Justifying Conditioning

Suppose that after learning B, we update Pr to $\Pr_B$. Some reasonable assumptions:

C1. $\Pr_B$ is a probability distribution

C2. $\Pr_B(B) = 1$

C3. If $C_1, C_2 \subseteq B$, then $\Pr_B(C_1)/\Pr_B(C_2) = \Pr(C_1)/\Pr(C_2)$.

- The relative probability of subsets of B doesn't change after learning B.

Theorem: C1–C3 force conditioning:

$$\Pr_B(A) = \Pr(A \cap B)/\Pr(B).$$

Proof: By C1, $\Pr_B(B) + \Pr_B(\overline{B}) = 1$.
By C2, $\Pr_B(B) = 1$, so $\Pr_B(\overline{B}) = 0$.

General property of probability:
If $C \subseteq C'$, then $\Pr_B(C) \leq \Pr_B(C')$ (Why?)
So if $C \subseteq \overline{B}$, then $\Pr_B(C) = 0$

By C1,

$$\Pr_B(A) = \Pr_B(A \cap B) + \Pr_B(A \cap \overline{B}) = \Pr_B(A \cap B).$$

Since $A \cap B \subseteq B$, by C3,

$$\Pr_B(A \cap B)/\Pr_B(B) = \Pr(A \cap B)/\Pr(B).$$

By C3, $\Pr_B(B) = 1$, so $\Pr_B(A \cap B) = \Pr(A \cap B)/\Pr(B)$.

The Second-Child Problem

Suppose that any child is equally likely to be male or female, and that the sex of any one child is independent of the sex of the other. You have an acquaintance and you know he has two children, but you don't know their sexes. Consider the following four cases:

1. You visit the acquaintance, and a boy walks into the room. The acquaintance says "That's my older child."
2. You visit the acquaintance, and a boy walks into the room. The acquaintance says "That's one of my children."
3. The acquaintance lives in a culture, where male children are always introduced first, in descending order of age, and then females are introduced. You visit the acquaintance, who says "Let me introduce you to my children." Then he calls "John [a boy], come here!"
4. You go to a parent-teacher meeting. The principal asks everyone who has at least one son to raise their hands. Your acquaintance does so.

In each case, what is the probability that the acquaintance's second child is a boy?

- The problem is to get the right sample space

Independence

Intuitively, events A and B are independent if they have no effect on each other.

This means that observing A should have no effect on the likelihood we ascribe to B, and similarly, observing B should have no effect on the likelihood we ascribe to A.

Thus, if $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$ and A is independent of B, we would expect

$$\Pr(B|A) = \Pr(B) \text{ and } \Pr(A|B) = \Pr(A).$$

Interestingly, one implies the other.

$\Pr(B|A) = \Pr(B)$ iff $\Pr(A \cap B)/\Pr(A) = \Pr(B)$ iff

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B).$$

Formally, we say A and B are *(probabilistically) independent* if

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B).$$

This definition makes sense even if $\Pr(A) = 0$ or $\Pr(B) = 0$.

Example

Alice has two coins, a fair one f and a loaded one l.

- l's probability of landing H is $p > 1/2$.

Alice picks f and flips it twice.

- What is the sample space?

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}.$$

- What is Pr?
- By symmetry this should be an equiprobable space.

Let $H_1 = \{(H, H), (H, T)\}$ and let $H_2 = \{(H, H), (T, H)\}$.

H_1 and H_2 are independent:

- $H_1 = \{(H, H), (H, T)\} \Rightarrow \Pr(H_1) = 2/4 = 1/2$.
- Similarly, $P(H_2) = 1/2$.
- $H_1 \cap H_2 = \{(H, H)\} \Rightarrow \Pr(H_1 \cap H_2) = 1/4$.
- So, $\Pr(H_1 \cap H_2) = \Pr(H_1) \cdot \Pr(H_2)$.

Alice next picks l and flips it twice.

- The sample space is the same as before:

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}.$$

- We now define Pr by *assuming* the flips are independent:
  - $\Pr\{(H,H)\} = \Pr(H_1 \cap H_2) := p^2$
  - $\Pr\{(H,T)\} = \Pr(H_1 \cap \bar{H}_2) := p(1-p)$
  - $\Pr\{(T,H)\} = \Pr(\bar{H}_1 \cap H_2) := (1-p)p$
  - $\Pr\{(T,T)\} = \Pr(\bar{H}_1 \cap \bar{H}_2) := (1-p)^2$.
- H_1 and H_2 are now independent by construction:

$$\begin{aligned}\Pr(H_1) &= \Pr\{(H,H),(H,T)\} = \\ &= p^2 + p(1-p) = p[p+(1-p)] = p.\end{aligned}$$

  Similarly, $\Pr(H_2) = \Pr\{(H,H),(T,H)\} = p$
  $\Pr(H_1 \cap H_2) = \Pr(H,H) = p^2$.

- For either coin, the two flips are independent.

What if Alice randomly picks a coin and flips it twice?

- What is the sample space?

$$\Omega = \{(f,(H,H)),(f,(H,T)),(f,(T,H)),(f,(T,T)),$$
$$(l,(H,H)),(l,(H,T)),(l,(T,H)),(l,(T,T))\}.$$

  The sample space has to specify which coin is picked!

- How do we construct Pr?
- E.g.: $\Pr(f, H, H)$ should be probability of getting the fair times the probability of getting heads with the fair coin: $1/2 \times 1/4$
- Follows from the following general result:

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

- So with F, L denoting the events f (respectively, l) was picked,

$$\begin{aligned}\Pr\{(f, (H, H))\} &= \Pr(F \cap (H_1 \cap H_2)) \\ &= \Pr(H_1 \cap H_2|F)\Pr(F) \\ &= 1/2 \cdot 1/2 \cdot 1/2.\end{aligned}$$

  Analogously, we have for example

$$\Pr\{(l, (H, T))\} = p(1-p) \cdot 1/2.$$

Are H_1 and H_2 independent now?

Claim. $\Pr(A) = \Pr(A|E)\Pr(E) + \Pr(A|\bar{E})\Pr(\bar{E})$

Proof. $A = (A \cap E) \cup (A \cap \bar{E})$, so

$$\Pr(A) = \Pr(A \cap E) + \Pr(A \cap \bar{E}).$$

$\Pr(H_1) = \Pr(H_1|F)\Pr(F) + \Pr(H_1|L)\Pr(L) = p/2 + 1/4.$

Similarly, $\Pr(H_2) = p/2 + 1/4$.

However,

$$\begin{aligned}
& \Pr(H_1 \cap H_2) \\
= \; & \Pr(H_1 \cap H_2|F)\Pr(F) + \Pr(H_1 \cap H_2|L)\Pr(L) \\
= \; & p^2/2 + 1/4 \cdot 1/2 \\
\neq \; & (p/2 + 1/4)^2 \\
= \; & \Pr(H_1) \cdot \Pr(H_2).
\end{aligned}$$

So H_1 and H_2 are dependent events.

Probability Trees

Suppose that the probability of rain tomorrow is .7. If it rains, then the probability that the game will be cancelled is .8; if it doesn't rain, then the probability that it will be cancelled is .1. What is the probability that the game will be played?

The situation can be described by a tree:

Similar trees can be used to describe

- Sequential decisions
- Randomized algorithms

Bayes' Theorem

Bayes Theorem: Let $A_1, \ldots, A_n$ be mutually exclusive and exhaustive events in a sample space S.

- That means $A_1 \cup \ldots \cup A_n = S$, and the A_i's are pairwise disjoint: $A_i \cap A_j = \emptyset$ if $i \neq j$.

Let B be any event in S. Then

$$\Pr(A_i|B) = \frac{\Pr(A_i)\Pr(B|A_i)}{\Sigma_{j=1}^n \Pr(A_j)\Pr(B|A_j)}.$$

Proof:

$$B = B \cap S = B \cap (\cup_{j=1}^n A_j) = \cup_{i=1}^n (B \cap A_j).$$

Therefore, $\Pr(B) = \Sigma_{j=1}^n \Pr(B \cap A_j)$.

Next, observe that $\Pr(B|A_i) = \Pr(A_i \cap B)/\Pr(A_i)$. Thus,

$$\Pr(A_i \cap B) = \Pr(B|A_i)\Pr(A_i).$$

Therefore,

$$\begin{aligned} \Pr(A_i|B) &= \frac{\Pr(A_i \cap B)}{\Pr(B)} \\ &= \frac{\Pr(B|A_i)\Pr(A_i)}{\Sigma_{j=1}^n \Pr(B \cap A_j)} \\ &= \frac{\Pr(B|A_i)\Pr(A_i)}{\Sigma_{j=1}^n \Pr(B|A_j)\Pr(A_j)} \end{aligned}$$

Example

In a certain county, 60% of registered voters are Republicans, 30% are Democrats, and 10% are Independents. 40% of Republicans oppose increased military spending, while 65% of the Democrats and 55% of the Independents oppose it. A registered voter writes a letter to the county paper, arguing against increased military spending. What is the probability that this voter is a Democrat?

$S = \{\text{registered voters}\}$
$A_1 = \{\text{registered Republicans}\}$
$A_2 = \{\text{registered Democrats}\}$
$A_3 = \{\text{registered independents}\}$
$B = \{\text{voters who oppose increased military spending}\}$

We want to know $\Pr(A_2|B)$.

We have

$$\begin{array}{lll} \Pr(A_1) = .6 & \Pr(A_2) = .3 & \Pr(A_3) = .1 \\ \Pr(B|A_1) = .4 & \Pr(B|A_2) = .65 & \Pr(B|A_3) = .55 \end{array}$$

Using Bayes' Theorem, we have:

$$\begin{aligned}
\Pr(A_2|B) &= \frac{\Pr(B|A_2)\times\Pr(A_2)}{\Pr(B|A_1)\times\Pr(A_1)+\Pr(B|A_2)\times\Pr(A_2)+\Pr(B|A_3)\times\Pr(A_3)} \\
&= \frac{.65\times.3}{(.4\times.6)+(.65\times.3)+(.55\times.1)} \\
&= \frac{.195}{.49} \\
&\approx .398
\end{aligned}$$

AIDS

Suppose we have a test that is 99% effective against AIDS. Suppose we also know that .3% of the population has AIDS. What is the probability that you have AIDS if you test positive?

$S = \{\text{all people}\}$ (in North America??)
$A_1 = \{\text{people with AIDS}\}$
$A_2 = \{\text{people who don't have AIDS}\}$ ($A_2 = \overline{A_1}$)
$B = \{\text{people who test positive}\}$

$$\Pr(A_1) = .003 \qquad \Pr(A_2) = .997$$

Since the test is 99% effective:

$$\Pr(B|A_1) = .99 \qquad \Pr(B|A_2) = .01$$

Using Bayes' Theorem again:

$$\begin{aligned} \Pr(A_1|B) &= \tfrac{.99\times.003}{(.99\times.003)+(.01\times.997)} \\ &\approx \tfrac{.003}{.003+.01} \\ &\approx .23 \end{aligned}$$

Averaging and Expectation

Suppose you toss a coin that's biased towards heads (Pr(heads) = 2/3) twice. How many heads do you expect to get?

- In mathematics-speak:
  What's the *expected number* of heads?

What about if you toss the coin k times?

What's the average weight of the people in this classroom?

- That's easy: add the weights and divide by the number of people in the class.

But what about if I tell you I'm going to toss a coin to determine which person in the class I'm going to choose; if it lands heads, I'll choose someone at random from the first aisle, and otherwise I'll choose someone at random from the last aisle.

- What's the expected weight?

Averaging makes sense if you use an equiprobable distribution; in general, we need to talk about *expectation*.

Random Variables

To deal with expectation, we formally associate with every element of a sample space a real number.

Definition: A *random variable* on sample space S is a function from S to the real numbers.

Example: Suppose we toss a biased coin ($\Pr(h) = 2/3$) twice. The sample space is:

- hh - Probability 4/9
- ht - Probability 2/9
- th - Probability 2/9
- tt - Probability 1/9

If we're interested in the number of heads, we would consider a random variable $\#H$ that counts the number of heads in each sequence:

$$\#H(hh) = 2; \quad \#H(ht) = \#H(th) = 1; \quad \#H(tt) = 0$$

Example: If we're interested in weights of people in the class, the sample space is people in the class, and we could have a random variable that associates with each person his or her weight.