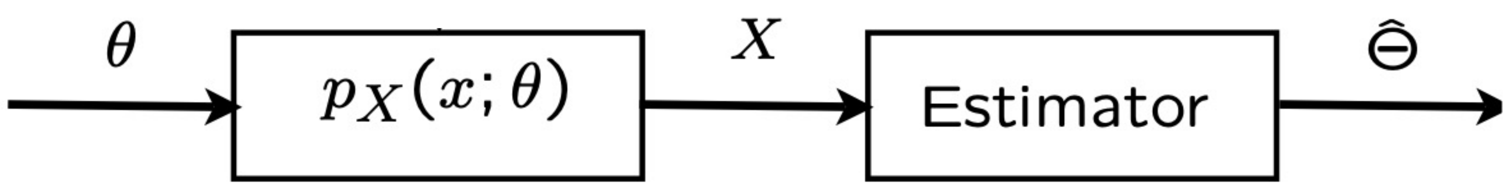
Bayesian Data Analysis

Prof. Pradeep Ravikumar
pradeepr@cs.cmu.edu

Recall: Frequentist Data Analysis

- Frequentist/Classical Data Analysis



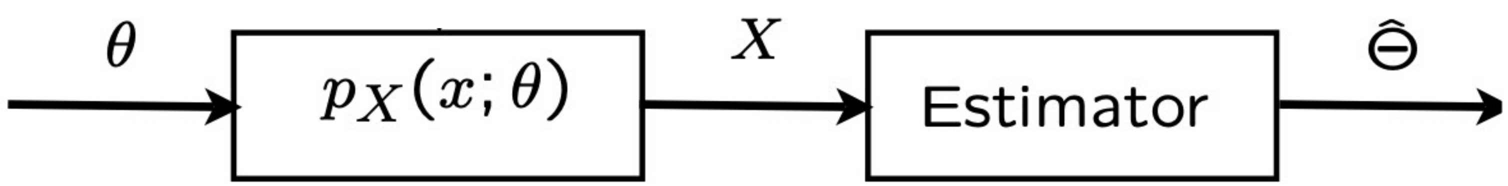
θ: unknown parameter (not a r.v.)

- E.g., θ = mass of electron

Example: we observe 10 coin flips from a biased coin, what is the bias of the coin θ?

Types of Data Analysis

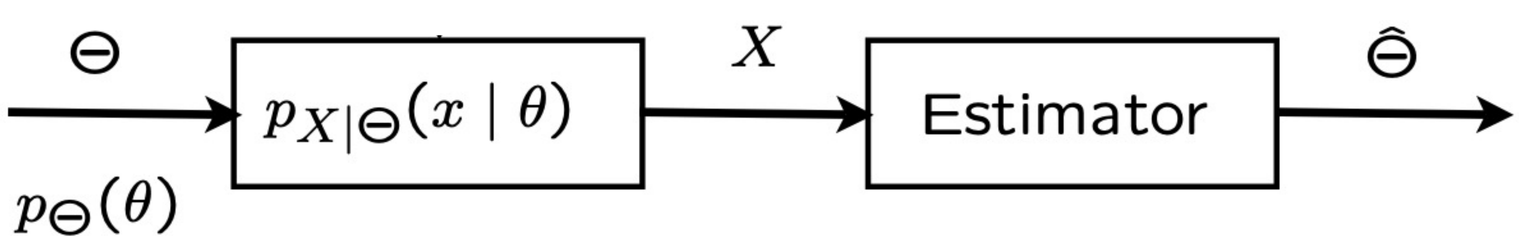
- Frequentist/Classical Data Analysis



θ: unknown parameter (not a r.v.)

- E.g., $\theta =$ mass of electron

- Bayesian Data Analysis



Parameter itself is random!

Bayesian vs Classical

- Classical: Say we have a model with mass of electron as parameter. We might not know the value, but it is nonetheless a constant.

- Bayesian: If we do not know its value completely, use a prior distribution reflecting what we do know.

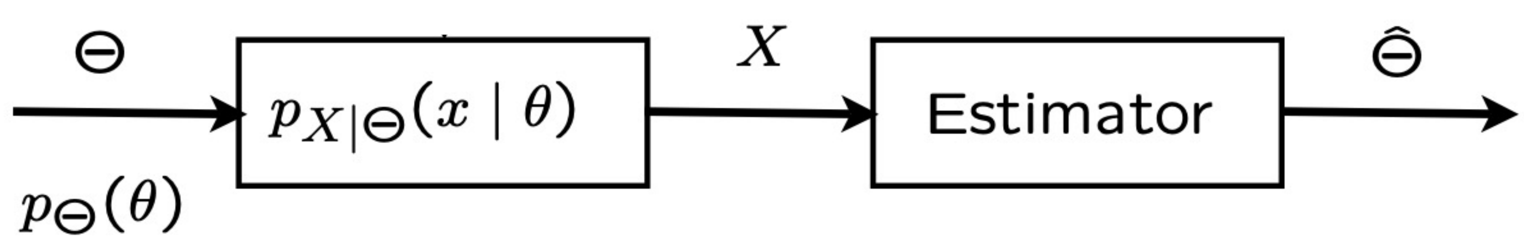
- Classical: Prior Distribution seems arbitrary.

- Bayesian: Every statistical method makes some choice* ; might as well use a prior to codify these choices.

- Classical: Bayesian methods too difficult to compute (practical considerations)

Bayesian Inference

- We start with a prior distribution p_Θ or f_Θ for the unknown random variable Θ.
- We have a model $p_{X|\Theta}$ or $f_{X|\Theta}$ of the observation vector X.
- After observing the value x of X, we form the posterior distribution of Θ, using the appropriate version of Bayes' rule.



Recall: Your first consulting job

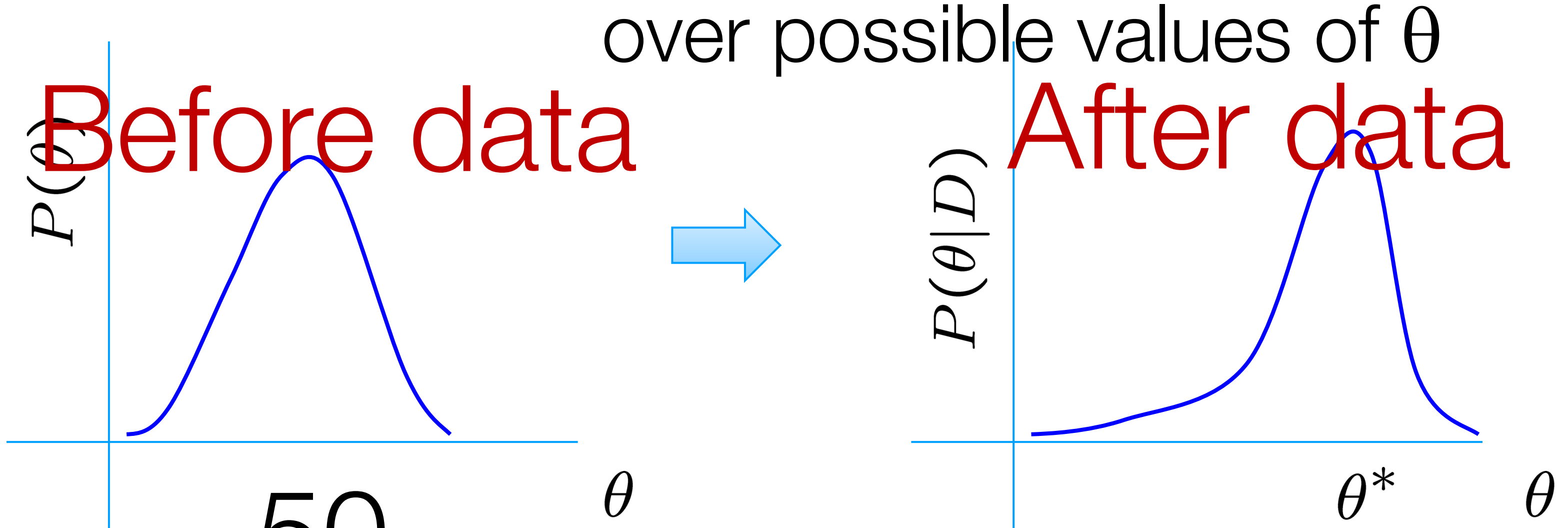
- A billionaire from the suburbs of Seattle asks you a question:
  - He says: I have a coin, if I flip it, what's the probability it will fall with the head up?
  - You say: Please flip it a few times:

  - You say: The probability is: **3/5** because… frequency of heads in all flips
  - **He says: But can I put money on this estimate?**
  - You say: ummm…. Maybe not.
    - Not enough flips (less than sample complexity)

What about prior knowledge?

- Billionaire says: Wait, I know that the coin is “close” to 50-50. What can you do for me now?
  - You say: I can learn it the Bayesian way…

- Rather than estimating a single θ, we obtain a distribution over possible values of θ



50
-
50

Bayesian Learning

- Use Bayes rule:

likelihood prior

$$P(\theta \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \theta) P(\theta)}{P(\mathcal{D})}$$

Para
met
ers

Data

Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53:370-418**

Bayesian Learning

- Use Bayes rule:

$$P(\theta \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \theta)P(\theta)}{P(\mathcal{D})}$$

- Or equivalently:

$$P(\theta \mid \mathcal{D}) \propto P(\mathcal{D} \mid \theta)P(\theta)$$

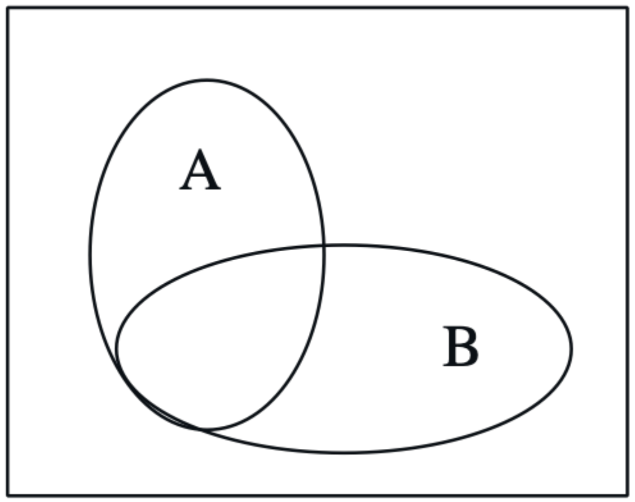
posterior likelihood prior

Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London,* **53:370-418**

Background: Bayes Rule

- Given the importance of the Bayes rule for Bayesian data analysis, let us review some key background on conditional probability, total probability theorem, and the Bayes Rule

Conditional Probability

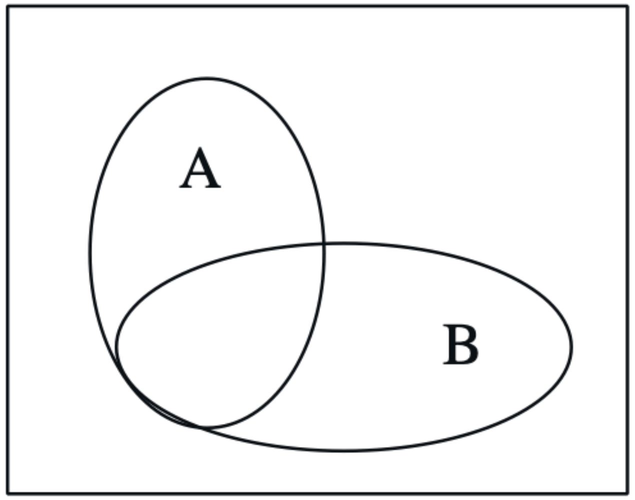


Outcome of experiment is within some event B.

What is the probability it is also in some event A?

- **P**(A | B) = conditional probability of event A, given an event B

Conditional Probability



Outcome of experiment is within some event B.

What is the probability it is also in some event A?

- **P**(A | B) = conditional probability of event A, given an event B

- **Definition:** Assuming P(B) not equal to zero,

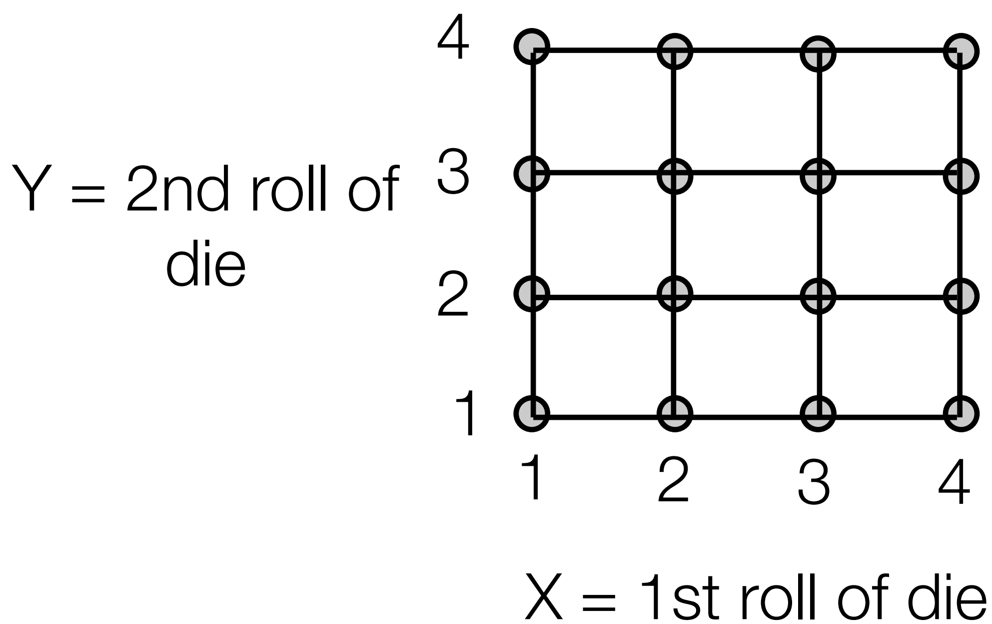
$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

$$\mathbf{P}(A \mid B) \text{ undefined if } \mathbf{P}(B) = 0$$

Example: Die Rolls

- In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is even. How likely is it that the first roll was a 6?
  - B = event that sum of two rolls is even = {(1,1), (1,3), (1,5), (2,2), ...}
    - There are 18 outcomes in B
  - Among outcomes in B, first roll is a six = {(6,2),(6,4),(6,6)}
    - There are 3 such outcomes
  - Intuitively :: P(A | B) should be equal to 3/18 = 1/6
  - Verify that the definition in previous page gives the same answer
    - i.e. P(A|B) = P(A intersection B)/P(B) = 1/6

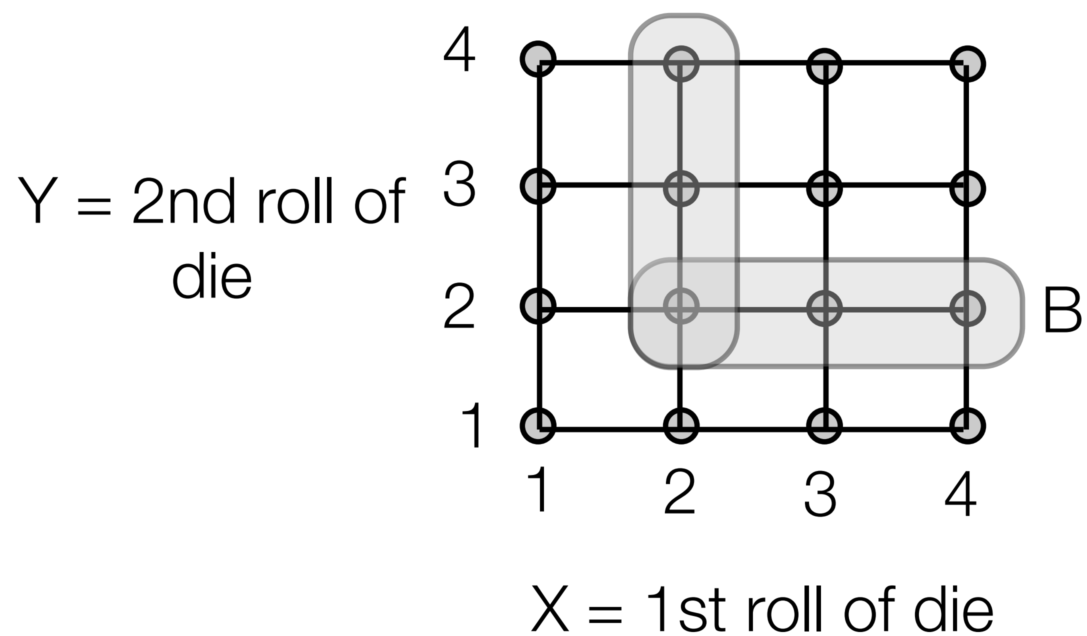
Example I



All outcomes are equally likely
Probability = 1/16

- Let B be the event: min(X,Y) = 2

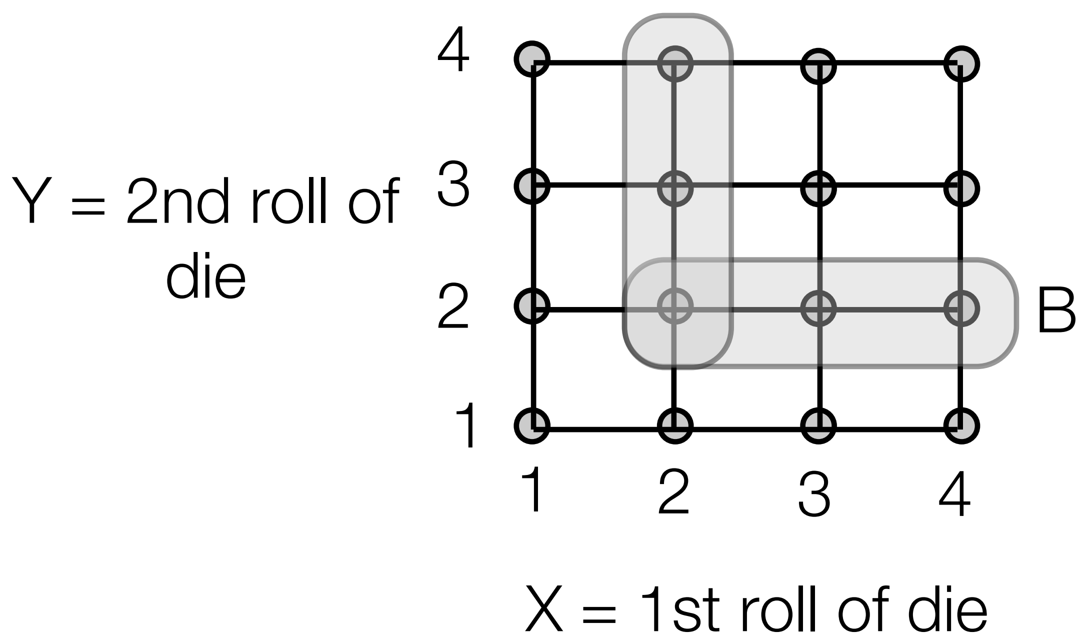
Example I



All outcomes are equally likely
Probability = 1/16

- Let B be the event: min(X,Y) = 2
- Let M = max(X,Y)
- P(M = 1 | B) = ... Count the number of outcomes in intersection
- P(M = 2 | B) =

Example I



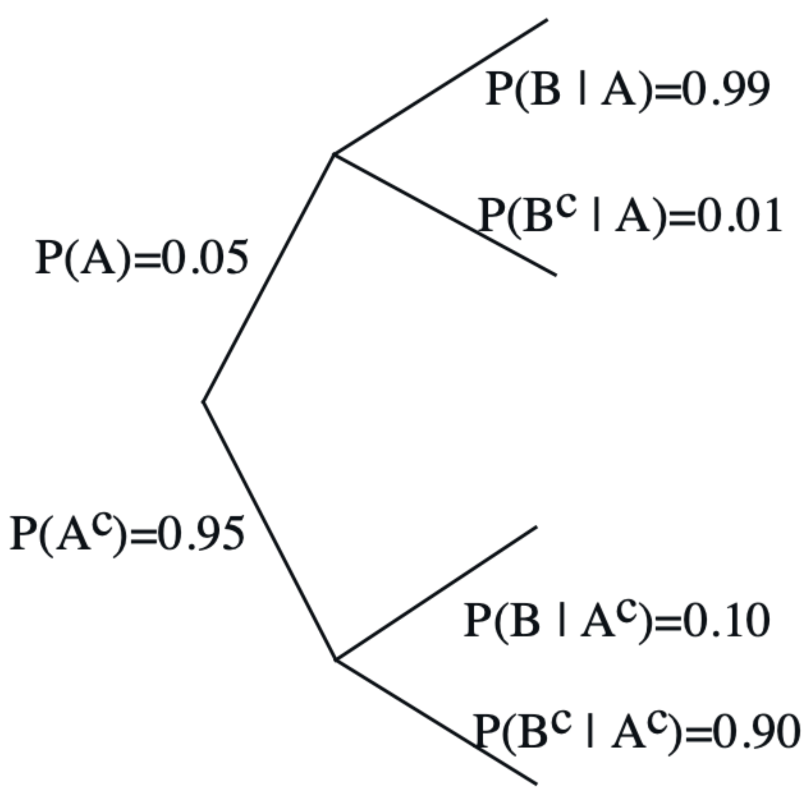
All outcomes are equally likely
Probability = 1/16

- Let B be the event: min(X,Y) = 2
- Let M = max(X,Y)
- P(M = 1 | B) = 0/5
- P(M = 2 | B) = 1/5

Models based on Conditional Probabilities

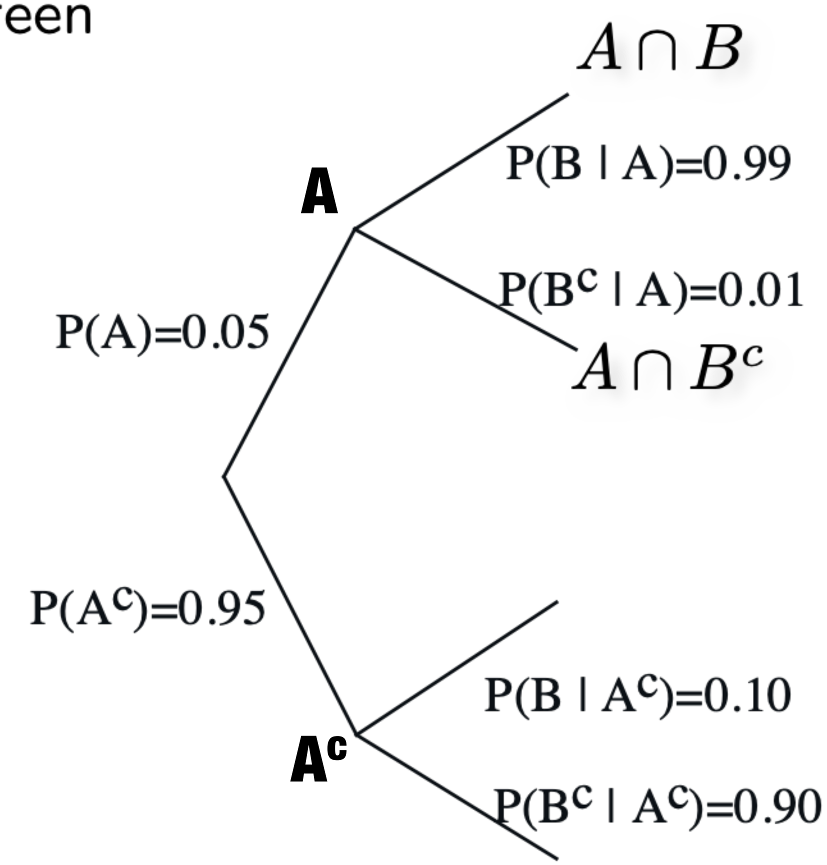
- Event A: Airplane is flying above
  Event B: Something registers on radar screen

A^c = no airplane
B^c = nothing on radar screen



Models based on Conditional Probabilities

- Event A: Airplane is flying above
  Event B: Something registers on radar screen

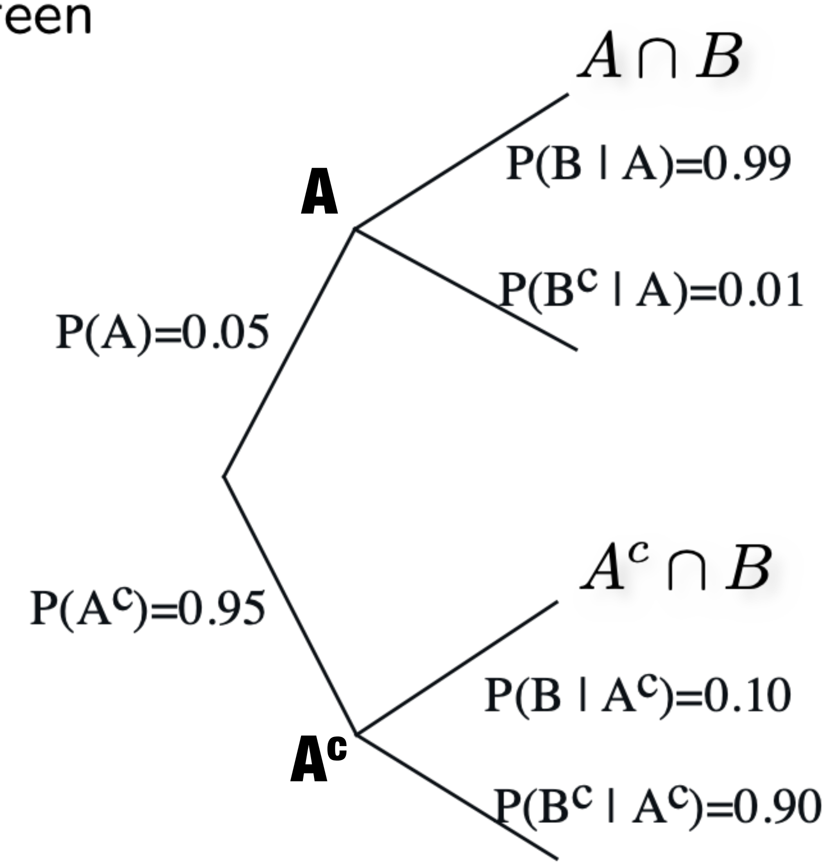


$\mathbf{P}(A \cap B) =$

$\mathbf{P}(A \mid B) =$

Models based on Conditional Probabilities

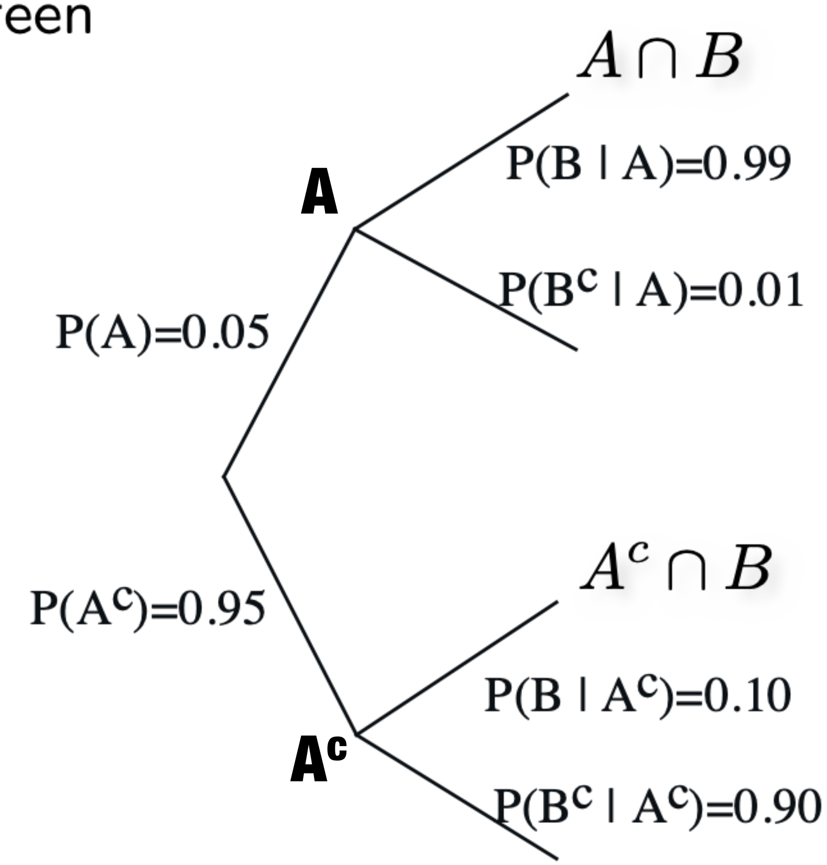
- Event A: Airplane is flying above
  Event B: Something registers on radar screen



$$\mathbf{P}(A \cap B) = \quad P(A)\ P(B|A) = 0.05 \times 0.99 = 0.0495$$

Models based on Conditional Probabilities

- Event A: Airplane is flying above
  Event B: Something registers on radar screen



$$\mathbf{P}(A \cap B) = \quad P(A)\; P(B|A) = 0.05 \times 0.99 = 0.0495$$

$$\mathbf{P}(A \mid B) = \quad \frac{P(A \cap B)}{P(B)} = \frac{0.0495}{0.1445} = 0.342$$

Example: Cards and Hearts

- Three cards are drawn from an ordinary 52-card deck without replacement (drawn cards are not placed back in the deck). We wish to find the probability that none of the three cards is a heart.
  - One approach is to use counting :: count the number of all card triplets that do not include a heart, and divide it with the number of all possible card triplets. But this is cumbersome.
  - We will use a stage-wise description of the experiment, and use the multiplication rule.

Example: Cards and Hearts

- Three cards are drawn from an ordinary 52-card deck without replacement (drawn cards are not placed back in the deck). We wish to find the probability that none of the three cards is a heart.

$$A_i = \{\text{the } i\text{th card is not a heart}\}, \qquad i = 1, 2, 3.$$

We will calculate $\mathbf{P}(A_1 \cap A_2 \cap A_3)$, the probability that none of the three cards is a heart, using the multiplication rule

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1)\mathbf{P}(A_3 \mid A_1 \cap A_2).$$

Example: Cards and Hearts

- Three cards are drawn from an ordinary 52-card deck without replacement (drawn cards are not placed back in the deck). We wish to find the probability that none of the three cards is a heart.

$$A_i = \{\text{the } i\text{th card is not a heart}\}, \qquad i = 1, 2, 3.$$

We will calculate $\mathbf{P}(A_1 \cap A_2 \cap A_3)$, the probability that none of the three cards is a heart, using the multiplication rule

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1)\mathbf{P}(A_3 \mid A_1 \cap A_2).$$

$\mathbf{P}(A_1) = \frac{39}{52}$, ... 39 non-heart cards in 52 card deck

$\mathbf{P}(A_2 \mid A_1) = \frac{38}{51}$. .... 38 non-heart cards in remaining 51 cards

$\mathbf{P}(A_3 \mid A_1 \cap A_2) = \frac{37}{50}$.

$$\boxed{\mathbf{P}(A_1 \cap A_2 \cap A_3) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}.}$$

Example: Defective Items

- A batch of one hundred items is inspected by testing four randomly selected items. If one of the four is defective, the batch is rejected. What is the probability that the batch is accepted if it contains five defectives?

Example: Defective Items

- A batch of one hundred items is inspected by testing four randomly selected items. If one of the four is defective, the batch is rejected. What is the probability that the batch is accepted if it contains five defectives?
    - A = event that batch is accepted
    - A_i = event that i-th item is not defective
    - $A = A_1 \cap A_2 \cap A_3 \cap A_4$

Example: Defective Items

- A batch of one hundred items is inspected by testing four randomly selected items. If one of the four is defective, the batch is rejected. What is the probability that the batch is accepted if it contains five defectives?
  - A = event that batch is accepted
  - A_i = event that i-th item is not defective $A = A_1 \cap A_2 \cap A_3 \cap A_4$
  - $A = A_1 \cap A_2 \cap A_3 \cap A_4$
  - Multiplication Rule:

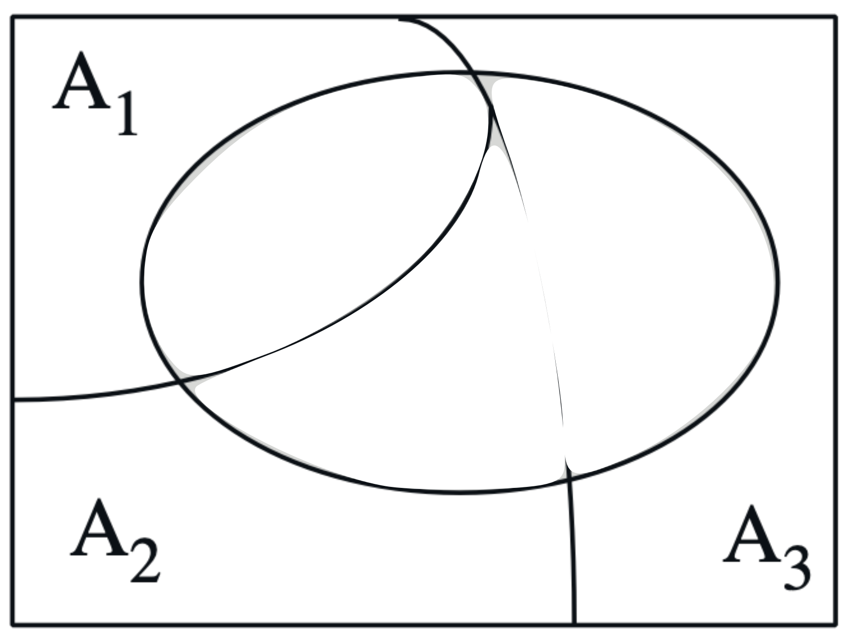
$$\mathbf{P}(A) = \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1)\mathbf{P}(A_3 \mid A_1 \cap A_2)\mathbf{P}(A_4 \mid A_1 \cap A_2 \cap A_3) = \frac{95}{100} \cdot \frac{94}{99} \cdot \frac{93}{98} \cdot \frac{92}{97} = 0.812.$$

Total Probability Theorem & Bayes Rule

- Bayesian Data Analysis relies on application of the Bayes Rule
- Let us now recap the Bayes Rule and the Total Probability Theorem that the Bayes Rule relies on

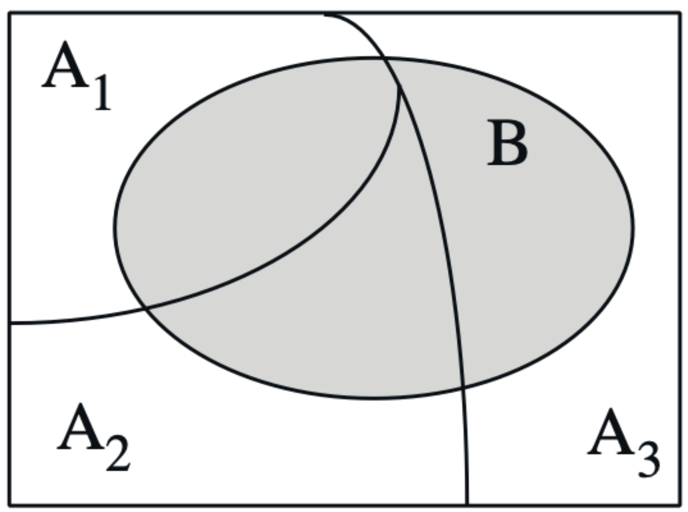
Total Probability Theorem

- Suppose we have a partition of sample space into A_1, A_2, A_3
- And that we are given $P(A_1)$, $P(A_2)$, $P(A_3)$



Total Probability Theorem

- We are given P(A1), P(A2), P(A3)
- Have $\mathbf{P}(B \mid A_i)$, for every i (for some event B)

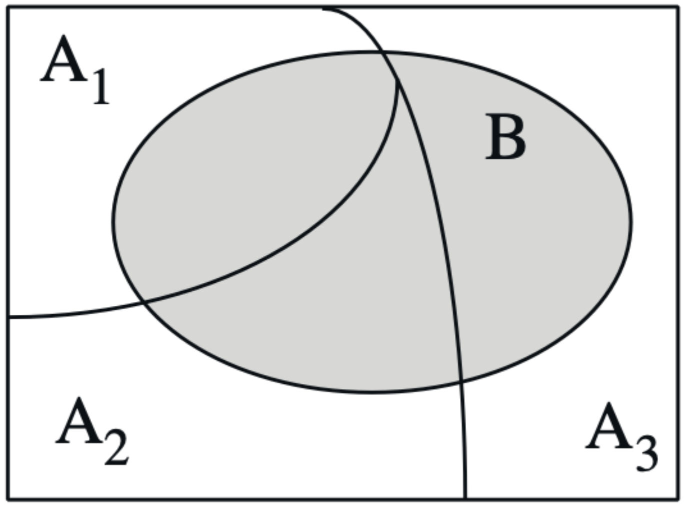


- One way of computing $\mathbf{P}(B)$:

$$\begin{aligned}\mathbf{P}(B) = \quad & \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) \\ + \; & \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) \\ + \; & \mathbf{P}(A_3)\mathbf{P}(B \mid A_3)\end{aligned}$$

Total Probability Theorem

- We are given $P(A_1)$, $P(A_2)$, $P(A_3)$
- Have $\mathbf{P}(B \mid A_i)$, for every i (for some event B)

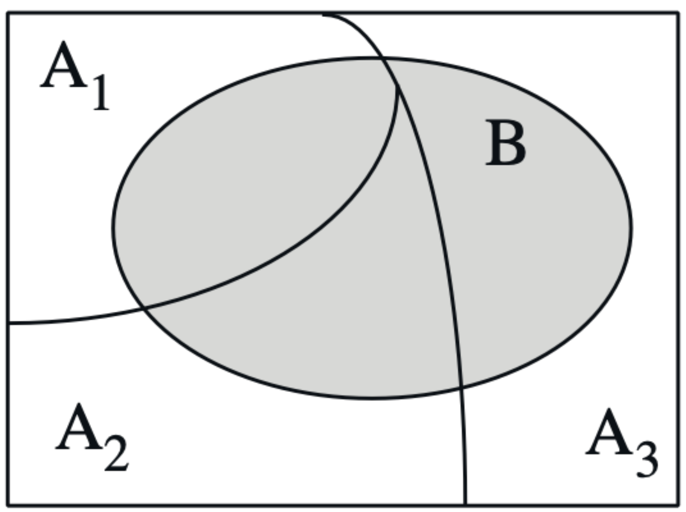


- One way of computing $\mathbf{P}(B)$:

$$\begin{aligned}\mathbf{P}(B) = \quad & \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) \\ + \; & \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) \\ + \; & \mathbf{P}(A_3)\mathbf{P}(B \mid A_3)\end{aligned}$$

Total Probability Theorem

- We are given $P(A_1)$, $P(A_2)$, $P(A_3)$
- Have $\mathbf{P}(B \mid A_i)$, for every i (for some event B)



Scenarios A_i

Probabilities of event B under each scenario

Overall probability of B is a weighted average!

- One way of computing $\mathbf{P}(B)$:

$$\begin{aligned}\mathbf{P}(B) = \quad & \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) \\ + \; & \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) \\ + \; & \mathbf{P}(A_3)\mathbf{P}(B \mid A_3)\end{aligned}$$

Example I

- You enter a chess tournament where your probability of winning a game is
  - 0.3 against half the players (call them type 1),
  - 0.4 against a quarter of the players (call them type 2), and
  - 0.5 against the remaining quarter of the players (call them type 3).
  - You play a game against a randomly chosen opponent. What is the probability of winning?

Example I

- You enter a chess tournament where your probability of winning a game is
  - 0.3 against half the players (call them type 1), 0.4 against a quarter of the players (call them type 2), and 0.5 against the remaining quarter of the players (call them type 3).
  - You play a game against a randomly chosen opponent. What is the probability of winning?

- Let A_i be the event of playing with opponent of type i.

$$\mathbf{P}(A_1) = 0.5, \qquad \mathbf{P}(A_2) = 0.25, \qquad \mathbf{P}(A_3) = 0.25.$$

Example I

- You enter a chess tournament where your probability of winning a game is
  - 0.3 against half the players (call them type 1), 0.4 against a quarter of the players (call them type 2), and 0.5 against the remaining quarter of the players (call them type 3).
  - You play a game against a randomly chosen opponent. What is the probability of winning?

- Let A_i be the event of playing with opponent of type i.

$$\mathbf{P}(A_1) = 0.5, \qquad \mathbf{P}(A_2) = 0.25, \qquad \mathbf{P}(A_3) = 0.25.$$

- Let B be event of winning.

$$\mathbf{P}(B \mid A_1) = 0.3, \qquad \mathbf{P}(B \mid A_2) = 0.4, \qquad \mathbf{P}(B \mid A_3) = 0.5.$$

Example I

- You enter a chess tournament where your probability of winning a game is
  - 0.3 against half the players (call them type 1), 0.4 against a quarter of the players (call them type 2), and 0.5 against the remaining quarter of the players (call them type 3).
  - You play a game against a randomly chosen opponent. What is the probability of winning?

- Let A_i be the event of playing with opponent of type i.

$$\mathbf{P}(A_1) = 0.5, \qquad \mathbf{P}(A_2) = 0.25, \qquad \mathbf{P}(A_3) = 0.25.$$

- Let B be event of winning.

$$\mathbf{P}(B \,|\, A_1) = 0.3, \qquad \mathbf{P}(B \,|\, A_2) = 0.4, \qquad \mathbf{P}(B \,|\, A_3) = 0.5.$$

$$\boxed{\begin{aligned}\mathbf{P}(B) &= \mathbf{P}(A_1)\mathbf{P}(B \,|\, A_1) + \mathbf{P}(A_2)\mathbf{P}(B \,|\, A_2) + \mathbf{P}(A_3)\mathbf{P}(B \,|\, A_3) \\ &= 0.5 \cdot 0.3 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5 \\ &= 0.375.\end{aligned}}$$

.... Total Probability Theorem

Example II

- You roll a fair four-sided die. If the result is 1 or 2, you roll once more but otherwise, you stop. What is the probability that the sum total of your rolls is at least 4?
  - Let A_i be the event that first roll is i; P(A_i) = 1/4
  - B be the event that sum >= 4
  - Given A_1 :: sum >= 4 iff second roll is 3 or 4
    - P(B|A_1) = 2/4 = 1/2
  - Given A_2 :: sum >= 4 iff second roll is 2, 3 or 4
    - P(B|A_1) = 3/4
  - Given A_3, A_4, you stop.
    - P(B|A_3) = 0, P(B|A_4) = 1

Example II

- You roll a fair four-sided die. If the result is 1 or 2, you roll once more but otherwise, you stop. What is the probability that the sum total of your rolls is at least 4?
    - Let A_i be the event that first roll is i; P(A_i) = 1/4
    - B be the event that sum <= 4

From previous slide ::

$$\mathbf{P}(B \mid A_1) = \frac{1}{2}, \qquad \mathbf{P}(B \mid A_2) = \frac{3}{4}, \qquad \mathbf{P}(B \mid A_3) = 0, \qquad \mathbf{P}(B \mid A_4) = 1.$$

Total Probability Theorem ::

$$\mathbf{P}(B) = \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 1 = \frac{9}{16}.$$

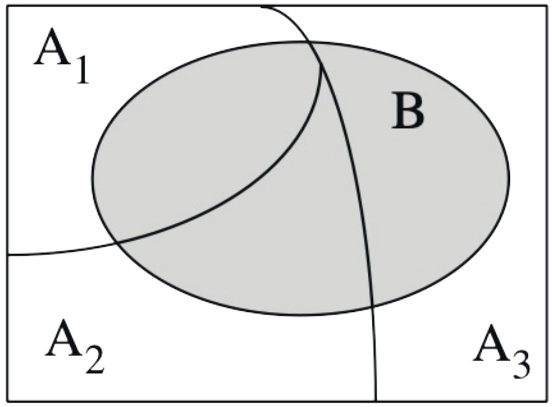
Example III: The Monty Hall Problem

- A prize is equally likely to be found behind any **one of three** closed doors in front of you. You point to one of the doors. The host opens for you **one of the remaining two** doors, after making sure that the prize is not behind it. At this point, you can (a) Stick to your initial choice, or (b) Switch to the other unopened door. Which is the best strategy?
  - Label the door you picked as 1, and the other two doors as 2 and 3
  - P(1 has prize, and host opens 2) = P(1 has prize) P(host opens 2| 1 has prize) = 1/3 x 1/2 = 1/6
  - P(1 has prize, and host opens 3) = 1/3 x 1/2 = 1/6
  - P(2 has prize, and host opens 3) = 1/3 x 1 = 1/3
  - P(3 has prize, and host opens 2) = 1/3 x 1 = 1/3
  - P(Win if you stay at 1) = 1/6 + 1/6 = 1/3
  - P(Win if you switch) = 1/3 + 1/3 = 2/3!

Bayes Rule

- Partition of sample space into A_1, A_2, A_3 :: "Prior Beliefs" $P(A_1)$, $P(A_2)$, $P(A_3)$

- We know $\mathbf{P}(B \mid A_i)$ for each i (for some event B)
- Wish to compute $\mathbf{P}(A_i \mid B)$
  - revise "beliefs", given that B occurred

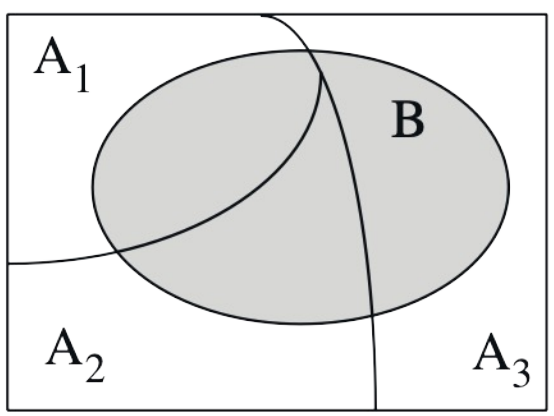


$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(B)}$$

Bayes Rule

- Partition of sample space into A_1, A_2, A_3 :: "Prior Beliefs" $P(A_1)$, $P(A_2)$, $P(A_3)$

- We know $\mathbf{P}(B \mid A_i)$ for each i (for some event B)
- Wish to compute $\mathbf{P}(A_i \mid B)$
  - revise "beliefs", given that B occurred



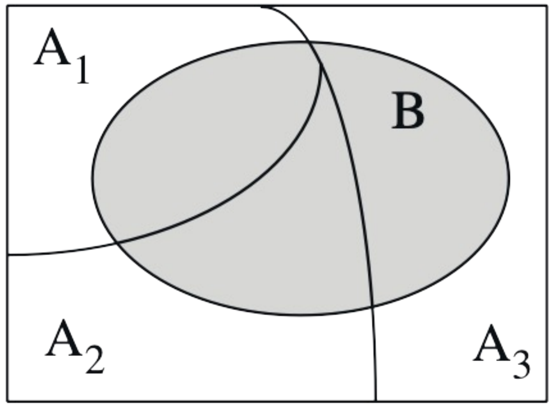
$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(B)}$$

$$= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)}$$

Bayes Rule

- Partition of sample space into A_1, A_2, A_3 :: "Prior Beliefs" $P(A_1)$, $P(A_2)$, $P(A_3)$

- We know $\mathbf{P}(B \mid A_i)$ for each i (for some event B)
- Wish to compute $\mathbf{P}(A_i \mid B)$
  - revise "beliefs", given that B occurred

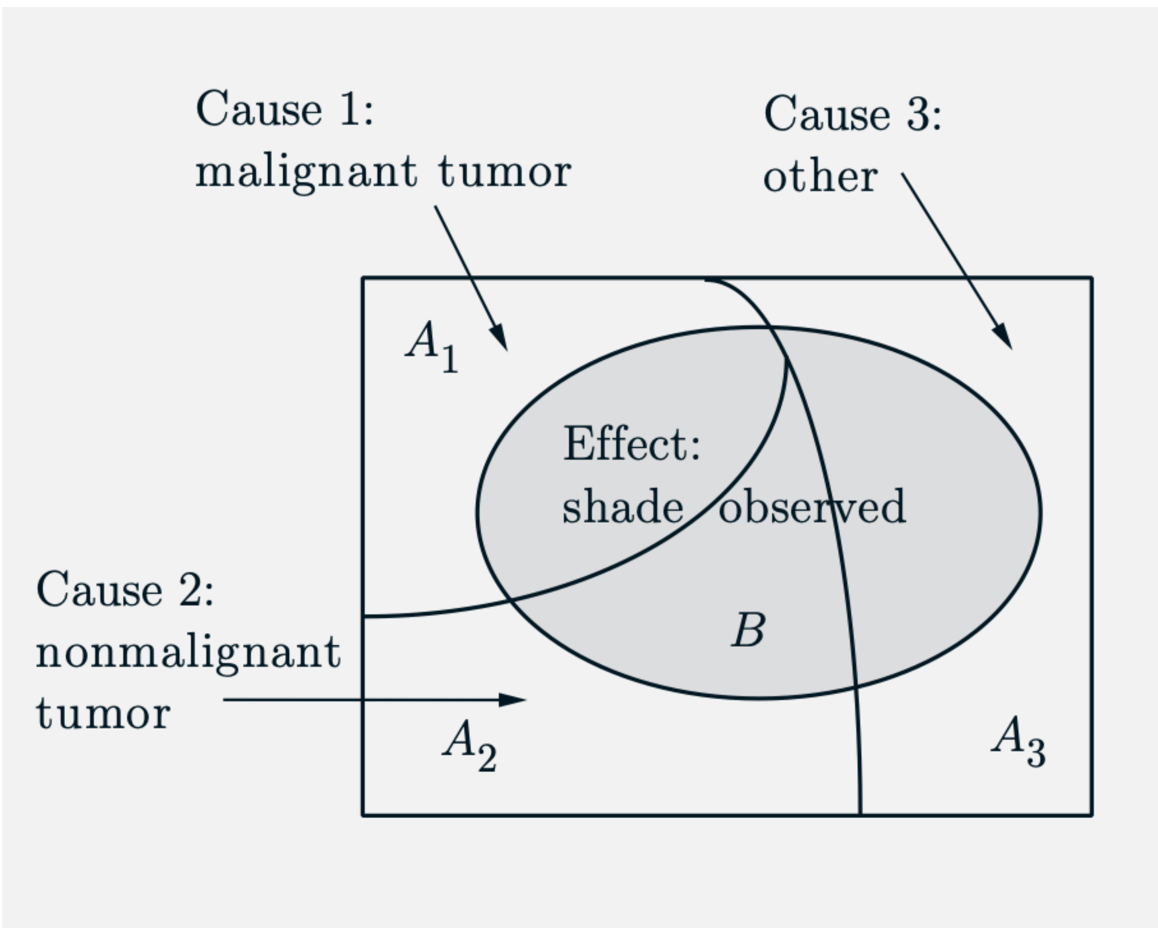


$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(B)}$$

$$= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)}$$

$$= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\sum_j \mathbf{P}(A_j)\mathbf{P}(B \mid A_j)}$$

Bayes Rule



$$
\begin{aligned}
\mathbf{P}(A_i \mid B) &= \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(B)} \\
&= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)} \\
&= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\sum_j \mathbf{P}(A_j)\mathbf{P}(B \mid A_j)}
\end{aligned}
$$

Given: (a) prior probabilities of causes,
(b) probability of effect given any of the causes

Obtain: probability of any of the causes given effect.