640:192:03 **How big is a ball?** **10/23/96**

I want to compute the volume of the unit ball in $\mathbb{R}^n$, n-dimensional space. The unit ball will be the collection of points whose distance to the origin is less than or equal to 1. Before starting the computation, we should probably define all the terms used in this paragraph!

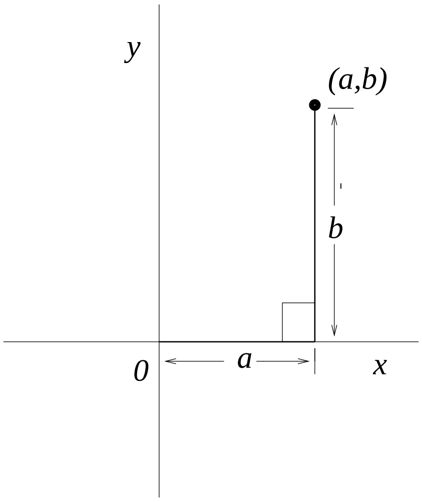
I wanted to do this computation because it is a significant application of a reduction formula obtained using integration by parts. It also provides some interesting and perhaps unsuspected asymptotic information about higher dimensional geometry. A few other interesting intellectual morsels are gained, including some hints about the factorial function and a method for approximating π. Learning more strengthens intuition!

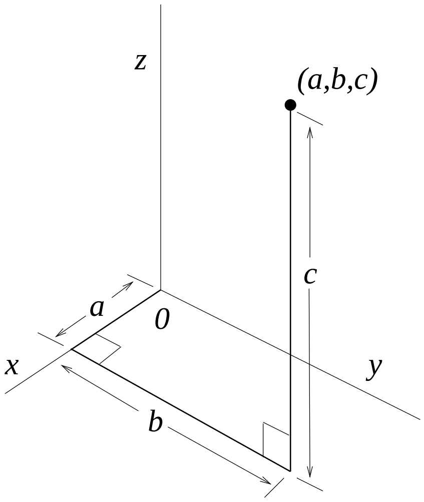
• *Some formulas to start with* •

Definition $\mathbb{R}^n$ is the collection of $n-$tuples of real numbers. A typical point in $\mathbb{R}^n$, P, has coordinates $(x_1, x_2, \ldots, x_n)$.

> Math is made up of definitions and deductions from the definitions, all structured to make sense of examples given by applications and by past work. It makes a great deal of sense to investigate the appropriateness of any definition, especially when seeking to link it to ideas already known!

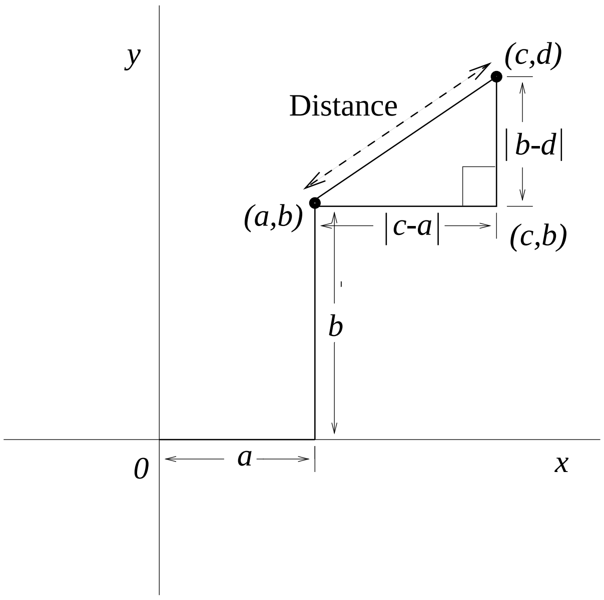
Discussion of the definition $\mathbb{R}^2$ is the Euclidean plane, as described to us by Descartes and similar thinkers: every point "is" a pair of numbers (a, b). The point can be located by moving a directed distance a along a first axis (the $x-$axis) and then moving perpendicular to that a directed distance b, as indicated below. $\mathbb{R}^3$ is three-dimensional space, and a point is a triple of numbers, (a, b, c). Such a point can be located using three mutually perpendicular axes as indicated in the picture below.



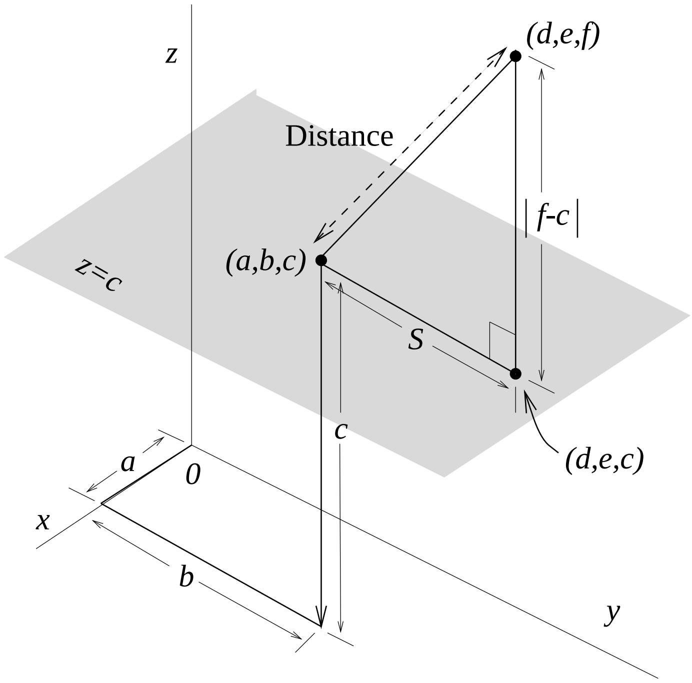


Definition If $P = (x_1, x_2, \ldots, x_n)$ and $Q = (y_1, y_2, \ldots, y_n)$ are two points in $\mathbb{R}^n$, then the *distance from P to Q* is $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$.

Discussion of the definition Of course if P and Q are points in $\mathbb{R}^2$, say $P = (a, b)$ and $Q = (c, d)$ respectively, then the distance we've known for years between P and Q is $\sqrt{(a-c)^2 + (b-d)^2}$. This "is" the distance because the Theorem of Pythagoras (see the picture below) combined with the observation that $\sqrt{(|c-a|)^2 + (|b-d|)^2} = \sqrt{(c-a)^2 + (b-d)^2}$ convince us that it is. It is Euclidean distance.



If P and Q are points of $\mathbb{R}^3$, with $P = (a,b,c)$ and $Q = (d,e,f)$, then a slightly more elaborate argument combined with a more complex picture is necessary. Here's the picture:



"Drop" a perpendicular from Q to the plane described by $z = c$. The perpendicular intersects the plane in the point whose coordinates at (d,e,c). The distance S in the diagram is a two-dimensional distance between points (a,b,c) and (d,e,c) given by $S = \sqrt{(e-b)^2 + (d-a)^2}$. The Theorem of Pythagoras again applies to give us the distance between P and Q as

$$\sqrt{S^2 + (|f-c|)^2} = \sqrt{(e-b)^2 + (d-a)^2 + (f-c)^2}$$

Notice first that we begin to run out of letters for two points after a few dimensions, so we should call the coordinates $x_1, x_2, \ldots$, etc. Second, pictures become difficult to draw (see Edward Abbot's *Flatland* for an implicit discussion about how to draw pictures in higher dimensions).

But these two geometrically reasonable demonstrations are supposed to convince you that the distance formula above is "correct": it is intuitive and obvious. It is only one of several candidates for distance in higher dimensions, but it is the one most often considered. The computation we are about to do, however, will show that some of the geometric intuition we have needs to be extended suitably. The intuition isn't wrong, but it needs more examples!

The *origin*, 0, in $\mathbb{R}^n$ has coordinates $(0, 0, \ldots, 0)$.

Definition The *unit ball in* $\mathbb{R}^n$ is the collection of points whose distance to the origin is less than or equal to 1. So $P = (x_1, x_2, \ldots, x_n)$ is in the ball if $x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2 \leq 1$.

The ball of radius R in n dimensions is defined analogously. $V_{n,R}$ will denote the n-dimensional volume of the ball of radius R. If c_n is the volume of the unit ball, then $V_{n,R}$ will be $c_n R^n$ because multiplication by R changes n-dimensional volume by a factor of R^n (which won't be hard to see from what follows). I want to identify the sequence of constants whose n^{th} term is c_n. Here are some volumes from a textbook:

n	$V_{n,R}$	c_n	Approx. value
1	$2R$	2	2.000
2	πR^2	π	3.142
3	$\frac{4}{3}\pi R^3$	$\frac{4}{3}\pi$	4.189
4	$\frac{1}{2}\pi^2 R^4$	$\frac{1}{2}\pi^2$	4.935
5	$\frac{8}{15}\pi^2 R^5$	$\frac{8}{15}\pi^2$	5.264
6	$\frac{1}{6}\pi^3 R^6$	$\frac{1}{6}\pi^3$	5.168

The entries for $n = 2$ and $n = 3$ are well-known. Some explanation might be needed for the others. The one-dimensional "ball" is the interval $[-R, R]$ whose length (or 1-dimensional "volume") is $2R$. The entry for dimension 4 was gotten earlier in class, and the entries for dimensions 5 and 6 were copied from a text. They will be verified below. The pattern of the entries is almost surely *not* clear. There are too many coincidences for small n (the "law of small numbers" [!]) and we'll need to be more systematic to discover what is happening.

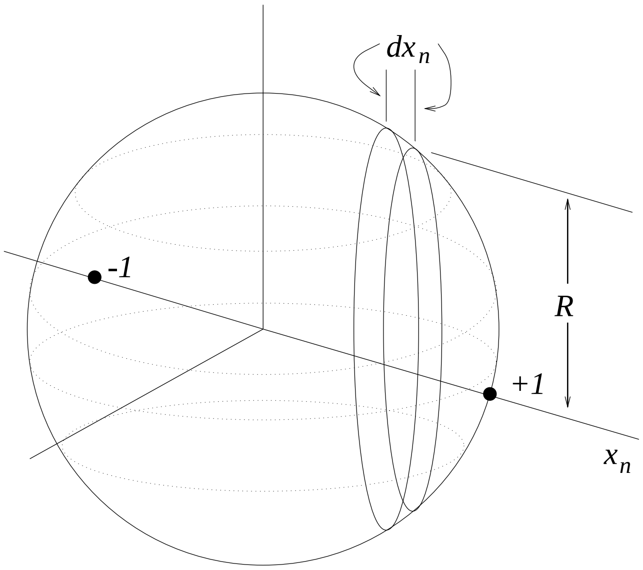
• *Volumes by slicing* •

Consider the inequality for the unit ball:

$$x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2 \leq 1$$

Since the left-hand side is a sum of squares, we see that x_n^2 must always be less than or equal to 1. That means x_n itself is in the interval $[-1, 1]$. What do $(n-1)$-dimensional cross-sections obtained by holding x_n fixed look like? Since $x_1^2+x_2^2+x_3^2+\ldots+x_{n-1}^2 \leq 1-x_n^2$

in such a slice, we see that the slice is an $(n-1)$-dimensional ball centered at the origin, with radius $R = \sqrt{1-x_n^2}$. Here's a "picture" of what the situation *might* look like:



The $(n-1)$-dimensional volume of the cross-section is $c_{n-1}R^{n-1}$ so the n-dimensional slice has approximate n-dimensional volume $c_n R^{n-1} dx_n$ and the total volume (inserting our formula for R) is

Equation **ONE** : $$c_n = \int_{x_n=-1}^{x_n=+1} c_{n-1}R^{n-1}\, dx_n = \int_{x_n=-1}^{x_n=+1} c_{n-1}\left(\sqrt{1-x_n^2}\right)^{n-1} dx_n$$

• ***Computing the integral (part 1)*** •

If we make the change of variables (substitution) $x_n = \sin\theta$ then

$$\begin{cases} dx_n \text{ becomes } \cos\theta\, d\theta \\ \sqrt{1-x_n^2} \text{ becomes } \cos\theta \\ x_n = 1 \text{ becomes } \theta = \dfrac{\pi}{2} \\ x_n = -1 \text{ becomes } \theta = -\dfrac{\pi}{2} \end{cases}$$

and equation **ONE** changes to the following equation:

Equation **TWO** : $$c_n = \left(\int_{-\pi/2}^{\pi/2} (\cos\theta)^n\, d\theta\right) c_{n-1}$$

which we'll now proceed to use. Note that **TWO** is valid when n is **an integer greater than 1** and that c_1 is $2R$. So when $n=2$ we get

$$c_2 = \left(\int_{-\pi/2}^{\pi/2} (\cos\theta)^2\, d\theta\right) c_1$$

so that, since $c_1 = 2$ and since $\int_{-\pi/2}^{\pi/2} (\cos\theta)^2\, d\theta = \frac{1}{2}\pi$ (using the double angle trick, as we have done several times in class), we can confirm that $c_2 = (\frac{1}{2})\pi \cdot 2 = \pi$. The next dimension is also fairly easy. We know that $\int_{-\pi/2}^{\pi/2} (\cos\theta)^3\, d\theta = \frac{4}{3}$ (a direct computation this integral with the substitution $u = \sin\theta$ and a trig identity) and we proved that $c_2 = \pi$. Therefore, **TWO** applies and $c_3 = \frac{4}{3} \cdot \pi = \frac{4}{3}\pi$, as expected.

Let's define I_n by the equation

$$I_n = \int_{-\pi/2}^{\pi/2} (\cos\theta)^n\, d\theta$$

since the integral will be important and referred to many times in this essay. Then **TWO** above translates to

Equation **THREE** : $\qquad c_n = I_n \cdot c_{n-1}$

which is valid for $n > 1$.

But let's try something a bit more challenging. We can try to compute c_5. By repeatedly using **THREE** we get the following sequence of equalities:

$$c_5 = I_5 \cdot c_4 = I_5 \cdot I_4 \cdot c_3 = I_5 \cdot I_4 \cdot I_3 \cdot c_2 = I_5 \cdot I_4 \cdot I_3 \cdot I_2 \cdot c_1$$

and now we stop. We know that $c_1 = 2$, and "all we need to do" to "compute" c_5 is to learn the value of the various integrals.

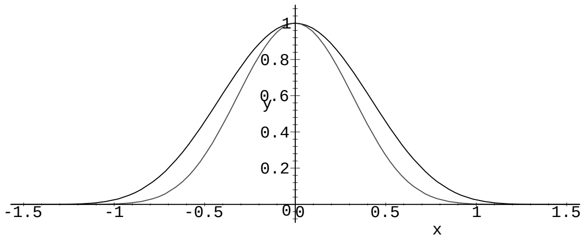
• ***A numerical interlude*** ... •

We can and will compute the integrals "exactly", but we can already get some numerical information which has rather disconcerting implications for the volumes of unit balls. Either hand computation or a short session with a computing device produces the following table of three decimal place numerical approximations to I_n for "low" n's (repeat: this is a table of values of integrals!):

n	I_n
1	2.000
2	1.571
3	1.333
4	1.178
5	1.067
6	.982

What can we learn? The numbers displayed are decreasing as n increases. The whole sequence of numbers $\{I_n\}$ is a decreasing sequence. This is actually not too hard to see. Remember that I_n is the integral of $(\cos x)^n$ over the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. In this interval, the values of cosine are between 0 and 1. Therefore the integral will be less than the area of the box whose base is π long and whose height is 1. But, in fact, the higher the power, the smaller the value of the function except at those places where cosine is either

1 ($x = 0$) or 0 ($x = \frac{\pi}{2}$ or $x = -\frac{\pi}{2}$). Higher powers make smaller integrals. Here is a graph of two powers of cosine over the interval. The larger one is $(\cos x)^6$, with integral (approximately) .982, and the smaller one is $(\cos x)^{10}$ with integral (approximately) .773 (picture and computations both produced by the computer program `maple`).



$(\cos x)^6$ and $(\cos x)^{10}$ on the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

These observations mean that for $n \geq 6$, the I_n's all will be less than .982, which means that for large n, many numbers less than .982 will be multiplied to get c_n. But powers of .982 decrease rapidly. Therefore, the sequence of c_n's must go to 0 very quickly. We can already see that the $n-$dimensional unit balls have *very* small volume when n is large! The strange unit balls are those in the first 5 dimensions, since they are the only ones whose volumes grow with n. After $n = 5$, the volume of the balls *shrinks* quite fast. We'll see how fast by doing precise computation of the I_n integrals, which in turn will allow us to precisely compute the c_n volumes.

• *Computing the integral (part 2)* •

We'll use integration by parts for the computation. The parts used aren't totally obvious, but they eventually yield a nice reduction formula.

$$\int_{-\pi/2}^{\pi/2} (\cos\theta)^n \, d\theta = (\cos\theta)^{n-1} \cdot \sin\theta \Big]_{-\pi/2}^{\pi/2} + (n-1)\int_{-\pi/2}^{\pi/2} (\cos\theta)^{n-2}(\sin\theta)^2 \, d\theta$$

$$\int u\,dv \quad = \quad uv \quad - \quad \int v\,du$$

$$\left.\begin{array}{r} u = (\cos\theta)^{n-1} \\ dv = \cos\theta \, d\theta \end{array}\right\}\left\{\begin{array}{l} du = (n-1)(\cos\theta)^{n-2}(-\sin\theta)\, d\theta \\ \quad v = \sin\theta \end{array}\right.$$

Note first that we get + in the actual application of integration by parts here because there are two minus signs, one from the integration by parts formula and one from the derivative of cosine. Also, the "penalty term" (what we pay for the privilege of exchanging the integrals – the stuff with the]) is actually 0 because $\sin\theta$ is 0 when θ is $\pm\pi/2$.

We can continue to massage the integral we've gotten by using $(\sin\theta)^2 = 1 - (\cos\theta)^2$ and by dropping the term which equals 0. We then get:

$$\begin{aligned} \int_{-\pi/2}^{\pi/2} (\cos\theta)^n \, d\theta &= (n-1)\int_{-\pi/2}^{\pi/2} (\cos\theta)^{n-2}\left(1 - (\cos\theta)^2\right)\, d\theta \\ &= (n-1)\int_{-\pi/2}^{\pi/2} (\cos\theta)^{n-2}\, d\theta - (n-1)\int_{-\pi/2}^{\pi/2} (\cos\theta)^n \, d\theta \end{aligned}$$

Now we can "solve" for c_n, the desired integral, by taking the second term on the right side of the equation to the left, adjusting the sign, and dividing by the resulting coefficient, n. We get:

$$\int_{-\pi/2}^{\pi/2} (\cos\theta)^n \, d\theta = \frac{n-1}{n} \int_{-\pi/2}^{\pi/2} (\cos\theta)^{n-2} \, d\theta$$

We can translate this:

$$I_n = \frac{n-1}{n} \, I_{n-2}$$

This reduction formula is valid if $n \geq 2$. We certainly know that $I_0 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos x)^0 \, dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 1 \, dx = \pi$ and $I_1 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos x)^1 \, dx = 2$. With these two values and the reduction formula we can exactly compute all of the I_n's. For example,

$$I_5 = \frac{4}{5} \cdot I_3 = \frac{4}{5} \cdot \frac{2}{3} \cdot I_1 = \left(\frac{4 \cdot 2}{5 \cdot 3}\right) 2$$

and

$$I_6 = \frac{5}{6} \cdot I_4 = \frac{5}{6} \cdot \frac{3}{4} \cdot I_2 = \frac{5}{6} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot I_0 = \left(\frac{5 \cdot 3 \cdot 1}{6 \cdot 4 \cdot 2}\right) \pi$$

These two computations provide examples of descriptions for I_n. The descriptions depend on the parity (even or odd) of n. Before I get bogged down in formulas, let's "check" by applying this to the computation of some volumes.

• ***The volume of unit balls*** •

We start with the formula we developed for dimension 5:

$$c_5 = I_5 \cdot I_4 \cdot I_3 \cdot I_2 \cdot c_1$$

Now let's replace the I_n's by their values:

$$c_5 = \overbrace{\left(\frac{4 \cdot 2}{5 \cdot 3}\right) 2}^{I_5} \cdot \overbrace{\left(\frac{3 \cdot 1}{4 \cdot 2}\right) \pi}^{I_4} \cdot \overbrace{\left(\frac{2}{3}\right) \cdot 2}^{I_3} \cdot \overbrace{\left(\frac{1}{2}\right) \pi}^{I_2} \cdot \overbrace{(2)}^{c_1} = \frac{8}{15}\pi^2$$

Let's check dimension 6:

$$c_6 = I_6 \cdot I_5 \cdot I_4 \cdot I_3 \cdot I_2 \cdot c_1$$

And replace each of the I_n's as before:

$$c_6 = \overbrace{\left(\frac{5 \cdot 3 \cdot 1}{6 \cdot 4 \cdot 2}\right) \pi}^{I_6} \cdot \overbrace{\left(\frac{4 \cdot 2}{5 \cdot 3}\right) 2}^{I_5} \cdot \overbrace{\left(\frac{3 \cdot 1}{4 \cdot 2}\right) \pi}^{I_4} \cdot \overbrace{\left(\frac{2}{3}\right) \cdot 2}^{I_3} \cdot \overbrace{\left(\frac{1}{2}\right) \pi}^{I_2} \cdot \overbrace{(2)}^{c_1} = \frac{1}{6}\pi^3$$

So we agree with the textbook answers for dimensions 5 and 6. In addition, if we examine the structure of the answers, we can see patterns which will clearly persist in general. Here the phrase "will clearly persist in general" is a substitution for "can be proved by mathematical induction."

Let's look at even n first to discover the general formula. In the answer for c_6 above, all of the fractions cancel except for

$$\frac{1}{6\cdot 4\cdot 2} = \frac{1}{(3\cdot 2\cdot 1)\cdot(2\cdot 2\cdot 2)} = \frac{1}{3!\cdot 2^3}$$

The π's and 2's alternate: there are three of each. Note that $6 = 2\cdot 3$. This suggests the following formula, which is correct:

$$\text{If } n \text{ is even, then } c_n = \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2}\right)!}$$

Odd dimensions seem a bit more complicated. Let's look at c_5 again, and group terms suggestively.

$$c_5 = \left(\frac{4\cdot 2}{5\cdot 3}\right)2\cdot\left(\frac{3\cdot 1}{4\cdot 2}\right)\pi\cdot\left(\frac{2}{3}\right)\cdot 2\cdot\left(\frac{1}{2}\right)\pi\cdot(2) = \left(\frac{2\cdot 2\cdot 2}{5\cdot 3\cdot 1}\right)\pi^2 = \frac{1}{\left(\frac{5}{2}\right)\cdot\left(\frac{3}{2}\right)\cdot\left(\frac{1}{2}\right)}\cdot\pi^{\frac{4}{2}}$$

My aim here is a simple formula for c_5. What's the bottom of the fraction look like? $\left(\frac{5}{2}\right)\cdot\left(\frac{3}{2}\right)\cdot\left(\frac{1}{2}\right)$ seems almost to be the beginning of a factorial – a collection of numbers decreasing by 1 and being multiplied. But we would have to consider factorials of *half* integers. There is a generally accepted interpolation to non-integer values of the factorial function. If you're willing to willing to accept (temporarily) the fact that

$$\left(\tfrac{1}{2}\right)! = \left(\tfrac{1}{2}\right)\cdot\sqrt{\pi}$$

that is, the factorial of $\frac{1}{2}$ is or should be $\left(\frac{1}{2}\right)\cdot\sqrt{\pi}$, then "clearly" $\left(\frac{5}{2}\right)! = \left(\frac{5}{2}\right)\cdot\left(\frac{3}{2}\right)! = \left(\frac{5}{2}\right)\cdot\left(\frac{3}{2}\right)\cdot\left(\frac{1}{2}\right)! = \left(\frac{5}{2}\right)\cdot\left(\frac{3}{2}\right)\cdot\left(\frac{1}{2}\right)\cdot\sqrt{\pi}$. Multiply the formula for c_5 by $\sqrt{\pi}$ on both the top and bottom:

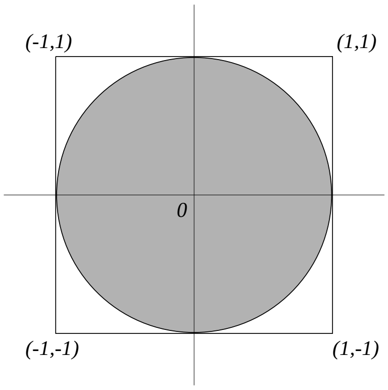
$$c_5 = \frac{1}{\left(\frac{5}{2}\right)\cdot\left(\frac{3}{2}\right)\cdot\left(\frac{1}{2}\right)}\cdot\pi^{\frac{4}{2}}\cdot\frac{\sqrt{\pi}}{\sqrt{\pi}} = \frac{\pi^{\frac{5}{2}}}{\left(\frac{5}{2}\right)!}$$

I will need to show you how to find factorials of non-integer positive real numbers, and I must verify the result I quoted above at that time. If you're willing to accept this future result on factorial interpolation, we get a neat formula, valid in all dimensions:

$$\text{If } n \text{ is } \textbf{any} \text{ positive integer, then } c_n = \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2}\right)!}$$

• ***A ball in a box*** •

The box just containing the unit circle in $\mathbb{R}^2$ is the square with corners $(1,1)$, $(-1,1)$, $(1,-1)$, and $(-1,-1)$.

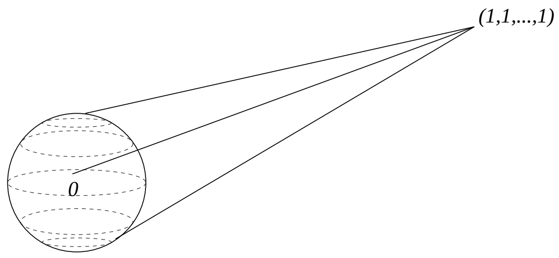


A deceptive two-dimensional picture

The proportion of the box occupied by the ball is $\frac{\pi}{4}$, approximately 78.5%. In $\mathbb{R}^3$, the box just containing the ball has 8 corners, all those points whose coordinates are either $+1$ or -1. The volume of that box is $2^3 = 8$, and the proportion of the box occupied by the unit ball is $\left(\frac{4}{3}\pi\right)/8$, about 52.4%. But remember that low dimensions are very exceptional in what we're doing here, indeed, they are almost weird!

In dimension n, the box just containing the unit ball has 2^n corners (all the points whose n coordinates are either $+1$ or -1), and every side has length 2. So the n-dimensional volume of this box is 2^n. The volume of the ball is given by the formula above. The top of the ball formula is $\pi^{\frac{n}{2}}$ which is $\left(\sqrt{\pi}\right)^n$, approximately $(1.772)^n$. Therefore the top grows more slowly than 2^n all by itself. But the factorial "downstairs" in the ball formula makes the proportion of the box occupied by the ball get small *very* quickly.

A example computed with the help of `maple` may illustrate this more concretely. The volume of the 56-dimensional ball is reported to be approximately $2.729 \cdot 10^{-16}$. The 56-dimensional box just enclosing this ball has 56-dimensional volume 2^{56}, which is approximately $7.206 \cdot 10^{16}$. The quotient of the volume of the ball divided by the volume of the box is approximately $3.788 \cdot 10^{-33}$, a *really* tiny number! What's happening "geometrically" and how can we hope to understand it? Here are some suggestions. A typical "corner" of the box is a point with coordinates $(\pm1, \pm1, \ldots, \pm1)$. The distance from the origin to this corner is $\sqrt{(\pm1)^2 + (\pm1)^2 + \ldots + (\pm1)^2} = \sqrt{56} > 7$ (since the coordinates of a point are a 56-tuple). The corners are very far from the origin, more than 7 times the radius of the unit ball, while the ball stays very close since points in the ball have distance ≤ 1 to the origin. Not only are the corners of the box far away, but there are many of them (2^{56}, in fact). So our low dimensional intuition has not seen enough examples. More directly: dimensions 2 and 3 are exceptional. Below is a poetic picture, an "impression", of part of what the ball in the box *might* look like if we had better perception. I tried to show only one corner and the ball. Of course, in the picture the ball is much too big and the corner is much too close!



A vision of higher dimensional truth (?)

As I mentioned in class, I first saw these computations significantly used in an electrical engineering text. Let me give a a brief and much simplified version of the discussion I read. The question of how to analyze "random" strings of signals was being explored. Each signal had "strength" between -1 and $+1$. One possible measure of the total strength of an n long string of signals was the square root of the sum of the squares of the strengths – our distance to the origin. In this context, a signal string of length 56 doesn't seem excessive at all to me! The analysis I described above was done in just a few lines of the text (many implied uses of "clearly" were needed!) and the notion of total signal strength gotten from Euclidean distance (the square root of a sum of squares) was then abandoned. An alternative measure of strength derived by considering boxes instead of balls was proposed and used: measure the strength of the signal by taking the maximum of the absolute values of the individual signals. In our language, the recommendation was to discard $\sqrt{\sum_{j=1}^{n}(x_j)^2}$ as a measure of the strength of the entire signal and to use $\max\{|x_j|, 1 \leq j \leq n\}$ instead.

• *Computing π strangely: an unexpected reward* •

We know I_5 and I_6:

$$I_5 = \left(\frac{4 \cdot 2}{5 \cdot 3}\right) 2$$

$$I_6 = \left(\frac{5 \cdot 3 \cdot 1}{6 \cdot 4 \cdot 2}\right) \pi$$

And here are I_7 and I_8:

$$I_7 = \left(\frac{6 \cdot 4 \cdot 2}{7 \cdot 5 \cdot 3}\right) 2$$

$$I_8 = \left(\frac{7 \cdot 5 \cdot 3 \cdot 1}{8 \cdot 6 \cdot 4 \cdot 2}\right) \pi$$

We also know that the integrals of higher powers are smaller:

$$I_8 \leq I_7 \leq I_6$$

Let's look at the values of just these definite integrals:

$$\left(\frac{7 \cdot 5 \cdot 3 \cdot 1}{8 \cdot 6 \cdot 4 \cdot 2}\right) \pi \leq \left(\frac{6 \cdot 4 \cdot 2}{7 \cdot 5 \cdot 3}\right) 2 \leq \left(\frac{5 \cdot 3 \cdot 1}{6 \cdot 4 \cdot 2}\right) \pi$$

Now "simplify" the first two expressions, putting $\frac{\pi}{2}$ on the left-hand side:

UP $$\frac{\pi}{2} \leq \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdot 8}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7\cdot 7}$$

Now "simplify" the last two expressions, putting $\frac{\pi}{2}$ on the right-hand side:

DOWN $$\frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7} \leq \frac{\pi}{2}$$

Incidentally, if you don't believe this, `maple` reports that the overestimate of $\frac{\pi}{2}$ shown in **UP** is approximately 1.672, and the underestimate of $\frac{\pi}{2}$ shown in **DOWN** is approximately 1.463; note that 1.571 is a three-decimal place approximation of $\frac{\pi}{2}$.

The quotient of the **UP** overestimate divided by the **DOWN** underestimate is

$$\frac{\frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdot 8}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7\cdot 7}}{\frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7}} = \frac{8}{7}$$

We can extend this reasoning by looking at the formulas for I_{2n} and I_{2n+1} and I_{2n+2}. The discrepancy between the over– and underestimates will be $\frac{2n+2}{2n+1}$ which surely approaches 1 as n gets large. Therefore [!!!],

$$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots$$

This "infinite product formula" is attributed to John Wallis (1616-1703). It converges very slowly to its limit. This isn't the formula used by the well-publicized computations which get billions of digits of π!

• *Interpolating factorials* •

Here's a simple integral which we can compute and then complicate:

$$\int_0^A e^{-x}dx = -e^{-x}\Big]_0^A = -e^{-A} - (-e^0) = 1 - e^{-A}$$

If we let $A \to \infty$, then $e^{-A} \to 0$ so the improper integral $\int_0^\infty e^{-x}dx$ converges, and it has value 1. We'll use this in combination with a reduction formula. If n is a positive integer consider:

$$\int_0^A x^n e^{-x}dx = -x^n e^{-x}\Big]_0^A + n\int_0^A x^{n-1}e^{-x}dx$$

$$\int u\,dv \quad = \quad uv \quad - \quad \int v\,du$$

$$\left.\begin{matrix} u = x^n \\ dv = e^{-x}dx \end{matrix}\right\}\left\{\begin{matrix} du = nx^{n-1}dx \\ v = -e^{-x} \end{matrix}\right.$$

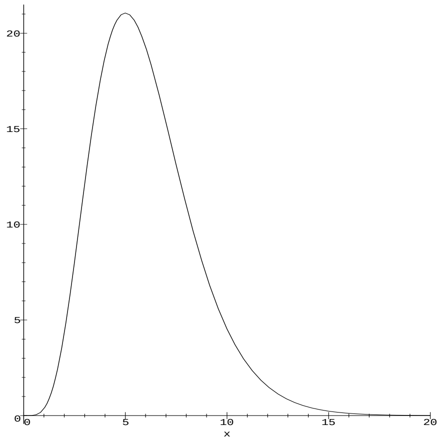
Again two $-$'s make a $+$: the $-$ from v and the $-$ from the integration by parts formula create a $+$ in our formula. Since n is a positive integer, both $x^n e^{-x}$ and $x^{n-1}e^{-x}$ are continuous in $[0, A]$. There is no "singularity" at 0. The lower limit of the] term contributes nothing, since $0^n e^{-0}$ is 0 if n is a positive integer.

What happens as $A \to \infty$? The upper limit of the] term gives $-A^n e^{-A}$. This has limit 0 since the exponential function grows faster than *any* power (just apply l'Hôpital's Rule n times).

If $\int_0^\infty x^{n-1}e^{-x}dx$ exists, then this argument shows that $\int_0^\infty x^n e^{-x}dx$ exists, and that $\int_0^\infty x^n e^{-x}dx = n\int_0^\infty x^{n-1}e^{-x}dx$. We know the value of this integral when $n = 0$ by the computation above. For larger integer n we now see

$$n! = \int_0^\infty x^n e^{-x}dx$$

when n is a positive integer.



$x^5 e^{-x}$ in the interval $[0, 20]$

Some simple facts about $x^n e^{-x}$ on $[0, \infty)$: it is always non-negative and is 0 only at 0. Its limit as $x \to \infty$ is 0. It has one critical point, a maximum, at $x = n$ where its value is $\left(\frac{n}{e}\right)^n$. It has two inflection points located at $n \pm \sqrt{n}$.

If n is positive, then the integral $\int_0^\infty x^n e^{-x}dx$ converges because when $x \geq 1$ we can compare it to the integral of a higher integer power which we already know converges. If we agree to make the following **definition**:

$$\text{If } n \geq 0 \text{ then } n! = \int_0^\infty x^n e^{-x}dx$$

then we first observe that this definition agrees with the usual one when n is a positive integer. Also, the integration by parts argument applies to show that $n! = n \cdot (n-1)!$ when $n \geq 1$.

One specific factorial value is

$$\left(\tfrac{1}{2}\right)! = \int_0^\infty \sqrt{x}\, e^{-x}dx$$

What is this number? We can try to approximate it by splitting the integral up into an infinite "tail" which is small and standard definite integral which can be approximated numerically.

$$\int_0^\infty \sqrt{x}\, e^{-x} dx = \int_0^{20} \sqrt{x}\, e^{-x} dx + \int_{20}^\infty \sqrt{x}\, e^{-x} dx$$

The infinite tail can be overestimated, and then the larger integral can be explicitly computed with one use of the reduction formula (the steps involving limits evaluating improper integrals are omitted):

$$\int_{20}^\infty \sqrt{x}\, e^{-x} dx < \int_{20}^\infty x e^{-x} dx = -xe^{-x} - e^{-x}\Big]_{20}^\infty = \frac{21}{e^{20}} \approx 4.238 \cdot 10^{-8}$$

Dropping the infinite tail will give an error which won't affect the first seven decimal places of the whole integral. Approximation by `maple` tell us that $\int_0^{20} \sqrt{x} e^{-x} dx$ is about .8862269 so therefore

$$\int_0^\infty \sqrt{x} e^{-x} dx \approx .8862269$$

But $(.8862269)^2 \cdot 4 = 3.141592473$ is interestingly near π. If what this suggests is true then we should believe that $\left(\frac{1}{2}\right)! = \left(\frac{1}{2}\right) \cdot \sqrt{\pi}$ as was declared in the ball computation. I'd like to prove this now.

An obvious preliminary substitution makes $\int_0^\infty \sqrt{x} e^{-x} dx$ slightly easier to work with: $x = s^2$ so $dx = 2s\, ds$. The integral then becomes $\int_0^\infty 2s^2\, e^{-s^2} ds$. Another application of integration by parts will turn this into one of the most famous integrals in mathematics. I'll be careful, and regard the improper integral as a limit of the proper integrals below:

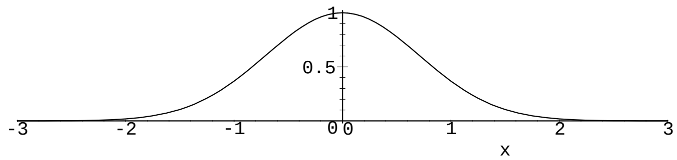
$$\begin{array}{ccccc} \displaystyle\int_0^A 2s^2\, e^{-s^2} ds & = & s \cdot -e^{-s^2}\Big]_0^A & + & \displaystyle\int_0^A e^{-s^2} ds \\ \displaystyle\int u\, dv & = & uv & - & \displaystyle\int v\, du \end{array}$$

$$\left.\begin{array}{r} u = s \\ dv = e^{-s^2} 2s\, ds \end{array}\right\} \left\{\begin{array}{r} du = ds \\ v = -e^{-s^2} \end{array}\right.$$

The] term is 0 when $s = 0$ and has limit 0 as $A \to \infty$ since e^{-A^2} dies off much more quickly than A grows. Our original integral has now been changed to $\int_0^\infty e^{-s^2} ds$. We'll double it using the evenness of s^2 and instead try to show that

$$\int_{-\infty}^\infty e^{-s^2} ds = \sqrt{\pi}$$

The function e^{-s^2} arises everywhere in probability and statistics. The presence of so many π's in statistical formulas is because of the value of this integral. Here's a picture of this famous "bell-shaped" curve:

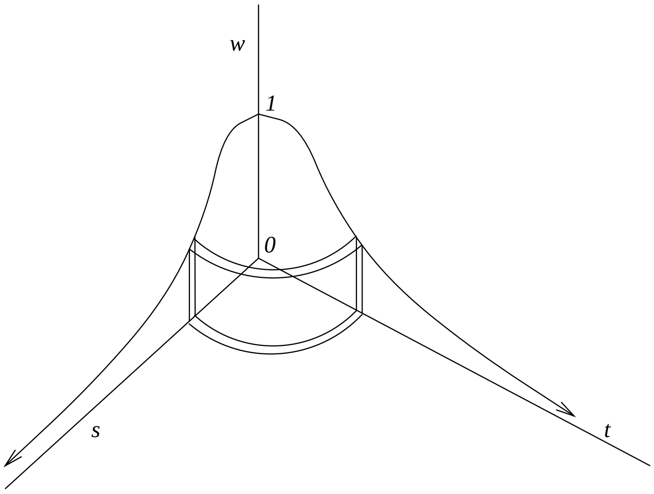


It has a maximum at 0 and inflection points at $\pm\frac{1}{\sqrt{2}}$.

There are many ways to conclude that $I = \int_{-\infty}^{\infty} e^{-s^2}\,ds$ is $\sqrt{\pi}$. Here is one of them.

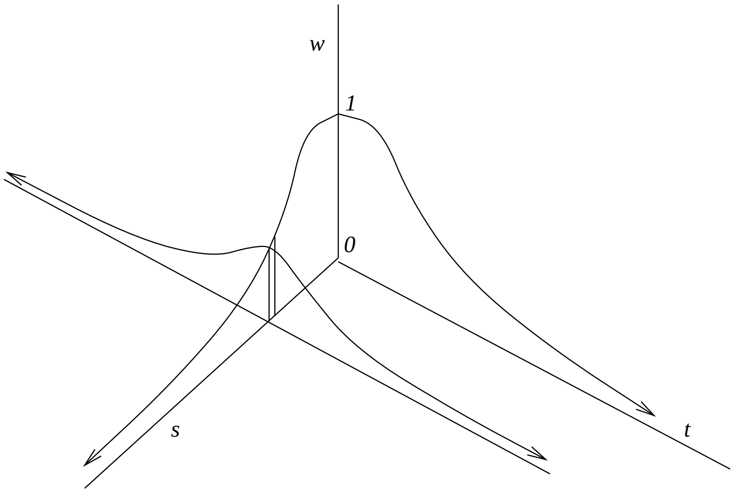
Revolve the curve $w = e^{-s^2}$ for $s \geq 0$ about the $w-$axis. The resulting solid of revolution has a volume, V, which can be computed with "thin shells".

$$V = \int_0^{\infty} 2\pi s e^{-s^2}\,ds = 2\pi \cdot -\frac{1}{2}e^{-s^2}\Big]_0^{\infty} = \pi$$

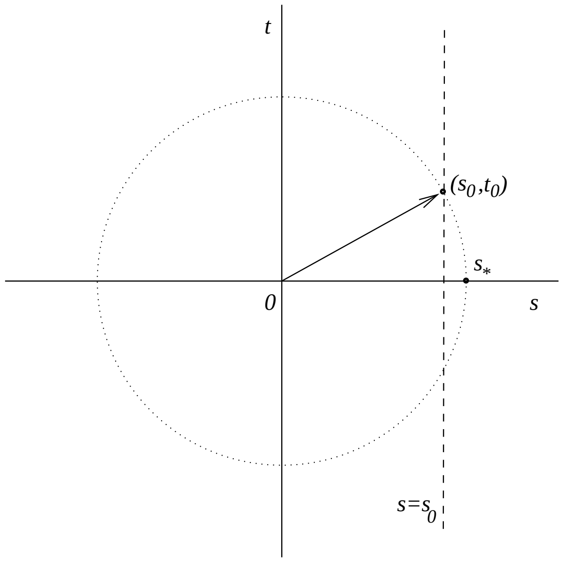


A quarter of the revolved curve, with a vertical strip ds thick forming a thin shell

Now let's compute V another way. We can slice the solid by planes perpendicular to the $s-$axis. We'll get a profile curve which will vary with s. A piece of the volume will be obtained by multiplying that profile curve by ds. What *is* the profile curve? A picture may help.



The profile curve looks a great deal like the original bell curve rescaled. This is correct, as the following more precise analysis shows. What's the height of the revolved curve over a point (s_0, t_0)? The height depends only on the distance of the point to the origin because the surface is composed of circles centered at $(0, 0)$. The height at (s_0, t_0) is the same height as the curve $w = e^{-s^2}$ has at $s_* = \sqrt{(s_0)^2 + (t_0)^2}$. Here is a picture looking *down* on the surface.



So the height of the surface at (s_0, t_0) is $e^{-\left(\sqrt{(s_0)^2+(t_0)^2}\,\right)^2} = e^{-(s_0)^2-(t_0)^2}$. The curve over the line $s = s_0$ has height over the point (s_0, t) the value $e^{-(s_0)^2-t^2} = e^{-(s_0)^2} \cdot e^{-t^2}$. Therefore the profile curve is the original curve rescaled: it is multiplied by $e^{-(s_0)^2}$. The area must be I, the original integral we want to compute, multiplied by $e^{-(s_0)^2}$. Let's move from the fixed cross-section at s_0 to any cross-section. The volume slice we want, which is the cross-sectional area times ds, will be $I \cdot e^{-s^2} ds$. We'll need to add these pieces of volume up from $s = -\infty$ to $s = \infty$. Therefore

$$V = \int_{-\infty}^{\infty} I \cdot e^{-s^2} ds = I \cdot \int_{-\infty}^{\infty} e^{-s^2} ds = I^2$$

Since we already know that $V = \pi$, we've confirmed that $I = \sqrt{\pi}$.

This rather involved demonstration of the value of I is an adaption of another way of computing I, using double integrals and polar coordinates. This method and others rely on the multiply/add property of the exponential function: $e^a \cdot e^b = e^{a+b}$. One of my colleagues says the tricks involved in evaluating this integral are almost "criminal acts" because they are so special and so clever.

History and notation This integral interpolation for factorials was discovered and investigated long ago. The resulting function was named the Gamma function, and the definition was given a shift in its argument. The Gamma function, $\Gamma(n)$ is defined by

$$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$$

so that $\Gamma(n) = (n-1)!$ when n is a positive integer. Please argue with Leonhard Euler (1707–1783) if you're unhappy about the Gamma function.

• ***How big are factorials?*** •

A silicon-based acquaintance of mine reports that

$$20! = 2432\ 90200\ 81766\ 40000$$

and that

$$50! = 30414\ 09320\ 17133\ 78043\ 61260\ 81660\ 64768\ 84437\ 76415\ 68960\ 51200\ 00000\ 00000$$

and also remarks that 100! has 158 digits. How "big" are these numbers? That is, can we describe how fast they grow in terms of other, more traditional functions? The factorial function does grow very fast, and in many applications having some idea of how fast it grows is quite useful. The following is true and is called **Stirling's Formula**:

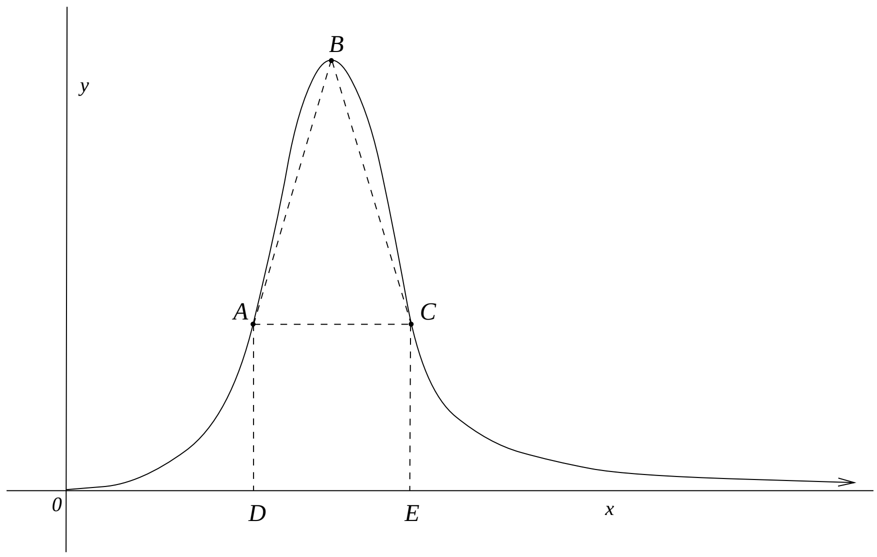
$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

For example, this approximation when $n = 20$ is about $2.4228 \cdot 10^{18}$ and when $n = 50$ it is about $3.0363 \cdot 10^{64}$. You can compare these numbers to the exact values above. Stirling's formula gives an underestimate for 20 and an overestimate for 50.

Where does such a result come from? Since we know that

$$n! = \int_0^\infty x^n e^{-x}\,dx$$

when n is a positive integer, we can use a picture to supply some evidence to make Stirling's formula credible. Here is a rather stylized picture of $x^n e^{-x}$:



B is the maximum of the curve, and has coordinates $(n, \left(\frac{n}{e}\right)^n)$. A and C are at the inflection points, and their coordinates are, respectively, $(n - \sqrt{n}, \text{something})$ and $(n + \sqrt{n}, \text{something else})$.

The picture shows $ACED$ as a rectangle. This isn't correct because AD and CE aren't equal, but their quotients do have limit 1 as $n \to \infty$. It's an asymptotic rectangle.

The total area is underestimated by the sum of the areas of $\triangle ABC$ plus the area of (almost) rectangle $ACED$. This is about $\sqrt{n} \cdot$ height of $B = \sqrt{n}\left(\frac{n}{e}\right)^n$.

You can already see some parts of Stirling's formula creeping in. In particular, you can already see that the factorial function (the Gamma function) is superexponential. A careful study of the integral will, in fact, disclose every detail and even information about the error in this approximation. You should read about this yourself: I'm tired!

Contents