Theorem of total probability

Let $B_1, B_2, \ldots, B_n$ form a partition of S
(i.e., they are mutually exclusive and collectively exhaustive)

We can decompose any event A into a collection of mutually exclusive events:
$(A \cap B_1)$, $(A \cap B_2)$, $\ldots$, $(A \cap B_n)$.

Because $A = A \cap S = A(B_1 \cup B_2 \cup \ldots \cup B_n)$

$$= (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$$

Since they're disjoint, we can compute

$$\boxed{P(A) = \sum_{i=1}^{n} P(A \cap B_i)}$$

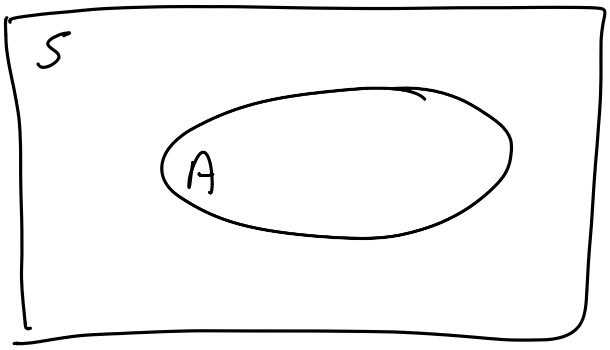
Applying the definition of conditional probability,

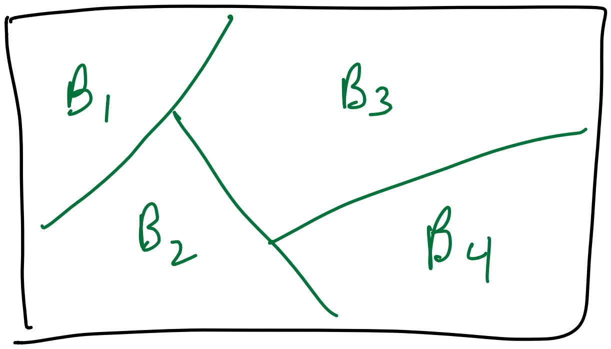
$$\boxed{P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)}$$

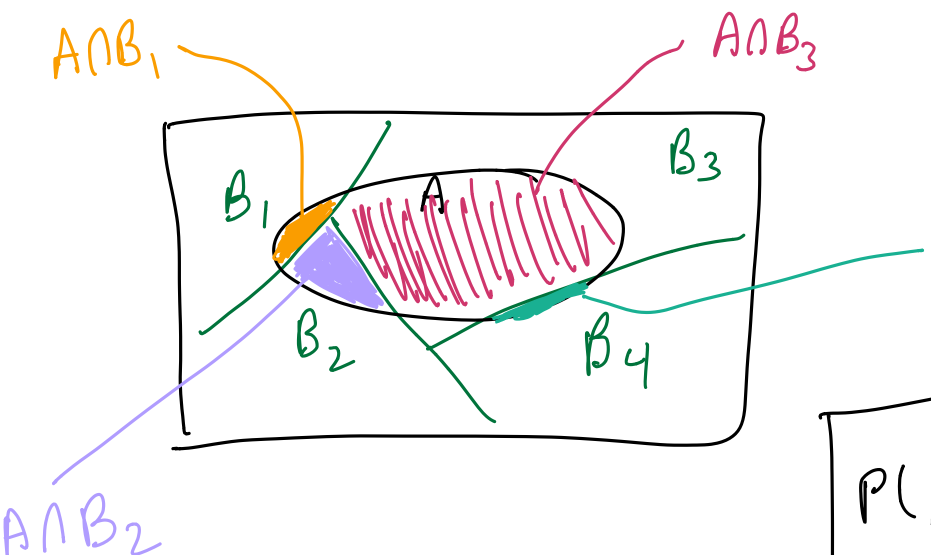
These are both the theorem of total probability

The theorem of total probability (pictorially)

B_i's form a partition







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

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

$$\boxed{P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)}$$

Example : Theorem of total probability

A company has 3 machines that make $1k\Omega$ resistors. Events B_1, B_2, B_3 represent those resistors made by machines 1, 2, 3 respectively.

80% of resistors from B_1 are within 50Ω of desired

90% " B_2 "

60% " B_3 "

and machine B_1 makes 3000 resistors per hour

B_2 " 4000 "

B_3 " 3000 "

All resistors are mixed together and a resistor is chosen at random.

What is the probability that a given resistor is within 50Ω of the desired value?

Answer start by defining events

Let $A = \{\text{resistor is within } 50\Omega \text{ of desired}\}$

Then translate the words into math:

$P(A|B_1) = 0.8$ $\quad P(A|B_2) = 0.9$ $\quad P(A|B_3) = 0.6$

The total # resistors produced in an hour $= 10{,}000$.

So $P(B_1) = 0.3$ $\quad P(B_2) = 0.4$ $\quad P(B_3) = 0.3$

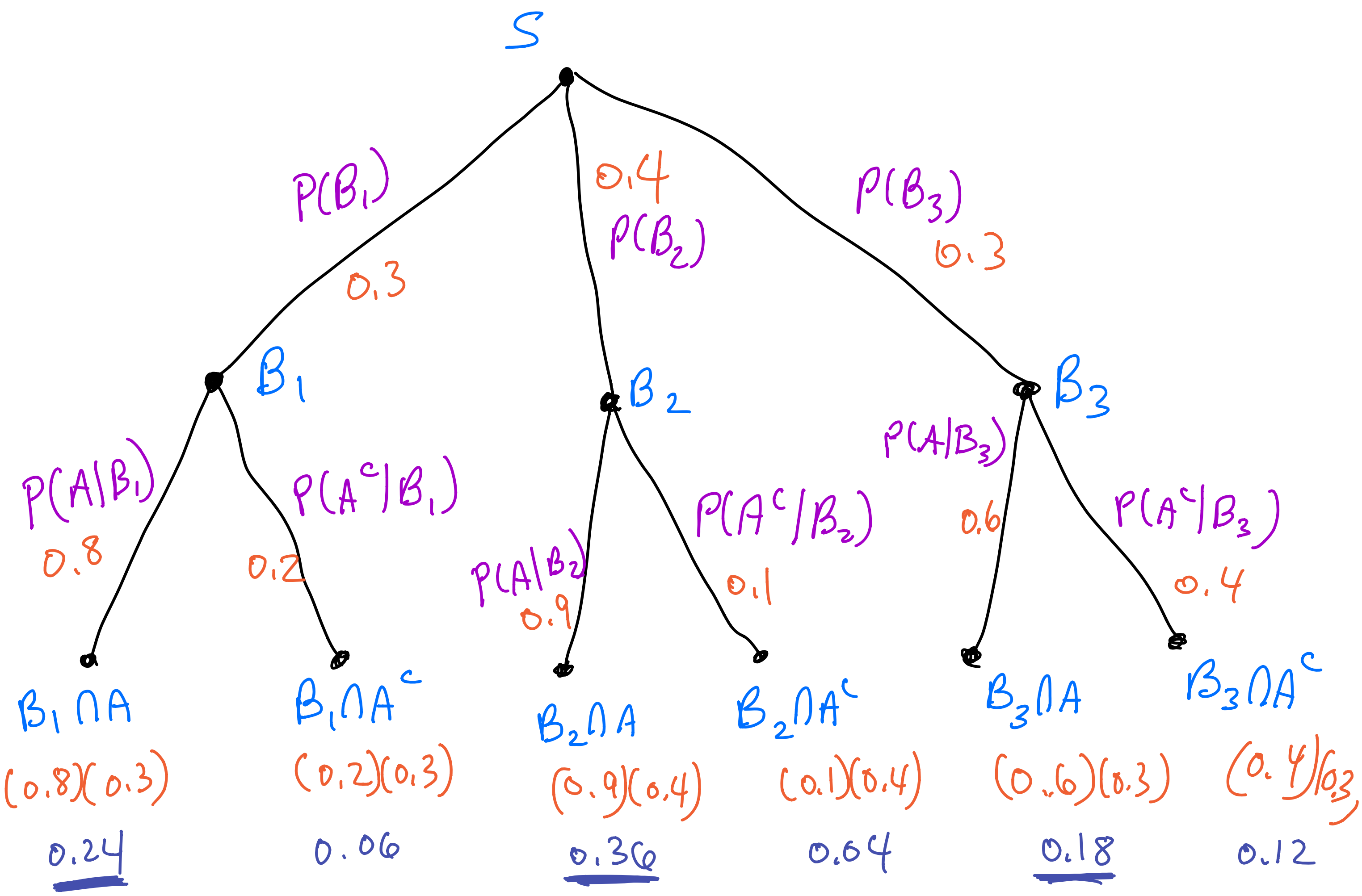
Then apply the theorem of total probability

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$
$$= (0.8)(0.3) + (0.9)(0.4) + (0.6)(0.3)$$
$$= \boxed{0.78}$$

For these 3 machines, overall 78% of resistors are within 5% of the desired value

---

An approach using a tree diagram



$$P(A) = 0.24 + 0.36 + 0.18 = 0.78$$

The tree diagram, explained

- The root of the tree (top or left) is the entire sample space
- Each node is an event
- Each set of branches partition the event at the previous node into disjoint sets
- Label each branch with a conditional probability. The conditioning event is the event at the previous node.

  (Note: $P(A) = P(A|S)$)
- Leaf nodes (at the bottom or right) are events of interest (example: $A \cap B$,)
- To get the probability of an event corresponding to a leaf node, follow the path from the leaf node to the root, multiplying probabilities

$\boxed{\text{Probabilities leaving each node sum to 1.}}$

$\boxed{\text{At each level of the tree, the probabilities of all events sum to 1}}$

One example of the power of probability models defined by conditional probability:

Return to the example of 3 machines making resistors.

We know that $P(A|B_1) = 0.8$, $P(A|B_2) = 0.9$, and $P(A|B_3) = 0.6$.

Originally, $P(B_1) = 0.3$ $P(B_2) = 0.4$, $P(B_3) = 0.3$

Now suppose machine 3 breaks completely and stops making any resistors.

In this new scenario, $P(B_1) = 3/7$, $P(B_2) = 4/7$.

What is the probability of event A now?

Answer:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$

$$= \frac{1}{7}\left((0.8)(3) + (0.9)4\right)$$

$$= \frac{2.4 + 3.6}{7} = \frac{6}{7}$$

Because machines 1 and 2 didn't change!

Example: Suppose 2 horses race against each other regularly, and it's observed that in rain, horse 1 wins 70% of the time, but in dry weather, horse 1 only wins 20% of the time.

a) The two horses are scheduled to race a week from now, and the forecast is 50% chance of rain. What's the probability horse 1 will win?

Answer: Let $H = \{\text{horse 1 wins}\}$ $R = \{\text{rain}\}$

We know $P(H|R) = 0.7 \Rightarrow P(H^c|R) = 0.3$

$P(H|R^c) = 0.2$ $\quad P(H^c|R^c) = 0.8$

By the theorem of total probability,

$$P(H) = P(H|R)P(R) + P(H|R^c)P(R^c)$$
$$= \frac{1}{2}(0.7 + 0.2) = \frac{9}{20} = \boxed{\frac{45}{100}}$$

b) The morning of the race, the forecast predicts 80% chance of rain. What's the revised probability horse 1 will win?

Answer: now $P(R) = 0.8$ and $P(R^c) = 0.2$

$$P(H) = (0.7)(0.8) + (0.2)(0.2) = \boxed{0.60}$$

Interpretation: Using conditional probabilities to define models can be very powerful

Bayes Rule

(Thomas Bayes was a minister, 1702–1761. He never published his work.)

Recall the definition of conditional probability:

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \quad \text{if } P(B) > 0$$

Multiplying both sides by $P(B)$, we have

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

so

$$\boxed{P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad \text{when } P(A)\neq 0 \text{ and } P(B)\neq 0}$$

This also extends to a partition of n sets:

$$\boxed{P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{k=1} P(A|B_k)P(B_k)}}$$

(note that in the denominator, we applied the theorem of total probability)

We will use versions of Bayes Rule throughout the course to learn from observing the outcome of an experiment.

Use Bayes Rule when you want to "flip" the conditioning, for example when the model describes $P(A|B)$ but you want to find $P(B|A)$

$P(B_i)$ are prior probabilities
(a priori) (before the experiment)

$P(B_i|A)$ are the posterior probabilities
(a posteriori) (after observing
the outcome
that the event
A occured during
the experiment)

A classic example where intuition often fails us

You have been tested for cancer, and you received a positive test result. However, only 0.5% of the population has this particular cancer.
The test is 98% accurate.
What are the chances you have cancer?

To define events, step back and consider the situation that existed <u>before</u> the experiment.
The experiment was to take the cancer test.
There were 2 uncertainties
a) you may or may not have cancer ⇒ event C
b) you may or may not receive a positive result ⇒ event P

$C = \{\text{cancer}\}$ $\quad P = \{\text{positive result}\}$

want $P(C|P)$ (since you <u>did</u> get a positive result)

Accuracy 98% ⇒ $P(P|C^c) = 0.02$
<u>and</u> $P(P^c|C) = 0.02$

Also, $P(C) = 0.005$ from the problem statement

$$P(C|P) = \frac{P(P|C)P(C)}{P(P)} = \frac{P(P|C)P(C)}{P(P|C)P(C) + P(P|C^c)P(C^c)}$$

$$= \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.02)(0.995)} \approx 0.198$$

About 20% chance you have cancer despite a positive test!

Comments on the conclusions drawn by Bayes Rule.

A lot depends on the prior probability, in this case, $P(c)$

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

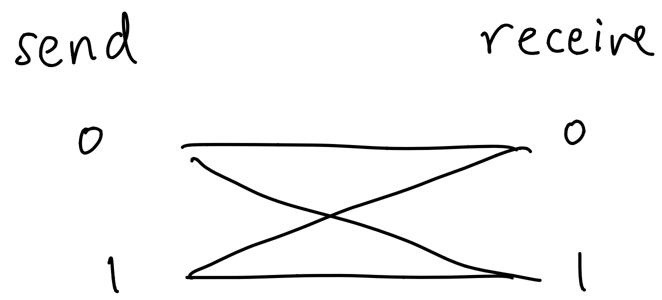
example 98% test accuracy

prior probability $P(c)$	posterior probability $P(c\|P)$
0.005	0.198
0.05	0.72
0.5	0.98

- The doctor might have had a hunch that you had cancer before ordering the test. So the $P(c)$ might have been higher.

- Even if $P(c)$ is small, and say $P(c|P) \approx 0.2$, we can increase our belief about the event c by doing a second test..

Example: Binary communications channel



Let $S_i = \{\text{send } i\}$ and $R_i = \{\text{receive } i\}$

In a real system, we could measure the probability of receiving a one when we send a zero, and that of receiving a zero when we send a one.

Suppose when we send a zero, it is correctly received 95% of the time, and when we send a zero it's correctly received 90% of the time.

Then $P(R_0 \mid S_0) = 0.95$ and $P(R_1 \mid S_1) = 0.1$

so $P(R_1 \mid S_0) = 0.05$ and $P(R_0 \mid S_1) = 0.9$

The prior probabilities $P(S_0)$ and $P(S_1)$ will affect the overall probability of error in this case.

$$\begin{aligned} P(\text{error}) &= P(\text{error} \mid S_0)\,P(S_0) + P(\text{error} \mid S_1)\,P(S_1) \\ &= P(R_1 \mid S_0)\,P(S_0) + P(R_0 \mid S_1)\,P(S_1) \\ &= P(R_1 \mid S_0)\,P(S_0) + P(R_0 \mid S_1)\left[1 - P(S_0)\right] \end{aligned}$$

Another quantity of potential interest:
what's the probability that if you received a one (R_1) that a one was actually sent?

Bayes Rule: $P(S_1|R_1) = \dfrac{P(R_1|S_1)\,P(S_1)}{P(R_1)}$ if $P(R_1) \neq 0$

So we need to find $P(R_1)$ using the theorem of total probability

$$P(R_1) = P(R_1|S_0)\,P(S_0) + P(R_1|S_1)\,P(S_1)$$

Suppose $P(S_0) = P(S_1) = 1/2$

$$P(R_1) = (0.05)\,1/2 + (0.90)\,1/2 = 0.475$$

And substituting this into Bayes Rule,

$$P(S_1|R_1) = \frac{(0.9)\,1/2}{0.475} = \frac{90}{95}$$

Also, $P(S_0|R_1) = 5/95$

$\Rightarrow$ a sensible result. more likely to have sent a one if we receive a one

BUT suppose the prior probabilities were different.

$P(S_0) = 0.99$ and $P(S_1) = 0.01$.

Using the same equations, $P(R_1) = 0.0585$

$P(S_1|R_1) = 0.154$

Even if we receive a one, it's more likely that a zero was sent!!