Chapter 5

Conditional Probability and Independence

Lecture 5 | Devore 9e: 2.4

Last week we finished the axioms: a probability space $(S, \mathcal{F}, \mathbb{P})$ and a function $\mathbb{P} : \mathcal{F} \to [0, 1]$. This week: what happens to $\mathbb{P}$ when partial information arrives.

5.1 Conditional probability

Example 5.1 (The motivating computation)**.** Toss a coin twice, $S = \{(H, H), (H, T), (T, H), (T, T)\}$. Then $\mathbb{P}(2 \text{ heads}) = \frac{1}{4}$.

Now suppose someone tells you *there was at least one head.* Three outcomes remain possible, all still equally likely, so

$$\mathbb{P}(2 \text{ heads} \mid \text{at least 1 head}) = \tfrac{1}{3}.$$

What happened: the information shrank the sample space from S to B, and we renormalized so that B has total probability 1.

Definition 5.2 (Conditional probability)**.** Let $(S, \mathcal{F}, \mathbb{P})$ be a probability space and $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. For any $A \in \mathcal{F}$, the **conditional probability of A given B** is

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Example 5.3 (Die)**.** Roll a die, $S = \{1, \dots, 6\}$. Let $A = \{\text{outcome is even}\} = \{2, 4, 6\}$ and $B = \{\text{outcome} > 3\} = \{4, 5, 6\}$. Then $A \cap B = \{4, 6\}$, so

$$\mathbb{P}(A \mid B) = \frac{2/6}{3/6} = \frac{2}{3}, \qquad \mathbb{P}(B \mid A) = \frac{2/6}{3/6} = \frac{2}{3}.$$

(The two happen to agree here because $\mathbb{P}(A) = \mathbb{P}(B)$; in general they differ, and confusing them is the single most common error in this course.)

Example 5.4 (Disease testing — set up now, solved in Lecture 7)**.** A disease affects 1% of the population. Let $A = \{\text{person has the disease}\}$, so $\mathbb{P}(A) = 0.01$ and $\mathbb{P}(A^c) = 0.99$. A test satisfies

- if the person has the disease, the result is positive 99% of the time: $\mathbb{P}(B \mid A) = 0.99$;
- if the person does not, the result is negative 95% of the time: $\mathbb{P}(B^c \mid A^c) = 0.95$;

where $B = \{\text{test result is positive}\}$. The patient's question is not $\mathbb{P}(B \mid A)$ — it is $\mathbb{P}(A \mid B)$. These are wildly different numbers, and Bayes' theorem is what converts one into the other.

5.2 Properties

Theorem 5.5 (Multiplication rule)**.** *If* $\mathbb{P}(A) > 0$ *and* $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A) = \mathbb{P}(B)\,\mathbb{P}(A \mid B).$$

More generally, for events $A_1, \ldots, A_n$ *with* $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$,

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 A_2) \cdots \mathbb{P}(A_n \mid A_1 \cdots A_{n-1}).$$

Here juxtaposition means intersection: $A_1 A_2 := A_1 \cap A_2$.

This is the tool for anything described *sequentially*: draws from an urn, cards dealt one at a time, components failing in order.

Theorem 5.6 (Conditioning gives a probability)**.** *Fix* B *with* $\mathbb{P}(B) > 0$. *Then* $A \mapsto \mathbb{P}(A \mid B)$ *is itself a probability on* $(S, \mathcal{F})$. *In particular*

$$\mathbb{P}(A \mid B) \geq 0, \qquad \mathbb{P}(S \mid B) = 1, \qquad \mathbb{P}(\emptyset \mid B) = 0, \qquad \mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B),$$

and for pairwise disjoint $A_1, A_2, \ldots$, $\mathbb{P}\big(\bigcup_i A_i \mid B\big) = \sum_i \mathbb{P}(A_i \mid B)$.

Everything you know about $\mathbb{P}$ therefore holds for $\mathbb{P}(\cdot \mid B)$ as well. Note carefully that this is a statement about the *first* slot only: $\mathbb{P}(A \mid B^c)$ has nothing to do with $\mathbb{P}(A \mid B)$, and $\mathbb{P}(A \mid B) + \mathbb{P}(A \mid B^c)$ is not 1 in general.

5.2.1 Pólya's urn model

Example 5.7 (Pólya's urn)**.** An urn holds r red and b black balls. Repeatedly: draw one ball at random, put it back, and add c more balls of the color just drawn. After n draws, what is the probability that the first n_1 draws are black and the remaining $n_2 = n - n_1$ are red?

Let A_i be the event that draw i comes out the prescribed color. The multiplication rule handles it, because after each draw we know exactly what the urn contains:

$$\mathbb{P}(A_1) = \frac{b}{b+r}, \qquad \mathbb{P}(A_2 \mid A_1) = \frac{b+c}{b+r+c}, \quad \mathbb{P}(A_3 \mid A_1 A_2) = \frac{b+2c}{b+r+2c}, \quad \ldots$$

$$\mathbb{P}(A_{n_1} \mid A_1 \cdots A_{n_1-1}) = \frac{b+(n_1-1)c}{b+r+(n_1-1)c},$$

and then the reds begin, with the urn already holding $b + n_1 c$ black balls:

$$\mathbb{P}(A_{n_1+1} \mid \cdots) = \frac{r}{b+r+n_1 c}, \qquad \mathbb{P}(A_{n_1+2} \mid \cdots) = \frac{r+c}{b+r+(n_1+1)c}, \ \ldots, \ \mathbb{P}(A_n \mid \cdots) = \frac{r+(n_2-1)c}{b+r+(n-1)c}.$$

Multiplying,

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \frac{\prod_{i=0}^{n_1-1}(b+ic)\ \prod_{j=0}^{n_2-1}(r+jc)}{\prod_{k=0}^{n-1}(b+r+kc)}.$$

Remark 5.8. The answer depends only on *how many* blacks and reds, not on the order — so any sequence with n_1 blacks and n_2 reds has the same probability. Pólya's urn is the standard model for "success breeds success" (contagion, popularity, reinforcement). Setting $c = 0$ recovers sampling with replacement.

5.3 Does conditioning raise or lower the probability?

Neither, in general. Draw one card from a deck, and let $A = \{\text{card is red}\}$, so $\mathbb{P}(A) = \frac{1}{2}$.

B	$\mathbb{P}(A \mid B)$	
card is a heart	1	information helps
card is a spade	0	information hurts
card is an ace	$\frac{1}{2}$	information is irrelevant

The third row is the interesting one, and it deserves a name.

5.4 Independence

Definition 5.9 (Independence of two events)**.** Events A and B are **independent**, written $A \perp\!\!\!\perp B$, if

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \iff \mathbb{P}(B \mid A) = \mathbb{P}(B) \iff \boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)}.$$

The boxed form is the official definition: it is symmetric in A and B, and it makes sense even when $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$.

Proposition 5.10. *If $A \perp\!\!\!\perp B$ then $A^c \perp\!\!\!\perp B$, $A \perp\!\!\!\perp B^c$, and $A^c \perp\!\!\!\perp B^c$.*

Proof. $\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(A^c)\mathbb{P}(B)$. Apply this twice for the last claim. □

Remark 5.11. Independent $\neq$ mutually exclusive. If $A \cap B = \emptyset$ and both have positive probability, then $\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)$, so disjoint events are as *dependent* as they get: knowing A happened tells you B did not.

Example 5.12 (Drawing with replacement)**.** A bag has a black and b white balls. Draw twice *with replacement*. Let $A = \{\text{first draw black}\}$, $B = \{\text{second draw black}\}$. Then

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\frac{a}{a+b} \cdot \frac{a}{a+b}}{\frac{a}{a+b}} = \frac{a}{a+b},$$

while, splitting on the first draw,

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) = \frac{a^2}{(a+b)^2} + \frac{ba}{(a+b)^2} = \frac{a(a+b)}{(a+b)^2} = \frac{a}{a+b}.$$

So $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ and the draws are independent — as they should be, since we put the ball back. Without replacement they are not.

5.5 More than two events

For three or more events there are two competing notions, and they are not the same.

Definition 5.13 (Pairwise vs. mutual independence)**.** Events $A_1, \dots, A_n$ are

- **pairwise independent** if $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all $i \neq j$;
- **mutually independent** if for *every* sub-collection $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

Mutual independence $\Rightarrow$ pairwise independence. The converse is **false**, and here is the standard counterexample — next lecture.