MthStat 465, Spring 2005, Lecture Number 13
Normal Approximation to the Binomial Distribution.

We show how the normal distribution arises as an approximation to the binomial distribution. This is summarized by the DeMoivre-Laplace Theorem, which we will take up in the next lecture.

1. Two building blocks

We need two results. The first is Stirling's formula:

Theorem 1 (Stirling's Formula)**.**

$$\lim_{N\to\infty} \frac{N!}{N^N e^{-N}\sqrt{2\pi}} = 1. \tag{1}$$

Proof. For a proof see Ash, *Basic Probability Theory.* □

We are concerned with the symmetric binomial probabilities. In particular, in $2N$ tosses of a fair coin, what is the probability that the number of heads equals the number of tails? Stirling's Formula suggests the following algebraic manipulation:

$$\begin{aligned}
\binom{2N}{N}\left(\frac{1}{2}\right)^{2N} &= \frac{(2N)!}{N!N!}\frac{1}{2^{2N}} \\
&= \frac{(2N)!}{(2N)^{2N}e^{-2N}\sqrt{2\pi 2N}}\left(\frac{N^N e^{-N}\sqrt{2\pi N}}{N!}\right)^2 \\
&\quad \times\frac{(2N)^{2N}e^{-2N}\sqrt{2\pi 2N}}{(N^N e^{-N}\sqrt{2\pi N})^2}\frac{1}{2^{2N}} \\
&= \frac{(2N)!}{(2N)^{2N}e^{-2N}\sqrt{2\pi 2N}}\left(\frac{N^N e^{-N}\sqrt{2\pi N}}{N!}\right)^2 \frac{1}{\sqrt{N\pi}}
\end{aligned}$$

from which we see that

$$\lim_{N\to\infty} \sqrt{N\pi}\binom{2N}{N}\left(\frac{1}{2}\right)^{2N} = 1. \tag{2}$$

We can restate (2) as

$$\binom{2N}{N}\left(\frac{1}{2}\right)^{2N} \approx \frac{1}{\sqrt{\pi N}}. \tag{3}$$

The other tool we need is that for $|x| << 1$ we have

$$\log(1+x) \approx x. \tag{4}$$

This should be familiar from calculus.

2. Where the normal distribution comes in

The idea is to measure all the other binomial probabilities as ratios with the left hand side of (3). We consider for each $k \in \{-N, \dots, N\}$ the ratio

$$\begin{aligned}
\frac{\begin{pmatrix} 2N \\ N+K \end{pmatrix} \left(\frac{1}{2}\right)^{2N}}{\begin{pmatrix} 2N \\ N \end{pmatrix} \left(\frac{1}{2}\right)^{2N}} &= \frac{\begin{pmatrix} 2N \\ N+K \end{pmatrix}}{\begin{pmatrix} 2N \\ N \end{pmatrix}} \\
&= \frac{(2N)!}{(N+k)!(N-k)!} \frac{N!N!}{(2N)!} \\
&= \frac{N!}{(N-k)!} \frac{N!}{(N+k)!} \\
&= \frac{N(N-1)\cdots(N-(k-1))}{(N+k)(N+k-1)\cdots(N+1)} \\
&= \frac{N(N-1)\cdots(N-(k-1))N^{-k}}{(N+k)(N+k-1)\cdots(N+1)N^{-k}} \\
&= \frac{(1-\frac{1}{N})\cdots(1-\frac{k-1}{N})}{(1+\frac{k}{N})(1+\frac{k-1}{N})\cdots(1+\frac{1}{N})}
\end{aligned}$$

so

$$\frac{\begin{pmatrix} 2N \\ N+K \end{pmatrix} \left(\frac{1}{2}\right)^{2N}}{\begin{pmatrix} 2N \\ N \end{pmatrix} \left(\frac{1}{2}\right)^{2N}} = \frac{(1-\frac{1}{N})\cdots(1-\frac{k-1}{N})}{(1+\frac{k}{N})(1+\frac{k-1}{N})\cdots(1+\frac{1}{N})}. \tag{5}$$

We now estimate each of these products by first estimating their logarithms using (4). The key here is that we have to make sure that $|k| << N$ so that the approximations are good enough.

$$\begin{aligned}
\log\left((1-\frac{1}{N})\cdots(1-\frac{k-1}{N})\right) &= \log\left(1-\frac{1}{N}\right) + \cdots + \log\left(1-\frac{k-1}{N}\right) \\
&\approx -\left(\frac{1}{N} + \cdots + \frac{k-1}{N}\right) \\
&= -\frac{(k-1)k}{2N} \\
&= \frac{-k^2+k}{2N}
\end{aligned}$$

so

$$(1-\frac{1}{N})\cdots(1-\frac{k-1}{N}) \approx \exp\left(\frac{-k^2+k}{2N}\right), \tag{6}$$

where $\exp(x) = e^x$. Similarly,

$$\begin{aligned}
\log\left((1+\frac{k}{N})\cdots(1+\frac{1}{N})\right) &= \log\left(1+\frac{k}{N}\right)+\cdots+\log\left(1+\frac{1}{N}\right) \\
&\approx \left(\frac{k}{N}+\cdots+\frac{1}{N}\right) \\
&= \frac{k(k+1)}{2N} \\
&= \frac{k^2+k}{2N},
\end{aligned}$$

so

$$(1+\frac{k}{N})\cdots(1+\frac{1}{N}) \approx \exp\left(\frac{k^2+k}{2N}\right). \tag{7}$$

Combining (5), (6) and (7) we have

$$\frac{\binom{2N}{N+K}\left(\frac{1}{2}\right)^{2N}}{\binom{2N}{N}\left(\frac{1}{2}\right)^{2N}} \approx \exp\left(-\frac{k^2}{N}\right),$$

or, more usefully,

$$\binom{2N}{N+K}\left(\frac{1}{2}\right)^{2N} \approx \binom{2N}{N}\left(\frac{1}{2}\right)^{2N}\exp\left(-\frac{k^2}{N}\right). \tag{8}$$

Combining (3) with (8) and cleverly re-arranging, we arrive at

$$\binom{2N}{N+K}\left(\frac{1}{2}\right)^{2N} = \frac{1}{\sqrt{\pi}}\exp\left(-\left(\frac{k}{\sqrt{N}}\right)^2\right)\frac{1}{\sqrt{N}}. \tag{9}$$

So, if we are looking at an expression of the form

$$\sum_{k=A} B\binom{2N}{N+K}\left(\frac{1}{2}\right)^{2N}$$

we see it can be approximated by a sum

$$\sum_{k=A}^{B}\frac{1}{\sqrt{\pi}}\exp\left(-\left(\frac{k}{\sqrt{N}}\right)^2\right)\frac{1}{\sqrt{N}}.$$

which is a Riemann sum for an integral of the form

$$\int_a^b \frac{1}{\sqrt{\pi}}e^{-x^2}\,dx,$$

provided we can show that all the approximations are good enough. The relationship between a and A is that $a = A/\sqrt{N}$ and, similarly, $b = B/\sqrt{N}$.