SOME GEOMETRY IN HIGH-DIMENSIONAL SPACES

MATH 527A

1. Introduction

Our geometric intuition is derived from three-dimensional space. Three coordinates suffice. Many objects of interest in analysis, however, require far more coordinates for a complete description. For example, a function f with domain $[-1, 1]$ is defined by infinitely many "coordinates" $f(t)$, one for each $t \in [-1, 1]$. Or, we could consider f as being determined by its Taylor series $\sum_{n=0}^{\infty} a_n t^n$ (when such a representation exists). In that case, the numbers $a_0, a_1, a_2, \ldots$ could be thought of as coordinates. Perhaps the association of Fourier coefficients (there are countably many of them) to a periodic function is familiar; those are again coordinates of a sort.

Strange Things can happen in infinite dimensions. One usually meets these, gradually (reluctantly?), in a course on Real Analysis or Functional Analysis. But infinite dimensional spaces need not always be completely mysterious; sometimes one lucks out and can watch a "counterintuitive" phenomenon developing in $\mathbb{R}^n$ for large n. This might be of use in one of several ways: perhaps the behavior for large but finite n is already useful, or one can deduce an interesting statement about $\lim_{n\to\infty}$ of something, or a peculiarity of infinite-dimensional spaces is illuminated.

I will describe some curious features of cubes and balls in $\mathbb{R}^n$, as $n \to \infty$. These illustrate a phenomenon called *concentration of measure*. It will turn out that the important *law of large numbers* from probability theory is just one manifestation of high-dimensional geometry.

Along the way, we will meet some standard analysis techniques. These may be familiar to varying degrees. I think it could be useful for you to see how multidimensional integrals, linear algebra, estimates, and asymptotics appear in the analysis of a concrete problem. A number of these matters are relegated to appendices. In the main part of the exposition, I try to focus (as much as possible) on geometric phenomena; it would be reasonable to read about those first, and only refer to an appendix for quick clarification as needed. Ultimately, you should also understand the appendices.

2. The Cube

2.1. **Volume of the cube.** $C^n(s)$ is the cube centered at the origin in $\mathbb{R}^n$ with sidelength $2s$. I.e.,

$$C^n(s) = \{(x_1, \ldots, x_n) \mid -s \leq x_j \leq s \text{ for all } j\}.$$

Its n-dimensional volume is (by definition)

$$\mathrm{Vol}(C^n(s)) = \underbrace{(2s) \times (2s) \times \cdots \times (2s)}_{n \text{ times}} = (2s)^n.$$

We have an obvious consequence:

Proposition 2.1. *As $n \to \infty$, the volume of $C^n(s)$ tends to zero if $s < \frac{1}{2}$, to ∞ if $s > \frac{1}{2}$, and it is always $= 1$ for $s = \frac{1}{2}$.*

From now on, $C^n(\frac{1}{2})$ will be my reference cube, and I will simply write C^n. So $\mathrm{Vol}(C^n) = 1$ for all n. Notice, however, that the point $(\frac{1}{2}, \ldots, \frac{1}{2})$ is a vertex of the cube, and it has distance $\sqrt{n}/2$ from the origin. So

Proposition 2.2. *The cube C^n has diameter $\sqrt{n}$, but volume 1.*

The mathematics is completely elementary, but I hope you agree that visualizing such behavior is rather more tricky. It gets worse.

2.2. **Concentration of volume.** I want to compare the volume of C^n to the volume of a subcube, $C^n(\frac{1}{2} - \frac{\epsilon}{2})$, where $\epsilon \in (0, 1)$ is given. We already know from Proposition 2.1 that the latter volume tends to zero as $n \to \infty$. Hence the "shell" between the two cubes contains most of the volume:

Proposition 2.3. *For every $\epsilon \in (0, 1)$,*

$$\textit{Volume of shell} = \mathrm{Vol}(C^n - C^n(\frac{1}{2} - \frac{\epsilon}{2}))$$

tends to 1 as $n \to \infty$.

In other words, as $n \to \infty$, there is no such thing as "a thin shell of small volume". *All* the volume of C^n escapes towards its surface, out of any prescribed subcube.

In order to understand, in some measure, how the volume concentrates at the surface of C^n, look again at

$$\mathrm{Vol}(C^n - C^n(\frac{1}{2} - \frac{\epsilon}{2})) = 1 - (1 - \epsilon)^n. \tag{2.1}$$

Of course, $(1-\epsilon)^n \to 0$ because $0 < 1-\epsilon < 1$. Now invoke Lemma A.1:

$$\lim_{n\to\infty}(1+\frac{a}{n})^n = e^a.$$

This suggests that we should let the ϵ in (2.1) vary with n. Instead of taking a subcube $C^n(\frac{1}{2}-\frac{\epsilon}{2})$ whose sides are in fixed ratio to the side (length = 1) of C^n, we expand the subcube, and shrink the shell, as n increases. In this way, the volume "trapped" in the shrinking shell does have a nonzero limit:

Proposition 2.4. *For every $t > 0$,*

$$\lim_{n\to\infty}\mathrm{Vol}(C^n - C^n(\frac{1}{2}-\frac{t}{2n})) = \lim_{n\to\infty} 1-(1-\frac{t}{n})^n = 1-e^{-t}$$

. [It is understood that n must be large enough to ensure $\frac{t}{2n} < \frac{1}{2}$].

How to think about this? Say you want to know what shell contains half the volume of C^n (for large n, of course). Since $1-e^{-t} = .5$ for $t = .69315\ldots$, you know that the cube of sidelength $1-\frac{.69\ldots}{2n}$ has volume about $\frac{1}{2}$, with the remaining volume $\frac{1}{2}$ contained in the very thin shell (of width $\frac{.69\ldots}{2n}$) between the subcube and our reference cube, C^n.

Later on, we will look at the cube a bit differently, and see that its volume *also* concentrates along the diagonal plane

$$x_1 + \cdots + x_n = 0.$$

This will be related to probability theory. First, however, we will study the ball; it is more symmetrical than the cube, and easier to deal with.

2.3. **Surface area of the cube.** In dimensions 1 and 2, "cubes" are not *cubes* in the everyday sense of the word. The "cube" $C^1(s)$ is a line segment of length $2s$. The "cube" $C^2(s)$ is a square, sidelength $2s$. It has four 1-dimensional sides, each of which is a copy of $C^1(s)$. The surface area of $C^2(s)$ is the total 1-dimensional "volume", in this case just ordinary length, of the four sides, to wit $4\times(2s)$.

$C^3(s)$ *is* the usual cube. Its six 2-dimensional sides are copies of $C^2(s)$, and the surface area of $C^3(s)$ is the sum of the 2-dimensional volumes, or the usual areas, of its sides: $6\times(2s)^2$.

The cube $C^4(s)$ in four dimensions has a number of 3-dimensional sides. Let's call them 3-*faces*. The "surface area" of $C^4(s)$ is really the sum of the 3-dimensional volumes of the 3-faces: $8\times(2s)^3$.

To see why this formula is correct, we need to describe the 3-faces. A 3-face is determined by setting one of the coordinates equal to its extreme value, $\pm s$, and allowing the other coordinates to vary in $[-s, s]$.

For example,

$$\{(s, x_2, x_3, x_4) \mid |x_j| \le s \text{ for } j = 2, 3, 4\}$$

is a 3-face. It is a copy of the cube

$$C^3(s) = \{(x_1, x_2, x_3) \mid |x_j| \le s \text{ for } j = 1, 2, 3\};$$

except the indices have changed (which is irrelevant). This object has the 3-dimensional volume $(2s)^3$. There are eight 3-faces, since any one of the four coordinates could be fixed at $+s$ or $-s$, with the other three coordinates allowed to vary in $[-s, s]$.

The "surface area" of $C^n(s)$ is the sum of the $(n-1)$-dimensional volumes of its $(n-1)$-faces. "Area" is not really a good term [1], but we will use it to distinguish between the n-dimensional volume of the solid n-dimensional object, and the $(n-1)$-dimensional volume of the boundary of the object. Later on, the same convention will apply to the ball.

Exercise 2.1. How many $(n-1)$-faces does $C^n(s)$ have? How many $(n-k)$-dimensional faces, for $0 \le k \le n$? (A 0-face is a vertex, and the n-face, by convention, is the whole solid cube). □

Exercise 2.2. For each $s \in (0, \infty)$, compare the behavior of the volume of $C^n(s)$ to the behavior of its surface area, as $n \to \infty$. The case $s = \frac{1}{2}$ should strike you as being counterintuitive. □

Exercise 2.3. Fix $\epsilon \in (0, s)$. Let S_ϵ be the slice

$$S_\epsilon = C^n(s) \cap \{(x_1, \ldots, x_{n-1}, z) \mid |z| < \epsilon\},$$

i.e. a sort of "equatorial slice" of the cube. How does the ratio

$$\frac{\mathrm{Vol}(S_\epsilon)}{\mathrm{Vol}(C^n(s))}$$

behave as $n \to \infty$? □

3. Volume of the Ball

The material about concentration of volume comes (at times verbatim) from P. Lévy, *Problèmes concrets d'analyse fonctionelle*, Gauthier-Villars (1951), p.209 ff. See also pp. 261-262 of the Math 527 notes for the computation of the volume of a ball in $\mathbb{R}^n$.

[1] something like "$(n-1)$-dimensional Lebesgue measure" might be preferred by cognoscenti

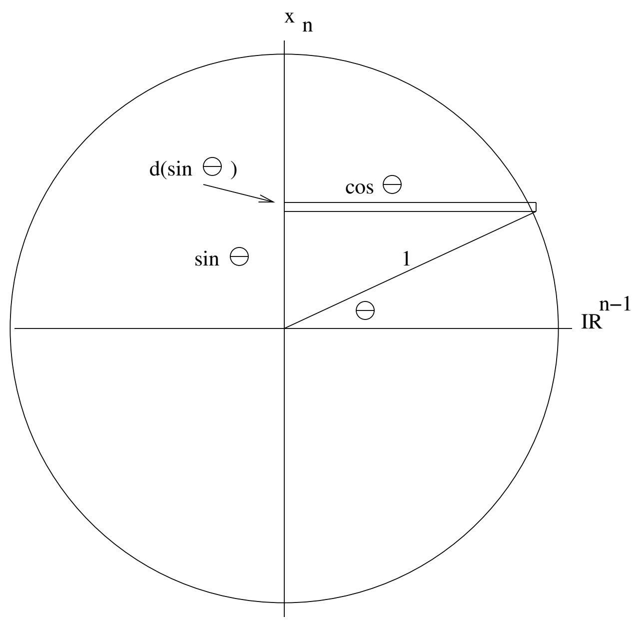

FIGURE 1. Volume of ball by slices

3.1. **Volume of the ball, part 1.** We write $B^n(R)$ for the *solid* ball, radius R, centered at the origin in $\mathbb{R}^n$. The term "ball" means "surface plus interior". The surface itself is called the "sphere", denoted by $S^{n-1}(R)$. The superscript $(n-1)$ signifies that the sphere *in* $\mathbb{R}^n$ has dimension $(n-1)$. Thus

$$B^n(R) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots x_n^2 \leq R\},$$
$$S^{n-1}(R) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots x_n^2 = R\}.$$

Example 3.1. To be clear about dimensions, consider

$$B^4(1) = \{(x_1, x_2, x_3, x_4) \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 \leq 1\} \subset \mathbb{R}^4.$$

This is 4-dimensional, in the sense that as long as $x_1^2 + x_2^2 + x_3^2 + x_4^2 < 1$, all four coordinates can be varied independently. However,

$$S^3(1) = \{(x_1, x_2, x_3, x_4) \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\} \subset \mathbb{R}^4.$$

This is 3-dimensional, in the sense that once three of the x_j are fixed, the remaining one is largely determined (up to ± sign, in this example). There are better definitions, but this should convey the idea. It is rather tricky to picture S^3, since we can't go into four dimensions, but there are ways. In any case, this section deals with the ball only. □

Now let K_n be the volume of the unit ball $B^n(1)$ in $\mathbb{R}^n$. Then the volume of $B^n(R)$ is $K_n R^n$. Let us calculate K_n (refer to Figure 1).

The circle represents the ball of radius 1. The horizontal axis is $\mathbb{R}^{n-1}$, the set of $(n-1)$-tuples $(x_1, \ldots, x_{n-1})$. The vertical axis is the last direction, x_n. If we were computing the volume of the ball in $\mathbb{R}^3$ (i.e., $n = 3$), we would add up the volumes of circular slices of radius $\cos\theta$ and thickness $d(\sin\theta)$:

$$K_3 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \pi \cos^2\theta \, d(\sin\theta).$$

Now this integrand $\pi\cos^2\theta$ is the *two*-dimensional volume of the *two*-dimensional ball (=disk) of radius $\cos\theta$. Note that it has the form

$$\text{volume of two-dimensional UNIT ball} \times \text{radius}^2 = \pi \times \cos^2\theta.$$

Using symmetry in θ, we may thus write

$$K_3 = 2\int_0^{\frac{\pi}{2}} K_2 \cos^3\theta \, d\theta.$$

The pattern continues for $n > 3$. I claim that $x_n = \sin\theta$, represented by a horizontal line segment in the figure, intersects $B^n(1)$ in an $(n-1)$-dimensional ball of radius $\cos\theta$. Indeed, the intersection is defined by

$$\begin{aligned} &x_1^2 + \cdots + x_{n-1}^2 + \sin^2\theta \le 1 \\ \Longrightarrow\ &x_1^2 + \cdots + x_{n-1}^2 \le 1 - \sin^2\theta = (\cos\theta)^2, \end{aligned} \tag{3.1}$$

and this is $B^{n-1}(\cos\theta)$. So, instead of a very thin slice with a two-dimensional disk as cross-section, we now have a very thin $(d(\sin\theta))$ slice whose cross-section is the ball of radius $\cos\theta$ in dimension $(n-1)$. Its volume is $K_{n-1}\cos^{n-1}\theta \times d(\sin\theta)$. Thus,

$$K_n = 2\int_0^{\frac{\pi}{2}} K_{n-1}\cos^n\theta \, d\theta = 2I_n K_{n-1}, \tag{3.2}$$

where

$$I_n \stackrel{\text{def}}{=} \int_0^{\frac{\pi}{2}} \cos^n\theta \, d\theta. \tag{3.3}$$

To make further progress towards a useful formula for K_n, we need to understand the integral I_n.

3.2. **The integral I_n.** Here we collect information about I_n. We return to the volume K_n in the next subsection.

Integration by parts in (3.3) gives (for $n > 1$)

$$\begin{aligned} I_n &= [\sin\theta\cos^{n-1}\theta]_0^{\frac{\pi}{2}} + (n-1)\int_0^{\frac{\pi}{2}} \cos^{n-2}\theta\sin^2\theta\,d\theta \\ &= (n-1)\int_0^{\frac{\pi}{2}} (\cos^{n-2}\theta - \cos^n\theta)\,d\theta \\ &= (n-1)(I_{n-2} - I_n). \end{aligned}$$

Hence

$$I_n = \frac{n-1}{n} I_{n-2}. \tag{3.4}$$

Since $I_0 = \frac{\pi}{2}$ and $I_1 = 1$, we can solve (3.4) recursively. Starting with I_0, we find $I_2, I_4, \ldots$, and from I_1 we get $I_3, I_5, \ldots$. The pattern is easy to ascertain:

$$I_{2p} = \frac{\pi}{2}\frac{1\cdot 3\cdot 5\cdots(2p-1)}{2\cdot 4\cdot 6\cdots(2p)}, \tag{3.5}$$

$$I_{2p+1} = \frac{2\cdot 4\cdot 6\cdots(2p)}{3\cdot 5\cdots(2p+1)}. \tag{3.6}$$

Remark 3.1. Standard notation: $1\cdot 3\cdot 5\cdot 7\cdots(2p+1) = (2p+1)!!$. □

We will need to know how I_n behaves as $n \to \infty$. Since the integrand $\cos^n\theta$ decreases with n for every $\theta \neq 0$, one expects that $I_n \to 0$. This is true, but we want to know how fast the decrease is. The next lemma provides the answer. We use the notation $a_n \sim b_n$ to signify that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$; this would be a good time to glance at Appendix B.

Lemma 3.1. $I_n \sim \sqrt{\frac{\pi}{2n}}$.

Proof. From (3.5) and (3.6), one sees that *for all positive* n,

$$I_n I_{n-1} = \frac{\pi}{2n}. \tag{3.7}$$

(Another proof: $nI_n = (n-1)I_{n-2}$ implies

$$nI_nI_{n-1} = (n-1)I_{n-2}I_{n-1}; \tag{3.8}$$

then writing (3.8) for with n replaced by $n-1, n-2, \ldots$, we get the string of equalities

$$nI_nI_{n-1} = (n-1)I_{n-1}I_{n-2} = (n-2)I_{n-2}I_{n-3} = \cdots = I_1I_0 = \frac{\pi}{2}.)$$

Here is the idea of the rest of the proof. First we note that by (3.4), $I_n \sim I_{n-2}$. Next we show that I_{n-1} is trapped between I_n and I_{n-2}.

It will follow that $I_{n-1} \sim I_n$, so that (3.7) gives $I_n^2 \sim \frac{\pi}{2n}$. Taking the square root will establish the Lemma. Let's do it more carefully.

Because $0 \le \cos\theta \le 1$, we have

$$\cos^n\theta < \cos^{n-1}\theta < \cos^{n-2}\theta$$

for $\theta \in (0, \frac{\pi}{2})$, and thus

$$I_n < I_{n-1} < I_{n-2}.$$

Divide by I_n and use (3.7):

$$1 < \frac{I_{n-1}}{I_n} < \frac{n}{n-1}.$$

Since the right side has limit 1, it follows that

$$\lim_{n\to\infty} \frac{I_{n-1}}{I_n} = 1.$$

Now multiply (3.7) by $\frac{I_n}{I_{n-1}}$ and rearrange:

$$\left(\frac{I_n}{\sqrt{\pi/(2n)}}\right)^2 = \frac{I_n}{I_{n-1}}.$$

The right side has limit $= 1$, hence so does the left side. Now take the square root. □

Remark 3.2. Lemma 3.1 has a remarkable consequence. Applied to $n = 2p+1$, it says that

$$\lim_{p\to\infty} (2p+1) \cdot I_{2p+1}^2 = \frac{\pi}{2},$$

or in longhand,

$$\lim_{p\to\infty} \frac{2^2 4^2 \cdots (2p)^2}{3^2 5^2 \cdots (2p-1)^2} \frac{1}{2p+1} = \frac{\pi}{2}. \tag{3.9}$$

There is a "dot-dot-dot" notation that is used for infinite products, analogous to the usual notation for infinite series (here ... stands for a certain limit of *partial products*, analogous to the familiar partial sums); one would write

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

This is called *Wallis' product*[2]. □

[2] J. Wallis, *Arithmetica infinitorum*, Oxford 1656 (!). The product can be recast as $\Pi_1^\infty(4n^2/(4n^2-1))$. By the way, Wallis was a codebreaker for Oliver Cromwell, and later chaplain to Charles II after the restoration

Let us now look at $\int_0^{\frac{\pi}{2}} \cos^n \theta \, d\theta$ more closely. The integrand is decreasing in θ, and for each $\theta \neq 0$ (where its value is $= 1$), it decreases to zero with n. So the main contribution to the integral should come from a neighborhood of $\theta = 0$. Indeed, if $\alpha > 0$ and $\theta \in (\alpha, \frac{\pi}{2}]$, then $\cos\theta < \cos\alpha$ and so

$$\int_\alpha^{\frac{\pi}{2}} \cos^n \theta \, d\theta < \int_\alpha^{\frac{\pi}{2}} \cos^n \alpha \, d\theta = (\frac{\pi}{2} - \alpha) \cos^n \alpha.$$

Now $\cos^n \alpha$ approaches zero geometrically, but we saw in Lemma 3.1 that I_n approaches zero only like $n^{-\frac{1}{2}}$. We are in the situation of the example in Appendix B ("geometric decrease beats power growth"), which gives $n^{\frac{1}{2}}(\cos\alpha)^n \to 0$:

$$\frac{1}{I_n} \int_\alpha^{\frac{\pi}{2}} \cos^n \theta \, d\theta \to 0.$$

We obtain:

Proposition 3.1. *For every* $\alpha \in (0, \frac{\pi}{2}]$,

$$I_n \sim \int_0^\alpha \cos^n \theta \, d\theta.$$

If, however, the limit of integration α is allowed to decrease as n increases, one can trap a proper fraction of the the total value of I_n in the shrinking interval of integration.

Consider

$$\int_0^{\frac{\beta}{\sqrt{n}}} \cos^n \theta \, d\theta = \frac{1}{\sqrt{n}} \int_0^\beta \cos^n(\frac{t}{\sqrt{n}}) \, dt.$$

From (A-6) in Appendix A,

$$\int_0^\beta \cos^n(\frac{t}{\sqrt{n}}) \, dt \to \int_0^\beta e^{-\frac{t^2}{2}} \, dt,$$

and we just saw that

$$I_n \sim \sqrt{\frac{\pi}{2n}}.$$

Thus

$$\lim_{n\to\infty} \frac{1}{I_n} \int_0^{\frac{\beta}{\sqrt{n}}} \cos^n \theta \, d\theta = \sqrt{\frac{2}{\pi}} \int_0^\beta e^{-\frac{t^2}{2}} \, dt. \tag{3.10}$$

Remark 3.3. By the time we get to (3.10), we no longer need $\beta \in (0, \pi/2)$. That restriction was used in the earlier observation that the integral over any finite subinterval $[0, \alpha] \subset [0, \pi/2]$ was asymptotically equivalent to I_n. Once we go to the shrinking upper limit $\alpha = \beta/\sqrt{n}$, then this upper limit will fall into $(0, \pi/2]$ when n is large enough, no matter the choice of β. $\square$

3.3. **Volume of the ball, part 2.** We had obtained the relation (3.2),

$$K_n = 2\int_0^{\frac{\pi}{2}} K_{n-1}\cos^n\theta\, d\theta = 2I_n K_{n-1}.$$

Replace n by $(n-1)$ to get $K_{n-1} = 2I_{n-1}K_{n-2}$, and thus

$$K_n = 4I_n I_{n-1} K_{n-2}.$$

Now recalling (3.7) from the preceding subsection, we obtain a recursion relation from which K_n can be determined:

$$K_n = \frac{2\pi}{n} K_{n-2}. \tag{3.11}$$

Knowing that $K_2 = \pi$ and $K_3 = \frac{4}{3}\pi$, we find $K_4, K_6, \ldots$ and $K_5, K_7, \ldots$, and recognize the pattern:

$$K_{2p} = \frac{\pi^p}{p!} \tag{3.12}$$

$$K_{2p+1} = \frac{2\cdot(2\pi)^p}{1\cdot 3\cdot 5\cdots(2p+1)}. \tag{3.13}$$

This formula has an amazing consequence:

Proposition 3.2. *For every $R > 0$, the volume of the ball of radius R in $\mathbb{R}^n$ tends to zero as $n \to \infty$.*

Proof. If $n = 2p$ is even, for example, the volume will be

$$K_{2p}R^{2p} = \frac{(\pi R^2)^p}{p!}.$$

Thus

$$K_{2p+2}R^{2p+2} = \frac{(\pi R^2)^2}{(p+2)(p+1)} K_{2p}R^{2p},$$

showing that the volume decreases rapidly, once p becomes sufficiently large. (If you notice that the $K_{2p}R^{2p}$ happen to be the coefficients in the Taylor series representation of e^x when $x = \pi R^2$, this argument should ring a bell). The proof for odd dimensions is similar. $\square$

As was seen in Proposition 2.1, the dependence of the volume of the cube $C^n(s)$ on s and n is quite different.

There is another curious relation between cubes and balls that should be noticed. The distance between the origin and the vertex $(s, \ldots, s)$ of $C^n(s)$ is $s\sqrt{n}$. Therefore, the radius ($=s\sqrt{n}$) of the smallest ball containing $C^n(s)$ tends to ∞ with n. Furthermore, if $s > \frac{1}{2}$, then $\mathrm{Vol}(B^n(s\sqrt{n})) \to \infty$ (since the volume of the inscribed cube tends to infinity). On the other hand, the largest ball contained in $C^n(s)$ will have radius s, which is independent of n. Even if $\mathrm{Vol}(C^n(s)) \to \infty$, the volume of this inscribed ball $B^n(s)$ tends to zero.

It is tempting to extrapolate to infinite dimensions. In $\mathbb{R}^\infty$ (whatever that might mean), a cube is not contained in any ball of finite radius. The cube with sides of length = 1 and volume = 1 contains the unit ball, which has volume = 0. Well, this last statement is mostly nonsense, but somewhere in it there is a little truth.

3.4. **Asymptotic behavior of the volume.** The formula for K_n involves factorials, which are approximated by Stirling's formula (C-2),

$$k! \sim \sqrt{2\pi k}\, k^k e^{-k}.$$

Proposition 3.3.

$$\mathrm{Vol}(B^n(R\sqrt{n})) \sim \frac{(2\pi e)^{\frac{n}{2}}}{\sqrt{n\pi}} R^n.$$

Proof. Apply Stirling's formula, first in even dimensions. From (3.12),

$$K_{2p} \sim \frac{\pi^p}{\sqrt{2\pi p}\, p^p e^{-p}};$$

if we set $p = n/2$ and do some rearranging (exercise), we get

$$K_n \sim \frac{(2\pi e)^{\frac{n}{2}}}{\sqrt{n\pi}(\sqrt{n})^n}. \tag{3.14}$$

Multiplying this by $(\text{radius})^n = (\sqrt{n}R)^n$, we get the desired formula.

The case of odd dimensions is something of a mess, and there is a subtle point. Begin by multiplying top and bottom of the expression (3.13) for K_{2p+1} by

$$2 \cdot 4 \cdot 6 \cdots (2p) = 2^p p!,$$

and also multiply from the beginning by the factor $(\sqrt{n})^n$ that will come from $(\sqrt{n}R)^n$, to get (recall that $n = 2p + 1$)

$$\frac{2(2\pi)^p\, 2^p p!}{(2p+1)!}(2p+1)^{\frac{2p+1}{2}}.$$

Next, substitute for the two factorials from Stirling's formula. The terms are ordered so that certain combinations are easier to see.

$$2(2\pi)^p\, 2^p \frac{\sqrt{2\pi p}}{\sqrt{2\pi(2p+1)}} \frac{e^{-p}}{e^{-2p-1}} \frac{p^p}{(2p+1)^{2p+1}} (2p+1)^{\frac{2p+1}{2}}.$$

First some obvious *algebraic* simplifications in the last expression:

$$2(2\pi)^p 2^p e^{p+1}\frac{\sqrt{p}}{\sqrt{2p+1}}\frac{p^p}{(2p+1)^p}(2p+1)^{-\frac{1}{2}}.$$

Moreover,

$$\frac{p^p}{(2p+1)^p}=\frac{1}{2^p}\left(\frac{p}{p+\frac{1}{2}}\right)^p=\frac{1}{2^p}\frac{1}{(1+\frac{1}{2p})^p},$$

and inserting this leads to

$$K_{2p+1}\sim 2(2\pi)^p e^{p+1}\frac{\sqrt{p}}{\sqrt{2p+1}}\frac{1}{(1+\frac{1}{2p})^p}(2p+1)^{-\frac{1}{2}}. \tag{3.15}$$

Now we do *asymptotic* simplifications as $p\to\infty$. The subtle point arises here. Let RHS stand for the right side of (3.15).

So far we know that

$$\lim_{p\to\infty}\frac{K_{2p+1}}{\text{RHS}}=1. \tag{3.16}$$

If we replace any part of RHS by an asymptotically equivalent expression, obtaining a modified expression RHS', say, we will still have

$$\lim_{p\to\infty}\frac{K_{2p+1}}{\text{RHS'}}=1.$$

For example,

$$\lim_{p\to\infty}\frac{\frac{\sqrt{p}}{\sqrt{2p+1}}}{\frac{1}{\sqrt{2}}}=1.$$

We can then modify (3.16):

$$\lim_{p\to\infty}\frac{K_{2p+1}}{\text{RHS}}\frac{\frac{\sqrt{p}}{\sqrt{2p+1}}}{\frac{1}{\sqrt{2}}}=1.$$

This gives a new denominator, and we get

$$\lim_{p\to\infty}\frac{K_{2p+1}}{2(2\pi)^p e^{p+1}\frac{1}{\sqrt{2}}\frac{1}{(1+\frac{1}{2p})^p}(2p+1)^{-\frac{1}{2}}}=1,$$

which we would write more briefly as

$$K_{2p+1}\sim 2(2\pi)^p e^{p+1}\frac{1}{\sqrt{2}}\frac{1}{(1+\frac{1}{2p})^p}(2p+1)^{-\frac{1}{2}}. \tag{3.17}$$

A comparison of (3.15) and (3.17) suggests that we have allowed p to tend to ∞ in some places, but not in others. Isn't this unethical? No it is not, as long as we are replacing one expression whose ratio with K_{2p+1} tends to 1 by another expression with the same property.

This same reasoning further allows us to substitute

$$\frac{1}{e^{\frac{1}{2}}} \text{ for } \frac{1}{(1+\frac{1}{2p})^p}.$$

By now, (3.15) has been simplified to

$$\frac{\sqrt{2}(2\pi)^p e^{p+\frac{1}{2}}}{(2p+1)^{\frac{1}{2}}}. \tag{3.18}$$

and it is an easy manipulation to turn this into the desired formula in terms of $n = 2p + 1$. □

Corollary 3.1.

$$\lim_{n\to\infty} \mathrm{Vol}(B^n(R\sqrt{n})) = \begin{cases} 0, & \text{if } R \leq \frac{1}{\sqrt{2\pi e}} \\ \infty, & \text{if } R > \frac{1}{\sqrt{2\pi e}}. \end{cases}$$

Proof. Write the asymptotic expression for the volume as

$$\frac{(\sqrt{2\pi e R^2})^n}{\sqrt{n\pi}},$$

and recall that when $2\pi e R^2 > 1$, the numerator grows much faster than the denominator. The case ≤ 1 is trivial. □

This is a little more like the behavior of the volume of the cube, especially if you think not about the side length $2s$ of $C^n(s)$, which remains fixed, but about the length $2s\sqrt{n}$ of its diagonal:

$$\text{Cube: diameter} = \text{constant } \times \sqrt{n},$$
$$\text{Ball: radius} = \text{constant } \times \sqrt{n}.$$

The behavior of the volume is determined by the value of the proportionality constant. Still, there is no way to get a nonzero, non-infinite limiting volume for the ball.

3.5. **Volume concentration near the surface.** Consider the two concentric balls $B^n(1)$ and $B^n(1-\epsilon)$. The ratio of volumes is

$$\frac{\mathrm{Vol}(B^n(1-\epsilon))}{\mathrm{Vol}(B^n(1))} = \frac{K_n(1-\epsilon)^n}{K_n} = (1-\epsilon)^n.$$

For every ϵ, this ratio tends to zero as $n \to \infty$, which means that every spherical shell, no matter how thin, will contain essentially the whole volume of $B^n(1)$. Of course, it should be remembered that $\lim \mathrm{Vol}(B^n(1)) = 0$, so the shell also has a small volume, it is just a significant fraction of the whole.

To see how the volume accumulates at the surface, we again let ϵ depend on n, as we did for the cube. Choosing $\epsilon = \frac{t}{n}$, we find that

$$\frac{\mathrm{Vol}(B^n(1-\frac{t}{n}))}{\mathrm{Vol}(B^n(1))} = \frac{K_n(1-\frac{t}{n})^n}{K_n} = (1-\frac{t}{n})^n \to e^{-t}.$$

The interpretation is exactly as for the cube (subsection 2.1).

3.6. **Volume concentration near the equator for $\mathbf{B^n(R)}$.** Recall that

$$\mathrm{Vol}(B^n(R)) \to 0$$

for every R. We ask: what fraction of this vanishing (as $n \to \infty$) volume accumulates near the equator of the ball? Let $\theta_1 < 0 < \theta_2$, and consider

$$\frac{R^n \int_{\theta_1}^{\theta_2} \cos^n\theta\, d\theta}{R^n \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}} = \frac{1}{2I_n}\left(\int_{\theta_1}^{0} \cos^n\theta\, d\theta + \int_0^{\theta_2} \cos^n\theta\, d\theta\right). \tag{3.19}$$

The numerator is the volume of the slice bounded by the hyperplanes[3] $x_n = R\sin\theta_1$, $x_n = R\sin\theta_2$ (see Figure 1). Since the factor R^n has cancelled, we will temporarily set $R = 1$ and work with the unit ball; later, R will be re-inserted by hand.

According to Proposition 3.1, the two integrals on the right side of (3.19) are each asymptotically equivalent to I_n. Hence:

Proposition 3.4. *For every $\theta_1 \in [-\frac{\pi}{2}, 0)$ and $\theta_2 \in (0, \frac{\pi}{2}]$, the fraction of the volume of $B^n(R)$ contained in the equatorial slice between the hyperplanes $x_n = \sin\theta_1$ and $x_n = \sin\theta_2$ tends to 1 as $n \to \infty$. If, however, $\theta_1 > 0$ ($\theta_1 < \theta_2$), then the fraction of volume in the slice tends to zero.*

As before, to trap a proper fraction of the volume, we must allow the limits of integration to approach 0 as $n \to \infty$. The results of Subsection 3.2 suggest that we take $\theta_1 = \frac{\beta_1}{\sqrt{n}}$ and $\theta_2 = \frac{\beta_2}{\sqrt{n}}$, $\beta_1 < \beta_2$. The signs of the β's will not matter now, only the fact that $\frac{\beta_j}{\sqrt{n}} \to 0$ at a cleverly chosen rate. Incidentally, since $\sin u \sim u$ for $u \to 0$, the bounding hyperplanes $x_n = \sin(\frac{\beta_j}{\sqrt{n}})$ are essentially $x_n = \frac{\beta_j}{\sqrt{n}}$ as n gets large.

From equation (3.10), we now obtain

[3] a *hyperplane* in $\mathbb{R}^n$ is an $n-1$-dimensional set defined by a single linear equation in $x_1, \ldots, x_n$. If this set does not contain the origin, it is sometimes called an *affine* hyperplane, but I won't make that distinction. "Hyper" means "one dimension down", in this context. The set defined by a single, generally nonlinear, equation $f(x_1, \ldots, x_n) = 0$ is a *hypersurface*. The sphere $S^{n-1}(R)$ is a hypersurface in $\mathbb{R}^n$.

Proposition 3.5. *Let $\beta_1 < \beta_2$. The fraction of the volume of $B^n(1)$ contained in the slice bounded by the hyperplanes $x_n = \sin(\frac{\beta_j}{\sqrt{n}})$, $j = 1, 2$, tends to*

$$\frac{1}{\sqrt{2\pi}} \int_{\beta_1}^{\beta_2} e^{-\frac{t^2}{2}}\, dt. \tag{3.20}$$

Remark 3.4. Proposition 3.5 remains true if one asks about the volume between the hyperplanes $x_n = \frac{\beta_j}{\sqrt{n}}$, $j = 1, 2$. That volume differs from the volume in the proposition by

$$\int_{\frac{\beta_1}{\sqrt{n}}}^{\sin^{-1}(\frac{\beta_1}{\sqrt{n}})} \cos^n \theta\, d\theta$$

and another similar integral, but the limits of integration approach each other as $n \to \infty$ and the integral tends to zero. The details are omitted. □

Remark 3.5. If you know some probability theory, you will have noticed that (3.20) represents an area under the standard normal curve. It is an often used fact that the value of (3.20) for $\beta_1 = -1.96$, $\beta_2 = +1.96$ is approximately .95. In other words, for large n all but about 5% of the volume of $B^n(R)$ will be found in the very thin slice between $x_n = \pm\frac{1.96R}{\sqrt{n}}$. □

Remark 3.6. By rotational symmetry, Proposition 3.5 will hold for *every equator*, not just the one perpendicular to the x_n-direction.

THINK ABOUT THAT FOR A MINUTE. Next, incorporate this in your thoughts: THE VOLUME ALSO CONCENTRATES NEAR THE SURFACE OF THE BALL. □

3.7. **Volume concentration near the equator for $\mathbf{B^n(R\sqrt{n})}$.** The results of the last subsection can be viewed from a slightly different angle, which is important enough to be emphasized in its own little subsection.

Instead of fixing R, and looking at the fraction of volume contained in the slice between[4] $x_n = \frac{\beta_1}{\sqrt{n}}$ and $x_n = \frac{\beta_2}{\sqrt{n}}$, one can let the radius of the ball become infinite at the rate $R\sqrt{n}$ and look at the slice between $x = \beta_1$ and $x_n = \beta_2$. Keep in mind that $\mathrm{Vol}(B^(R\sqrt{n})) \to 0$ or ∞, so we may be talking about a fraction of a tiny volume or of a huge volume.

[4]it is simpler to use $\beta/\sqrt{n}$ instead of $\sin(\beta/\sqrt{n})$; the two are asymptotically equivalent

Proposition 3.6. *Let $\beta_1 < \beta_2$. The fraction of the volume of $B^n(R\sqrt{n})$ contained in the slice bounded by the hyperplanes $x_n = \beta_j, j = 1, 2$, tends to*

$$\frac{1}{\sqrt{2\pi}} \int_{\beta_1}^{\beta_2} e^{-\frac{t^2}{2}}\, dt. \tag{3.21}$$

Exercise 3.1. Prove this proposition. □

4. Surface area of the ball

4.1. **A formula for surface area.** Recall that the surface of a "ball" is a "sphere". We speak of the "area" of a sphere simply to distinguish that quantity from the "volume" of the whole ball. But keep in mind (Example 3.1) that the sphere $S^3(R)$ is itself 3-dimensional (sitting in the 4-dimensional $\mathbb{R}^4$). What we refer to as its "area" is actually its 3-*dimensional volume.* Likewise, the "area" of $S^{n-1}(R) \subset \mathbb{R}^n$ is understood to mean its $(n-1)$-dimensional volume.

The surface area of $B^n(R)$ is easily expressed in terms of $\mathrm{Vol}(B^n(R))$. Consider the n-dimensional volume of the spherical shell between the balls of radii $R + \delta R$ and R:

$$\begin{aligned} \mathrm{Vol}(B^n(R+\delta R)) - \mathrm{Vol}(B^n(R)) &= K_n[(R+\delta R)^n - R^n] \\ &= K_n[nR^{n-1}\delta R + \cdots], \end{aligned}$$

with $\cdots$ denoting terms in $(\delta R)^2$ and higher. Think of this volume as being approximately

$$\text{volume of shell} = \text{area of } S^{n-1}(R) \ \times \ \text{thickness } \delta R \text{ of shell}, \tag{4.1}$$

i.e.

$$\text{volume of shell } = nK_nR^{n-1} \times \delta R + \cdots .$$

(4.1) becomes more accurate as $\delta R \to 0$. Hence the area of S^{n-1} is

$$\frac{d}{dR}\mathrm{Vol}(B^n(R)) = nK_nR^{n-1}. \tag{4.2}$$

This type of argument would indeed recover a formula which you know from calculus, for the area of a region D on a surface (in $\mathbb{R}^3$) given by $z = f(x, y)$:

$$\iint_D \sqrt{1 + f_x^2 + f_y^2}\, dx\, dy.$$

But we will need to get at the area of the sphere by a different, seemingly more rigorous approach anyway.

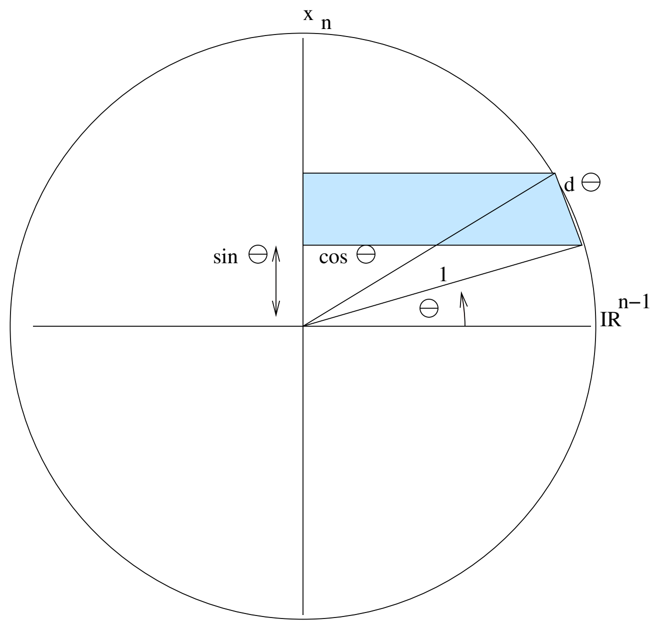

FIGURE 2. Surface area by slices

Let L_{n-1} denote the surface area of the unit ball $B^n(1)$ in $\mathbb{R}^n$. We know from general principles that the ball $B^n(R)$ will have surface area $L_{n-1} \times R^{n-1}$; this will be used shortly. Now refer to Figure 2, which again represents the unit ball $B^n(1)$ in $\mathbb{R}^n$. (If you are at all uneasy with the use of "n" in what follows, work through the steps using $n = 3$).

Recall from (3.1) that the line segment $x_n = \sin\theta$ represents the ball $B^{n-1}(\cos\theta)$; as just remarked, it will have surface area $L_{n-2} \times (\cos\theta)^{n-2}$. This is $(n-2)$-dimensional volume. The outside arc of the shaded area is the surface of a cylinder put together as follows: the base is $B^{n-1}(\cos\theta)$ and the edge (surface) of the base is $S^{n-2}(\cos\theta)$; the height is $d\theta$. For small $d\theta$, the side of the cylinder is nearly perpendicular to the base, and then the outside arc of the shaded region represents a contribution

$$\text{circumference} \times \text{height} \; = L_{n-2}(\cos\theta)^{n-2} \times d\theta$$

to the total surface area. Therefore

$$L_{n-1} = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} L_{n-2} \cos^{n-2}\theta \, d\theta = 2I_{n-2}L_{n-2}. \tag{4.3}$$

To obtain a formula for L_{n-1}, we can proceed in three ways. First, in the formula $L_{n-1} = nK_n$, which is (4.2) at $R = 1$, use the known expressions for K_n. Second, in (4.3), start with $n = 3$ and $L_{3-2} = L_1 =$

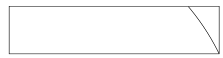

FIGURE 3. Volume vs. surface area

2π, and use the known expressions for I_{n-2}. Third, in (4.3), replace L_{n-2} by $2I_{n-3}L_{n-3}$ (which is (4.3) after the replacement $(n-1) \to (n-2)$), to get

$$L_{n-1} = 4I_{n-2}I_{n-3}L_{n-3} = \frac{2\pi}{n-2}L_{n-3} \quad \text{(last step by (3.7))}.$$

Then use $L_1 = 2\pi$ and $L_2 = 4\pi$ to get started on a recursive determination of L_{n-1}. Presumably, all these methods lead to the same answer:

Proposition 4.1.

$$L_{2p} = 4\pi\,\frac{(2\pi)^{p-1}}{(2p-1)!!}$$

$$L_{2p-1} = 2\pi\,\frac{\pi^p}{p!}.$$

Remark. It is worth understanding exactly why $\cos^n\theta$ becomes $cos^{n-2}\theta$, and not $\cos^{n-1}\theta$, which would be analogous to the change in the exponent of R.

For the volume, we had

$$\text{volume of base of slice } \propto (\text{radius})^{n-1}$$

which was then multiplied by

$$\text{thickness of slice } = d(\sin\theta) = \cos\theta\, d\theta.$$

For the surface area, it was

$$\text{surface area of base of slice } \propto (\text{radius})^{n-2}$$

(here the power does just drop by 1), but this was only multiplied by the "arclength" $d\theta$. So the difference is the bit of rectangle that sticks out beyond the circular arc in Figure 3. □

Exercise 4.1. Compare the $n \to \infty$ behavior of the volume and surface area of $B^n(R\sqrt{n})$, for all $R > 0$. One value of R should be peculiar. □

4.2. **Surface area concentration near the equator.** The surface area of $B^n(R)$ concentrates near the (or any) equator. The calculation has already been done. Look at (3.19). If R^n is replaced by R^{n-1}, and $\cos^n \theta$ is replaced by $cos^{n-2}\theta$, (3.19) represents the fraction of surface area of $B^n(R)$ contained between the hyperplanes $x_n = \sin\theta_j$, $j = 1, 2$. Everything we know about the $n \to \infty$ limits remains true, word for word. I will not restate the results, you do it:

Exercise 4.2. State precisely the proposition about surface area concentration near the equator. □

5. The weak law of large numbers

In this section, concentration of volume for the cube about its "equator" is shown to be an expression of one of the basic theorems of probability theory. I will use several technical probability definitions in an intuitive way; otherwise, there would be too many digressions.

5.1. **The probability setting.** Let $C^1 = C^1$ again be the interval $(\frac{1}{2}) = [-\frac{1}{2}, \frac{1}{2}]$, and consider the experiment of "choosing a number at random from this interval". All results are to be considered equally likely. More precisely, the probability that a randomly chosen number x_1 falls into a subinterval $[a, b] \subseteqq [-\frac{1}{2}, \frac{1}{2}]$ is *defined* to be its length, $b - a$. We write $P(x_1 \in [a, b]) = b - a$. (So $P(x_1 \in [-\frac{1}{2}, \frac{1}{2}]) = 1$). In a setup like this, where the possible choices form a continuum, the probability that a random x_1 will be exactly equal to a given number is *zero*. For example, $P(x_1 = 0) = 0$. A justification might be that you would never know x_1 to all its decimal places. More mathematically, this is the only possible logical consequence of our starting definition of the probability: $P(x_1 \in [0, 0]) = 0 - 0 = 0$.

Now we go to C^n (unit side length, remember), and define the probability that a randomly chosen point lies in a region $D \subseteqq C^n$ to be the volume, $\mathrm{Vol}(D)$. Again, the whole cube has probability 1. One can imagine picking $\mathbf{x} = (x_1, \ldots, x_n)$ "as a whole", e. g. by throwing a dart into the n-dimensional cube.

Or—and this is what I will do—one can build up the random $\mathbf{x}$ one coordinate at a time. First choose $x_1 \in C^1$ at random. Then, independently of this first choice, pick $x_2 \in C^1$. Continue. The probability that an $\mathbf{x}$ built in such a way lies in D is again $\mathrm{Vol}(D)$. This is not easy to prove with full mathematical rigor, one essentially needs the theory of Lebesgue measure. I will illustrate for the case $n = 2$ and $D =$ a rectangle. The key technical notion, which cannot be avoided, is *independence.*

Summary. An "experiment" with a number (maybe infinite) of possible outcomes is performed. An *event* A is a subset of the set of possible outcomes. A *probability measure* assigns numbers between 0 and 1 to events. The interpretation of "$P(A) = p$" is that if the "experiment" is repeated a large number N times, then about pN of the results will fit the criteria defining A. (Every single statement I just made requires clarification and much more mathematical precision.)

Example 5.1. Suppose a 6-sided die is tossed $60,000$ times. Let

$$A = \text{"2, 4, or 6 shows"}, \text{ and } B = \text{"1, 3, 4, or 6 shows"}.$$

Assuming that each of the six outcomes $1, 2, 3, 4, 5, 6$ is equally likely, we have $P(A) = \frac{1}{2}$ and $P(B) = \frac{2}{3}$. This means, intuitively, that A should happen on about half, or $30,000$ of the tosses. Of these $30,000$, about one third each will result in in $2, 4, 6$, so B happens on $2/3$ or $20,000$ tosses. Tosses resulting in 1 or 3 are no longer possible, since we are restricting attention to outcomes where A has occurred.

In other words, the probability of B was $\frac{2}{3}$ at the outset, and it was still $\frac{2}{3}$ after the results were restricted to the outcomes $2, 4, 6$ in A. One says:

$$\text{the probability of } B \text{ given } A = \text{the probability of } B. \tag{5.1}$$

We cannot make a better prediction about the occurrence of B amongst the outcomes satisfying A than we could about the occurrence of B amongst all possible outcomes. □

Example 5.2. Suppose now B is "1,3, or 5 shows". Then none of the $30,000$ tosses fitting the description of A will fit the description of B. Thus:

$$\text{the probability of } B \text{ given } A = 0.$$

If we know that A has occurred, we know with certainty that B has not occurred. □

We consider the A and B in Example 5.1 to be *independent*. Example 5.2 illustrates an extreme case of non-independence. Another extreme case would be $B =$ "1, 2, 4, or 6 shows". If we know that A has happened, it is certain that B has also happened, or

$$\text{the probability of } B \text{ given } A = 1.$$

The technical meaning of *independence* is a rephrasing of (5.1):

$$P(A \text{ and } B \text{ both happen }) = P(A)P(B). \tag{5.2}$$

To understand this, rewrite it as (if $P(A) \neq 0$)

$$\frac{P(A \text{ and } B \text{ both happen })}{P(A)} = P(B). \tag{5.3}$$

The left side is the fraction of the outcomes satisfying A for which B also happens. In the setting of Example 5.1, it would be

$$\frac{20,000}{30,000} = \frac{\frac{1}{3} \times 60,000}{\frac{1}{2} \times 60,000} = \frac{2}{3} = P(B).$$

Example 5.3. We will pick a point $\mathbf{x} \in C^2$ at random, by choosing its two coordinates independently. Let A_1 be the event "$x_1 \in [a_1, b_1]$", and let A_2 be the event "$x_2 \in [a_2, b_2]$". We have $P(A_1) = b_1 - a_1$, $P(A_2) = b_2 - a_2$. The choices of x_1 and x_2 are to be made independently. This means that we *require* that

$$P(x_1 \in [a_1, b_1] \text{ and } x_2 \in [a_2, b_2]) = (b_2 - a_2)(b_1 - a_1).$$

But this is the same as saying that

$$\begin{aligned} P(\mathbf{x} = (x_1, x_2) \in &\text{ rectangle with base } [a_1, b_1] \text{ and height } [a_2, b_2]) \\ &= \mathrm{Vol}(\text{rectangle}). \end{aligned}$$

The general statement "$P(\mathbf{x} \in D) = \mathrm{Vol}(D)$" is deduced by approximating the region D by a union of small rectangles (compare the definition of a multiple integral as limit of Riemann sums). □

Laws of large numbers give precise meaning to statements that "something should happen on average". For example, if we choose, randomly and independently, a large number N of points in $C^1 = [-\frac{1}{2}, \frac{1}{2}]$, say $a_1, a_2, \ldots, a_N$, any positive a_j should sooner or later be nearly balanced by a negative one, and the average

$$\frac{a_1 + a_2 + \cdots + a_N}{N} \tag{5.4}$$

should be close to zero. The larger N is, the better the "law of averages" should apply, and this ratio should be closer to zero.

Now, as just explained, these independently chosen a_j can be thought of as the coordinates of a randomly picked point $\mathbf{a} = (a_1, \ldots, a_N)$ in the high-dimensional cube C^N. One version of the "law of averages" says that the volume of the region in C^N where (5.4) is *not* close to zero becomes negligible as $N \to \infty$. In other words, we have concentration of volume again:

$$\mathrm{Vol}\left(\{\mathbf{a} = (a_1, \ldots, a_N) \in C^N \mid |\frac{a_1 + a_2 + \cdots + a_N}{N}| \text{ is small}\} \right) \to 1.$$

I now turn to the geometric problem, and revert to my usual notation $(x_1, \ldots, x_n) \in C^n$. The interest is in concentration of volume about the plane $\Pi_0 : x_1 + \cdots + x_n = 0$. Notice that this plane has normal $(1, 1, \ldots, 1)$, and that the line along the normal is in fact the main diagonal of the cube, passing through the corner $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$. The distance to the corner is $\frac{1}{2}\sqrt{n}$. We want to look at slices bounded by planes $\Pi_{\pm\eta} : x_1 + \ldots = \pm\eta$; possibly η will depend on n—we shall see.

The honest approach would be to find a nice, more or less explicit, formula for the volume of such an "equatorial" slice, as there was for a ball, and then to apply the geometric fact of concentration of volume to obtain a probabilistic application. However, I confused myself trying to find a volume formula; there are too many corners when you slice the cube by a plane Π_η, and I gave up. The less satisfying approach is to prove the probabilistic result as it is done in probability texts (using a simple inequality, see 527 Notes, p. 386), and to reinterpret it in geometric language. This is what I have to do.

5.2. **The weak law of large numbers.** For $\mathbf{x} = (x_1, \ldots, x_n) \in C^n$, set

$$S_n(\mathbf{x}) = x_1 + \cdots + x_n;$$

this is a function whose domain is the cube. I will sometimes omit the argument $\mathbf{x}$. Further, given $\epsilon > 0$, define the "equatorial slice"

$$\mathcal{S}_\epsilon = \{\mathbf{x} \in C^n \mid |\frac{S_n(\mathbf{x})}{n}| < \epsilon\}. \tag{5.5}$$

Theorem 5.1 (Weak law of large numers)**.** *For every $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathrm{Vol}(\mathcal{S}_\epsilon) = 1.$$

Equivalently,

$$\lim_{n\to\infty} \mathrm{Vol}\bigg(\{\mathbf{x} \in C^n \mid |\frac{S_n(\mathbf{x})}{n}| > \epsilon\}\bigg) = 0. \tag{5.6}$$

Remark 5.1. The "weak law" is really a much more general theorem. I have stated it in the context of our cube example. But the x_j might also be chosen at random from the two-element set $\{-1, 1\}$, with $P(-1) = P(1) = \frac{1}{2}$. These two numbers could signify "tails" and "heads" in a toss of a fair coin. We would then be working with a discrete cube, $\{-1, 1\}^n$. Theorem 5.1 gives one precise meaning to the law of averages for coin tosses: there are, on average, about as many heads as there are tails. □

Remark 5.2. If there is a weak law, there should be a strong law, and there is. It is set in the limiting cube, C^∞, where now $\mathbf{x} = (x_1, x_2, \ldots)$ is an ∞-tuple. It says that those $\mathbf{x}$ for which

$$\lim_{N\to\infty} \frac{x_1 + \cdots + x_N}{N} = 0 \tag{5.7}$$

occupy total volume 1 in C^∞. Evidently, some more theory is needed here, to give meaning to the notion of volume in C^∞, and to clarify the nature of possible sets of zero volume (where the limit (5.7) fails to hold). □

The proof of Theorem 5.1 is based on a useful inequality, which is stated here for the cube C^n, although any region of integration could be used. For typing convenience, I use a single integral sign to stand for an n-fold multiple integral, and I abbreviate $dx_1 \cdots dx_n = d^n\mathbf{x}$:

$$\underbrace{\int \cdots \int}_{C^n} \cdots \, dx_1 \cdots dx_n \equiv \int_{C^n} \cdots \, d^n\mathbf{x}.$$

Lemma 5.1 (Chebyshev's inequality)**.** *Let $F : C^n \to \mathbb{R}$ be a function for which $\sigma^2 \stackrel{\text{def}}{=} \int F(\mathbf{x})^2 \, d^n\mathbf{x}$ is finite. For $\delta > 0$, let*

$$A_\delta = \{\mathbf{x} \in C^n \mid |F(\mathbf{x})| > \delta\}.$$

Then

$$\operatorname{Vol}(A_\delta) < \frac{\sigma^2}{\delta^2}. \tag{5.8}$$

Proof. (of lemma) Since $A_\delta \subsetneqq C^n$,

$$\int_{A_\delta} F(\mathbf{x})^2 \, d^n\mathbf{x} \leq \int_{C^n} F(\mathbf{x})^2 \, d^n\mathbf{x} = \sigma^2.$$

For $\mathbf{x} \in A_\delta$ we have $(\delta)^2 < (F(\mathbf{x}))^2$, so

$$\int_{A_\delta} \delta^2 \, d^n\mathbf{x} < \int_{A_\delta} F(\mathbf{x})^2 \, d^n\mathbf{x}.$$

The left side is just

$$\delta^2 \int_{A_\delta} d^n\mathbf{x} = \delta^2 \operatorname{Vol}(A_\delta).$$

Combine this string of inequalities to get

$$\delta^2 \operatorname{Vol}(A_\delta) < \sigma^2,$$

as desired. □

Proof. (of Theorem) We apply Chebyshev's inequality to $F = S_n$. First, we note that

$$(S_n)^2 = \sum_{j=1}^{n} x_j^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} 2x_i x_j.$$

The integral over C^n of the double sum vanishes, since in the iterated multiple integral,

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} x_i \, dx_i = 0.$$

A typical contribution from the simple sum is

$$\int_{C^n} x_1^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \cdots \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} x_1^2 \, dx_1 \right) dx_2 \dots dx_n = \frac{1}{12}.$$

Thus,

$$\sigma^2 = \frac{n}{12}.$$

Chebyshev's inequality now gives

$$\mathrm{Vol}\Big(\{\mathbf{x} \mid |S_n(\mathbf{x})| > \delta\} \Big) < \frac{n}{12\delta^2}.$$

Set $\delta = \epsilon n$ in this relation, and you get

$$\mathrm{Vol}\Big(\{\mathbf{x} \mid \frac{|S_n(\mathbf{x})|}{n} > \epsilon\} \Big) < \frac{1}{12n\epsilon^2}. \tag{5.9}$$

This tends to zero as $n \to \infty$, and the theorem is proved. □

5.3. **Connection with concentration of volume.** We have seen that the volume of the region $\subsetneqq C^n$ where $\frac{|S_n|}{n} < \epsilon$ tends to 1 as $n \to \infty$. The geometric meaning of this fact still needs a little more clarification.

To this end, we switch to a coordinate system in which one direction is perpendicular to the planes $\Pi_\eta : x_1 + \cdots + x_n = \eta$. This new direction is analogous to the x_n-direction in our discussion of the ball.

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the standard basis,

$$\mathbf{e}_j = (0, \dots, 0, \underbrace{1}_{j^{th} \text{ place}}, 0, \dots, 0).$$

Thus,

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

Now let $\mathbf{f}_n$ be the unit vector

$$\frac{1}{\sqrt{n}}(1, \dots, 1) = \frac{1}{\sqrt{n}}(\mathbf{e}_1 + \cdots + \mathbf{e}_n)$$

in the direction normal to Π_0. The orthogonal complement to $\mathbf{f}_n$ is of course the $(n-1)$-dimensional vector subspace

$$\Pi_0 = \{\mathbf{x} \in \mathbb{R}^n \mid x_1 + \cdots + x_n = 0\}.$$

The process of *Gram-Schmidt orthogonaliztion* in linear algebra guarantees that we can start with the unit vector $\mathbf{f}_n$, and construct orthonormal vectors $\mathbf{f}_1, \ldots, \mathbf{f}_{n-1}$ which, together with $\mathbf{f}_n$, provide a new orthonormal basis of $\mathbb{R}^n$. (These $\mathbf{f}_j$ are far from unique; certainly one possible set can be written down, but that would be more confusing than helpful).

The point is that a given $\mathbf{x}$ in the cube now has two coordinate expressions:

$$x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n = y_1\mathbf{f}_1 + \cdots + y_{n-1}\mathbf{f}_{n-1} + y_n\mathbf{f}_n.$$

Take inner products of both sides with $\mathbf{f}_n$, and remember that the $\mathbf{f}_j$ are orthonormal. You find

$$y_n = \frac{1}{\sqrt{n}}(x_1 + \cdots + x_n) = \frac{1}{\sqrt{n}} S_n. \tag{5.10}$$

Now go back to (5.9). Inserting (5.10), we convert it into

$$\mathrm{Vol}\Big(\{\mathbf{y} \mid |y_n| > \epsilon\sqrt{n}\} \Big) < \frac{1}{12n\epsilon^2}.$$

Finally, notice that the corner of the cube, where all $x_j = \frac{1}{2}$, corresponds to $y_n = \frac{\sqrt{n}}{2}$. We may therefore phrase the weak law of large numbers in more geometric language, as follows.

Proposition 5.1. *The diagonal of C^n has length $\sqrt{n}$. The fraction of the (unit) volume of C^n contained in the slice $\mathcal{S}_{\epsilon\sqrt{n}}$ (see (5.5)) approaches 1 as $n \to \infty$.* □

In other words, a slice $\mathcal{S}_{\epsilon\sqrt{n}}$ whose width is an arbitrarily small fraction of the length of the diagonal of C^n will eventually enclose almost all the volume of C^n.

Remark 5.3. Recall that the volume of C^n *also* concentrates near the surface! □

Exercise 5.1. Verify that one possible choice for the new basis vectors $\mathbf{f}_j$ is, for $j = 1, \ldots, n-1$

$$\mathbf{f}_j = (\underbrace{\frac{1}{\sqrt{j(j+1)}}, \ldots, \frac{1}{\sqrt{j(j+1)}}}_{j \text{ terms}}, -\frac{j}{\sqrt{j(j+1)}}, 0, \ldots, 0).$$

□

A. Appendix: compound interest

In the main part of these notes, we need two facts. First, equation (A-2) below, which often arises in a discussion of compound interest in beginning calculus. It is fairly easy to prove. Second, the rather more subtle formula

$$\lim_{n\to\infty}\int_0^\beta \cos^n(\frac{t}{\sqrt{n}})\,dt = \int_0^\beta \lim_{n\to\infty}\cos^n(\frac{t}{\sqrt{n}})\,dt = \int_0^\beta e^{-\frac{t^2}{2}}\,dt, \tag{A-1}$$

which is a consequence of (A-2) and some other stuff. The first subsection explains the ideas, and everything afterwards fills in the technical mathematical details. There are lots of them, because they will reappear during the Math 527 year, and you might as well get a brief introduction. The details could well detract from the important issues on first reading. Study the earlier material, and then go through the details in the appendices later (but do go through them).

A.1. The ideas.

The first result should be familiar.

Lemma A.1. *For every $a \in \mathbb{R}$,*

$$\lim_{n\to\infty}(1+\frac{a}{n})^n = e^a. \tag{A-2}$$

I will actually require a consequence of a refinement of Lemma A.1. This generalization of Lemma A.1 says, in a mathematically precise way, that if $b_n \to 0$, then we still have

$$\lim_{n\to\infty}(1+\frac{a+b_n}{n})^n = e^a, \tag{A-3}$$

and that the approach to the limit is equally fast for all a in any fixed finite interval $[-A, A]$. (Technical term: the convergence is *uniform*). The relevant application is this unexpected corollary:

Corollary A.1.

$$\lim_{n\to\infty}\cos^n(\frac{t}{\sqrt{n}}) = e^{-\frac{t^2}{2}}. \tag{A-4}$$

Convergence is uniform *on finite intervals, in this sense: given $\beta > 0$ and $\epsilon > 0$, there is an integer $N > 0$ depending on β and ϵ* but not on *t, such that $n > N$ implies*

$$|\cos^n(\frac{t}{\sqrt{n}}) - e^{-\frac{t^2}{2}}| < \epsilon \text{ for all } t \in [-\beta, \beta]. \tag{A-5}$$

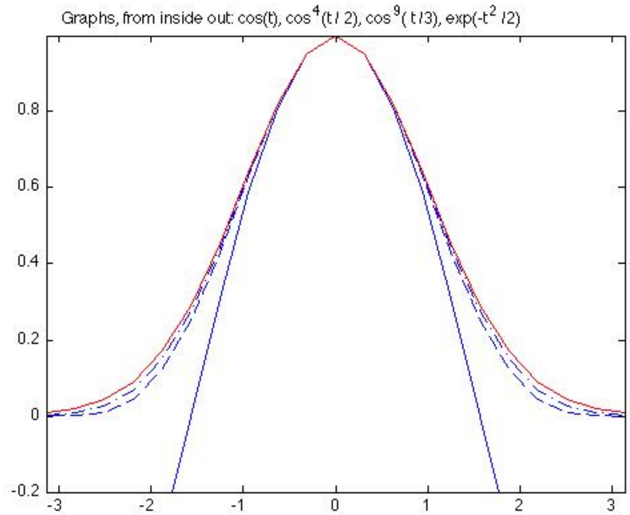

FIGURE 4. Illustration of (A-4)

Shelve that definition for the moment, and let me reveal why (A-4) should be true and why I want this "uniform" convergence. From the Taylor series for $\cos u$ we have

$$\cos u = 1 - \frac{1}{2}u^2 + \cdots .$$

Suppose all the "$\cdots$" terms are just not there. Then by (A-2),

$$(\cos(\frac{t}{\sqrt{n}}))^n = (1 - \frac{t^2}{2n})^n$$

which has limit $e^{-t^2/2}$! But the "$\cdots$" *are* there, in the guise of the correction term in Taylor's formula. So one encounters the more complicated limit (A-3).

I will then want to conclude that

$$\text{(A-6)} \qquad \lim_{n\to\infty} \int_0^\beta \cos^n(\frac{t}{\sqrt{n}})\, dt \overset{\downarrow}{=} \int_0^\beta \lim_{n\to\infty} \cos^n(\frac{t}{\sqrt{n}})\, dt = \int_0^\beta e^{-\frac{t^2}{2}}\, dt.$$

This seems plausible, given (A-4), but it is not always true that the limit of the integrals of a sequence of functions is equal to the integral of the limit function. In other words, equality $\overset{\downarrow}{=}$ may fail for some sequences of functions.

It helps to look at some graphs.

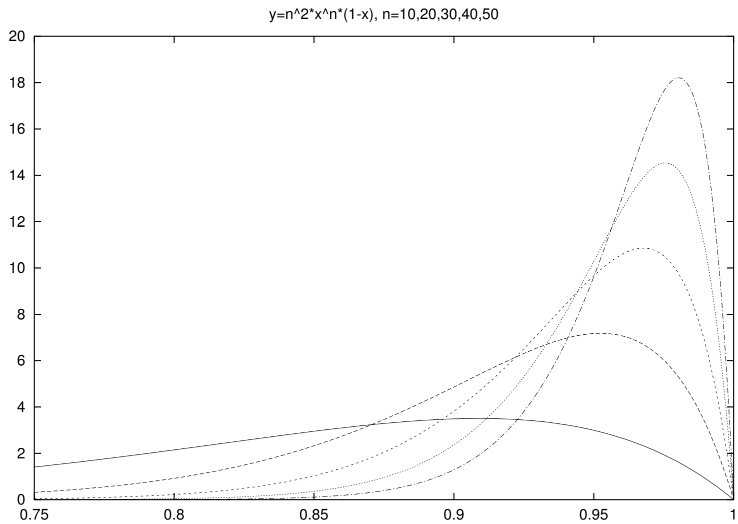

FIGURE 5. $f_n \to 0$, but $\int f_n \nrightarrow 0$

Figure 4 shows the graphs of $\cos^n(t/\sqrt{n})$ for $n = 1, 4, 9$, and the graph of $\exp(-t^2/2)$ (it is the outermost curve). The cosine powers get very close to each other, and to the exponential, over the whole interval $[-\pi, \pi]$. Take an arbitrarily narrow strip along the curve $\exp(-t^2/2)$. All the $\cos^n(t/\sqrt{n})$, from some n on, will fit into that strip[5].

This is the meaning of "$f_n \to f$ uniformly". Pick any narrow strip about the graph of the limit function f. Then the graphs of f_n fit into that strip, from some N onwards.

Clearly, the areas under the cosine curves approach the area under the exponential curve. This is an obvious, and important, consequence of "uniform convergence".

Figure 5 shows the graphs of the functions $f_n(x) = n^2 x^n (1 - x)$, defined for $0 \le x \le 1$. Clearly, $f_n(0) = f_n(1) = 0$ for all n, and for each individual $x \in (0, 1)$, the "geometric decay beats power growth" principle implies that $f_n(x) \to 0$ (see the beginning of Appendix B, if you are not sure what I mean). So $f_n(x) \to 0$ for every $x \in [0, 1]$. If the convergence $f_n \to 0$ were uniform on the closed interval $[0, 1]$, then–by definition of uniform convergence–you could pick an arbitrarily narrow strip about $y = 0$, say $|y| < \epsilon$, and all f_n, from some n on, would fit into that strip. This does not happen here. $f_n(x) \to 0$ at different rates for different x.

[5] the convenient strip has the form $\{(t, y) \mid \; |y - \exp(-t^2/2)| < \epsilon\}$. Compare this with equation (A-5).

This affects the behavior of the integrals $\int f_n$. We have

$$\int_0^1 n^2\,(x^n - x^{n+1})\,dx = n^2\,(\frac{1}{n+1} - \frac{1}{n+2}) = \frac{n^2}{(n+1)(n+2)} \to 1.$$

Again: even though $f_n(x) \to 0$ for every individual x, the integrals of the f_n do not approach zero. An area of size close to 1 migrates to the right. The graph becomes very high over a very narrow interval, and the integrals remain finite even though f_n becomes very small over the rest of $[0,1]$.

We express this by saying that the identity

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = \int_0^1 \lim_{n\to\infty} f_n(x)\,dx \tag{A-7}$$

does not hold. But if you look back at (A-4) and (A-1), you see that the property (A-7) is the key to the example.

Just to confuse matters: interchange of limit and integral may be valid even for sequences $\{f_n\}$ that do *not* converge uniformly. Take

$$f_n(x) = nx^n(1-x),\ 0 \le x \le 1.$$

As for $n^2x^n(1-x)$, one has $f_n(x) \to 0$ for each x, i.e.

$$\lim_{n\to\infty} f_n(x) = 0.$$

Because

$$\int_0^1 f_n(x)\,dx = \frac{n}{(n+1)(n+2)},$$

it is true that

$$\lim_{n\to\infty} \int f_n(x)\,dx = 0 = \int_0^1 \lim_{n\to\infty} f_n(x)\,dx.$$

However, convergence is not uniform. You should check this by drawing graphs of f_n for a few values of n. A more mathematical way to see it is to find the maximum value of f_n. It is $(n/(n+1))^{n+1}$, assumed at $x = n/(n+1)$. Curiously enough, the maximum value approaches $\exp(-1)$. No matter how large you take n, there will always be a bump in the graph that sticks up to about .36, preventing the graph from fitting into a narrow strip about the limit function, $f(x) \equiv 0$. Check this.

A.2. **Sequences of numbers.** First, a review of some simple, and likely familiar, definitions and facts about limits of sequences of numbers.

Definition A.1. A sequence $\{c_n\}_{n=1}^{\infty}$ *converges to* c as $n \to \infty$ if for every $\epsilon > 0$ there is an integer $N(\epsilon) > 0$ (i.e., depending in general on ϵ) such that

$$n > N(\epsilon) \text{ implies } |c_n - c| < \epsilon.$$

□

We will want to plug sequences into continuous functions, and use the following property (sometimes taken as definition of continuity, because it is easy to picture).

$$\text{If } F \text{ is continuous, then } c_n \to c \Longrightarrow F(c_n) \to F(c). \tag{A-8}$$

We will prove later that

$$n \ln(1 + \frac{a}{n}) \to a. \tag{A-9}$$

Taking F to be the exponential function exp, $c_n = n \ln(1 + \frac{a}{n})$, and $c = a$, we then obtain from (A-8) and (A-9) that

$$(1 + \frac{a}{n})^n = \exp[n \ln(1 + \frac{a}{n})] \to e^a, \tag{A-10}$$

which is one of our goals.

While (A-8) is a very practical definition of continuity, the "official" one can be adapted to more general settings; we will do later in the course.

Definition A.2. A function $F : \mathbb{R} \to \mathbb{R}$ is *continuous* at $c \in \mathbb{R}$, if for every $\epsilon > 0$ there is a $\delta(\epsilon)$ (i.e., depending in general on ϵ), such that

$$|x - c| < \delta(\epsilon) \text{ implies } |F(x) - F(c)| < \epsilon.$$

□

If f satisfies this condition, one can deduce the property (A-8). If you are not familiar with this style of proof, read it.

Claim:

Suppose that F is continuous at c, in the sense of Definition A.2. If $c_n \to c$, then $F(c_n) \to F(c)$.

We need to show that given $\epsilon > 0$, there is an $N(\epsilon) > 0$, such that $n > N(\epsilon)$ implies $|F(c_n) - F(c)| < \epsilon$. So let $\epsilon > 0$ be given. According to Definition A.2, we are guaranteed a $\delta(\epsilon)$ so that

$$|x - c| < \delta(\epsilon) \text{ implies } |F(x) - F(c)| < \epsilon.$$

Now use Definition A.1 of convergence, but substitute our δ for the ϵ in the definition: There is an $N(\delta) > 0$ such that

$$n > N(\delta) \text{ implies } |c_n - c| < \delta;$$

this N depends on δ, and through δ also on ϵ. The implications

$$n > N(\delta(\epsilon)) \Longrightarrow |c_n - c| < \delta \Longrightarrow |F(c_n) - F(c)| < \epsilon$$

establish the claim. □

A.3. **Sequences of functions.** So much for sequences of *numbers*, and how they behave when plugged into a continuous function; however, we will need to work with sequences of *functions*.

Let us return to uniform convergence. The picture one should have in mind is, as explained above:

(A-11) "the graphs of f_n are close to the graph of $\lim f_n$ over the whole interval"

Definition A.3. A sequence of functions $\{f_n\}_{n=1}^{\infty}$ *converges uniformly* to the function f on the interval $[-A, A]$, if for each $\epsilon > 0$ there is an N such that

$$n > N \quad \text{implies} \quad |f_n(a) - f(a)| < \epsilon, \quad \text{for all} \quad a \in [-A, A].$$

($A = \infty$ could occur, or the interval might be replaced by one of the more general form $[\alpha, \beta]$). □

Example A.1. Remember that the functions $f_n(x) = n^2 x^n (1 - x)$ do not approach the identically zero function uniformly on $[0, 1]$. It was easy to see that the verbal description (A-11) is violated, but perhaps it is less clear how to deal with the ϵ's and N's. If we want to show that the criterion of the definition is not satisfied, we must show that its negation holds:

"there exists $\epsilon > 0$, such that for every $N > 0$, there exist $n > N$ and $a \in [0, 1]$ for which $|f_n(a) - 0| > \epsilon$".

ASIDE: Since negating more involved statements could, later in the course, be more tricky, a brief elaboration might be warranted.

The definition is built from "for all" and "there exists" statements, nested in a Russian-doll sort of way.

1) for all $\epsilon > 0$, the following is true:
2) there exists an N for which the following is true:
3) for all $n > N$ the following is true:
4) for all $a \in [0, 1]$, the following is true:
5) $|f_n(a) - f(a)| < \epsilon$.

The negation of "X is true for all" is "X is false for at least one", and the negation of "there exists one for which X is true" is "X is false for all".

Now let us negate the sequence of statements 1)-5).

There is an $\epsilon > 0$ for which 2) is false.
Thus, for that ϵ and every choice of $N > 0$, 3) is false. Choose an N arbitrarily and continue.
For that ϵ and the chosen N, there is an $n > N$ such that 4) is false.
For that ϵ, the chosen N, and that n, there is an $a \in [0, 1]$ such that 5) is false.

Thus, for that ϵ and every choice of N, and that n and that a, we have $|f_n(a) - f(a)| \geq \epsilon$.

Let us show that our sequence f_n does not converge uniformly. Take $\epsilon = 1$. (The argument can be adapted to your favorite ϵ.)

The maximum value of f_n is attained at $a = n/(n+1)$, and equals

$$\max(f_n) = n\left(\frac{n}{n+1}\right)^{n+1} = n\left(\left(\frac{n+1}{n}\right)^{(n+1)}\right)^{-1},$$

which tends to ∞ with n, because the factor next to the n tends to $\exp(-1)$, by (A-10). From some n on, say starting with M, this maximum value is > 2. Then, given an arbitrary N, pick $n > N$ so large (larger than M) that $\max(f_n) > 2$. Now recall that 2 is bigger than 1. Ergo, $\max(f_n) > \epsilon$ and convergence is not uniform.

Note that this argument simply says, in a roundabout way, that for every large n, the graph of f_n goes up higher than 1, somewhere. So it will never fit into a width 1 strip[6]. □

This next proposition is one of our goals[7].

Proposition A.1. *Suppose $f_n \to f$ uniformly on the* finite *interval $[\alpha, \beta]$. Then*

$$\int_\alpha^\beta f_n(a)\, da \to \int_\alpha^\beta f(a)\, da.$$

Proof. Let $\epsilon > 0$ be given. We have

$$\left|\int_\alpha^\beta f_n(a)\, da - \int_\alpha^\beta f(a)\, da\right| = \left|\int_\alpha^\beta (f_n(a) - f(a))\, da\right| \leq \int_\alpha^\beta |f_n(a) - f(a)|\, da.$$

By uniform convergence, there is an $N > 0$ such that

$$|f_n(a) - f(a)| < \frac{\epsilon}{\beta - \alpha},\ a \in [\alpha, \beta].$$

So

$$\int_\alpha^\beta |f_n(a) - f(a)|\, da < \int_\alpha^\beta \frac{\epsilon}{\beta - \alpha}\, da = \epsilon,$$

as desired. □

Exercise A.1. The proof used the inequality $|\int g| \leq \int |g|$. Why is this true?

---

[6] For every really really large n, the graph goes up higher than 10^{23}, and will not fit into an $\epsilon = 10^{23}$-strip. That is another possible ϵ

[7] If you know that the integral may not be defined for some wild functions—I am not worrying about this. Our functions are nice.

Remark A.1. The integration interval has to be finite for this particular proof to make sense, since you need $(\beta - \alpha)$ to be a finite number. In fact, the proposition is no longer true on an infinite integration interval. Let

$$f_n(a) = \begin{cases} \frac{1}{n}, & \text{for } n \le a \le 2n, \\ 0, & \text{otherwise.} \end{cases}$$

Then $f_n \to 0$ uniformly on all of $\mathbb{R}$, while

$$\int_{-\infty}^{\infty} f_n(a)\, da = 1 \quad \text{for all } n = 1, 2, 3, \ldots.$$

□

There is one final uniform convergence issue. Earlier, we had $c_n \to c$ and concluded that (F continuous) $F(c_n) \to F(c)$. Now, if $f_n \to f$ uniformly, does $F \circ f_n \to F \circ f$ uniformly ($F \circ f$ is the composition, $F(f(a))$)?

It is really tedious to write down the full statement, so let me just indicate the idea. Suppose the Mean Value Theorem applies to F, and that $|F'(c)| <$ some K on a big interval. Thus,

$$F(y) - F(x) = F'(c)(y - x), \text{ or } |F(y) - F(x)| \le K|y - x|$$

for some $c \in [x, y]$. It follows that

$$|F(f_n(a)) - F(f(a))| \le K|f_n(a) - f(a)|,$$

and if n is large enough, the right side will be as small as you please.

Of course, one should worry about ranges and domains; the argument that goes into F is the value that comes out of f_n or f. The details are left as exercise.

Proposition A.2. *Suppose the functions f_n are continuous on $[-A, A]$ (or $[\alpha, \beta]$, more generally) and that they converge uniformly to the continuous function f on that interval*[8]*. Suppose the range of the f_n and f lies in some open subinterval of $[-C, C]$. Let $F : [-C, C] \to \mathbb{R}$ be continuously differentiable, and suppose there is a number K such that $|F'(c)| < K$ for $c \in [-C, C]$. Then $F \circ f_n \to F \circ f$ uniformly on $[-A, A]$.*

[8]it turns out that such a limiting function f is automatically continuous

A.4. **Compound interest proofs.**

Lemma A.2. *Fix $A > 0$. Let $\{b_k\}$ be a sequence of functions converging uniformly to zero on $[-A, A]$. Then for every $\epsilon > 0$, there is an integer $N > 0$ (depending on $A, \{b_k\}, \epsilon$) such that $n > N$ implies*

$$|e^a - (1 + \frac{a + b_n(a)}{n})^n| < \epsilon, \text{ for all } a \in [-A, A].$$

Instead of proving this Lemma all at once, I will start with the simplest case, and repeat the argument with small modifications. This is a bit tedious, but conceivably the strategy can be helpful to you if you haven't spent a lot of time with proofs.

Proof. Let us first prove Lemma A.1. Assertion (A-2) will follow from the relation obtained by taking logarithms,

$$\lim_{n\to\infty} n \ln(1 + \frac{a}{n}) = a. \tag{A-12}$$

Recall the Taylor series for $\ln(1 + x)$, valid for $|x| < 1$:

$$\ln(1 + x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots .$$

We use this in the left side of (A-3).

$$n \ln(1 + \frac{a}{n}) = n(\frac{a}{n} + \frac{a^2}{2n^2} + \frac{a^3}{3n^3} + \cdots)$$

$$= a + \overbrace{\frac{a^2}{n}(\frac{1}{2} + \frac{a}{3n} + \frac{a^2}{4n^2} + \cdots)}^{\rho_n}. \tag{A-13}$$

We need to show that the error term ρ_n can be made arbitrarily small by taking n large enough. (Caution: the "n" is *not* the summation index! You have a different series for each n). Estimate:

$$\left|\frac{1}{2} + \frac{a}{3n} + \frac{a^2}{4n^2} + \cdots\right|$$

$$\leq \frac{1}{2} + \frac{|a|}{3n} + \frac{|a|^2}{4n^2} + \cdots \qquad \text{(triangle inequality)}$$

$$\leq 1 + \frac{|a|}{n} + (\frac{|a|}{n})^2 + (\frac{|a|}{n})^3 + \cdots \qquad \text{(decrease denominator)}. \tag{A-14}$$

This last series is the familiar geometric series, $\sum r^k$, with $r = \frac{|a|}{n}$ in our case. It converges when $0 < r < 1$, and this will happen as soon as $\frac{|a|}{n} < 1$. Then the sum is $1/(1 - r)$.

Now let $\epsilon > 0$ be given. Choose N_0 (it will depend on a and ϵ) so that $n > N_0$ implies $\frac{|a|}{n} < \frac{1}{2}$. Then the geometric series converges, and moreover

$$\frac{1}{1 - \frac{|a|}{n}} < 2.$$

The error term is then estimated by

$$|\rho_n| < \frac{|a|^2}{n} \frac{1}{1 - \frac{|a|}{n}} < 2\frac{|a|^2}{n}.$$

Next, choose $N \geq N_0$ so that $n > N$ implies

$$2\frac{|a|^2}{n} < \epsilon.$$

For $n > N$, therefore, $|\rho_n| < \epsilon$, or

$$|\ln(e^a) - \ln(1 + \frac{a}{n})^n| < \epsilon.$$

This is the translation of

$$\lim_{n\to\infty} \ln(1 + \frac{a}{n})^n = \ln(e^a) \qquad \text{(A-15)}$$

into official mathematics. The desired (A-2) now follows by taking exponentials.

Next case. If you can follow the rest of the proof, you win a prize. I have likely made it uglier than necessary.

So far, a has been a single fixed number. If, however, it is allowed to range over an interval $[-A, A]$, we are asking about the convergence of $f_n(a) = n \ln(1 + \frac{a}{n})$ to $f(a) = a$. The convergence is uniform. To see this, we replace the upper bounds in the proof just concluded by their largest possible value. Take N_0 (it will depend on ϵ and now on A) so large that for $n > N_0$,

$$\frac{1}{1 - \frac{A}{n}} < 2$$

(the same as $\frac{A}{n} < \frac{1}{2}$). Then

$$\frac{1}{1 - \frac{|a|}{n}} \leq \frac{1}{1 - \frac{A}{n}} < 2, \text{ for } |a| \leq A.$$

Finally, given $\epsilon > 0$, choose $N \geq N_0$ so that for $n > N$

$$2\frac{A^2}{n} < \epsilon.$$

Then the error term from (A-13), which is now written $\rho_n(a)$ since a is no longer fixed, satisfies

$$|\rho_n(a)| < \frac{|a|^2}{n}\frac{1}{1-\frac{|a|}{n}} \leq \frac{A^2}{n}\frac{1}{1-\frac{A}{n}} < 2\frac{A^2}{n} < \epsilon.$$

Thus we have

$$\text{for all } a \in [-A, A], |a - n\ln(1+\frac{a}{n})| < \epsilon \text{ when } n > N,$$

which is uniform convergence. Applying Proposition A.2, we conclude that

$$(1+\frac{a}{n})^n \to e^a, \quad \text{uniformly on } [-A, A].$$

Last case: the general statement. Instead of (A-13), we have

$$\text{(A-16)} \quad n\ln(1+\frac{a+b_n(a)}{n}) = a + b_n(a) + \overbrace{\frac{a_n^2}{n}(\frac{1}{2}+\frac{a_n}{3n}+\frac{a_n^2}{4n^2}+\cdots)}^{\rho_n(a)},$$

where a_n is used as abbreviation for $a + b_n(a)$ in the sum.

Choose $0 < \epsilon < 1$. (If we can make something less than any such ϵ, we can obviously make it less than big ϵ's). Because $b_n \to 0$ uniformly, there is an N_0 such that $n > N_0$ implies $|b_n(a)| < \frac{\epsilon}{2} < 1$ for all $a \in [-A, A]$ (the reason for taking $1/2$ of ϵ becomes clear shortly). For such n, we have

$$|a_n| = |a + b_n(a)| \leq |a| + |b_n(a)| < |a| + 1 \leq A + 1.$$

Next, take $N \geq N_0$ so that simultaneously

$$\frac{1}{1-\frac{A+1}{n}} < 2$$

and

$$2\frac{(A+1)^2}{n} < \frac{\epsilon}{2}$$

when $n > N$. Just as before,

$$|\rho_n(a)| < \frac{\epsilon}{2}.$$

One arrives at

$$n\ln(1+\frac{a+b_n(a)}{n}) = a + b_n(a) + \rho_n(a),$$

where

$$|\rho_n(a)| < \frac{\epsilon}{2}, \text{ for all } a \in [-A, A] \text{ and } n > N.$$

Therefore

$$|a - n\ln(1 + \frac{a + b_n(a)}{n})| < |\rho_n(a)| + |b_n(a)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Another application of Proposition A.2 proves Lemma A.2 in full glory. □

A.5. **The corollary.** It is now time to prove Corollary A.1. Recall, from Calculus, Taylor's formula with remainder (also called Lagrange's formula, or Lagrange's remainder form):

$$f(u) = f(0) + f'(0)u + \frac{1}{2!}f''(0)u^2 + \frac{1}{3!}f'''(0)u^3 + \frac{1}{4!}f^{(\mathrm{iv})}(c)u^4; \tag{A-17}$$

c is some number known to lie between 0 and u. Applied to cos, this gives

$$\cos u = 1 - \frac{1}{2}u^2 + \frac{1}{24}(\cos c)\, u^4.$$

Set $u = \frac{t}{\sqrt{n}}$:

$$\cos(\frac{t}{\sqrt{n}}) = 1 - \frac{t^2}{2n} + \frac{1}{24}(\cos c)\frac{t^4}{n^2}. \tag{A-18}$$

Here, c will lie between 0 and $\frac{t}{\sqrt{n}}$, and will depend on t and n. Writing the right side of (A-18) in the form

$$1 + \frac{-t^2/2 + (1/24)(\cos c)(t^4/n)}{n},$$

we see that we can identify $-t^2/2$ with a and $(1/24)(\cos c)(t^4/n)$ with $b_n(a)$ in Lemma A.2. Since $|t| \leq \beta$ and $|\cos c| \leq 1$, it is also true that

$$|(1/24)(\cos c)(t^4/n)| \leq \frac{1}{n}\frac{\beta^4}{24},$$

from which it is clear that $b_n \to 0$ uniformly on $[-\beta, \beta]$. Thus, Lemma A.2 applies, and we have proved the corollary.

Primarily, we wanted the limiting relation (A-6) for the integrals. This immediately follows from Proposition A.1.

B. Appendix: Asymptotic equivalence

Here is an example. Let $r \in (0, 1)$ and $p > 0$ be fixed. We know that as $n \to \infty$, both r^n and n^{-p} tend to zero. Do they approach zero at the same rate? One way to see that the answer is "no" is to look at the ratio

$$\frac{r^n}{n^{-p}} = n^p r^n.$$

This has limit zero[9]. On the other hand,

$$n^{-p} \text{ and } (\sqrt{n^2+n}+5)^{-p}$$

tend to zero at the same rate, since

$$\frac{n^{-p}}{(\sqrt{n^2+n}+5)^{-p}} = \left(\frac{\sqrt{n^2+n}+5}{n}\right)^p = (\sqrt{1+\frac{1}{n}}+\frac{5}{n^2})^p,$$

which has limit 1.

This leads to a definition.

Definition B.1. If $\{a_n\}, \{b_n\}$ are two sequences for which

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1, \tag{B-1}$$

we say that the sequences are *asymptotically equivalent*, and write

$$a_n \sim b_n \text{ (as } n \to \infty \text{, usually left unsaid).} \tag{B-2}$$

□

In the examples above, the two sequences both had limit zero. They might just as well both tend to infinity, in which case the statement (B-2) would mean that they grow at the same rate. The same notation is applied to functions: if $f(x), g(x) \to 0$ or $f(x), g(x) \to \infty$ as $x \to A$ (A could be finite or $\pm\infty$) and

$$\lim_{x\to A} \frac{f(x)}{g(x)} = 1, \tag{B-3}$$

we write

$$f \sim g \text{ as } x \to A.$$

Here, we usually insert "as $x \to A$", since x might be asked to approach any one of many possible values A. In the case of sequences, the subscript n *has* to go to ∞ (unless $n \to -\infty$), so it is hardly ever necessary to mention the fate of n.

Remark B.1. If a_n and b_n both have the same finite, nonzero limit L as $n \to \infty$, relation (B-1) is automatic and contains no information. What might be useful is a comparison of the *rate of approach to the limit*, i.e. consideration of the ratio

$$\frac{a_n - L}{b_n - L}.$$

□

[9] The logarithm of the right side is $p \ln n + n \ln r$. Recall that n increases much faster than $\ln n$, so the $n \ln r$ dominates and it tends to $-\infty$. So the log of the right side tends to $-\infty$, and the right side tends to zero. Calculus students learn that "geometric decay beats power growth".

Remark B.2. The concept of asymptotic equivalence, as defined above, is only the most restrictive of a number of related (and important) notions. One might wonder when the ratios in (B-2) or (B-3) remain bounded–in that case, the sequences or functions are comparable in a weaker sense. This is written: $f = O(g)$ (read: "f is big Oh of g"). Or one might want the ratio to approach zero, signifying that the numerator decays more rapidly than the denominator. Written: $f = o(g)$ ("f is little oh of g"). For example, the assertion

$$e^x = 1 + x + o(x) \text{ as } x \to 0$$

means that the error term in the approximation of e^x by $1 + x$ tends to zero faster than x, for small x. We will not need these variants, but they are constantly used in applied analysis. □

C. APPENDIX: STIRLING'S FORMULA

C.1. **The formula.** It is difficult to get a handle on the growth of $n!$ as n increases. For example, $90! \sim 1.48 \times 10^{138}$, $100! \sim 9.32 \times 10^{157}$ (my computer calculated $\sum_1^{90} \ln k$, and then exponentiated). Is there a pattern? In probability theory one asks about the binomial coefficient

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2}, \tag{C-1}$$

which counts the number of outcomes of the experiment of tossing a coin $2n$ times in which exactly n heads (and, of course, also exactly n tails) occur[10]. How big is (C-1)? One must replace the factorials by something manageable, and Stirling's formula provides the tool (proof below).

Proposition C.1.

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}. \tag{C-2}$$

It is important to understand what (C-2) does claim, and what it does not claim. According to our definition of "$\sim$",

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

It is not true that the *difference*

$$|n! - \sqrt{2\pi n}\, n^n e^{-n}|$$

is at all small. What is small is the relative error, since

$$\lim_{n\to\infty} \frac{|n! - \sqrt{2\pi n}\, n^n e^{-n}|}{n!} = \lim_{n\to\infty} |1 - \frac{\sqrt{2\pi n}\, n^n e^{-n}}{n!}| = |1-1| = 0.$$

[10] $2^{-2n}\binom{2n}{n}$ is also the coefficient of x^n in the Taylor series of $1/\sqrt{1-x}$

A table will illustrate this.

n	5	10	25
$n!$	120	$3,628,800$	1.551×10^{25}
Stirling	118.02	$3,598,695$	1.546×10^{25}
Ratio	1.0168	1.0084	1.0033
Rel. Error	.0168	.0084	.0033
Difference	1.98	$30,104$	5.162×10^{22}

Remark C.1. Stirling's formula (C-2) can be improved. It is possible to show that

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}[1 + \frac{1}{12n} + \frac{1}{288n^2} + O(n^{-3})]$$

using the "big Oh" notation from Remark B.2. Thus $O(n^{-3})$ stands for numbers T_n satisfying

$$|T_n/n^{-3}| = |n^3\, T_n| < C, n \to \infty,$$

for some fixed (but perhaps unknown) number C. The T_n's are therefore smallish, but notice that they are multiplied by the huge prefactor outside the $[\ldots]$. Big Oh can be further replaced by $\frac{-139}{51840n^3} + O(n^{-4})$, etc. *ad nauseum.* Math 583 deals with techniques for obtaining "asymptotic expansions" like this. □

Example C.1. To illustrate Stirling's formula, let us find the asymptotic form of the binomial coefficient (C-1).

$$\text{(C-3)} \qquad \frac{(2n)!}{(n!)^2} \sim \frac{\sqrt{2\pi(2n)}\,(2n)^{2n}e^{-2n}}{(\sqrt{2\pi n}\, n^n e^{-n})^2} = \frac{2^{2n}}{\sqrt{\pi n}}.$$

This applies in probability theory as follows. There are altogether 2^{2n} outcomes of the experiment of tossing a coin $2n$ times, i.e., 2^{2n} different sequences of "H" and "T" (heads or tails). If the coin is fair, each sequence has probability $1/2^{2n}$. Therefore, the probability of getting exactly n heads is

$$\frac{1}{2^{2n}}\binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}}.$$

E.g., the probability of getting exactly 500 heads in 1000 tosses is about $1/\sqrt{1000\pi} \sim .000318$. Not likely. So the question arises: what is the "law of averages", precisely? One version of it, illustrating high-dimensional geometry, is presented in Section 5, see Remark 5.1. □

Exercise C.1. Use Lemma 3.1 to derive the asymptotic approximation in (C-3), thereby circumventing Stirling's formula. □

C.2. **Proof of Stirling's formula.** This subsection is tacked on only because a proof of this useful formula is rarely encountered except, perhaps, in a more advanced course that includes asymptotic methods. One standard approach is via the Γ-*function*[11],

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

One shows, using integration by parts, that $\Gamma(n+1) = n!$, for $n = 0, 1, 2, \ldots$. On the other hand, there are techniques for approximating integrals like this one for large values of x, and those yield Stirling's formula (as well as the improvements in Remark C.1). There is, however, an "elementary" proof as well (elementary $\neq$ easy), which relies on something called the *Euler summation formula* and Wallis' product (3.9). I follow K. Knopp, *Theory and Application of Infinite Series*, §64.

We want to prove that

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

or the equivalent statement for the logarithms,

$$\lim_{n\to\infty}\Big(\sum_{k=1}^{n} \ln k - \ln\sqrt{2\pi} - (n+\frac{1}{2})\ln n + n\Big) = 0. \tag{C-4}$$

The sum

$$\sum_{k=1}^{n} \ln k$$

is roughly comparable to

$$\int_1^n \ln x\,dx = (x\ln x - x)\Big|_1^n = n\ln n - n + 1.$$

We begin to see where Stirling might come from, but evidently the approximation of sum by integral is too crude. That is where Euler's more refined method comes in. We need the "greatest integer function", defined by

$$[x] = \text{ greatest integer } \leq x.$$

Proposition C.2 (Euler summation formula)**.** *Let f be continuously differentiable on $[1, \infty)$. Abbreviate $f(k) = f_k$. Then for $n = 1, 2, \ldots$,*

(C-5)

$$f_1 + f_2 + \cdots + f_n = \int_1^n f(x)\,dx + \frac{1}{2}(f_1 + f_n) + \int_1^n (x - [x] - \frac{1}{2})f'(x)\,dx.$$

[11] 527 Notes, pp. 262-263

Notice that the first two terms on the right already give (for $f = \ln$)

$$(n \ln n - n + 1) + \frac{1}{2}(0 + \ln n) = (n + \frac{1}{2}) \ln n - n + 1.$$

Proof. For $k = 1, 2, \ldots, n-1$, we have by integration by parts:

(C-6)

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

Since $k \le x \le k+1$ in the left side of (C-6), we may replace the k in the integrand by $[x]$ (the fact that $[x] \ne k$ at the upper limit, where $x = k + 1 = [x]$, does not affect the value of the integral). Thus

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

Summing these equations over k, we get on the left side

$$\frac{1}{2}(f_1 + f_2) + \frac{1}{2}(f_2 + f_3) + \cdots + \frac{1}{2}(f_{n-1} + f_n) = \frac{1}{2} f_1 + f_2 + f_3 + \cdots + f_{n-1} + \frac{1}{2} f_n,$$

and on the right side

$$\int_1^n f(x)\, dx + \int_1^n (x - [x] - \frac{1}{2}) f'(x)\, dx.$$

Add $\frac{1}{2}(f_1 + f_n)$ to both sides to get Euler's formula. □

Notice that $P_1(x) \stackrel{\text{def}}{=} x - [x] - \frac{1}{2}$ is a periodic function, period 1, with jump discontinuities at the integers (Figure 6). We are about to integrate by part in the integral $\int P_1 \cdot f'$ in Euler's formula; in preparation observe that P_1 is the derivative of the periodic *bounded* function (Figure 7)

$$P_2(x) \stackrel{\text{def}}{=} \frac{1}{2}(x - [x])^2 - \frac{1}{2}(x - [x]) + \frac{1}{12}.$$

P_2 has corners at the integers, as expected, since P_1 has jumps there. Note also that $P_2(k) = 1/12$ for $k = 0, 1, 2, \ldots$.

Lemma C.1. *If f'' exists, then*

$$\int_1^n P_1(x) f'(x)\, dx = (P_2(x) f'(x))\Big|_1^n + \int_1^n P_2(x) f''(x)\, dx.$$

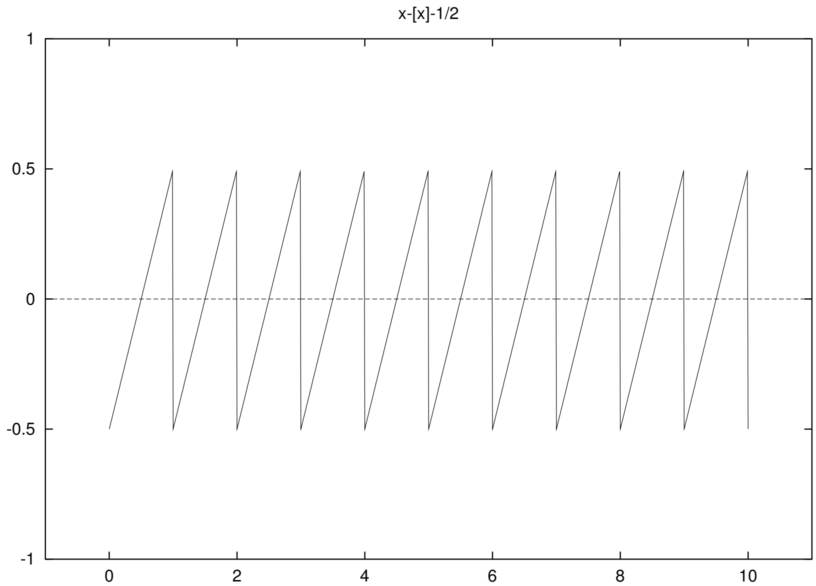

FIGURE 6. The function P_1 in Euler's formula

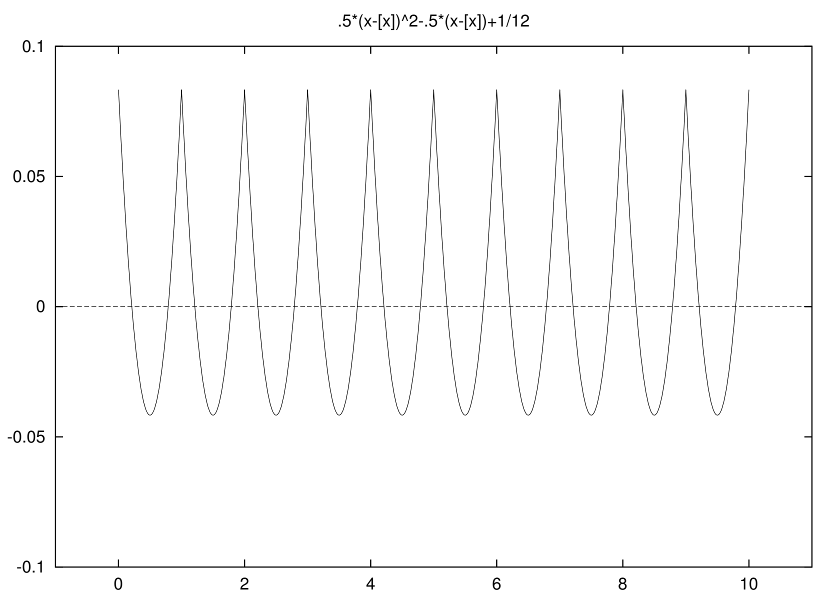

FIGURE 7. Antiderivative of P_1

Let us now set $f(x) = \ln x$. We already have gotten to

(C-7) $$\sum_{k=1}^{n} \ln k = (n + \frac{1}{2}) \ln n - n + \gamma_n,$$

where, by the lemma,

$$\gamma_n = \left.\frac{P_2(x)}{x}\right|_1^n + \int_1^n \frac{P_2(x)}{x^2}\,dx.$$

The integral converges as $n \to \infty$, by comparison with

$$\int_1^\infty \frac{1}{x^2}\,dx.$$

The term before the integral is

$$\text{(C-8)}\qquad \frac{1}{n}P_2(n) - P_2(1) = \frac{1}{12n} - \frac{1}{12} \to -\frac{1}{12},\ n \to \infty.$$

We conclude that γ_n has a limit γ as $n \to \infty$. The last step is to show that $\gamma = \ln\sqrt{2\pi}$.

The method strikes me as being exceedingly tricky. You have to know what you are looking for, or be very clever. Go back to (3.9),

$$\lim_{p\to\infty} \frac{2^2 4^2 \cdots (2p)^2}{3^2 5^2 \cdots (2p-1)^2}\frac{1}{2p+1} = \frac{\pi}{2}$$

or

$$\text{(C-9)}\qquad \lim_{p\to\infty} \frac{2\cdot 4\cdot 6\cdots(2p)}{3\cdot 5\cdots(2p-1)}\frac{1}{\sqrt{2p+1}} = \sqrt{\frac{\pi}{2}}.$$

We are about to work γ into this.

By (C-7),

$$\begin{aligned} 2\ln(2\cdot 4\cdots(2p)) &= 2\ln(2^p p!) = 2p\ln 2 + 2\ln p! \\ &= 2p\ln 2 + \underbrace{(2p+1)\ln p - 2p + 2\gamma_p}_{\text{from (C-7)}} \\ &= (2p+1)\ln 2p - 2p - \ln 2 + 2\gamma_p. \end{aligned}$$

Also,

$$\ln(2p+1)! = (2p + \frac{3}{2})\ln(2p+1) - (2p+1) + \gamma_{2p+1}.$$

Subtract. On the left side, you get a difference of logs, which is the log of the quotient

$$\frac{2^2 4^2 \cdots (2p)^2}{1\cdot 2\cdot 3\cdot 4\cdots(2p+1)} = \frac{2\cdot 4\cdot 6\cdots(2p)}{1\cdot 3\cdot 5\cdots(2p-1)}\frac{1}{2p+1}.$$

The expression on the right is rearranged a bit (check it), resulting in

$$\begin{aligned} \ln\left(\frac{2\cdot 4\cdot 6\cdots(2p)}{1\cdot 3\cdot 5\cdots(2p-1)}\frac{1}{2p+1}\right) &= (2p+1)\ln\left(1 - \frac{1}{2p+1}\right) \\ &\quad - \frac{1}{2}\ln(2p+1) + 1 - \ln 2 + 2\gamma_p - \gamma_{2p+1}. \end{aligned}$$

Move $-\frac{1}{2}\ln(2p+1)$ to the left side (where it produces a $\frac{1}{\sqrt{2p+1}}$ inside the log).

Now everything in sight has a limit. The left side tends to $\ln\sqrt{\frac{\pi}{2}}$ by (C-9), the term

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

tends to -1 by the compund interest formula (A-9), and

$$2\gamma_p - \gamma_{2p+1} \to 2\gamma - \gamma = \gamma.$$

(Here it is important to know that the sequence $\{\gamma_n\}$ has a limit—that was the reason for bringing in $P_2(x)$). Thus

$$\ln\sqrt{\frac{\pi}{2}} = -1 + 1 - \ln 2 + \gamma,$$

from which (!)

$$\gamma = \ln\sqrt{2\pi}.$$

Remark C.2. There is an "improved" Euler formula. Namely, one can continue with the integration by parts:

$$\begin{aligned}
\int_1^n P_1 f' &= [P_2 f']_1^n - \int_1^n P_2 f'' \\
&= \frac{1}{12} f'|_1^n - [P_3 f'']_1^n + \int_1^n P_3 f''' \\
&= \frac{1}{12} f'|_1^n - [P_3 f'']_1^n + \int_1^n P_3 f''';
\end{aligned}$$

here

$$P_3(x) = \frac{1}{6}(x-[x])^3 - \frac{1}{4}(x-[x])^2 + \frac{1}{12}(x-[x])$$

is the bounded periodic antiderivative of P_2. In this way, one gets the $\frac{1}{12n}$ term, and subsequent terms, in the improved Stirling formula mentioned in Remark C.1. You can see the $\frac{1}{12n}$ appearing in (C-8). □