Probability and Statistics: A Primer for Beginners and Pre-Beginners

The Journey Begins: Probability Theory

Part Two: The Axioms, Explained

Primary reference: Casella-Berger 2nd Edition

Left you on a bit of a cliffhanger last time…

You got those axioms you wanted! But what do they mean?

Well, like I said, they don't really mean anything other than what they state. Axiomatically, probability is just a function P(x), where $\mathcal{B}$ (all open sets in Ω) is the domain (possible inputs to the function) and [0,1] is the range (possible outputs of the function).

This probability function is defined by three axioms, or rules:

Axioms of Probability

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$

Simple enough, this just means the probability of an event can't be negative.

2. $P(\Omega) = 1$

The probability of the sample space is 1. Pretty intuitive, since the sample space contains all possible outcomes of an experiment

3. If sets $A_1, A_2, A_3, \ldots \in \mathcal{B}$ are pairwise disjoint, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

A little more complicated but it just means that if events don't overlap at all, then the probability of the combined (union) events is the sum of the probabilities of the individual events.

Kinda…anti-climactic

Maybe at first glance, but we can do some pretty crazy things with these. Lets bring back our friends the coins from the first section on set theory:

$$\Omega = \{ H , T \}$$

Now, <u>outside the axioms</u>, we could say, intuitively, and assuming a fair coin, the probabilities of flipping heads and tails are equal:

$$P(H) = P(T)$$

Kinda…anti-climactic

But we also know that heads and tails are disjoint (intersection is ∅), and they partition Ω (union is Ω in addition to being disjoint):

$$\Omega = H \cup T \qquad \emptyset = H \cap T$$

So by axiom 2 (because they partition the sample space):

$$P(H \cup T) = P(\Omega) = 1$$

And by axiom 3 (because they are disjoint):

$$P(H \cup T) = P(H) + P(T)$$

Getting interesting…

Lets put this together:

(Axiom 3) (Axiom 2)

$$P(H \cup T) = P(H) + P(T) = P(\Omega) = 1$$

$$\Rightarrow P(H) + P(T) = 1$$

But earlier we posited <u>outside the axioms</u>:

$$P(H) = P(T)$$

So close…

So we can substitute tails with a second heads:

$$P(H) + P(T) = P(H) + P(H) = 1$$

$$\Rightarrow 2P(H) = 1$$

$$\Rightarrow P(H) = \frac{1}{2} = 0.5$$

That was a lot of work to calculate a 50-50 chance of landing heads-up! But remember, that equality of probability was based on our own assumption of a fair coin. An unfair coin could have a probability of landing heads-up equal to, say, 0.2.

Are we gonna have to do that every time?!

Nah, we can derive a definition of probability that doesn't have us referencing the axioms all the time, which is great because experiments get a lot more complicated than a coin toss!

So let's make some assumptions and then calculate the probability of some event A:

1. $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n\}$ is finite
2. Our old friend $\mathcal{B}$ is any collection of subsets of Ω.
3. We've got some nonnegative numbers $p_1, p_2, p_3, \ldots, p_n$ that sum to 1.

$$P(A) = \sum_{\{i:\omega_i \in A\}} p_i$$

This bottom part looks tricky, but its pretty simple. The sum only includes values of p_i for which ω_i, which we know is part of the sample space and thus a possible outcome, is an element in A.

Can we do an example to show how this thing works?

Anything for you, buddy. Lets assume set $A_1 = \{\omega_1, \omega_2\}$

$$P(A_1) = \sum_{\{i:\omega_i \in A_1\}} p_i = p_1 + p_2$$

$(\omega_1, \omega_2 \in A_1)$

So what's the point?

The point is that this compact little function satisfies the axioms of probability, and now, brace yourself, because we're gonna prove it!

Axiom 1. P(A) ≥ 0 for all $A \in \mathcal{B}$. All p_i are nonnegative, so their sum and thus P(A) ≥ 0.

Axiom 2. P(Ω) =1. Let's plug it into the function and see what happens!

$$P(\Omega) = \sum_{\{i:\omega_i \in \Omega\}} p_i = \sum_{i=1}^{n} p_i = 1$$

(all ω_i are in Ω, so we sum all p_i)

($p_1, p_2, p_3, \dots, p_n$ sum to 1)

So what's the point?

Proving that the function satisfies Axiom 3 gets a little complicated, so we'll take it nice and slow.

Axiom 3. If sets $A_1, A_2, A_3, \dots, A_k \in \mathcal{B}$ are pairwise disjoint, then $P(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i)$
We're swapping infinity for k since $\mathcal{B}$ contains finite sets, so we're only looking at finite unions.

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{\{i:\omega_i \in \bigcup_{i=1}^{k} A_i\}} p_i = \sum_{i=1}^{k} \sum_{\{j:\omega_j \in A_i\}} p_j = \sum_{i=1}^{k} P(A_i) \Rightarrow P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i)$$

(definition of the function)

We broke the earlier summation down into two parts. The inner summation sums over p_j for all j such that ω_j is an element of the set A_i. Basically, it sums p for each ω in A_i

The outer summation just adds up the sums calculated for each A_i in the inner summation. It's like looping over a loop in programming.

But the inner summation, $\sum_{\{j:\omega_j \in A_i\}} p_j$, is just the function P(A_i), so now we are summing the probabilities of each event in the union! Axiom 3 is proven!

A little more explanation…

The double summation may still read like Greek to you, and to be fair, there's definitely some Greek in there, so lets break that part down a little more and do it in a slightly different order to get rid of the double summation.

Axiom 3. If sets $A_1, A_2, A_3, \dots, A_k \in \mathcal{B}$ are pairwise disjoint, then $\mathrm{P}(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i)$

(definition of our probability function)

$$P(\bigcup_{i=1}^{k} A_i) = \sum_{\{i:\omega_i \in \bigcup_{i=1}^{k} A_i\}} p_i$$

(breaking down the summation to one summation per set A_i)

$$= \sum_{\{j:\omega_j \in A_1\}} p_j + \sum_{\{j:\omega_j \in A_2\}} p_j + \sum_{\{j:\omega_j \in A_3\}} p_j + \cdots + \sum_{\{j:\omega_j \in A_k\}} p_j$$

(those summations, though, are just the definition of our probability function for each set A_i)

$$= P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_k)$$

$$= \sum_{i=1}^{k} P(A_i)$$

So we just sum the probabilities of the sets, and this is the result we were trying to prove!

The secret bonus axiom

Axiom 3. If sets $A_1, A_2, A_3, \ldots \in \mathcal{B}$ are pairwise disjoint, then $\mathrm{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Like we said earlier, this axiom is a little more complicated than the others. For some statisticians back in the 1970's, a little more complicated was *too much more* complicated. So they rejected it. You can totally do that with axioms.

They substituted a different axiom that's way simpler, called the *axiom of finite additivity*:

Alternative Axiom 3. If sets $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint,
$P(A \cup B) = P(A) + P(B)$