Chapter 5

Normal approximation to the Binomial

In 1733, Abraham de Moivre presented an approximation to the Binomial distribution. He later appended the derivation of his approximation to the solution of a problem asking for the calculation of an expected value for a particular game. He posed the rhetorical question of how we might show that experimental proportions should be close to their expected values.

A passage from De Moivre

Corollary.

From this it follows, that if after taking a great number of Experiments, it should be perceived that the happenings and failings have been nearly in a certain proportion, such as of 2 to 1, it may safely be concluded that the Probabilities of happening or failing at any one time assigned will be very near in that proportion, and that the greater the number of Experiments has been, so much nearer the Truth will the conjectures be that are derived from them.

But suppose it should be said, that notwithstanding the reasonableness of building Conjectures upon Observations, still considering the great Power of Chance, Events might at long run fall out in a different proportion from the real Bent which they have to happen one way or the other; and that supposing for Instance that an Event might as easily happen as not happen, whether after three thousand Experiments it may not be possible it should have happened two thousand times and failed a thousand; and that therefore the Odds against so great a variation from Equality should be assigned, whereby the Mind would be the better disposed in the Conclusions derived from the Experiments.

In answer to this, I'll take the liberty to say, that this is the hardest Problem that can be proposed on the Subject of Chance, for which reason I have reserved it for the last, but I hope to be forgiven if my Solution is not fitted to the capacity of all Readers; however I shall derive from it some Conclusions that may be of use to every body: in order thereto, I shall here translate a Paper of mine which was printed November 12, 1733, and communicated to some Friends, but never yet made public, reserving to myself the right of enlarging my own Thoughts, as occasion shall require.

Novemb. 12, 1733

A Method of approximating the Sum of the Terms of the Binomial $\overline{a+b}\backslash^n$ expanded into a Series, from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments.

Altho' the Solution of problems of Chance often requires that several Terms of the Binomial $\overline{a+b}\backslash^n$ be added together, nevertheless in very high Powers the thing appears so laborious, and of so great difficulty, that few people have undertaken that Task; for besides **James** *and* **Nicolas Bernouilli**, *two great Mathematicians, I know of no body that has attempted it; in which, tho' they have shown very great skill, and have the praise that is due to their Industry, yet some things were further required; for what they have done is not so much an Approximation as the determining very wide limits, within which they demonstrated that the Sum of the Terms was contained. Now the method ...*

A. De Moivre, *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*, 3rd edition (1756), pages 242–243. (Photographic reprint of final edition by Chelsea Publishing Company, 1967.)

This Chapter will explain de Moivre's approximation.

Suppose X_n has a Bin(n, p) distribution. That is,

$$b_n(k) := \mathbb{P}\{X_n = k\} = \binom{n}{k} p^k q^{n-k} \qquad \text{for } k = 0, 1, \ldots, n, \text{ where } q = 1 - p,$$

What does the distribution look like?

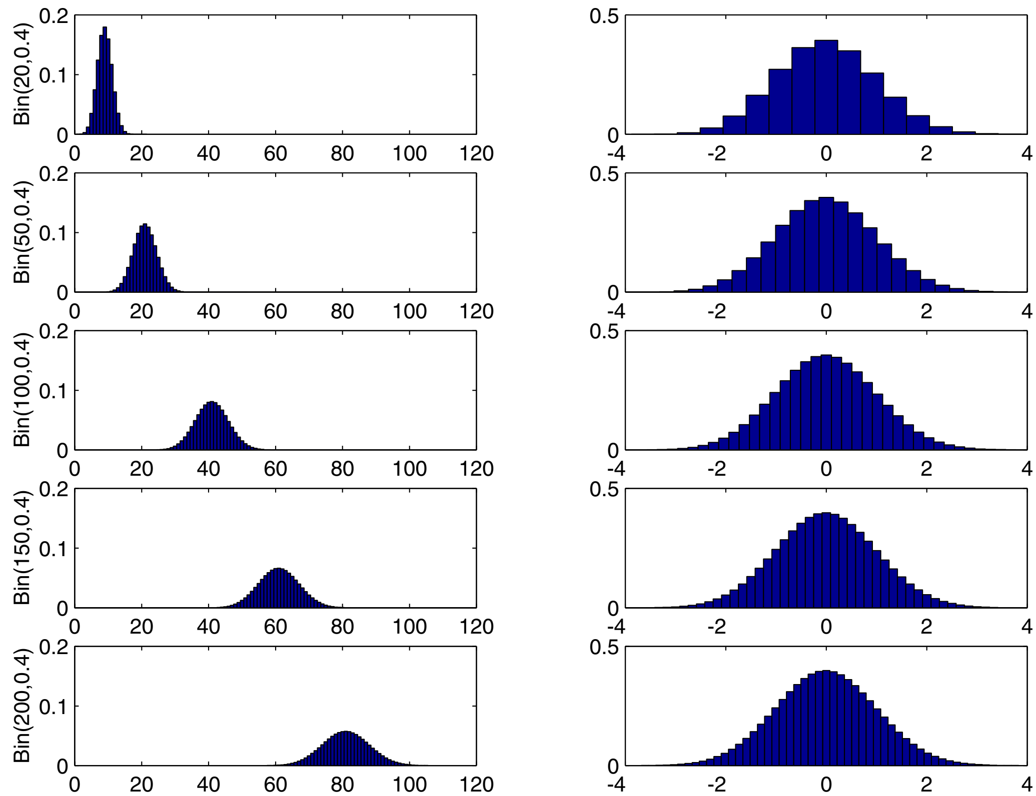
Recall that Tchebychev's inequality suggests the distribution should be clustered around the expected value, np, with a spread determined by the standard deviation, $\sigma_n = \sqrt{npq}$. Also, from Problem sheet 4, you know that the probabilities $b_n(k)$ for $k = 0, 1, \ldots, n$ achieve their maximum vaue at a value $k_{\max}$ close to np. Moreover, the values of $b(k)$ are increasing for $k < k_{\max}$ and decreasing for $k > k_{\max}$, facts that follow from the simple expression for the ratio of successive terms:

<5.1> $$\frac{b_n(k)}{b_n(k-1)} = \frac{(n-k+1)p}{kq} \qquad \text{for } k = 1, 2, \ldots, n.$$

The plots on the left-hand side of the next display, for the Bin$(n, 0.4)$ distribution with $n = 20, 50, 100, 150, 200$, illustrate this behavior. Each plot shows bars of height $b_n(k)$ and width 1, centered at k. The maxima occur near $n \times 0.4$ for each plot. As n increases, the spread also increases, reflecting the increase in the standard deviations $\sigma_n = \sqrt{npq}$ for $p = 0.4$. Each of the shaded regions has area

$$\sum\nolimits_{k=0}^{n} b_n(k) = 1 \qquad \text{for various } n.$$

The location and scaling effects of the increasing expected values and standard deviations (with $p = 0.4$ and various n) is removed from the plots on the right-hand side of the display. Each plot is shifted to bring the location of the maximum to 0 and the horizontal scale is multiplied by a factor $1/\sigma_n$. Now a bar of height $\sigma_n \times b_n(k)$ with width $1/\sigma_n$ is centered at $(k - np)/\sigma_n$. The plots all have similar shapes. Each shaded region still has area 1.



Notice how the plots on the right settle down to a symmetric 'bell-shaped' curve. The shape of the "standardized" Binomial quickly stablizes as n increases.

De Moivre established this stability mathematically, by showing that

$$\mathbb{P}\{X = k_{\max} + m\} \approx b(k_{\max}) \exp\left(-\frac{m^2}{2npq}\right).$$

Here, and subsequently, I translate de Moivre's results into modern notation. Also, I omit the subscript n from the $b_n(k)$ symbol, because n will stay fixed duringthe explanation.

De Moivre's approximation is largely explained by two simple facts:

$$\log(1+x) \approx x \qquad \text{for } x \text{ near } 0,$$
$$1 + 2 + 3 + \ldots + m = \tfrac{1}{2}m(m+1) \approx \tfrac{1}{2}m^2.$$

The ratio in equality <5.1> takes a simpler form if we replace k by $k_{\max} + i$.

$$\frac{b(k_{\max}+i)}{b(k_{\max}+i-1)} = \frac{(n-k_{\max}-i+1)p}{(k_{\max}+i)q} \approx \frac{(nq-i)p}{(np+i)q} = \frac{1-i/(nq)}{1+i/(np)}.$$

The logarithm of the last ratio equals

$$\log\left(1-\frac{i}{nq}\right) - \log\left(1+\frac{i}{np}\right) \approx -\frac{i}{nq} - \frac{i}{np} = -\frac{i}{npq}.$$

By summing such terms we get an approximation for the logarithm of the ratio of the individual Binomial probabilities to their largest term. For example, if $m \geq 1$ and $k_{\max} + m \leq n$,

$$\log\frac{b(k_{\max}+m)}{b(k_{\max})} = \log\left(\frac{b(k_{\max}+1)}{b(k_{\max})} \times \frac{b(k_{\max}+2)}{b(k_{\max}+1)} \times \ldots \times \frac{b(k_{\max}+m)}{b(k_{\max}+m-1)}\right)$$

$$
\begin{aligned}
&= \log \frac{b(k_{\max}+1)}{b(k_{\max})} + \log \frac{b(k_{\max}+2)}{b(k_{\max}+1)} + \ldots + \log \frac{b(k_{\max}+m)}{b(k_{\max}+m-1)} \\
&\approx \frac{-1-2-\ldots-m}{npq} \\
&\approx -\tfrac{1}{2}\frac{m^2}{npq}.
\end{aligned}
$$

Thus

$$\mathbb{P}\{X = k_{\max}+m\} \approx b(k_{\max}) \exp\left(-\frac{m^2}{2npq}\right) \qquad \text{for } m \text{ not too large.}$$

An analogous approximation holds for $0 \le k_{\max}+m \le k_{\max}$.

Using the fact that the probabilities sum to 1, de Moivre was also able to show for $p = 1/2$ that the $b(k_{\max})$ should decrease like $2/(B\sqrt{n})$, for a constant B that he was initially only able to express as an infinite sum. Referring to his calculation of the ratio of the maximum term in the expansion of $(1+1)^n$ to the sum, 2^n, he wrote (page 244 of the *Doctrine of Chances*):

> *When I first began that inquiry, I contented myself to determine at large the Value of B, which was done by the addition of some Terms of the above-written Series; but as I perceived that it converged but slowly, and seeing at the same time that what I had done answered my purpose tolerably well, I desisted from proceeding further till my worthy and learned Friend Mr. James Stirling, who had applied himself after me to that inquiry, found that the Quantity B did denote the Square-root of the Circumference of a Circle whose Radius is Unity, so that if that Circumference be called c, the Ratio of the middle Term to the Sum of all the Terms will be expressed by $\frac{2}{\sqrt{nc}}$.*

For positive integers n, the Stirling formula asserts that

$$n! \approx \sqrt{2\pi}\, n^{n+1/2} e^{-n}$$

in the sense that the ratio of both sides tends to 1 as n goes to infinity. (See the derivation at the end of this Chapter.) Consequently,

$$
\begin{aligned}
b(k) &= \frac{n!}{k!(n-k)!} p^k q^{n-k} \\
&\approx \frac{n^{n+1/2} e^{-n+k+n-k}}{\sqrt{2\pi}\,(np)^{k+1/2}(nq)^{n-k+1/2}} p^k q^{n-k} \qquad \text{if } k \approx np \\
&= \frac{1}{\sqrt{2\pi npq}}
\end{aligned}
$$

De Moivre's approximation becomes, for general p,

$$\mathbb{P}\{X_n = k_{\max}+m\} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{m^2}{2npq}\right),$$

or, substituting np for $k_{\max}$ and writing k for $k_{\max}+m$,

$$\mathbb{P}\{X_n = k\} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right) = \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left(-\frac{(k-np)^2}{2\sigma_n^2}\right).$$

That is, $\mathbb{P}\{X_n = k\}$ is approximately equal to the area under the smooth curve

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left(-\frac{(x-np)^2}{2\sigma_n^2}\right),$$

for the interval $k - 1/2 \le x \le k + 1/2$. (The length of the interval is 1, so it does not appear in the previous display.)

Similarly, for each pair of integers with $0 \le a < b \le n$,

$$\mathbb{P}\{a \le X_n \le b\} = \sum\nolimits_{k=a}^{b} b_n(k) \approx \sum\nolimits_{k=a}^{b} \int_{k-1/2}^{k+1/2} f(x)\,dx = \int_{a-1/2}^{b+1/2} f(x)\,dx.$$

A change of variables, $y = (x - np)/\sigma_n$, simplifies the last integral to

$$\frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-y^2/2} dy \qquad \text{where } \alpha = (a - np - 1/2)/\sigma_n \text{ and } \beta = (b - np + 1/2)/\sigma_n.$$

REMARK. It usually makes little difference to the approximation if we omit the $\pm 1/2$ terms from the definitions of α and β.

How does one actually perform a normal approximation? Back in the olden days, one would interpolate from a table of values for the function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy,$$

which was found in most statistics texts. For example, if X has a Bin(100, 1/2) distribution,

$$\mathbb{P}\{45 \le X \le 55\} \approx \Phi\left(\frac{55.5 - 50}{5}\right) - \Phi\left(\frac{44.5 - 50}{5}\right) \approx 0.8643 - 0.1356 = 0.7287$$

These days, I would just calculate in R:

```
> pnorm(55.5, mean = 50, sd = 5) - pnorm(44.5, mean = 50, sd = 5)
[1] 0.7286679
```

or use another very accurate, built-in approximation:

```
> pbinom(55,size = 100, prob = 0.5) - pbinom(44,size = 100, prob = 0.5)
[1] 0.728747
```

THE MYSTERIOUS $\sqrt{2\pi}$

Notice that, for the Binomial$(n, 1/2)$ distribution with n very large,

$$1 = \mathbb{P}\{0 \le X_n \le n\} \approx \Phi(\sqrt{n}) - \Phi(-\sqrt{n}) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy$$

In fact, the constant $C = \int_{-\infty}^{\infty} \exp(-x^2/2)\,dx$ is exactly equal to $\sqrt{2\pi}$. Equivalently, the constant $C^2 = \iint \exp(-(x^2 + y^2)/2)\,dx\,dy$ equal to 2π. (Here, and subsequently, the double integral runs over the whole plane.) We can evaluate this double integral by using a small Calculus trick.

Using the fact that

$$\int_0^\infty \mathbb{I}\{r \le z\} e^{-z}\,dz = e^{-r} \qquad \text{for } r > 0,$$

we may rewrite C^2 as a triple integral: replace r by $(x^2 + y^2)/2$, then substitute into the double integral to get

$$C^2 = \iint \left(\int_0^\infty \mathbb{I}\{x^2 + y^2 \le 2z\} e^{-z}\,dz \right) dx\,dy = \int_0^\infty \left(\iint \mathbb{I}\{x^2 + y^2 \le 2z\}\,dx\,dy \right) e^{-z}\,dz.$$

With the change in the order of integration, the double integral is now calculating the area of a circle centered at the origin and with radius $\sqrt{2z}$. The triple integral reduces to

$$\int_0^\infty \pi \left(\sqrt{2z}\right)^2 e^{-z}\,dz = \int_0^\infty \pi 2z e^{-z}\,dz = 2\pi.$$

That is, $C = \sqrt{2\pi}$, as asserted.

Stirling's Formula

For positive integers n, the formula asserts that

<5.2> $$n! \approx \sqrt{2\pi}\, n^{n+1/2} \exp(-n),$$

in the sense that the ratio of both sides tends to 1 as n goes to infinity.

As the first step towards a proof, write

$$\log n! = \log 1 + \log 2 + \ldots + \log n$$

as a sum of integrals of indicator functions:

$$\log n! = \sum_{i=1}^{n} \int_1^n \mathbb{I}\{1 \le x < i\} \frac{1}{x}\, dx = \int_1^n \sum_{i=1}^{n} \mathbb{I}\{1 \le x < i\} \frac{1}{x}\, dx$$

The sum of indicator functions counts the number of integers in the range $1, 2, \ldots, n$ that are greater than x. It equals $n - \lfloor x \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x. The difference $\psi(x) = x - \lfloor x \rfloor$ lies in the range $[0, 1)$; it gives the fractional part of x.

The integral representating $\log(n!)$ is equal

$$\int_1^n \frac{n - \lfloor x \rfloor}{x}\, dx = \int_1^n \frac{n - x + \psi(x)}{x}\, dx = n \log n - n + \int_1^n \frac{\psi(x)}{x}\, dx.$$

The last integral diverges as n tends to infinity, because the contribution from the interval $[i, i+1)$ equals

$$\int_i^{i+1} \frac{x - i}{x}\, dx = \int_0^1 \frac{t}{t+i}\, dt \approx \frac{1}{2i}.$$

For the approximation I have treated the $t + i$ in the denominator as approximately equal to i and then noted that $\int_0^1 t\, dt = 1/2$. The sum of the contributions from the integral involving ψ increases like $\frac{1}{2} \log n$.

It seems we have to subtract off an extra $\frac{1}{2} \log n = \frac{1}{2} \int_1^n \frac{1}{x}\, dx$ to keep the remainder term under control. Splitting the integral into contributions from intervals $[i, i+1)$, we then get

<5.3> $$\log(n!) - (n + 1/2) \log n - n = \sum_{i=1}^{n} \int_0^1 \frac{t - 1/2}{t+i}\, dt$$

With the subtraction of the $1/2$ we will get some cancellation between the negative contribution for $0 \le t \le 1/2$ and the positive contribution for $1/2 < t \le 1$.

Make the change of variable $s = 1/2 - t$ for the integral over $[0, 1/2]$, and the change of variable $s = t - 1/2$ over $(1/2, 1]$.

$$\begin{aligned} \int_0^1 \frac{t - 1/2}{t+i}\, dt &= \int_0^{1/2} \frac{-s}{i + 1/2 - s}\, ds + \int_0^{1/2} \frac{s}{i + 1/2 + s}\, ds \\ &= -2 \int_0^{1/2} \frac{s^2}{(i + 1/2)^2 - s^2}\, ds. \end{aligned}$$

The last expression is bounded in absolute value by i^{-2}. The sum of the integrals forms a convergent series. That is, for some constant c,

$$\int_1^n \frac{\psi(x) - 1/2}{x}\, dx \to c \qquad \text{as } n \to \infty.$$

Equivalently, from <5.3>,

$$\frac{n!}{n^{n+1/2} e^{-n}} \to e^c \qquad \text{as } n \to \infty$$

This result is equivalent to formula <5.2>, except for the identification of e^c as the constant $\sqrt{2\pi}$. See the discussion on the next page for a way of deriving the value of the constant.

For an alternative derivation of Stirling's formula, see Feller *An Introduction to Probability Theory and Its Applications*, volume I, third edition, pages 52–53.