Stirling's Formula with Error Bounds

Jerry Alan Veeh

September 16, 2014

Contents

1 Introduction

The objective here is to provide a self-contained proof of Stirling's approximation

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

and also to examine refinements which provide the bounds

$$n^n e^{-n}\sqrt{2\pi\left(n+\frac{1}{6}\right)} \le n! \le n^n e^{-n}\sqrt{2\pi\left(n+\frac{1}{5}\right)}$$

for $n \ge 1$. Bounds of this form are often attributed to R. W. Gosper in his paper *Decision Procedure for Indefinite Hypergeometric Summation* that appeared in *Proceedings of the National Academy of Sciences* volume 75, number 1, pages 40-42, 1978. But the bounds do not appear there.

Any derivation of Stirling's approximation can not be completely elementary since the factor of π arises for non-elementary reasons. The approach taken here makes use of the gamma function, whose definition and basic properties are useful in other contexts. The approximation will then be proved using Laplace's method.

2 The Gamma Function

The gamma function is defined for $x > 0$ by the formula

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt.$$

There are two potential problems with the integral in this definition.

First, for $0 < x < 1$, $\lim_{t \downarrow 0} t^{x-1} e^{-t} = \infty$ and there is the potential for the integral to fail to converge near the lower endpoint of integration. Fortunately simple estimation gives

$$\int_0^1 t^{x-1} e^{-t}\, dt \le \int_0^1 t^{x-1}\, dt = 1/x < \infty$$

since $x > 0$ and this potential problem dissolves.

The second concern is about integration of a positive function over an infinite interval. Intuitively, the exponential function decays for large values of t so much faster than the power function grows that the exponential function should cause convergence. For $0 < x \le 1$ the simple bound $t^{x-1} e^{-t} \le e^{-t}$ for $t \ge 1$ confirms this intuition. If $x > 1$, write $e^{-t} = e^{-t/2} e^{-t/2}$ and compute the maximum value of $A(t) = t^{x-1} e^{-t/2}$ for $t > 0$. Simple calculus shows that $A'(t) = 0$ only at $t = 2x - 2 > 0$ and so the maximum value of $A(t)$ is $A(2x-2)$. Thus

$$\int_0^\infty t^{x-1} e^{-t}\, dt \le \int_0^\infty A(2x-2) e^{-t/2}\, dt = 2 \cdot A(2x-2) < \infty$$

and the second concern vanishes too.

The key property of the gamma function used here is

$$\Gamma(x+1) = x\Gamma(x)$$

for any $x > 0$. This property is established by a simple integration by parts:

$$\begin{aligned} \Gamma(x+1) &= \int_0^\infty t^x e^{-t}\, dt \\ &= -t^x e^{-t} \Big|_{t=0}^{\infty} + \int_0^\infty x t^{x-1} e^{-t}\, dt \\ &= x\Gamma(x). \end{aligned}$$

Since $\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1$ by direct evaluation, $\Gamma(2) = \Gamma(1+1) = 1\Gamma(1) = 1$, $\Gamma(3) = \Gamma(2+1) = 2\Gamma(2) = 2$, and a simple induction argument gives $\Gamma(n+1) = n!$. More explicitly,

$$n! = \int_0^\infty t^n e^{-t}\, dt$$

and this will be the starting point for the derivation of Stirling's approximation.

There is one additional fact about the gamma function that is needed here, but which will be developed in greater generality than necessary. Making the change of variable $x^2 = t$ in the definition of the gamma function gives

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}\, dt = 2\int_0^\infty x^{2a-1} e^{-x^2}\, dx.$$

Using a similar expression for $\Gamma(b)$ but integrating with respect to a variable y and then multiplying gives

$$\Gamma(a)\Gamma(b) = 4\int_0^\infty \int_0^\infty x^{2a-1} y^{2b-1} e^{-(x^2+y^2)}\, dxdy.$$

Now change to polar coordinates $x = r\cos\theta$ and $y = r\sin\theta$ and recongnize the integral with respect to r as a gamma function to obtain

$$\Gamma(a)\Gamma(b) = 2\Gamma(a+b)\int_0^{\pi/2} \cos^{2a-1}\theta \sin^{2b-1}\theta\, d\theta.$$

Of particular interest here will be $\Gamma(1/2) = 2\int_0^\infty e^{-x^2}\,dx = \int_{-\infty}^\infty e^{-x^2}\,dx$ since the integrand is symmetric about the origin. Substituting $a = b = 1/2$ yields

$$\Gamma(1/2)\Gamma(1/2) = 2\Gamma(1)\int_0^{\pi/2} d\theta = \pi$$

from which $\Gamma(1/2) = \int_{-\infty}^\infty e^{-x^2}\,dx = \sqrt{\pi}$.

As a final note, the change of variable $x = \cos^2\theta$ in the integral on the right just above produces

$$\Gamma(a)\Gamma(b) = \Gamma(a+b)\int_0^1 x^{a-1}(1-x)^{b-1}\,dx.$$

The integral on the right defines the beta function.

3 Taylor's Expansion with Error Estimate, Concave and Convex Functions

For a function f with continuous derivative, the Fundamental Theorem of Calculus gives

$$f(x) = f(a) + \int_a^x f'(t)\,dt.$$

If f' also has a continuous derivative, $f'(t) = f'(a) + \int_a^t f''(u)\,du$, also by the Fundamental Theorem. Substituting this expression for f' into the integral above and simplifying gives

$$f(x) = f(a) + f'(a)(x-a) + \int_a^x \int_a^t f''(u)\,du\,dt.$$

Now if $m \le f''(u) \le M$,

$$f(a) + f'(a)(x-a) + m(x-a)^2/2 \le f(x) \le f(a) + f'(a)(x-a) + M(x-a)^2/2$$

by evaluating the iterated integrals. This type of estimate will be used below.

A function f is convex if $f''(x) \ge 0$ for all x. (Some calculus books say such an f is concave up.) Using the left side of the previous inequality with $m = 0$ gives

$$f(a) + f'(a)(x-a) \le f(x)$$

where $f(a)+f'(a)(x-a)$ is the tangent line to f at $x=a$. So a convex function always lies above any of its tangent lines. A function f is concave if $f''(x) \leq 0$ for all x. (Some calculus books say such an f is concave down.) Using the right side of the previous inequality with $M=0$ gives

$$f(x) \leq f(a)+f'(a)(x-a)$$

so a concave function always lies below any of its tangent lines. This fact will be used below.

The preceding is all that will be required to establish Stirling's approximation, but the upper and lower bounds will require 4 additional estimates.

The first two estimates come from a straightforward application of the Fundamental Theorem. For $x>0$,

$$\begin{aligned} \ln(1+x) &= \ln(1+0)+\int_0^x \frac{1}{1+t}\,dt \\ &\leq \int_0^x \frac{1}{1+0}\,dt \\ &= x \end{aligned}$$

and similarly, for $x>0$,

$$\begin{aligned} \ln(1+x) &= \ln(1+0)+\int_0^x \frac{1}{1+t}\,dt \\ &\geq \int_0^x \frac{1}{1+x}\,dt \\ &= \frac{x}{1+x}. \end{aligned}$$

These are the first two estimates.

The next two estimates require more care. From the geometric series formula $\frac{1}{1-r} = 1+r+r^2+r^3+\cdots$, term by term integration with respect to r from $r=0$ to $r=x$ gives, for $0<x<1$,

$$-\ln(1-x) = x+x^2/2+x^3/3+\cdots.$$

Similarly, integrating the geometric series $\frac{1}{1+r} = 1-r+r^2-r^3+\cdots$ over the same interval of r values gives, for $0<x<1$

$$\ln(1+x) = x-x^2/2+x^3/3-x^4/4+\cdots.$$

Since $\ln\left(\frac{1-x}{1+x}\right) = \ln(1-x)-\ln(1+x)$, substitution gives

$$\ln\left(\frac{1-x}{1+x}\right) = -2x-2x^3/3-2x^5/5-\cdots.$$

Finally,

$$\frac{1}{2x}\ln\left(\frac{1-x}{1+x}\right) = -1-x^2/3-x^4/5-x^6/7-\cdots.$$

The last two estimates needed below are upper and lower bounds on the left side of this equation. Now subtracting less only makes the sum larger, so

$$\begin{aligned}
\frac{1}{2x}\ln\left(\frac{1-x}{1+x}\right) &= -1-x^2/3-x^4/5-x^6/7-\cdots \\
&\leq -1-x^2/3-\left(x^2/3\right)^2-\left(x^2/3\right)^3-\cdots \\
&= -1-\frac{x^2/3}{1-x^2/3} \\
&= -1-\frac{1}{3x^{-2}-1}
\end{aligned}$$

by applying the geometric series formula. A lower bound follows similarly by

$$\begin{aligned}
\frac{1}{2x}\ln\left(\frac{1-x}{1+x}\right) &= -1-x^2/3-x^4/5-x^6/7-\cdots \\
&\geq -1-x^2/3-x^4/3-x^6/3-\cdots \\
&= -1-\frac{1}{3}\frac{x^2}{1-x^2} \\
&= -1-\frac{1}{3(x^{-2}-1)}.
\end{aligned}$$

This completes the derivation of the 4 estimates.

4 Laplace's Method

Laplace is credited with the following method of estimating the value of integrals depending on a large parameter. In the present case

$$n! = \int_0^\infty t^n e^{-t}\,dt = \int_0^\infty e^{n\ln t - t}\,dt$$

and n is the large parameter. Simple calculus shows that $L(t) = n\ln t - t$ has $L'(t) = n/t - 1$ and $L''(t) = -n/t^2 < 0$. Thus L has a maximum at $t = n$ and is concave.

Laplace's method is motivated by the Taylor series expansion of L to the quadratic term: $L(t) = L(n) + L'(n)(t-n) + L''(n)(t-n)^2/2 + \cdots = n\ln n - n - (t-n)^2/2n + \cdots$. From this expansion the function $e^{L(t)}$ decays very rapidly as $|t-n|$ increases. On this basis, most of the value of the integral should be captured in a small interval around $t = n$. The bulk of the work consists of tightening up this informal reasoning.

Since $L''(t) = -n/t^2$, the interval around $t = n$ will be selected to be of the form $rn \leq t \leq n/r$, for some $0 < r < 1$ to be specified later. Then on this interval $-1/nr^2 \leq L''(t) \leq -r^2/n$ and the Taylor expansion with error estimate gives

$$n\ln n - n - (t-n)^2/2nr^2 \leq L(t) \leq n\ln n - n - r^2(t-n)^2/2n.$$

So

$$\int_{rn}^{n/r} e^{L(t)}\,dt \quad\leq\quad \int_{rn}^{n/r} e^{n\ln n - n - r^2(t-n)^2/2n}\,dt$$

$$
\begin{aligned}
&= n^n e^{-n} \int_{rn}^{n/r} e^{-r^2(t-n)^2/2n}\, dt \\
&= n^n e^{-n} \sqrt{2n} \frac{1}{r} \int_{\sqrt{n/2}r(r-1)}^{\sqrt{n/2}r(1/r-1)} e^{-x^2}\, dx
\end{aligned}
$$

after making the substitution $x = r(t-n)/\sqrt{2n}$. Similarly,

$$
\begin{aligned}
\int_{rn}^{n/r} e^{L(t)}\, dt &\geq \int_{rn}^{n/r} e^{n\ln n - n - (t-n)^2/2nr^2}\, dt \\
&= n^n e^{-n} \sqrt{2n} r \int_{\sqrt{n/2}(r-1)/r}^{\sqrt{n/2}(1/r-1)/r} e^{-x^2}\, dx
\end{aligned}
$$

after making the substitution $x = (t-n)/r\sqrt{2n}$. These estimates have also exposed all of the key components of Stirling's approximation.

To conclude Laplace's method

$$
\begin{aligned}
\frac{n!}{n^n e^{-n}\sqrt{2n}} &= \frac{e^{-L(n)}}{\sqrt{2n}} \int_0^{\infty} e^{L(t)}\, dt \\
&= \frac{e^{-L(n)}}{\sqrt{2n}} \int_0^{rn} e^{L(t)}\, dt \\
&\quad + \frac{e^{-L(n)}}{\sqrt{2n}} \int_{rn}^{n/r} e^{L(t)}\, dt \\
&\quad + \frac{e^{-L(n)}}{\sqrt{2n}} \int_{n/r}^{\infty} e^{L(t)}\, dt.
\end{aligned}
$$

To estimate the first and third integral, use the fact that L is concave and thus lies below any tangent line. For the first integral, use the tangent line at $t = rn$ to obtain

$$
\begin{aligned}
\frac{e^{-L(n)}}{\sqrt{2n}} \int_0^{rn} e^{L(t)}\, dt &\leq \frac{e^{-L(n)}}{\sqrt{2n}} \int_0^{rn} e^{L(rn)+L'(rn)(t-rn)}\, dt \\
&= \frac{e^{L(rn)-L(n)}}{\sqrt{2n}} \int_0^{rn} e^{L'(rn)(t-rn)}\, dt \\
&= \frac{e^{L(rn)-L(n)}}{L'(rn)\sqrt{2n}} \left(1 - e^{-L'(rn)rn}\right) \\
&\leq \frac{1}{L'(rn)\sqrt{2n}} \\
&\to 0
\end{aligned}
$$

where the fact that $L(n)$ is the maximum value of L has been used to see that $L(rn) - L(n) < 0$ and $L'(rn) = 1/r - 1 > 0$. The third integral is estimated using the tangent line at $t = n/r$ in a similar way:

$$
\frac{e^{-L(n)}}{\sqrt{2n}} \int_{n/r}^{\infty} e^{L(t)}\, dt \leq \frac{e^{-L(n)}}{\sqrt{2n}} \int_{n/r}^{\infty} e^{L(n/r)+L'(n/r)(t-n/r)}\, dt
$$

$$
\begin{aligned}
&= \frac{e^{L(n/r)-L(n)}}{\sqrt{2n}} \int_{n/r}^{\infty} e^{L'(n/r)(t-n/r)}\,dt \\
&= -\frac{e^{L(n/r)-L(n)}}{L'(n/r)\sqrt{2n}} \\
&\to 0
\end{aligned}
$$

where again the fact that $L(n)$ is the maximum value of L has been used to see that $L(n) > L(n/r)$ and the fact that $L'(n/r) = r - 1 < 0$ has been used to evaluate the integral. Using these limits and the earlier estimates of the second integral shows that all limit points of $\frac{n!}{n^n e^{-n}\sqrt{2n}}$ as $n \to \infty$ lie between r and $1/r$ times the integral $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$. Letting $r \uparrow 1$ shows that

$$
\lim_{n\to\infty} \frac{n!}{n^n e^{-n}\sqrt{2n}} = \sqrt{\pi}
$$

completing the proof of Stirling's approximation.

5 Some Improvements

The asymptotics provided by Stirling's approximation can be replaced by lower and upper bounds that are reasonably tight with a little additonal effort. A lower bound will be established first.

For some $0 < a < 1$ define

$$
A_n = \frac{n!}{n^n e^{-n}\sqrt{2\pi(n+a)}}.
$$

By Stirling's approximation, $A_n \to 1$ as $n \to \infty$. Now if in fact for an appropriate choice of a, $A_n \downarrow 1$ then $A_n \geq 1$ for $n \geq 1$ and

$$
n! \geq n^n e^{-n}\sqrt{2\pi(n+a)}
$$

provides a lower bound for $n \geq 1$. This monotone convergence will be established if $A_{n+1}/A_n \leq 1$ for $n \geq 1$, or equivalently, if $\ln A_{n+1} - \ln A_n \leq 0$.

Simple computation shows that for $n \geq 1$,

$$
\ln A_{n+1} - \ln A_n = \left(n+\frac{1}{2}\right)\ln\left(\frac{n}{n+1}\right) + 1 + \frac{1}{2}\ln\left(\frac{n+1}{n}\cdot\frac{n+a}{n+1+a}\right).
$$

Now the first factor on the right can be written as

$$
\left(n+\frac{1}{2}\right)\ln\left(\frac{n}{n+1}\right) = \frac{2n+1}{2}\ln\left(\frac{1-\frac{1}{2n+1}}{1+\frac{1}{2n+1}}\right)
$$

which is of the form $\frac{1}{2x}\ln\left(\frac{1-x}{1+x}\right)$ with $x = 1/(2n+1)$. So using one of the four estimates found above gives

$$
\left(n+\frac{1}{2}\right)\ln\left(\frac{n}{n+1}\right) = \frac{2n+1}{2}\ln\left(\frac{1-\frac{1}{2n+1}}{1+\frac{1}{2n+1}}\right)
$$

$$
\begin{aligned}
&\leq -1 - \frac{1}{3(1/(2n+1))^{-2}-1} \\
&= -1 - \frac{1}{3(2n+1)^2-1}.
\end{aligned}
$$

Similarly, the third factor above transforms to

$$
\begin{aligned}
\frac{1}{2}\ln\left(\frac{n+1}{n}\frac{n+a}{n+1+a}\right) &= \frac{1}{2}\ln\left(\frac{1+\frac{a}{n}}{1+\frac{a}{n+1}}\right) \\
&= \frac{1}{2}\ln\left(\frac{1+\frac{a}{n+1}+\frac{a}{n}-\frac{a}{n+1}}{1+\frac{a}{n+1}}\right) \\
&= \frac{1}{2}\ln\left(1+\frac{a}{n(n+1)+na}\right)
\end{aligned}
$$

from which, using the earlier estimate $\ln(1+x) \leq x$ gives

$$\frac{1}{2}\ln\left(\frac{n+1}{n}\frac{n+a}{n+1+a}\right) \leq \frac{1}{2}\frac{a}{n(n+1)+na}.$$

Using these two inequalites with the first expression above gives

$$\ln A_{n+1} - \ln A_n \leq -1 - \frac{1}{3(2n+1)^2-1} + 1 + \frac{1}{2}\frac{a}{n(n+1)+na}$$

so that the difference of logs is non-positive provided a is chosen so that

$$-\frac{1}{3(2n+1)^2-1} + \frac{1}{2}\frac{a}{n(n+1)+na} \leq 0$$

for all $n \geq 1$. Rearrangment of this inequality gives the requirement that

$$a \leq \frac{n^2+n}{6n^2+5n+1}$$

for $n \geq 1$. The fraction on the right is clearly near $1/6$. In fact, $6(n^2+n) = 6n^2+5n+n \geq 6n^2+5n+1$ for $n \geq 1$, so the right side is at least $1/6$ for all $n \geq 1$ and equality holds when $n = 1$. Thus the sequence A_n will be non-increasing as long as $a \leq 1/6$. This proves that

$$n! \geq n^n e^{-n}\sqrt{2\pi\left(n+\frac{1}{6}\right)}$$

for $n \geq 1$.

A similar approach can be used to find an upper bound of the same form. Define

$$B_n = \frac{n!}{n^n e^{-n}\sqrt{2\pi(n+b)}}$$

for some $0 < b < 1$ and seek to find the smallest b for which $B_n \uparrow 1$. This will be the case provided $\ln B_{n+1} - \ln B_n \geq 0$. Using the same algebraic manipulations as earlier gives

$$\begin{aligned}
\ln B_{n+1} - \ln B_n &= \left(n+\frac{1}{2}\right)\ln\left(\frac{n}{n+1}\right)+1+\frac{1}{2}\ln\left(\frac{n+1}{n}\frac{n+b}{n+1+b}\right) \\
&= \frac{2n+1}{2}\ln\left(\frac{1-\frac{1}{2n+1}}{1+\frac{1}{2n+1}}\right)+1+\frac{1}{2}\ln\left(1+\frac{b}{n(n+1)+nb}\right).
\end{aligned}$$

Now using the lower bound on the first term found earlier gives

$$\begin{aligned}
\frac{2n+1}{2}\ln\left(\frac{1-\frac{1}{2n+1}}{1+\frac{1}{2n+1}}\right) &\geq -1-\frac{1}{3(1/(2n+1))^{-2}-1)} \\
&= -1-\frac{1}{12n(n+1)}.
\end{aligned}$$

Similarly, using the bound $\ln(1+x) \geq x/(1+x)$ found earlier on the third term gives

$$\begin{aligned}
\frac{1}{2}\ln\left(\frac{n+1}{n}\frac{n+b}{n+1+b}\right) &\geq \frac{1}{2}\frac{\frac{b}{n(n+1)+nb}}{1+\frac{b}{n(n+1)+nb}} \\
&= \frac{1}{2}\cdot\frac{b}{n(n+1)+(n+1)b} \\
&= \frac{1}{2}\cdot\frac{b}{(n+1)(n+b)}.
\end{aligned}$$

Substitution yields

$$\ln B_{n+1} - \ln B_n \geq -\frac{1}{12n(n+1)}+\frac{1}{2}\frac{b}{(n+1)(n+b)}$$

so that B_n is non-decreasing if the right side of this inequality is non-negative. Simplification reduces this to the requirement that

$$b \geq \frac{n}{6n-1}.$$

Clearly the fraction on the right takes is largest value, $1/5$, when $n = 1$. Thus B_n is increasing as long as $b \geq 1/5$, which implies the inequality

$$n! \leq n^n e^{-n}\sqrt{2\pi\left(n+\frac{1}{5}\right)}$$

for $n \geq 1$.

A moments reflection shows that what has actually been proved is that for $n \geq k$,

$$n! \leq n^n e^{-n}\sqrt{2\pi\left(n+\frac{k}{6k-1}\right)}.$$

By choosing $k = 3$ and manually checking the inequality for $n = 1$ and $n = 2$,

$$n! \leq n^n e^{-n} \sqrt{2\pi \left(n + \frac{3}{17} \right)}.$$

for $n \geq 1$.

Also, since $\frac{k}{6k-1} \to \frac{1}{6}$ as $k \to \infty$, the lower bound can not be improved. This suggests that the lower bound provides a much better approximation to $n!$ than Stirling's approximation while being no harder to compute.

Also note that the upper bound on $n!$ also holds for $n = 0$, by direct computation. A lower bound on $n!$ for $n \geq 0$ can be obtained by replacing $1/6$ by $1/2\pi$ in the expression for the lower bound obtained here.