The Central Limit Theorem

MAS 372 Mathematical Statistics
Sample Paper
Jane D. Student
January 2018

Careful plots of the probability histograms for binomial distributions (say, for the number of heads in n tosses of a fair coin) suggest that there is some limit histogram as $n \to \infty$. Indeed this is exactly what de Moivre noticed in the mid-1700s. In the late 1700s and early 1800s, Laplace and Gauss derived the formula for this limit histogram curve, called the ***normal distribution***, the ***bell curve***, the the ***Laplacian*** distribution, or the ***Gaussian*** distribution. Here is a way to see how to find the curve that roughly parallels Gauss' published account[1].

Consider the probability histogram for the number of heads in n tosses of a fair coin. For a given number of heads m, the probability histogram has a rectangle R with a base of width 1, centered at m, and height $\binom{n}{m}\frac{1}{2^n}$, which is the probability of getting m heads in n tosses. Now let $x = x(m)$ denote the standardized version of m, that is,

$$x = \frac{m - E(m)}{\mathrm{SD}(m)} = \frac{m - n/2}{\sqrt{n}/2} = \frac{2m - n}{\sqrt{n}}.$$

In the histogram with standard units, the rectangle R transforms into R' with base of width $2/\sqrt{n}$ and height $\binom{n}{m}\frac{1}{2^n}\frac{\sqrt{n}}{2}$ (the total area remains the same).

Let $y = y(x)$ denote the unknown continuous limit curve of the histogram for coin tosses in standard units, that is,

$$y(x) = \lim_{n\to\infty} \binom{n}{m} \frac{1}{2^n} \frac{\sqrt{n}}{2}. \tag{1}$$

Notice that $m = m(x) = \frac{x\sqrt{n}+n}{2}$ is not constant, and worse, this value is not likely to be an integer. We could hope that

$$y(x) = \lim_{n\to\infty} \binom{n}{\lfloor \frac{x\sqrt{n}+n}{2} \rfloor} \frac{1}{2^n} \frac{\sqrt{n}}{2}.$$

might work, but even so, the above limit turns out to be resistant to direct attack. So we need an indirect method. Success comes, as it so often does, by considering rates of change. We will continue to work with the approximation $w_n(x)$ for $y(x)$ given by

$$w_n(x) = \binom{n}{\lfloor \frac{x\sqrt{n}+n}{2} \rfloor} \frac{1}{2^n} \frac{\sqrt{n}}{2}.$$

It is natural in science to examine *relative change* of a function, which is the quantity $\frac{\Delta y}{y} \approx \frac{y'\,dx}{y}$. This in turn motivates examining the function y'/y. The idea is that if you can discover a law of the form

$$\frac{y'}{y} = f(x)$$

then you can integrate to get a solution.

$$\ln y = \int f(x)\,dx + C$$

$$y = Ae^{\int f(x)\,dx}$$

We will show that the limit function y in (1) satisfies $y'/y = -x$ by showing that an approximation for the left side approaches x as $n \to \infty$, that is,

$$\lim_{n\to\infty} \frac{w_n'}{w_n} = -x. \tag{2}$$

First we need to approximate w_n' from it's discrete values. Using x values $x(u)$ and $x(u+1)$ (for some integer $u = \lfloor m(x) \rfloor$, so u depends on x and n) we have

$$w_n' \approx \frac{\Delta w_n}{\Delta x} = \frac{w_n(x(u+1)) - w_n(x(u))}{1/(\sqrt{n}/2)} \approx \frac{\left[\binom{n}{u+1} - \binom{n}{u}\right]}{2^n} \frac{n}{4}. \tag{3}$$

Then we have

$$\frac{w_n'}{w_n} \approx \frac{\binom{n}{u+1} - \binom{n}{u}}{\binom{n}{u}} \frac{\sqrt{n}}{2} = \frac{n - 2u - 1}{u+1} \frac{\sqrt{n}}{2}.$$

Now we substitute $u = \frac{x\sqrt{n}+n}{2}$ and we have (after simplifying)

$$\frac{w_n'}{w_n} \approx -\frac{xn+1}{n + x\sqrt{n} + 2}$$

so we see that $w_n'/w_n \to -x$ as $n \to \infty$. Thus it is reasonable to suppose that if the limit curve y exists, then it must satisfy

$$\frac{y'}{y} = -x \tag{4}$$

from which we readily find the solution

$$y = Ae^{-x^2/2}. \tag{5}$$

A standard exercise in multivariable calculus shows that

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}$$

so we have $A = 1/\sqrt{2\pi}$. This is the normal curve.

This concludes an outline of a method for deriving the standard normal distribution as a limit of binomial distributions using a method that is in the spirit of that used by Gauss.

References

[1] C.F. Gauss. *Theoria motvs corporvm coelestivm in sectionibvs conicis Solem ambientivm [Theory of the motion of the heavenly bodies moving about the Sun in conic sections]*. 1809. Sections 175,176,177 develop the formula for the normal standard normal distribution.