

Counting Elementary Events

- We still consider the same random variable: **what was the outcome when I rolled the die 5 times**
- How many ways are there of selecting 3 elementary events out of Ω ?
- First, select the first element: there's 6^5 ways
- Next, select the second element: there's $6^5 - 1$ ways
- Finally, select the last element: there's $6^5 - 2$ ways
- We selected the items in order - but we are putting them unordered in a bag, there are 6 ways of ordering them, so we have $(6^5)(6^5 - 1)(6^5 - 2)/6$

Counting Elementary Events

In general, if we are selecting
k elements out of a set of n

n choose k

$$\binom{n}{k} = n \times (n-1) \times \dots \times (n-k+1) / k!$$

$$= \frac{n!}{k!(n-k)!}$$

Recall

$$n! = n \times (n-1) \times \dots \times 1$$

Application: the Binomial Theorem

$$(x+y)^n = \underbrace{(x+y) \times \dots \times (x+y)}_{n \text{ times}}$$

- We consider all the possible combinations of products of n terms
- There is one way of making x^n
- There are ? ways of making $y x^{n-1}$

Application: the Binomial Theorem

$$(x+y)^n = \underbrace{(x+y) \times \dots \times (x+y)}_{n \text{ times}}$$

- We consider all the possible combinations of products of n terms
- There is one way of making x^n
- There are n ways of making $y x^{n-1}$
- There are ? ways of making $y^2 x^{n-2}$

Application: the Binomial Theorem

$$(x+y)^n = \underbrace{(x+y) \times \dots \times (x+y)}_{n \text{ times}}$$

Note: read pages 6-10
of your
book along with this

- We consider all the possible combinations of products of n terms
- There is one way of making x^n
- There are n ways of making $y x^{n-1}$
- There are $n(n-1)/2$ ways of making $y^2 x^{n-2}$
- In general, there are $\binom{n}{k}$ ways of making $y^k x^{n-k}$

Think - Pair Share

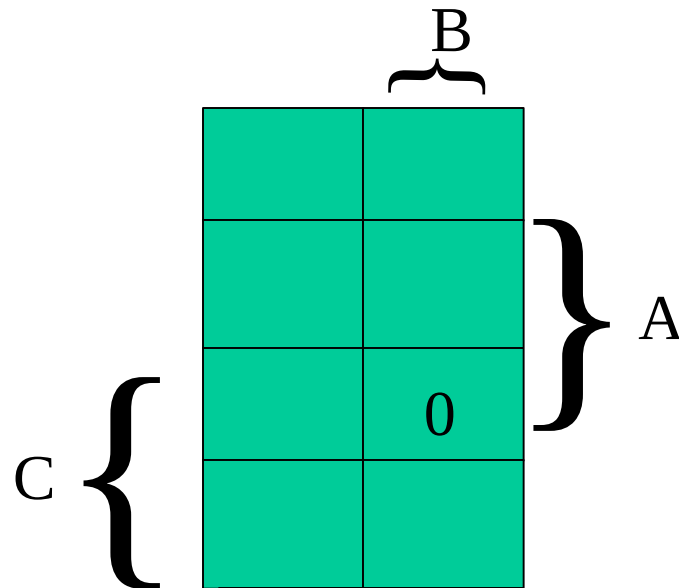
- Give the Karnaugh map

Putting it all together

- Example: I'm trying to debug a program, A. I am having memory leaks. A calls B and C. There is a register of my memory (register k) which is affected if *both* B and C have memory leaks (if only B or C overflow, the register is unaffected). My experiment is running A and checking the registers before and after. I check register k and it was unaffected
- Event A: A overflows. Event B: B overflows. Event C: C overflows.
- The outcome of the experiment was: if A then B and C did not occur together

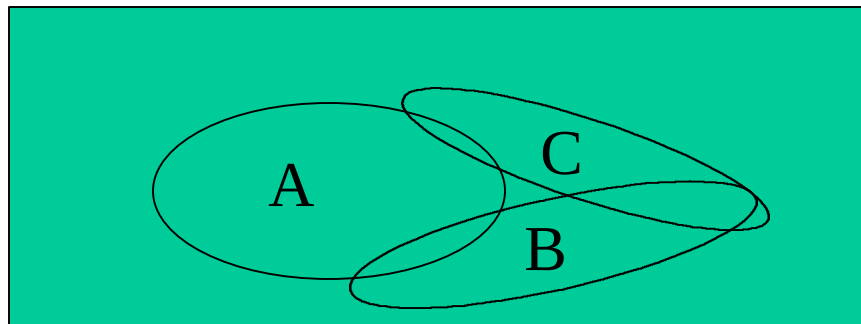
Think - Pair Share

- Give the Karnaugh map



Putting it all together

- Event: if A then B and C did not occur together
- Set-theoretic notation: $(A \cap (B \cap C)^C) \cup A^C$
- Chocolate interpretation: *a square of chocolate in piece A is not in both pieces B and C*
- Venn diagram:



Does an event with probability 0 never happen?

- In practice, we do not have the accuracy to distinguish between an event which has probability 0 and probability 10^{-40}
- If think of a “relative frequency” interpretation (later in the course, we’ll give some backbone to the relative frequency interpretation), then the two are basically the same
- Later we’ll look at continuous outcomes and re-visit the concept of events with probability 0

Outline

- Random variables: getting a bit more precise
- Conditioning

Random Variable

- We have developed an intuitive feel for what is a random variable
- We have up to now mapped a random variable to an experiment - the outcome of the experiment determines the value the random variable takes
- More formal definition: **real-valued function defined on a sample space**
- Why do this? **I can deal with random variables (r.v.s) abstractly (I no longer have to think about dice, rain, etc...) and r.v.s are clearly linked to sample spaces**

Reality Check

But things may not be real valued!

- In some cases (points from rolling a die), the outcome is a real anyway
- What if the outcome is whether I get beans or carrots today?

Reality Check

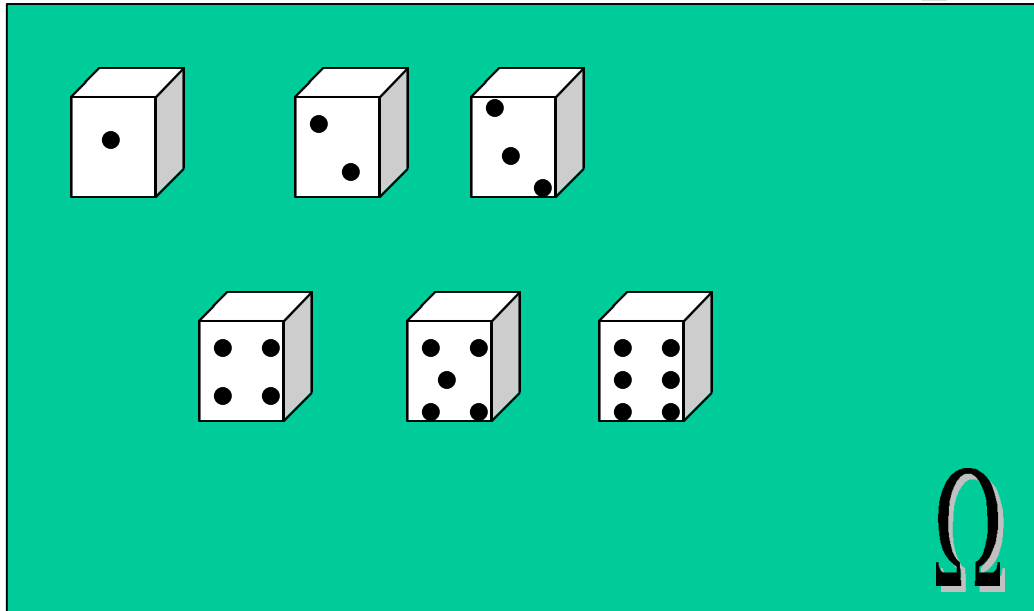
- We do a mapping: for instance map beans to a 1 and carrots to a 0

Look at Section 4.1 in Ross and pp.65-68 in your notes carefully.

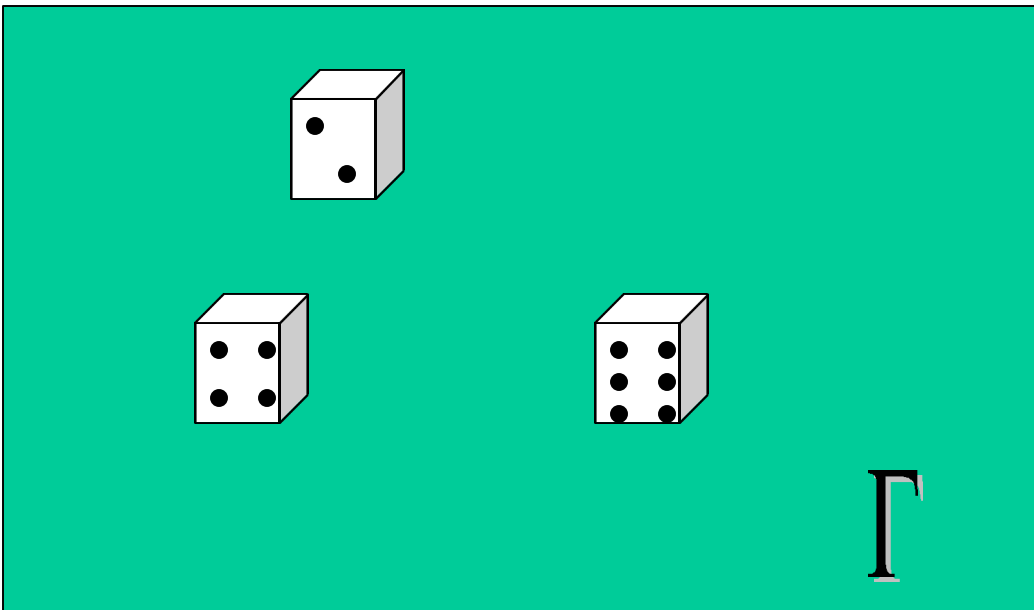
Why is the fact that
r.v.s are linked to sample spaces
an advantage of our new definition?

- We may want to define sample spaces and random variables on those sample spaces which are not simple experiments
- Example: consider the random variable A: **whether or not I get a 4 from rolling a die**
- **Remember:** we want it to be a real function, so say that A is 1 if we get a 4 from rolling a die and 0 otherwise
- Let's consider 2 different sample spaces ...

Different sample spaces



Recall our old sample set
We define it as
Set or possible outcomes:
1 through 6



New sample set
We define it as
Set or possible **even**
outcomes:
2 , 4, 6

Random variables and sample spaces

- I can define A on either of the sample spaces Ω and Γ
- But clearly - it is a different random variable depending on which sample space I define it on!

Think - Pair Share

- What is the probability that the random variable A takes on the value 1 if we define it on Ω ?
- What if we define it on Γ ?
- Hint: just guess right now

Think - Pair Share

- What is the probability that the random variable A takes on the value 1 if we define it on Ω ?
- $1/6$
- What if we define it on Γ ?
- $1/3!$

What's going on?

- The sample space Γ is itself a subset of Ω
- But recall that a subset of Ω is an **event**
- So we could define A either on Ω itself or on an **event** defined on Ω
- The random variable does not have a life of its own, it must be specifically linked to the sample space

Outline

- Random variables: getting a bit more precise
- Conditioning

Conditioning - you just did it!

- When we consider the random variable A in the new sample space given by the event Γ , we say that we **condition** on Γ (or **given** Γ)
- We write it as $A \mid \Gamma$
- We have to **renormalize**
- Chocolate interpretation: in our original chocolate bar, we had 6 squares, of 1/6 lb each
- Now we have 3 squares, but the weight of those 3 squares must still be 1 lb - so each square is 1/3 lb

How do we normalize?

- Remember our axioms: the probabilities must add up to one (**weight of bar of chocolate is 1 lb**)
- We guessed the probability of each elementary event conditioned on Γ to be $1/3$
- $P(A=1) = P((A=1) \cap \Gamma) = 1/6$ (in the original Ω)
- $P(\Gamma) = 1/2$
- $P((A=1) | \Gamma) = 1/3 = P((A=1) \cap \Gamma) / P(\Gamma)$
- The normalizing factor is **$1/P(\Gamma)$**

Reality Check

- What happened to the elementary events in Γ which are not in Ω ? Or - what is $P(\text{outcome of a die roll has } 5 \mid \Gamma)$?

Reality Check

- What happened to the elementary events in Ω which are not in Γ ? Or - what is $P(\text{outcome of a die roll has } 5 \mid \Gamma)$?

It's 0!

- The elementary events in Γ which are not in Ω are part of $\emptyset$ when we condition on (or given) Γ

Let us refine our model

- Why do we write $P((A=1) \cap \Gamma) / P(\Gamma)$ instead of just $P(A=1) / P(\Gamma)$?
- Consider another example - let us look at the random variable B: **B is 1 if the roll of a die yields a number less than 4 and 0 otherwise**

Think - Pair Share

- $P(B=1) = ?$
- $P((B=1) \cap \Gamma) = ?$
- $P((B=1) \mid \Gamma) = ?$

Think - Pair Share

- $P(B=1) = 1/2$
- $P((B=1) \cap \Gamma) = ?$
- $P((B=1) \mid \Gamma) = ?$

Think - Pair Share

- $P(B=1) = 1/2$
- $P((B=1) \cap \Gamma) = 1/6$: the only way is to have 2 be the outcome of the roll
- $P((B=1) | \Gamma) = ?$

Think - Pair Share

- $P(B=1) = 1/2$
- $P((B=1) \cap \Gamma) = 1/6$: the only way is to have 2 be the outcome of the roll
- $P((B=1) \mid \Gamma) = 1/3$

Study Self-Test

- What is the probability of Γ happening conditioned on B?
- What is the probability of Γ happening conditioned on A?

Read carefully over the examples in pages 17-19 of your notes

Overview

- Last time: we figured out that
 - $P(A \mid B) = P(A \cap B) / P(B)$
 - it's important to have $P(A \cap B)$ and not just $P(A)$ in the numerator
 - the denominator is a normalization factor - event B is our new sample space

Think - Pair Share

- $P(B=1) = 1/2$
- $P((B=1) \cap \Gamma) = 1/6$: the only way is to have 2 be the outcome of the roll
- $P((B=1) | \Gamma) = 1/3$
- Remember $P(\Gamma) = 1/2$, so $P(B=1) / P(\Gamma) = 1$, which is clearly not $P((B=1) | \Gamma)$!

This lecture ...

- Interpret conditioning
- Look at theorem of total probability
- Bayes's Formula

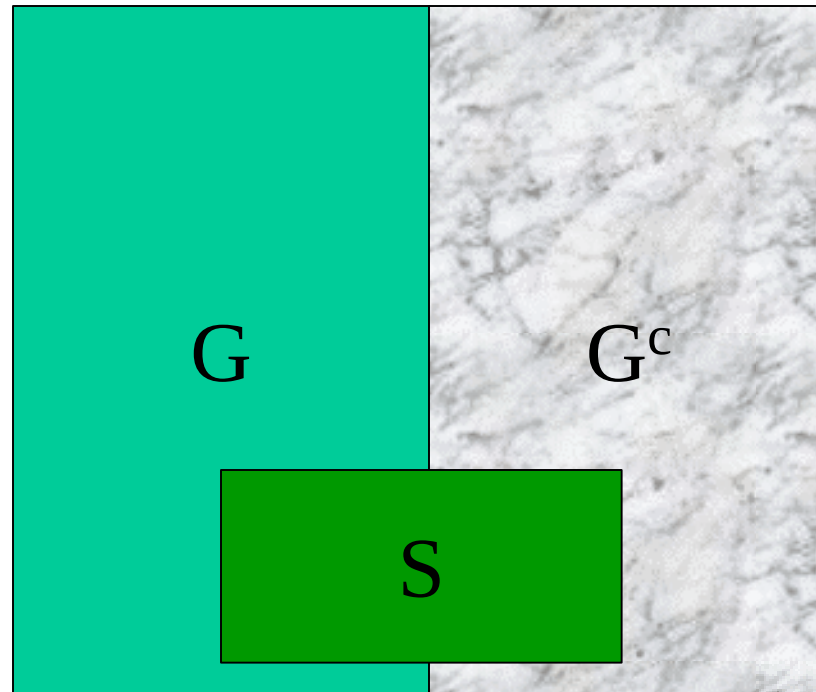
Conditioning as a “what if?” question

- Often is useful to figure out events under different scenarios
- What is the probability that the deficit will be a surplus in 2002 (event S)?
- Two scenarios:
 - economy stays good: G
 - economy tanks: G^c
- We can partition our W into G and G^c

Conditioning as a “what if?” question

$$P(S) = P(S \cap G) + P(S \cap G^c)$$

Theorem of
Total
Probability



Result
depends
strongly on
 $P(G)$!

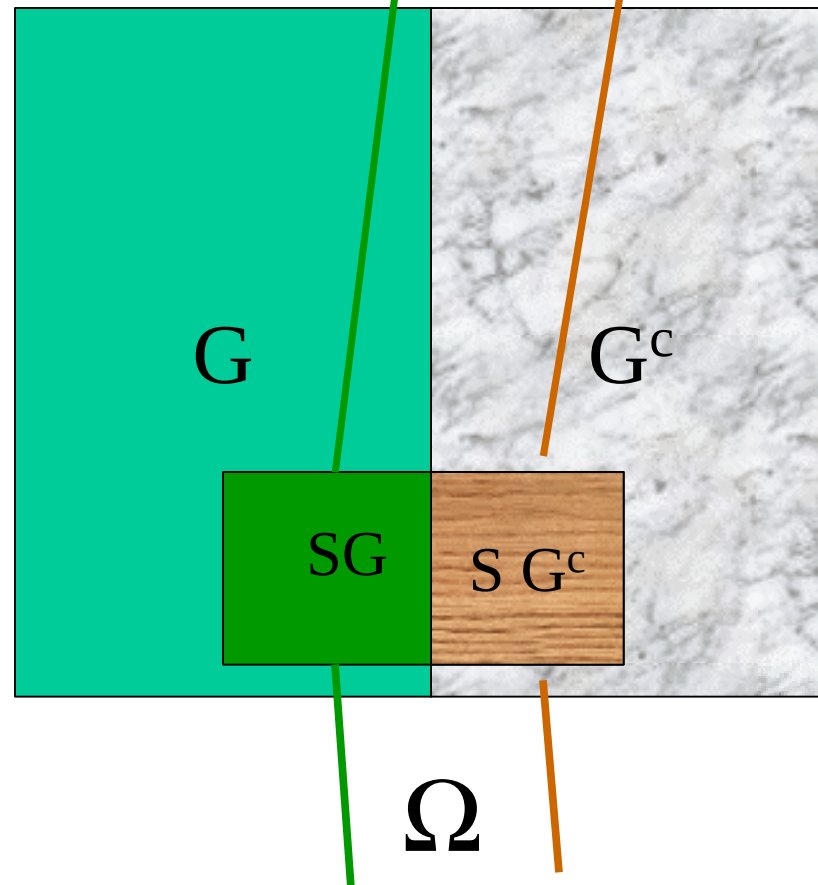
Ω

$$P(S) = P(S \mid G) P(G) + P(S \mid G^c) P(G^c)$$

We have come up with a partition of S

$$P(S) = P(S \cap G) + P(S \cap G^c)$$

Theorem of
Total
Probability



$$P(S) = P(S | G) P(G) + P(S | G^c) P(G^c)$$

Last Time ...

- Interpret conditioning
- Theorem of total probability

This Lecture

- Continue case study: our chip testing example
- Bayes's Theorem
- Bayesian case study: clinical trials

Think-Pair Share

- We use certain sequences (pseudo-random) to test a chip
- From looking at our chips, we know that the probability that the chip is good is 0.999
- The probability that a bad chip looks OK in the test is 0.001
- If my chip looks OK in the test, what is the probability that it is actually OK?

Think-Pair Share

- From looking at our chips, we know that the probability that the chip is good is 0.999
- $P(G^c) = 0.001$, $P(G) = 0.999$
- The probability that a bad chip looks OK in the test is 0.001 (if the chip is OK, the test is always OK)
- If my chip looks OK in the test, what is the probability that it is actually OK?

Think-Pair Share

- From looking at our chips, we know that the probability that the chip is good is 0.999
- $P(G^c) = 0.001$, $P(G) = 0.999$
- The probability that a bad chip looks OK in the test is 0.001 (if the chip is OK, the test is always OK)
 $P(\text{OK} \mid G^c) = 0.001$, so $(\text{OK} \cap G^c) = 10^{-6}$
- $P(\text{OK} \mid G) = 1$, so $(\text{OK} \cap G) = 0.999$
- If my chip looks OK in the test, what is the probability that it is actually OK?

Graphical interpretation



Think - Pair Share

- $P(G^c) = 0.001$, $P(G) = 0.999$
- $P(\text{OK} \mid G^c) = 0.001$, so $P(\text{OK} \cap G^c) = 10^{-6}$
- $P(\text{OK} \mid G) = 1$, so $P(\text{OK} \cap G) = 0.999$
- If my chip looks OK in the test, what is the probability that it is actually OK? $P(G \mid \text{OK})$
- **Method 1:** $P(G \mid \text{OK}) = P(G \cap \text{OK}) / P(\text{OK})$
- $P(\text{OK}) = P(\text{OK} \cap G^c) + P(\text{OK} \cap G)$ (Theorem of Total Probability)
- $P(G \mid \text{OK}) = 0.999 / (0.999 + 10^{-6})$

Think - Pair Share

- $P(G^c) = 0.001$, $P(G) = 0.999$
- $P(\text{OK} \mid G^c) = 0.001$, so $P(\text{OK} \cap G^c) = 10^{-6}$
- $P(\text{OK} \mid G) = 1$, so $P(\text{OK} \cap G) = 0.999$
- If my chip looks OK in the test, what is the probability that it is actually OK (i.e. G)? $P(G \mid \text{OK})$
- **Method 2:** $P(G^c \mid \text{OK}) = P(G^c \cap \text{OK}) / P(\text{OK})$
- $P(G^c \mid \text{OK}) = 10^{-6} / (0.999 + 10^{-6})$
- $P(G \mid \text{OK}) = 1 - P(G^c \mid \text{OK}) = (0.999 + 10^{-6} - 10^{-6}) / (0.999 + 10^{-6}) = 0.999 / (0.999 + 10^{-6})$ **CHECKS!**

Read notes pages
17-24.

Bayes's Formula

- We simply use substitution
- $P(A | B) = P(A \cap B) / P(B) = P(A) P(B | A) / P(B)$
- so $P(A | B) / P(B | A) = P(A) / P(B)$
- We can also use the theorem of total probability in the denominator
- $$P(A | B) = P(A) P(B | A) / (P(B | A) P(A) + P(B|A^C) P(A^C))$$

Is it more than just re-arranging terms?

- No, not really
- But sometimes it makes our job much easier, depending on what information we're given
- Example: we are evaluating a new medical test, and need to design clinical trials. We decide to run the test on people with and without disease D and we find that the test being positive (event T) has $P(T \mid D)$ probability when someone has the disease and $P(T \mid D^c)$ probability when someone is healthy

Example for Bayes's formula

- We decide to use the test to diagnose disease D
- Which quantity are we interested in?
- How do we use the results from our medical trials to get those quantities?

Example for Bayes's formula

- We decide to use the test to diagnose disease D
- Which quantity are we interested in?
- We want to know $P(D | T)$ and $P(D | T^c)$
- How do we use the results from our medical trials to get those quantities?

Think - Pair Share

Example for Bayes's formula

- We decide to use the test to diagnose disease D
- Which quantity are we interested in?
- We want to know $P(D | T)$ and $P(D | T^c)$
- How do we use the results from our medical trials to get those quantities?
- We're missing critical data! What is $P(D)$ and $P(D^c)$? These are the *a priori* probabilities.

How do we apply our test?

- We go back to our trials and figure out that $d\%$ of the people in the study had disease D and $(100-d)\%$ did not have disease D. Now what?

How do we apply our test?

- We go back to our trials and figure out that $d\%$ of the people in the study had disease D and $(100-d)\%$ did not have disease D . Now what?
- We say that $P(D) = d/100$ and $P(D^c) = (100-d)/100$ and apply Bayes's formula
- $P(D | T) = P(D) P(T | D) / (P(T | D) P(D) + P(T|D^c) P(D^c))$

How do we apply our test?

- We now decide to figure out how common this disease is in the population at large, so we decide to use our test to figure this out. Are my results any good?

How do we apply our test?

- We now decide to figure out how common this disease is in the population at large, so we decide to use our test to figure this out. Are my results any good?
- $P(D | T) = P(D) P(T | D) / (P(T | D) P(D) + P(T|D^C) P(D^C))$
- If the test was run on a population which was sicker to start out with, our test will not be applicable! BEWARE OF CIRCULAR REASONING!

How do we apply our test?

- $P(D | T) = \frac{P(D)P(T | D)}{(P(T | D)P(D) + P(T | D^c)P(D^c))}$
- $y = \frac{x \cdot a}{a \cdot x + b \cdot (1-x)}$
- Recall $P(T | D) \neq 1 - P(T | D^c)$
- $\frac{dy}{dx} = \frac{a(ax - bx + b) - (a-b)ax}{(ax + b(1-x))^2}$
 $= \frac{ab}{(ax + b(1-x))^2}$
 > 0

Overview

- Last time:
 - our chip testing example, continued
 - Bayes's Theorem
 - Bayesian case study: clinical trials
- This time:
 - decision under uncertainty: decision trees
 - beginning of hypothesis testing

How do we use conditional probabilities to guide our decisions?

- Recall: conditioning can be thought of as a “what if?” question. It is useful to figure out events under different scenarios
- We can use conditioning to figure out our options

Decision Trees

- A student has to decide whether to take a good job offer now, or wait to hear from Grad School. The job offer is for 60 K. If she passes it up, she will only get to make 50 K. If she gets into grad school, she will definitely choose to go and she'll make 70 K. Suppose the student only wants to maximize her revenue, what will she do?

Model

- Event J: takes the good job now.
- Event G: gets into grad school.
- We consider the random variable S, which is the salary
- We are given $P(S = 60 \text{ K} \mid J)$
- $P(S = 70 \text{ K} \mid G)$
- $P(S = 50 \text{ K} \mid J^c G^c)$

Model

- Event J: takes the good job now.
- Event G: gets into grad school.
- We consider the random variable S, which is the salary
- We are given $P(S = 60 \text{ K} \mid J) = 1$
- $P(S = 70 \text{ K} \mid G) = 1$
- $P(S = 50 \text{ K} \mid J^c G^c) = 1$

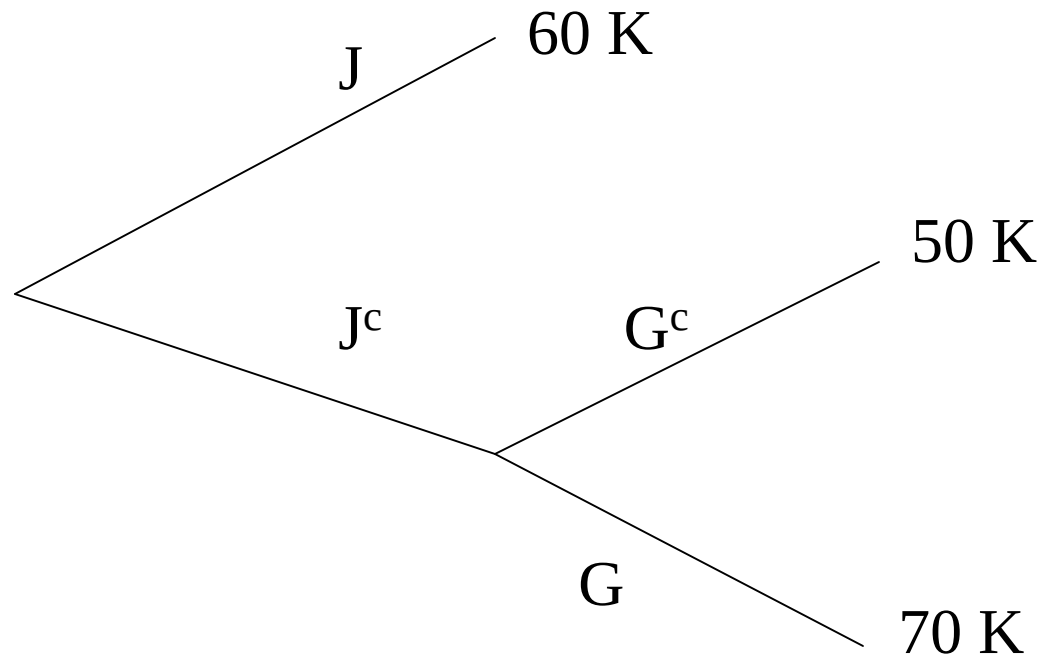
Think - Pair Share

- What are the probabilities we need?

Think - Pair Share

- What are the probabilities we need?
- Just $P(G)$!

How do we model it?



Reality Check

- What is the student trying to figure out?

Reality Check

- What is the student trying to maximize?
- $P(G) \times 70 + (1 - P(G)) \times 50 <> 60$
- Later we'll firm up this intuitive concept of expected value

Self-Study: figure out for what value of $P(G)$ she should wait to see if she gets into grad school.

The story continues ...

- Suppose now the student's classmate is trying to figure out if she got the good job offer. He knows that she got a job offer but he doesn't know whether it was a 60 K job offer. He doesn't want to ask. He simply observes whether or not she takes the job offer. He knows her chances of getting into grad school are 60 %. He knows her chances of **getting the good job offer are 30%**. What is he going to infer about the kind of job offer she got?

Hypothesis Testing

- Let us now consider cases when we cannot always make decisions with certainty
- We want to test a hypothesis:
 - whether or not my radar detected a target
 - whether or not a chip is good
 - whether or not a new treatment is effective.

Maximum Likelihood

- Pick the one that maximizes the probability of what we observed
- Hypotheses $H_1, H_2, H_3 \dots$
- We observe A_j
- We select the H_i for which $P(A_j)$ is highest
- We have different probability measures to use for each H_i

In terms of Conditional Probabilities

- We look at the H_i maximizing $P(A_j | H_i)$
- We call H_i the true hypothesis
- How would we calculate our probability of error?

$$\sum_{\text{all } H_m} P(H_m) \sum_{\substack{\text{all } A_j \text{ such that } P(A_j | H_k) > P(A_j | H_m) \\ \text{for some } k}} P(A_j | H_m)$$

Example

- A target is either present or absent for a radar measurement
- The reading is HI when the target is present 95 % of the time
- The reading is HI when the target is not present 20 % of the time.

Think - Pair Share

- Do we know the probability that the reading is LO when the target is absent?
- Do we know the probability that the reading is LO when the target is present?

Think - Pair Share

- Do we know the probability that the reading is LO when the target is absent?
- 80 %
- Do we know the probability that the reading is LO when the target is present?
- 5 %.

How do I do ML detection?

- Just pick best option to maximize the likelihood
- How sensitive is our decision to probabilities?
- How about our probability of error?