Chan and Khandani
Axioms of Probability

An Introduction

This chapter's title sounds like the album name of a shitty math rock band.

That being said - last chapter, we learned how to count. But we never really explored the probability of, well, anything. So why did we even learn how to count?

What we actually did was learn how to construct the sample space - the set of all possibilities - of a given experiment. Knowing the sample space paves the way for us to actually calculate probability.

Definitions of Probabalistic Concepts

Unfortunately, with basically every course, we've gotta slog through some terminology. What we can do, is make it quick.

Sample space - the set of all possibilities

Elementary event - a single possibility within the sample space (a set with one element)

Event - zero or more possibilities within the sample space

Random experiment - outcome is subject to chance

Relative frequency - what we defined earlier as "ratio"
- the limit of $\frac{\#\text{ of event occurrences}}{\text{tries}}$

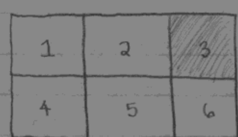
Statistical regularity - a property of random experiments that says an experiment's events have a defined relative frequency

Phew, glad that's over with.

Unions, Intersections, and Complements of Events

We'll start with complements because those are the simplest.

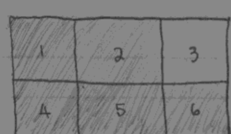
ex.



Let's take a dice roll: there are 6 distinct possibilities of the roll.

Let's define the event E_1 as the event "we rolled a 3". I've represented our sample space as the box, and split it into each outcome, highlighting the set E_1.

E_1^c, the complement of E_1, is simply every OTHER event that isn't contained within E_1.

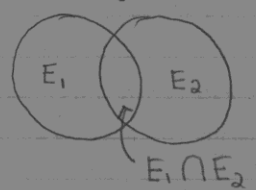


As such,

$E_1^c = \{1, 2, 4, 5, 6\}$, given that $E_1 = \{3\}$

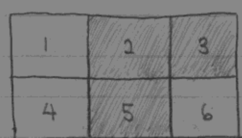
So what is an intersection? Simply put, it is the set of shared elementary events from two events; the set of all elementary events that would satisfy all events involved.

Denoted as an upside-down u ($\cap$)*, an intersection is best graphically depicted as the well-known Venn Diagram.

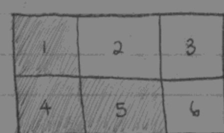


E_1 and E_2 are both events of the same sample space, and they share some elementary events. Those shared elementary events ARE the set of $E_1 \cap E_2$.

ex. We'll use the same dice roll example here.



$E_1 = \{2, 3, 5\}$



$E_2 = \{1, 4, 5\}$

We've got 2 events here, E_1, and E_2. The only shared elementary event is 5,

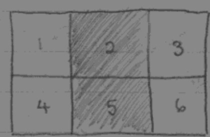
$$\therefore \; E_1 \cap E_2 = \{5\}$$

Meaning the only way for us to have E_1 AND E_2 occur with a single throw is to roll a 5.

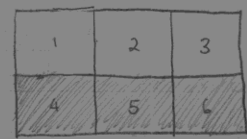
Finally, we have unions, denoted as $\cup$ (what a surprise). A union is the combination of multiple events. Where intersections dealt with trying to satisfy all events, a union produces a set that represents all outcomes that will satisfy <u>at least one</u> of the events in the union.

*or sometimes simply $E_1 E_2 E_3 \ldots$

ex.



$E_1 = \{2, 5\}$



$E_2 = \{4, 5, 6\}$

To take the union of two events, we construct a new set that consists of every unique elementary event found in each event.

Though E_1 and E_2 share the elementary event 5, it only appears once in the union.

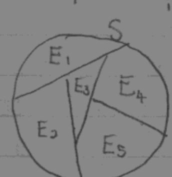
$$E_1 \cup E_2 = \{2, 4, 5, 6\}$$

So in order to satisfy at least one event with a single roll, we can roll any of 2, 4, 5, or 6.

<u>Partitioning a Sample Space</u>

Finally, we can say a number of events partition the sample space if they have no events in common, and adding them all up creates the sample space itself.

Formally; $\bigcup_{i=1}^{n} E_i = S$ and $E_i \cap E_j = \emptyset$, $i \neq j$



Partitioning a sample space simply means being able to chop it up into a bunch of pieces.

If you could accomplish this physically, you have successfully partitioned a sample space

Properties of Unions and Intersections

Both unions and intersections exhibit commutativity.

$$E \cup F = F \cup E \qquad E \cap F = F \cap E$$

Like addition, the order doesn't matter in the calculation. The resulting sets of both operations depend only on the given events.

ex. $$E = \{3, 4, 5\} \qquad F = \{5, 6\}$$

$$E \cup F = \{3, 4, 5, 6\} \qquad E \cap F = \{5\}$$
$$F \cup E = \{5, 6, 3, 4\} \qquad F \cap E = \{5\}$$

In the same vein, both operations exhibit associativity.

$$(E \cup F) \cup G = E \cup (F \cup G) \qquad (E \cap F) \cap G = E \cap (F \cap G)$$

As associativity is also an order-related property, it is easy to see that unions / intersections are associative as well.

ex. $E = \{1, 6, 7\} \qquad F = \{6, 7, 9\} \qquad G = \{1, 10\}$

$$\begin{aligned}(E \cup F) \cup G &= \{1, 6, 7, 9\} \cup \{1, 10\} \\ &= \{1, 6, 7, 9, 10\}\end{aligned}$$

$$\begin{aligned}E \cup (F \cup G) &= \{1, 6, 7\} \cup \{1, 6, 7, 9, 10\} \\ &= \{1, 6, 7, 9, 10\}\end{aligned}$$

$$\begin{aligned}(E \cap F) \cap G &= \{6, 7\} \cap (1, 10) \\ &= \varnothing\end{aligned}$$

$E \cap (F \cap G) = \{1, 6, 7\} \cap \emptyset$
$= \emptyset$

Finally, like multiplication, unions and intersections exhibit distributivity.

$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ $\quad$ $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$

ex. $E = \{5, 7, 10, 11\}$ $\quad$ $F = \{5, 6, 10, 11\}$
$G = \{1, 5, 8\}$

$(E \cup F) \cap G = \{5, 6, 7, 10, 11\} \cap \{1, 5, 8\}$
$= \{5\}$

$(E \cap G) \cup (F \cap G) = \{5\} \cup \{5\}$
$= \{5\}$

$(E \cap F) \cup G = \{5, 10, 11\} \cup \{1, 5, 8\}$
$= \{1, 5, 8, 10, 11\}$

$(E \cup G) \cap (F \cup G) = \{1, 5, 7, 8, 10, 11\} \cap \{1, 5, 6, 8, 10, 11\}$
$= \{1, 5, 8, 10, 11\}$

De Morgan's Law

Does this sound familiar? It should; we learnt it in 124 with boolean operators:

$(A \text{ and } B)^c = A^c \text{ or } B^c$

When we think about unions and intersections, aren't they essentially the same things?

Unions create sets that indicate how to satisfy ANY event involved, ie. E_1 OR E_2 OR $E_3 \ldots$

Intersections create sets that indicate how to satisfy ALL events involved, ie. E_1 AND E_2 AND $E_3 \ldots$

Now, doesn't it make sense that we can apply De Morgan's Law to our and/or operations?

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c \qquad \left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

ex. Given that $S = \{1, 2, 3, 4, 5\}$
$E = \{3, 4, 5\}$
$F = \{1, 5\}$

What is $(E \cup F)^c$?

$$\begin{aligned}(E \cup F)^c &= E^c \cap F^c \\ &= \{1, 2\} \cap \{2, 3, 4\} \\ &= \{2\}\end{aligned}$$

On a somewhat unrelated note, any given E and E^c partition a sample space. As such, we can write any event F of the same sample space as

$$F = FE^c \cup FE, \text{ where } (FE^c) \cap (FE) = \emptyset$$

which may come in handy later on for all the weird boolean algebra / proof questions we always seem to get.

Simple Propositions

Let's walk through some examples.

ex. $P(E^c) = 1 - P(E)$

Translation: the probability of an event's complement is
1 − the probability of an event itself

This one is fairly intuitive - since E and E^c partition the sample space, if E takes up some portion of the space, the probability of E not occurring must take up all the REMAINING space within the sample space.

ex. $E \subset F \rightarrow P(E) \leq P(F)$

Translation: If the event E is a subset of F, the chance of it occurring is less than or equal to the chance of F occurring.

Since subsets can only contain the same elementary events their superset contains, the chance of the subset's events occurring are equal (if the subset and superset are the same) or less than the superset's (if the subset only contains some elements, not all, of the superset).

ex. $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Translation: The probability of either E or F occurring is the sum of the probabilities of each event occurring individually, minus the probability of both occurring.

The reason why we subtract the probability of the intersection is because we don't want to over-count the shared probabilities.

$E = \{1, 2, 3, 4\}$ If each elementary event has
$F = \{2, 3, 4, 5\}$ a probability of 0.2, not

subtracting the intersection would give us an overall probability of 1.6, which obviously isn't possible. As such, by subtracting the $P(\{2,3,4\}) = 0.6$, we get the correct probability of the union - 1.

ex. $$P(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \ldots$$

$$+ (-1)^{r+1} \sum_{i_1 < i_2 < \ldots i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) + \ldots$$

$$+ (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$$

where this sum is over $\binom{n}{r}$ terms.

This is a multi-event extension of the previous probability of a union, and in all honesty, is incredibly confusing. It might be useful if you could even read the damn equation, but I sure as hell can't, so let me show you the method I've been using.

Calculating the Probability of the Union of Multiple Events

This method breaks down to one very simple thing: finding the probability of intersections of events. Let's start.

① ADD the probabilities of all events individually
② Find every combination of 2 events, take their intersections, and SUBTRACT those probabilities to the running total
③ Find every combination of 3 events, take their intersections, and ADD those probabilities to the running total
④ Continue to add / subtract probabilities
- → subtract if the number of events in the intersection is even
- → add if the number of events is odd

That's it! This is literally the same thing as the equation from the page prior. Let's go through an example.

ex. Let's assume there's some experiment that randomly generates a number from 1 to 10 - 0.1 probability each.

$E_1 = \{2, 3, 4, 5, 6\}$ $\quad$ $E_2 = \{1, 3, 5, 7\}$

$E_3 = \{3, 6, 9\}$ $\quad$ $E_4 = \{8, 10\}$

$E_1 \cup E_2 \cup E_3 \cup E_4 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, in other words, produces the entire sample space

Obviously, that must mean the overall probability is 1, so we can confirm our final answer with this.

① $P(E_1) = 0.5$ $\quad$ $P(E_2) = 0.4$

$P(E_3) = 0.3$ $\quad$ $P(E_4) = 0.2$

Total $= 1.4$

② $P(E_1 \cap E_2) = P(\{5\}) = 0.1$ $\quad$ $P(E_1 \cap E_3) = P(\{3, 6\}) = 0.2$

$P(E_1 \cap E_4) = P(\emptyset) = 0$ $\quad$ $P(E_2 \cap E_3) = P(\{3\}) = 0.1$

$P(E_2 \cap E_4) = P(\emptyset) = 0$ $\quad$ $P(E_3 \cap E_4) = P(\emptyset) = 0$

Total $= 1.4 - (0.1 + 0.2 + 0.1) = 1$

③ $P(E_1 \cap E_2 \cap E_3) = P(\emptyset) = 0$ $\quad$ $P(E_1 \cap E_2 \cap E_4) = P(\emptyset) = 0$

$P(E_2 \cap E_3 \cap E_4) = P(\emptyset) = 0$

Total $= 1 + (0) = 1$

④ $P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(\emptyset) = 0$

Total $= 1$

Voila! It all checks out.

Questions

We've explored the actual concepts themselves but very little about answering problems. There's some tricky terminology and wording that crop up, and sometimes it's difficult to determine what the question wants you to do.

ex. A coin is tossed until you get the same result twice in a row for the first time. Every possible outcome that takes n tosses to complete has its probability described as $1/2^{n-1}$.

i) Describe the sample space.

This is the first oddity - what does 'describe' mean? In actuality, there are a few distinct properties of a particular sample space.

- how many outcomes?
- what does a single elementary event within a sample space look like?
- what is the probability of each elementary event?

Let's answer these in order:

a) There are an infinite number of outcomes, ranging from $n=2$ to $n=\infty$: $[2, \infty)$, as you could conceivably toss a coin and have an infinite number of alternating heads/tails before the same result appears twice in a row.

b) In this case, an outcome is technically just n = something, but it's easier to understand if we lay it out visually:

n = 2: HH TT
n = 3: HTT THH
n = 4: HTHH THTT
n = x: (HT...) HH (TH...) TT

c) As described in the question, the probability of a specific outcome n is $1/2^{n-1}$.

ii) What is the probability of the experiment ending before the sixth toss?

So this means we want n = 2 OR 3 OR 4 OR 5, in other words, the union of these events.

$P(2) = 1/4$ $P(3) = 1/8$ $P(4) = 1/16$ $P(5) = 1/32$

with the resulting probability being $15/32$.

But remember, there's a $1/2^{n-1}$ chance for an outcome to end at n tosses - but there's two distinct ways for n = 2 to end: HH or TT. This is the case with all other outcomes.

As such, the actual probability is $15/16$,* without needing to add/subtract anything extra as all our events are disjoint.

*double

ii) What is the probability of an outcome requiring an even number of tosses?

Again, our disjoint probabilities:

$$\frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \ldots = \sum_{n=2} \frac{1}{2^{n-1}} \quad \forall n \% 2 = 0$$

$\uparrow$ $n=2$ $\quad \uparrow$ $n=4$ $\quad \uparrow$ $n=6$

$$= \frac{1/4}{1 - 1/4}$$

$$= \frac{1}{3}$$

And once more, since there are two sides to a coin, our overall probability resolves to $2/3$.

Let's do one more.

ex. Two players A and B have a and b dollars, respectively. They flip a coin - on a head, A gets one dollar from B and vice versa. They continue until one goes bankrupt, requiring them to sell a kidney to make up for their losses.

Determine the probability that A goes bankrupt and is required to sell a kidney.

The first, most important, thing to realize is that whatever equation we construct to solve the problem has to work at any point in A and B's game.

This is a bit confusing, but hopefully it'll clear up as we move along.

So let's assume at some point in the game

A: n dollars
B: a+b-n dollars

(We can do a more formal proof of why B has that much money but it's inconsequential to the final answer)

From here, we can let $P(n)$ be the probability that A eventually loses, starting at n dollars. With the next flip, A will either give away or receive a dollar, with a 50% chance (obviously, it's a coin flip).

So there are two scenarios after the coin flip:

W = A wins this flip, but loses eventually (A: n+1)
L = A loses this flip, and loses eventually (A: n-1)

$$\therefore \quad P(W) = \tfrac{1}{2} P(n+1)$$
$$P(L) = \tfrac{1}{2} P(n-1)$$

Since winning and losing the flip are disjoint events,

$$P(A \text{ loses}) = P(W) + P(L)$$
$$P(n) = \tfrac{1}{2} P(n+1) + \tfrac{1}{2} P(n-1)$$

Remember the term recurrence relation? This is what we've built here - an equation that references itself. To solve this, we have to then build a characteristic equation:

$$P(n) = \tfrac{1}{2}P(n+1) + \tfrac{1}{2}P(n-1)$$
$$0 = -P(n) + \tfrac{1}{2}P(n+1) + \tfrac{1}{2}P(n-1)$$
$$0 = \tfrac{1}{2}P(n+1) - P(n) + \tfrac{1}{2}P(n-1)$$
$$0 = P(n+1) - 2P(n) + P(n-1)$$
$$0 = \lambda^2 - 2\lambda + 1$$

Now, we solve for the characteristic roots.

$$0 = (\lambda-1)(\lambda-1)$$
$$\therefore\ \lambda = 1, \text{ twice}$$

Remember there are 3 cases of what to do when we find roots*:

Case 1: distinct real roots

$$y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} + \dots$$

Case 2: complex roots $(\lambda \pm \mu j)$

$$y(t) = c_1 e^{\lambda t}\cos(\mu t) + c_2 e^{\lambda t}\sin(\mu t)$$
(by Euler's formula)

Case 3: repeated roots

$$y(t) = c_1 e^{\lambda t} + c_2 t e^{\lambda t} + \dots$$

Christ, even I wasn't expecting an ECE205 recap in the 2nd chapter of probability.

We're gonna utilize case 3

*general equations for solutions

$$P(n) = c_1 e^n + c_2 n e^n$$

Now, we have to solve for the constants c_1 and c_2.
We need two base cases.

$P(0) = 1$: If A starts off with 0 dollars, it's certain that he's headed to the surgery room

$P(a+b) = 0$: If A has all the money available, there's no way he'll lose.

Let's plug these in.

$$P(0) = c_1 e^0 + c_2 (0) e^0$$
$$1 = c_1$$

$$P(a+b) = c_1 e^{a+b} + c_2 (a+b) e^{a+b}$$
$$0 = e^{a+b} (c_1 + (a+b) c_2)$$
$$0 = 1 + (a+b) c_2$$
$$\frac{-1}{a+b} = c_2$$

And finally, we can construct our equation to calculate the probability of A losing.

$$P(n) = 1 - \frac{n}{a+b}$$

$$\therefore P(a) = 1 - \frac{a}{a+b}$$
$$= \frac{a+b}{a+b} - \frac{a}{a+b}$$
$$= \frac{b}{a-b}$$