Factorials and Stirling's Approximation

Mathias Winther Madsen

January 15, 2015

The **factorials** are the numbers

$$n! \;=\; 1 \cdot 2 \cdot 3 \cdots n.$$

For instance, "four factorial" is $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$. By convention, $0! = 1$.

The factorials have a number of combinatorical interpretations. They count, for instance, the number of permutations of n objects, such as

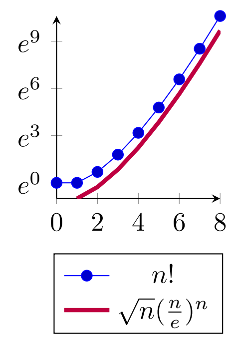
$$123, 132, 213, 231, 312, 321$$

for $3! = 6$. They also count the number of ways you can put n objects into n boxes such that each box contains exactly one object.

The factorials can be approximated by **Stirling's approximation**,

$$n! \;\approx\; C \times \sqrt{n} \left(\frac{n}{e}\right)^n,$$

where e is the base of the natural logarithm and C is a constant whose exact value we'll worry about later.



This approximation is the topic of this note. We will see why it holds, how good it is, and what inequalities it gives rise to. In particular, we will show that when $C \approx 2.507$, this approximation provides a lower bound, while when $C \approx 2.719$, it provides an upper bound. We arrive at this result after a series of approximations of increasing quality.

1 A Lower Bound

We first prove a weak version of Stirling's approximation which does not contain the square root term.

Theorem 1.

$$e\left(\frac{n}{e}\right)^n \;\leq\; n!$$

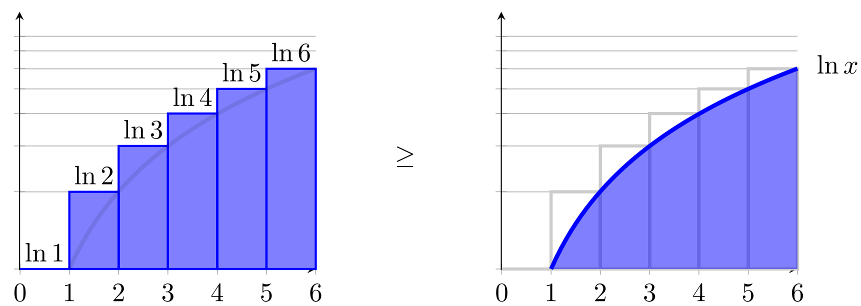
Proof. On a logarithmic scale, the factorial is a sum rather than a product:

$$\ln(n!) \;= \ln 1 + \ln 2 + \ln 3 + \cdots + \ln n.$$

We will approximate this sum by a continuous integral:

$$\sum_{x=1}^{n} \ln x \;\approx\; \int_0^n \ln x \, dx.$$

Consider therefore a series of boxes of width 1 and height $\ln x$. Since the logarithm is an increasing function, these boxes completely cover the area under the graph of $\ln x$:



The area of the first box is 0, so we only have to worry about the area under the graph from $x = 1$ onwards. In algebraic terms, this means that

$$\sum_{x=1}^{n} \ln x \;=\; \sum_{x=2}^{n} \ln x \;\geq\; \int_1^n \ln x \, dx \;=\; n \ln n - n + 1.$$

By exponentiating, this proves the bound $e(n/e)^n \leq n!$. $\square$

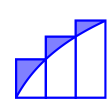
2 An Upper Bound

In the proof of the previous theorem, squeezed a smooth curve underneath a staircase function. We will now add a correction term accounts for the gap between these two curves.

Theorem 2.

$$n! \;\leq\; \sqrt{n}e\left(\frac{n}{e}\right)^n.$$

Proof. As we saw in the proof of the previous theorem, the value of $\ln(n!)$ can be represented as the area under a staircase function s, and the graph of ln is separated from this staircase by a series of roughly triangular figures. We will now approximate the area of these figures by pretending they are perfectly triangular.

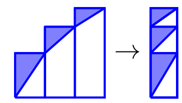


The height of the kth such triangle is $\ln(k+1) - \ln(k)$. The total area of all the triangles is thus

$$\frac{\ln 2 - \ln 1}{2} + \frac{\ln 3 - \ln 2}{2} + \frac{\ln 4 - \ln 3}{2} + \cdots + \frac{\ln n - \ln(n-1)}{2}.$$

This is a telescoping sum, and almost all its terms cancel, leaving only

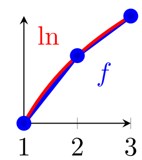
$$\frac{\ln n}{2}.$$



We could also have seen this directly by stacking the all triangles on top of each other in order to obtain a single stack of height $\ln n$.

We now have the following improved approximation of $n!$:

$$\ln(n!) \;\approx\; \frac{\ln n}{2} + n\ln n - n + 1.$$



This is an overstimate. The natural logarithm is a strictly concave function ($\ln'' < 0$), so any straight line between two points on its graph thus lie strictly below it. This means that our triangles overlap slightly with the area under the graph, and we have thus added a tiny bit too much area when we added the corrective term $1/2 \ln n$. Hence,

$$\ln(n!) \;\leq\; \frac{\ln n}{2} + n\ln n - n + 1,$$

and the upper bound $n! \leq \sqrt{n}e(n/e)^n$ follows by exponentiation. □

3 An Improved Lower Bound

The previous theorem established that $n! \leq e\sqrt{n}(n/e)^n$. We will now show that there is a lower bound that only differs by a constant from this upper bound. Proving this will require us to show consider the limit behavior of the error committed by the upper bound.

Theorem 3.

$$\frac{e^{11/12}}{4\sqrt{2}} \times \sqrt{n}\left(\frac{n}{e}\right)^n \;\leq\; n!$$

Proof. In the previous theorem, we estimated the size of the box $\ln(k+1)$ by an integral plus a triangle,

$$\ln(k+1) \;\approx\; \int_k^{k+1} \ln(x)\,dx + \frac{\ln(k+1) - \ln(k)}{2},$$

We will now upper-bound the error of this approximation,

$$b(k) \;=\; \int_k^{k+1} \ln(x)\,dx + \frac{\ln(k+1) - \ln(k)}{2} - \ln(k+1).$$

By subtracting this upper bound from $\ln(k+1)$, we can then produce an improved lower bound on $\ln(k+1)$ and thus on $\ln(n!)$.

Fortunately, we can compute the exact value of $b(k)$. Since

$$\int_k^{k+1} \ln(x)\,dx \;=\; (k+1)\ln(k+1) - k\ln k - 1,$$

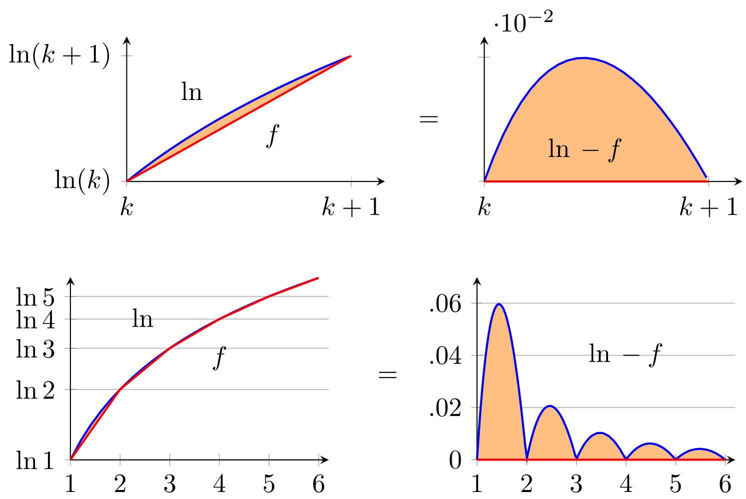
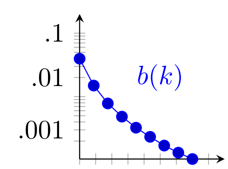


Figure 1: Graphical interpretation of the error $b(k)$, with $k = 1, 2, 3, \ldots, n-1$.



a bit of fearless algebra shows that

$$b(k) \;=\; \left(k + \frac{1}{2}\right) \ln\left(1 + \frac{1}{k}\right) - 1.$$

We could continue to work with this exact error term, but it is a bit clunky. We will therefore use a polynomial approximation of it instead.

To do so, we first diferentiate the function $\ln(1 + x)$ over and over in order to find its Taylor expansion around the point $x = 0$:

$$\ln(1 + x) \;=\; x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \frac{1}{5}x^5 - \frac{1}{6}x^6 + \cdots$$

Setting $k = 1/x$ and writing $b(k)$ as

$$b\,(1/x) \;=\; \frac{1}{x}\ln(1 + x) + \frac{1}{2}\ln(1 + x) - 1,$$

we thus see that the Taylor expansion of b is

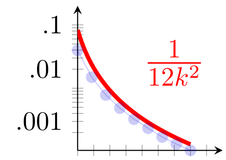
$$b\,(1/x) \;=\; \frac{1}{12}x^2 - \frac{1}{12}x^3 + \frac{3}{40}x^4 - \frac{1}{15}x^5 + \cdots$$

The fact that this series has no terms of order zero and one agrees with the fact that it expresses the error committed by a linear approximation.

Since $k \geq 1$, we are only interested in the behavior of this function on the unit interval. The third term in this series consequently always falls somewhere between 0 and $-1/12$. By Taylor's theorem, the same bounds hold for the entire tail of this expansion for $0 \leq x \leq 1$. This gives us the upper bound

$$b(k) \;\leq\; \frac{1}{12k^2}.$$

This bound could be strenghthened arbitrarily by including a larger number of terms, but this will do for our purposes.



Using this upper bound for the individual error terms, we can now estimate the total error by an integral:

$$\sum_{k=1}^{n-1} b(k) \;\leq\; \sum_{k=1}^{n-1} \frac{1}{12k^2} \;\leq\; \int_0^{n-1} \frac{1}{12x^2}\,dx.$$

However, this integral is divergent between 0 and 1, so we need to compute the exact value of the first term explicitly:

$$\sum_{k=1}^{n-1} b(k) \;\leq\; b(1) + \int_1^{n-1} \frac{1}{12x^2}\,dx \;=\; b(1) + \frac{1}{12} - \frac{1}{12(n-1)}.$$

This converges to $b(1) - 1/12$ for $n \to \infty$, and $b(1) = {}^3\!/_2 \ln 2 - 1$. The total error is thus bounded by the constant

$$\sum_{k=1}^{n-1} b(k) \;\leq\; \frac{3}{2}\ln 2 - \frac{11}{12} \;\approx 0.123.$$

It follows that our upper bound for $\ln(n!)$ overshoots its target by at most ${}^3\!/_2 \ln 2 - {}^{11}\!/_{12}$, and the estimate $e \times \sqrt{n}(n/e)^n$ similarly too large by a factor of at most

$$\exp\left(\sum_{k=1}^{n-1} b(k)\right) \;\leq\; \frac{2^{3/2}}{e^{11/12}} \;=\; \frac{4\sqrt{2}}{e^{11/12}} \;\approx\; 1.131.$$

By dividing by this number, we turn the upper bound $n! \leq e \times \sqrt{n}(n/e)^n$ into the lower bound

$$\frac{e^{23/12}}{4\sqrt{2}} \times \sqrt{n}\left(\frac{n}{e}\right)^n \;\leq\; n!$$

which is roughly $2.404 \times \sqrt{n}(n/e)^n$. □

4 Pushing the Envelope

The previous theorem corrected the error committed by the approximation

$$\ln(n!) \;\approx\; n\ln n - n + 1 - \frac{1}{2}\ln n$$

by subtracting the sum $\sum_{k=1}^{\infty} b(k)$. This sum, in turn, was estimated by splitting it up into one exact computation and a tail approximation:

$$\sum_{k=1}^{\infty} b(k) \;\leq\; b(1) + \int_m^{\infty} b(x)\,dx.$$

However, since we are splitting up the sum anyway, we might as well let the exact part to eat a bit more into the approximate one:

$$\sum_{k=1}^{\infty} b(k) \;\le\; \sum_{k=1}^{m} b(k) + \int_{m}^{\infty} b(x)\,dx.$$

This way, we compute the exact value of the partial sum $b(1)+b(2)+\cdots+b(m)$ and only use the approximation to bound the sum of the remaining terms. The higher we set the cut-off point m, the accurately we estimate the sum.

A few approximations of higer accuracy, where the tail integrals were computed with $b(k) \le 1/(12k^2)$, are given in the following table:

m	1	2	3	4	5	6	7	$\cdots$	100
$\sum_{k=1}^{m} b(k)$	.040	.054	.061	.065	.068	.070	.071	$\cdots$	.081
$\int_{m}^{\infty} b(x)\,dx$	.083	.042	.028	.021	.017	.014	.012	$\cdots$	.001
$\sim \sum_{k=1}^{\infty} b(k)$	.123	.095	.088	.086	.083	.083	.083	$\cdots$	.081

As this table suggests, the total area of all the gaps between f and ln converge to a number equal to about 0.081. Since $e/e^{0.081} \approx 2.507$, it follows that $n!$ is always squeezed inside the sandwich

$$2.507 \times \sqrt{n}\left(\frac{n}{e}\right)^n \;\le\; n! \;\le\; 2.718 \times \sqrt{n}\left(\frac{n}{e}\right)^n.$$

The larger of these constants cannot be pushed lower than e without spoiling the inequality for $n = 1$. The smaller of these constants cannot be pushed higher than $\sqrt{2\pi} \approx 2.507$, which is the limiting value of the ratio between $n!$ and $\sqrt{n}(n/e)^n$. Showing that this is indeed the best possible value for the lower bound involves a number of additional concepts, and we are not going to prove it here.

Finally, we could of course use approximations of more complicated forms instead of insisting on using $\sqrt{n}(n/e)^n$ times a constant. If we are willing to make life a bit more complicated this was, we can achieve arbitrarily good approximations to $n!$ by additional polynomial corrections, such as

$$n! \;\approx\; \sqrt{2\pi}\left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \cdots\right) \times \sqrt{n}\left(\frac{n}{e}\right)^n.$$

For our purposes, however, such additional corrections will not be necessary.