2. Conditional Probability and Bayes' Rule

$$P(A|B) \neq P(B|A)$$

How should we update our beliefs in light of the evidence we observe? *Bayes' rule* is an extremely useful theorem that helps us perform such updates.

Together, Bayes' rule and the law of total probability can be used to solve a very wide variety of problems.

2.1 The importance of thinking conditionally

A useful perspective is that all probabilities are *conditional*

A and B events, $P(B) > 0$, then *conditional probability* of A given B:

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

A is the event whose uncertainty we want to update, and B is the evidence we observe.

$P(A)$ is called *prior probability* of A.

$P(A|B)$ is called *posterior probability* of A.

$P(A|B)$ is the probability of A given the evidence B, **not** the probability of some weird entity called $A|B$.

Note:

1. It's extremely important to be careful about which events to put on which side of the conditioning bar. Confusing these two quantities is called the *prosecutor's fallacy*.

2. Both $P(A|B)$ and $P(B|A)$ make sense. We are considering what **information** observing one event provides about another event, not whether one event **causes** another.

Frequentist interpretation:

The conditional probability of A given B: it is the fraction of times that A occurs, restricting attention to the trials where B occurs.

2.3 Bayes' rule and the law of total probability

Just move the denominator in the definition to the other side of the equation:

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

It often turns out to be possible to find conditional probabilities without going back to the definition.

Same for n events:

$\forall A_1, \ldots A_n$

$P(A_i, A_2, \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)\ldots P(A_n|A_1 \ldots A_{n-1})$

Then **Bayes' rule:**

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

The **law of total probability (LOTP)** relates conditional probability to unconditional probability:

$A_1, \ldots, A_n$ a partition of the sample space S (A_i disjoint, their union is S), $P(A_i) > 0 \forall i$ then:

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

Proof:

Decompsition of B:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)$$

then

$$P(B) = P(B \cap A_1) + \ldots + P(B \cap A_n) = P(B|A_1)P(A_1) \ldots P(B|A_n)P(A_n)$$

The choice of how to divide up the sample space is crucial!

2.4 Conditional probabilities are probabilities

When we condition on an event E, we update our beliefs to be consistent with this knowledge, effectively putting ourselves in a **universe where we know that E occurred**.

Bayes' rule with extra conditioning

$P(A \cap E) > 0$ and $P(B \cap E) > 0$

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

LOTP with extra conditioning

$A_1, \ldots, A_n$ is a partition of S, $P(A_i \cap E) > 0 \forall i$

$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E)$$

2.5 Independence of events

A and B are independent if

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

if $P(A) > 0$ and $P(B) > 0$,

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

Two events are independent if we can obtain the probability of their intersection by multiplying their individual probabilities.

If A is independent of B, then B is independent of A.

Independence is not disjointness!!!

If A and B are disjoint, $P(A \cap B) = 0$

Disjoint events can be independent only if $P(A) = 0$ or $P(B) = 0$.

Knowing that A occurs tells us that B definitely did not occur.

Independence of three or more events!

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

$$P(A \cap C) = P(A)P(C),$$

$$P(B \cap C) = P(B)P(C).$$

If these three conditions are hold: A, B, C are *pairwise independent* $\neq$ *independence*.

For example, it is possible that A, B occurred would give us knowledge about C
.

For full independence, 4th condition:

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Conditional independence:

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

Conditional independence $\neq$ *independence*

(example with biased and fair coin)

Independence $\neq$ *conditional independence*

2.6 Conditioning as a problem-solving tool

2.6.1 Strategy: condition on what you wish you knew

Example: Monty Hall

Without loss of generality, we can assume the contestant picked door 1

C_i the event that the car is behind i-th door

$$P(getcar) = P(getcar|C_1)\frac{1}{3} + P(getcar|C_2)\frac{1}{3} + P(getcar|C_3)\frac{1}{3}$$

Switching strategy: if the car behind 1, switching will fail $P(getcar|C_1) = 0$.

If the car behind 2 or 3, because Monty always reveals a goat, switching will succeed. Thus,

$$P(getcar) = 0\frac{1}{3} + 1\frac{1}{3} + 1\frac{1}{3} = \frac{2}{3}$$

So *unconditional probability* of success is $2/3$ (when following the switching strategy), let's also show that the *conditional probability* of success for switching, given the information that Monty provides, is also $2/3$.

Let M_j be the event that Monty opens j-th door. Then:

$$P(getcar) = P(getcar|M_2)P(M_2) + P(getcar|M_3)P(M_3),$$

by symmetry, $P(M_2) = P(M_3) = 1/3$, $P(getcar|M_2) = P(getcar|M_3)$

so $P(getcar|M_2) = P(getcar|M_3) = 2/3$

Bayes' rule:

Suppose that Monty opens door 2,

$$P(C_1|M_2) = \frac{P(M_2|C_1)P(C_1)}{P(M_2)} = \frac{(1/2)(1/3)}{1/2} = \frac{1}{3}$$

So there is 1/3 chance that original choice was correct, which means that 2/3 chance that switching strategy was better.

2.6.2 Strategy: condition on the first step

Example: Branching process

A single amoeba, Bobo, lives in a pond. At minute, Bobo will die, or do nothing or split to two amoebas with probabilities 1/3.

What is the probability that the amoeba population will eventually die out?

D is the evenr that the population dies out, B_i is the event that Bobo will turn into $i = 0, 1, 2$ amoebas.

$$P(D) = \frac{P(D|B_0)}{3} + \frac{P(D|B_1)}{3} + \frac{P(D|B_2)}{3}$$

$P(D|B_0) = 1, P(D|B_0) = P(D),$ (we're back to where we started)
$P(D|B_2) = P(D)^2$, (two independent original problems). So:

$$P(D) = 1/3 + P(D)/3 + P(D)^2/3$$

Solving: $P(D) = 1$

Example: Gambler's ruin

A and B make a sequence of $1 bets. A has probability p of winning, B has $q = 1 - p$. A starts with $\$i$ B starts with $\$N - i$

Game ends when A or B is ruined.

What is the probability that A wins the game?

After the first step, it's exactly the same game, except that A's wealth is now either $i + 1$ or $i - 1$.

p_i is the probability that A wins, given that A starts with i dollars. W is the event that A wins.

By LOTP,

$$p_i = P(W|A\ starts\ at\ i,\ wins)p + P(W|A\ starts\ at\ i,\ loses)q =$$

$$= P(W|A\ starts\ at\ i+1)p + P(W|A\ starts\ at\ i-1)q =$$

$$= p_{i+1}p + p_{i-1}q.$$

This is true $\forall i \in [1, N-1]$, boundary conditions $p_0 = 0, p_N = 1$

Solving this as *difference equation* to obtain p_i:

$$if\ p \neq 1/2,\ p_i = \frac{1-(q/p)^i}{1-(q/p)^N},$$

$$if\ p = 1/2,\ p_i = \frac{i}{N}.$$

Example: Simpson's paradox

For events A, B, and C, we say that we have a *Simpson's paradox* if:

$$P(A|B,C) < P(A|B^C,C)$$

$$P(A|B,C^C) < P(A|B^C,C^C)$$

but

$$P(A|B) > P(A|B^C).$$

Aggregation across different types of surgeries presents a misleading picture of the doctors' abilities because we lose the information about which doctor tends to perform which type of surgery.