De Moivre's Formula

Peter G. Doyle

Version 1.0.2, 26 May 2010
GNU FDL*

One of the earliest triumphs of probability theory was de Moivre's discovery in 1756 of the way the 'bell curve' governs coin tossing (peek ahead to Figure 1). We discuss here the simplest case of this wide-reaching phenomenon, using ideas that can be extended to more general cases. The tools we need are calculus, and specifically Taylor series; some trigonometry(!); and some simple combinatorial probability of the kind taught in high school algebra courses.

If you glance ahead, you'll see that there are a lot of formulas coming up. Lest you lose heart, you should know that in the end, everything comes down to the fact that if you plot a high power of $\cos(x)$, the result looks very much like a bell curve (peek ahead at Figure 2). And that's because

$$\cos(x)^n = (1 - \frac{x^2}{2} + \ldots)^n \approx (1 - \frac{x^2}{2})^n = (1 - \frac{1}{n}\frac{nx^2}{2})^n \approx e^{-nx^2/2}.$$

Here's we've used the Taylor series for $\cos(x)$, and the formula for continuously compounded interest:

$$\lim_{n\to\infty}(1 + \frac{a}{n})^n = e^a.$$

There's nothing very special about $\cos(x)$ here: A bell curve emerges when you plot powers of any generic function with a maximum at 0. This universality of the bell curve when taking powers dictates the universality of

*Copyright (C) 2006 Peter G. Doyle. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, as published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

the bell curve in probability through the agency of the exponential function, which converts addition into multiplication, and repeated addition into taking powers.

One warning: An honest approach to de Moivre's formula introduces the function $\cos(x)$ by way of the complex-valued exponential function $e^{ix} = \cos(x) + i\sin(x)$. These complex exponentials are the standard tools of 'Fourier analysis', which is what we'll secretly be using here. To keep the discussion elementary, we'll skip the introduction, and just start working with $\cos(x)$. Once the elementary discussion is finished, we'll go back over the argument, and show how much simpler it becomes when we take a more honest approach using complex exponentials. Hopefully the effect will be to convince you of the wisdom of John Conway's adage, 'Imaginary numbers are real! Complex numbers are simple!'

An even-more-honest approach to de Moivre's formula uses, in addition to complex numbers, basic probability concepts that go beyond what is taught in high school. We'll save this fancier approach for last, because even our most elementary approach contains the seeds of the key ideas.

Elementary approach

Peter and Paul play a game called *heads or tails*. In this game, a fair coin is tossed n times. Each time heads comes up Peter wins 1 penny from Paul, and each time tails comes up Peter loses 1 penny to Paul. For example, with $n = 40$, one possible outcome would be

$$\text{THTHHHHTTHTHHTTHHTTTTHHHTHHTHHHTHHHTTTHH}.$$

In this case Peter has won 6 pennies from Paul, because heads has come up $23 = \frac{40+6}{2}$ times and tails $17 = \frac{40-6}{2}$ times.

Is this outcome typical? Well, there are 2^{40} possible outcomes for a sequence of 40 tosses, each outcome having probability $1/2^{40}$. The number of outcomes with 23 heads and 17 tails is

$$\binom{40}{23} = \frac{40!}{23!17!} = 88732378800.$$

The total probability for Peter to net 6 pennies is

$$\frac{1}{2^{40}}\binom{40}{23} = \frac{88732378800}{1099511627776} = \frac{5545773675}{68719476736} \approx 0.080701628394308499992.$$

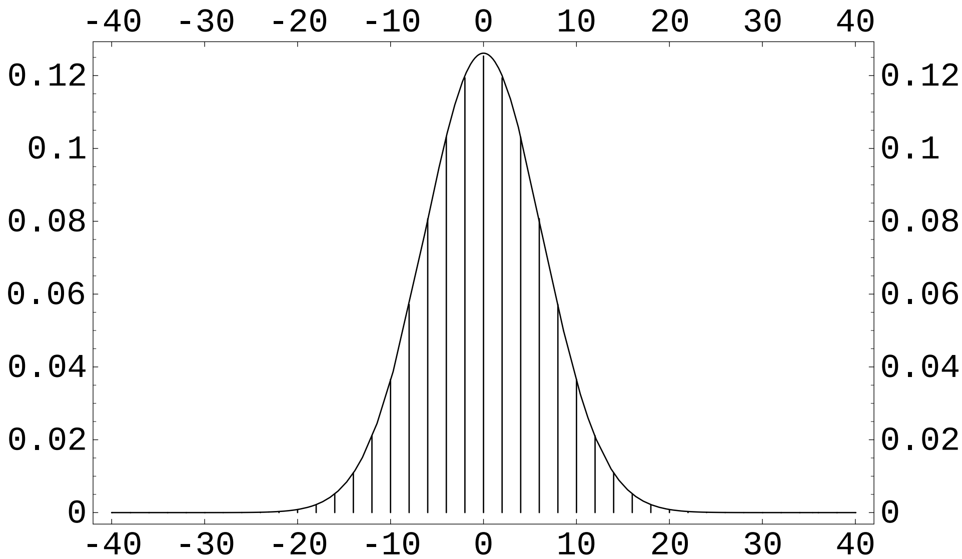

Figure 1: $w_{n,j}$ compared to a bell curve.

So yes, the outcome appears to be fairly typical, at least as far as Peter's net winnings go.

In general, the probability $w_{n,j}$ of netting j pennies in n tosses will be

$$w_{n,j} = \frac{1}{2^n}\binom{n}{\frac{n+j}{2}}$$

as long as $n + j$ is even, i.e., as long as n and j have the same parity (either both even or both odd). If we interpret $\binom{n}{k}$ to be 0 when k is a not an integer, then this formula holds for all n and j regardless of parity.

Using this formula, we can plot the values of $w_{40,j}$—see Figure 1. Note the remarkable agreement with a suitably scaled version of the 'bell-shaped curve' $y = e^{-x^2}$. The approximation to a bell curve gets better and better as n increases. We want to formulate and prove a statement indicating how and why this happens.

The values of $w_{n,j}$ for small n, j are shown in Table 1. Note how the numerators form a sideways Pascal's triangle. Taking the denominators into account, each entry in the table (other than the entry $w_{0,0} = 1$) is the average of the two numbers directly to the southwest and northwest of it. (In the case of entries for $w_{n,\pm n}$ bear in mind that blank entries in the table indicate

j \ n	0	1	2	3	4
5					
4					$\frac{1}{16}$
3				$\frac{1}{8}$	
2			$\frac{1}{4}$		$\frac{4}{16}$
1		$\frac{1}{2}$		$\frac{3}{8}$	
0	1		$\frac{2}{4}$		$\frac{6}{16}$
-1		$\frac{1}{2}$		$\frac{3}{8}$	
-2			$\frac{1}{4}$		$\frac{4}{16}$
-3				$\frac{1}{8}$	
-4					$\frac{1}{16}$
-5					

Table 1: Probability of winning j pennies in n tosses. Entries with value 0 are left blank.

the value 0.) The reason for this averaging property is that, in order to wind up ahead j after n flips, Peter must either have been ahead $j-1$ after $n-1$ flips, and then won the last flip; or have been ahead $j+1$, and then lost.

Now, here's where the trigonometry comes in.

$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b),$$

and

$$\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b),$$

so

$$\cos(a)\cos(b) = \frac{1}{2}(\cos(a+b) + \cos(a-b)).$$

Thus

$$\cos(jx)\cos(x) = \frac{1}{2}\cos((j+1)x) + \frac{1}{2}\cos((j-1)x).$$

According to this formula, multiplying by $\cos(x)$ takes $\cos(jx)$ and replaces it with a half-and-half combination of $\cos((j+1)x)$ and $\cos((j-1)x)$. This is very much like what happens to Peter's net winnings in the heads-and-tails game.

$1\cos(0x)$	$\frac{1}{2}\cos(1x)$ $+$ $\frac{1}{2}\cos(-1x)$	$\frac{1}{4}\cos(2x)$ $+$ $\frac{2}{4}\cos(0x)$ $+$ $\frac{1}{4}\cos(-2x)$	$\frac{1}{8}\cos(3x)$ $+$ $\frac{3}{8}\cos(1x)$ $+$ $\frac{3}{8}\cos(-1x)$ $+$ $\frac{1}{8}\cos(-3x)$	$\frac{1}{16}\cos(4x)$ $+$ $\frac{4}{16}\cos(2x)$ $+$ $\frac{6}{16}\cos(0x)$ $+$ $\frac{4}{16}\cos(-2x)$ $+$ $\frac{1}{16}\cos(-4x)$
$\cos(x)^0$	$\cos(x)^1$	$\cos(x)^2$	$\cos(x)^3$	$\cos(x)^4$

Table 2: Succesive powers of $\cos(x)$.

To exploit this connection, write $1 = \cos(0x)$; multiply the equation by $\cos(x)$; and expand the right-hand side according to the formula above:

$$\cos(x) = \cos(0x)\cos(x) = \frac{1}{2}\cos(1x) + \frac{1}{2}\cos(-1x).$$

Since $\cos(1x) = \cos(-1x) = \cos(x)$, this is hardly a surprise! Let's keep going.

$$\begin{aligned}
\cos(x)^2 &= (\frac{1}{2}\cos(1x) + \frac{1}{2}\cos(-1x))\cos(x) \\
&= \frac{1}{2}\cos(1x)\cos(x) + \frac{1}{2}\cos(-1x)\cos(x) \\
&= \frac{1}{2}(\frac{1}{2}\cos(2x) + \frac{1}{2}\cos(0x)) + \frac{1}{2}(\cos(0x) + \frac{1}{2}\cos(-2x)) \\
&= \frac{1}{4}\cos(2x) + \frac{2}{4}\cos(0x) + \frac{1}{4}\cos(-2x).
\end{aligned}$$

We could simplify this to the familiar formula $\cos(x)^2 = \frac{1}{2}\cos(2x) + \frac{1}{2}$, but instead we plow ahead to compute higher powers of $\cos(x)$. The results are shown shown in a suggestive way in Table 2. Comparing Table 2 with Table 1 shows that

$$\cos(x)^n = \sum_{j=-n}^{n} w_{n,j}\cos(jx).$$

So the probabilites $w_{n,j}$ are lurking in the powers of $\cos(x)$. Look at Figure 2, which shows plots of some of these powers. There's our bell

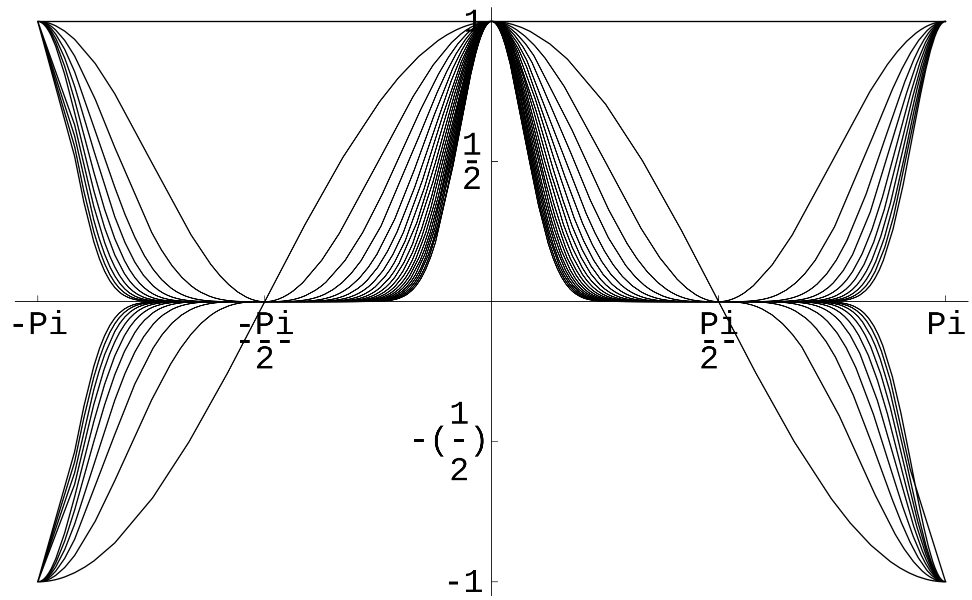

Figure 2: Powers of $\cos(x)$ on the interval $[-\pi, \pi]$.

curve! But no, that can't be our bell curve: Ours involves values of $w_{n,j}$ for varying j, while this one involve values of $\cos(x)^n$ for varying x. To get to our bell curve, we will need to extract the coefficients $w_{n,j}$ from the linear combination $\cos(x)^n = \sum_{j=-n}^{n} w_{n,j} \cos(jx)$.

To see how this can be done, look at

$$\cos(x)^4 = \frac{1}{16}\cos(4x) + \frac{4}{16}\cos(2x) + \frac{6}{16} + \frac{4}{16}\cos(-2x) + \frac{1}{16}\cos(-4x).$$

Here's we've simplified the middle term $\frac{6}{16}\cos(0x)$ to emphasize that it is constant. If we integrate this expression from $-\pi$ to π, this constant term is the only term that will survive, while the oscillatory terms will disappear, because

$$\int_{-\pi}^{\pi} \cos(kx)dx = \begin{cases} 2\pi & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Thus

$$\int_{-\pi}^{\pi} \cos(x)^4 = 0 + 0 + \frac{6}{16}2\pi + 0 + 0;$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(x)^4 = \frac{6}{16}.$$

Using this same technique, we can extract all the 'break-even' probabilities $w_{n,0}$:

$$w_{n,0} = \frac{1}{2^n}\binom{n}{\frac{n}{2}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(x)^n dx.$$

Note that this formula works just fine even when n is odd and $w_{n,0} = 0$.

This technique of extracting the constant part of a periodic function by averaging over a period can be modified to extract coefficients other than that of the constant term. This can be see most clearly when working with complex exponentials, but we can get by without them by observing that

$$\begin{aligned}\int_{-\pi}^{\pi} \cos(kx)\cos(lx)dx &= \int_{-\pi}^{\pi} \frac{1}{2}(\cos((k+l)x) + \cos((k-l)x))dx \\ &= \begin{cases} \pi & \text{if } |k| = |l| > 0 \\ 2\pi & \text{if } k = l = 0 \\ 0 & \text{if } |k| \neq |l| \end{cases}.\end{aligned}$$

After some fuss, this yields

$$w_{n,j} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(jx)\cos(x)^n dx.$$

To avoid the fuss, we can see this result directly by recognizing that the right-hand side gives the break-even probability in a game where Peter starts out down by j pennies, and this probability is equal to $w_{n,j}$.

Let us now restrict ourselves to the case where $n + j$ is even. In this case the integrand $\cos(jx)\cos(x)^n$ repeats with period π, rather than the nominal period 2π, so we can cut the range of integration in half and write

$$w_{n,j} = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \cos(jx)\cos(x)^n dx.$$

Figure 2 shows the powers of $\cos(x)$ on the interval $[-\pi, \pi]$. As we observed before, these graphs look like bell curves on the interval $[-\pi/2, \pi/2]$. And here's why:

$$\cos(x)^n = (1 - \frac{x^2}{2} + \ldots)^n \approx (1 - \frac{x^2}{2})^n = (1 - \frac{1}{n}\frac{nx^2}{2})^n \approx e^{-nx^2/2}.$$

The last approximation comes from the formula for continuously compounded interest:

$$\lim_{n\to\infty}(1 + \frac{a}{n})^n = e^a.$$

The first approximation comes from the Taylor series for $\cos(x)$, and is only good for small values of x. But for larger x in our interval the values of $\cos(x)^n$ and $e^{-nx^2/2}$ are both tiny when n is large. The upshot is that even for quite modest n the approximation

$$\cos(x)^n \approx e^{-nx^2/2}$$

holds very accurately for all x in the interval $[-\pi/2, \pi/2]$. We'll spare ourselves working out the exact error bounds.

The graph $y = e^{-nx^2/2} = e^{-(\sqrt{\frac{n}{2}}x)^2}$ is just the graph $y = e^{-x^2}$ squashed horizontally by a factor of $\sqrt{\frac{n}{2}}$, so our graphs of $\cos(x)^2$ really are approaching standard bell-shaped curves.

Now to get to the bell curve for the values of $w_{n,j}$, we just have to integrate.

$$\begin{aligned} w_{n,j} &= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \cos(jx)\cos(x)^n dx \\ &\approx \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \cos(jx) e^{-nx^2/2} dx \\ &\approx \frac{1}{\pi}\int_{-\infty}^{\infty} \cos(jx) e^{-nx^2/2} dx. \end{aligned}$$

Can we do this integral? We can, and we will.

But for now, we'll just ask the computer:

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \cos(jx) e^{-nx^2/2} dx = \sqrt{\frac{2}{\pi n}} e^{\frac{-j^2}{2n}}.$$

This result follows by changing variables in the formula

$$\int_{-\infty}^{\infty} \cos(ax) e^{-x^2/2} dx = \sqrt{2\pi} e^{-a^2/2},$$

which is the formula we'll want to prove, after we've taken a minute to savor

DE MOIVRE'S FORMULA. If $n + j$ is even,

$$w_{n,j} = \frac{1}{2^n}\binom{n}{\frac{n+j}{2}} \approx 2 \cdot \frac{1}{\sqrt{2\pi n}} e^{\frac{-j^2}{2n}}.$$

In particular, if n is even

$$w_{n,0} = \frac{1}{2^n}\binom{n}{n/2} = \frac{1}{2^n}\frac{n!}{(n/2)!^2} \approx 2 \cdot \frac{1}{\sqrt{2\pi n}}$$

QUESTION: Why did we pull that factor of 2 out in front?

OK, now back to that integral. Let

$$f(a) = \int_{-\infty}^{\infty} \cos(ax) e^{-x^2/2} dx.$$

Differentiate under the integral sign:

$$\frac{d}{da} f(a) = \frac{d}{da} \int_{-\infty}^{\infty} \cos(ax) e^{-x^2/2} dx = -\int_{-\infty}^{\infty} \sin(ax) x e^{-x^2/2} dx.$$

Integrate by parts:

$$\frac{d}{da} f(a) = \int_{-\infty}^{\infty} \frac{1}{a} \cos(ax) e^{-x^2/2} dx = \frac{1}{a} f(a).$$

The general solution to this differential equation is $ce^{-a^2/2}$, as we can verify by differentiation. So $f(a) = f(0)e^{-a^2/2}$, with $f(0) = \int_{-\infty}^{\infty} e^{-x^2/2} dx$.

So all we need to show is that

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

As you may know, there is a famous trick to show this. We'll come back to that in a moment. But first we want to show how to establish the value of $f(0)$ in a more straight-forward way.

Without filling in the value of $f(0)$, we can still write a version of de Moivre's formula:

$$w_{n,j} \approx \frac{f(0)}{\pi\sqrt{n}} e^{\frac{-j^2}{2n}}$$

when $n + j$ is even. In fact, this is basically what de Moivre first did. Then we can write

$$1 = \sum_j w_{n,j} \approx \frac{1}{2} \sum_j \frac{f(0)}{\pi\sqrt{n}} e^{\frac{-j^2}{2n}}.$$

Note well the $\frac{1}{2}$ in front of the sum: It's there because our approximation only works when $n+j$ is even. Every other value of j has $w_{n,j} = 0$. If we were working with the more explicit version of de Moivre's formula that we wrote down previously, the $\frac{1}{2}$ would cancel the 2 we pulled out in front—which (in case you're still wondering) is why we pulled that 2 out front in the first place.

Continue by approximating the sum with an integral, and then change variables:

$$1 \approx \frac{1}{2}\int_{-\infty}^{\infty} \frac{f(0)}{\pi\sqrt{n}} e^{\frac{-x^2}{2n}} dx = \frac{f(0)}{2\pi}\int_{-\infty}^{\infty} e^{-x^2/2} dx = \frac{f(0)^2}{2\pi},$$

so

$$1 \approx \frac{f(0)^2}{2\pi}.$$

The approximations we've made in establishing this approximate equality become exact in the limit of large n, and neither side involves n, so the equality must be exact:

$$1 = \frac{f(0)^2}{2\pi},$$

so

$$f(0) = \int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

That was the straight-forward proof. Here's the promised trick.

$$\begin{aligned} f(0)^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2/2} e^{-y^2/2} dx dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy. \end{aligned}$$

Change to polar coordinates:

$$f(0)^2 = \int_{0}^{\infty} 2\pi r e^{-r^2/2} dr = 2\pi.$$

Pretty slick, hey?

More honest approach

$$p = \frac{1}{2}; \quad q = 1 - p = \frac{1}{2}.$$

$$b_{n,k} = \binom{n}{k} p^k q^{n-k} = \frac{1}{2^n}\binom{n}{k}.$$

$$e^{ix} = \cos(x) + i\sin(x).$$

$$\cos(x) = \frac{1}{2}(e^{ix} + e^{-ix}).$$

$$\cos(x)^n = \frac{1}{2^n}\sum_k \binom{n}{k}(e^{ix})^k(e^{-ix})^{n-k} = \sum_k b_{n,k}e^{i(2k-n)x}.$$

$$\int_{-\pi}^{\pi} e^{ilx}e^{-imx}dx = \begin{cases} 2\pi & l = m \\ 0 & l \neq m \end{cases}.$$

$$\begin{aligned} b_{n,k} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(x)^n e^{-i(2k-n)x}dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(x)^n \cos((2k-n)x)dx \\ &= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \cos(x)^n \cos((2k-n)x)dx. \end{aligned}$$

$$\cos(x)^n = (1-\frac{x^2}{2}+\ldots)^n \approx (1-\frac{x^2}{2})^n = (1-\frac{1}{n}\frac{nx^2}{2})^n \approx e^{-nx^2/2}, \quad -\pi/2 \leq x \leq \pi/2.$$

$$\begin{aligned} b_{n,k} &\approx \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} e^{-nx^2/2}\cos((2k-n)x)dx \\ &\approx \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-nx^2/2}\cos((2k-n)x)dx \\ &= \frac{1}{\pi\sqrt{n}}\int_{-\infty}^{\infty} e^{-x^2/2}\cos(\frac{2k-n}{\sqrt{n}}x)dx. \end{aligned}$$

$$\int_{-\infty}^{\infty} e^{-x^2/2}\cos(ax)dx = \sqrt{2\pi}e^{-a^2/2}.$$

$$b_{n,k} \approx \frac{1}{\pi\sqrt{n}}\sqrt{2\pi}e^{-\frac{(2k-n)^2}{2n}} = \frac{1}{\sqrt{2\pi n/4}}e^{-\frac{(k-n/2)^2}{2n/4}}.$$

The last equation is written so as to conform to the more general formula

$$b_{n,p,k} \approx \frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}}.$$

To prove this more general formula, we can modify the foregoing argument using $pe^{ix}+qe^{-ix}$ in place of $\cos(x) = \frac{1}{2}(e^{ix}+e^{-ix})$ —and kiss the real numbers goodbye. We'll do something akin to this in the next section.

Even-more-honest approach

The natural logarithm $\log x$ was introduced by Napier in 1614 as a way of turning multiplication into addition. Reducing multiplication to addition simplified computations done by hand. The inverse of the logarithm is the exponential e^x: It turns addition into multiplication. Reducing addition to multiplication simplifies analyzing distributions of sums of random variables. For this purpose it is most convenient to use the complex exponential $e^{ix} = \cos x + i \sin x$, introduced by Euler in 1748. This approach via complex exponentials is the 'method of Fourier transforms', but in keeping with the usual probabilist's lingo we'll call it the 'method of characteristic functions'.

For any random variable Y, define its *characteristic function*

$$f_Y(x) = \mathrm{E}(e^{ixY}).$$

Note that

$$f_Y(0) = \mathrm{E}(e^{i0Y}) = \mathrm{E}(1) = 1.$$

As our prime example, if HorT is a heads-or-tails random variable taking values ± 1 with probability $\frac{1}{2}$,

$$f_{\mathrm{HorT}}(x) = \frac{1}{2}e^{ix} + \frac{1}{2}e^{-ix} = \cos(x).$$

Note that $\cos(x)$ has period 2π, i.e. $\cos(x + 2\pi) = \cos(x)$. This periodicity $f_Y(x + 2\pi) = f_Y(x)$ holds for any Y that takes on only integer values, since in that case

$$e^{i(x+2\pi)Y} = e^{ixY} e^{2\pi iY} = e^{ixY}.$$

Now let's watch the exponential turn addition of independent random variables into multiplication of their characteristic functions. If Y and Z are independent,

$$f_{Y+Z}(x) = \mathrm{E}(e^{i(Y+Z)}) = \mathrm{E}(e^{iY} e^{iX}) = \mathrm{E}(e^i Y)\mathrm{E}(e^{iZ}) = f_Y(x) f_Z(x),$$

where we've used the important (and Fundamentally Mysterious) fact that the expected value of a product is the product of the expected values *when the variables involved are independent*.

If $S_n = Y_1 + \ldots + Y_n$ is the sum of n independent random variables having the same distribution as Y, then

$$f_{S_n}(x) = f_Y(x)^n.$$

In the case that S_n represents Peter's fortune after n rounds of heads-or-tails,

$$f_{S_n}(x) = f_{\mathrm{HorT}}(x)^n = \cos(x)^n.$$

The mean and variance of a random variable are easy to extract from its characteristic function. Suppose Y has mean μ and variance v. Then

$$\frac{d}{dx} f_Y(x) = \mathrm{E}(\frac{d}{dx} e^{ixY}) = \mathrm{E}(iY \frac{d}{dx} e^{ixY})$$

and

$$\frac{d^2}{dx^2} f_Y(x) = \mathrm{E}(\frac{d^2}{dx^2} e^{ixY}) = \mathrm{E}(-Y^2 \frac{d^2}{dx^2} e^{ixY}).$$

Setting $x = 0$ yields

$$\frac{d}{dx} f_Y(x)|_{x=0} = \mathrm{E}(iY) = i\mu$$

and

$$\frac{d^2}{dx^2} f_Y(x)|_{x=0} = \mathrm{E}(-Y^2) = -(v + \mu^2).$$

Thus μ and v determine the second-order Taylor approximation to $f_Y(x)$

$$f_Y(x) \approx f_Y(0) + \frac{d}{dx} f_Y(x)|_{x=0}\, x + \frac{d^2}{dx^2} f_Y(x)|_{x=0}\, \frac{x^2}{2} = 1 + i\mu x - (v + \mu^2)\frac{x^2}{2}.$$

Let's check these computations in the case of the heads-or-tails random variable HorT, which has mean 0 and variance 1:

$$\frac{d}{dx} \cos(x)|_{x=0} = 0;$$

$$\frac{d^2}{dx^2} \cos(x)|_{x=0} = -1;$$

$$\cos(x) \approx 1 + i0x - \frac{x^2}{2}.$$

Much of the foregoing discussion would go through if we had used the real-valued exponential e^x in place of the complex exponential e^{ix}, and defined the characteristic function to be $\mathrm{E}(e^{xY})$. Yes, there would have been some convergence problems if the values Y can assume aren't bounded. But the real virtue of using complex exponentials lies in the ease with which we can go backward from the characteristic function $f_Y(x)$ to the distribution of Y.

To show how this works, we'll restrict to an integer-valued Y taking on value j with probability p_j, so that

$$f_Y(x) = \sum_{n=-\infty}^{\infty} p_j e^{ijx}.$$

As we've already observed, in this case f_Y has period 2π. If we average f_Y over a period, we get p_0:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f_Y(x)dx = p_0.$$

That's because, when averaged over a period, the terms $p_j e^{ijx}$ with $j \neq 0$ all vanish, leaving only the constant term. And if we multiply f_Y by e^{-ijx} before averaging, we can move the term $p_j e^{ijx}$ into the limelight, and get

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f_Y(x)e^{-ijx}dx = p_j.$$

So for integer-valued random variables, it's easy to extract all the probabilities p_j from the characteristic function.

NOTE: Something fundamentally similar works for general random variables, to allow us to read off the distribution of the random variable from the characteristic function: To bandy words around, we work with the 'Fourier transform' instead of 'Fourier series'. This approach is needed to prove the general Central Limit Theorem, so it's good to know about. But we've no need for it here.

Now here comes de Moivre's formula. Let S_n represent Peter's fortune after n rounds of heads-or-tails. And let's allow the possibility that the coin being used may be biased, with probability p of coming up heads and probability $q = 1-p$ of coming up tails. Then Peter's winning Y for a single flip has characteristic function

$$f_Y(x) = pe^{ix} + qe^{-ix}.$$

The main thing we'll need to use about the characteristic function is the approximation

$$f_Y(x) \approx 1 + i\mu x - (v + \mu^2)\frac{x^2}{2},$$

with

$$\mu = p \cdot 1 + q \cdot (-1) = 1 - 2p$$

and

$$v = 4pq.$$

We propose to estimate p_j using our general formula

$$p_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_Y(x)^n e^{-ijx} dx.$$

If $n+j$ is even then the integrand has period π, so we can restrict our average to the interval $-\pi/2 \leq x \leq \pi/2$. On this interval

$$f_Y(x)^n \approx (1 + i\mu x - (v + \mu^2)\frac{x^2}{2})^n \approx e^{in\mu x - nv\frac{x^2}{2}} = e^{-nv\frac{x^2}{2}} e^{in\mu x}$$

so

$$\begin{aligned} p_j &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} e^{-nv\frac{x^2}{2}} e^{i(n\mu - j)x} dx \\ &\approx \frac{1}{\pi} \int_{-\infty}^{\infty} e^{-nv\frac{x^2}{2}} e^{i(n\mu - j)x} dx. \end{aligned}$$

This is an integral which (by now) we know how to do:

$$p_j \approx \frac{2}{\sqrt{2\pi v}} e^{-\frac{(n\mu - j)^2}{2nv}}.$$

For heads-or-tails with a loaded coin, $\mu = 2p - 1$ and $v = 4pq$, and p_j is the binomial probability

$$b_{n,p,k} = \binom{n}{k} p^k q^{n-k},$$

where $k = \frac{n+j}{2}$ is the number of times Peter must win in order to be ahead by j after n rounds. In terms of k, $j = 2k - n$. Plugging this into our approximation, along with $\mu = 2p - 1$ and $v = 4pq$ gives

$$b_{n,p,k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(np-k)^2}{2npq}}.$$

PROBLEM: Observe how the 2 in the numerator of our previous formula has now disappeared. Elucidate!

Stirling's formula

Stirling's formula, published by James Stirling in 1730, gives an approximation to the factorial function:

$$n! \approx (n/e)^n \sqrt{2\pi n}.$$

It holds in the sense that the ratio of the two sides approaches 1.

According to a hoary anecdote, a student taking his Ph.D. qualifying exam from Jacob Tamarkin at Brown was asked to state Stirling's formula. 'I don't know it,' the student admitted, 'But I know where to find it when I need it.' 'Young man,' said Tamarkin, 'You need it now!'

Peter Major of the Hungarian Academy of Sciences has pointed out where we can find Stirling's formula when we need it 'now': In the central term of a Poisson distribution with parameter n. The Poisson distribution with parameter $\lambda \geq 0$ is

$$p_{\lambda,k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

The mean and variance are both λ. Set $\lambda = n$, and look at the central term

$$p_{n,n} = e^{-n} \frac{n^n}{n!}.$$

When n gets large the central limit theorem kicks in to give

$$p_{n,n} \approx \frac{1}{\sqrt{2\pi n}}.$$

So

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

which yields Stirling's formula above.

For this mnemonic to work, we need to (1) remember (or be able to derive) the Poisson distribution; (2) know (or be able to figure out) the mean and variance; and (3) be able to apply the central limit theorem. A lot could be said about all this. Let's try to say only a little.

(1) One way to remember the Poisson distribution involves writing

$$1 = e^{-\lambda}(1 + \lambda + \lambda^2/2! + \lambda^3/3! + \ldots).$$

(2) A Poisson random variable with parameter n is the sum of n independent Poisson random variables with parameter 1, so mean and variance grow linearly with n, and it's natural to suppose that the constant of proportionality is 1.

(3) The main trick to remembering the central limit theorem is to get the normalizing constant right in the standard Gaussian density $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. For this, we can use the polar coordinates trick to show $\int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}$. Of course, we also need to remember whether we really want that 2 in the denominator of the exponent $-x^2/2$. To check, we can integrate $x^2e^{-x^2}$ by parts. As soon as we start the calculation, we'll remember why we want that 2 in the denominator.

If all else fails, we can just buckle down and adapt the proof of the central limit theorem to the case at hand. This turns out to be easy, as we'll see in a jiffy. This approach has the advantage of anticipating what would doubtless have been Tamarkin's next demand: Prove Stirling's formula!

Derivation of Stirling's formula

$$p_{\lambda,k} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

$$\begin{aligned}
\sum p_{\lambda,k}e^{ikx} &= e^{-\lambda}\sum \frac{\lambda^k}{k!}e^{ikx} \\
&= e^{-\lambda}\sum \frac{(\lambda e^{ix})^k}{k!} \\
&= e^{-\lambda}e^{\lambda e^{ix}} \\
&= e^{\lambda(e^{ix}-1)}.
\end{aligned}$$

$$p_{\lambda,k} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\lambda(e^{ix}-1)}e^{-ikx}dx.$$

$$\begin{aligned}
p_{n,n} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{n(e^{ix}-1-ix)}dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{n(1+ix-\frac{x^2}{2}+\ldots-1-ix)}dx \\
&\approx \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-n\frac{x^2}{2}}dx
\end{aligned}$$

$$\begin{aligned}
&\approx \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-n\frac{x^2}{2}}dx \\
&= \frac{1}{2\pi\sqrt{n}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx \\
&= \frac{\sqrt{2\pi}}{2\pi\sqrt{n}} \\
&= \frac{1}{\sqrt{2\pi n}}.
\end{aligned}$$

$$e^{-n}\frac{n^n}{n!} \approx \frac{1}{\sqrt{2\pi n}}.$$

$$n! \approx (n/e)^n\sqrt{2\pi n}.$$

Stirling's series

We can get an approximation to $n!$ that works better for large n by keeping more terms in the expansion for e^{-ix}.

$$e^{ix} = 1 + ix - \frac{x^2}{2} - i\frac{x^3}{6} + \frac{x^4}{24} + \dots$$

$$\begin{aligned}
p_{n,n} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{n(e^{ix}-1-ix)}dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-n\frac{x^2}{2}}e^{-in\frac{x^3}{6}}e^{n\frac{x^4}{24}}\cdots dx \\
&= \frac{1}{2\pi\sqrt{n}}\int_{-\pi}^{\pi} e^{-n\frac{x^2}{2}}(1 - i\frac{x^2}{6\sqrt{n}} - \frac{x^6}{72n} + \dots)(1 + \frac{x^4}{24} + \dots)\cdots dx \\
&\approx \frac{1}{2\pi\sqrt{n}}\int_{-\pi}^{\pi} e^{-n\frac{x^2}{2}}(1 - i\frac{x^3}{6\sqrt{n}} + \frac{x^4}{4!n} - \frac{x^6}{72n})dx \\
&\approx \frac{1}{2\pi\sqrt{n}}\int_{-\infty}^{\infty} e^{-n\frac{x^2}{2}}(1 - i\frac{x^3}{6\sqrt{n}} + \frac{x^4}{4!n} - \frac{x^6}{72n})dx.
\end{aligned}$$

Note that in this expansion, we've kept only terms whose denominators grow at most linearly in n. The imaginary term drops $-i\frac{x^3}{6\sqrt{n}}$ drops out (it's odd), and the others succumb to integration by parts:

$$p_{n,n} \approx \frac{1}{2\pi\sqrt{n}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}(1 + \frac{x^4}{24n} - \frac{x^6}{72n})dx$$

$$
\begin{aligned}
&= \frac{1}{2\pi\sqrt{n}}\sqrt{2\pi}(1+\frac{3}{24n}-\frac{3\cdot 5}{72n}) \\
&= \frac{1}{\sqrt{2\pi n}}(1-\frac{1}{12n}).
\end{aligned}
$$

$$
n! \approx (n/e)^n\sqrt{2\pi n}\frac{1}{1-\frac{1}{12n}} \approx (n/e)^n\sqrt{2\pi n}(1+\frac{1}{12n}).
$$

Keeping terms with denominators of higher order in n will yield further terms of 'Stirling's series'

$$
n! \approx (n/e)^n\sqrt{2\pi n}(1+\frac{1}{12n}+\frac{1}{288n^2}-\frac{139}{51840n^3}-\frac{571}{2488320n^4}+\ldots).
$$