The central limit theorem - how Laplace actually proved it

Peter Haggstrom
www.gotohaggstrom.com
mathsatbondibeach@gmail.com

October 17, 2015

1 Introduction

The central limit theorem has its genesis in work done by de Moivre for the simpler binomial case where the probability of success equalled $\frac{1}{2}$. De Moivre started of course with a binomial expression and used Stirling's approximation to $n!$ to arrive at the required probabilities. In his book "The Doctrine of Chances" de Moivre said that "If a binomial $1 + 1$ is raised to a very high power n, the ratio which the middle term has to the sum of all the terms, that is, to 2^n, may be expressed by the fraction $\frac{2A(n-1)^n}{n^n\sqrt{n-1}}$ where A represents the number of which the hyperbolic logarithm is $\frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680}$.." [1] De Moivre then makes the following estimates:

$$\frac{2A(n-1)^n}{n^n\sqrt{n-1}} = \frac{2A}{\sqrt{n-1}}(1-\frac{1}{n})^n \sim \frac{2A}{\sqrt{n-1}}e^{-1} \sim \frac{2A}{\sqrt{n}}e^{-1} \tag{1}$$

In modern terminology:

$$\ln A = \frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680}\dots \text{ ie } A = e^{\frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680}\cdots} \tag{2}$$

De Moivre then says if $\ln B = -1 + \frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680}\dots$ (1) can be written as:

$$\frac{2}{B^*\sqrt{n}} \tag{3}$$

where $B^* = B^{-1}$

He then says that his friend James Stirling had found that B did denote the square root of the circumference of circle of radius unity" [1, page 244]. This of course is 2π which was denoted by c in De Moivre's book. Thus (3) becomes:

$$\frac{2}{\sqrt{2\pi n}} \tag{4}$$

Karl Pearson has made a case for giving De Moivre credit for discovering the normal curve rather than Laplace or Gauss and that what we now call Stirling's formula should be called the De Moivre-Stirling formula since Stirlimg contributed the 2π. [2] De Moivre's development was essentially based on equally likely events ie $p = q = \frac{1}{2}$ whereas Laplace generalised the binomial approximation for arbitrary p and q.

2 The details of Laplace's original proof

Thanks to Google we have Laplace's original treatise *Théorie Analytique des Probabilitiés* in digitised form [3]. The relevant part is Chapter III of Book II. It is of course in French. A translation by Richard J Pulskamp of Chapter III can be found at [4] along with a translation of some relevant preliminary material. I will follow the translation by Richard J Pulskamp and stick to Laplace's symbols. The chapter is titled "On the laws of probability which result from the indefinite multiplication of events" and Laplace sets up the basic model as follows [4,page 10]. He invites us to conside the way in which two simple events, one of which must occur at each trial (a Bernoulli event in modern language), devleop when we multiply the number of trials ie make n large. He says that "it is clear that the event of which the facility is greatest must probably arrive more often in a given number of trials, and we are carried naturally to think that by repeating the trials a very great number of times, each of these events will arrive proportionally to its facility, that we will be able thus to discover by experience. We will demonstrate analytically this important theorem". [4, page 10]

Laplace lets p and $1 - p$ be the respective probabilities of two events a and b so that in $x + x'$ trials the probability that the event a will arrive x times and the event b will arrive x' times will be:

$$\frac{1.2.3\ldots(x + x')}{1.2.3\ldots x.1.2.3\ldots x'}\, p^x(1 - p)^{x'} \tag{5}$$

(5) is simply the $(x' + 1)^{st}$ term of the binomial $[p + (1 - p)]^{x+x'}$. Note here that the $(x' + 1)^{st}$ term is obtained by putting $j = x'$ in the sum $\sum_{j=0}^{x+x'} \binom{x+x'}{j} p^j (1 - p)^{x+x'-j}$ making it the $(x' + 1)^{st}$ term . Laplace then considers the greatest of the terms in (5) to be designated by k. He says the anterior term to k will be:

$$\frac{kp}{1-p}\,\frac{x'}{x+1} \tag{6}$$

while the following term will be:

$$\frac{k(1-p)}{p}\,\frac{x}{x'+1} \tag{7}$$

These are derived as follows. The term immediately before k is:

$$\begin{aligned}\binom{x+x'}{x'-1}p^{x+1}(1-p)^{x'-1} &= \frac{(x+x')!}{(x+1)!\,(x'-1)!}\,p^{x}(1-p)^{x'}\,\frac{p}{1-p}\\ &= \underbrace{\frac{(x+x')!}{x!\,x'!}\,p^{x}(1-p)^{x'}}_{k}\,\frac{x'}{x+1}\,\frac{p}{1-p}\\ &= \frac{kp}{1-p}\,\frac{x'}{x+1}\end{aligned} \tag{8}$$

The term immediately after k is:

$$\begin{aligned}\binom{x+x'}{x'+1}p^{x-1}(1-p)^{x'+1} &= \frac{(x+x')!}{(x-1)!\,(x'+1)!}\,p^{x}(1-p)^{x'}\,\frac{1-p}{p}\\ &= \underbrace{\frac{(x+x')!}{x!\,x'!}\,p^{x}(1-p)^{x'}}_{k}\,\frac{x}{x'+1}\,\frac{1-p}{p}\\ &= \frac{k(1-p)}{p}\,\frac{x}{x'+1}\end{aligned} \tag{9}$$

The next step in Laplace's argument is the proposition that in order for k to be the greatest term it is necessary that the following inequalities apply:

$$\frac{x}{x'+1} < \frac{p}{1-p} < \frac{x+1}{x'} \tag{10}$$

(10) follows from the two inequalities:

$$k > \frac{kp}{1-p}\,\frac{x'}{x+1} \quad\Longrightarrow\quad \frac{x+1}{x'} > \frac{p}{1-p} \tag{11}$$

and

$$k > \frac{k(1-p)}{p}\frac{x}{x'+1} \implies \frac{p}{1-p} > \frac{x}{x'+1} \tag{12}$$

Letting $x + x' = n$ we then have the following:

$$\frac{x}{x'+1} < \frac{p}{1-p} \implies x(1-p) < p(n-x+1) \implies x < (n+1)p \tag{13}$$

and

$$\frac{p}{1-p} < \frac{x+1}{x'} \implies p(n-x) < x - px + 1 - p \implies (n+1)p - 1 < x \tag{14}$$

Thus x will be the greatest number comprehended within $(n+1)p$ ie

$$(n+1)p - 1 < x < (n+1)p \tag{15}$$

Laplace then expressed x as:

$$x = (n+1)p - s \tag{16}$$

where $0 < s < 1$. The following relationships then follow immediately from (16):

$$p = \frac{x+s}{n+1} \tag{17}$$

$$1 - p = 1 - \frac{x+s}{n+1} = \frac{n+1-x-s}{n+1} = \frac{x'+1-s}{n+1} \tag{18}$$

$$\frac{p}{1-p} = \frac{x+s}{x'+1-s} \tag{19}$$

Up to this point Laplace has not relied upon the size of x and x' but now he supposes that both are very great numbers so that we will have the following approximation:

$$\frac{p}{1-p} = \frac{x}{x'} \tag{20}$$

In other words "the exponents of p and $1-p$ in the greatest term of the binomial are quite nearly in the ratio of these quantities; so that, of all the combinations which are able to take place in a very great number n of trials, the most probable is that in which each event is repeated proportionally to its probability." [4, pages 10-11]

The next phase of the proof involves moving away from the central (or greatest) term by l terms. The l^{th} term after the greatest is given by:

$$\frac{n!}{1.2..3 \ldots (x-l).1.2.3 \ldots (x^{'}+l)} p^{x-l} p^{x^{'}+l} \tag{21}$$

Laplace now uses Stirling's approximation to $n!$ in the following form:

$$n! = n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi} (1 + \frac{1}{12n} + \ldots) \tag{22}$$

This leads to:

$$\begin{aligned} \frac{1}{1.2.3 \ldots (x-l)} &= \frac{1}{(x-l)^{x-l+\frac{1}{2}} e^{-(x-l)} \sqrt{2\pi} (1 + \frac{1}{12n} + \ldots)} \\ &= \frac{(x-l)^{l-x-\frac{1}{2}} e^{x-l}}{\sqrt{2\pi}} (1 - \frac{1}{12(x-l)} + \ldots) \end{aligned} \tag{23}$$

Note here that by long division $\frac{1}{1+u} = 1 - u + u^2 - \ldots$.

Completely analogously:

$$\frac{1}{1.2.3 \ldots (x^{'}+l)} == \frac{(x^{'}+l)^{-x^{'}-l-\frac{1}{2}} e^{x^{'}+l}}{\sqrt{2\pi}} (1 - \frac{1}{12(x^{'}+l)} + \ldots) \tag{24}$$

Laplace now takes the natural logarithm (what he calls the "hyperbolic logarithm") of $(x-l)^{l-x-\frac{1}{2}}$ and $(x^{'}+l)^{-x^{'}-l-\frac{1}{2}}$. Thus:

$$\ln\Big((x-l)^{l-x-\frac{1}{2}}\Big) = \Big(l - x - \frac{1}{2}\Big)\Big[\ln x + \ln(1 - \frac{l}{x})\Big] \tag{25}$$

Next Laplace simply expands $\ln(1 - \frac{l}{x})$ as a Taylor's series as follows:

$$\ln\Big(1 - \frac{l}{x}\Big) = -\frac{l}{x} - \frac{l^2}{2x^2} - \frac{l^3}{3x^3} - \frac{l^4}{4x^4} - \ldots \tag{26}$$

This is verified as follows. If we let $f(u) = \ln(1+u)$ and expand as a Taylor series about $u = 0$ we have:

$$
\begin{aligned}
\ln(1+u) = f(u) =& f(0) + u f^{(1)}(0) + \frac{u^2}{2!} f^{(2)}(0) + \frac{u^3}{3!} f^{(3)}(0) + \frac{u^4}{4!} f^{(4)}(0) + \dots \\
=& \ln 1 + u - \frac{u^2}{2} + \frac{2u^3}{3!} - \frac{6u^4}{4!} + \dots \\
=& u - \frac{u^2}{2} + \frac{u^3}{3} - \frac{u^4}{4} - \dots
\end{aligned} \tag{27}
$$

Letting $u = \frac{-l}{x}$ we get the desired result.

Going back to (25):

$$
\begin{aligned}
\Big(l - x - \frac{1}{2}\Big)\Big[\ln x + \ln(1 - \frac{l}{x})\Big] =& \Big(l - x - \frac{1}{2}\Big) \ln x + \Big(l - x - \frac{1}{2}\Big)\Big(-\frac{l}{x} - \frac{l^2}{2x^2} - \frac{l^3}{3x^3} - \frac{l^4}{4x^4} - \dots\Big) \\
=& \Big(l - x - \frac{1}{2}\Big) \ln x + \Big(x - l + \frac{1}{2}\Big)\Big(\frac{l}{x} + \frac{l^2}{2x^2} + \frac{l^3}{3x^3} + \frac{l^4}{4x^4} + \dots\Big)
\end{aligned} \tag{28}
$$

Now in estimating (28) Laplace makes the obvious decision because n is large to neglect quantities of order $\frac{1}{n}$ and he further supposes that l^2 is not greater than order n. This latter decision is broadly justified on the basis that the central term is of order $\frac{1}{\sqrt{n}}$ - indeed Wallis' formula gives us $\frac{1}{\sqrt{\pi n}}$. To get a fixed area around the central term one would need to go order $\sqrt{n}$ either side. Crudely to get a total probability of 1 you would need to go $\frac{\sqrt{n}}{2}$ either side of the central term which has height $\frac{1}{\sqrt{n}}$. On this basis he is able to neglect terms of order $\frac{l^4}{x^3}$ because x and x' are of order n. If we expand the last product in (28) we get:

$$
\begin{aligned}
\Big(x - l + \frac{1}{2}\Big)\Big(\frac{l}{x} + \frac{l^2}{2x^2} + \frac{l^3}{3x^3} + \frac{l^4}{4x^4} + \dots\Big) =& l + \frac{l^2}{2x} + \frac{l^3}{3x^2} + \underbrace{\frac{l^4}{4x^3}}_{1} - \frac{l^2}{x} - \frac{l^3}{2x^2} - \underbrace{\frac{l^4}{3x^3}}_{1} - \underbrace{\frac{l^5}{4x^4}}_{2} + \frac{l}{2x} + \underbrace{\frac{l^2}{4x^2}}_{1} + \\
& \underbrace{\frac{l^3}{6x^3}}_{2} + \underbrace{\frac{l^4}{8x^4}}_{3} + \dots
\end{aligned} \tag{29}
$$

Using the rules stated by Laplace above we can ignore most of the terms on the RHS of (29). Terms marked 1 have order $\frac{1}{n}$, terms marked with 2 have order $\frac{1}{n\sqrt{n}}$ which is less than order $\frac{1}{n}$ and terms marked with 3 have order $\frac{1}{n^2}$ which is less than order $\frac{1}{n}$. Hence we are left with:

$$\begin{aligned}\left(x-l+\frac{1}{2}\right)\left(\frac{l}{x}+\frac{l^2}{2x^2}+\frac{l^3}{3x^3}+\frac{l^4}{4x^4}+\dots\right) &= l+\frac{l}{2x}+\frac{l^2}{2x}-\frac{l^2}{x}+\frac{l^3}{3x^2}-\frac{l^3}{2x^2}\\ &= l+\frac{l}{2x}-\frac{l^2}{2x}-\frac{l^3}{6x^2}\end{aligned} \tag{30}$$

Thus (25) becomes:

$$\left(l-x-\frac{1}{2}\right)\left[\ln x+\ln(1-\frac{l}{x})\right]=\left(l-x-\frac{1}{2}\right)\ln x+l+\frac{l}{2x}-\frac{l^2}{2x}-\frac{l^3}{6x^2} \tag{31}$$

Laplace then obtains the following expression by exponentiation:

$$(x-l)^{l-x-\frac{1}{2}}=e^{l-\frac{l^2}{2x}}\,x^{l-x-\frac{1}{2}}\left(1+\frac{1}{2x}-\frac{l^3}{6x^2}\right) \tag{32}$$

This results follows from (25) and the use of an approximation by ignoring terms of order $\frac{1}{n}$:

$$\begin{aligned}(x-l)^{l-x-\frac{1}{2}} &= e^{(l-x-\frac{1}{2})\,\ln x+l+\frac{l}{2x}-\frac{l^2}{2x}-\frac{l^3}{6x^2}}\\ &= e^{(l-x-\frac{1}{2})\,\ln x}\;e^{l+\frac{l}{2x}-\frac{l^2}{2x}-\frac{l^3}{6x^2}}\\ &= e^{\ln x^{l-x-\frac{1}{2}}}\;e^{l-\frac{l^2}{2x}}\,e^{\frac{l}{2x}-\frac{l^3}{6x^2}}\\ &= x^{l-x-\frac{1}{2}}\;e^{l-\frac{l^2}{2x}}\left[1+\frac{l}{2x}-\frac{l^3}{6x^2}+\frac{1}{2!}(\frac{l}{2x}-\frac{l^3}{6x^2})^2+\dots\right]\\ &= x^{l-x-\frac{1}{2}}\;e^{l-\frac{l^2}{2x}}\left[1+\underbrace{\frac{l}{2x}-\frac{l^3}{6x^2}}_{\text{terms of order }\frac{1}{\sqrt{n}}}+\frac{1}{2}\underbrace{(\frac{l^2}{4x^2}-\frac{l^4}{6x^3}+\frac{l^6}{36x^4})}_{\text{terms of order }\frac{1}{n}}+\dots\right]\\ &= e^{l-\frac{l^2}{2x}}\,x^{l-x-\frac{1}{2}}\left[1+\frac{l}{2x}-\frac{l^3}{6x^2}\right]\quad\text{ignoring terms of order }\frac{1}{n}\end{aligned} \tag{33}$$

Note that the terms of order $\frac{1}{\sqrt{n}}$ are greater than those of order $\frac{1}{n}$. By completely analogous reasoning (being careful with the signs) one also gets:

$$(x^{'}+l)^{-l-x^{'}-\frac{1}{2}}=e^{-l-\frac{l^2}{2x'}}\,x^{'\;-l-x'-\frac{1}{2}}\left[1-\frac{l}{2x'}+\frac{l^3}{6x'^2}\right] \tag{34}$$

Laplace's next step is to observe that since $p = \frac{x+s}{n+1}$ (see (17)) where $0 < s < 1$, by making:

$$p = \frac{x-z}{n} \tag{35}$$

z will be contained in the limits:

$$\frac{x}{n+1} \quad \text{and} \; -\frac{(n-x)}{n+1} \tag{36}$$

and consequently (ignoring the sign) less than 1. To show this we note the following:

$$\begin{aligned} \frac{x+s}{n+1} &= \frac{x-z}{n} \\ \frac{x}{n+1} - \frac{x}{n} + \frac{s}{n+1} &= -\frac{z}{n} \\ \frac{s}{n+1} - \frac{x}{n(n+1)} &= -\frac{z}{n} \\ \therefore z = \frac{x}{n+1} - \frac{ns}{n+1} \end{aligned} \tag{37}$$

When $s = 0$ the bound is $\frac{x}{n+1}$ and when $s = 1$ the bound is $\frac{x-n}{n+1} = -\frac{(n-x)}{n+1}$ but because $0 < s < 1$, z is between these bounds (save for sign).

Now because $p = \frac{x-z}{n}$ it follows that $1 - p = 1 - \frac{x-z}{n} = \frac{n-x+z}{n} = \frac{x'+z}{n}$. Laplace now uses these expressions to provide the following :

$$p^{x-l}\,(1-p)^{x'+l} = \frac{x^{x-l}\,x'^{\,x'+l}}{n^n}\left(1 + \frac{nzl}{xx'}\right) \tag{38}$$

To demonstrate this derivation involves the multiplication of the binomial expressions $(x-z)^{x-l}$ and $(x'+z)^{x'+l}$ and then ignoring terms of order $\frac{1}{n}$. Thus:

$$
\begin{aligned}
(x-z)^{x-l}(x'+z)^{x'+l} =& \Big(x^{x-l} - (x-l)x^{x-l-1}z + \frac{(x-l)(x-l-1)}{2!}x^{x-l-2}z^2 - \dots\Big)\times \\
& \Big(x'^{\,x'+l} + (x'+l)x'^{\,x'+l-1}z + \frac{(x'+l)(x'+l-1)}{2!}x'^{\,x'+l-2}z^2 - \dots\Big) \\
=& x^{x-l}x'^{\,x'+l}\Big[1 + \underbrace{(x'+l)\frac{z}{x'}}_{1} + \underbrace{\frac{(x'+l)(x'+l-1)}{2!}\frac{z^2}{x'^2}}_{3} \underbrace{-(x-l)\frac{z}{x}}_{2} \underbrace{-(x-l)(x'+l)\frac{z^2}{xx'}}_{4} \\
& \underbrace{-\frac{(x-l)(x'+l)(x'+l-1)}{2!}\frac{z^3}{xx'^2}}_{6} + \underbrace{\frac{(x-l)(x-l-1)}{2!}\frac{z^2}{x^2}}_{5} + \underbrace{\frac{(x-l)(x-l-1)(x'+l)}{2!}\frac{z^3}{x^2x'}}_{7} \\
& + \underbrace{\frac{(x-l)(x-l-1)(x'+l)(x'+l-1)}{2!2!}\frac{z^4}{x^2x'^2}}_{8}\Big]
\end{aligned}
\tag{39}
$$

In equation (38) the term $\frac{nzl}{xx'}$ has order $\frac{n\sqrt{n}}{n^2}$ = order $\frac{1}{\sqrt{n}}$ which, as observed before, is bigger than order $\frac{1}{n}$. Note that $|z| < 1$ so it is of order 1. The term $\frac{nzl}{xx'}$ is derived by adding 1 and 2 in equation (39) ie:

$$
(x'+l)\frac{z}{x'} - (x'-l)\frac{z}{x} = z\Big(\frac{xx' + xl - xx' + x'l}{xx'}\Big) = \frac{znl}{xx'} \quad \text{since } x + x' = n \tag{40}
$$

In what follows we ignore z because it is of order 1 although as the powers increase its effect in combination with the other terms ensures the higher powers have less effect. In equation (39) when we add the terms labelled $3, 4, 5$ we get something of order $\frac{1}{n}$, which can be seen as follows (noting that l is of order $\sqrt{n}$).

$$
\begin{aligned}
& \frac{(n+\sqrt{n})(n+\sqrt{n}-1)}{2n^2} - \frac{(n-\sqrt{n})(n+\sqrt{n})}{n^2} + \frac{(n-\sqrt{n})(n-\sqrt{n}-1)}{2n^2} \\
=& \frac{n^2 + 2n\sqrt{n} + n - n - \sqrt{n} - 2n^2 + 2n + n^2 - 2n\sqrt{n} + n - n + \sqrt{n}}{2n^2} = \frac{1}{n}
\end{aligned}
\tag{41}
$$

Terms 6 and 7 in equation (39) which involve z^3 , when added, have an effect of order $\frac{1}{n}$ as can be seen below:

$$
\begin{aligned}
& \frac{(n-\sqrt{n})(n-\sqrt{n}-1)(n+\sqrt{n})}{2n^3} - \frac{(n-\sqrt{n})(n+\sqrt{n})(n+\sqrt{n})}{2n^3} \\
=& \frac{(n^2-n)}{2n^3}\Big[n - \sqrt{n} - 1 - n + \sqrt{n}\Big] = -\frac{(n-1)}{2n^2} \quad \text{ie order } \frac{1}{n}
\end{aligned}
\tag{42}
$$

Of course, the above development does not show the effect of all the terms involving z^3 simply because I stopped at z^2 in the expansions at the start of (39). When you take into account the two additional terms involving z^3 the overall effect is of order $\frac{1}{n^2\sqrt{n}}$ which is less than order $\frac{1}{n}$. Similarly the effect of the terms of z^4 (such as term 8 in equation (39)) and higher is below order $\frac{1}{n}$.

Having obtained (40) Laplace then says that:

$$\frac{n!}{1.2.3\dots(x-l).1.2.3\dots(x-l)}\,p^{x-l}\,(1-p)^{x'+l} = \frac{\sqrt{n}\,e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}\Big[1+\frac{nzl}{xx'}+\frac{l(x-x')}{2xx'}-\frac{l^3}{6x^2}+\frac{l^3}{6x'^{\,2}}\Big] \tag{43}$$

Laplace has assembled all the components to derive (43). - see (22),(23),(24),(32),(34) and (38). Using those components the LHS of (43) is:

$$\begin{aligned} &n^{n+\frac{1}{2}}\,e^{-n}\,\sqrt{2\pi}\Big(1+\frac{1}{12n}\Big)e^{l-\frac{l^2}{2x}}\,x^{l-x-\frac{1}{2}}\Big(1+\frac{l}{2x}-\frac{l^3}{6x^2}\Big)\frac{e^{x-l}}{\sqrt{2\pi}}\Big(1-\frac{1}{12(x-l)}\Big)\\ &\times e^{-l-\frac{l^2}{2x'}}\,x'^{\,-l-x'-\frac{1}{2}}\Big(1-\frac{l}{2x'}+\frac{l^3}{6x'^{\,2}}\Big)\frac{e^{x'+l}}{\sqrt{2\pi}}\Big(1-\frac{1}{12(x'+l)}\Big)\frac{x^{x-l}\,x'^{\,x'+l}}{n^n}\Big(1+\frac{nzl}{xx'}\Big)\end{aligned} \tag{44}$$

The expression in (44) is messy and it is easy to get lost so I will simplify it in parts as follows. First we can ignore these terms:

$$\Big(1+\frac{1}{12n}\Big)\Big(1-\frac{1}{12(x-l)}\Big)\Big(1-\frac{1}{12(x'+l)}\Big) \tag{45}$$

each of which is of order $\Big(1+\text{order}\Big(\frac{1}{n}\Big)\Big)$. Hence the product is bounded by 1 . Some further simplification gives:

$$\begin{aligned} &n^{n+\frac{1}{2}}\,e^{-n}\,\sqrt{2\pi}\,e^{l-\frac{l^2}{2x}}\,x^{l-x-\frac{1}{2}}\,\frac{e^{x-l}}{\sqrt{2\pi}}\,e^{-l-\frac{l^2}{2x'}}\,x'^{\,-l-x'-\frac{1}{2}}\,\frac{e^{x'+l}}{\sqrt{2\pi}}\,\frac{x^{x-l}\,x'^{\,x'+l}}{n^n}\\ &=\frac{\sqrt{n}\,e^{-n}}{\sqrt{2\pi}\,\sqrt{xx'}}\,e^{-\frac{l^2}{2}\left(\frac{1}{x}+\frac{1}{x'}\right)}\,e^{x+x'}\\ &=\frac{\sqrt{n}\,e^{-n}\,e^{n}}{\sqrt{\pi}\,\sqrt{2xx'}}\,e^{-\frac{nl^2}{2xx'}}\\ &=\frac{\sqrt{n}}{\sqrt{\pi}\,\sqrt{2xx'}}\,e^{-\frac{nl^2}{2xx'}}\end{aligned} \tag{46}$$

Finally we come to the following three terms:

$$\begin{aligned}&\Big(1+\frac{l}{2x}-\frac{l^3}{6x^2}\Big)\Big(1-\frac{l}{2x'}+\frac{l^3}{6x'^2}\Big)\Big(1+\frac{nzl}{xx'}\Big)\\ =&\Big(1-\frac{l}{2x'}+\frac{l^3}{6x'^2}+\frac{l}{2x}-\underbrace{\frac{l^2}{4xx'}}_{1}+\underbrace{\frac{l^4}{12xx'^2}}_{1}-\frac{l^3}{6x^2}+\underbrace{\frac{l^4}{12x^2x'}}_{1}-\underbrace{\frac{l^6}{36x^2x'^2}}_{1}\Big)\Big(1+\frac{nzl}{xx'}\Big)\end{aligned} \tag{47}$$

The terms marked "1" are of order $\frac{1}{n}$ and can be ignored leaving:

$$\begin{aligned}\Big(1+\frac{l}{2}\Big(\frac{1}{x}-\frac{1}{x'}\Big)-\frac{l^3}{6x^2}+\frac{l^3}{6x'^2}\Big)\Big(1+\frac{nzl}{xx'}\Big)&=\Big(1+\frac{l(x'-x)}{2xx'}-\frac{l^3}{6x^2}+\frac{l^3}{6x'^2}\Big)\Big(1+\frac{nzl}{xx'}\Big)\\ &=\Big(1+\frac{nzl}{xx'}+\frac{l(x'-x)}{2xx'}-\frac{l^3}{6x^2}+\frac{l^3}{6x'^2}\Big)+\text{ terms of order }\frac{1}{n}\end{aligned} \tag{48}$$

Thus the LHS of (43) simplifies in the way claimed by Laplace:

$$\frac{\sqrt{n}\,e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}\Big[1+\frac{nzl}{xx'}+\frac{l(x-x')}{2xx'}-\frac{l^3}{6x^2}+\frac{l^3}{6x'^2}\Big] \tag{49}$$

Equation (49) is beginning to look familiar but there is still some way to go. Laplace says that "we will have the term anterior to the greatest term and which is extended from it at the distance *l*, by making *l* negative in this equation (ie (49)): by uniting next these two terms, their sums will be":

$$\frac{2\sqrt{n}\,e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}} \tag{50}$$

All Laplace is doing is moving away from the central term either side by *l* terms. When *l* is made negative in (49) and you get:

$$\frac{\sqrt{n}\,e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}\Big[1-\frac{nzl}{xx'}-\frac{l(x-x')}{2xx'}+\frac{l^3}{6x^2}-\frac{l^3}{6x'^2}\Big] \tag{51}$$

and so when you add (49) and (51) you get (50).

Laplace now moves to the sum of terms in (50) and states that the "finite integral

$$\sum \frac{2\sqrt{n}\, e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}} \tag{52}$$

taken from $l = 0$ inclusively, will express therefore the sum of all the terms of the binomial $[p + (1 - p)]^n$, comprehended between the two terms, of which we have p^{x+l} for factor, and the other has p^{x-l} for factor, and which are equidistant from the greatest term: but it is necessary to subtract from this sum the greatest term which is evidently contained twice."

Note that the greatest term occurs when $l = 0$ and hence is $\frac{\sqrt{n}}{\sqrt{\pi}\sqrt{2xx'}}$. To move from the sum in (52) to an integral Laplace uses finite difference techniques he has previously developed. Thus he makes the following statement. "Now, in order to have this finite integral, we will observe that we have, by §10 of Book 1, y being a function of l,

$$\sum y = \frac{1}{e^{\frac{dy}{dl}} - 1} = \left(\frac{dy}{dl}\right)^{-1} - \frac{1}{2}\left(\frac{dy}{dl}\right)^{0} + \frac{1}{12}\frac{dy}{dl} + \dots \tag{53}$$

whence we deduce, by the preceding section,

$$\sum y = \int y\, dl - \frac{1}{2}y + \frac{1}{12}\frac{dy}{dl} + \dots + constant \tag{54}$$

y being here equal to $\frac{2\sqrt{n}\, e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}$, the successive differentials of y acquire for factor $\frac{nl}{2xx'}$, and its powers." Note that $\left(\frac{dy}{dl}\right)^0 = y$.

Laplace had already developed the finite difference machinery earlier in his book (see [3], [4]) but rather than retrace all of that I will simply extract the material necessary to make sense of the relevant derivations.

2.1 Finite difference operator theory

We start with the basic operator Δ which acts on functions as follows:

$$\Delta f(x) = f(x + h) - f(x) \tag{55}$$

It follows that:

$$(1 + \Delta) f(x) = f(x) + f(x + h) - f(x) = f(x + h) \tag{56}$$

Now Taylor's theorem is used to expand $f(x+h)$ as follows:

$$\begin{aligned} f(x+h) =& f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \ldots \\ =& (1 + hD + \frac{h^2 D^2}{2!} + \ldots) f(x) \\ =& e^{hD} f(x) \end{aligned} \tag{57}$$

Note here that D is the derivative operator so that $D = \frac{d}{dx}$ and $D^2 = \frac{d^2}{dx^2}$ etc. Thus (56) and (57) lead to the following operator quality:

$$1 + \Delta = e^{hD} \tag{58}$$

Thus we can formally write:

$$\begin{aligned} \Delta^{-1} =& \frac{1}{e^{hD} - 1} \\ =& \frac{1}{1 + hD + \frac{h^2 D^2}{2!} + \frac{h^3 D^3}{3!} + \cdots - 1} \\ =& \frac{1}{hD + \frac{h^2 D^2}{2!} + \frac{h^3 D^3}{3!} + \ldots} \\ =& \frac{1}{hD} - \frac{1}{2} + \frac{hD}{12} - \frac{h^3 D^3}{720} + \ldots \end{aligned} \tag{59}$$

The last line in (59) is obtained by simple long division.

The meaning of the operator $\frac{1}{D} = D^{-1}$ is obtained by considering functions $f(x)$ and $F(x)$ such that:

$$F(x) = D^{-1} f(x) \tag{60}$$

If we apply D to both sides of (60) we have:

$$D[F(x)] = D[D^{-1}]f(x) = (DD^{-1})f(x) = f(x) \tag{61}$$

In other words $F(x)$ is that function whose derivative is $f(x)$ which is just another way of saying the following:

$$F(x) = \int f(x)\, dx + \text{constant} \tag{62}$$

Therefore:

$$D^{-1}f(x) = \int f(x)\,dx + \text{constant} \tag{63}$$

We can therefore interpret $D^{-1} = \frac{1}{D}$ as an integral operator ie:

$$D^{-1} = \frac{1}{D} = \int (\;)\;dx \tag{64}$$

If we suppose that $F(x)$ and $f(x)$ are functions such that:

$$\begin{aligned} \frac{\Delta}{\Delta x}F(x) = f(x) \quad &\text{or} \\ \Delta\, F(x) = h\, f(x) \quad &\text{where } h = \Delta x \end{aligned} \tag{65}$$

then applying the inverse operator Δ^{-1} to both sides of (65) gives::

$$F(x) = \Delta^{-1}[hf(x)] \tag{66}$$

It can be established that Δ^{-1} is a linear operator and if two functions have the same difference then they can differ ar most by an arbitrary periodic constant $C(x)$ such that $C(x + h) = C(x)$.

If we assume that $F(x)$ and $f(x)$ are such that:

$$F(x) = D^{-1}f(x) \tag{67}$$

Then:

$$\begin{aligned} D[F(x)] =& D[D^{-1}f(x)] \\ =& (DD^{-1})f(x) \\ =& f(x) \end{aligned} \tag{68}$$

In other words (68) says that F(x) is a that function whose derivative is $f(x)$. From integral calculus we know that:

$$F(x) = \int f(x)\,dx + constant \tag{69}$$

Therefore:

$$D^{-1}f(x) = \int f(x)\,dx + constant \tag{70}$$

or, in terms of operators:

$$D^{-1} = \int (\ )\ dx \tag{71}$$

By analogy then with integration (see (66)):

$$\Delta^{-1}[h\,f(x)] = \sum f(x)\,h + C_1(x) \quad \text{where } C_1(x) \text{ is a constant} \tag{72}$$

On dividing by h we have:

$$\Delta^{-1}[f(x)] = \sum f(x) + C(x) \tag{73}$$

Laplace's actual development of the finite difference approach can be found in [4], pages 5-9. The application of the theory to $y(l) = \frac{2\sqrt{n}\,e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}$ is equation (53). Recasting equation (53) in terms of the operator set out in (59) we see that, with $h = 1$ in (59):

$$\begin{aligned}
\Delta^{-1}y(l) &= \sum_l y(l) + constant \\
&= \frac{1}{e^D - 1}\,y(l) \\
&= \frac{1}{e^{\frac{dy}{dl}} - 1}\,y(l) \\
&= \Big[\frac{1}{hD} - \frac{1}{2} + \frac{hD}{12} - \frac{h^3D^3}{720} + \dots\Big]\,y(l)
\end{aligned} \tag{74}$$

Laplace's equation (54) the follows from (74) ie

$$\begin{aligned}
\sum y(l) &= \frac{1}{D}y(l) - \frac{1}{2}y(l) + \frac{D}{12}y(l + \cdots + \text{constant} \\
&= \int y(l)\,dl - \frac{1}{2}y(l) + \frac{1}{12}\frac{dy}{dl} + \cdots + \text{constant}
\end{aligned} \tag{75}$$

More details of the theory of finite differences can be found in [5].

At this point Laplace ignores some further terms based on the behaviour of derivatives of $y(l) = \frac{2\sqrt{n}\, e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}$ bearing in mind that it is supposed that l is of order $\sqrt{n}$. Thus $y' = \frac{-2n\sqrt{n}l}{\sqrt{\pi}\,\sqrt{2xx'}\,2xx'}\, e^{\frac{-nl^2}{2xx'}}$ which is of order $\frac{1}{n}$. Note that the exponential term is of order $e^{-\frac{1}{2}}$ so the size of the whole term is determined by the behaviour of the multiplying factor at each differentiation. Thus higher derivatives of y are smaller and Laplace ignores the terms from the first derivative onwards in (75) (or his (54)).

The next step in Laplace's argument is that "by starting with l the two finite and infinitely small integrals and designating by Y the greatest term of the binomial" we have:

$$\sum y = \int y\, dl - \frac{1}{2}y + \frac{1}{2}Y \tag{76}$$

The jump from (54) to (76) is justified as follows. He says that the "sum of all the terms of the binomial $[p + (1-p)]^n$ contained between the two terms equidistant from the greatest term by the number l being equal to $\sum y - \frac{1}{2}Y$, it will be:

$$\int y\, dl - \frac{1}{2}y \quad '' \tag{77}$$

In Laplace's terminology $\frac{Y}{2} = \frac{1}{2}y(0) = \frac{\sqrt{n}}{\sqrt{\pi}\,\sqrt{2xx'}}$ ie this is the central term which is the greatest. This term is included twice in the LHS of (76) - see (52) and the comments immediately following it. Thus:

$$\sum y - \frac{1}{2}Y = \int y\, dl - \frac{1}{2}y \tag{78}$$

The final step in Laplace's argument is that he says that "if one adds there (ie to (78)) the sum of these extreme terms, one will have, for the sum of all these terms:

$$\int y\, dl + \frac{1}{2}y \tag{79}$$

By making the substitution:

$$t = \frac{l\sqrt{n}}{\sqrt{2xx'}} \tag{80}$$

the sum therefore becomes:

$$\frac{2}{\sqrt{\pi}} \int e^{-t^2}\, dt + \frac{\sqrt{n}}{\sqrt{\pi}\,\sqrt{2xx'}}\, e^{-t^2} \tag{81}$$

Laplace notes that the terms being neglected are of order $\frac{1}{n}$ so that as n tends to infinity the approximation becomes better.

The jump from (79) to (80) requires some comment. Laplace is basically adding the extreme terms at $+l$ and $-l$ ie $\frac{1}{2}y(l) = \frac{1}{2}y(-l)$ where $y(l) = \frac{2\sqrt{n}\, e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}$. He is justified in doing this because when you look at the approximations made in getting to (49) and (51) (note z can be negative) the end points may not be hit and adding them in ensures that they are.

Finally Laplace notes that since $x = np + z$ with $|z| < 1$ one has:

$$\frac{x+l}{n} - p = \frac{np+z+l}{n} - p = \frac{l+z}{n} = \frac{t\sqrt{2xx'}}{n\sqrt{n}} + \frac{z}{n} \tag{82}$$

so that (82) "expresses the probability that the difference between the ratio of the number of times that the event a must arrive to the total number of trials, and the facility p of the event, is contained within the limits:

$$\pm\frac{t\sqrt{2xx'}}{n\sqrt{n}} + \frac{z}{n} \quad " \tag{83}$$

Laplace gives a practical demonstration of his approximation by supposing that the "facilities of the births of boys and girls are in the ratio of 18 to 17, and that there are born in one year 14000 infants: one demands the probability that the number of boys will not surpass 7363 and will not be less that 7037".

In this case:

$$\begin{aligned} p &= \frac{18}{35} \\ x &= 7200 \\ x' &= 6800 \\ n &= 14000 \\ l &= 163 \end{aligned} \tag{84}$$

Laplace says that (81) "gives quite nearly 0.994303 for the sought probability." We can verify this as follows. Using (80):

$$t = \frac{163 \times \sqrt{14000}}{\sqrt{2 \times 7200 \times 6800}} \tag{85}$$

Thus (81) becomes:

$$\frac{2}{\sqrt{\pi}} \int_0^t e^{-y^2} \, dy + \frac{\sqrt{14000}}{\sqrt{\pi} \, \sqrt{2 \times 7200 \times 6800}} \, e^{-t^2} \tag{86}$$

and when one uses Mathematica to perform the calculation the result is 0.994306 - a difference of 0.000003. Not bad for log tables and numerical approximation of the integral !!

The exact probability based on the binomial development is as follows. Let:

$$m = \sum_{k=0}^{163} \frac{14000!}{(7200-k)! \, (6800+k)!} \, p^{7200-k} \, (1-p)^{6800+k} \tag{87}$$

Thus m represents the right tail centred at $x = 7200$ and the left tail is by symmetry the same. Note that the central term is counted twice when we add the expressions for the two tails. and must be subtracted Thus the required probability is:

$$2m - \frac{14000!}{7200! \, 6800!} \, p^{7200} \, (1-p)^{6800} \tag{88}$$

When (88) is run through Mathematica the result is 0.994232.

When we recast Laplace's symbols in modern terms we can see that he knew that the expression for the maximum ordinate of the distribution is:

$$y_0 = \frac{1}{\sigma \sqrt{2\pi}} \tag{89}$$

where $\sigma = \sqrt{npq} = \sqrt{np(1-p)}$ is the standard deviation.

This is seen as follows noting that $p = \frac{x}{n}$ and $q = 1 - p = \frac{n-x}{n} = \frac{x'}{n}$. We start with $y(l) = \frac{2\sqrt{n}\, e^{-\frac{nl^2}{2xx'}}}{\sqrt{\pi}\,\sqrt{2xx'}}$ so the central term is:

$$
\begin{aligned}
\frac{1}{2}y(0) =& \frac{1}{2}\frac{2\sqrt{n}}{\sqrt{\pi}\sqrt{2xx'}} \\
=& \frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{n^2\left(\frac{x}{n}\right)\left(\frac{n-x}{n}\right)}} \\
=& \frac{1}{\sqrt{2\pi}}\frac{\sqrt{n}}{\sqrt{n}\sqrt{npq}} \\
=& \frac{1}{\sigma\sqrt{2\pi}}
\end{aligned}
\tag{90}
$$

3 Final observations

Laplace's derivation represents a fundamental shift in the level of sophistication that could be applied to probabilistic calculations. The huge literature on the central limit theorem reflects the significance of his achievement. The modern proofs of the central limit theorem tend to involve Fourier theory but one can find proofs in the spirit of Laplace's original proof in Leo Breiman's book "Probability" [6]. Steve Dunbar at the University of Nebraska-Lincoln has taken Breiman's proof and fleshed out the details - see [7].

Without the work done by De Moivre and Stirling in relation to an approximation for $n!$ the path to the final form of (81) would have been more difficult. Gven the somewhat tedious approximation arguments used by Laplace and his use of finite difference operator theory it is easy to see why his original proof is never given (at least to my knowledge) in undergraduate classes.

4 References

1. A scanned version of the book cn be found at `http://www.ime.usp.br/~walterfm/cursos/mac5796/DoctrineOfChances.pdf` - see pages 243-244.

2. Karl Pearson, "Historical note on the Origin of the Normal Curve of Errors", *Biometrika*,Vol. 16, No. 3/4 (Dec., 1924), pp. 402-404

3. `https://books.google.com.au/books?id=6MRLAAAAMAAJ&ie=ISO-8859-1&hl=en`

4. http://cerebro.xu.edu/math/Sources/Laplace/Source%20Book%20Selections.pdf

5. Murray R. Spiegel, Calculus of Finite Differences and Difference Equations , Schaum's Outline Series in Mathematics, McGraw Hill,1971.

6. Leo Breiman, Probability , Classics in Applied Mathematics, SIAM, 1992, pages 7-9.

7. http://www.math.unl.edu/~sdunbar1/ProbabilityTheory/Lessons/BernoulliTrials/DeMoivreLaplaceCLT/demoivrelaplaceclt.xml

5 History

Created 23/08/2015
Updated 24/08/2015 - added modern description of central term in terms of standard deviation.
Updated 17/10/2015 - corrected two scurrilous typos on the first page!!