Probability and Independence

Robert Y. Lewis

CS 0220 2024

April 8, 2024

Overview

Probability refresher

A *sample space* is a set of possible *outcomes*. An *event* is a subset of the sample space. A *probability function* assigns each outcome a probability between 0 and 1.

Conditional probability

Definition: The conditional probability of event A given event B (with $\Pr[B] > 0$) is:

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Conceptually, if we limit ourselves to the outcomes in B, how likely is an outcome in A?

Example:

- A: Die shows a number divisible by 3. $\Pr[A] = 1/3$. (3 and 6 from the six possibilities.)
- B: Die shows an odd number. $\Pr[B] = 1/2$.
- What does $\Pr[A \cap B]$ mean? Die shows an odd number divisible by 3. $\Pr[A \cap B] = 1/6$ (only 3).
- What does $\Pr[A|B]$ mean? Die shows a number divisible by 3 given that it's odd. $\Pr[A|B] = 1/3$ (probability of picking 3 from 1, 3, 5). Also, $\frac{1}{6}/\frac{1}{2} = \frac{1}{6} \times 2 = \frac{1}{3}$.

Independence

Definition: Event A is *independent of* event B iff

$$Pr[A|B] = \Pr[A].$$

If $\Pr[B] = 0$, we say it is independent of any other event including itself.

Example:

- A: Die shows the maximum or minimum number. $\Pr[A] = 1/3$. (1, 6 from the six possibilities.)
- B: Die shows an odd number. $\Pr[B] = 1/2$.
- $\Pr[A|B] = 1/3$. (1 from 1,3,5.) So, A and B are independent.
- C: Die shows an even number. $\Pr[C] = 1/2$.
- Are B and C independent? No, $\Pr[C|B] = 0 \neq \Pr[C]$. Common misconception. Independent does not mean disjoint.

Independent events multiply

Theorem: A is independent of B iff

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B].$$

Proof: By cases.

- Case 1: If $\Pr[A] = 0$ or $\Pr[B] = 0$, then $\Pr[A \cap B] = 0$. Equality and independence are both achieved.
- Case 2: Otherwise,

$$\begin{aligned} A \text{ is independent of } B &\iff \Pr[A|B] = \Pr[A] \text{ (def. independence)} \\ &\iff \Pr[A \cap B]/Pr[B] = Pr[A] \text{ (def. cond. prob.)} \\ &\iff \Pr[A \cap B] = Pr[A] \cdot Pr[B] \text{ (multiplying across)} \end{aligned}$$

Independent coin flips

I flip a fair coin twice.

- *A*: first flip is heads.
- *B*: second flip is heads.

Independent events?

Hash collisions

I have a perfect hash function and two pieces of data a and b to insert into a hashtable.

- A: a has a hash collision.
- B: b has a hash collision.

Independent events?

Independence: "knowledge about one event does not give us knowledge about another."

Mutual Independence

Definition: A set of events $E_1, E_2, \ldots, E_n$ is *mutually independent* iff for all subsets $S \subseteq [1, n]$,

$$\Pr\left[\bigcap_{j \in S} E_j\right] = \prod_{j \in S} \Pr[E_j].$$

Example: If we toss n fair coins, the tosses are mutually independent iff for every subset of m coins, the probability that every coin in the subset comes up heads is 2^{-m}.

Pairwise independence isn't mutual independence

If A is independent of B and C, and B and C are independent of each other, how could A, B, and C not be independent??

Example: Jania, Tyler, Carmen each pick a bit 0/1 uniformly at random.

- A: Jania + Tyler $\equiv 1 \pmod 2$
- B: Tyler + Carmen $\equiv 1 \pmod 2$
- C: Jania + Carmen $\equiv 1 \pmod 2$

Claim 1: These events are all pairwise independent.

For example, $\Pr[A] = 1/2$. $\Pr[A|B] = \frac{1}{4}/\frac{1}{2} = 1/2$.

Claim 2: These events are not mutually independent.

$\Pr[A \cap B \cap C] = 0$. Not $1/8$!

k-wise does not imply $(k+1)$-wise mutual independence.

What are we even measuring?

What does it mean to say:

- The probability that a sequence of 5 coin flips will be all heads is $1/32$?
- The probability that it will rain tomorrow is .8?
- The probability that $2^{6972607} - 1$ is a prime number is...?

Two interpretations: *frequentist* vs *Bayesian*.

Frequentism vs Bayesianism

The frequentist: probability statements only apply to *repeatable* events. Coin flip example makes sense: if we repeatedly sample from the sample space, the "all heads" event will happen in about $1/32$ of the samples. Rain question is meaningless. Prime number question is either 0 or 1.

The Bayesian: probability statements describe a subjective *belief*/level of confidence. I'd take either side of a bet with 1:32 odds on the coin flip sequence. I think it's four times more likely to rain tomorrow than not. There's an objective answer to the prime number question, but based on my current information/computation ability, ...

Mathematical probability: tries to be agnostic between these two interpretations.