Supplement 5: Stirling's Approximation to the Factorial

S5.1 Stirling's approximation

In this supplement, we prove **Stirling's approximation** to the factorial. It is a remarkably righteous piece of analysis. Recall that

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1.$$

This is well approximated by

$$s(n) = \sqrt{2\pi}\,\sqrt{n}\,n^n\,e^{-n} = \sqrt{2\pi}\,n^{n+(1/2)}\,e^{-n}.$$

Stirling's approximation is often stated as

$$n! \sim s(n),$$

where the notation $a_n \sim b_n$ simply means that $\lim_{n\to\infty} a_n/b_n = 1$. But we can be much more precise than this.

S5.1.1 Stirling's approximation *For $n \geqslant 1$,*

$$n! = \sqrt{2\pi}\;n^{n+(1/2)}e^{-n}\;e^{\varepsilon_n} \tag{1}$$

where $\varepsilon_n \to 0$ as $n \to \infty$. Indeed,

$$\frac{1}{12n+1} < \varepsilon_n < \frac{1}{12n}.$$

The ratio $n!/s(n) = e^{\varepsilon_n}$ could also be (and usually is) written as $1+\eta_n$. Both expressions have the feature that $\lim_{\varepsilon\to 0} e^{\varepsilon} = \lim_{\eta\to 0} 1+\eta = 1$. I find the exponential form a little more convenient. Note that $s(n)$ is always an underestimate of $n!$.

Table S5.1 shows that this is a very good approximation, even for small values of n.

S5.2 Proof, Part I

This proof is based in part on that of Herbert Robbins [13]. He in turn attributes the main geometric idea to Georges Darmois [4, pp. 315–317]. We start with the logarithm of $n!$:

$$\ln n! = \sum_{k=1}^{n} \ln k = \sum_{k=1}^{n-1} \ln(k+1). \tag{2}$$

Think of $\ln(k+1)$ as the area of a rectangle of height $\ln(k+1)$ with base $[k, k+1]$. Now the logarithm function is nearly linear, so the area of this rectangle is approximately the area under the graph of the logarithm plus a triangle. See Figure S5.1. The graph of the logarithm is shown in red. The area I_k under the graph is composed of two pieces, the cyan quadrilateral and the pink "crescent" designated D_k. More precisely, let

$$I_k = \int_k^{k+1} \ln x\,dx,$$

n	$n!$	$s(n)$	$r(n) = n!/s(n)$	$\varepsilon_n = \ln r(n)$	$\eta_n = r(n) - 1$
1	1	0.922137	1.08444	0.0810615	0.0844376
2	2	1.919	1.04221	0.0413407	0.0422071
3	6	5.83621	1.02806	0.0276779	0.0280645
4	24	23.5062	1.02101	0.0207907	0.0210083
5	120	118.019	1.01678	0.0166447	0.016784
6	720	710.078	1.01397	0.0138761	0.0139728
7	5040	4980.4	1.01197	0.0118967	0.0119678
8	$40,320$	$39,902.4$	1.01047	0.0104113	0.0104657
9	$362,880$	$359,537.0$	1.0093	0.00925546	0.00929843
10	$3,628,800$	$3,598,700.0$	1.00837	8.33056×10^{-3}	8.36536×10^{-3}
100	–	9.32485×10^{157}	1.00083	8.33331×10^{-4}	8.33678×10^{-4}
1000	–	4.02354×10^{2567}	1.00008	8.33333×10^{-5}	8.33368×10^{-5}
10^4	–	$2.84624 \times 10^{35,659}$	1.00001	8.33333×10^{-6}	8.33337×10^{-6}
10^5	–	$2.82423 \times 10^{456,573}$	1.	8.33333×10^{-7}	8.33334×10^{-7}

Table S5.1. Comparison of $n!$ and $s(n) = \sqrt{2\pi n}\, n^n\, e^{-n}$. The ratio $n!/s(n)$ is denoted $r(n)$. To save space, the values of $n!$ have been omitted for larger n. (Computed by Mathematica 11.2.)

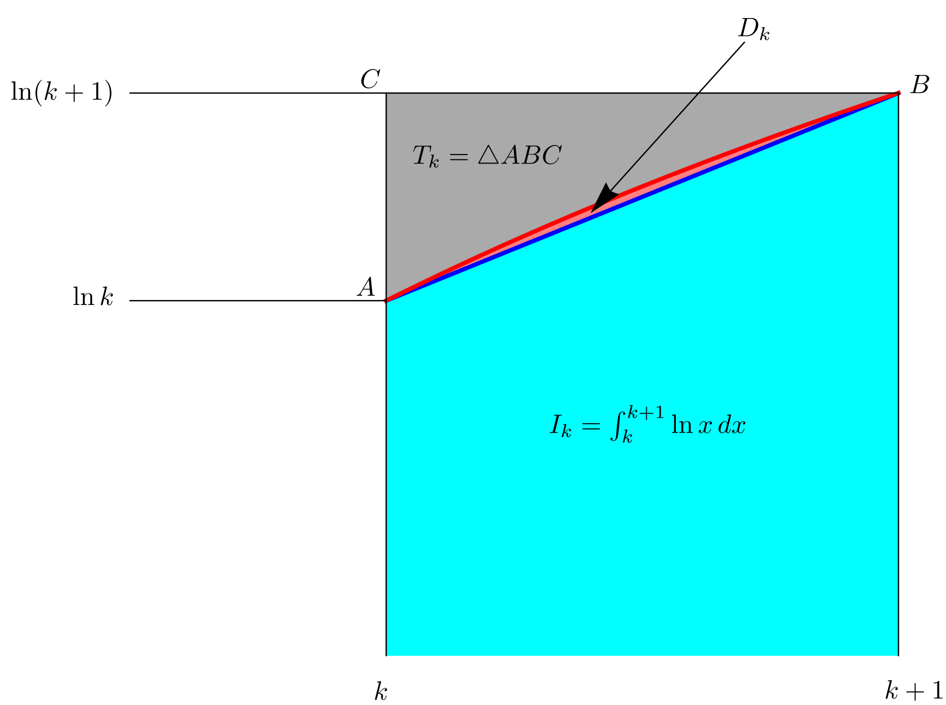

Figure S5.1.

and denote the area of the triangle ABC by

$$T_k = (\ln(k+1) - \ln k)/2.$$

Then the sum $I_k + T_k$ overstates the area of the rectangle by double counting the area D_k bounded by the graph of the logarithm and the graph of its linear interpolation, namely

$$D_k = I_k - T_k - \ln k.$$

In other words,

$$\ln(k+1) = I_k + T_k - D_k.$$

So by (2),

$$\begin{aligned} \ln n! = \sum_{k=1}^{n-1} \ln(k+1) &= \sum_{k=1}^{n-1} I_k + \sum_{k=1}^{n-1} T_k - \sum_{k=1}^{n-1} D_k \\ &= \int_1^n \ln x \, dx + \frac{1}{2} \ln n - \sum_{k=1}^{n-1} D_k. \end{aligned} \tag{3}$$

(Note that $\sum_k T_k$ is a telescoping sum.) Now the indefinite integral of $\ln x$ is $x \ln x - x$, so

$$\int_1^n \ln x \, dx = n \ln n - n + 1,$$

and (3) becomes

$$\ln n! = \left(n + \frac{1}{2}\right) \ln n - n + 1 - \sum_{k=1}^{n-1} D_k. \tag{4}$$

Now let's take a break, and exponentiate this to get

$$n! = n^{n+(1/2)} e^{-n} \, C_n = C_n \, e^{-n} n^n \sqrt{n}$$

where $C_n = \exp(1 - \sum_{k=1}^{n-1} D_k)$. This is beginning to look like Stirling's approximation (1), so we might be on the right track. Let's look more closely at each term D_k.

The area D_k is the discrepancy between the integral of $\ln x$ from k to $k+1$ and the irregular quadrilateral with vertexes at $(k, 0)$, $(k, \ln k)$, $\big(k+1, \ln(k+1)\big)$, $(k+1, 0)$. (Again refer to Figure S5.1.) So using the integral of the logarithm, we get

$$\begin{aligned} D_k &= \int_k^{k+1} \ln x \, dx - \left[\ln k + \frac{1}{2}\big(\ln(k+1) - \ln k\big)\right] \\ &= \Big[(k+1)\ln(k+1) - (k+1) - k \ln k + k\Big] - \left[\ln k + \frac{1}{2}\big(\ln(k+1) - \ln k\big)\right] \\ &= \frac{2k+1}{2}\big(\ln(k+1) - \ln k\big) - 1 \\ &= \frac{2k+1}{2} \ln\left(\frac{k+1}{k}\right) - 1. \end{aligned}$$

The next step is one that I confess would not have occurred to me. Let

$$x_k = \frac{1}{2k+1}, \quad \text{so that } \frac{1+x_k}{1-x_k} = \frac{k+1}{k},$$

and rewrite the above as

$$D_k = \frac{1}{2x_k} \ln\left(\frac{1+x_k}{1-x_k}\right) - 1. \tag{5}$$

Now we use the series

$$\ln\Big(\frac{1+x}{1-x}\Big) = 2\Big(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots\Big),$$

which is valid for $|x| < 1$. (This follows from the infinite Taylor series for $\ln(1+x) - \ln(1-x)$, see Apostol [1, equation 10.32, p. 390].) Then (6) can be written as

$$D_k = \frac{1}{3(2k+1)^2} + \frac{1}{5(2k+1)^4} + \frac{1}{7(2k+1)^6} + \cdots . \tag{6}$$

An upper bound on D_k is given by

$$\begin{aligned} D_k &= \frac{1}{3(2k+1)^2} + \frac{1}{5(2k+1)^4} + \frac{1}{7(2k+1)^6} + \cdots \\ &< \frac{1}{3(2k+1)^2} + \frac{1}{3(2k+1)^4} + \frac{1}{3(2k+1)^6} + \cdots \\ &= \frac{1}{3}\left[\frac{1}{(2k+1)^2} + \frac{1}{(2k+1)^4} + \frac{1}{(2k+1)^6} + \cdots\right]. \end{aligned}$$

The term in brackets is the geometric series $\theta + \theta^2 + \theta^3 + \cdots$, where $\theta = 1/(2k+1)^2 < 1$. This series has sum $\theta/(1-\theta)$, so

$$D_k < \frac{1}{3}\frac{1}{(2k+1)^2 - 1} = \frac{1}{3}\,\frac{1}{4k(k+1)} = \frac{1}{12}\Big(\frac{1}{k} - \frac{1}{k+1}\Big). \tag{7}$$

This is a telescoping sequence, so it follows that the infinite sum

$$D = \sum_{k=1}^{\infty} D_k$$

is convergent and its tails satisfy

$$\sum_{k=n}^{\infty} D_k < \frac{1}{12}\sum_{k=n}^{\infty}\Big(\frac{1}{k} - \frac{1}{k+1}\Big) = \frac{1}{12n}. \tag{8}$$

Likewise we can use (6) to get a lower bound for D_k:

$$\begin{aligned} D_l &= \frac{1}{3(2k+1)^2} + \frac{1}{5(2k+1)^4} + \frac{1}{7(2k+1)^6} + \cdots \\ &> \frac{1}{3(2k+1)^2} + \frac{1}{9(2k+1)^4} + \frac{1}{27(2k+1)^6} + \cdots . \end{aligned}$$

This is another geometric series $\theta + \theta^2 + \theta^3 + \cdots$, where $\theta = 1/3(2k+1)^2$. This series has sum $\theta/(1-\theta)$, so

$$\begin{aligned} D_k &> \frac{1}{3(2k+1)^2 - 1} \\ &= \frac{1}{12k^2 + 12k + 2} \\ &> \frac{1}{12k^2 + 12k + 2k + 1 + \frac{1}{12}} \\ &= \frac{1}{12}\,\frac{1}{\big(k + (1/12)\big)\big(k + 1 + (1/12)\big)} \\ &= \frac{1}{12}\Big(\frac{1}{k + (1/12)} - \frac{1}{k + 1 + (1/12)}\Big). \end{aligned} \tag{9}$$

Again we have found a telescoping sequence, so the tails satisfy

$$\sum_{k=n}^{\infty} D_k > \frac{1}{12}\sum_{k=n}^{\infty}\Big(\frac{1}{k+(1/12)} - \frac{1}{k+1+(1/12)}\Big) = \frac{1}{12}\,\frac{1}{n+(1/12)} = \frac{1}{12n+1}. \tag{10}$$

Now go back to equation (4) and rewrite it as

$$\begin{aligned}\ln n! &= \Big(n+\frac{1}{2}\Big)\ln n - n + 1 - \sum_{k=1}^{n-1} D_k \\ &= \Big(n+\frac{1}{2}\Big)\ln n - n + 1 - D + \sum_{k=n}^{\infty} D_k. \end{aligned}\tag{11}$$

In other words, $\Big(n+\frac{1}{2}\Big)\ln n - n + 1 - D$ underestimates $\ln n!$ by the amount

$$\varepsilon_n = \sum_{k=n}^{\infty} D_k,$$

where by (8) and (10) we have

$$\frac{1}{12n+1} < \varepsilon_n < \frac{1}{12n}.$$

Exponentiating (11) gives the following.

There is a constant $C > 0$ such that for each $n > 1$,

$$n! = C\, n^{n+(1/2)} e^{-n}\, e^{\varepsilon_n}, \tag{12}$$

where

$$C = e^{1-D},$$

so

$$e^{11/12} > C > e^{12/13},$$

and for each n,

$$\frac{1}{12n+1} < \varepsilon_n < \frac{1}{12n}. \tag{13}$$

S5.2.1 Remark Inequalities (13) imply that $1 < e^{\varepsilon_n}$, so that $s(n) = C\, n^{n+(1/2)} e^{-n}$ always underestimates $n!$, but still $n! \sim s(n)$. Thus the approximation could be improved by adding a factor $e^{1/(1+12n)}$, but that would destroy its simple beauty.

For many purposes, e.g., Example 3.8.1 or Proposition S6.6.4, we do not need to know the value of the constant C. This is where Robbins left off, writing, "The constant C [...] may be shown by one of the usual methods to have the value $\sqrt{2\pi}$." We take this up next.

Let C_n denote the ratio

$$C_n = \frac{n!}{n^{n+1/2}e^{-n}}.$$

That is, $C_n = C\, e^{\varepsilon_n}$, so inequalities (13) imply that $C_n \to C$.

In Lemma S5.3.1 in the next section we prove that

$$\pi n \leqslant \left(\frac{2^{2n}(n!)^2}{(2n)!}\right)^2 \leqslant \pi\big(n+(1/2)\big), \tag{14}$$

which Apostol refers to as Wallis' inequality. Following Apostol [2, p. 616], substitute

$$C_n n^{n+1/2} e^{-n} = n!$$

in (14) to get

$$\pi n \leqslant \left(\frac{2^{2n} n^{2n+1} e^{-2n} C_n^2}{(2n)^{2n+1} e^{-2n} C_{2n}} \right)^2 \leqslant \pi\big(n + (1/2)\big).$$

Dividing by n, this is equivalent to

$$\pi \leqslant \frac{C_n^4}{2C_{2n}^2} \leqslant \pi\big(1 + (1/2n)\big).$$

Letting $n \to \infty$, since $C_n \to C$ we see that

$$\pi \leqslant \frac{C^2}{2} \leqslant \pi,$$

or

$$C = \sqrt{2\pi}.$$

This completes the proof of Stirling's approximation, except for Wallis' inequality.

S5.3 Wallis' Inequality

S5.3.1 Lemma (Wallis' inequality) *For $n \geqslant 1$,*

$$\pi n \leqslant \left(\frac{2^{2n} (n!)^2}{(2n)!} \right)^2 \leqslant \pi\big(n + (1/2)\big).$$

Proof: (Cf. Apostol [2, pp. 617–618].) For $n = 0, 1, \dots,$ define

$$I_n = \int_0^{\pi/2} \sin^n t \, dt.$$

Note that

$$I_0 = \frac{\pi}{2}, \quad \text{and} \quad I_1 = 1.$$

Note that the integral denoted by I_k in this section is not the same as the integral denoted by I_k in the previous section. My bad.

Now we evaluate I_n recursively. Start by cleverly using the identity

$$\begin{aligned} \frac{d}{dt}(\cos t \, \sin^{n+1} t) &= -\sin^{n+2} t + \cos^2 t \, (n+1) \sin^n t && \text{(chain rule)} \\ &= -\sin^{n+2} + (1 - \sin^2 t)(n+1) \sin^n t && (\cos^2 + \sin^2 = 1) \\ &= (n+1) \sin^n t - (n+2) \sin^{n+2} t && \text{(regroup)} \end{aligned}$$

Integrating this over the interval $[0, \pi/2]$ gives

$$0 = \cos t \, \sin^{n+1} t \Big|_0^{\pi/2} = (n+1) I_n - (n+2) I_{n+2},$$

so

$$I_{n+2} = \frac{n+1}{n+2} I_n. \tag{15}$$

Since we know I_0 and I_1, we can use (15) to compute each I_n. For even n, say $n = 2k - 2$, equation (15) becomes

$$\frac{I_{2k}}{I_{2k-2}} = \frac{2k-1}{2k} = \frac{2k(2k-1)}{(2k)^2},$$

so

$$\prod_{k=1}^{n} \frac{I_{2k}}{I_{2k-2}} = \prod_{k=1}^{n} \frac{(2k-1)2k}{(2k)^2} = \frac{(1\cdot 2)(3\cdot 4)\cdots(2n-1)\cdot 2n}{2^{2n} n! n!} = \frac{(2n)!}{2^{2n}(n!)^2}.$$

Now the left-hand side telescopes to I_{2n}/I_0, so since $I_0 = \pi/2$ we get

$$I_{2n} = \frac{(2n)!}{2^{2n}(n!)^2} \frac{\pi}{2}. \tag{16}$$

For odd n, say $n = 2k-1$, equation (15) becomes

$$\frac{I_{2k+1}}{I_{2k-1}} = \frac{2k}{2k+1} = \frac{(2k)^2}{2k(2k-1)},$$

so

$$\prod_{k=1}^{n} \frac{I_{2k+1}}{I_{2k-1}} = \prod_{k=1}^{n} \frac{(2k)^2}{2k(2k-1)} = \frac{2^{2n}(n!)^2}{(2n)!} = \frac{1}{2n+1} \frac{\pi}{2} \frac{1}{I_{2n}},$$

where the last equality uses (16). This left-hand side telescopes to I_{2n+1}/I_1, so since $I_1 = 1$ we get

$$I_{2n} I_{2n+1} = \frac{1}{2n+1} \frac{\pi}{2}. \tag{17}$$

From (15), we have

$$I_{2n-1} = I_{2n+1} \frac{2n+1}{2n}.$$

Substitute this into (17) to get

$$I_{2n} I_{2n-1} = I_{2n} I_{2n+1} \frac{2n+1}{2n} = \frac{\pi}{2(2n+1)} \frac{2n+1}{2n} = \frac{\pi}{4n}. \tag{18}$$

Observe that for $0 \leqslant t \leqslant \pi/2$, we have $0 \leqslant \sin t \leqslant 1$ so $\sin^n t$ is decreasing in n for each t. Thus $I_0 > I_1 > I_2 > \cdots > 0$. As a result, for each $n \geqslant 1$,

$$\frac{1}{I_{2n-1}} < \frac{1}{I_{2n}} < \frac{1}{I_{2n+1}}.$$

Multiplying by the positive quantity $1/I_{2n}$ above yields

$$\frac{1}{I_{2n} I_{2n-1}} < \frac{1}{I_{2n}^2} < \frac{1}{I_{2n} I_{2n+1}}.$$

Using equations (18), (16) and (17), this becomes

$$\frac{4n}{\pi} < \left(\frac{2^{2n}(n!)^2}{(2n)!} \right)^2 \frac{4}{\pi^2} < \frac{2(2n+1)}{\pi}.$$

Now multiply these inequalities $\pi^2/4$ to get Wallis' inequality. Whew! ∎

S5.4 Historical notes

Abraham De Moivre published his *Miscellanea Analytica* in 1730. According to Karl Pearson [11], the rare 1733 edition had an extra supplement, which included equation (12), and Pearson had access to the later edition. De Moivre calculated C to be approximately 2.5074. It remained for James Stirling to show that $C = \sqrt{2\pi} \approx 2.50663$. Stirling based his argument on John Wallis' infinite product for π, which Wallis discovered in 1655 while working on $\int_0^\pi \sin^n t\, dt$.

Pearson believed that Stirling's contribution was not sufficient to warrant naming the result after him. Oh, well. Apostol's derivation uses what he term's Wallis' inequality, rather than an infinite product. Apostol comes to (12) via Euler's summation formula. The Robbins approach has the advantage of a tighter bound on the error terms ε_n. Apostol's approach shows only $0 < \varepsilon_n < 1/(8n)$.

You may find other proofs in Feller [8, p. 52] or [6, 7] or Ash [3, pp.43–45], or Diaconis and Freedman [5], or the exercises in Pitman [12, p. 136]. Flajolet and Sedgewick [9] offer five different proofs. I thank Jim Tao for this last reference.

Bibliography

[1] T. M. Apostol. 1967. *Calculus, Volume I: One-variable calculus with an introduction to linear algebra*, 2d. ed. New York: John Wiley & Sons.

[2] ——— . 1969. *Calculus, Volume II: Mulit-variable calculus and linear algebra, with applications to differential euations and probability*, 2d. ed. Waltham, Massachusetts: Blaisdell.

[3] R. B. Ash. 2008. *Basic probability theory*. Mineola, New York: Dover. Reprint of the 1970 edition published by John Wiley and Sons.

[4] G. Darmois. 1928. *Statistique mathématique*. Number 27 in Encyclopédie scientifique: Bibliothèque de mathématiques appliquées. Paris: G. Doin. http://books.google.com/books?id=E781AQAAIAAJ

[5] P. Diaconis and D. Freedman. 1986. An elementary proof of Stirling's formula. *American Mathematical Monthly* 93(2):123–125. http://www.jstor.org/stable/2322709

[6] W. Feller. 1967. A direct proof of Stirling's formula. *American Mathematical Monthly* 74(10):1223–1225. http://www.jstor.org/stable/2315671

[7] ——— . 1968. Correction to "A direct proof of Stirling's formula". *American Mathematical Monthly* 75(5):518. http://www.jstor.org/stable/2314719

[8] ——— . 1968. *An introduction to probability theory and its applications*, 3d. ed., volume 1. New York: Wiley.

[9] P. Flajolet and R. Sedgewick. 2009. *Analytic combinatorics*. Cambridge: Cambridge University Press. http://ac.cs.princeton.edu/home/AC.pdf

[10] L. M. Le Cam. 1986. The central limit theorem around 1935. *Statistical Science* 1(1):78–91. DOI: 10.1214/ss/1177013818

[11] K. Pearson. 1924. Historical note on the origin of the normal curve of errors. *Biometrika* 16(3/4):402–404. http://www.jstor.org/stable/2331714

[12] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[13] H. Robbins. 1955. A remark on Stirling's formula. *American Mathematical Monthly* 62(1):26–29. http://www.jstor.org/stable/2308012