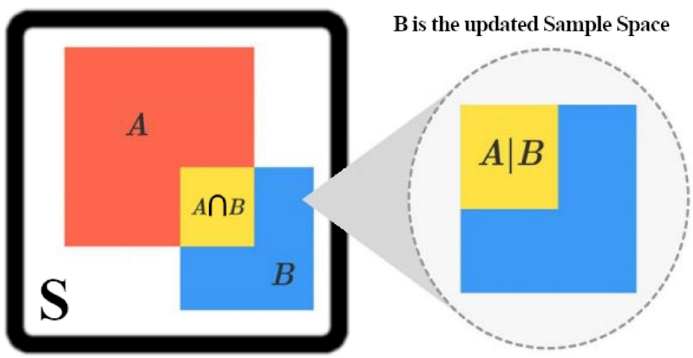
Probability Theory

Textbook: *Introduction to Probability* by Blitzstein and Hwang

Previous Lecture

- Conditional Probability

§2.2 - Definition & Intuition

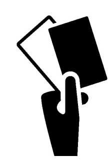


$P(A)$ being updated to $P(A|B)$ if we know B occured.

Def (**Conditional Prob**): If A, B are events w/$P(B) > 0$, then the conditional prob of A given B, denoted $P(A|B)$ is: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

B is the evidence we observe (or treat as a given).

We call $P(A)$ the *prior probability* of A and $P(A|B)$ the *posterior probability* of A ("prior" means before updating based on evidence B, and "posterior" means after updating based on evidence B).



Ex (**Two Cards**): Two cards are drawn randomly from a shuffled deck, one at a time w/out replacement. Let A be event that the first card is a heart, and B be event that the second card is red. Find $P(A|B)$ and $P(B|A)$.

Solution: By definition of conditional prob, we need $P(A \cap B)$, $P(A)$, and $P(B)$.

By naive definition and multiplication rule, $P(A \cap B) = \frac{13}{52} \cdot \frac{25}{51} = \frac{25}{204}$, since a favorable outcome is choosing one of 13 hearts, then any of the remaining 25 red cards.

Note: $P(A) = \frac{1}{4}$ since the four suits are equally likely.

And $P(B) = \frac{26}{52} \cdot \frac{51}{51} = \frac{1}{2}$, since there are 26 favorable possibilities for the second card, and for each of those, the first card can be any other card (recall chronological order isn't needed for multiplication rule).

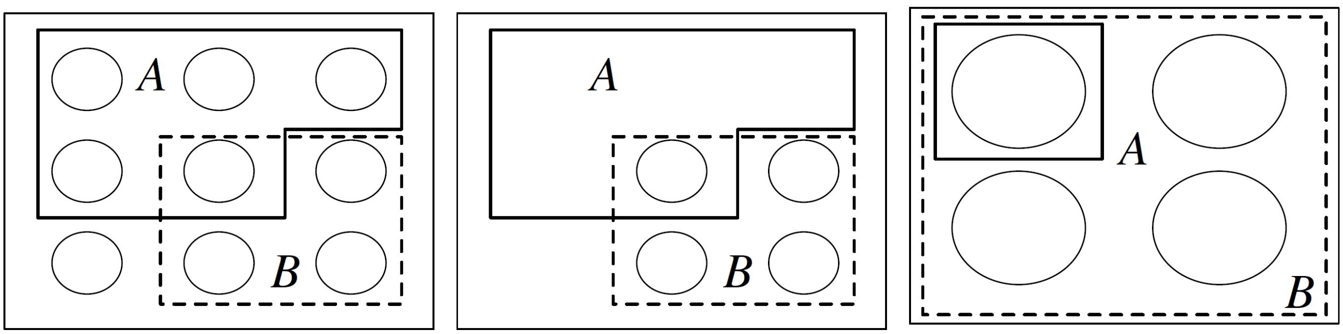
A neater way to see $P(B) = \frac{1}{2}$ is by symmetry. Note that before doing the experiment, the second card is equally likely to be any card in the deck.

We now have the ingredients for conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{25/204}{1/2} = \frac{25}{102} \approx 0.245,$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{25/204}{1/4} = \frac{25}{51} \approx 0.490. \qquad \square$$

Intuition for P(A|B)



(a) (b) (c)

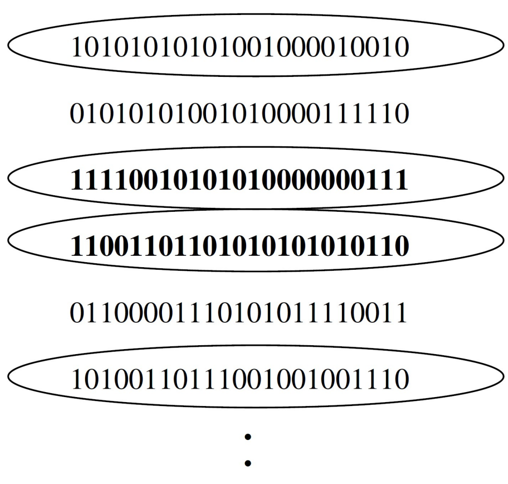
From left to right:

(a) Events A, B are subsets of sample space.

(b) Because we know B occurred, get rid of outcomes in B^c.

(c) In the restricted sample space, renormalize so total mass is still 1.

Intuition (**Frequentist Interpretation**). Frequentist interpretation of prob is based on relative frequency over a large # of repeated trials. Imagine repeating our experiment many times, generating a long list of observed outcomes. The conditional prob of A given B can then be thought of in a natural way: it's the *fraction of times that A occurs, when restricting our attention to trials where B occurs*.



B events are circled, A events are bolded.

$P(A|B)$ is the long-run relative frequency of the repetitions where A occurs, within subset of repetitions where B occurs.

Harvard Video: youtube.com/watch?v=P7NE4WF8j-Q&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=5

(FYI: This is sufficient to complete HW 2)

---

§2.3 - Bayes' Rule & the Law of Total Probability (LOTP)

Below we learn Baye's rule, which is very useful in computing conditional probability.
But first, let's build a few building blocks underpinning the rule.

Let A and B be events.

From definition of conditional prob ($P(A|B) = \frac{P(A \cap B)}{P(B)}$), if we multiply both sides by $P(B)$, we get $P(A|B)P(B) = P(A \cap B)$.

Switching the roles of A and B in the definition gives us a different relation: $P(B|A) = \frac{P(A \cap B)}{P(A)}$.
If we then multiply both sides by $P(A)$, we get $P(B|A)P(A) = P(A \cap B)$.

Notice that the RHS of these two equations are the same, in other words: $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$.

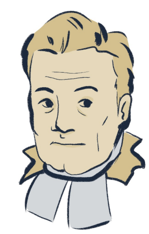
Thm (**Prob of Intersection of Two Events**): For any events A and B w/positive probs, $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$.

This result might seem fairly simple to calculate, but ends up being very powerful. We can generalize it to any # of rvs:

❗ **Notation**: $P(A,B)$ means $P(A \cap B)$. Similarly, $P(B|A_1, A_2, \ldots, A_n) = P(B|A_1 \cap A_2 \cap \ldots \cap A_n)$.

Thm (**Prob of Intersection of n Events**): For any events $A_1, \ldots, A_n$ with $P(A_1, \ldots, A_{n-1}) > 0$,
$P(A_1, \ldots, A_n) = P(A_n|A_{n-1}, \ldots, A_1) \ldots P(A_3|A_2, A_1)P(A_2|A_1)P(A_1)$.

Bayes' Rule



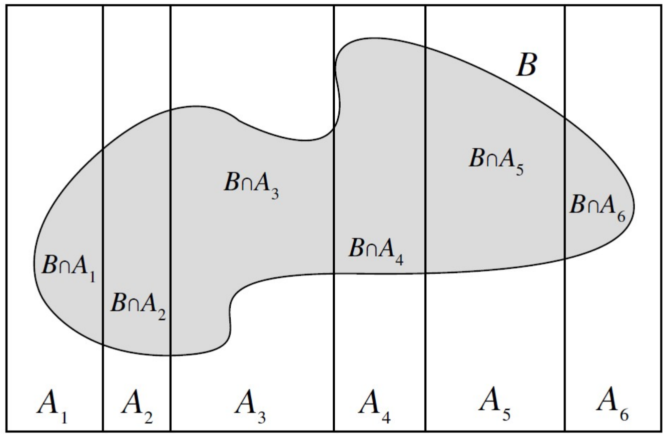
Thm (**Bayes' Rule**): $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

Proof: "Prob of Intersection of Two Events," then divide by $P(B)$. ■

Extremely powerful tool for calculating real-world probs, as we'll see in example below.
Often times when using this, we also need another tool called the **Law of Total Prob** (**LOTP**).

Thm (**Law of Total Prob**, **LOTP**): Let $A_1, \ldots, A_n$ be a partition of S
(i.e., A_i are disjoint and their union is S), with $P(A_i) > 0$ for all i. Then for any B, $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$.

Proof: Since A_i form a partition of S, then $B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)$.



By Disjoint Axiom, we add the probs to get: $P(B) = P(B \cap A_1) + P(B \cap A_2) + \ldots + P(B \cap A_n)$.

Now we apply "Prob of Intersection of Two Events" to each $P(B \cap A_i)$:
$P(B) = P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)$. ■

Ex (**Random Coin**): You have one fair coin which lands Heads with prob $\frac{1}{2}$,
and one biased coin which lands Heads with prob $\frac{3}{4}$.
You pick one coin at random and flip it three times. It lands Heads all three times.
What's the prob the coin you picked is the fair one?

Solution: Let A be the event that the chosen coin lands Heads three times.
Let F be that we picked the fair coin.
We are interested in $P(F|A)$, but it's easier to find $P(A|F)$ and $P(A|F^c)$.

This suggests using Bayes' rule and LOTP.

So we have: $P(F|A) = \frac{P(A|F)P(F)}{P(A)}$ (**Bayes' Rule**)

$= \frac{P(A|F)P(F)}{P(A|F)P(F)+P(A|F^c)P(F^c)}$ (**LOTP**)

$= \frac{\left(\frac{1}{2}\right)^3 \cdot \frac{1}{2}}{\left(\frac{1}{2}\right)^3 \cdot \frac{1}{2} + \left(\frac{3}{4}\right)^3 \cdot \frac{1}{2}} \approx 0.23.$ □

Note: Before flipping, we thought: $P(F) = P(F^c) = \frac{1}{2}$.
Upon observing HHH, it becomes more likely we've chosen the biased coin: $P(F|A) \approx 0.23$.

Activity 5

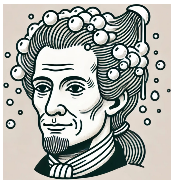
Harvard Video: youtube.com/watch?v=JzDvVgNDxo8&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=6

§2.4 - Conditional Probs are Probs

When we condition on E, we put ourselves in the universe where we know E occurred.
Within this universe, the laws of prob operate just as before.
Conditional prob satisfies all axioms/properties of prob!

- Conditional probs are between 0 and 1.
- $P(S|E) = 1$, $P(\emptyset|E) = 0$.
- If $A_1, A_2, \ldots$ are disjoint, then: $P(\cup_{j=1}^{\infty} A_j|E) = \sum_{j=1}^{\infty} P(A_j|E)$.
- $P(A^c|E) = 1 - P(A|E)$.
- Inclusion-Exclusion: $P(A \cup B|E) = P(A|E) + P(B|E) - P(A \cap B|E)$.

Thm (**Bayes' Rule w/Extra Conditioning**). Provided that $P(A \cap E) > 0$ and $P(B \cap E) > 0$, we have: $P(A|B,E) = \frac{P(B|A,E)P(A|E)}{P(B|E)}$.



Thm (**LOTP w/extra conditioning**). Let $A_1, \ldots, A_n$ be a partition of S.
Provided $P(A_i \cap E) > 0$ for all i, we have: $P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E)$.

What did we learn?

- Conditional Prob: Prior prob, Posterior prob
- Bayes' Rule
- LOTP
- Conditional Probs are Probs