100$

Figure 3: The CLT in pictures—several examples showing the pdf of the sum of n independent exponential distributions with $\lambda = 1$, along with the pdf of the normal distribution with the same mean and variance. The sum of n independent exponentials is called the Erlang distribution—it is the waiting time for the first n events to occur in the Poisson distribution.

New Variables from Old Variables

If one random variable is obtained by combining some others, there are a couple of ways one might go about figuring out the new variables pdf and cdf. For example, suppose $0 \leq X$ has pdf $f(x)$ and cdf $F(x)$, and we want to compute the pdf and cdf of $\sqrt{X}$. How can we do it?

Let us start with the cdf of $\sqrt{X}$, because this is easier. If $Z = \sqrt{X}$, then the cdf

$$G(z) = p(Z < z) = p(\sqrt{X} < z) = p(X < z^2) = F(z^2).$$

Then we can easily recover the pdf of $\sqrt{X}$ by taking the derivative of $G(z^2)$. For example, if X is uniformly distributed between 0 and 1, we get that the cdf $G(z)$ is

$$G(z) = \begin{cases} 0 & \text{if } z < 0 \text{ or } z > 1. \\ z^2 & \text{if } 0 \leq z \leq 1. \end{cases}$$

So, the pdf of $\sqrt{X}$ must be:

$$g(z) = \begin{cases} 0 & \text{if } z < 0 \text{ or } z > 1. \\ 2z & \text{if } 0 \leq z \leq 1. \end{cases}$$

We can also calculate the pdf of Z directly. Observe that the probability that Z is in the rectangle near z is $g(z)\,dz$, and this should be the same as as the probability that X is in the rectangle under x, which has area $f(x)\,dx$. So we get that

$$g(z) = \frac{dx}{dz} \cdot f(x).$$

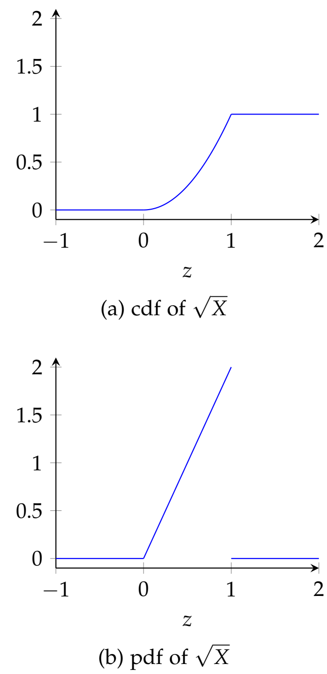
Since $X = Z^2$ in the area of interest, this gives that

$$g(z) = 2z \cdot f(z^2),$$

which gives exactly the same pdf.

Similarly, if X, Y are independent, and have pdfs $f(x)$ and $h(y)$, we can compute the pdf $g(z)$ of $Z = X + Y$ using the formula:

$$g(z) = \int_{-\infty}^{\infty} f(x)h(z-y)\,dy.$$



(a) cdf of $\sqrt{X}$

(b) pdf of $\sqrt{X}$

Figure 4: The cdf and pdf of $\sqrt{X}$, when X is uniform between 0 and 1.