Stirling's formula!

N. David Mermin
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853

(Received 15 April 1983; accepted for publication 18 May 1983)

Stirling's approximation to $n!$ and other estimates, some cruder, some more refined, are developed along surprisingly elementary lines. The aim is to shed some light on why these approximations work so well, for students using them to study entropy and irreversibility in such simple statistical models as might be examined in a general education physics course. At the heart of the argument is an elementary and useful representation of *all* the corrections to Stirling's approximation, with which even many experts seem to be unacquainted.

Stirling's approximation to the factorial function,

$$n! \sim \sqrt{2\pi n}(n/e)^n, \tag{1}$$

plays a central role in any number of investigations of statistical physics, and is invaluable in the kinds of simple probabilistic studies that can convey to students in a general education course the nature of entropy and irreversibility. Unfortunately, the usual derivations of (1) are inaccessible to such students and even to many beginning physics majors. One can, of course, simply verify its remarkably accurate performance,[1] but the better students are bound to find this frustrating: Why is it that Stirling's formula works as well as it does?

I provide here an elementary answer to this question that can be adapted to give convincing explanations at a range of levels of mathematical innocence. For the crudest argument it is only necessary to know the elementary definition of the number e that arises in the theory of compound interest. Students who also know that the natural logarithm has the expansion

$$\ln(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \ldots \tag{2}$$

can be given a really intimate glimpse into the workings of Stiring's formula, while those who are willing to approximate a few simple sums by integrals can acquire a level of understanding possessed, I suspect, by few professionals.

Stiring's formula begins to yield up its secrets with the observation[2] than $n!$ can evidently be written in the form

$$n! = (\tfrac{1}{2})(\tfrac{2}{3})^2(\tfrac{3}{4})^3(\tfrac{4}{5})^4 \cdots [(n-1)/n]^{n-1} n^n. \tag{3}$$

This can be written in the equivalent form,

$$n! = n^n \Big/ \prod_{j=1}^{n-1}\left(1+\frac{1}{j}\right)^j, \tag{4}$$

which immediately calls to mind the definition of e as the limiting value for large j of $(1+1/j)^j$:

$$e \sim (1+1/j)^j. \tag{5}$$

The denominator of (4) is thus a product of successively better and better approximants to e. If we were to estimate each term in the product by its limiting value e, we would extract from (4) the estimate

$$n! \sim n^n/e^{n-1}. \tag{6}$$

For many purposes (6) is as good as one needs.[3] One can, however, do very much better if one realizes that although $(1+1/j)^j$ is indeed a good approximation to e for large j, the refinement

$$e \sim (1+1/j)^{j+1/2} \tag{7}$$

does spectacularly better. The superiority of (7) to (5) is evident from Table I, and is easy to understand with the aid of the expansion (2). We have, on the one hand, the large j expansion

$$\begin{aligned}(1+1/j)^j/e &= \exp[j\ln(1+1/j) - 1] \\ &= \exp[j(1/j - 1/2j^2 + 1/3j^3 - \ldots) - 1] \\ &\sim e^{-1/2j},\end{aligned} \tag{8}$$

while, on the other hand, the ratio of the estimate (7) to e is given by

$$\begin{aligned}(1+1/j)^{j+1/2}/e &= \exp[(j+\tfrac{1}{2})\ln(1+1/j) - 1] \\ &= \exp[(j+\tfrac{1}{2})(1/j - 1/2j^2 + 1/3j^3 - \ldots) - 1] \\ &\sim e^{1/12j^2}.\end{aligned} \tag{9}$$

By stepping up the exponent by $\frac{1}{2}$, we have brought about the vanishing of the correction term in $1/j$, thereby reducing the error to one of order $1/j^2$ (with the agreeable coefficient of $\frac{1}{12}$ appearing as a gratuitous bonus).

We can exploit the superiority of the estimate (7) of e by modifying (3) to the equally evident identity.

$$n! = (\tfrac{1}{2})^{1.5}(\tfrac{2}{3})^{2.5}(\tfrac{3}{4})^{3.5} \times (\tfrac{4}{5})^{4.5} \ldots [(n-1)/n]^{n-1/2} n^{n+1/2},$$

or, equivalently,

$$n! = n^{n+1/2} \Big/ \prod_{j=1}^{n-1}\left(1+\frac{1}{j}\right)^{j+1/2}. \tag{10}$$

If we now approximate each term in the denominator by e we arrive at the considerably improved estimate

$$n! \sim n^{n+1/2}/e^{n-1}. \tag{11}$$

This differs from Stirling's formula only in the replacement of $\sqrt{2\pi} = 2.50662\ldots$ by $e = 2.71828\ldots$, thereby overesti-

Table I. The compound interest approximants to e (first column) and a much more rapidly convergent sequence (second).

n	$(1+1/n)^n$	$(1+1/n)^{n+1/2}$
1	2	2.828...
10	2.59...	2.7203...
100	2.704...	2.71830...
1000	2.7169...	2.718282...
10000	2.71814...	2.71828183...
100000	2.718268...	2.718281828...

mating the correct asymptotic form, but only by about $8\frac{1}{2}\%$.

The final step is suggested by writing the exact relation (10) as the approximate one (11) divided by the required correction:

$$n! = \frac{n^{n+1/2}}{e^{n-1}} \Big/ \prod_{j=1}^{n-1}\left(\frac{(1+1/j)^{j+1/2}}{e}\right). \tag{12}$$

Note that when n is large the product is exceedingly insensitive to the actual value of its upper limit, since increasing n by one augments the product by a factor $(1+1/n)^{n+1/2}/e$ which for example, differs from unity by only a part in a billion when n is 10 000. Thus for large n we will do better if we estimate the product in the denominator not by unity, as in (10), but by its limiting value for infinite[4] n. To correct the error brought about by this extension of the product, we can introduce in the numerator the product of all the terms by which we have augmented the denominator, thereby maintaining the identity (12) but in the slightly modified form:

$$n! = \frac{n^{n+1/2}}{e^{n-1}} \frac{\prod_{j=n}^{\infty}[(1+1/j)^{j+1/2}/e]}{\prod_{j=1}^{\infty}[(1+1/j)^{j+1/2}/e]}. \tag{13}$$

The product in the denominator of (13) is some definite number C, independent of n; the product in the numerator depends on n, but approaches unity as n becomes large. If (13) is to agree with Stirling's formula (1) for large n, it must then be that

$$C = \prod_{j=1}^{\infty}\left(\frac{(1+1/j)^{j+1/2}}{e}\right) = \frac{e}{\sqrt{2\pi}} = 1.084437551.... \tag{14}$$

If we insert this value into (13), with the understanding that π is some constant temporarily defined as $e^2/2C^2$, whose value we must eventually demonstrate to be 3.1415926535..., we arrive at the *exact identity*[5]:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \prod_{j=n}^{\infty}\left(\frac{(1+1/j)^{j+1/2}}{e}\right). \tag{15}$$

This identity is the centerpiece of this note. It cannot be new, but knowledge of it seems to have been lost by current generations of physicists. I have failed to turn up any reference to it after a few hours hunting in the Mathematics Library, though I did find some people getting rather close in 1877.[6]

The celebrated success of Stirling's formula, even for small values of n, is now easy to account for. When $n = 1$ Stirling's formula falls short of the exact (15) by the product from 1 to infinity, which is just the constant $C = 1.0844...$, whose closeness to unity is a measure of the excellence of the approximations (7) to e. When n is 2, the correction factor is the product from 2 to infinity—i.e., the worst of the estimates, $(1+1)^{1.5} = 2.828...$ no longer appears, and the correction drops down to $Ce/2\sqrt{2} = 1.0422....$ With each additional step n takes through the integers that factor in the product farthest from unity drops out, yielding smaller and smaller successive percent corrections.

Two points remain to be made:

(a) The fact that the number π appearing in (15) has the value 3.145926535... can be established in a suitably elementary manner.

(b) The exact form (15) can be used to give with minimum effort the leading terms in a series of improved approximations to $n!$

Point (b) is of no importance in any physical application I am aware of, but it is fun to show students how much better one can do with only a little more effort. Furthermore the results of such an extension can be put to use in establishing point (a). For this reason I attend first to point (b).

For the simplest estimate of the correction (15) gives to Stirling's formula, we need only use (9) to estimate the leading deviation of $(1+1)/j^{j+1/2}$ from e. This immediately gives as the correction factor

$$K = \prod_{j=n}^{\infty}\left(\frac{(1+1/j)^{j+1/2}}{e}\right) = \exp\left(\sum_{j=n}^{\infty}\frac{1}{12j^2}\right). \tag{16}$$

Estimating the sum by the corresponding integral, we have at once:

$$n! \sim \sqrt{2\pi n}(n/e)^n e^{1/12n}, \tag{17}$$

whose spectacular improvement on Stirling's formula students should be urged to investigate for themselves.

Equation (17) does better than one might think it has any right to do. The reason is that (17) gets correct not only the leading correction factor $(1+1/12n)$, but also the next term in the series, $1/288n^2$. The full correction factor is, in fact, the exponential of a series in odd inverse powers of n. To see this and extract the next few terms as well, we write the full correction factor in the form

$$K = \left(\frac{(1+1/j)^{j+1/2}}{e}\right) = \exp\left\{\sum_{j=n}^{\infty}\left[(j+\frac{1}{2})\ln\left(1+\frac{1}{j}\right)-1\right]\right\}. \tag{18}$$

The hidden structure of (18) is exposed by shifting from an integral variable of summation to one that assumes values halfway between the integers. Introducing the notational convention:

$$\sum_{n+1/2}^{\infty} f(m) = f(n+\tfrac{1}{2}) + f(n+\tfrac{3}{2}) + f(n+\tfrac{5}{2}) + \cdots, \tag{19}$$

and letting $j+\frac{1}{2}$ be a new variable m that assumes half-integral values, we rewrite (18):

$$K = \exp\left\{\sum_{n+1/2}^{\infty}\left[m\ln\left(\frac{1+1/2m}{1-1/2m}\right)-1\right]\right\}. \tag{20}$$

Expanding separately the logarithms of $1 \pm 1/2m$ then gives

$$K = \exp\left[\sum_{n+1/2}^{\infty}\left(\frac{1}{3(2m)^2}+\frac{1}{5(2m)^4}+\frac{1}{7(2m)^6}+\cdots\right)\right]. \tag{21}$$

By introducing the half-integral summation variable we have thus disposed of the odd terms in the expansion of the logarithm. But this is not the only service it performs for us: It is also tailor made for estimating the error we make when we go on to approximate the sums in (21) by integrals. To see this, start with the elementary identity

$$\int_n^{\infty} f(x)dx = \sum_{n+1/2}^{\infty}\int_{-1/2}^{1/2} f(m+x)dx, \tag{22}$$

and expand each $f(m+x)$ about $x=0$:

$$f(m+x) = f(m) + xf'(m) + (x^2/2)f''(m) + (x^3/6)f'''(m) + ... \tag{23}$$

Note that only the even terms survive the integration over the symmetric interval $(-\frac{1}{2},\frac{1}{2})$. The leading term gives the sum we wish to evaluate; the remaining terms give corrections to its approximation by the integral:

$$\sum_{n+1/2}^{\infty} f(m) = \int_n^{\infty} f(x)dx - \sum_{n+1/2}^{\infty} \frac{1}{3!} f''(m)/2^2 - \sum_{n+1/2}^{\infty} \frac{1}{5!} f^{iv}(m)/2^4 - \sum_{n+1/2}^{\infty} \frac{1}{7!} f^{vi}(m)/2^6 - \dots \quad (24)$$

The expansion (24) can be used to evaluate the sums in (21) to give corrections to (17) beyond the leading term $1/12n$. Suppose, for example, we desire the correction up to and including terms of order $1/n^5$. Then the terms beyond $1/m^6$ in (21) can be ignored, and we can estimate the m^{-6}, m^{-4}, and m^{-2} terms succesively as follows.

We only need the leading term in (24) to establish to the required degree of accuracy that

$$\sum_{n+1/2}^{\infty} \frac{1}{m^6} \sim \int_n^{\infty} \frac{dx}{x^6} = \frac{1}{5n^5}. \quad (25)$$

The term in m^{-4} requires the first two terms in (24):

$$\sum_{n+1/2}^{\infty} \frac{1}{m^4} \sim \int_m^{\infty} \frac{dx}{x^4} - \frac{5}{6} \sum_{n+1/2}^{\infty} \frac{1}{m^6} = 1/3n^3 - 1/6n^5, \quad (26)$$

where we have used the result (25) to evaluate the sum of m^{-6}. Finally, the term in m^{-2} requires the first three terms in (24):

$$\sum_{n+1/2}^{\infty} \frac{1}{m^2} \sim \int_n^{\infty} \frac{dx}{x^2} - \frac{1}{4} \sum_{n+1/2}^{\infty} \frac{1}{m^4} - \frac{1}{16} \sum_{n+1/2}^{\infty} \frac{1}{m^6} = 1/n - 1/12n^3 + 7/240n^5, \quad (27)$$

where we use the results (25) and (26) to evaluate the sums of m^{-4} and m^{-6}.

Using these three results to evaluate the first three terms in the expression (21) for the correction factor K to Stirling's formula leads immediately to

$$n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \exp\left(\frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5}\right), \quad (28)$$

where the error is now of order $1/n^7$ in the argument of the exponential. It is the rare student who is not deeply moved by the results of testing (28) on a programmable pocket calculator for successive values of n, starting with unity. I know of no more vivid demonstration of the power of a very little bit of elementary analysis.

But what about the pi? One could, of course, determine the constant in (28) to a precision well beyond anything one would ever require by, for example, fitting it so that (28) gave $n!$ exactly for $n = 10$ (which gives π to ten place accuracy). I find the following approach, however, more adaptable to numerical computation, more esthetic, and more in line with the original statistical motivation for the whole investigation.

Consider the quantity

$$P = (2n)!/2^{2n}(n!)^2. \quad (29)$$

This is, of course, the probability of getting an exact 50–50 split in $2n$ tosses of a balanced coin. By multiplying out everything in (29) we find that

$$P = \frac{1\cdot 3\cdot 5\cdot 7\cdots(2n-1)}{2\cdot 4\cdot 6\cdot 8\cdots(2n)}. \quad (30)$$

On the other hand, using (28) to estimate the value of P for large n gives

$$P \sim \frac{1}{\sqrt{\pi n}} \exp\left(-\frac{1}{8n} + \frac{1}{192n^3} - \frac{1}{640n^5}\right). \quad (31)$$

Students who know about Wallis's product for pi,

$$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\frac{8}{7}\cdots \quad (32)$$

can be immediately persuaded by a comparison of (31) (*without* the exponential correction) and (30) that our pi is *the* pi. Those who do not could be invited to evaluate the Wallis product numerically, but its convergence is exasperatingly slow. It is better to use the correction factor in (31) to estimate the unevaluated remainder in Wallis's product:

$$\pi = 4\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\frac{8}{7}\cdots\frac{(2n-2)(2n)}{(2n-1)(2n-1)} \times\exp(-1/4n + 1/96n^3 - 1/320n^5). \quad (33)$$

This produces values of π well beyond anything most physicists would care to remember. (For $n = 10$, 3.1415926528···; for $n = 50$, 3.141592653589784···.)

There are, of course, students who will want to know *why* Wallis's product should have anything to do with the circumference of a circle—indeed, they are the ones whose persistent questioning inspired this essay in the first place, and they deserve a better answer. For them the approach I prefer—the one I find (dare I say it?) the most *physical*—is to let the answer emerge in the subsequent investigation of entropy.

Elementary insights into irreversibility can be had from studying the statistics of coin tossing. For this one needs to know not only the probability (29) of an even break in $2n$ tosses, but, more generally, the probability of the number of heads or tails deviating from n by an amount x. When n is large one can use Stirling's formula (without the correction factor, but with our undetermined pi) to evaluate the appropriate binomial coefficients, and convince the class that the resulting expression is Gaussian when x is of order $n^{1/2}$, and utterly negligible when x is any greater. Determining the value of *our* pi then reduces to the problem of normalizing a Gaussian distribution, which is done in the usual way, squaring the normalization integral and evaluating it in polar coordinates. *The* pi comes from the angular part, the deed is accomplished, and the course can march on into the study of irreversibility, securely based on the rock solid foundation provided by Stirling's self-evident approximation to the elementary identity (15).

[1]This has been the subject of recent comment in this Journal. See, for example, S. A. Feller and E. Kaspar, Am. J. Phys. **50**, 682 (1982); or Y. Weissman, Am. J. Phys. **51**, 9 (1983). See also several letters to the editor in the Sept. 1983 issue.

[2]Although this article addresses teachers, not students, I have sketched the argument along lines one might present to a relatively innocent class, relegating more learned asides to footnotes. Readers in a hurry should skip directly to Eq. (15) and its accompanying note.

[3]The approximation to $n!$ is poor, since cumulative errors lead to a divergent correction factor, as remarked upon in footnote 4, below. The approximation to $\ln(n!)/n$, however, gets quite good as n increases since it

averages over a set of increasingly good approximants to ln(*e*).

[4]That the infinite product converges is an immediate consequence of the expansion (9), since the sum of $1/j^2$ converges. Note that this convergence would fail had we tried to base an analogous argument on the inferior estimate (8), owing to the divergence of the sum of $1/j$. The degree to which this entire point should be soft-pedaled in a general education course is best left to the discretion of the instructor. It would never occur to many such students that the infinite products in (13) *could* diverge, but the morality of relying on such innocence must be weighed in the conscience of each pedagogue.

[5]For the sophisticate who already knows Stirling's approximation it should be immediately evident that (15) is exact: (a) the product converges as a direct consequence of (9); (b) the value of (15) for $n+1$ is directly and easily shown to be $n+1$ times its value for n so the entire expression is indeed a constant times $n!$, and (c) the constant is determined by the requirement that (15) agree with (1) in the large n limit.

[6]See J. W. Glaisher, Q. J. Math. **15**, 57 (1877), and note the comment by Cayley at the end.

SOLUTION TO PROBLEM ON PAGE 339

The differential scattering cross section for transition from initial momentum state **k** to final momentum state **k**′, in Born approximation, is given by (in SI units)

$$\left(\frac{d\sigma}{d\Omega}\right)_{\mathbf{k}\to\mathbf{k}'} = \left(\frac{me}{2\pi\epsilon_0\hbar^2(\mathbf{k}-\mathbf{k}')^2}\right)^2 |F(\mathbf{k}-\mathbf{k}')|^2, \qquad (1)$$

where the form factor is $F(\mathbf{q}) = \int \rho(\mathbf{r}')\exp(i\mathbf{q}\cdot\mathbf{r}')dv'$. To solve this integral we choose x', y', and variables r', ϕ' as defined in Fig. 1. Therefore,

$$F(\mathbf{q}) = \rho_0 \int_0^R r'\,dr' \times \int_0^{2\pi} \exp(-ikr'\sin\theta\cos\phi')d\phi'. \qquad (2)$$

Using the following integrals involving Bessel functions[1]

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp[i\cos(\theta-\alpha)]d\theta,$$

$$xJ_1(x) = \int_0^x x'J_0(x')dx',$$

the form factor is obtained as $F(\mathbf{q}) = 2QJ_1(kR\sin\theta)/kR\sin\theta$. Therefore, the differential scattering cross section is

$$\frac{d\sigma}{d\Omega} = \left(\frac{meQ}{2\pi\epsilon_0\hbar^2\mathbf{q}^2}\right)^2 \left|\frac{2J_1(kR\sin\theta)}{kR\sin\theta}\right|^2.$$

The intensity of Fraunhofer diffraction from a circular aperture[1] is

$$I(\theta) = I_0|2J_1(kR\sin\theta)/kR\sin\theta|^2.$$

To envisage the similarity, recall that a convenient way of obtaining a physical picture of scattering is to use Huygens's principle[2] to get Eq. (1). It is visualized that the incident plane wave ψ_i is scattered from various volume elements like dv' (see Fig. 2). Scattering at each volume element is Rutherford type (point scattering) and, therefore, contributes an amplitude proportional to $\rho(\mathbf{r}')dv'/\mathbf{q}^2$. Furthermore, the incident wave reaching at point **r**′ has an additional phase $\exp(i\mathbf{k}\cdot\mathbf{r}')$ with respect to the wave reaching an arbitrary origin. The spherical wave originating at point **r**′ carries this information. The contribution of dv' to the scattered wave is $d\psi_s = (A/r)e^{ikr}e^{i\mathbf{k}\cdot\mathbf{r}'}$, where $A = C\rho(\mathbf{r}')dv'/\mathbf{q}^2$. The proportionality constant C may be set equal to $(me/2\pi\epsilon_0\hbar^2)$. The scattered wave $\psi_s = \psi_{\text{total}} - \psi_i$ results from a coherent superposition of individual waves $d\psi_s$,

$$\psi_s = \frac{e^{ikR}}{R}\left(\frac{C}{\mathbf{q}^2}\int \rho(\mathbf{r}')\exp(i\mathbf{q}\cdot\mathbf{r}')dv'\right).$$

Comparision with $\psi_{\text{total}} = \psi_i + f(\mathbf{q})\exp(ikR)/R$ leads to Eq. (1). This picture corresponds exactly to the Fraunhofer diffraction in optics. The analogy works because of the assumption that incident wave is not greatly disturbed by scattering (first Born approximation). It also illustrates interference effects in scattering.

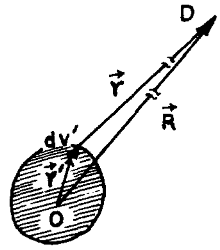

Fig. 2. Diffraction scattering. The boundary condition is that $r' \ll R$. Therefore, with $\mathbf{k}' = k\hat{R}$, $kr = kR - \mathbf{k}'\cdot\mathbf{r}'$.

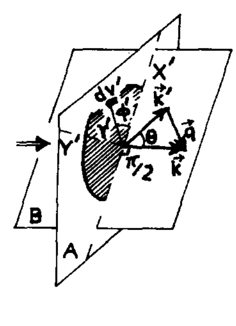

Fig. 1. Choice of coordinates x', y' and variables r', ϕ' to obtain Eq. (2). Plane A contains circularly symmetric charge distribution. The vectors **k**, **k**′, and **q** are coplaner and define plane B.

Sardar Singh and N. K. Sharma
Department of Physics
University of Rajasthan
Jaipur 302 004, India

[1]M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1970), pp. 395–396; see Eqs. (9), (12), and (14).

[2]D. Bohm, *Quantum Theory* (Prentice-Hall, New York, 1951), p. 546.