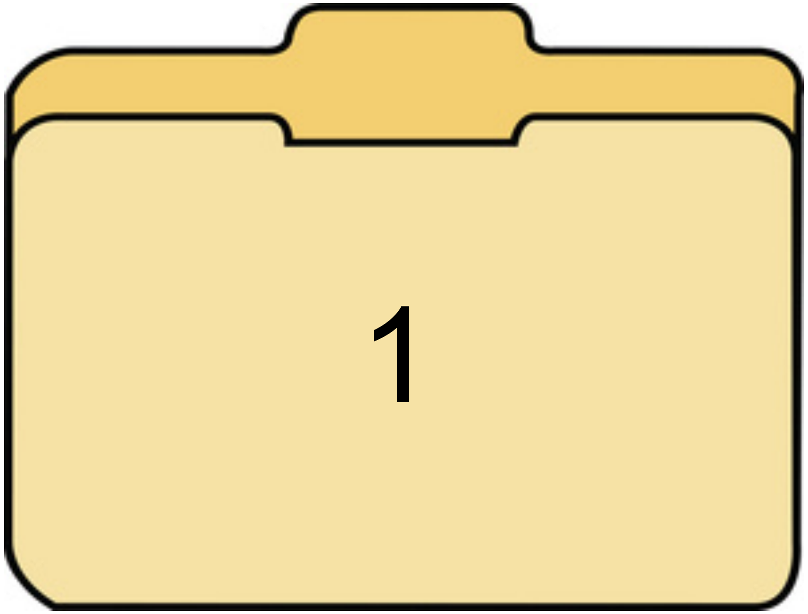
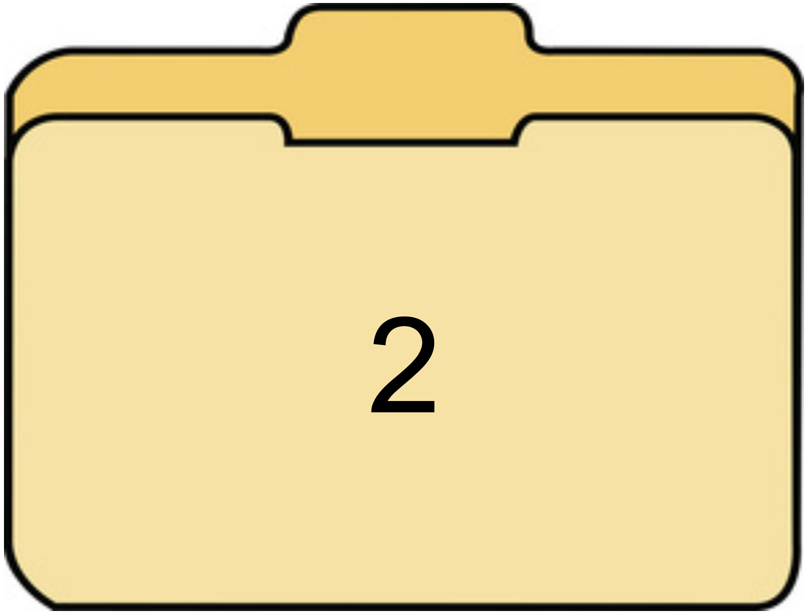
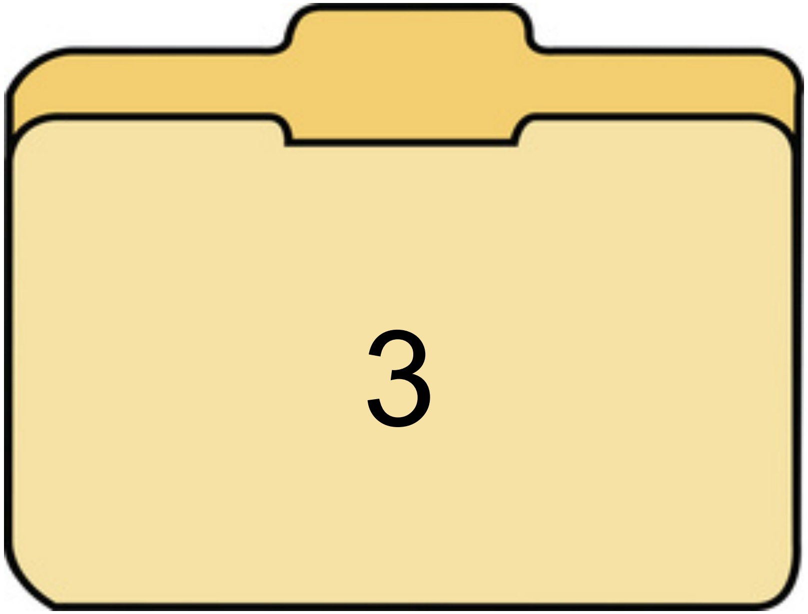
Bayes' Law and Bayes Nets

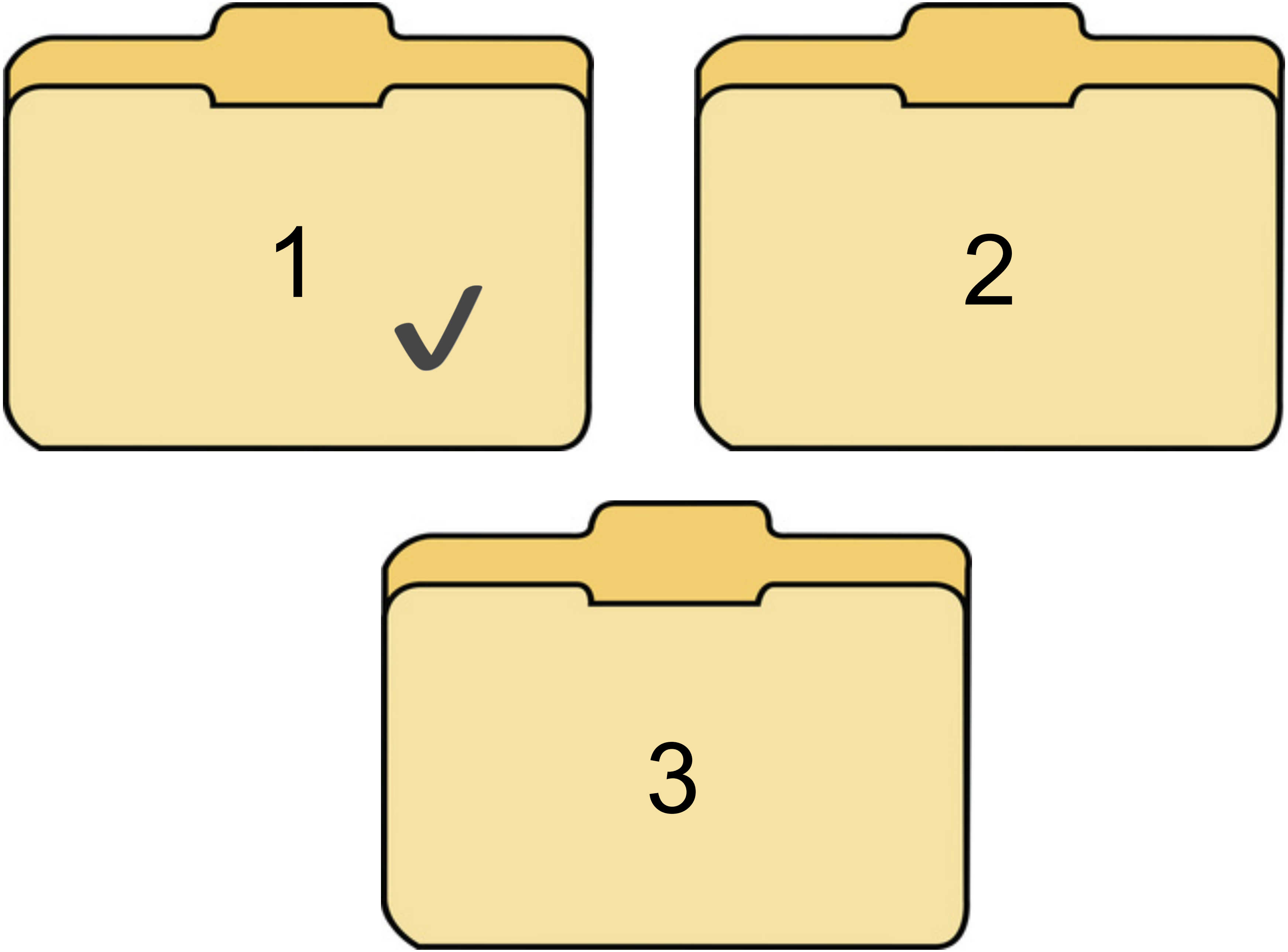
Announcements

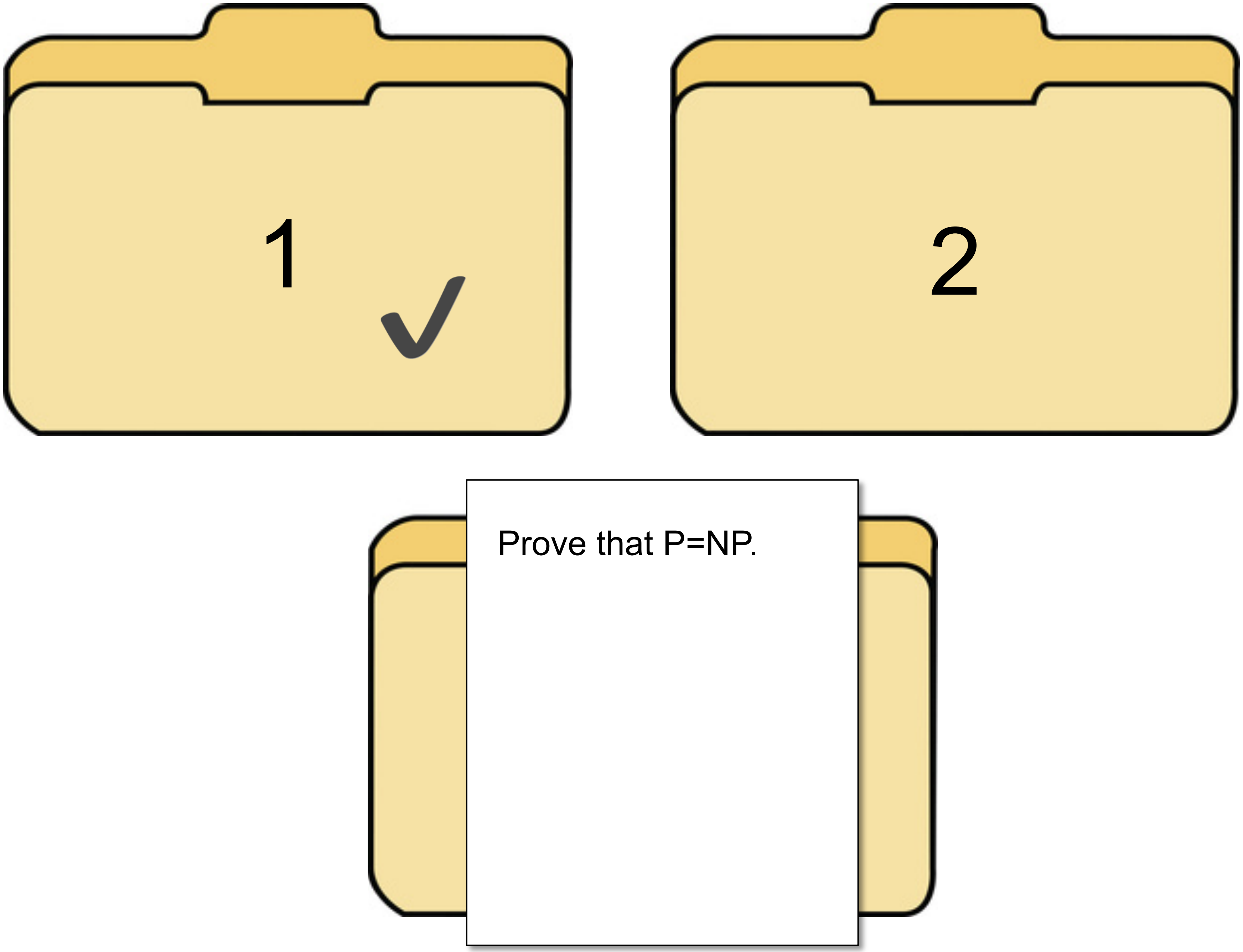
- Don’t forget to reach out to your team members for A1.


1


2


3


1
2
3


1
2
Prove that P=NP.

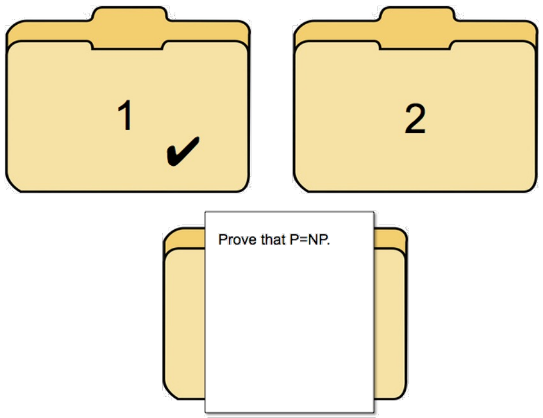
Assumptions

- Easy exam randomly placed in one of the 3 folders
- The teacher always reveals a hard exam
  - If the student chooses a hard exam, the teacher reveals the other hard exam
  - If the student chooses an easy exam, the teacher reveals one of the hard exams, chosen at random

Using Bayes' law…

Given that #1 was chosen by the student,

P(2 easy | 3 shown) = P(3 shown | 2 easy) P(2 easy) / P(3 shown)



P(2 easy) = ?

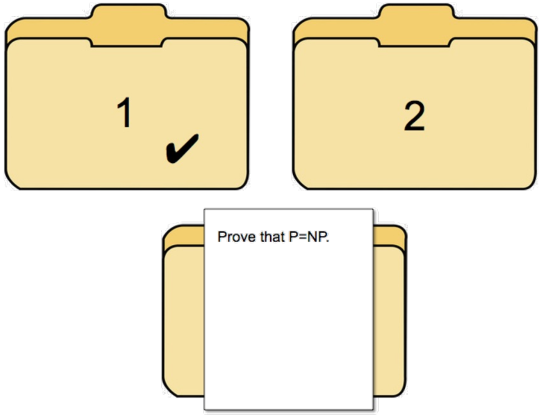
P(3 shown | 2 easy) = ?

P(3 shown) = ?

Using Bayes' law…

Given that #1 was chosen by the student,

P(2 easy | 3 shown) = P(3 shown | 2 easy) P(2 easy) / P(3 shown)



P(2 easy) = 1/3

P(3 shown | 2 easy) = ?

P(3 shown) = ?

Using Bayes' law…

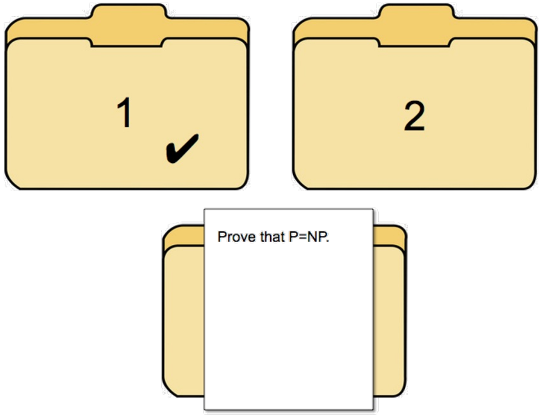
Given that #1 was chosen by the student,

P(2 easy | 3 shown) = P(3 shown | 2 easy) P(2 easy) / P(3 shown)

P(2 easy) = 1/3

P(3 shown | 2 easy) = 1

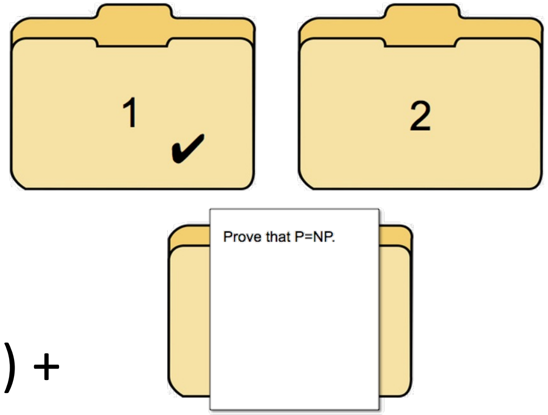
P(3 shown) = ?



Using Bayes' law…

Given that #1 was chosen by the student,

P(2 easy | 3 shown) = P(3 shown | 2 easy) P(2 easy) / P(3 shown)



P(2 easy) = 1/3

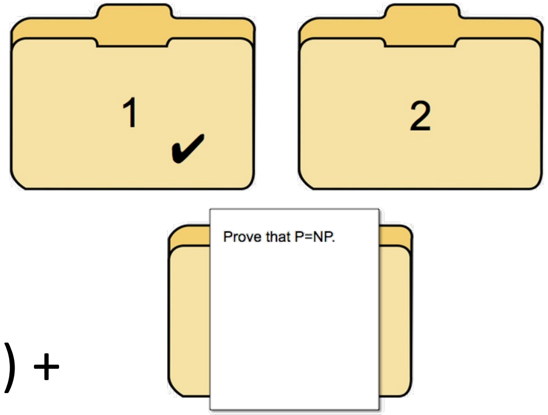
P(3 shown | 2 easy) = 1

P(3 shown) = P(1 easy) P(3 shown | 1 easy) +
P(2 easy) P(3 shown | 2 easy) +
P(3 easy) P(3 shown | 3 easy)

Using Bayes' law…

Given that #1 was chosen by the student,

P(2 easy | 3 shown) = P(3 shown | 2 easy) P(2 easy) / P(3 shown)



P(2 easy) = 1/3

P(3 shown | 2 easy) = 1

P(3 shown) = P(1 easy) P(3 shown | 1 easy) +

P(2 easy) P(3 shown | 2 easy) +

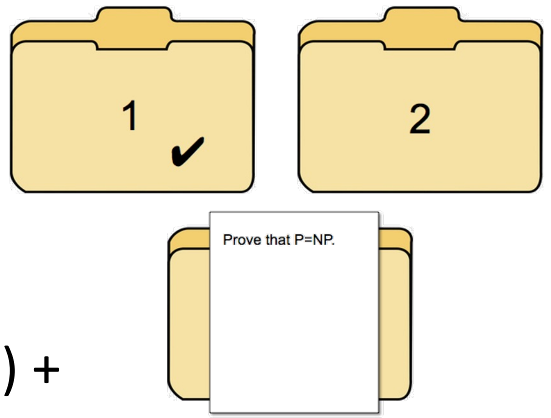
P(3 easy) P(3 shown | 3 easy)

= (1/3) (1/2) + (1/3) (1) + (1/3) (0) = 1/2

Using Bayes' law…

Given that #1 was chosen by the student,

$$P(2 \text{ easy} \mid 3 \text{ shown}) = P(3 \text{ shown} \mid 2 \text{ easy})\, P(2 \text{ easy}) \,/\, P(3 \text{ shown})$$



$$P(2 \text{ easy}) = 1/3$$

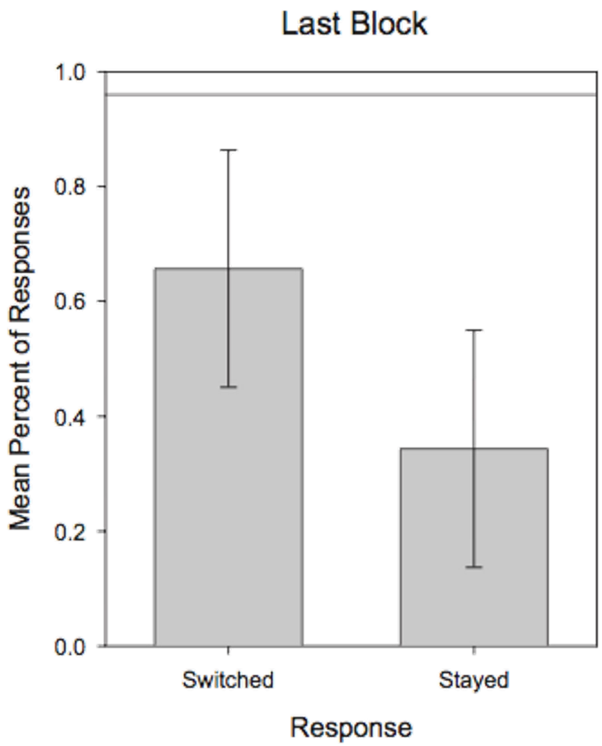
$$P(3 \text{ shown} \mid 2 \text{ easy}) = 1$$

$$\begin{aligned} P(3 \text{ shown}) = {} & P(1 \text{ easy})\, P(3 \text{ shown} \mid 1 \text{ easy}) + \\ & P(2 \text{ easy})\, P(3 \text{ shown} \mid 2 \text{ easy}) + \\ & P(3 \text{ easy})\, P(3 \text{ shown} \mid 3 \text{ easy}) \\ = {} & (1/3)\,(1/2) + (1/3)\,(1) + (1/3)\,(0) = 1/2 \end{aligned}$$

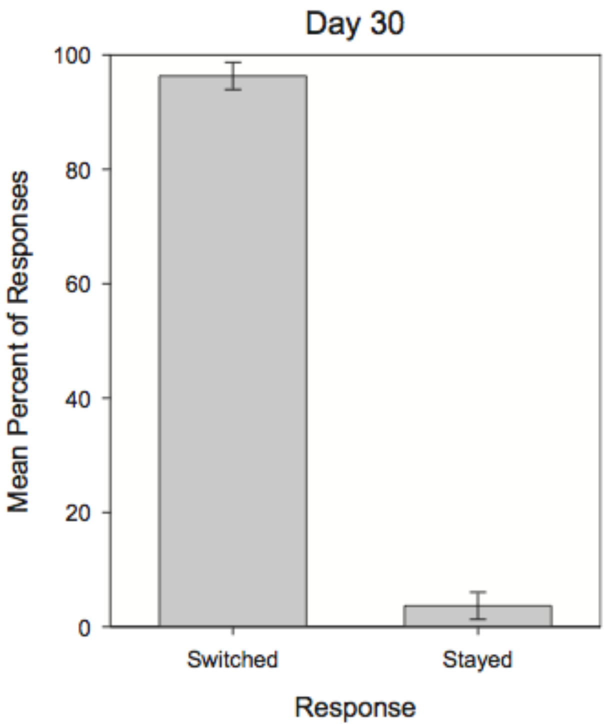
$$P(2 \text{ easy} \mid 3 \text{ shown}) = (1)(1/3) \,/\, (1/2) = 2/3$$

Monty Hall Problem

Behavior of undergrads after 200 trials:



Behavior of pigeons:



Herbranson & Schroeder 2010

Bayes’ Law Example #3b

- Suppose that you have a friend, Mary, who lives in Seattle. She tells the truth 80% of the time and lies the other 20% of the time. The weather in Seattle on any given day is either sunny or cloudy; 30% of days are sunny and 70% are cloudy.

- Suppose that you have another friend in Seattle, Sue, who tells the truth 90% of the time. Sue tells you that the weather is cloudy, while on the same day Mary says that it is sunny. What is the probability that the weather is actually sunny? You may assume that whether or not Sue lies is independent of whether or not Mary lies.

Solution #3b

In addition to events C, S, Mc, and Ms, defined as in the last problem, let Sc and Ss denote the events that Sue says cloudy and sunny, respectively. From Bayes' Law we have:

$$P(S|M_s, S_c) = \frac{P(M_s, S_c|S)P(S)}{P(M_s, S_c)}$$

The assumption that Mary and Sue decide whether or not to lie independently lets us factor P (Ms, Sc|S) as,

$$P(M_s, S_c|S) = P(M_s|S)P(S_c|S) = (0.8)(0.1) = 0.08$$

$$\begin{aligned} P(M_s, S_c) = P(M_s, S_c|S)P(S) + P(M_s, S_c|C)P(C) = \\ P(M_s|S)P(S_c|S)P(S) + P(M_s|C)P(S_c|C)P(C) = \\ (0.8)(0.1)(0.3) + (0.2)(0.9)(0.7) = 0.15 \end{aligned}$$

$$P(S|M_s, S_c) = \frac{(0.08)(0.3)}{0.15} = 0.16$$

Random variables

- A *random variable* is like a multi-valued "attribute" that can be used to specify sets of outcomes
  - Formally, it's a function that associates an attribute with each outcome, typically $f : S \to \mathbb{N}$ or $f : S \to \mathbb{R}$
- E.g. Roll a die 5 times. Let random variable X be the number of 1's rolled.
  - Now P(X=2) denotes probability of rolling two 1's, i.e.

$$P(X = 2) = P(\{a \in S | f(a) = 2\})$$

where *f(a)* counts the number of 1's in a given outcome, i.e. f(11111)=5, f(11114)=4, …

Marginal distributions

- A probability distribution over the values of a random variable is called a *marginal distribution*
- E.g. Roll a die 5 times. Let random variable X be the number of 1's rolled.
  - P(X) is the marginal distribution over X

Joint distributions

- A probability distribution over multiple random variables is called a *joint distribution*
- E.g. Roll a die 5 times. Let random variable X be the number of 1’s rolled, and Y be the *sum* of the rolls.
  - P(X,Y) is the joint distribution, i.e. $P : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$
  - E.g. P(X=3, Y=6) refers to the probability that a given roll has 3 1’s <u>and</u> sums to 6

An example

- Given set of students, define random variables X for Intelligence and Y for grade in class
  - Joint distribution P(X,Y) given by:

		Intelligence	
		low	*high*
Grade	*A*	0.07	0.18
	B	0.28	0.09
	C	0.35	0.03

  - What are the marginal distributions P(X) and P(Y)?

Conditional probabilities

- Conditional probabilities can also be written using random variables
  - Denoted P(X|Y) for random variables X and Y
  - For any value of Y, this is a distribution over X

- Example: What's P(X=low | Y=B)?

		Intelligence	
		low	*high*
	A	0.07	0.18
Grade	*B*	0.28	0.09
	C	0.35	0.03

Expectation

- The expected value of a (discrete) random variable X,

$$E[X] = \sum_{x \in Val(X)} x P(x)$$

  - E.g. Let X be the amount of money you win in the $50 million Powerball Jackpot (Sept 2019). The probability of winning is about 1 in 300 million. What is E[X]?

Statistical Inference

Inference: Using distributions to answer questions

- Several main types of queries
- Type 1: Probability queries
  - Calculate distribution over some random variables given observed values for others
  - e.g. calculate P(**Y** | **X** = **x**), where **X** and **Y** are sets of random variables
  - There might be a set of other random variables **Z** that are neither the query variables nor the observations; these can be "marginalized out":

$$P(\mathbf{Y}|\mathbf{X}=\mathbf{x}) = \sum_{\mathbf{Z}} P(\mathbf{Y},\mathbf{Z}|\mathbf{X}=\mathbf{x})$$

Using distributions to answer questions

- Type 2: Expectation queries
  - Calculate the expected value "on average" for some random variables, given observed values for others
  - i.e. **E[Y | E=e]**

- Type 3: Maximum A Priori (MAP) queries
  - Calculate the most likely values for some random variables, given observed values for others,

$$\arg\max_{\mathbf{y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{E} = \mathbf{e})$$

Back to AI…

- In AI we often want to predict an unknown answer given known answers to past problems
  - E.g., Given current weather observations, will it rain later?
- Whether it will rain (R) may depend on hundreds or thousands of observations, V_1, V_2, … V_{1000}
  - Temperatures across U.S., moisture in atmosphere, etc…
- Given enough days of data, we could directly estimate a probability function $P(R \mid V_1, V_2, \ldots, V_{1000})$
  - Then problem would be solved!
  - How many days of data would you need?

A huge problem

- Say all variables of $(R, V_1, V_2, \ldots, V_{1000})$ are binary
  - Need at least 2^{1000} days of data just to observe all possible combinations of the variables
  - Need to observe multiple days for each combination of variables to estimate conditional probability robustly
  - Simply impossible from a computational, representational, or intuitive point of view
- This seemed fatal for the first ~30 years of AI research
  - Graphical models are a framework for avoiding this problem by making assumptions about the structure of a model

Next class

- More Bayes Nets Inference