

Math Review

This note is meant to remind you of relevant mathematical tools and methods that we will be actively using throughout the course. The presentation is formal enough for our purposes but omits some other important details. The topics are not connected with one another, so you may only read sections that are relevant to you, in any order.

1 Crash course on probability theory

The primary goal of this section is to remind you how to work with continuously distributed random variables. It does so by building analogies to relevant rules for discrete random variables. To illustrate the rules, we will use two following random variables:

- $x \in \{1, 2, 3\}$ with *probability distribution function* $p(x) = x/6$, meaning that $x = 1$ with probability $1/6$; $x = 2$ w.p. $2/6$ and $x = 3$ w.p. $3/6$.
- $y \in [0, 1]$ with *probability density function* $f(y) = 2y$.

Probability density function is the analog of probability distribution function for continuously distributed random variables. We need it because in most continuous distributions, the *probability* of any given value occurring is exactly zero (unless this distribution has *atoms* – values that occur with strictly positive probability).

One way to interpret pdf $f(y)$ (“pdf” will always refer to density, not distribution function) is saying it represents *relative likelihoods*. For example, $y = 0.5$ with density $f(0.5) = 1$ is twice more likely to occur than $y = 0.25$ with density $f(0.25) = 0.5$, but twice less likely than $y = 1$. Finally, just like probabilities of all values x can take must sum up to one: $\sum_{x \in \{1, 2, 3\}} p(x) = 1$, the density over the range of y must integrate to one: $\int_0^1 f(y) dy = 1$.

1.1 Mathematical expectations

How can we find the mathematical expectation – average value – of a random variable? For discrete random variables, we weigh each value with the probability of that value occurring:

$$\begin{aligned}\mathbb{E}[x] &= \sum_{x \in \{1, 2, 3\}} x \cdot p(x) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{3}{6} \\ &= \frac{14}{6} \approx 2.33\end{aligned}$$

To find an expectation of a continuous random variable, we do the exact same thing, except we use integral instead of the sum (since an integral is pretty much a sum of a continuum of infinitesimally small terms):

$$\begin{aligned}\mathbb{E}[y] &= \int_0^1 [y \cdot f(y)] dy \\ &= \int_0^1 [2y^2] dy \\ &= \left. \frac{2}{3} y^3 \right|_{y=0}^1 = \frac{2}{3} - 0 = \frac{2}{3}\end{aligned}$$

1.2 Events

An *event* is a collection of realizations of a random variable. For example, if air temperature t (in degrees C) is the only unknown (random) variable, then the event “it will be warm tomorrow” means something like “ $t \geq 15$ tomorrow”.

If we know the distribution of the random variable, we can calculate probabilities of events. For example, probability of event $E_x \equiv \{x \geq 2\}$ for random variable x defined above is given by

$$\mathbb{P}(E_x) = \sum_{x \in E_x} p(x) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$$

For continuous random variables, we again substitute the sum with the integral. For example, probability of event $E_y \equiv \{y \geq 0.5\}$ for y defined above is

$$\mathbb{P}(E_y) = \int_{y \in E_y} f(y) dy = \int_{y=0.5}^1 f(y) dy = \int_{y=0.5}^1 2y dy = y^2 \Big|_{y=0.5}^1 = 1 - 0.25 = 0.75$$

1.3 Bayes' rule and Conditional expectations

We often need to calculate the conditional expectation of a random variable given some event. E.g., we do not know the exact realization of x , but we know that $x \geq 2$. Then we can find the expectation of x conditional on $x \geq 2$ by weighing all possible values of $x \geq 2$ by their respective conditional probabilities (you may recognize this expression as the Bayes' rule)¹

$$\mathbb{P}(x|x \geq 2) = \frac{p(x)}{\mathbb{P}\{x \geq 2\}},$$

thus obtaining

$$\mathbb{E}[x|x \geq 2] = \sum_{x \geq 2} \frac{p(x)}{\mathbb{P}\{x \geq 2\}} \cdot x = \frac{\frac{2}{6}}{\frac{2}{6} + \frac{3}{6}} \cdot 2 + \frac{\frac{3}{6}}{\frac{2}{6} + \frac{3}{6}} \cdot 3.$$

Note that the denominator does not depend on the realization of x , so we can take it out of the sum. This means the conditional expectation is equivalent to weighing all possible values of $x \geq 2$ by their respective probabilities, and then normalizing the resulting sum by the probability of the whole event $x \geq 2$:

$$\begin{aligned} \mathbb{E}[x|x \geq 2] &= \frac{\sum_{x \geq 2} x \cdot p(x)}{\mathbb{P}\{x \geq 2\}} \\ &= \frac{\sum_{x \geq 2} x \cdot p(x)}{\sum_{x \geq 2} p(x)} = \frac{\frac{1}{3} \cdot 2 + \frac{1}{2} \cdot 3}{\frac{1}{3} + \frac{1}{2}} = \frac{13}{5} = 2.6. \end{aligned}$$

For continuous random variables, we again simply substitute the sums for integrals. Taking the same event $E_y = \{y \geq 0.5\}$ as an example, we have

$$\mathbb{E}[y|E_y] = \frac{\int_{y \in E_y} [y \cdot f(y)] dy}{\int_{y \in E_y} f(y) dy} = \frac{\int_{0.5}^1 2y^2 dy}{\int_{0.5}^1 2y dy} = \frac{\frac{2}{3} y^3 \Big|_{y=0.5}^1}{y^2 \Big|_{y=0.5}^1} = \frac{\frac{2}{3} \cdot (1 - \frac{1}{8})}{1 - \frac{1}{4}} = \frac{7}{9}.$$

¹Here I use $p(x)$ to denote the probability of a given realization x and \mathbb{P} to denote the probability of a given event, as generated by p . This distinction is not meaningful, and you can use either p , or \mathbb{P} everywhere instead. My goal here was to show that normally, once the probabilities of different realizations, $p(\cdot)$, are given as a primitive, you can calculate the probabilities of all other respective events, $\mathbb{P}(\cdot)$. Mathematically, however, the two are equivalent, since p is already implied to be a probability measure on the whole Borel sigma-algebra on the set of realizations of x .

1.4 Expectations of functions

What if you need to calculate an expectation of some function $u(y)$? You do the very same thing: you weigh various values of $u(y)$ by the probabilities of the respective realizations of y . The general formula is

$$\mathbb{E}[u(y)|E_y] = \frac{\int_{y \in E_y} [u(y) \cdot f(y)] dy}{\int_{y \in E_y} f(y) dy}.$$

(Can you see how it simplifies in case of an unconditional expectation? Can you see how it will look like for a discrete distribution?)

To summarize all lessons of this sections, consider the following slightly elaborate example. Suppose that you can pay \$0.5 to participate in a lottery which pays $w = y^2$ if $y \geq 1/2$ and pays $w = 0$ otherwise (y here has the same distribution as above). How do you compute the average payoff of such a lottery?

$$\begin{aligned} \mathbb{E}[w] &= P\left(y \geq \frac{1}{2}\right) \cdot \mathbb{E}\left[y^2 \mid y \geq \frac{1}{2}\right] + P\left(y < \frac{1}{2}\right) \cdot \mathbb{E}\left[0 \mid y < \frac{1}{2}\right] \\ &= \int_{\frac{1}{2}}^1 f(y) dy \cdot \frac{\int_{\frac{1}{2}}^1 [y^2 \cdot f(y)] dy}{\int_{\frac{1}{2}}^1 f(y) dy} + \int_0^{\frac{1}{2}} f(y) dy \cdot \frac{\int_0^{\frac{1}{2}} [0 \cdot f(y)] dy}{\int_0^{\frac{1}{2}} f(y) dy} \\ &= \int_{\frac{1}{2}}^1 [y^2 \cdot f(y)] dy + 0 = \int_{\frac{1}{2}}^1 2y^3 dy \\ &= \frac{1}{2} y^4 \Big|_{y=\frac{1}{2}}^1 = \frac{1}{2} \left(1 - \frac{1}{16}\right) = \frac{15}{32} \end{aligned}$$

So if you are risk-neutral and, thus, evaluate lotteries by their expected payoff, then lottery w is not worth taking, because its average payoff is $\frac{15}{32}$, which is less than the price $\frac{1}{2} = \frac{16}{32}$ required to buy it.

2 Leibniz rule

The Leibniz rule tells you how to take the derivative of a function that is an integral. We will only use one very special case of it. Therefore, I show you this special case and give the full rule only for sake of completeness below.

What we want to look at are functions like

$$f(x) \equiv \int_0^x g(z) dz.$$

What is the derivative of f (w.r.t. x)? The Leibniz rule says that

$$f'(x) = g(x)$$

The way you should think about this is that the integral of a function is the area below the graph of this function. So, draw some function g now. Seriously, do it! The integral $\int_0^x g(z) dz$ is the area between the axis and the function $g(z)$ from $z = 0$ to some $z = x$ (take some $x > 0$ and shade this area in the graph you just drew). The derivative of $f(x)$ with respect to x answers to the following question: “How does the size of the shaded area change if you make x a bit bigger?” From the graph, it should be clear that if you increase x by dx , the area gets approximately $g(x) \cdot dx$ bigger. This means that $f'(x) = g(x)$ (increasing x by one unit increases f by $g(x)$).

Similar intuition implies that the derivative of

$$h(x) \equiv \int_x^1 g(z)dz$$

is $h'(x) = -g(x)$.

The full Leibniz rule (that we will not really use) says that the derivative of

$$f(x) \equiv \int_{a(x)}^{b(x)} g(x, z)dz$$

is $f'(x) = -a'(x)g(x, a(x)) + b'(x)g(x, b(x)) + \int_{a(x)}^{b(x)} g'_x(x, z)dz$.

3 Fundamental Theorem of Calculus

This theorem states what you almost surely know, at least informally: “integration and taking derivatives are opposite operations”. The theorem consists of two parts, relating to the two possible directions of the “round trip”: part 1 is about the derivative of the integral, while part 2 is about the integral of a derivative.²

Theorem 1 (Fundamental Theorem of Calculus).

1. Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded and continuous almost everywhere. For any $x \in [a, b]$, define $F(x)$ as

$$F(x) = F(a) + \int_a^x f(y)dy. \quad (1)$$

Then F is continuous on $[a, b]$, and its derivative $F'(x)$ exists and equals $f(x)$ at all continuity points of f .

2. Let $F : [a, b] \rightarrow \mathbb{R}$ be continuously differentiable. For any $x \in [a, b]$, define $f(x)$ as $f(x) = F'(x)$. Then (1) holds.

4 Integration by parts

Integration is difficult, mkay?³ Integration by parts is one trick that sometimes helps make it a little easier. For example, if you are asked to compute

$$\int_4^5 xe^x dx,$$

finding the “antiderivative” of xe^x is not a trivial task. But you can use integration by parts to go around it. The rule itself looks as

$$\int_a^b u(x)v'(x)dx = [u(x)v(x)]|_a^b - \int_a^b u'(x)v(x)dx, \quad (2)$$

where u and v are some functions of x .

Let us try to apply it to our problem. Set $u(x) = x$ and $v'(x) = e^x$, then $u'(x) = 1$ and $v(x) = e^x$ (since

²See also: <https://www.smbc-comics.com/comic/fundamental-2>.

³<https://xkcd.com/2117/>

$\frac{d}{dx}e^x = e^x$). Plugging all of these in, we get that

$$\begin{aligned}\int_4^5 xe^x dx &= [xe^x]_4^5 - \int_4^5 e^x dx \\ &= 5e^5 - 4e^4 - e^x \Big|_4^5 \\ &= 5e^5 - 4e^4 - (e^5 - e^4) = 4e^5 - 3e^4\end{aligned}$$

How did we know how to split xe^x into $u(x)$ and $v'(x)$? Trial and error really. (What happens if you try to do it the other way?)

For another example, see if you can use integration by parts to calculate $\int_a^b x^2 \log(x) dx$.

4.1 Double integral example

Integration by parts can sometimes transform double integrals to simple integrals. For example, consider the following integral:

$$\int_0^1 \left(2x \int_a^x f(y) dy \right) dx,$$

where $f(y)$ is some arbitrary function. Here we can set $u(x) = \int_a^x f(y) dy$ and $v'(x) = 2x$, which would yield $u'(x) = f(x)$ (remember Leibniz rule?) and $v(x) = x^2$. Plugging all of these into the integration by parts expression yields

$$\begin{aligned}\int_0^1 \left(2x \int_a^x f(y) dy \right) dx &= \left[x^2 \int_a^x f(y) dy \right]_{x=0}^1 - \int_0^1 f(x) x^2 dx \\ &= \int_a^1 f(y) dy - \int_0^1 f(x) x^2 dx,\end{aligned}$$

which looks much nicer! Note also that both x and y in the final expression are just integration variables which we can relabel freely. For example, we can relabel y to x so that the final answer is

$$\int_a^1 f(x) dx - \int_0^1 f(x) x^2 dx.$$

4.2 Why does integration by parts work?

If integration by parts looks arbitrary and mysterious, it is only because it disguised itself well. Truth is you know it well, since it is nothing more than product differentiation rule:

$$(u(x)v(x))' = u'(x)v(x) + u(x)v'(x)$$

To see the connection, integrate both sides of this expression from a to b :

$$\int_a^b (u(x)v(x))' dx = \int_a^b u'(x)v(x) dx + \int_a^b u(x)v'(x) dx.$$

Here we already split the right-hand side integral into two. Notice that on the left-hand side we have an integral of the derivative – the two operations which “cancel each other out” (see the fundamental theorem

of calculus). Meaning that what we have on the LHS is exactly

$$\int_a^b (u(x)v(x))' dx = [u(x)v(x)]_a^b.$$

So by rearranging the terms a bit we get precisely the integration by parts rule (2):

$$\begin{aligned} [u(x)v(x)]_a^b &= \int_a^b u'(x)v(x)dx + \int_a^b u(x)v'(x)dx \\ \Leftrightarrow \int_a^b u(x)v'(x)dx &= [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)dx. \end{aligned}$$