Stirling's Formula and DeMoivre-Laplace Central Limit Theorem

Márton Balázs [1] and Bálint Tóth[1]

October 6, 2014

[1]University of Bristol / Budapest University of Technology and Economics

Using *Stirling's formula* we prove one of the most important theorems in probability theory, the *DeMoivre-Laplace Theorem*. The statement will be that under the appropriate (and different from the one in the Poisson approximation!) scaling the Binomial distribution converges to Normal. First we prove Stirling's formula for approximating factorials. This will be needed to estimate the binomial coefficient in the Binomial distribution.

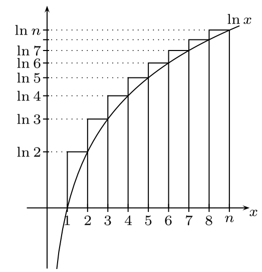
Theorem 1 (Stirling's-formula) *For large integer n's we have*

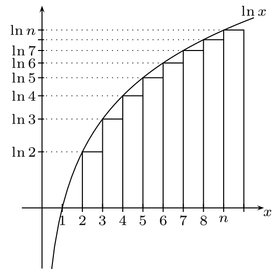
$$n! \simeq \frac{n^n}{\mathrm{e}^n}\cdot\sqrt{2\pi n}, \qquad \textit{to be more precise } \forall n>0 \qquad 1<\frac{n!}{\frac{n^n}{\mathrm{e}^n}\cdot\sqrt{2\pi n}}<\mathrm{e}^{\frac{1}{12n}}.$$

For large n's, the error term that bounds the fraction is approximately $1+\frac{1}{12n}$.

Proof. The value of π will not result from this proof yet. We will use the DeMoivre-Laplace Theorem to see this. Taking the logarithm of $n!$, $\ln n! = \sum_{k=1}^{n}\ln k$. We start with some intuition that explains why things are done later the way as shown. Comparing areas on the pictures, we have

$$\int_0^n \ln x\,\mathrm{d}x \le \sum_{k=1}^{n}\ln k \le \int_1^{n+1}\ln x\,\mathrm{d}x :$$





Computing the integrals,

$$n\ln n - n \le \ln n! \le (n+1)\ln(n+1) - n.$$

Our approximating formula will be between these two bounds. We pick $(n+1/2)\ln n - n$, and look at the difference

$$d_n = \ln n! - \Big[\Big(n+\frac{1}{2}\Big)\ln n - n\Big].$$

We show that

- d_n is monotone decreasing, thus converges to a limit d,
- for every n, we also have $d<d_n<d+\frac{1}{12n}$.

Once we have these properties, the statement follows:

$$\mathrm{e}^d < \mathrm{e}^{d_n} = \frac{n!}{\frac{n^{n+1/2}}{\mathrm{e}^n}} < \mathrm{e}^d\cdot\mathrm{e}^{\frac{1}{12n}}, \text{ that is,}$$

$$1<\frac{n!}{\frac{n^n}{\mathrm{e}^n}\cdot\sqrt{2\widehat{\pi} n}}<\mathrm{e}^{\frac{1}{12n}}$$

with some unknown constant $\widehat{\pi}$.

To see the two properties, we look at the "decrement" of d_n:

$$\begin{aligned}
d_n - d_{n+1} &= -\ln(n+1) - \Big(n+\frac{1}{2}\Big)\ln n + \Big(n+\frac{3}{2}\Big)\ln(n+1) - 1\\
&= \Big(n+\frac{1}{2}\Big)\ln\frac{n+1}{n} - 1\\
&= \Big(n+\frac{1}{2}\Big)\ln\frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}} - 1\\
&= \frac{1}{2t}\ln\frac{1+t}{1-t} - 1\\
&= \frac{1}{2t}\big[\ln(1+t)-\ln(1-t)\big] - 1,
\end{aligned}$$

where we intorduced the new variable $t = 1/[2n+1]$. Integrating the sum formula

$$\frac{1}{1-t} = \sum_{k=0}^{\infty} t^k$$

for the geometric series, and looking at the substitution at $t = 0$ we easily derive the Taylor series of $\ln(1-t)$:

$$\ln(1-t) = -\sum_{k=1}^{\infty} \frac{t^k}{k}, \qquad \text{from which} \qquad \ln(1+t) = -\sum_{k=1}^{\infty} (-1)^k \frac{t^k}{k}$$

also follows. Using this,

$$d_n - d_{n+1} = \frac{1}{2t} \sum_{k=1}^{\infty} \bigl(1 - (-1)^k\bigr) \cdot \frac{t^k}{k} - 1 = \frac{1}{t} \sum_{\ell=0}^{\infty} \frac{t^{2\ell+1}}{2\ell+1} - 1 = \frac{1}{t} \sum_{\ell=1}^{\infty} \frac{t^{2\ell+1}}{2\ell+1},$$

which shows that d_n is decreasing, therefore it has a limit d. We proceed with the sum on the right hand-side:

$$d_n - d_{n+1} < \sum_{\ell=1}^{\infty} \frac{t^{2\ell}}{3} = \frac{1}{3} \cdot \frac{t^2}{1-t^2} = \frac{1}{3} \cdot \frac{1}{(2n+1)^2 - 1} = \frac{1}{12} \cdot \frac{1}{n^2+n} = \frac{1}{12} \cdot \frac{1}{n} - \frac{1}{12} \cdot \frac{1}{n+1},$$

from which $d_n - \frac{1}{12n}$ is increasing. Thus

$$d_n - d = \lim_{m\to\infty} (d_n - d_m) = \lim_{m\to\infty} \Bigl(d_n - \frac{1}{12n} - d_m + \frac{1}{12m} + \frac{1}{12n} - \frac{1}{12m}\Bigr) < \frac{1}{12n} - \lim_{m\to\infty} \frac{1}{12m} = \frac{1}{12n}.$$

□

Let now $X \sim \mathrm{Binom}(n, p)$, p fixed, and take n to infinity. (This is very different from the Poisson approximation, where p went to zero like the reciprocal of n.) Below we use $q = 1 - p$. The standardised version

$$\frac{X - \mathbf{E}X}{\mathbf{D}X} = \frac{X - np}{\sqrt{npq}}$$

of the Binomial variable has zero mean and standard deviation one. The statement of the DeMoivre-Laplace Theorem will be that as $n \to \infty$, this standardized Binomial distribution converges to Standard Normal. However, as the Binomial is discrete and the Normal is continuous, it is not even clear at first reading how to formulate such a statement precisely. Consider the probability mass function of the Binomial variable X. As X is integer-valued, it is natural to associate the integer k with the real interval $[k - 1/2,\, k + 1/2)$:

$$p_X(k) = \mathbf{P}\{X = k\} = \mathbf{P}\Bigl\{k - \frac{1}{2} \le X < k + \frac{1}{2}\Bigr\} = \mathbf{P}\Bigl\{\frac{k - 1/2 - np}{\sqrt{npq}} \le \frac{X - np}{\sqrt{npq}} < \frac{k + 1/2 - np}{\sqrt{npq}}\Bigr\}.$$

The length $1/\sqrt{npq}$ of this rescaled interval goes to zero. By the intuitive meaning of the probability density function, we are now able to formulate what is meant by the convergence of the rescaled Binomial to the Normal distribution: for large n we expect

$$p_X(k) = \mathbf{P}\Bigl\{\frac{k - 1/2 - np}{\sqrt{npq}} \le \frac{X - np}{\sqrt{npq}} < \frac{k + 1/2 - np}{\sqrt{npq}}\Bigr\} \simeq \varphi\Bigl(\frac{k - np}{\sqrt{npq}}\Bigr) \cdot \frac{1}{\sqrt{npq}},$$

where φ is the Standard Normal density. This is the essence of the statement of the DeMoivre-Laplace Theorem:

Theorem 2 (DeMoivre-Laplace) *Let $X \sim Binom(n, p)$, where p is fixed. Let A_n be nondecreasing such that $\lim_{n\to\infty} A_n/n^{1/6} = 0$. Then*

$$\max_{|k-np| < \sqrt{n}\cdot A_n} \Bigl| \frac{p_X(k) \cdot \sqrt{npq}}{\varphi\bigl(\frac{k-np}{\sqrt{npq}}\bigr)} - 1 \Bigr| = \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr)$$

that is, the left hand-side stays bounded in n after dividing it by $A_n^3/\sqrt{n}$.

Proof. It will be a consequence of this theorem that the constant $\widehat{\pi}$ equals π, we do not need to assume this. Approximating the Binomial mass function via Stirling's formula,

$$\begin{aligned} p_X(k) &= \frac{n! \cdot p^k \cdot q^{n-k}}{k! \cdot (n-k)!} = \frac{n^n \cdot p^k \cdot q^{n-k}}{k^k \cdot (n-k)^{n-k}} \cdot \frac{\sqrt{2\widehat{\pi} n}}{\sqrt{2\widehat{\pi} k} \cdot \sqrt{2\widehat{\pi}(n-k)}} \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{1}{n}\Bigr)\Bigr) \\ &= \frac{1}{\sqrt{2\widehat{\pi} npq}} \cdot \frac{1}{\bigl(\frac{n-k}{nq}\bigr)^{n-k} \bigl(\frac{k}{np}\bigr)^k} \cdot \frac{1}{\bigl(\frac{n-k}{nq}\bigr)^{1/2} \bigl(\frac{k}{np}\bigr)^{1/2}} \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{1}{n}\Bigr)\Bigr) \\ &= \frac{1}{\sqrt{2\widehat{\pi} npq}} \cdot \underbrace{\frac{1}{\bigl(1 - \frac{k-np}{nq}\bigr)^{n-k} \bigl(1 + \frac{k-np}{np}\bigr)^k}}_{(I)} \cdot \underbrace{\frac{1}{\bigl(1 - \frac{k-np}{nq}\bigr)^{1/2} \bigl(1 + \frac{k-np}{np}\bigr)^{1/2}}}_{(II)} \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{1}{n}\Bigr)\Bigr). \end{aligned}$$

The advantage here is that according to the condition in the max, we have $[k - np]/n < A_n/\sqrt{n}$ which goes to zero. Therefore taking logarithm of the term (I) and expanding it to second order we have

$$\begin{aligned} \ln(I) &= (k-n) \cdot \ln\Bigl(1 - \frac{k-np}{nq}\Bigr) - k \cdot \ln\Bigl(1 + \frac{k-np}{np}\Bigr) \\ &= -(k-n) \cdot \frac{k-np}{nq} - k \cdot \frac{k-np}{np} - \frac{1}{2} \cdot (k-n) \cdot \Bigl(\frac{k-np}{nq}\Bigr)^2 + \frac{1}{2} \cdot k \cdot \Bigl(\frac{k-np}{np}\Bigr)^2 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}^3}\Bigr) \cdot n \\ &= \frac{(n-k)p - kq}{npq} \cdot (k-np) + \frac{(n-k)p^2 + kq^2}{2n^2p^2q^2} \cdot (k-np)^2 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr) \\ &= \frac{np - k\overbrace{(p+q)}^{1}}{npq} \cdot (k-np) + \frac{np^2 - k(p-q)\overbrace{(p+q)}^{1}}{2n^2p^2q^2} \cdot (k-np)^2 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr) \\ &= \frac{-2npq + np^2 - k(p-q)}{2n^2p^2q^2} \cdot (k-np)^2 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr) \\ &= \frac{-npq + np(p-q) - k(p-q)}{2n^2p^2q^2} \cdot (k-np)^2 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr) \\ &= -\frac{(k-np)^2}{2npq} - \frac{(p-q)(k-np)^3}{2n^2p^2q^2} + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr) \\ &= -\frac{(k-np)^2}{2npq} + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr), \end{aligned}$$

since the second term is also $\mathcal{O}(A_n^3/\sqrt{n})$. Take the exponential again:

$$(I) = \mathrm{e}^{-\frac{(k-np)^2}{2npq}} \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr)\Bigr).$$

In the term (II) we only look at the constant part of the series expansion, with first order error bound:

$$(II) = 1 + \mathcal{O}\Bigl(\frac{A_n}{\sqrt{n}}\Bigr).$$

Combining the details,

$$\begin{aligned} p_X(k) &= \frac{1}{\sqrt{2\widehat{\pi} npq}} \cdot \mathrm{e}^{-\frac{(k-np)^2}{2npq}} \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr)\Bigr) \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{A_n}{\sqrt{n}}\Bigr)\Bigr) \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{1}{n}\Bigr)\Bigr) \\ &= \frac{1}{\sqrt{2\widehat{\pi} npq}} \cdot \mathrm{e}^{-\frac{(k-np)^2}{2npq}} \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr) + \mathcal{O}\Bigl(\frac{A_n}{\sqrt{n}}\Bigr) + \mathcal{O}\Bigl(\frac{1}{n}\Bigr)\Bigr) \\ &= \frac{1}{\sqrt{2\widehat{\pi} npq}} \cdot \mathrm{e}^{-\frac{(k-np)^2}{2npq}} \cdot \Bigl(1 + \mathcal{O}\Bigl(\frac{A_n^3}{\sqrt{n}}\Bigr)\Bigr) \end{aligned}$$

where we picked the largest of the error bounds in the last step. □

In the DeMoivre-Laplace Theorem mass functions and densitiy functions are featured, thus we have a *local limit theorem*. Notice that fine asymptotics and error bounds are also part of the statement. The version comparing distribution functions is the *global limit theorem*:

Theorem 3 (Global form of the DeMoivre-Laplace Theorem) *Let p be fixed, and $X \sim Binom(n, p)$. Then for all $a < b$ fixed reals, we have*

$$\lim_{n\to\infty} \mathbf{P}\Bigl\{a \le \frac{X - np}{\sqrt{npq}} < b\Bigr\} = \Phi(b) - \Phi(a).$$

Proof. By the triangle inequality,

$$\begin{aligned} \Bigl|\mathbf{P}\Bigl\{a \le \frac{X-np}{\sqrt{npq}} < b\Bigr\} - \bigl(\Phi(b) - \Phi(a)\bigr)\Bigr| &\le \Bigl|\sum_{k=\lceil a\sqrt{npq}+np\rceil}^{\lceil b\sqrt{npq}-1+np\rceil} p_X(k) - \int_a^b \varphi(x)\,\mathrm{d}x\Bigr| \\ &\le \Bigl|\sum_{k=\lceil a\sqrt{npq}+np\rceil}^{\lceil b\sqrt{npq}-1+np\rceil} \Bigl(p_X(k) - \frac{1}{\sqrt{npq}} \cdot \varphi\Bigl(\frac{k-np}{\sqrt{npq}}\Bigr)\Bigr)\Bigr| \qquad (1) \\ &\quad + \Bigl|\sum_{k=\lceil a\sqrt{npq}+np\rceil}^{\lceil b\sqrt{npq}-1+np\rceil} \frac{1}{\sqrt{npq}} \cdot \varphi\Bigl(\frac{k-np}{\sqrt{npq}}\Bigr) - \int_a^b \varphi(x)\,\mathrm{d}x\Bigr|. \end{aligned}$$

The second term is the difference between the integral of the function φ and the Riemann sum of this integral, therefore this term goes to zero. For the first term, let $c > \max(b\sqrt{pq},\, a\sqrt{pq})$. Rewriting the DeMoivre-Laplace Theorem,

$$\max_{|k-np|<c\sqrt{n}} \Big| p_X(k) - \frac{1}{\sqrt{npq}} \cdot \varphi\Big(\frac{k-np}{\sqrt{npq}}\Big) \Big| = \max_{|k-np|<c\sqrt{n}} \Big| \frac{p_X(k)\cdot\sqrt{npq}}{\varphi\big(\frac{k-np}{\sqrt{npq}}\big)} - 1 \Big| \cdot \frac{1}{\sqrt{npq}} \cdot \varphi\Big(\frac{k-np}{\sqrt{npq}}\Big)$$
$$\le \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big) \cdot \frac{1}{\sqrt{2\widehat{\pi} npq}} = \mathcal{O}\Big(\frac{1}{n}\Big)$$

for all k's in (1). The sum (1) has only $\mathcal{O}(\sqrt{n})$ many summands, thus the sum of the error bounds is still in the order of $\mathcal{O}(1/\sqrt{n})$.

The global theorem is now proved for the density $\varphi(x) = (1/\sqrt{2\widehat{\pi}}) \cdot e^{-x^2/2}$. However, it then follows that this function φ is a proper density function, from which we also conclude that $\widehat{\pi} = \pi$. □

We remark here that the global theorem is a special case of the so-called *Central Limit Theorem.*

Example 4 *The ideal size of a course is 150 students. On average 30% of those accepted will enroll, therefore the organisers accept 450 students. What is the probability that more than 150 students enroll?*

Assuming independence, the number of those who enroll is $X \sim \mathrm{Binom}(450,\, 0.3)$. Using the global theorem, $(np = 135,\, \sqrt{npq} \simeq 9.72)$

$$\mathbf{P}\{X > 150\} = \mathbf{P}\{X > 150.5\} \simeq \mathbf{P}\Big\{\frac{X-135}{9.72} > \frac{150.5-135}{9.72}\Big\} \simeq 1 - \Phi\Big(\frac{150.5-135}{9.72}\Big) \simeq 1 - \Phi(1.59) \simeq 0.0559.$$

Example 5 *Flipping a fair coin 40 times, what is the probability that exactly 20 Heads result?*

The answer to this question is of course $\binom{40}{20} \cdot (1/2)^{40} \simeq 0.1254$. An approxmative answer can be given using the DeMoivre-Laplace Theorem:

$$p(20) \simeq \frac{1}{\sqrt{npq}} \cdot \varphi\Big(\frac{20-np}{\sqrt{npq}}\Big) = \frac{1}{\sqrt{20\pi}} e^0 \simeq 0.1262.$$

We can also use the global version of the theorem, X is the number of Heads that result:

$$\mathbf{P}\{X = 20\} = \mathbf{P}\{19.5 \le X < 20.5\} = \mathbf{P}\Big\{\frac{19.5-np}{\sqrt{npq}} \le \frac{X-np}{\sqrt{npq}} < \frac{20.5-np}{\sqrt{npq}}\Big\}$$
$$= \mathbf{P}\Big\{-\frac{0.5}{\sqrt{10}} \le \frac{X-20}{\sqrt{10}} < \frac{0.5}{\sqrt{10}}\Big\}$$
$$\simeq \Phi\Big(\frac{0.5}{\sqrt{10}}\Big) - \Phi\Big(-\frac{0.5}{\sqrt{10}}\Big)$$
$$= 2\Phi\Big(\frac{0.5}{\sqrt{10}}\Big) - 1 \simeq 0.1256.$$

Below we plot the graphs connected to this example. The solid line represents the standard normal density function (the denominator in the DeMoivre-Laplace Theorem). Dots plot $\sqrt{npq} = \sqrt{10}$ times the probability that the Binomial random variable takes on value $\sqrt{npq} \cdot x + np = \sqrt{10} \cdot x + 20$ when this is an integer (this is the numerator in the DeMoivre Laplace Theorem). One can hardly spot any difference.