Chapter 6

Independent Trials, Total Probability, Bayes' Theorem

Lecture 6 | Devore 9e: 2.4–2.5

Last time: conditional probability, the multiplication rule, and independence of two events. We ended by defining pairwise and mutual independence and claiming they differ. Let us settle that first.

6.1 Pairwise does not imply mutual

Example 6.1 (Bernstein's tetrahedron)**.** A regular tetrahedron has four faces, painted

$$\text{red}, \quad \text{white}, \quad \text{black}, \quad \text{red} + \text{white} + \text{black (all three)}.$$

Roll it and look at the face it lands on. Let R, W, K be the events that the resting face contains red, white, black respectively. Each color appears on exactly two of the four faces, so

$$\mathbb{P}(R) = \mathbb{P}(W) = \mathbb{P}(K) = \tfrac{2}{4} = \tfrac{1}{2}.$$

Two colors appear together on exactly one face (the tricolor one), so

$$\mathbb{P}(R \cap W) = \mathbb{P}(R \cap K) = \mathbb{P}(W \cap K) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2}.$$

So the three events are **pairwise independent**. But

$$\mathbb{P}(R \cap W \cap K) = \tfrac{1}{4} \neq \tfrac{1}{8} = \mathbb{P}(R)\mathbb{P}(W)\mathbb{P}(K),$$

so they are **not mutually independent**.

The moral: pairwise independence lets you multiply two at a time and nothing more. Whenever a problem says "independent" about three or more events, it means mutually independent.

6.2 Independent trials

So far independence has been a property of events inside one experiment. Usually we want it to be a property of *experiments*: repeated coin tosses, repeated measurements, repeated draws. That requires building a new probability space out of two old ones.

Let Experiment 1 have space $(S_1, \mathcal{F}_1, \mathbb{P}_1)$ and Experiment 2 have $(S_2, \mathcal{F}_2, \mathbb{P}_2)$. Take the **product**

$$S := S_1 \times S_2, \qquad \mathcal{F} := \mathcal{F}_1 \times \mathcal{F}_2,$$

and define $\mathbb{P}$ on rectangles $A = A_1 \times A_2$ by

$$\mathbb{P}(A_1 \times A_2) := \mathbb{P}_1(A_1) \cdot \mathbb{P}_2(A_2).$$

Definition 6.2 (Independent trials)**.** With the above $\mathbb{P}$, the triple $(S, \mathcal{F}, \mathbb{P})$ is a probability space, and Experiments 1 and 2 are called **independent trials**.

Example 6.3. $S_1 = \{1, 2, 3, 4, 5, 6\}$ (a die), $S_2 = \{\text{red}, \text{black}\}$ (a card color). Then $S_1 \times S_2 = \{(1, \text{red}), (1, \text{black}), (2, \text{red}), \dots\}$, with twelve outcomes, and $\mathbb{P}\big((3, \text{red})\big) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$.

Example 6.4 (Calm gambler)**.** One game: $\mathbb{P}(W) = 0.6$, $\mathbb{P}(L) = 0.4$, and games are independent trials. Play three games, so $S = \{W, L\}^3$. Then

$$\mathbb{P}(WWW) = \mathbb{P}(W)\mathbb{P}(W \mid W)\mathbb{P}(W \mid WW) = \mathbb{P}(W)^3 = 0.6^3 = 0.216,$$

$$\mathbb{P}(LLW) = 0.4 \times 0.4 \times 0.6 = 0.096.$$

Independence is exactly what lets us drop the conditioning bars.

Example 6.5 (Emotional gambler)**.** The same chances in the *first* game, $\mathbb{P}(W) = 0.6$, $\mathbb{P}(L) = 0.4$, but the player is rattled by wins and motivated by losses:

$$\mathbb{P}(W \mid W) = 0.2, \quad \mathbb{P}(L \mid W) = 0.8, \qquad \mathbb{P}(W \mid L) = 0.8, \quad \mathbb{P}(L \mid L) = 0.2,$$

and the player reacts only to the game just played, so $\mathbb{P}(W \mid WW) = \mathbb{P}(W \mid W)$ and so on. These are *not* independent trials, so we must keep the multiplication rule in full:

$$\mathbb{P}(WWW) = 0.6 \times 0.2 \times 0.2 = 0.024, \qquad \mathbb{P}(LLW) = 0.4 \times 0.2 \times 0.8 = 0.064.$$

Compare with the calm gambler: same first-game odds, very different three-game behavior.

6.3 The law of total probability

Theorem 6.6 (Law of total probability)**.** *Let $\{A_i\}_{i \geq 1}$ be a partition of S with $\mathbb{P}(A_i) > 0$ for every i. Then for any event B,*

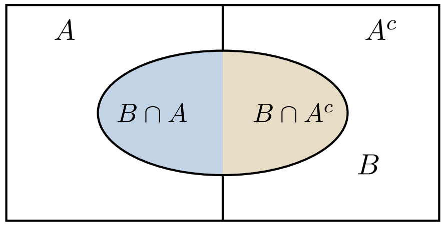
$$\mathbb{P}(B) = \sum_i \mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i).$$

In the simplest case ($A_1 = A$, $A_2 = A^c$): $\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B \mid A) + \mathbb{P}(A^c)\mathbb{P}(B \mid A^c)$.

Proof. Since $\bigcup_i A_i = S$,

$$\mathbb{P}(B) = \mathbb{P}\Big(B \cap \bigcup_i A_i\Big) = \mathbb{P}\Big(\bigcup_i (B \cap A_i)\Big) = \sum_i \mathbb{P}(B \cap A_i) = \sum_i \mathbb{P}(A_i)\mathbb{P}(B \mid A_i),$$

where the third equality is countable additivity: the sets $B \cap A_i$ are pairwise disjoint because the A_i are. $\square$



Remark 6.7. This is the workhorse. Whenever a problem has a hidden "which case are we in?" — which machine made the part, which route the person took, which box the coin came from — condition on the cases and add up.

Example 6.8 (Lost umbrella)**.** You lost your umbrella in one of three places, with prior probabilities

$$\mathbb{P}(A_1) = 0.50 \text{ (library)}, \quad \mathbb{P}(A_2) = 0.30 \text{ (classroom)}, \quad \mathbb{P}(A_3) = 0.20 \text{ (bus)},$$

and if it is there, the chance you actually recover it is

$$\mathbb{P}(B \mid A_1) = 0.80, \qquad \mathbb{P}(B \mid A_2) = 0.60, \qquad \mathbb{P}(B \mid A_3) = 0.05,$$

where $B = \{\text{umbrella is found}\}$. Then

$$\mathbb{P}(B) = 0.50(0.80) + 0.30(0.60) + 0.20(0.05) = 0.40 + 0.18 + 0.01 = 0.59.$$

6.4 Bayes' theorem

The law of total probability computes $\mathbb{P}(B)$ from the pieces. Bayes' theorem runs the arrow backwards: given that B happened, which piece were we in?

Theorem 6.9 (Bayes' theorem)**.** *Let $\{A_i\}$ be a partition of S with $\mathbb{P}(A_i) > 0$, and let $\mathbb{P}(B) > 0$. Then*

$$\boxed{\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i)}{\sum_j \mathbb{P}(A_j)\,\mathbb{P}(B \mid A_j)}}$$

Proof. By the definition of conditional probability and the multiplication rule,

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i)\mathbb{P}(B \mid A_i)}{\mathbb{P}(B)},$$

and expand the denominator with the law of total probability. □

6.4.1 Vocabulary

$\mathbb{P}(A_i)$	**prior**	what you believed before the evidence
$\mathbb{P}(B \mid A_i)$	**likelihood**	how well cause i explains the evidence
$\mathbb{P}(A_i \mid B)$	**posterior**	what you believe after the evidence

Example 6.10 (Lost umbrella, continued)**.** With $\mathbb{P}(B) = 0.59$ from above,

$$\mathbb{P}(A_1 \mid B) = \frac{0.40}{0.59} \approx 0.678, \quad \mathbb{P}(A_2 \mid B) = \frac{0.18}{0.59} \approx 0.305, \quad \mathbb{P}(A_3 \mid B) = \frac{0.01}{0.59} \approx 0.017.$$

The bus started at 20% and dropped to under 2%: recovering it at all is strong evidence it was never on the bus.

Example 6.11 (n heads in a row)**.** A coin is two-headed with probability $\mathbb{P}(A) = 0.05$, and fair with probability $\mathbb{P}(A^c) = 0.95$. You toss it n times and get all heads; call that B. Then $\mathbb{P}(B \mid A) = 1$ and $\mathbb{P}(B \mid A^c) = (1/2)^n$, so

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c)} = \frac{1 \cdot 0.05}{1 \cdot 0.05 + (1/2)^n \cdot 0.95} \xrightarrow[n\to\infty]{} 1.$$

n	1	2	3	5	10	20
$\mathbb{P}(A \mid B)$	0.095	0.174	0.296	0.627	0.982	0.99998

Remark 6.12. Even a heavily skeptical prior is overwhelmed by enough evidence — but notice how slowly it moves at first. After three heads the two-headed hypothesis is still under 30%. Cheap evidence does not flip strong priors; that is the whole content of the disease-testing example we set up last lecture and finish next time.