2: Conditional Probability, and Independence

PSTAT 120A: Summer 2022

Ethan P. Marzban
June 22, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability; Probability Measure $\mathbb{P}$
- Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$
- Classical Definition of Probability
- Probability Rules (e.g. Complement Rule, Set Difference Rule, etc.)

Conditional Probability

Leadup

- Given an event A, the quantity $\mathbb{P}(A)$ represents our beliefs about the event A.
- Suppose we get some more information in the form of another event B.
- How, if at all, do our beliefs about A Change?

- As an example: suppose we want to estimate the chance of rain. In the absence of any information, we might say that the chance of rain tomorrow is 50%.
- But, we know that it is summer, in Santa Barbara; thus, we intuitively feel that the true chance of rain should probably be lower than 50%.

Proposition

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an event $B \in \mathcal{F}$ such that $\mathbb{P}(B) \neq 0$, the probability measure $\mathbb{P}_B : \mathcal{F} \to \mathbb{R}$ defined by

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

is a valid probability measure.

- I won't prove this (instead I've made it an extra problem on Worksheet 2!)
- Often times, instead of writing $\mathbb{P}_B(A)$ we will write $\mathbb{P}(A \mid B)$, read "the probability of A **given** B."
- $\mathbb{P}(A \mid B)$ represents an **updating** of our beliefs on A, in the presence of B.
  - Why is this an "updating?" Well, really $\mathbb{P}_B(A)$ is the proportion of B that is explained by A.
  - Sometimes read "if B, then A."

Example

Suppose I randomly select a number from the set $[|1:100|]$ (this is a shorthand notation for $\{1, 2, \cdots, 100\}$). Define the events A and B as follows:

$$A := \{\text{the number I selected was strictly greater than 50}\}$$
$$B := \{\text{the number I selected was a multiple of 5}\}$$

- Because the selection is done "randomly," we can use the classical definition of probability.
  - There are 50 numbers greater than 50 (that are in the set $[|1:100|]$), meaning $\mathbb{P}(A) = 50/100 = 1/2$.
  - There are 20 multiples of 5 in the set $[|1:100|]$, meaning $\mathbb{P}(B) = 20/100 = 1/5$.
- Additionally, $A \cap B$ represents the event "the number I selected was both greater than 50 and a multiple of 5." There are 10 multiples of 5 that are greater than 50; therefore $\mathbb{P}(A \cap B) = 10/100 = 1/10$.
- Thus, putting everything together,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/10}{1/5} = \frac{5}{10} = \frac{1}{2}$$

Multiplication Rule

- Our notion of conditional probability gives us a way of computing probabilities of intersections: since
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$
we can multiply both sides by $\mathbb{P}(B)$ to obtain:

Formula: The Multiplication Rule

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two events $A, B \in \mathcal{F}$ with $\mathbb{P}(B) \neq 0$,

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

- As an example: if A and B are two events with $\mathbb{P}(A) = 2/5$ and $\mathbb{P}(B \mid A) = 1/4$, then $\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A) = (1/4)(2/5) = 1/10$

Example

A recent survey at the *Isla Vista Co-Op* revealed that 50% of customers buy bread. Of those customers who buy bread, 20% buy cheese.

- **Always define notation first!** Let B denote "customer buys bread" and C denote "customer buys cheese." Then the problm tells us

$$\mathbb{P}(B) = 0.5; \quad \mathbb{P}(C \mid B) = 0.2$$

- We seek $\mathbb{P}(B \cap C)$. Since $\mathbb{P}(B \cap C) = \mathbb{P}(C \mid B) \cdot \mathbb{P}(B)$, we conclude that the proportion of customers who buy bread and cheese is

$$(0.2) \cdot (0.5) = \boxed{10\%}$$

Partitions (Again?)

- Now that we have the multiplication rule, we can derive a very useful formula.
- Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an event $A \in \mathcal{F}$.
- Consider another event $B \in \mathcal{F}$, and say we want to compute $\mathbb{P}(A)$.
- It is either the case that A happened along with B, or it happened along with not-B. That is,

$$A = [A \cap B] \cup [A \cap B^{\complement}]$$

- Taking the probability of both sides, and invoking the third axiom of probability, we find

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^{\complement})$$

Partitions (Again?)

- Let's generalize this further. Suppose we have a partition $\{B_i\}_{i=1}^{\infty}$ of Ω. Then:
  - Either A happened along with B_1,
  - ... or B_2,
  - ... or B_3,
  - and so on and so forth.
- Therefore,

$$A = \bigcup_{i=1}^{\infty} (A \cap B_i)$$

and, taking the probability of both sides,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i)$$

- Since $\mathbb{P}(A \cap B_i) = \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)$, we can rewrite this as:

Formula: The Law of Total Probability

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)$$

Leadup

- Let's go back to our definition of conditional probability:

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

- Note that $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap A)$.
- By the multiplication rule, $\mathbb{P}(B \cap A) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A)$.
- Hence, we have derived the following result:

Formula: Bayes' Theorem

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

- Colloquially, Bayes' Rule gives us a way of "reversing the order" of a conditional. This is especially useful when we have some sort of temporality.
- Oftentimes, we will use the Law of Total Probability in the denominator of Bayes' Rule.

Example: On the Chalkboard

In *Gauchoville*, motherboards are manufactured by three companies (called A, B, and C). 20% of motherboards manufactured in factory A are defective; 30% of those manufactured in factory B are defective, and 10% of those manufactured in factory C are defective. Additionally, Factory A is responsible for 10% of the motherboards sold in *Gauchoville*, B is responsible for 50%, and C is responsible for the remaining 40%.

(a) If a motherboard is selected at random, what is the probability that it is defective?

(b) Suppose that a randomly selected board was defective. What is the probability that it came from factory A?

The Multiplication Rule, Generalized

Formula: Multiplication Rule for n events

If $A_1, \cdots, A_n$ are events (and all of the conditional probabilities below are well-defined), we have

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \times \mathbb{P}(A_2 \mid A_1) \times \mathbb{P}(A_3 \mid A_1 \cap A_2) \times \cdots \times \mathbb{P}(A_n \mid A_1, A_2, \cdots, A_{n-1})$$

Example (2.7 from ASV)

Suppose an urn contains 8 red balls and 4 white balls. Draw four balls at random, without replacement; what is the probability that the first two draws are red and the third and fourth draws are white?

- Define

$$R_i := i^{\text{th}} \text{ ball was red}; \quad W_i := i^{\text{th}} \text{ ball was white}$$

  so that the quantity we seek can be written as

$$\mathbb{P}(R_1 \cap R_2 \cap W_3 \cap W_4)$$

- By the multiplication rule,

$$\mathbb{P}(R_1 \cap R_2 \cap W_3 \cap W_4) = \mathbb{P}(R_1) \times \mathbb{P}(R_2 \mid R_1) \times \mathbb{P}(W_3 \mid R_1 \cap R_2) \times \mathbb{P}(W_4 \mid R_1 \cap R_2 \cap R_3)$$

Example (2.7 from ASV)

Suppose an urn contains 8 red balls and 4 white balls. Draw four balls without replacement; what is the probability that the first two draws are red and the third and fourth draws are white?

- $\mathbb{P}(R_1) = \frac{8}{12}$, since there are initially 12 balls of which 8 are red (note that we used the classical definition of probability here, since our selection was done “at random”)
- $\mathbb{P}(R_2 \mid R_1)$ denotes the probability of drawing a red ball second, after having drawn a red ball first. Since we drew a red ball first, there are only 11 balls remaining of which 7 are red; hence $\mathbb{P}(R_2 \mid R_1) = \frac{7}{11}$.
- $\mathbb{P}(W_3 \mid R_1 \cap R_2)$ denotes the probability of drawing a white ball third, after having drawn a red ball first and another red ball second. Since we have already drawn two balls, both of which were red, we have a total of 10 marbles of which 4 are white; hence $\mathbb{P}(W_3 \mid R_1 \cap R_2) = \frac{4}{10}$
- $\mathbb{P}(W_4 \mid R_1 \cap R_2 \cap W_3)$ denotes the probability of drawing a white ball fourth, after having drawn a red ball first followed by another red ball followed by a white ball. There are 9 balls remaining of which 3 are white; hence $\mathbb{P}(W_4 \mid R_1 \cap R_2 \cap R_3) = \frac{3}{9}$
- Hence, putting everything together,

$$\mathbb{P}(R_1 \cap R_2 \cap W_3 \cap W_4) = \frac{8}{12} \times \frac{7}{11} \times \frac{4}{10} \times \frac{3}{9} = \frac{28}{495}$$

Independence

Leadup

- Recall that $\mathbb{P}(A)$ represents our beliefs on an event A.
- Additionally, $\mathbb{P}(A \mid B)$ represents our updated beliefs on A, in the presence of B.
- What if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$? In other words, our beliefs about A are completely unchanged by B.
- That is, A and B are *unaffected* by each other... they are **independent** of each other!

Definition: Independence

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two events $A, B \in \mathcal{F}$, we say that A and B are **independent** (notated $A \perp B$) if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, or, equivalently, if $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

An equivalent condition for independence is $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Example

Suppose A and B are events with $\mathbb{P}(A) = 0.2,\ \mathbb{P}(B) = 0.3$, and $\mathbb{P}(A \cap B) = 0.1$. Are A and B independent?

- No, because $\mathbb{P}(A \cap B) = 0.1 \neq 0.2 \cdot 0.3 = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Independence of Multiple Events

Definition: Independence of n Events

We say that a sequence of events $A_1, \ldots, A_n$ are **independent** (or **mutually independent**) if, for *every* subsequence $A_{i_1}, \ldots, A_{i_k}$, with $2 \leq k \leq n$ and $1 \leq i_1 < i_1 < \cdots < i_k \leq n$, we have

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

Independence of 4 events:

- $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$
- $\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C)$
- $\mathbb{P}(A \cap D) = \mathbb{P}(A) \cdot \mathbb{P}(D)$
- $\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C)$
- $\mathbb{P}(B \cap D) = \mathbb{P}(B) \cdot \mathbb{P}(D)$
- $\mathbb{P}(C \cap D) = \mathbb{P}(C) \cdot \mathbb{P}(D)$

two-way intersections

three-way intersections

- $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$
- $\mathbb{P}(A \cap B \cap D) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(D)$
- $\mathbb{P}(A \cap C \cap D) = \mathbb{P}(A) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D)$
- $\mathbb{P}(B \cap C \cap D) = \mathbb{P}(B) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D)$

- $\mathbb{P}(A \cap B \cap C \cap D) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D)$

four-way intersections

Example

Suppose I toss a fair coin three times, and I define G_i to be the event "the i^{th} toss landed tails" (for $i = 1, 2, 3$).

Show that G_1, G_2, and G_3 are independent.

Extensions/Modifications of Independence

Pairwise Independence

- Independence is a very strong condition!
- There exists a weaker form of independence:

Definition: Pairwise Independence

A sequence of events $A_1, A_2, \cdots$ is said to be **pairwise independent** if $A_i \perp A_j$ for any $i \neq j$.

- Note that independence implied pairwise independence, but not vice-versa.

Example

Suppose (again) I toss a fair coin three times, and define events

$$A := \{\text{I observe exactly one tails among the first two coinflips}\}$$
$$B := \{\text{I observe exactly one tails among the last two coinflips}\}$$
$$C := \{\text{I observe exactly one tails among the first and third coinflip}\}$$

Are A, B, C independent? Are they pairwise independent?

- At the beginning of this lecture, we defined a new probability measure; that of conditional probability. It makes sense to posit a notion of independence that utilizes the conditional probability measure:

Definition: Conditional Independence

Let $A_1, \cdots, A_n$ be events and B be an event with $\mathbb{P}(B) \neq 0$. Then we say that the events $A_1, \cdots, A_n$ are **conditionally independent, given $\boldsymbol{B}$**, if for any $k \in \{2, \cdots, k\}$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k} \mid B) = \mathbb{P}(A_{i_1} \mid B) \times \cdots \times \mathbb{P}(A_{i_k} \mid B)$$

Example (ASV 2.38)

Suppose 90% of coins in circulation are fair and 10% are biased coins that give tails with probability 3/5. I have a random coin and I flip it twice. Denote by A_1 the event that the first flip yields tails and by A_2 the event that the second flip yields tails.

- Let F denote "coin is fair" and B denote "coin is biased."
- It seems reasonable enough to assume that *for a given coin* the probability of tails does not change between the first and second flip. In other words,

$$\mathbb{P}(A_1 \mid F) = \mathbb{P}(A_2 \mid F) = \frac{1}{2}; \quad \mathbb{P}(A_1 \mid B) = \mathbb{P}(A_2 \mid B) = \frac{3}{5}$$

- Then, by the Law of Total Probability,

$$\mathbb{P}(A_i) = \mathbb{P}(A_i \mid F)\mathbb{P}(F) + \mathbb{P}(A_i \mid B)\mathbb{P}(B) = \frac{1}{2} \cdot \frac{9}{10} + \frac{3}{4} \cdot \frac{1}{10} = \frac{51}{100}$$

for $i = 1, 2$

Example (ASV 2.38)

Suppose 90% of coins in circulation are fair and 10% are biased coins that give tails with probability 3/5. I have a random coin and I flip it twice. Denote by A_1 the event that the first flip yields tails and by A_2 the event that the second flip yields tails.

- It also seems reasonable enough to assume that successive flips *of a given coin* are independent. In other words, we assume we have conditional independence:

$$\mathbb{P}(A_1 \cap A_2 \mid F) = \mathbb{P}(A_1 \mid F)\mathbb{P}(A_2 \mid F); \quad \mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

- Hence, by the Law of Total Probability,

$$\begin{aligned}
\mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(A_1 \cap A_2 \mid F)\mathbb{P}(F) + \mathbb{P}(A_1 \cap A_2 \mid B)\mathbb{P}(B) \\
&= \mathbb{P}(A_1 \mid F)\mathbb{P}(A_2 \mid F)\mathbb{P}(F) + \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)\mathbb{P}(B) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{9}{10} + \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{1}{10} = \frac{261}{1000}
\end{aligned}$$

Example (ASV 2.38)

Suppose 90% of coins in circulation are fair and 10% are biased coins that give tails with probability 3/5. I have a random coin and I flip it twice. Denote by A_1 the event that the first flip yields tails and by A_2 the event that the second flip yields tails.

- So, we have

$$\mathbb{P}(A_1) \times \mathbb{P}(A_2) = \left(\frac{51}{100}\right)^2$$

$$\mathbb{P}(A_1 \cap A_2) = \frac{261}{1000}$$

  which shows that A_1 and A_2 are **not** independent, despite the fact that they are (by our construction) conditionally independent given the fairness or biasedness of the coin.

- Intuitively, this is because the first flip gives us information about the coin we hold (i.e. with regards to whether or not it is fair). This information will clearly alter our believes about the second flip.

The "Other Direction"

- In our previous example, we saw that two conditionally independent events may not be mutually independent.
- Unsurprisingly, this shows that we can sometimes "inject" independence by conditioning!
- In other words, we can start out with two dependent events, but find a specific event B. such that we have conditional independent given B.
- If you're curious, I direct you to Example 2.40 from ASV [I'll try and post a copy of this example online shortly]