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

The Journey Begins: Probability Theory

Part Three: Axioms and Consequences

Primary reference: Casella-Berger 2nd Edition

Once more: Axioms of Probability

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

2. P(Ω) = 1

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

Why am I showing you this again? Because, dear reader, there is yet more these beauties can do! Let's use them to prove some other basic properties.

…but you're still gonna help visualize this stuff, right?

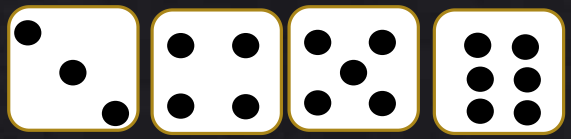
Of course! Throughout this lesson, we'll look at properties dealing with the complements A and A^C, and with the more general events A and B. So for A and A^C, let's use the coins:

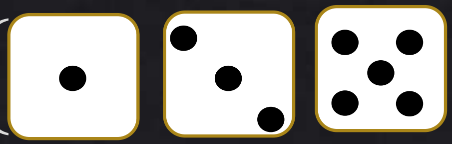
$A = \{H\}$ $\qquad A^C = \{T\}$

For more general events A and B, let's use faces of a die as shown below. Note, however, that these events are not disjoint and their union doesn't include every element in the sample space. Neither of these properties is necessary. We're just trying to generalize:

A = {  }

B = {  }

A^C = {  }

B^C = {  }

Oh, the properties we'll prove!

I. $P(A^C) = 1 - P(A)$: the probability of the complement of A is just 1 minus the probability of A

Remember Axiom 2: P(Ω) = 1, and the axiom of finite additivity: If sets $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint, then $P(A \cup B) = P(A) + P(B)$

Well, $A \in \mathcal{B}$ and $A^C \in \mathcal{B}$ are disjoint, so $P(A \cup A^C) = P(A) + P(A^C)$, and also remember that $A \cup A^C = \Omega$. Put it all together and:

$$1 = \mathrm{P}(\Omega) = P(A \cup A^C) = P(A) + P(A^C) \Rightarrow P(A) + P(A^C) = 1 \Rightarrow P(A^C) = 1 - P(A)$$

(Axiom 2) **(definition of complement)** **(Axiom 3)** **(rearrange)**

Oh, the properties we'll prove! (with pictures)

I. $P(T) = 1 - P(H)$: the probability of the complement of H is just 1 minus the probability of H

Remember Axiom 2: $P(\Omega) = 1$, and the axiom of finite additivity: If sets $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint, then $P(A \cup B) = P(A) + P(B)$

Well $H \in \mathcal{B}$ and $T \in \mathcal{B}$ are disjoint, so $P(H \cup T) = P(H) + P(T)$, and also remember that $H \cup T = \Omega$. Put it all together and:

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

(Axiom 2) (definition of complement) (Axiom 3) (rearrange)

Oh, the properties we'll prove!

II. $P(A) \leq 1$

Well, we just proved that $P(A^C) = 1 - P(A)$, and since $A^C \in \mathcal{B}$ then by axiom 1, $P(A^C) \geq 0$, and so P(A) cannot be greater than 1, since if P(A) was greater than 1, P(A^C) would have to be negative.

Oh, the properties we'll prove! (with pictures)

II. P(H) ≤ 1

Well, we just proved that P(T) $= 1 - P$(H), and since T $\in \mathcal{B}$ then by axiom 1, P(T) ≥ 0, and so P(H) cannot be greater than 1, since if P(H) was greater than 1, P(T) would have to be negative.

Oh, the properties we'll prove!

III. $P(\emptyset) = 0$

This one is my favorite because even though it's super easy, I still choked when asked the question during my graduate exam.

$\emptyset \cap \Omega = \emptyset$, so $\emptyset$ and Ω are disjoint (and again both are elements of $\mathcal{B}$). By the axiom of finite additivity, then, $\mathrm{P}(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega)$. But at the same time, $\emptyset \cup \Omega = \Omega$, and by axiom 2, $P(\Omega) = 1$. Put it all together:

(by definition) **(Axiom 3)** **(Axiom 2)**

$$P(\Omega) = \mathrm{P}(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega) = P(\emptyset) + 1$$

$$\Rightarrow P(\Omega) = 1 = P(\emptyset) + 1$$

$$\Rightarrow 1 = P(\emptyset) + 1 \Rightarrow P(\emptyset) = 0$$

Those were kind of easy… Why aren't *they* axioms?

You're right: they're pretty simple, but they don't need to be axioms, since we're able to prove them using axioms we already have.

And trust me on this: things are gonna get a little more complicated with these next three properties. I'm gonna call these properties a, b, and, c, because there's a lot of callbacks in these three.

Oh, the (more difficult) properties we'll prove!

a. $P(B \cap A^c) = P(B) - P(B \cap A)$

Oh man, what are we gonna do?! Our axioms don't even HAVE intersections in them! Okay, don't panic, lets just arrange this in a different way and hope something comes to us…

$$P(B \cap A) + P(B \cap A^c) = P(B)$$

Hey, wait a minute, isn't there something funny about $B \cap A$ and $B \cap A^c$? Let's take the intersection of these two and investigate. By that ol' associative property, we have:

$$(B \cap A) \cap (B \cap A^c) = (B \cap (A \cap B \cap A^c)$$

Oh, the (more difficult) properties we'll prove!

(by definition of complementation)

But then, by that ol' commutative property, we got:

$$B \cap (A \cap B \cap A^c) = B \cap (B \cap A \cap A^c) = B \cap (B \cap \emptyset) = B \cap \emptyset = \emptyset$$

So $(B \cap A) \cap (B \cap A^c) = \emptyset$ and thus $B \cap A$ and $B \cap A^c$ are disjoint! That means we get to use that beautiful *axiom of finite additivity!* (though we are kind of going in reverse this time)

$$P(B \cap A) + P(B \cap A^c) = P((B \cap A) \cup (B \cap A^c)) = P(B \cap (A \cup A^c))$$

(Distributive property!!!!!)

Oh, the (more difficult) properties we'll prove!

But we remember (hopefully) that an event and its complement partition the sample space:

$$P\left(B \cap (A \cup A^c)\right) = P(B \cap \Omega) = P(B)$$

(Sample Space Ω)

Ok, that was tough and took a LOT of steps, so we're gonna combine them all into one succinct proof, and you can try to follow along:

Put it all together!

What we want to prove:

$$P(B \cap A^c) = P(B) - P(B \cap A) \Rightarrow P(B \cap A^c) + P(B \cap A) = P(B)$$

How we prove it:

(Axiom 3 applies, as shown earlier)

(Distributive Property)

(Properties of the sample space)

$$P(B \cap A^c) + P(B \cap A) = P((B \cap A^c) \cup (B \cap A)) = P(B \cap (A^c \cup A)) = P(B \cap \Omega) = P(B)$$

$$\Rightarrow P(B \cap A^c) + P(B \cap A) = P(B)$$

$$\Rightarrow P(B \cap A^c) = P(B) - P(B \cap A)$$

Using dice to illustrate a.

a. $P(B \cap A^c) = P(B) - P(B \cap A)$

We're not going to do the proof with dice because I value my sanity, but we will illustrate the property:

Oh, the (more difficult) properties we'll prove!

b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(Just different ways to depict the sample space used here)

$$P(A \cup B) = P(A \cup (B \cap \Omega)) = P(A \cup (B \cap (A \cup A^c)))$$
$$= P\big(A \cup (B \cap A) \cup (B \cap A^c)\big) = P\big(A \cup (B \cap A^c)\big)$$

(Distributive property!!!!!)

$A \cup (B \cap A) = (A \cap B) \cup (A \cap A) = A$

All right, so we established that $P(A \cup B) = P\big(A \cup (B \cap A^c)\big)$, and if you're getting used to this pattern, you might be wanting to see if maybe A and $B \cap A^c$ are disjoint so let's take the intersection:

$$A \cap (B \cap A^c) = (A \cap A^c) \cap B = \emptyset \cap B = \emptyset$$

Aha! They're disjoint, so we can use axiom 3!!!

Oh, the (more difficult) properties we'll prove!

(Good ol' Axiom 3!)

$$P(A \cup B) = P\big(A \cup (B \cap A^c)\big) = P(A) + P(B \cap A^c)$$

Ok, so it looks like we're already pretty close, but we're stuck with this weird $P(B \cap A^c)$ term. Or, rather, we *would be* if we hadn't proven earlier that $P(B \cap A^c) = P(B) - P(B \cap A)$! Let's take it on home!

$$P(A \cup B) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(B \cap A)$$

Oh yeah, we did it!

And now, b. with dice!

b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

P(⚀ ⚁ ∪ ⚁ ⚃ ⚅) =P(⚀ ⚁) +P(⚁ ⚃ ⚅) - P(⚀ ⚁ ∩ ⚁ ⚃ ⚅)

⟹ P(⚀ ⚁ ⚃ ⚅) = P(⚀ ⚁) + P(⚁ ⚃ ⚅) - P(⚁)

Notice how this property accounts for the double-counting caused by the overlapping ⚁

Oh, the (more difficult) properties we'll prove!

c. $If\ A \subset B, then\ P(A) \leq P(B)$

(is a subset of)

Whoa, set notation? Since when are we combining THAT with probability functions? Well, it doesn't really matter, this is pretty simple: If A is a subset of B, then every element of A is also an element of B so $A \cap B = A$!

(Axiom 1) (property a) ($A \cap B = A$)

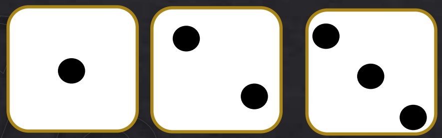
$$0 \leq P(B \cap A^c) = P(B) - P(B \cap A) = P(B) - P(A)$$

$$\Rightarrow 0 \leq P(B) - P(A) \Rightarrow P(A) \leq P(B)$$

Subset property, with dice

c. $If\ A \subset B, then\ P(A) \le P(B)$

For this one we need to quickly define a set of which A is a subset, so lets define C:

A = {  } ⊂ C = {  }

$Since\ A \subset C, then\ P(A) \le P(C)$

P() ≤ P()