02-680 Module 13

Essentials of Mathematics and Statistics

Aidan Jan

November 11, 2025

Conditional Probability

Conditional probabiity restricts the sample space. If we know onw thing happened, what's the probability of another?

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Using the Axiom of Total Probability:

$$P(A|\Omega) = \frac{P(A \cap \Omega)}{P(\Omega)} = \frac{P(A)}{1}$$

Let $A_{\text{firstheads}}$ be the event that the first coin is heads. Let $A_{\text{allheads}} = \langle \text{Heads, Heads} \rangle$: both coins are heads. Computing conditional probability:

$$P(A_{\text{allheads}}|A_{\text{firstheads}}) = \frac{P(A_{\text{allheads}} \cap A_{\text{firstheads}})}{P(A_{firstheads})} = \frac{1}{2}$$

This is based on the fact that $A_{\text{allheads}} \subseteq A_{\text{firstheads}}$, and interpreted as "the probability you end up with both coin flips being heads, given that the first coin was heads."

Sum Rule

How do we calculate the probability one person has flu given that the person has a cough?

- F represents flu.
- C represents cough.
- We want $P(F|C)$. We can write:

$$P(F|C) = \frac{P(F \cap C)}{P(C)}$$

But, how do we find $P(C)$? We have:

$$P(C) = \frac{\text{Num people with a cough}}{\text{Num people in population}}$$

If we have some part of the space $A_1, A_2, \ldots, A_n$ (remember a partition is a set of sets that are all disjoint), but

$$\bigcup_{i=1}^{n} A_i = \Omega$$

then we can say the following:

$$P(C) = P(C \cap F) + P(C \cup \bar{F}) = 0.3 + 0.15$$

Therefore,

$$P(F|C) = \frac{P(C \cap F)}{P(C)} = \frac{0.3}{0.45} = \frac{2}{3}$$

Product Rule

The product rule states that

$$P(A \cap B) = P(A|B)P(B)$$

Chain Rule

The chain rule is an expansion of the product rule for multiple events.

$$\begin{aligned}
&P(D_1 \cap D_2 \cap \cdots \cap D_n) \\
&= P(D_n|D_1 \cap D_2 \cap \cdots \cap D_{n-1})P(D_1 \cap D_2 \cap \cdots \cap D_{n-1}) \\
&= \ldots \\
&= P(D_1)P(D_2|D_1)P(D_3|D_1 \cap D_2)\ldots P(D_n|D_1 \cap D_2 \cap \cdots \cap D_{n-1})
\end{aligned}$$

Independence

Events A and B are **independent** if

$$P(A \cap B) = P(A) \cdot P(B)$$

We also know that, for any pair of A and B, $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$. Hence, we can conclude that if A and B are independent:

$$P(A|B) = P(A)$$

They are conditionally independent if

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

Conditional Independence vs. Independence

- Independence: $P(A \cap B) = P(A)P(B)$
- Conditional Independence: $P(A \cap B|C) = P(A|C)P(B|C)$

For example, suppose we have two coins, one regular, one two-headed. We randomly choose one coin and flip it twice. Let

- A: First toss is heads
- B: Second toss is heads
- C: We chose the regular coin

In this case, A and B are conditionally independent give C, but are not independent.

Properties of Conditional Probabiltiy

$P(\cdot|B)$ satisfies the three axioms of probability (for fixed B):

- Non-negativity: $P(A|B) \geq 0$
- Normalization: $P(\Omega|B) = 1$
- Countable additivity of mutually disjoint events

$$P\left(\bigcap_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

Note that, in general,

$$P(A|B \cup C) \neq P(A|B) + P(A|C)$$

Bayes' Rule

For some probability space $(\Omega, \mathcal{A}, P)$, and for some event that already happened $B \in \mathcal{A}$. Then the probability that some event $A \in \mathcal{A}$ also happens:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where $p(B) > 0$. For events A and B we have $p(A \cap B) = p(A|B)p(B) = p(B|A)p(A)$.

Each part of Bayes' Rule has a name. We call

- $p(A)$ - the prior, the initial belief before seeing evidence
- $p(B|A)$ - the likelihood, indicates how likely evidence is given the hypothesis
- $p(A|B)$ - the posterior, the updated belief after incorporating evidence.

Another Form of Bayes' Rule

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)p(A)}{p(B|A)p(A) + p(B|\bar{A})p(\bar{A})}$$

Inverse Probability Reasoning

Bayes' Rule helps us go backward from effect to possible causes using known likelihoods.

Forward probability:

$$P(\text{effect}|\text{cause})$$

Backward probability:

$$P(\text{cause}|\text{effect})$$

Real-world Application

Suppose we want to estimate the probability that a person has breast cancer given they carry a BRCA1 mutation:

$$P(\text{BreastCancer}|\text{BRCA1+}) = \frac{P(\text{BRCA1+}|\text{BreastCancer}) \cdot P(\text{BreastCancer})}{P(\text{BRCA1+})}$$

- Prior: baseline population risk of breast cancer
- Likelihood: how frequently BRCA1 mutation occurs among cancer patients
- Posterior: updated risk given the test result

This is the mathematical foundation for clinical genetic testing and personalized medicine.