

# Calculus and Probability Review

CS 70 STAFF

December 15, 2019

This is a quick refresher of single and multi variable calculus in the context of probability. If any material needs refreshing, [1] is the canonical calculus textbook, while Khan Academy has fantastic video explanations. Finally, this is a reminder that **calculus will not be over-emphasized on the final**; after all, this is a probability course, not a calculus course.

## 1 Derivatives

Single variable calculus starts with the notion of a derivative, which represents the *instantaneous* rate of change of a function  $f(x)$ . Formally,

**Definition 1** (Derivative). The **derivative** of a function  $f(x)$  is defined to be

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

We sometimes also denote it by  $\frac{df}{dx}$  or  $\frac{d}{dx}f(x)$  to emphasize that the derivative is with respect to  $x$ .

The idea is that this value gives us the slope of the secant line passing through  $(x, f(x))$  and  $(x+h, f(x+h))$ , and as  $h$  becomes smaller and smaller, the secant line becomes the tangent line at the point  $x$ .

This definition is seldomly used once we know the basics of dealing with derivatives though. That comes in the form of the following common forms and properties. Assume  $c \in \mathbb{R}$  is a constant.

- $\frac{d}{dx}c = 0$ .
- $\frac{d}{dx}x^n = nx^{n-1}$ .
- $\frac{d}{dx}\sin x = \cos x$ .
- $\frac{d}{dx}\cos x = -\sin x$ .
- $\frac{d}{dx}e^x = e^x$  (the only such function!).
- $\frac{d}{dx}c^x = c^x \ln c$ .
- $\frac{d}{dx}\ln x = \frac{1}{x}$ .

Of course, derivatives add and subtract, and with the next two rules, you're ready to differentiate nearly any function you come across.

**Lemma 2** (Product Rule). For any two functions  $f(x)$  and  $g(x)$ ,

$$\frac{d}{dx}(f \cdot g)(x) = f'(x)g(x) + f(x)g'(x)$$

or in other words,

$$(fg)' = f'g + fg'.$$

**Lemma 3** (Chain Rule). Suppose I have a function  $f(x) = g(h(x))$ . Then

$$f'(x) = g'(h(x)) \cdot h'(x).$$

Equivalently,

$$\frac{df}{dx} = \frac{df}{dh} \frac{dh}{dx}.$$

In other words, I keep unrolling my derivatives from outside in when I use the chain rule.

**Example 1.1.** Compute the derivative of  $f(x) = x^2 e^{2x}$ .

*Solution.* Applying the product rule followed by the chain rule gives

$$f'(x) = (x^2)' e^{2x} + x^2 (e^{2x})' = 2x e^{2x} + x^2 e^{2x} (2x)' = 2x e^{2x} + 2x^2 e^{2x}.$$

□

**Exercise 1.2.** Find the derivatives of the following functions:

(a)  $F(x) = 5(5 - x)^4$

(b)  $F(x) = e^{-3x}$

(c)  $F(x) = \sqrt{x^3}$

## 2 Integrals

### 2.1 Definitions

Complementing the derivative is first the *definite* integral, which represents the signed<sup>1</sup> area under the curve. The nice interpretations and motivations for this come from physics, so I'll move on and give an informal definition.

**Definition 4** (Definite Integral). The **definite integral**  $\int_a^b f(x) dx$  is the signed area of the region of the  $xy$ -plane bounded by the graph of  $f$ , the  $x$ -axis, and the vertical lines  $x = a$ ,  $x = b$ .

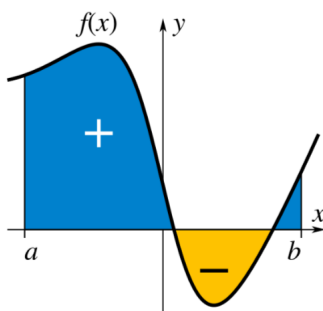


Figure 1: An example of such an integral.

It turns out that integration and differentiation are inverse operations. So we have the notion of an *antiderivative*, a function  $F$  whose derivative is a given function  $f$ , which is also sometimes called an *indefinite* integral.

**Definition 5** (Indefinite Integral). An antiderivative  $F(x)$  of  $f(x)$  is called an **indefinite integral**, and written as

$$F(x) = \int f(x) dx.$$

<sup>1</sup>Meaning positive if we're above the  $x$ -axis and negative otherwise.

One important point to note is that since constants vanish under differentiation, given an antiderivative  $F(x)$ , any other function  $F(x) + c$  where  $c$  is a constant is also an antiderivative. Never forget to plus  $c$ !

In the midst of all this back and forth, it's still not clear at all how we'd compute the area under the curve in order to actually find the definite integral. This is where the Fundamental Theorem of Calculus comes into play; it creates the connection between the definite and indefinite integrals (and therefore between definite integrals and derivatives).

**Theorem 6** (Fundamental Theorem of Calculus). If  $f$  is a real-valued, integrable function on  $[a, b]$ , and  $F$  is an antiderivative of  $f$  on  $[a, b]$  (so that  $F' = f$ ), then

$$\int_a^b f(x)dx = F(b) - F(a).$$

This theorem is paramount for going back and forth between integrals and derivatives. If you think about it, this theorem is what lets us use CDFs to evaluate our pdfs. So under the surface, problems like Example 4.3 are really using the fundamental theorem somewhere.

Here are some common integrals you should know:

- $\int x^n dx = \frac{x^{n+1}}{n+1} + C$  for  $n \neq -1$ .
- $\int \frac{1}{x} dx = \ln|x| + C$
- $\int e^x dx = e^x + C$
- $\int \sin x dx = -\cos x + C$
- $\int \cos x dx = \sin x + C$

**Example 2.1.** Compute  $\int_0^4 x^2 dx$ .

*Solution.* By inspection (or by looking at the list above) we find that an antiderivative is  $F(x) = \frac{x^3}{3}$ , and so by the fundamental theorem, our answer is just  $F(4) - F(0)$ . Written more concisely,

$$\int_0^4 x^2 dx = \frac{x^3}{3} \Big|_0^4 = \frac{4^3}{3} - \frac{0}{3} = \frac{64}{3}.$$

□

## 2.2 Integrating techniques: $u$ -sub, integration by parts, and others

There's two common techniques you'll use when solving integrals. The first is pretty simple. It's utilizing a change of variables, otherwise known as  $u$ -substitution.

**Example 2.2.** Compute  $\int_0^5 e^{-6x} dx$ .

*Solution.* This isn't an integral of one of the above forms since we have extra coefficients in the exponent, so we'll perform a change of variables to get it to look like it. Let  $u = -6x$ , so  $du = -6dx$  or  $-\frac{1}{6}du = dx$ . Now if we change variables, we will get an antiderivative of

$$\int e^{-6x} dx = \int -\frac{1}{6}e^u du = -\frac{1}{6}e^u + C = -\frac{1}{6}e^{-6x} + C.$$

Plugging this in gives

$$\int_0^5 e^{-6x} dx = -\frac{1}{6}e^{-6x} \Big|_0^5 = -\frac{1}{6}e^{-30} + \frac{1}{6}.$$

□

The second technique is *integration by parts*, which is not as straightforward. The idea is that we want to be able to integrate products of functions that aren't as nice. This reminds us of the product rule for derivatives, which states that if I had two functions  $u(x)$  and  $v(x)$ , then

$$(uv)' = u'v + uv'.$$

Integrating both sides with respect to  $x$  gives

$$u(x)v(x) = \int (u(x)v(x))' dx = \int u'(x)v(x) dx + \int u(x)v'(x) dx$$

If we use differentials and let  $du = u'(x)dx$  and  $dv = v'(x)dx$ ,<sup>2</sup> we can write this in a format that's more recognizable, namely

$$u(x)v(x) = \int v(x)du + \int u(x)dv \implies \boxed{\int u dv = uv - \int v du}.$$

So now we can take an integral that's hard to compute, and flip around what we're integrating to make it potentially easier to simplify. It's easiest to understand through an example.

**Example 2.3.** Compute  $\int xe^{3x} dx$ .

*Solution.* If I'm applying integration by parts, I need to find one expression to take the derivative of (which will be  $u$ ) and another to take the integral of (which will be  $dv$ ) to simplify. Here, I'll let  $u = x$  and  $dv = e^{3x} dx$ , so that  $du = dx$  and  $v = \frac{1}{3}e^{3x}$ . Then applying integration by parts gives

$$\int xe^{3x} = \int u dv = uv - \int v du = \frac{1}{3}xe^{3x} - \int \frac{1}{3}e^{3x} dx = \frac{1}{3}xe^{3x} - \frac{1}{9}e^{3x} + C.$$

□

It's not always obvious what I should be setting  $u$  and  $dv$  to typically. A good rule of thumb is to set  $u$  to the first term you see in this list:

- Logarithm
- Inverse trig function
- Algebraic function (e.x. polynomials)
- Trig function
- Exponential function

Here's another example.

**Example 2.4.** Compute  $\int \ln x dx$ .

*Solution.* You might think that we can't apply integration by parts here since we only have one term. But in fact, let  $u = \ln x$  and  $dv = dx$  so that  $du = \frac{1}{x}dx$  and  $v = x$ . Then

$$\int \ln x dx = \int u dv = uv - \int v du = x \ln x - \int x \cdot \frac{1}{x} dx = x \ln x - x + C.$$

□

You should also search for the tabular method if you want to speed things up when doing multiple integrations by parts (Khan academy is a good place to watch).

**Exercise 2.5.** Compute  $\int x^2 e^{3x} dx$ .

<sup>2</sup>A differential is change in the function based on the dependent variable (how far up we move based on horizontal movement and the slope).

### 3 Series

A quick informal definition so everyone's on the same page here.

**Definition 7 (Series).** Given some sequence of terms  $(a_1, a_2, \dots, a_n)$ , the **series** corresponding to this sequence is the sum

$$a_1 + a_2 + \dots + a_n.$$

If the sequence is infinite, then the series will contain infinitely many summands.

Series come up often in discrete probability when we ask what the probability of some range of events is. So it's useful to know how to sum the two main types of series and some tricks we can do with them.

There's two main series to know: the geometric series, and the Taylor series.

In a geometric series, we have an initial term  $a$ , and every term after that is some common ratio,  $r$ , times the previous terms. So a typical finite geometric series looks something like

$$S = a + ar + ar^2 + \dots + ar^k$$

where  $S$  is the final value of the series. To calculate this, notice that

$$rS = ar + ar^2 + \dots + ar^k + ar^{k+1},$$

so if we subtract one from the other, all the middle terms cancel out, leaving us with

$$S - rS = a - ar^{k+1} \implies \boxed{S = \frac{a(1 - r^{k+1})}{1 - r}}.$$

If we take  $k \rightarrow \infty$ , then we get an infinite geometric series, for which

$$\sum_{i=0}^{\infty} ar^i = \frac{a}{1 - r}$$

but only when  $|r| < 1$  (otherwise the series won't converge). Here's an example that utilizes geometric series for a very fitting distribution.

**Example 3.1.** Show that the sum of the probabilities of a random variable  $X \sim \text{Geo}(p)$  is 1.

*Solution.* Recall that  $\mathbb{P}(X = k) = (1 - p)^{k-1}p$ . Then

$$\sum_{k=1}^{\infty} \mathbb{P}(X = k) = \sum_{k=1}^{\infty} (1 - p)^{k-1}p.$$

But this is a *geometric* series with initial term  $p$  and ratio  $1 - p$ , so the sum is  $\frac{p}{1 - (1 - p)} = 1$ .  $\square$

The idea of a Taylor series is to approximate a function  $f(x)$  by its derivatives. This looks like

$$f(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \frac{1}{3!}f'''(0)x^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!}f^{(k)}(0)x^k$$

where  $f^{(k)}$  is the  $k$ th derivative of  $f$ . The only one you'll need for this class is  $e^x$ :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

One last interesting note is that geometric series are *also* Taylor series. If we let  $f(r) = \frac{a}{1 - r}$ , then the derivatives are

$$f(r) = \frac{a}{1-r}, \quad f'(r) = \frac{a}{(1-r)^2}, \quad f''(r) = 2! \cdot \frac{a}{(1-r)^3}, \quad \dots, \quad f^{(k)}(r) = k! \cdot \frac{a}{(1-r)^{k+1}}$$

so

$$f(0) = a, \quad f'(0) = 1! \cdot a, \quad f''(0) = 2! \cdot a, \quad \dots, \quad f^{(k)}(0) = k! \cdot a.$$

Plugging this into the Taylor series expansion gives

$$f(r) = \frac{a}{1-r} = a + ar + ar^2 + \dots$$

which is the geometric distribution as expected.

**Exercise 3.2.** Show that the sum of the probabilities of a Poisson random variable  $X \sim \text{Poisson}(\lambda)$  is 1.

#### 4 Intermission: Calculus and Series in Probability

Now that we've covered some calculus, let's look at how the different results we've seen so far are used in probability.

Here's two problems that make use of the Taylor series for  $e^x$ .

**Example 4.1** (Dis 12 #3). Let  $X \sim \text{Geo}(p)$  and  $Y \sim \text{Poisson}(\lambda)$  be independent random variables. Compute  $\mathbb{P}(X > Y)$ .

*Solution.* Let's condition on  $Y$  so that we can use  $\mathbb{P}(X > k) = (1-p)^k$ . This gives

$$\begin{aligned} \mathbb{P}(X > Y) &= \sum_{y=0}^{\infty} \mathbb{P}(Y = y) \mathbb{P}(X > y | Y = y) \\ &= \sum_{y=0}^{\infty} \frac{e^{-\lambda} \lambda^y}{y!} \cdot (1-p)^y \\ &= e^{-\lambda} \sum_{y=0}^{\infty} \frac{(\lambda(1-p))^y}{y!} \end{aligned}$$

But this sum is the Taylor series for  $e^{\lambda(1-p)}$ , so the probability is just  $e^{-\lambda} e^{\lambda(1-p)} = e^{-\lambda p}$ .  $\square$

This somewhat surprising result was derived in Lecture 21.

**Example 4.2.** Let  $X_n$  be a random variable counting the number of fixed points of a permutation on  $n$  variables. What is the distribution of  $X_n$  as  $n \rightarrow \infty$ ?

*Solution.* We wish to compute  $\mathbb{P}(X_n = j)$ , so we count the number of permutations with  $j$  fixed points. We need to choose  $j$  points to be fixed and the rest to not have any fixed points, which means they form a derangement, so there are  $\binom{n}{j} D_{n-j}$  such permutations. Hence

$$\mathbb{P}(X_n = j) = \frac{\binom{n}{j} D_{n-j}}{n!}.$$

Recalling that  $D_j = j! \sum_{k=0}^j \frac{(-1)^k}{k!}$ , we find

$$\begin{aligned} \mathbb{P}(X_n = j) &= \frac{1}{n!} \left[ \frac{n!}{j!(n-j)!} (n-j)! \sum_{k=0}^{n-j} \frac{(-1)^k}{k!} \right] \\ &= \frac{1}{j!} \sum_{k=0}^{n-j} \frac{(-1)^k}{k!}. \end{aligned}$$

But when  $n \rightarrow \infty$ , this is the Taylor Series for  $e^{-1}$ , so

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = j) = \frac{e^{-1} 1^j}{j!}$$

which is the pdf of a Poisson(1) distribution.  $\square$

This result also agrees with the fact that  $\mathbb{E}[X_n] = 1$ , which is true by linearity of expectation.

Here's an idea similar to Discussion 13, Problem 2 but in a different context. Under the hood, you should realize that we can do this cdf manipulation thanks to the Fundamental Theorem of Calculus.

**Example 4.3.** Let  $X, Y$  be a random variables where  $Y = X^2$ . Express the pdf of  $Y$ ,  $f_Y$ , in terms of the pdf of  $X$ ,  $f_X$ .

*Solution.* The first step should usually be to compute the cdf of  $Y$  in terms of  $X$ 's cdf. This gives

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(X^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x}).$$

From here, we differentiate with respect to  $X$  (being careful to use the chain rule) and get

$$f_Y(x) = \frac{dF_Y}{dx} = \frac{1}{2\sqrt{x}} F'_X(\sqrt{x}) - \left(-\frac{1}{2\sqrt{x}}\right) F'_X(-\sqrt{x}) = \frac{1}{2\sqrt{x}} (f_X(\sqrt{x}) + f_X(-\sqrt{x}))$$

$\square$

## 5 Double Integrals

Now we're finally talking about multivariable calculus. Before we start, we should ask the question: what even is a double integral?

We should first think about what a single integral represented. In summary, we defined it as

Finding an **area** under  $f(x)$  over a **one-dim. interval**  $[a, b]$  on the  **$x$ -axis**:  $A = \int_a^b f(x) dx$ .

The double integral will be the natural generalization of this. Instead of integrating an area over an interval, we will be integrating a volume over a region in 2D space.

Finding a **volume** under  $f(x, y)$  over a **two-dim. region**  $\mathcal{R}$  on the  **$xy$ -plane**:  $V = \iint_{\mathcal{R}} f(x, y) dA$ .

The hardest part of solving double integrals is often figuring out the region of integration and setting up the limits. After that, it's just single-variable calculus. Let's look at an example.

**Example 5.1.** Integrate  $f(x, y) = x^2 + y^2$  over the  $4 \times 4$  square centered at  $(0, 0)$ .

*Solution.* Before we begin, take a look at this visualization of our integral.

As you can tell, we're looking for the volume under the curve. Our region of integration is the region bounded by the lines  $x = -2, x = 2, y = -2, y = 2$ .

There's two ways in which we can integrate over this region: we can either integrate with respect to  $x$  first, or with respect to  $y$  first. Either way is the same in this case since everything is symmetric, so I'll set it up as the former. This means our integral is

$$\int_{-2}^2 \int_{-2}^2 x^2 + y^2 dx dy.$$

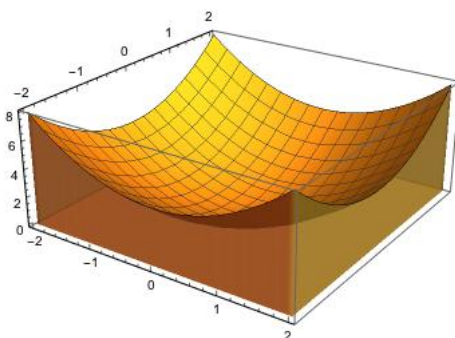


Figure 2: A visualization of the integral of  $f(x, y) = x^2 + y^2$  over the square  $[-2, 2] \times [-2, 2]$ .

In order to integrate this, we treat the inside integral like a single integral and treat everything that's not  $x$  like a constant. This gives

$$\int_{-2}^2 \int_{-2}^2 x^2 + y^2 \, dx \, dy = \int_{-2}^2 \left( \frac{x^3}{3} + xy^2 \right) \Big|_{x=-2}^2 \, dy = \int_{-2}^2 \frac{16}{3} + 4y^2 \, dy$$

Now that we've gotten rid of our  $x$  variable (as expected), all that's left is to finish the single integral, which gives us

$$\int_{-2}^2 \frac{16}{3} + 4y^2 \, dy = \left( \frac{16}{3}y + \frac{4}{3}y^3 \right) \Big|_{-2}^2 = \frac{128}{3}.$$

□

Here's one where the region of integration is trickier.

**Example 5.2.** Integrate  $ye^{-x}$  over the *triangle* enclosed by  $y = x$ ,  $y = 3$ , and  $x = 0$ .

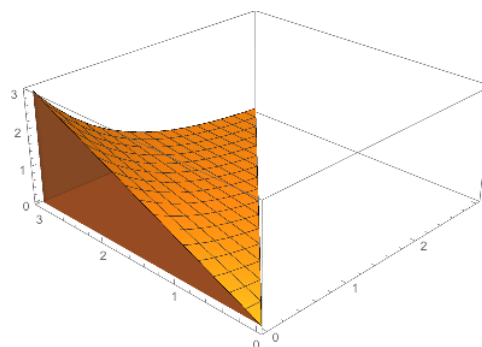
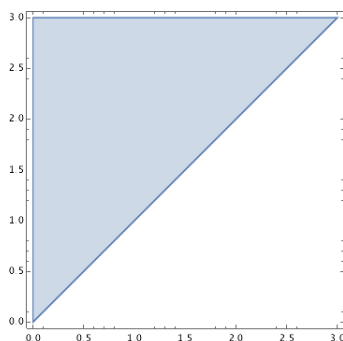


Figure 3: The boundary and the integral of  $f(x, y) = ye^{-x}$  over the region in Example 5.2.

*Solution.* We already know the region is a triangle, but we need to figure out how to integrate over it. If we first integrate over  $y$  then  $x$ , then the limits of our outer integral should go from 0 to 3. In this case, we then have the variable  $x$  to work with to set up our limits for  $y$ . Using the equations, we find that we should integrate from  $x$  to 3. Setting up our integral and simplifying gives

$$\begin{aligned} \int_0^3 \int_x^3 ye^{-x} \, dy \, dx &= \int_0^3 \frac{1}{2} (y^2 e^{-x}) \Big|_{y=x}^3 \, dx \\ &= \frac{1}{2} \int_0^3 9e^{-x} - x^2 e^{-x} \, dx \end{aligned}$$



which would require the use of *two* integrations by parts to evaluate; yuck! Let's see what happens if we set up our integral the other way. The outer limits are still from 0 to 3, but the inner ones are now from 0 to  $y$  (convince yourself of this), so the integral is

$$\begin{aligned} \int_0^3 \int_0^y ye^{-x} dx dy &= \int_0^3 (-ye^{-x}) \Big|_{x=0}^y dy = \int_0^3 -ye^{-y} + y dy \\ &= \int_0^3 y dy - \left( -ye^{-y} \Big|_0^3 + \int_0^3 e^{-y} dy \right) \\ &= \frac{9}{2} - (-3e^{-3} + 1 - e^{-3}) = \boxed{\frac{7}{2} + \frac{4}{e^3}}. \end{aligned}$$

□

We don't always have to integrate using  $x$  and  $y$  though. Sometimes it's more natural for us to use *polar* coordinates where we label every point as  $(r, \theta)$ ,  $r$  being its distance from the origin and  $\theta$  being its angle with the  $x$ -axis.

**Example 5.3.** Integrate  $f(x, y) = x^2 + y^2$  over the circle of radius 2 centered at  $(0, 0)$ .

*Solution.* The first step is to translate our function in terms of  $r$  and  $\theta$ . We know that  $x^2 + y^2 = r^2$ , so  $f(r, \theta) = r^2$ . Our region is  $0 \leq r \leq 2$ , so our integral is

$$\int_0^{2\pi} \int_0^2 r^2 dr d\theta = \int_0^{2\pi} \frac{r^3}{3} \Big|_{r=0}^2 d\theta = \int_0^{2\pi} \frac{8}{3} d\theta = \frac{16}{3}\pi.$$

□

We did something similar with Buffon's needle, where we integrated with respect to  $y$  and  $\theta$ . All that matters is that we find the appropriate regions and limits to integrate over.

**Exercise 5.4.** Integrate  $e^{x+y}$  over the rectangular region determined by the lines  $y = 0$ ,  $x = 0$ ,  $y = 3$ ,  $x = 2$ .

**Exercise 5.5.** Integrate  $-\frac{xe^x}{y^2}$  over the region defined by the lines  $x = 5$ ,  $y = -x$ ,  $y = x$ .

**Exercise 5.6.** Suppose  $X \sim \text{Exp}(\lambda_1)$ ,  $Y \sim \text{Exp}(\lambda_2)$  are independent exponential distributions so that  $f(x, y) = \lambda_1 \lambda_2 e^{-(\lambda_1 + \lambda_2)x}$  for  $x, y \geq 0$  is the density of their joint distribution. Show that the integral of this density is 1.

## References

- [1] James Stewart, *Calculus 7 ed.* Cengage Learning, 2012.