Lecture 8
The arc sine law.
Random walk.
The ballot theorem.

Shlomo Sternberg

The arc sine law.

The probability distribution with density

$$\sigma(x) = \frac{1}{\pi\sqrt{x(1-x)}}$$

is called the **arc sine law** in probability theory because, if I is the interval $I = [0, u]$ then

$$\mathrm{Prob}\ x \in I = \mathrm{Prob}\ 0 \le x \le u = \int_0^u \frac{1}{\pi\sqrt{x(1-x)}} = \frac{2}{\pi} \arcsin\sqrt{u}. \tag{1}$$

$$\text{Prob } x \in I = \text{Prob } 0 \le x \le u = \int_0^u \frac{1}{\pi\sqrt{x(1-x)}} = \frac{2}{\pi} \arcsin \sqrt{u}. \qquad (1)$$

We have already verified this integration because $I = h(J)$ where

$$h(t) = \sin^2 \frac{\pi t}{2}, \quad J = [0, v], \; h(v) = u,$$

and the probability measure we are studying is the push forward of the uniform distribution. So

$$\text{Prob } h(t) \in I = \text{Prob } t \in J = v.$$

The arc sine law plays a crucial role in the theory of fluctuations in random walks. As a cultural diversion we explain some of the key ideas, in today's lecture.

Suppose that there is an ideal coin tossing game in which each player wins or loses a unit amount with (independent) probability $\frac{1}{2}$ at each throw. Let $S_0 = 0, S_1, S_2, \ldots$ denote the successive cumulative gains (or losses) of the first player. We can think of the values of these cumulative gains as being marked off on a vertical s-axis, and representing the position of a particle which moves up or down with probability $\frac{1}{2}$ at each (discrete) time unit .

Let

$$\alpha_{2k,2n}$$

denote the *probability that up to and including time* $2n$*, the last visit to the origin occurred at time* $2k$. Let

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}. \tag{2}$$

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}. \qquad (2)$$

So $u_{2\nu}$ represents the probability that exactly ν out of the first 2ν steps were in the positive direction, and the rest in the negative direction. In other words, $u_{2\nu}$ is the probability that the particle has returned to the origin at time 2ν.

Using Stirling's formula.

We can find a simple approximation to $u_{2\nu}$ using Stirling's formula for an approximation to the factorial:

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

where the $\sim$ signifies that the ratio of the two sides tends to one as n tends to infinity.

I will discuss Stirling's formula at the end of today's lecture. In the meanwhile, let's take it for granted.

Then

$$\begin{aligned} u_{2\nu} &= 2^{-2\nu}\frac{(2\nu)!}{(\nu!)^2} \\ &\sim 2^{-2\nu}\frac{\sqrt{2\pi}(2\nu)^{2\nu+\frac{1}{2}}e^{-2\nu}}{2\pi\nu^{2\nu+1}e^{-2\nu}} \\ &= \frac{1}{\sqrt{\pi\nu}}. \end{aligned}$$

The results we wish to prove in the next few slides are

We have

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}, \tag{3}$$

so we have the asymptotic approximation

$$\alpha_{2k,2n} \sim \frac{1}{\pi\sqrt{k(n-k)}}. \tag{4}$$

If we set

$$x_k = \frac{k}{n}$$

then we can write

$$\alpha_{2k,2n} \sim \frac{1}{n}\sigma(x_k). \tag{5}$$

Thus, for fixed $0 < x < 1$ and n sufficiently large

$$\sum_{k<xn} \alpha_{2k,2n} \doteq \frac{2}{\pi} \arcsin \sqrt{x}. \tag{6}$$

The probability that in the time interval from 0 to $2n$ the particle spends $2k$ time units on the positive side and $2n-2k$ time units on the negative side equals $\alpha_{2k,2n}$. In particular, if $0<x<1$ the probability that the fraction k/n of time units spent on the positive be less than x tends to $\frac{2}{\pi}\arcsin\sqrt{x}$ as $n\to\infty$.

Let us call the value of S_{2n} for any given realization of the random walk, the **terminal point**. Of course, the particle may well have visited this terminal point earlier in the walk, and we can ask when it first reaches its terminal point.

The probability that the first visit to the terminal point occurs at time $2k$ is given by $\alpha_{2k,2n}$.

We can also ask for the first time that the particle reaches its maximum value: We say that the **first maximum occurs at time** l if

$$S_0 < S_l, S_1 < S_l, \ldots S_{l-1} < S_l, \qquad S_{l+1} \leq S_l, S_{l+2} \leq S_l, \ldots S_{2n} \leq S_l. \tag{7}$$

If $0 < l < 2n$ the probability that the first maximum occurs at $l = 2k$ or $l = 2k+1$ is given by $\frac{1}{2}\alpha_{2k,2n}$. For $l = 0$ this probability is given by u_{2n} and if $l = 2n$ it is given by $\frac{1}{2}u_{2n}$.

Before proving these various facts, let us discuss a few of their implications which some people find counterintuitive. For example, because of the shape of the density, σ, the last result implies that the maximal accumulated gain is much more likely to occur very near to the beginning or to the end of a coin tossing game rather than somewhere in the middle. The fourth assertion implies that the probability that the first visit to the terminal point occurs at time $2k$ is that same as the probability that it occurs at time $2n-2k$ and that very early first visits and very late first visits are much more probable than first visits some time in the middle.

In order to get a better feeling for the assertion of the first few assertions, let us tabulate the values of $\frac{2}{\pi} \arcsin \sqrt{x}$ for $0 \leq x \leq \frac{1}{2}$.

x	$\frac{2}{\pi} \arcsin \sqrt{x}$	x	$\frac{2}{\pi} \arcsin \sqrt{x}$
0.05	0.144	0.30	0.369
0.10	0.205	0.35	0.403
0.15	0.253	0.40	0.236
0.20	0.295	0.45	0.468
0.25	0.333	0.50	0.500

This table, in conjunction with our first two assertions says that if a great many coin tossing games are conducted every second, day and night for a hundred days, then in about 14.4 percent of the cases,the lead will not change after day five.

The proof of all our assertions hinges on three lemmas. Let us graph (by a polygonal path) the walk of a particle. So a "path" is a broken line segment made up of segments of slope ± 1 joining integral points to integral points in the plane (with the time or $t-$axis horizontal and the $s-$axis vertical). If $A = (a, \alpha)$ is a point, we let $A' = (a, -\alpha)$ denote its image under reflection in the $t-$axis.

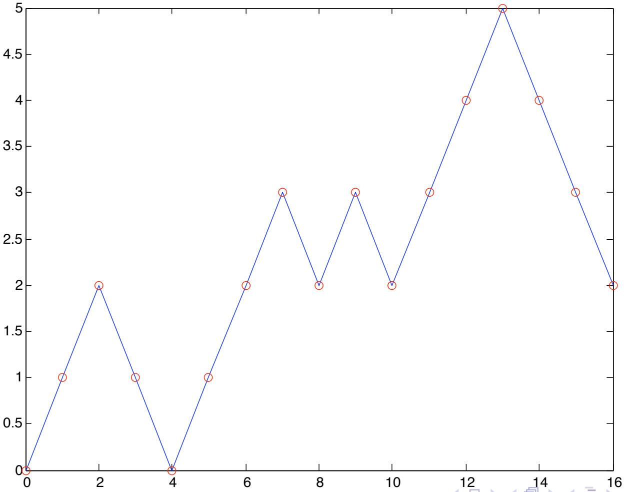
Lemma

The reflection principle. *Let* $A = (a, \alpha), B = (b, \beta)$ *be points in the first quadrant with* $b > a \geq 0, \alpha > 0, \beta > 0$. *The number of paths from* A *to* B *which touch or cross the* $t-$ *axis equals the number of all paths from* A' *to* B.

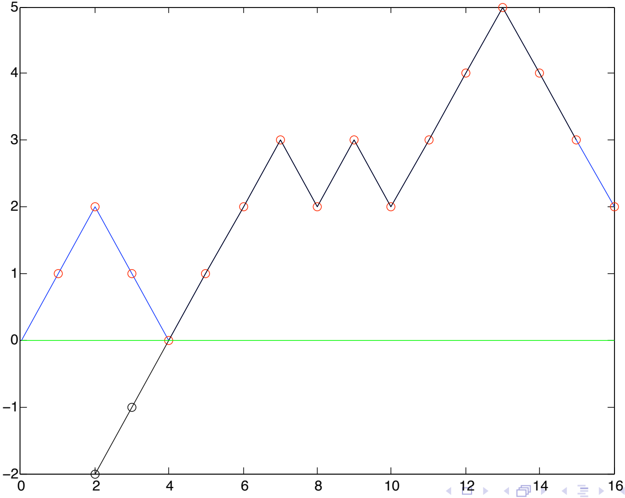
Proof.

For any path from A to B which touches the horizontal axis, let t be the abscissa of the first point of contact. Reflect the portion of the path from A to $T = (t, 0)$ relative to the horizontal axis. This reflected portion is a path from A' to T, and continues to give a path from A' to B. This procedure assigns to each path from A to B which touches the axis, a path from A' to B. This assignment is bijective: Any path from A' to B must cross the $t-$axis. Reflect the portion up to the first crossing to get a touching path from A to B. This is the inverse assignment. □

Original path.



Reflected path.



A path with n steps will join $(0,0)$ to (n,x) if and only if it has p steps of slope $+1$ and q steps of slope -1 where

$$p+q=n, \quad p-q=x.$$

The number of such paths is the number of ways of picking the positions of the p steps of positive slope and so the number of paths joining $(0,0)$ to (n,x) is

$$N_{n,x}=\left(\begin{array}{c} p+q \\ p \end{array}\right)=\left(\begin{array}{c} n \\ \frac{n+x}{2} \end{array}\right).$$

It is understood that this formula means that $N_{n,x}=0$ when there are no paths joining the origin to (n,x).

Lemma

The ballot theorem. *Let n and x be positive integers. There are exactly*

$$\frac{x}{n} N_{n,x}$$

paths which lie strictly above the t axis for $t > 0$ and join $(0,0)$ to (n,x).

Proof of the ballot theorem.

$$N_{n,x} = \begin{pmatrix} p+q \\ p \end{pmatrix} = \begin{pmatrix} n \\ \frac{n+x}{2} \end{pmatrix}.$$

There are as many paths joining $(0,0)$ to (n,x) which are strictly above the x-axis as there are paths joining $(1,1)$ to (n,x) which do *not* touch or cross the $t-$axis. This is the same as the total number of paths which join $(1,1)$ to (n,x) less the number of paths which do touch or cross. By the reflection principle, the number of paths which do touch or cross is the same as the number of paths joining $(1,-1)$ to (n,x) which is $N_{n-1,x+1}$.

Completion of the proof of the ballot theorem.

Thus, with p and q as above, the number of paths which lie strictly above the t axis for $t > 0$ and which join $(0,0)$ to (n,x) is

$$\begin{aligned} N_{n-1,x-1} - N_{n-1,x+1} &= \binom{p+q-1}{p-1} - \binom{p+q-1}{p} \\ &= \frac{(p+q-1)!}{(p-1)!(q-1)!}\left[\frac{1}{q} - \frac{1}{p}\right] \\ &= \frac{p-q}{p+q} \times \frac{(p+q)!}{p!q!} \\ &= \frac{x}{n} N_{n,x} \quad \square \end{aligned}$$

Why is this called the Ballot Theorem?

The reason that this lemma is called the Ballot Theorem is that it asserts that if candidate P gets p votes, and candidate Q gets q votes in an election where the probability of each vote is independently $\frac{1}{2}$, and if P wins, i.e. if $p > q$, then the probability that throughout the counting there are more votes for P than for Q is given by

$$\frac{p-q}{p+q}.$$

Here is our last lemma:

Lemma

The probability that from time 1 to time $2n$ the particle stays strictly positive is given by $\frac{1}{2}u_{2n}$. In symbols,

$$\text{Prob}\,\{S_1 > 0, \ldots, S_{2n} > 0\} = \frac{1}{2}u_{2n}. \tag{8}$$

So

$$\text{Prob}\,\{S_1 \neq 0, \ldots, S_{2n} \neq 0\} = u_{2n}. \tag{9}$$

Also

$$\text{Prob}\,\{S_1 \geq 0, \ldots, S_{2n} \geq 0\} = u_{2n}. \tag{10}$$

By considering the possible positive values of S_{2n} which can range from 2 to $2n$ we have $\mathrm{Prob}\,\{S_1 > 0, \dots, S_{2n} > 0\}$

$$\begin{aligned} &= \sum_{r=1}^{n} \mathrm{Prob}\,\{S_1 > 0, \dots, S_{2n} = 2r\} \\ &= 2^{-2n} \sum_{r=1}^{n} (N_{2n-1,2r-1} - N_{2n-1,2r+1}) \\ &= 2^{-2n} (N_{2n-1,1} - N_{2n-1,3} + N_{2n-1,3} - N_{2n-1,5} + \cdots) \\ &= 2^{-2n} N_{2n-1,1}. \end{aligned}$$

The passage from the first line to the second is the reflection principle, as in our proof of the Ballot Theorem, from the third to the fourth is because the sum telescopes to $N_{2n-1,1} - N_{2n-1,2n+1}$ and $N_{2n-1,2n+1} = 0$ because you can't get from 0 to $2n + 1$ in $2n - 1$ steps.

Proof of $\mathrm{Prob}\,\{S_1 > 0, \ldots, S_{2n} > 0\} = \frac{1}{2}u_{2n}$. (8)

So

$$\begin{aligned}\mathrm{Prob}\,\{S_1 > 0, \ldots, S_{2n} > 0\} &= 2^{-2n}N_{2n-1,1} \\ &= \frac{1}{2}p_{2n-1,1} \\ &= \frac{1}{2}u_{2n}.\end{aligned}$$

The $p_{2n-1,1}$ on the next to the last line is the probability of ending up at $(2n-1,1)$ starting from $(0,0)$. The last equality is simply the assertion that to reach zero at time $2n$ we must be at ± 1 at time $2n-1$ (each of these has equal probability, $p_{2n-1,1}$) and for each alternative there is a 50 percent chance of getting to zero on the next step. This proves (8).

Proof of $\text{Prob}\,\{S_1 \neq 0, \ldots, S_{2n} \neq 0\} = u_{2n}$. (9)

Since a path which never touches the $t-$axis must be always above or always below the $t-$axis, (9) follows immediately from (8).

Proof of $\text{Prob}\,\{S_1 \geq 0, \dots, S_{2n} \geq 0\} = u_{2n}$. (10).

Observe that a path which is strictly above the axis from time 1 on, must pass through the point $(1,1)$ and then stay above the horizontal line $s = 1$. The probability of going to the point $(1,1)$ at the first step is $\frac{1}{2}$, and then the probability of remaining above the new horizontal axis is $\text{Prob}\,\{S_1 \geq 0, \dots, S_{2n-1} \geq 0\}$. But since $2n - 1$ is odd, if $S_{2n-1} \geq 0$ then $S_{2n} \geq 0$. So, by (8) we have

$$\begin{aligned}
\frac{1}{2}u_{2n} &= \text{Prob}\,\{S_1 > 0, \dots, S_{2n} > 0\} \\
&= \frac{1}{2}\text{Prob}\,\{S_1 \geq 0, \dots, S_{2n-1} \geq 0\} \\
&= \frac{1}{2}\text{Prob}\,\{S_1 \geq 0, \dots, S_{2n-1} \geq 0, S_{2n} \geq 0\},
\end{aligned}$$

completing the proof of the lemma.

Proof of $\alpha_{2k,2n} = u_{2k}u_{2n-2k}$. (3)

We can now turn to the proofs of the assertions about the last visit probabilities, $\alpha_{2k,2n}$.

To say that the last visit to the origin occurred at time $2k$ means that

$$S_{2k} = 0$$

and

$$S_j \neq 0, \quad j = 2k+1, \ldots, 2n.$$

$\alpha_{2k,2n} = u_{2k}u_{2n-2k}.$ (3)

Recall that

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}$$

is the probability that the particle has returned to the origin at time 2ν.

By definition, the first $2k$ positions can be chosen in $2^{2k}u_{2k}$ ways to satisfy the condition $S_{2k} = 0$. Taking the point $(2k, 0)$ as our new origin, (9) says that there are $2^{2n-2k}u_{2n-2k}$ ways of choosing the last $2n - 2k$ steps so as to satisfy the condition $S_j \neq 0, \quad j = 2k+1, \ldots, 2n$. Multiplying and then dividing the result by 2^{2n} proves (3).

To prove:

The probability that in the time interval from 0 to $2n$ the particle spends $2k$ time units on the positive side and $2n - 2k$ time units on the negative side equals $\alpha_{2k,2n}$. In particular, if $0 < x < 1$ the probability that the fraction k/n of time units spent on the positive side be less than x tends to $\frac{2}{\pi} \arcsin \sqrt{x}$ as $n \to \infty$.

Proof.

We consider paths of $2n$ steps and let $b_{2k,2n}$ denote the probability that exactly $2k$ sides lie above the $t-$axis. We want to show that

$$b_{2k,2n} = \alpha_{2k,2n}.$$

For the case $k = n$ we have $\alpha_{2n,2n} = u_0 u_{2n} = u_{2n}$ and $b_{2n,2n}$ is the probability that the path lies entirely above the axis. So our assertion reduces to (10) which we have already proved. By symmetry, the probability of the path lying entirely below the the axis is the same as the probability of the path lying entirely above it, so $b_{0,2n} = \alpha_{0,2n}$ as well.

So we need prove our assertion for $1 \leq k \leq n-1$. In this situation, a return to the origin must occur. Suppose that the first return to the origin occurs at time $2r$. There are then two possibilities: the entire path from the origin to $(2r, 0)$ is either above the axis or below the axis. If it is above the axis, then $r \leq k \leq n-1$, and the section of the path beyond $(2r, 0)$ has $2k - 2r$ edges above the $t-$axis. The number of such paths is

$$\frac{1}{2} 2^{2r} f_{2r} 2^{2n-2r} b_{2k-2r,2n-2r}$$

where f_{2r} denotes the *probability of first return* at time $2r$:

$$f_{2r} = \mathrm{Prob}\,\{S_1 \neq 0, \ldots, S_{2r-1} \neq 0, S_{2r} = 0\}.$$

If the first portion of the path up to $2r$ is spent below the axis, the the remaining path has exactly $2k$ edges above the axis, so $n - r \geq k$ and the number of such paths is

$$\frac{1}{2} 2^{2r} f_{2r} 2^{2n-2r} b_{2k,2n-2r}.$$

So we get the recursion relation

$$b_{2k,2n} = \frac{1}{2}\sum_{r=1}^{k} f_{2r} b_{2k-2r,2n-2r} + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r} b_{2k,2n-2r} \quad 1 \le k \le n-1. \tag{11}$$

Now we proceed by induction on n. We know that $b_{2k,2n} = u_{2k}u_{2n-2k} = \frac{1}{2}$ when $n = 1$. Assuming the result up through $n-1$, the recursion formula (11) becomes

$$b_{2k,2n} = \frac{1}{2}u_{2n-2k}\sum_{r=1}^{k} f_{2r} u_{2k-2r} + \frac{1}{2}u_{2k}\sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r}. \tag{12}$$

But we claim that the probabilities of return and the probabilities of first return are related by

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \cdots + f_{2n} u_0. \tag{13}$$

Indeed, if a return occurs at time $2n$, then there must be a first return at some time $2r \leq 2n$ and then a return in $2n - 2r$ units of time, and the sum in (13) is over the possible times of first return. If we substitute (13) into the first sum in (12) it becomes u_{2k} while substituting (13) into the second term yields u_{2n-2k}. Thus (12) becomes

$$b_{2k,2n} = u_{2k} u_{2n-2k}$$

which is our desired result.

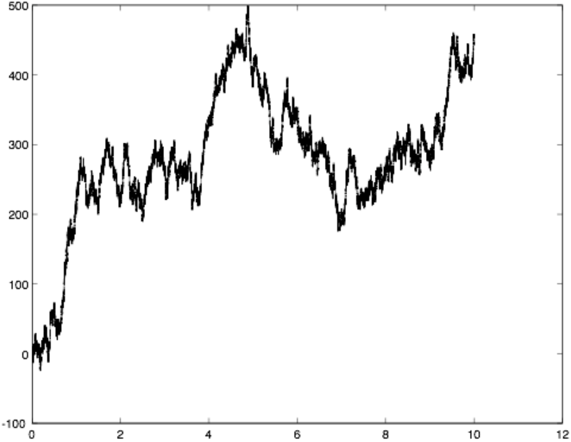
To prove: The probability that the first visit to the terminal point occurs at time $2k$ is given by $\alpha_{2k,2n}$.

The probability in the title is the probability that $S_{2k} = S_{2n}$ but $S_j \neq S_{2n}$ for $j < 2k$. Reading the path rotated through 180° about the end point, and with the endpoint shifted to the origin, this is clearly the same as the probability that $2n - 2k$ is the last visit to the origin. $\square$

The first visit to the maximum.

The probability that the maximum is achieved at 0 is the probability that $S_1 \leq 0, \ldots, S_{2n} \leq 0$ which is u_{2n} by (10). The probability that the maximum is first obtained at the terminal point, is, after rotation and translation, the same as the probability that $S_1 > 0, \ldots, S_{2n} > 0$ which is $\frac{1}{2}u_{2n}$ by (8). If the maximum occurs first at some time l in the middle, we combine these results for the two portions of the path - before and after time ℓ - together with (3) to complete the proof. □

The next slide shows a computer generated random walk with 100,000 steps. The last zero is at time 3783. For the remaining 96,217 steps the path is positive. According to the arc sine law, with probability $1/5$, the particle will spend about 97.6 percent of its time on one side of the origin.


500
400
300
200
100
0
-100
0
2
4
6
8
10
12

The Beta distributions.

The arc sine law is the special case $a = b = \frac{1}{2}$ of the Beta distribution with parameters a, b which has probability density proportional to

$$t^{a-1}(1-t)^{b-1}.$$

So long as $a > 0$ and $b > 0$ the integral

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$$

converges, and was evaluated by Euler to be

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

where Γ is Euler's Gamma function. So the Beta distributions with $a > 0, b > 0$ are given by

$$\frac{1}{B(a,b)} t^{a-1}(1-t)^{b-1}.$$

We characterized the arc sine law ($a = b = \frac{1}{2}$) as being the unique probability density invariant under L_4. The case $a = b = 0$, where the integral does not converge, also has an interesting characterization as an invariant density.

Consider transformations of the form

$$t \mapsto \frac{at+b}{ct+d}$$

where the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is invertible. Suppose we require that the transformation preserve the origin and the point $t = 1$. Preserving the origin requires that $b = 0$, while preserving the point $t = 1$ requires that $a = c + d$. Since $b = 0$ we must have $ad \neq 0$ for the matrix to be invertible. Since multiplying all the entries of the matrix by the same non-zero scalar does not change the transformation, we may as well assume that $d = 1$, and hence the family transformations we are looking at are

$$\phi_a : \quad t \mapsto \frac{at}{(a-1)t+1}, \quad a \neq 0.$$

Notice that

Our claim is that, up to scalar multiple, the density

$$\rho(t) = \frac{1}{t(1-t)}$$

is the unique density such that the measure

$$\rho(t)dt$$

is invariant under all the transformations ϕ_a.

Indeed,

$$\phi_a'(t) = \frac{a}{[1-t+at]^2}$$

so the condition of invariance is

$$\frac{a}{[1-t+at]^2}\rho(\phi_a(t)) = \rho(t).$$

Let us normalize ρ by

$$\rho\left(\frac{1}{2}\right) = 4.$$

Then

$$s = \phi_a\left(\frac{1}{2}\right) \Leftrightarrow s = \frac{a}{1+a} \Leftrightarrow a = \frac{s}{1-s}.$$

So taking $t = \frac{1}{2}$ in the condition for invariance and a as above, we get

$$\rho(s) = 4((1-s)/s)[\frac{1}{2} + \frac{1}{2}\frac{s}{1-s}]^2 = \frac{1}{s(1-s)}.$$

This elementary geometrical fact - that $1/t(1-t)$ is the unique density (up to scalar multiple) which is invariant under all the ϕ_a - was given a deep philosophical interpretation by Jaynes, [**?**]:

Suppose we have a possible event which may or may not occur, and we have a population of individuals each of whom has a clear opinion (based on ingrained prejudice, say, from reading the newspapers or watching television) of the probability of the event being true. So Mr. A assigns probability p(A) to the event E being true and (1-p(A)) as the probability of its not being true, while Mr. B assigns probability P(B) to its being true and (1-p(B)) to its not being true and so on.

Suppose an additional piece of information comes in, which would have a (conditional) probability x of being generated if E were true and y of this information being generated if E were not true. We assume that both x and y are positive, and that every individual thinks rationally in the sense that on the advent of this new information he changes his probability estimate in accordance with Bayes' law, which says that the posterior probability p' is given in terms of the prior probability p by

$$p' = \frac{px}{px + (1-p)y} = \phi_a(p) \text{ where } a := \frac{x}{y}.$$

We might say that the population as a whole has been "invariantly prejudiced" if any such additional evidence does not change the proportion of people within the population whose belief lies in a small interval. Then the density describing this state of knowledge (or rather of ignorance) must be the density

$$\rho(p) = \frac{1}{p(1-p)}.$$

According to this reasoning of Jaynes, we take the above density to describe the prior probability an individual (thought of as a population of subprocessors in his brain) would assign to the probability of an outcome of a given experiment. If a series of experiments then yielded M successes and N failures, Bayes' theorem (in its continuous version) would then yield the posterior distribution of probability assignments as being proportional to

$$p^{M-1}(1-p)^{N-1}$$

the Beta distribution with parameters M, N.

I am going to give two rather different looking proofs Stirling's formula. Both illustrate important methods in asymptotic analysis.

The Euler-Maclauren summation formula.

Let f be a continuously differentiable function on $[0, n]$. Integration by parts shows that for every integer $k = 0, 1, \ldots, n-1$ we have

$$\int_k^{k+1} [x - k - \frac{1}{2}] f'(x) dx = \frac{1}{2}[f_k + f_{k+1}] - \int_k^{k+1} f(x) dx$$

where we have written f_k for $f(k)$ to shorten the formula.

Letting $[x]$ denote the largest integer $\leq x$ we can write this as

$$\frac{1}{2}[f_k + f_{k+1}] = \int_k^{k+1} f(x) dx + \int_k^{k+1} [x - [x] - \frac{1}{2}] f'(x) dx.$$

$$\frac{1}{2}[f_k + f_{k+1}] = \int_k^{k+1} f(x)dx + \int_k^{k+1} [x - [x] - \frac{1}{2}]f'(x)dx.$$

Summing this from 0 to $n-1$ and adding $\frac{1}{2}[f_0 + f_n]$ gives (a baby version of) the **Euler-Maclauren formula**

$$f_0 + \cdots + f_n = \frac{1}{2}[f_0 + f_n] + \int_0^n f(x)dx + \int_0^n P_1(x)f'(x)dx$$

where

$$P_1(x) := x - [x] - \frac{1}{2}.$$

The function $P_1(x)$ is periodic of period one and is continuous except at the integers, and its integral over any interval of length one is zero.

$$P_1(x) := x - [x] - \frac{1}{2}.$$

We will let P_2 be an indefinite integral of P_1. To be specific, let us define

$$P_2(x) := \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{12} \quad 0 \leq x \leq 1$$

and extended to be periodic of period one. Its integral over any interval of length one is then also zero.

Let us apply the Euler-Maclauren summation formula to the function $f(x) = \frac{1}{1+x}$ and replace n by $n-1$ in the formula. We get

$$\log 1 + \log 2 + \cdot + \log n = \int_1^n \log x dx + \frac{1}{2}\log n + \int_1^n \frac{P_1(x)}{x}dx$$

$$= (n+\frac{1}{2})\log n - (n-1) + \int_1^n \frac{P_1(x)}{x}dx.$$

Integration by parts gives

$$\int_1^n \frac{P_1(x)}{x}dx = \frac{P_2(x)}{x}\Big|_1^n + \int_1^n \frac{P_2(x)}{x^2}dx,$$

which shows that the integral converges as $n \to \infty$.

So $\log(n!) = \left(n + \frac{1}{2}\right) - n + c_n$ where $c_n \to c$ for some value c as $n \to \infty$. So

$$n! \sim Ce^{n+\frac{1}{2}}e^{-n}$$

for some constant C.

We have shown that

$$n! \sim Ce^{n+\frac{1}{2}}e^{-n}$$

for some constant C.

Stirling's formula says that $C = \sqrt{2\pi}$. But we can conclude this from what we already know. For if we go back and do our computations with a "general C" we will find that the only way we could get a probability (i.e. that the appropriate integrals are 1) is if $C = \sqrt{2\pi}$.

Here is another proof:

We will use Euler's Gamma function:

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

so that

$$\Gamma(n+1) = n!.$$

We are going to make a change of variable in

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

Setting $t = s\tau$ this becomes

$$s^{s+1} \int_0^\infty \tau^s e^{-\tau s} d\tau = s^{s+1} e^{-s} \int_0^\infty e^{-s(\tau - 1 - \log \tau)} d\tau.$$

So

$$\Gamma(s+1) = s^{s+1}e^{-s}\int_0^\infty e^{-sf(\tau)}d\tau$$

where

$$f(\tau) = \tau - 1 - \log\tau.$$

Here is the idea: The function f vanishes at $\tau = 1$ and achieves a minimum there, tending to ∞ as $\tau \to 0$ or as $\tau \to \infty$. So for large values of s we expect that the contribution to the integral will come from τ near 1. Near 1 we will approximate f by its quadratic term which will then lead to Gaussian integral.

The following details are taken from Courant & Hilbert *Methods of Mathematical Physics* Vol 1 pp. 522-524.

Let $0 < \epsilon < \frac{1}{2}$. For $\frac{1}{2} < \tau < 1$ we have

$$\tau - 1 - \log \tau = \int_\tau^1 \left(\frac{1}{u} - 1 \right) du \geq \int_\tau^1 (1-u) du$$

$$= \frac{1}{2}(\tau - 1)^2 \geq \frac{1}{8}(\tau - 1)^2.$$

On the interval $(0, 1-\epsilon)$ the integrand $e^{-sf(\tau)}$ is less than its maximum value which is $e^{-s\epsilon^2/8}$. So

$$\int_0^{1-\epsilon} e^{-sf(\tau)} d\tau \leq e^{-s\epsilon^2/8}.$$

Similarly, for $1 \le \tau \le 4$

$$\tau - 1 - \log \tau = \int_1^\tau \left(1 - \frac{1}{u}\right) du \ge 1\frac{1}{4}\int_1^\tau (u-1)du = \frac{1}{8}(\tau - 1)^2.$$

Again replacing the integrand by its largest value shows that

$$\int_{1+\epsilon}^4 e^{-sf(\tau)} d\tau \le 3e^{-s\epsilon^2/8}.$$

For $\tau \ge 4$, $\tau - 1 - \log \tau \ge \frac{3}{4}\tau - \log \tau \ge \frac{1}{4}\tau$. Hence for $s > 4$

$$\int_4^\infty e^{-sf(\tau)} d\tau < \int_4^\infty e^{-s\tau/4} d\tau < e^{-s} < e^{-s\epsilon^2/8}.$$

So if we take $\epsilon = s^{-2/5}$ we will have

$$e^s s^{-s-1}\Gamma(s+1) = \int_{1-\epsilon}^{1+\epsilon} e^{-sf(\tau)} d\tau + O(e^{-s^{1/5}/8}).$$

So we are left with the study of the integral $\int_{1-\epsilon}^{1+\epsilon} e^{-sf(\tau)}d\tau$.
Now $f(1) = f'(1) = 0$ and $f''(1) = 1$. So we can write

$$f(\tau) = \frac{1}{2}(\tau-1)^2+(\tau-1)^3\psi(\tau) \quad \text{where} \quad |\psi(\tau)| \leq M \text{ on } \frac{1}{2} \leq \tau \leq \frac{3}{2}.$$

So on this interval we have

$$e^{-s(\tau-1)^2/2}e^{-Ms^{-1/5}} \leq e^{-sf(\tau)} \leq e^{-s(\tau-1)^2/2}e^{Ms^{-1/5}}$$

and

$$e^{-sf(\tau)} = e^{-s(\tau-1)^2/2}(1 + O(s^{-1/5})).$$

So

$$\int_{1-\epsilon}^{1+\epsilon} e^{-sf(\tau)}d\tau = (1 + O(s^{-1/5}))\int_{-\epsilon}^{\epsilon} e^{-su^2/2}du.$$

We have shown that

$$\int_{1-\epsilon}^{1+\epsilon} e^{-sf(\tau)}d\tau = (1+O(s^{-1/5}))\int_{-\epsilon}^{\epsilon} e^{-su^2/2}du.$$

If we make the change of variables $v = s^{\frac{1}{2}}u$ the integral on the right becomes

$$s^{-1/2}\int_{-\epsilon s^{\frac{1}{2}}}^{\epsilon s^{\frac{1}{2}}} e^{-v^2/2}dv$$

and

$$\int_{-\epsilon s^{\frac{1}{2}}}^{\epsilon s^{\frac{1}{2}}} e^{-v^2/2}dv \to \int_{-\infty}^{\infty} e^{-v^2/2}dv = \sqrt{2\pi}. \quad \square$$