

2: Conditional Probability

Spring 2021

Matthew Blackwell

Gov 2002 (Harvard)

Roadmap

1. Conditional Probability
2. Bayes's Rule
3. Independence

1/ Conditional Probability

Conditional probability

- **Conditional probability:** if we know that B has occurred, what is the probability of A ?
 - Conditioning our analysis on B having occurred.
- Examples:
 - What is probability of two states going to war **if** they are both democracies?
 - What is the probability of a judge ruling in a pro-choice direction **conditional** on having daughters?
 - What is the probability that there will be a coup in a country **conditional** on having a presidential system?
- Conditional probability is the cornerstone of quantitative social science.

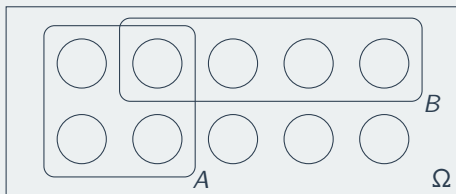
Conditional Probability definition

- Definition: If $\mathbb{P}(B) > 0$ then the **conditional probability** of A given B is

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

- How often A and B occur divided by how often B occurs.
- **WARNING!** $\mathbb{P}(A|B)$ does **not**, in general, equal $\mathbb{P}(B|A)$.
 - $\mathbb{P}(\text{smart} \mid \text{in gov 2002})$ is high
 - $\mathbb{P}(\text{in gov 2002} \mid \text{smart})$ is low.
 - There are many many smart people who are not in this class!
 - Also known as the **prosecutor's fallacy**

Intuition



Examples

$A = \{\text{you get an A grade}\}$ $B = \{\text{everyone gets an A grade}\}$

- If B occurs then A must also occur, so $\Pr(A|B) = 1$.
 - Does this mean that $\Pr(B|A) = 1$ as well?
- Now let $A = \{\text{you get a B grade}\}$.
 - The intersection $A \cap B = \emptyset$, so that $\Pr(A|B) = 0$.
 - Intuitively, it's because B occurring precludes A from occurring.

U.S. Senate example

	Democrats	Republicans	Independents	Total
Men	39	42	2	83
Women	12	5	0	17
Total	51	47	2	100

- Choose one senator at random from this population
- What is the probability that a randomly selected Republican is a Woman:

$$\bullet \mathbb{P}(\text{Woman} \mid \text{Republican}) = \frac{\mathbb{P}(\text{Woman} \cap \text{Republican})}{\mathbb{P}(\text{Republican})} = \frac{5/100}{47/100} = \frac{5}{47} = 0.106$$

- Choose two senators at random:
 - $\mathbb{P}(2 \text{ women} \mid \text{one draw is a woman})?$
 - $\mathbb{P}(2 \text{ women} \mid \text{one draw is a Liz Warren})?$

Conditional probabilities are probabilities

- Condition probabilities $\mathbb{P}(A|B)$ are valid probability functions:
 1. $\mathbb{P}(A|B) \geq 0$
 2. $\mathbb{P}(\Omega|B) = 1$
 3. If A_1 and A_2 are disjoint, then $\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B)$
- \rightsquigarrow rules of probability apply to left-hand side of conditioning bar (A)
 - All probabilities **normalized** to event B , $\mathbb{P}(B|B) = 1$.
- Not for right-hand side, so even if B and C are disjoint,

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

Joint probabilities from conditionals

- **Joint probabilities:** probability of two events occurring (intersections)
 - Often replace \cap with commas: $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A, B, C)$
- Definition of conditional prob. implies:

$$\mathbb{P}(A \cap B) \equiv \mathbb{P}(A, B) = \mathbb{P}(B)\mathbb{P}(A | B) = \mathbb{P}(A)\mathbb{P}(B | A)$$

- What about three events?

$$\mathbb{P}(A, B, C) = \mathbb{P}(A)\mathbb{P}(B | A)\mathbb{P}(C | A, B)$$

- Generalize to the intersection of N events:

$$\mathbb{P}(A_1, \dots, A_N) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1, A_2) \cdots \mathbb{P}(A_N | A_1, \dots, A_{N-1})$$

Joint probabilities, example

- Draw three cards at random from a deck without replacement.
- What's the probability that we draw three Aces?

$$\mathbb{P}(\text{Ace}_1 \cap \text{Ace}_2 \cap \text{Ace}_3) = \mathbb{P}(\text{Ace}_1)\mathbb{P}(\text{Ace}_2 \mid \text{Ace}_1)\mathbb{P}(\text{Ace}_3 \mid \text{Ace}_2 \cap \text{Ace}_1)$$

- What are these probabilities?
 - 4 Aces to pick out of 52 cards $\rightsquigarrow \mathbb{P}(\text{Ace}_1) = \frac{4}{52}$
 - 3 Aces left in the 51 remaining cards $\rightsquigarrow \mathbb{P}(\text{Ace}_2 \mid \text{Ace}_1) = \frac{3}{51}$
 - 2 Aces left in the 50 remaining cards $\rightsquigarrow \mathbb{P}(\text{Ace}_3 \mid \text{Ace}_2 \cap \text{Ace}_1) = \frac{2}{50}$
- Thus, $\mathbb{P}(\text{Ace}_1 \cap \text{Ace}_2 \cap \text{Ace}_3) = \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} = 0.00018$

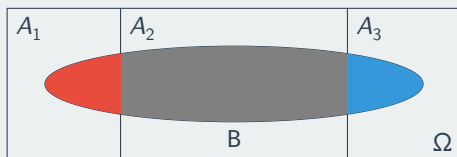
Probability of war resolution

- Suppose we observed country-dyads over 3 years
- In each year the dyad can be at war (W_t) or at peace (P_t).
- What's the probability that a war starts in year 1 ends after 2 years?

$$\mathbb{P}(W_1, W_2, P_3) = \mathbb{P}(W_1)\mathbb{P}(W_2 | W_1)\mathbb{P}(P_3 | W_1, W_2)$$

- **Actual Research QuestionTM**: modeling the continuation probability of war, $\mathbb{P}(W_2 | W_1)$ and the probability of conflict resolution, $\mathbb{P}(P_3 | W_1, W_2)$.

Law of Total Probability



- Often we only have disaggregated probabilities.
 - B = sampling a Trump supporter from either Cambridge or Somerville.
 - We know the prop. of Trump supporters in each city from precinct data.
 - How to calculate the overall probability of B ?
- A **partition** is a set of mutually disjoint events whose union is Ω .
- The **law of total probability** (LTP) states if A_1, \dots, A_k is a partition:

$$\mathbb{P}(B) = \sum_{i=1}^k \mathbb{P}(B | A_i) \mathbb{P}(A_i)$$

- Overall probability = weighted sum of within-partition probabilities
- Weights are the probability of the particular partition

A mixture of cities

- Randomly drawing voters from either Cambridge or Somerville:
 - Camb. had 50k voters and Somer. had around 40k, so:
 - $\Pr(\text{Camb.}) = 0.56$ and so $\Pr(\text{Somer.}) = 0.44$
- The state provides the following election results for each city:
 - $\Pr(\text{Trump}|\text{Camb.}) = 0.066$
 - $\Pr(\text{Trump}|\text{Somer.}) = 0.103$
- To get the overall turnout rate, $\mathbb{P}(\text{Trump})$, we can apply the LTP:

$$\begin{aligned}\Pr(\text{Trump}) &= \Pr(\text{Trump}|\text{Camb.}) \Pr(\text{Camb.}) + \Pr(\text{Trump}|\text{Somer.}) \Pr(\text{Somer.}) \\ &= 0.066 \times 0.56 + 0.103 \times 0.44 \\ &= 0.082\end{aligned}$$

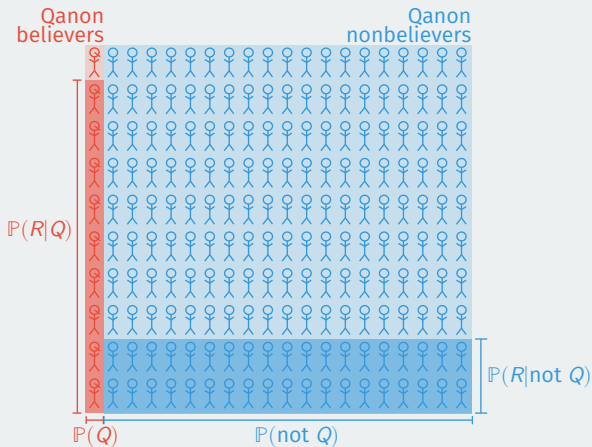
2/ Bayes's Rule



You meet a man named Steve and he tells you that he is a Republican. You have been interested in meeting someone who believes in the QAnon conspiracy theory. Given what you know about Steve, would you guess that he believes in QAnon or not?

- Common response: probably believes in QAnon since believers tend to be Republicans.
- **Base rate fallacy:** ignores how uncommon QAnon believers are!

Visualizing QAnon support



$$\text{Chance a random Republican believes QAnon} = \frac{P(R|Q)P(Q)}{P(R|Q)P(Q) + P(R|\text{not } Q)P(\text{not } Q)}$$

Bayes' rule



- Reverend Thomas Bayes (1701–61): English minister and statistician
- **Bayes' rule:** if $\mathbb{P}(B) > 0$, then:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B | A)\mathbb{P}(A)}{\mathbb{P}(B | A)\mathbb{P}(A) + \mathbb{P}(B | A^c)\mathbb{P}(A^c)}$$

Why is Bayes' rule useful?

- What is the probability of some hypothesis given some evidence?
 - $\mathbb{P}(\text{QAnon} \mid \text{Republican})$?
- Often easier to know probability of evidence given hypothesis.
 - $\mathbb{P}(\text{Republican} \mid \text{QAnon})$
- Combine this with the **prior probability** of the hypothesis.
 - Prior: $\mathbb{P}(\text{QAnon})$
 - **Posterior**: $\mathbb{P}(\text{QAnon} \mid \text{Republican})$
- Applying Bayes' rule is often called **updating the prior**.
 - $\mathbb{P}(\text{QAnon}) \rightsquigarrow \mathbb{P}(\text{QAnon} \mid \text{Republican})$
 - How does the evidence change the chance of the hypothesis being true?

Uses of Bayes' rule

- Medical testing:
 - Want to know: $\mathbb{P}(\text{Disease} \mid \text{Test Positive})$
 - Have: $\mathbb{P}(\text{Test Positive} \mid \text{Disease})$ and $\mathbb{P}(\text{Disease})$
- Predicting traits from names:
 - Want to know: $\mathbb{P}(\text{African American} \mid \text{Last Name})$
 - Have: $\mathbb{P}(\text{Last Name} \mid \text{African American})$ and $\mathbb{P}(\text{African American})$
- Spam filtering:
 - Want to know: $\mathbb{P}(\text{Spam} \mid \text{Email text})$
 - Have: $\mathbb{P}(\text{Email text} \mid \text{Spam})$ and $\mathbb{P}(\text{Spam})$

Medical tests

- Suppose you go and get a COVID-19 test and it comes back positive!
 - Let a positive test be PT .
- What's the probability you actually have COVID-19?
 - Let having COVID be labeled C .
 - Question: What is $\mathbb{P}(C | PT)$?
- Components for calculating Bayes' rule:
 - $\mathbb{P}(PT|C) = 0.8$: true positive rate
 - $\mathbb{P}(PT | C^c) = 0.005$: false positive rate
 - $\mathbb{P}(C) = 0.007$ rough prevalence of active COVID cases.

Applying Bayes' rule to COVID tests

- Use the law of total probability to get the denominator:

$$\begin{aligned}\mathbb{P}(PT) &= \mathbb{P}(PT | C)\mathbb{P}(C) + \mathbb{P}(PT|C^c)\mathbb{P}(C^c) \\ &= (0.8 \times 0.007) + (0.005 \times 0.993) \\ &= 0.011\end{aligned}$$

- Now plug in all values to Bayes' rule:

$$\mathbb{P}(C | PT) = \frac{\mathbb{P}(PT | C)\mathbb{P}(C)}{\mathbb{P}(PT)} = \frac{0.8 \times 0.007}{0.011} \approx 0.53$$

- If false positive rate goes up to 1% $\rightsquigarrow \mathbb{P}(C | PT) \approx 0.36$

3/ Independence

Independence

- Heart of Bayes's rule: knowing B occurs often changes probability of A .
 - What if B provides no information? \rightsquigarrow independence
- Two events A and B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
 - Sometimes written as $A \perp\!\!\!\perp B$
 - **Symmetric:** $A \perp\!\!\!\perp B$ equivalent to $B \perp\!\!\!\perp A$
 - Events that are not independent are **dependent**.

- Important consequence: if $A \perp\!\!\!\perp B$ and $\mathbb{P}(B) > 0$ then:

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

- Knowing B occurs has no impact on the probability of A .
 - Works other way too: if $\mathbb{P}(A) > 0$ and $A \perp\!\!\!\perp B \rightsquigarrow \mathbb{P}(B | A) = \mathbb{P}(B)$.
- Common misunderstanding: **independent is different than disjoint!**
 - Mutually exclusive events provide information!

Independence example

- If we have a gathering of size n drawn randomly from population of MA with current COVID infection rate of 1.37%, what's the probability someone in attendance is infected?
- When seeing “prob. of at least one” \rightsquigarrow work with complement:

$$\begin{aligned} & \mathbb{P}(\text{At least one COVID case at gathering}) \\ &= 1 - \mathbb{P}(\text{No COVID cases at gathering}) \end{aligned}$$

Independence and random sampling

- How we draw the random sample matters:
 - Sample $n > 1$ with replacement \rightsquigarrow independent events
 - Sample $n > 1$ without replacement \rightsquigarrow dependent events
- Sampling with replacement n for gathering:

$$\begin{aligned}\mathbb{P}(\text{No COVID cases at gathering}) &= \mathbb{P}(\text{No COVID for Person 1} \cap \dots \cap \text{No COVID for Person } n) \\ &= \mathbb{P}(\text{No COVID for Person 1}) \dots \mathbb{P}(\text{No COVID for Person } n) \\ &= (1 - 0.007)^n\end{aligned}$$

- Using the complement:

$$\mathbb{P}(\text{At least one COVID case at gathering}) = 1 - (1 - 0.007)^n$$

- $n = 5 \rightsquigarrow$ prob of 0.035
- $n = 100 \rightsquigarrow$ prob of 0.5

Conditional independence

- A and B are **conditionally independent** given E if

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

- Massively important in statistics and causal inference.
- **Warning:** independence \neq conditional independence.
 - Cond. ind. $\not\Rightarrow$ ind.: flipping a coin with unknown bias.
 - Ind. $\not\Rightarrow$ cond. ind.: test scores, athletics, and college admission.