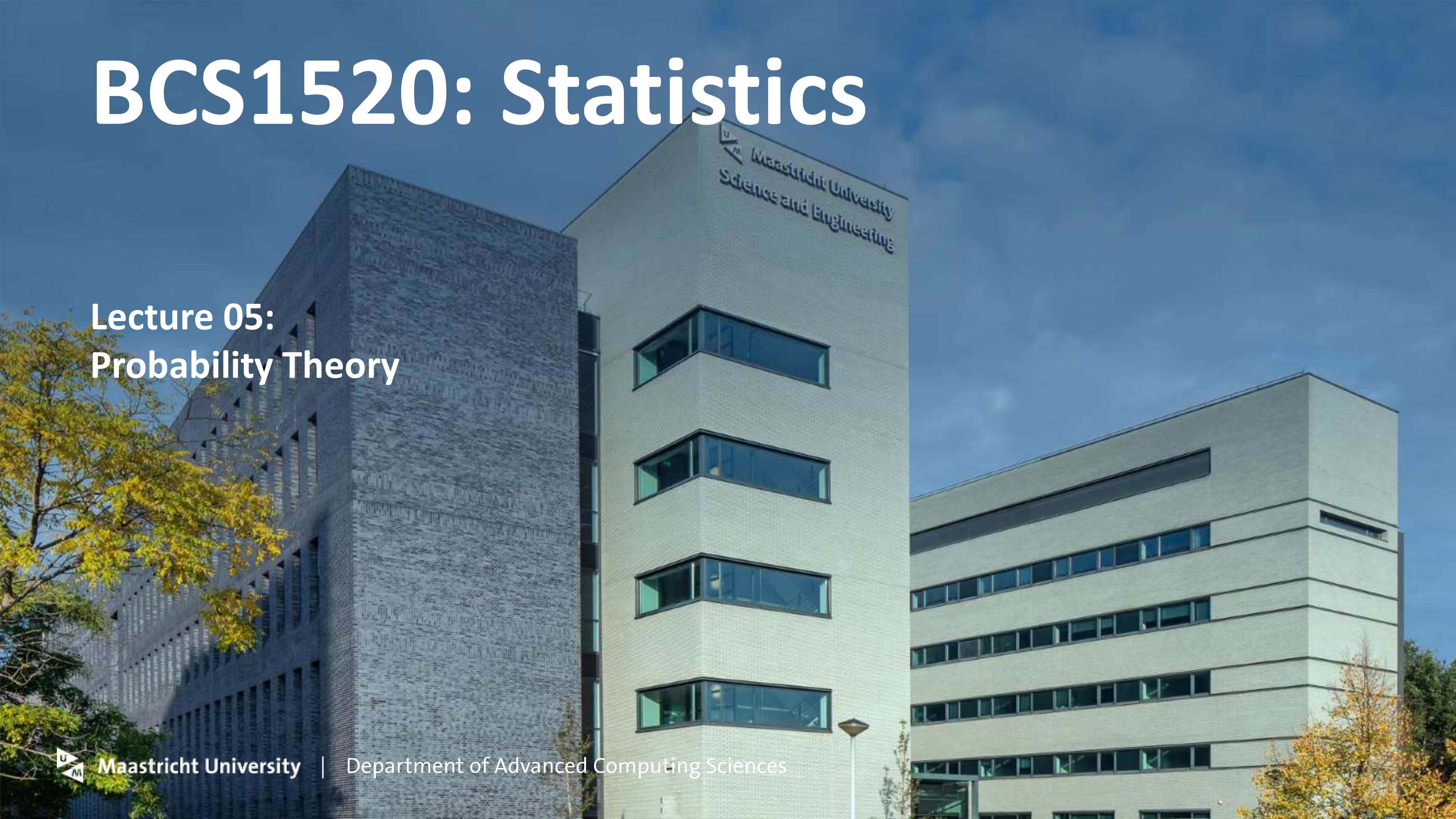
BCS1520: Statistics

Lecture 05:
Probability Theory



The origins of probability theory: Gambling

- Let’s bet.

- What odds would you give me for:
  - Getting a six when you throw a *fair* die at most 4 times?

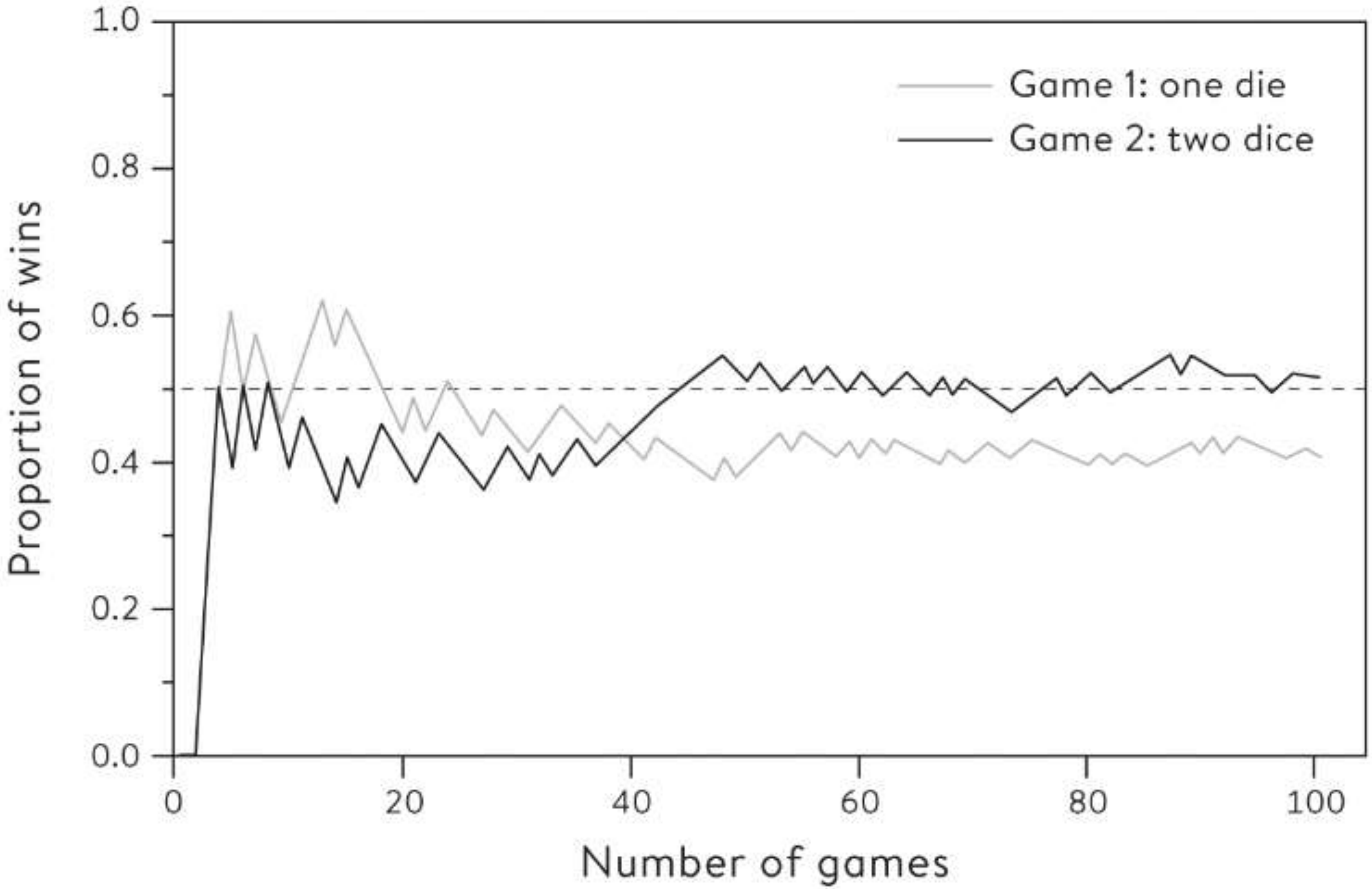
    vs
  - Getting two sixes when you throw a pair of dice 24 times?

- Which is more likely? Which one would you put your money on?

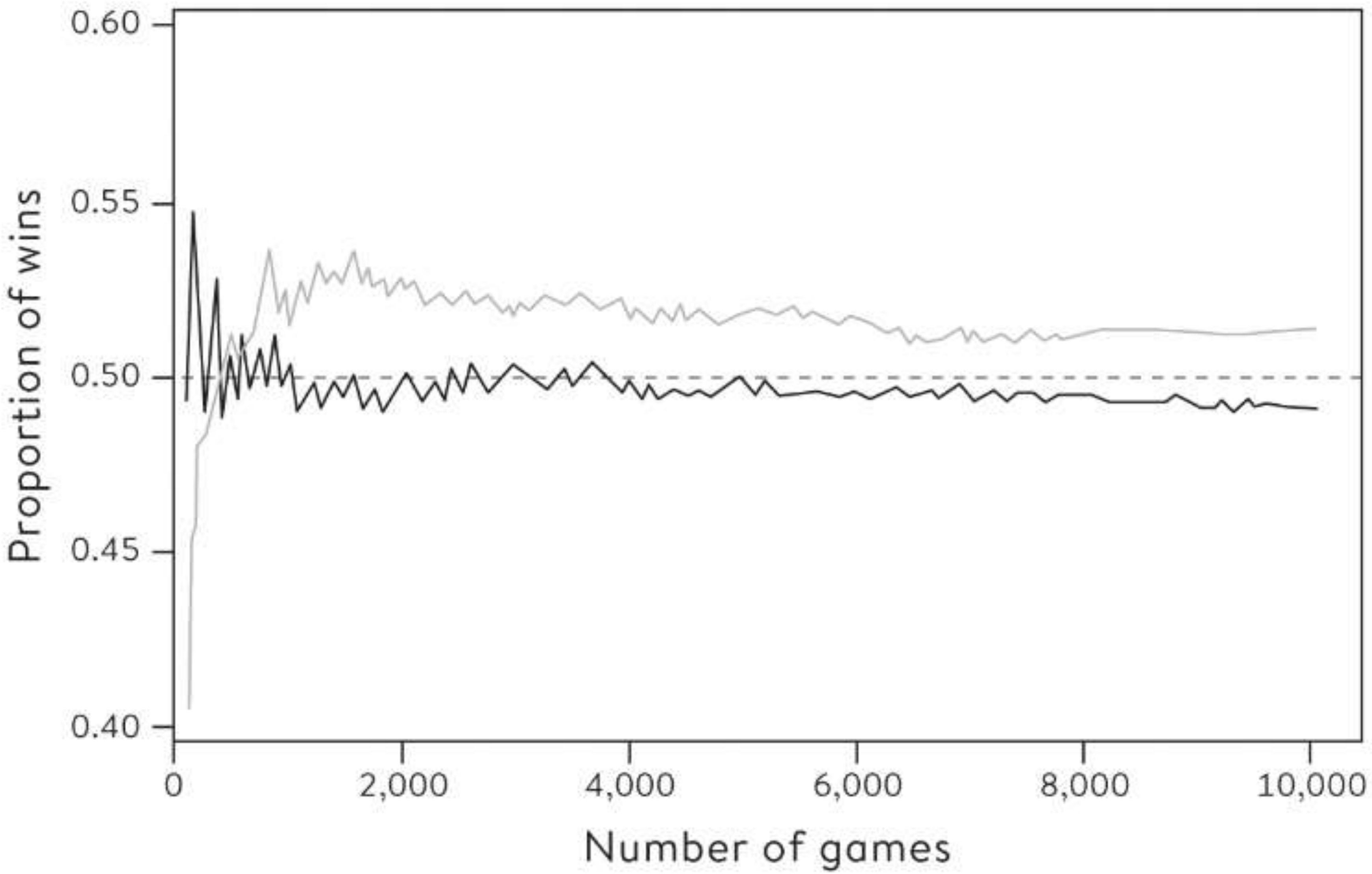
Chevalier de Mere's guess

- In Game 1, he thought since there seemed to be equal chance for the 6 always, so 1 in 6 times, then the probability was $4 * (1/6) = 2/3$.
- And for Game 2, he thought since the chance of two sixes at once is 1 in 36, then the probability was $24 * (1/36) = 2/3$, so that both games looked the same to him.
- Now what would you put your money on?

Simulated Probability



Simulated Probability: Longer Run Sims



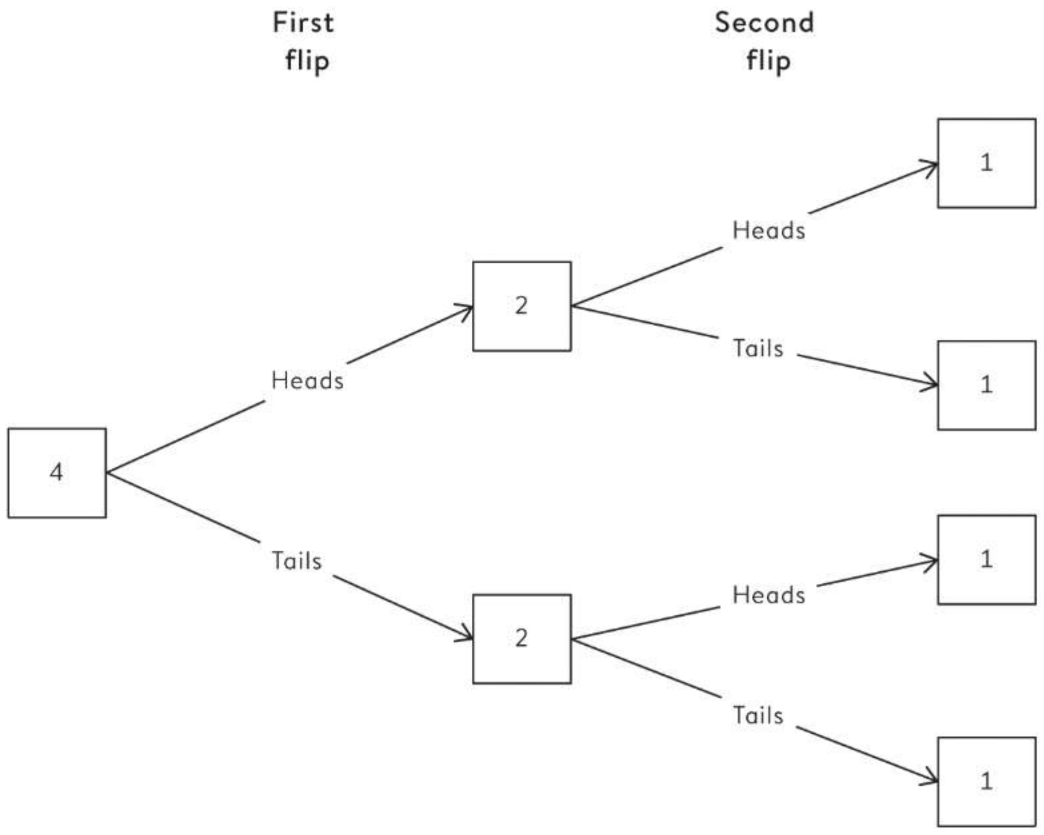
Probability is Tricky and Unintuitive

- How do you understand probability? Can you measure it out in the world like it is a real thing?

- We can measure distances, temperatures, voltages, where are the probabilities?

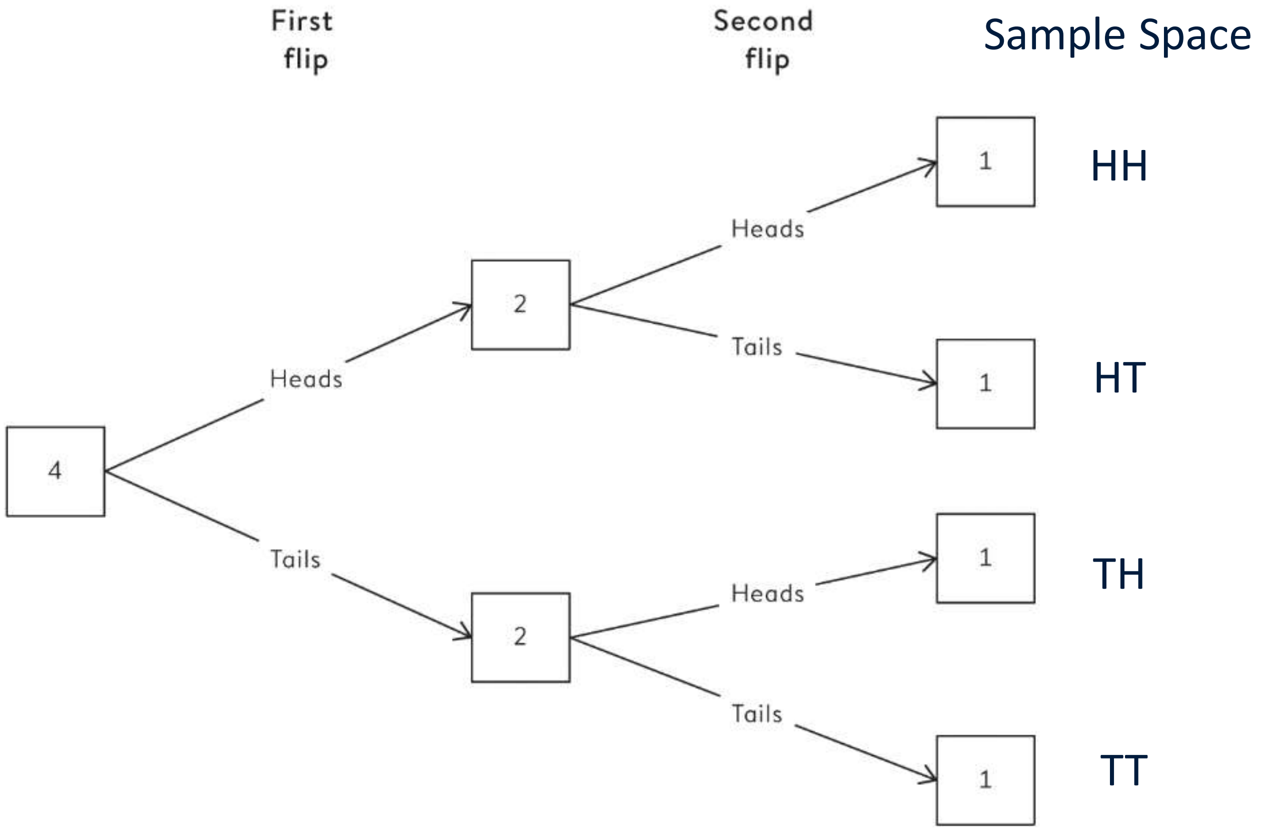
Thinking of probability as expected frequency

- What is the probability of 2 heads when flipping 2 coins?



Thinking of probability as expected frequency

- Can turn expected frequencies into a probability tree: How?



Some useful jargon

Experiment

An action whose outcome is determined by chance

Example: Flipping a coin

Sample Space

The set of possible outcomes of an experiment (denoted by S)

Example: the number of spots on each side of the die

$$S = \{HH, HT, TH, TT\}$$

Event

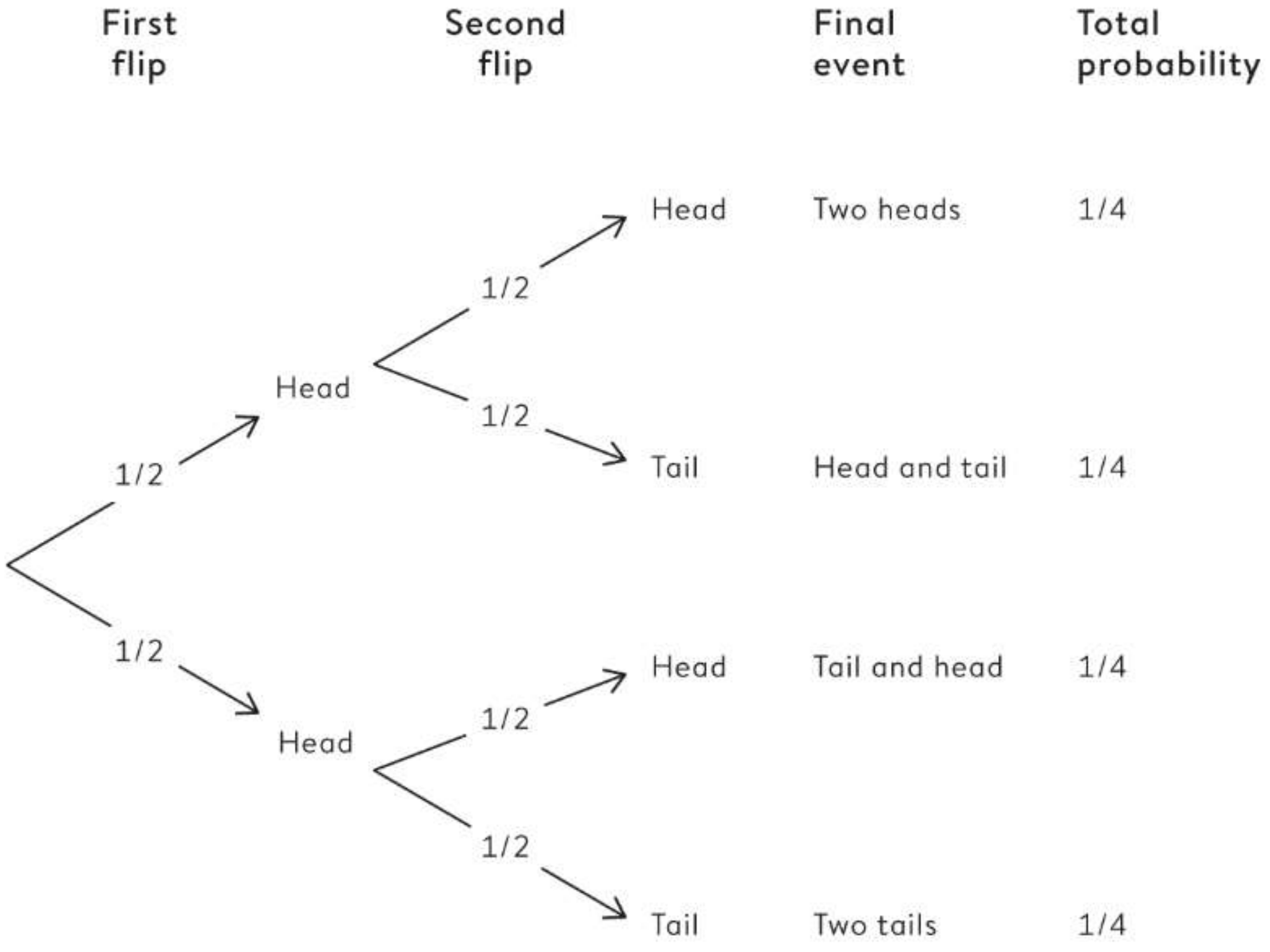
A subset of the sample space (denoted by $E \subseteq S$)

Example: the outcome of the experiment has two heads.

$$E = \{HH\}$$

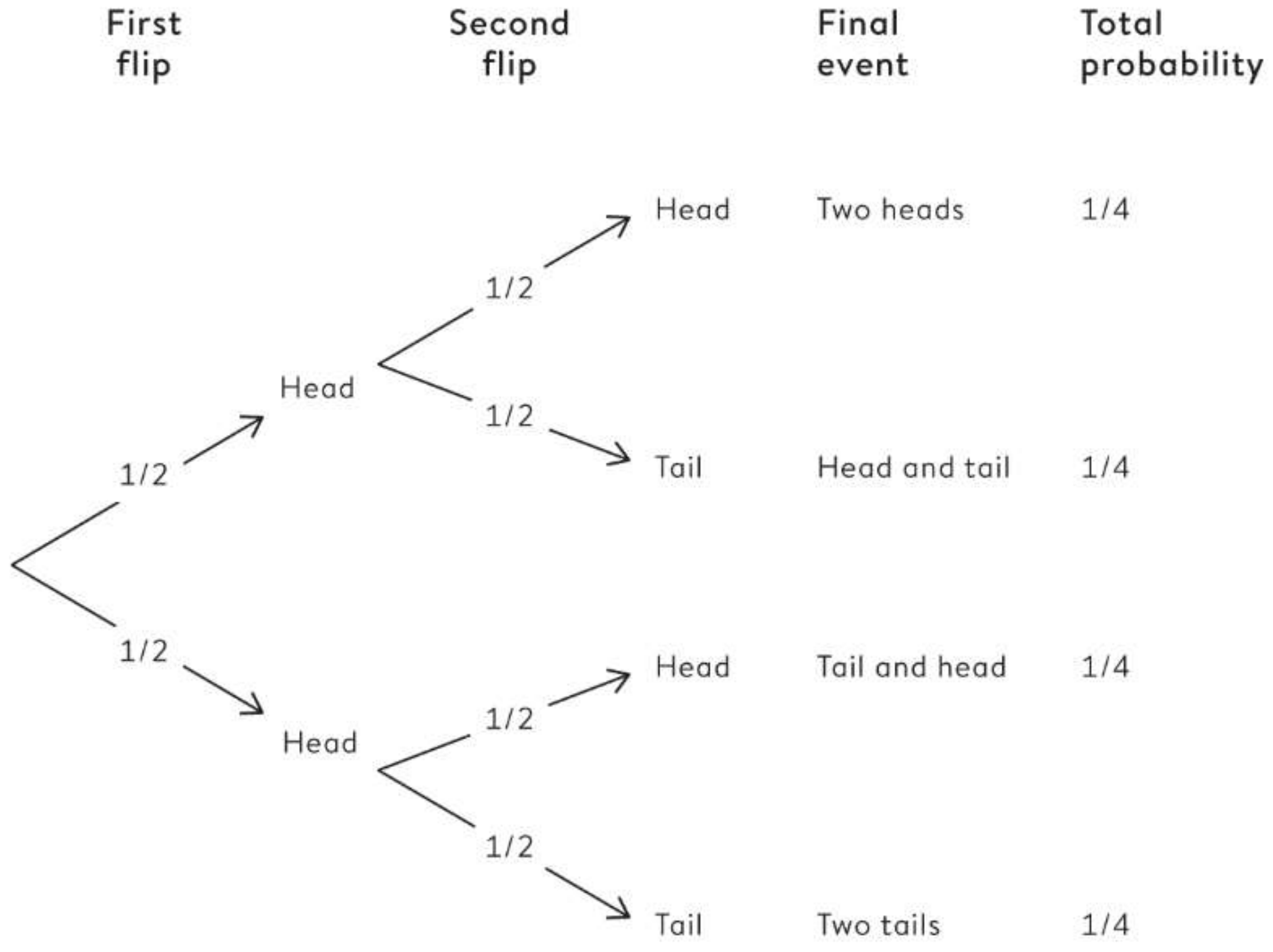
Thinking of probability as expected frequency

- Can turn expected frequencies into a probability tree
- Now, probability along any branch is multiplication of probabilities along it



Thinking of probability as expected frequency

- What if we first flipped a coin, then rolled a fair die?



Thinking of probability as expected frequency

Experiment: Draw a card out a full deck of cards

Sample space S: the cards in a full deck of cards.

$$S = \{♣2, ♣3, \dots, ♣\mathrm{A}, \Diamond 2, \dots, \Diamond\mathrm{A}, \heartsuit 2, \dots, \heartsuit\mathrm{A}, ♠2, \dots, ♠\mathrm{K}, ♠\mathrm{A}\}$$

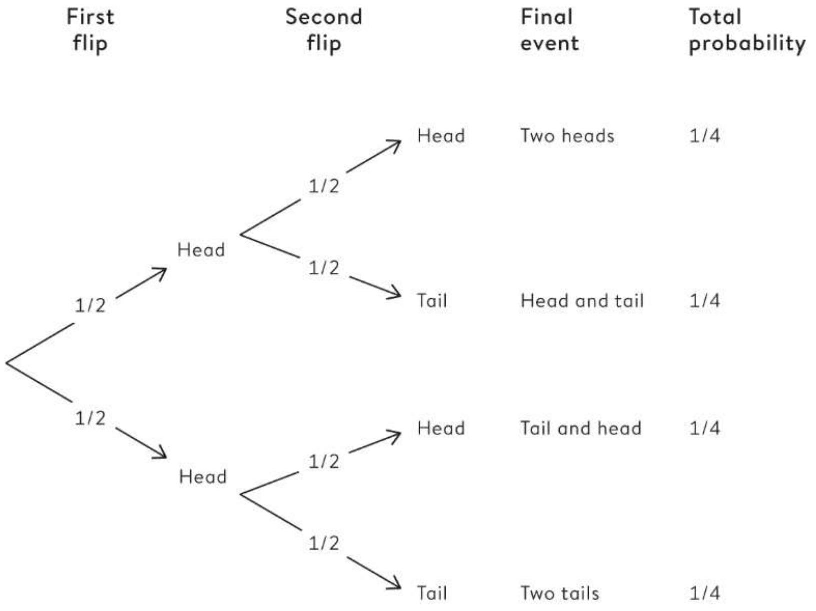
(52 outcomes)

Event E: the card drawn from the deck is a queen.

$$E = \{♣\mathrm{Q}, \Diamond\mathrm{Q}, \heartsuit\mathrm{Q}, ♠\mathrm{Q}\}$$

What rules of probability do we see from the tree?

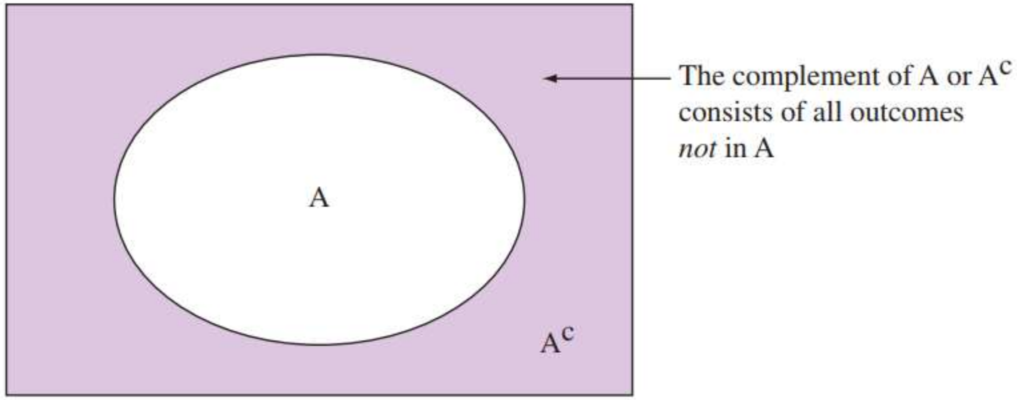
1. The probability of an *event E* (e.g. one head = {HT, TH}) is a number between 0 and 1. 0 is only for impossible events (e.g. 0 heads, 0 tails) and 1 for certain events (at least one head or tail).
   We represent this as $P(E) \in [0,1]$.



Events and their Complements

2. Given the sample space of events (flipping 2 coins), the probability of an event not happening is (1 - the probability of it happening). Probability of not two tails = 1 – ¼ = ¾ .

$$P(\sim A) = 1 - P(A)$$
$$P(A^c) = 1 - P(A)$$



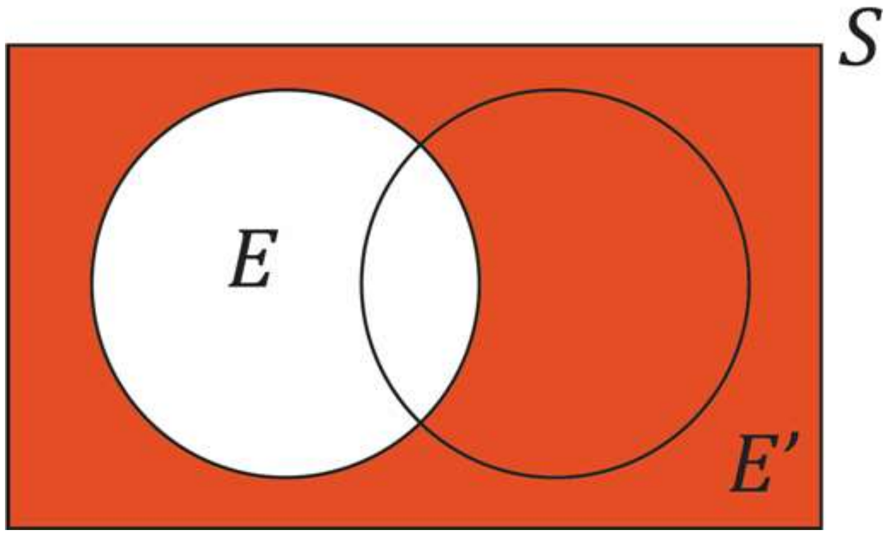
Events and their Complements

Complement of an Event E (denoted by E', or E^C)

Example: Rolling a single die

$$S = \{1,2,3,4,5,6\},$$
$$E = \{1,3,5\}$$

Figure: E' highlighted



Overlapping Event Probabilities

Intersection of Events E and F (denoted by $E \cap F$)

Example: Rolling a single die

$$S = \{1,2,3,4,5,6\},$$
$$E = \{1,3,5\}, F = \{3,6\}$$
$$E \cap F = \{3\}, P(E \cap F) = \frac{1}{6}$$

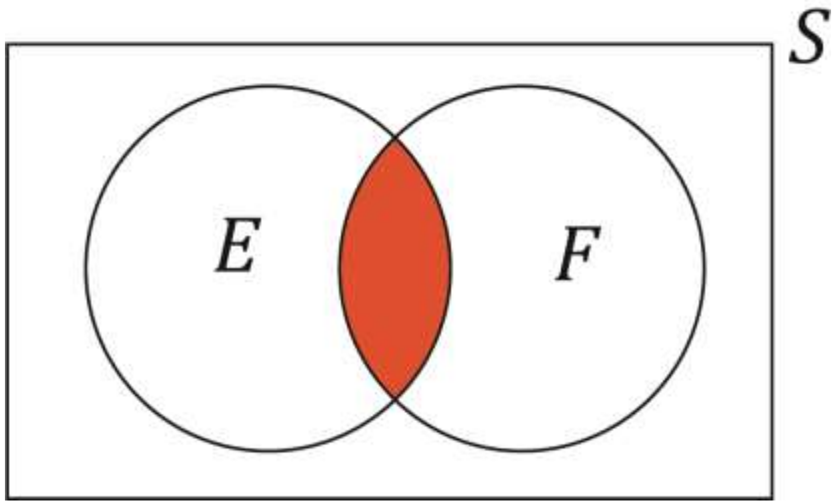


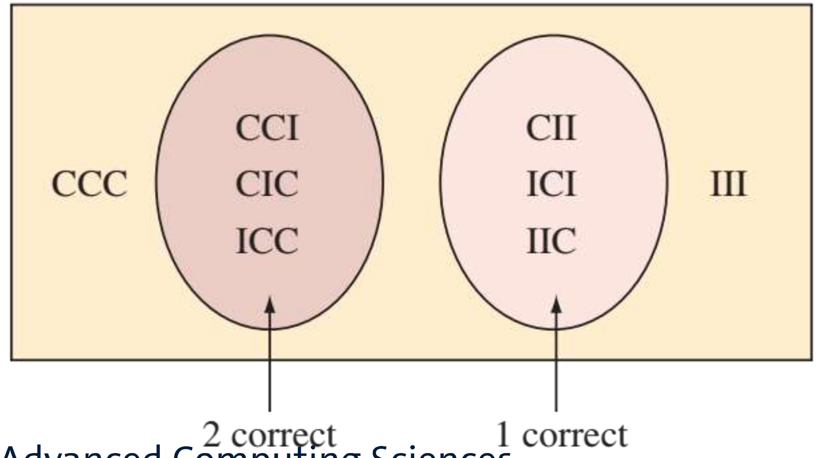
Figure: $E \cap F$ highlighted

Mutually exclusive/Disjoint Event Probabilities can be added (OR rule)

3. Probabilities of events that are mutually exclusive (no overlapping events) can be added together to get the total probability.

$P(E_1 \cap E_2) = \emptyset\ (Null\ event)$ then:

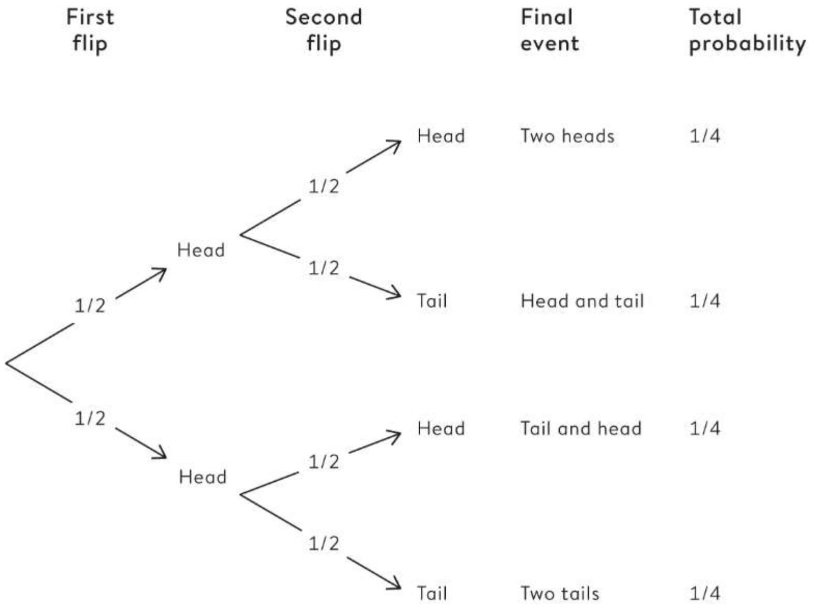
$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$



Independent event probabilities can be multiplied

4. Multiple probabilities for the total probability of a sequence of events (that are happening together).

$$P(HH) = P(H) * P(H) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$



Events are sets, consequently, we also have:

Union of Events E and F (denoted by $E \cup F$)

Example: Rolling a single die

$$S = \{1,2,3,4,5,6\},$$
$$E = \{1,3,5\}, F = \{3,6\}$$
$$E \cup F = \{1,3,5,6\}$$

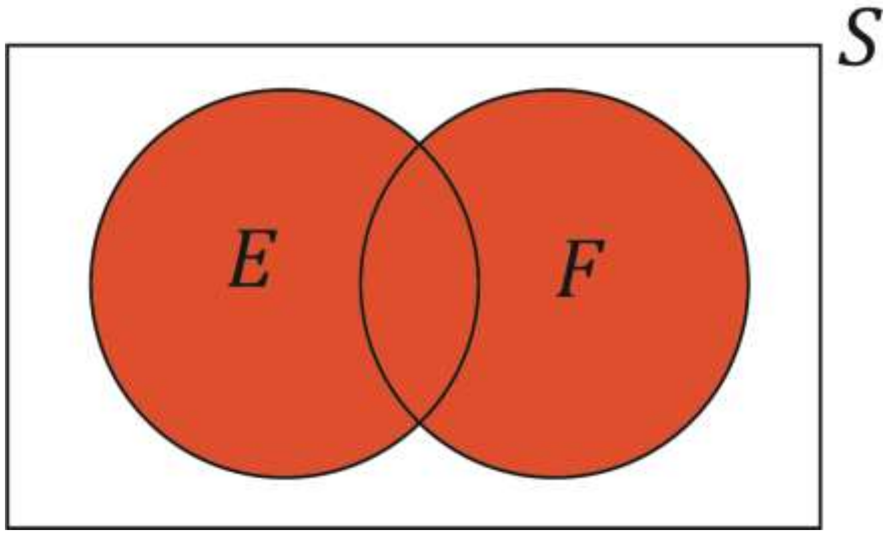
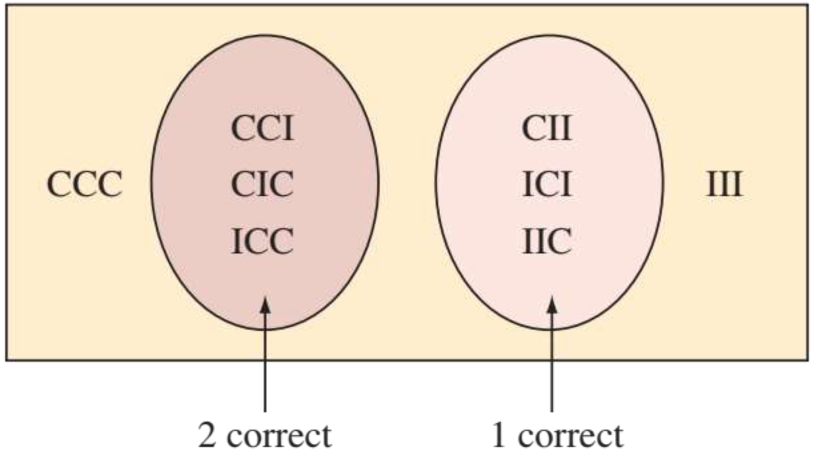
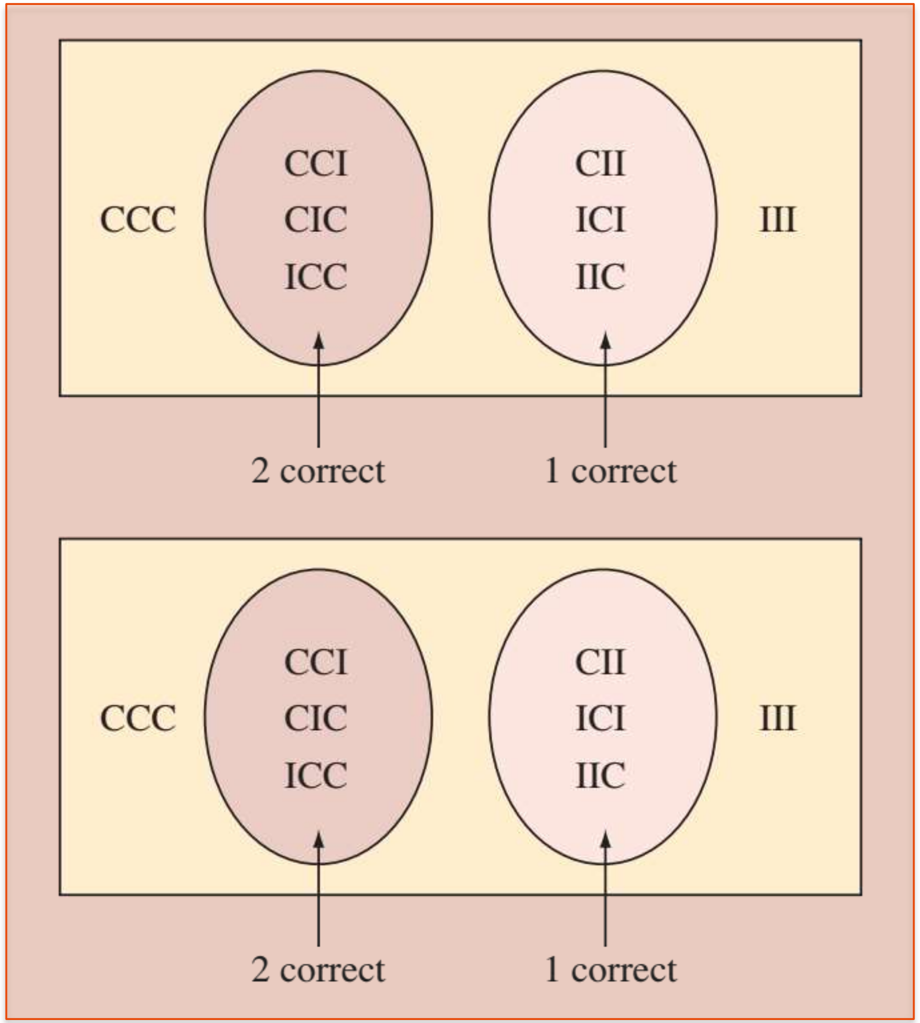


Figure: $E \cup F$ highlighted

Mutual Exclusion vs Independence

- What is the difference?





English	**Sets**
Events and occurrences	
sample space	S
s is a possible outcome	$s \in S$
A is an event	$A \subseteq S$
A occurred	$s_{\text{actual}} \in A$
something must happen	$s_{\text{actual}} \in S$

New events from old events	
A or B (inclusive)	$A \cup B$
A and B	$A \cap B$
not A	A^c
A or B, but not both	$(A \cap B^c) \cup (A^c \cap B)$
at least one of $A_1, \ldots, A_n$	$A_1 \cup \cdots \cup A_n$
all of $A_1, \ldots, A_n$	$A_1 \cap \cdots \cap A_n$

Relationships between events	
A implies B	$A \subseteq B$
A and B are mutually exclusive	$A \cap B = \emptyset$
$A_1, \dots, A_n$ are a partition of S	$A_1 \cup \cdots \cup A_n = S, A_i \cap A_j = \emptyset$ for $i \neq j$

The origins of probability theory: Gambling

- Can you now calculate what the optimal odds should be for the following? Assuming we played it 1000 times.

  Getting a six when you throw a *fair* die at most 4 times
  vs
  Getting two sixes when you throw a pair of dice 24 times?

Coin flipping is always biased!

Heads or Tails: Pure Chance?

18 oktober 2023

When you flip a coin, will it land more often on the same side it started? A well-known physics model suggests it will. Now, for the first time, scientists have gathered robust data to back up this hypothesis. They collected data from 350,757 coin tosses, including 12-hour coin-toss marathons. If you start with the head side up, the coin also more frequently ends up with the head side up.

Probability Theory Axioms

Probability function

A function that maps each possible event E to a number in $[0, 1]$.

Properties of probability functions:

1. $P(S) = 1$
2. $P(\emptyset) = 0$
3. $0 \leq P(E) \leq 1 \forall E \subseteq S$

Probability of an event

The probability of an event E is the sum of the probabilities of the elements in E (denoted by $P(E)$)

Probability functions (example)

A fair die

$$S = \{1,2,3,4,5,6\}, P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6}$$

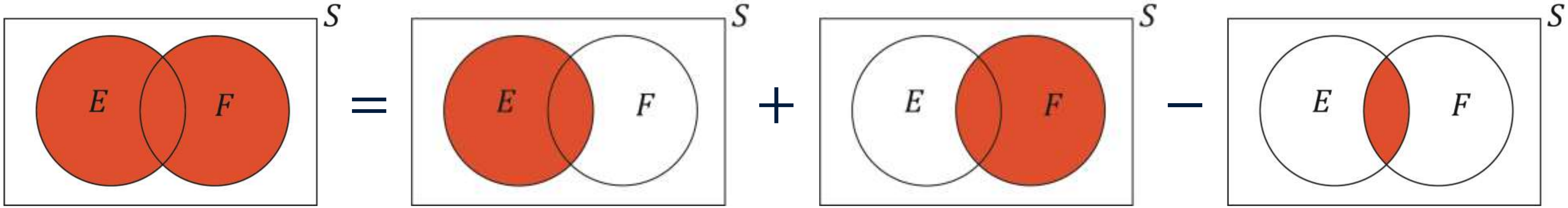
An unfair die

$$S = \{1,2,3,4,5,6\}, P(\{1\}) = \cdots = P(\{5\}) = \frac{1}{10}, P(\{6\}) = \frac{1}{2}$$

Probability Theory Basics

Rules of calculating probabilities

- $P(E) = \frac{|E|}{|S|}$, if each element in S has equal probability, where $|\cdot|$ denotes the cardinality (how many?) of the set.

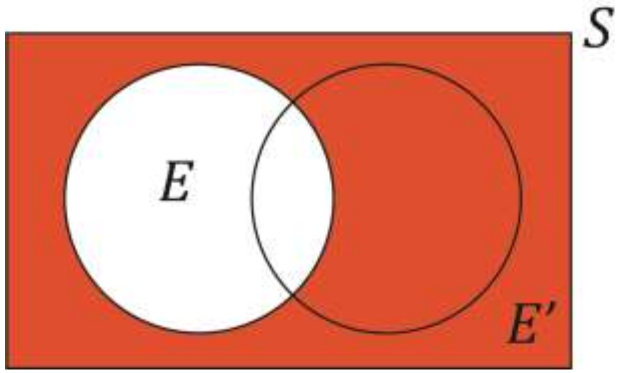


Probability Theory Basics

Rules of calculating probabilities

- $P(E) = \frac{|E|}{|S|}$, if each element in S has equal probability, where $|\cdot|$ denotes the cardinality of the set.
- Given events E and F: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- For disjoint events E and F: $P(E \cup F) = P(E) + P(F)$
- For any event E: $P(E) = 1 - P(E')$

Proof: $1 = P(S)$



Calculating Probabilities (example)

Example: Rolling a fair die

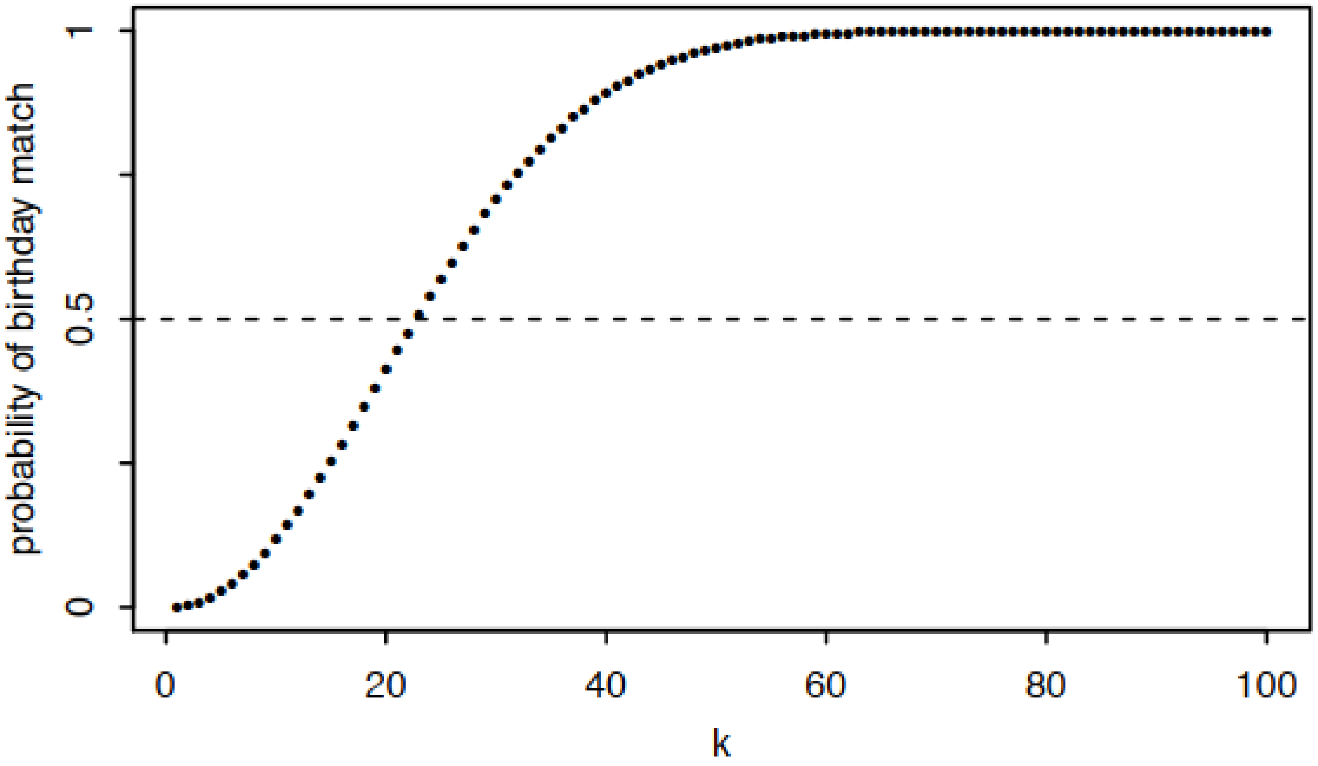
$S = \{1,2,3,4,5,6\}$, events: $A = \{1\}, B = \{2,4,6\}, C = \{1,2,3\}$

- $P(A) =$
- $P(B \cup C) =$
- $P(A \cup B)$
- $P(C') =$

The Birthday Problem

- K people in a room, considering only 365 days in a year, what is probability two or more people have the same birthday?
- Hint: Use the complement rule.

The Birthday Problem



Everything is Conditional!

- Probability of heads being 0.5 is conditional on a truly fair coin toss.
- Probability of a die being 1/6 is conditional on a *fair* die.

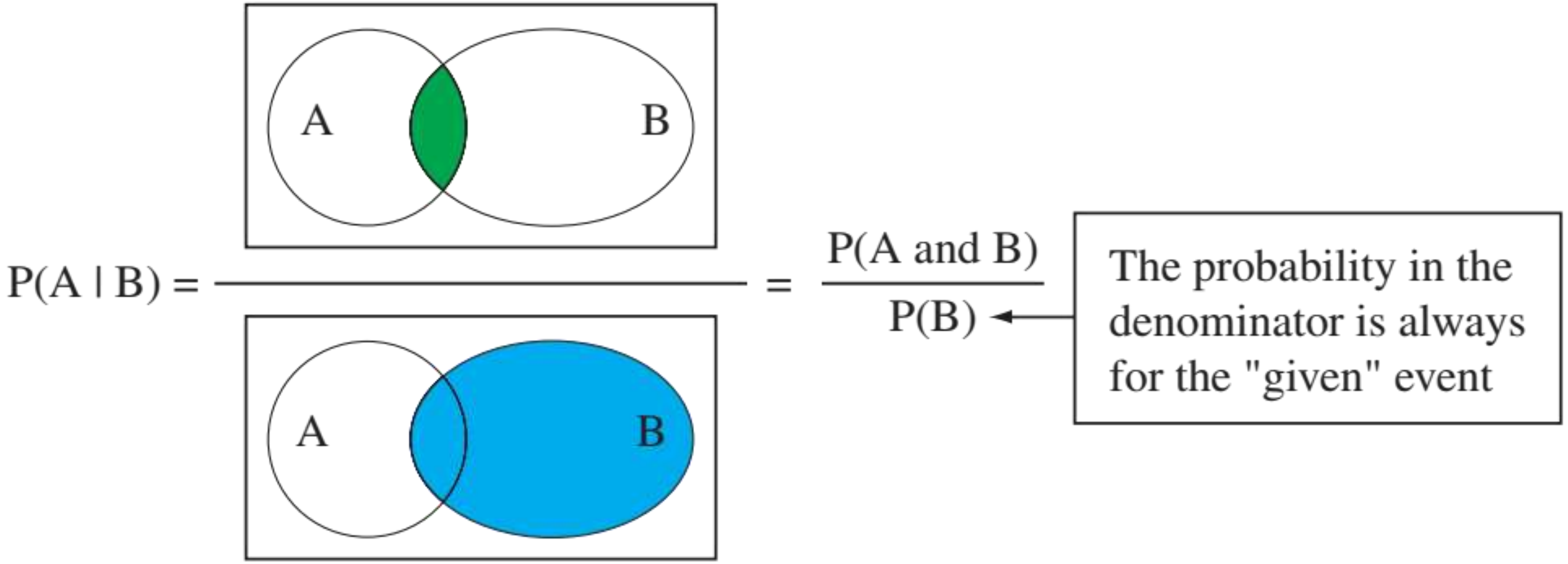
- What is a conditional probability?
- For two events A and B, the conditional probability

$$P(A|B) = \frac{P(A \text{ } and \text{ } B)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

Everything is Conditional!

- For two events A and B, the conditional probability

$$P(A|B) = \frac{P(A \; and \; B)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$



Everything is Conditional!

- For two events A and B, the conditional probability

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

- $P(A \cap B) = P(B) * P(A|B) = P(A) * P(B|A)$
- We can generalize this by recursive application:
- $P(A \cap B \cap C) = P(A) * P(B|A) * P(C|A,B) =$

$$P(B) * P(C|B) * P(A|B,C)$$

- (treat $B \cap C$ as one event E first)

Everything is Conditional!

- For two events A and B, the conditional probability when A and B are independent is:

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)} = \frac{P(A) * P(B)}{P(B)} = P(A)$$

Everything is Conditional! Or Independent.

Independent events

Events A and B are called independent if

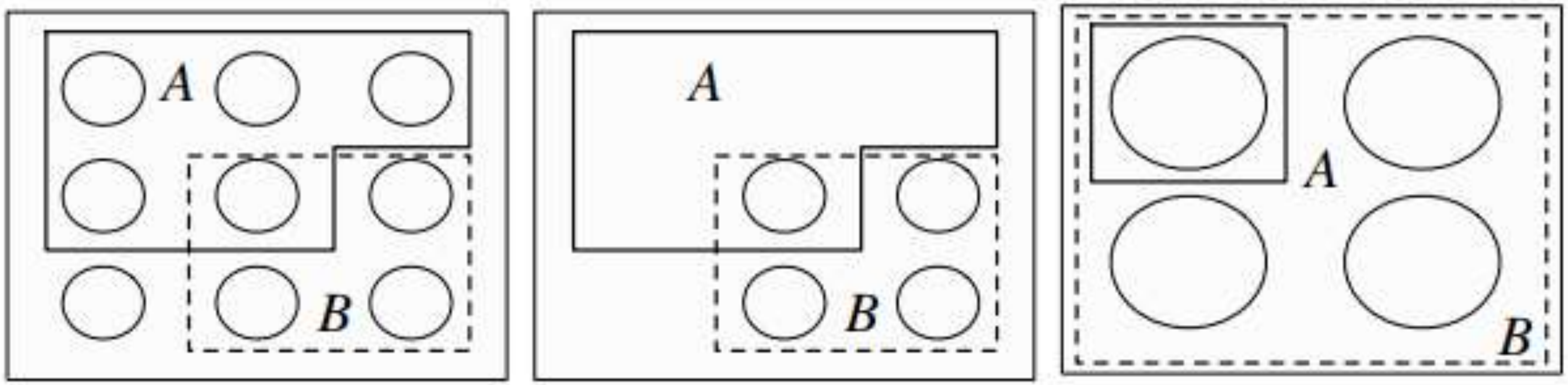
a) $P(A \mid B) = P(A)$ or

b) $P(B \mid A) = P(B)$

e

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

A Visual Interpretation of Conditioning



Conditional Probabilities

- When screening for breast cancer, mammography is roughly 90% accurate, in the sense that 90% of women with cancer, and 90% of women without cancer, will be correctly classified.

- Suppose 1% of women being screened actually have cancer: what is the probability that a randomly chosen woman will have a positive mammogram, and if she does, what is the chance she really has cancer?

Conditional Probabilities

- When screening for breast cancer, mammography is roughly 90% accurate, in the sense that 90% of women with cancer, and 90% of women without cancer, will be correctly classified.

- Suppose 1% of women being screened actually have cancer: **what is the probability that a randomly chosen woman will have a positive mammogram,** and if she does, what is the chance she really has cancer?

- $P(M^+) = ?$

Conditional Probabilities

- When screening for breast cancer, mammography is roughly 90% accurate, in the sense that 90% of women with cancer, and 90% of women without cancer, will be correctly classified.

- Suppose 1% of women being screened actually have cancer: **what is the probability that a randomly chosen woman will have a positive mammogram,** and if she does, what is the chance she really has cancer?

- $P(M^+) = P(M^+ \cap BC^-) + P(M^+ \cap BC^+)$
- $\qquad\qquad = P(M^+|BC^-) * P(BC^-) + P(M^+|BC^+) * P(BC^+)$
- $\qquad\qquad = 0.1 * 0.99 + 0.9 * 0.01 = \sim 0.11$

Conditional Probabilities

- When screening for breast cancer, mammography is roughly 90% accurate, in the sense that 90% of women with cancer, and 90% of women without cancer, will be correctly classified.

- Suppose 1% of women being screened actually have cancer: what is the probability that a randomly chosen woman will have a positive mammogram, and **if she does, what is the chance she really has cancer?**

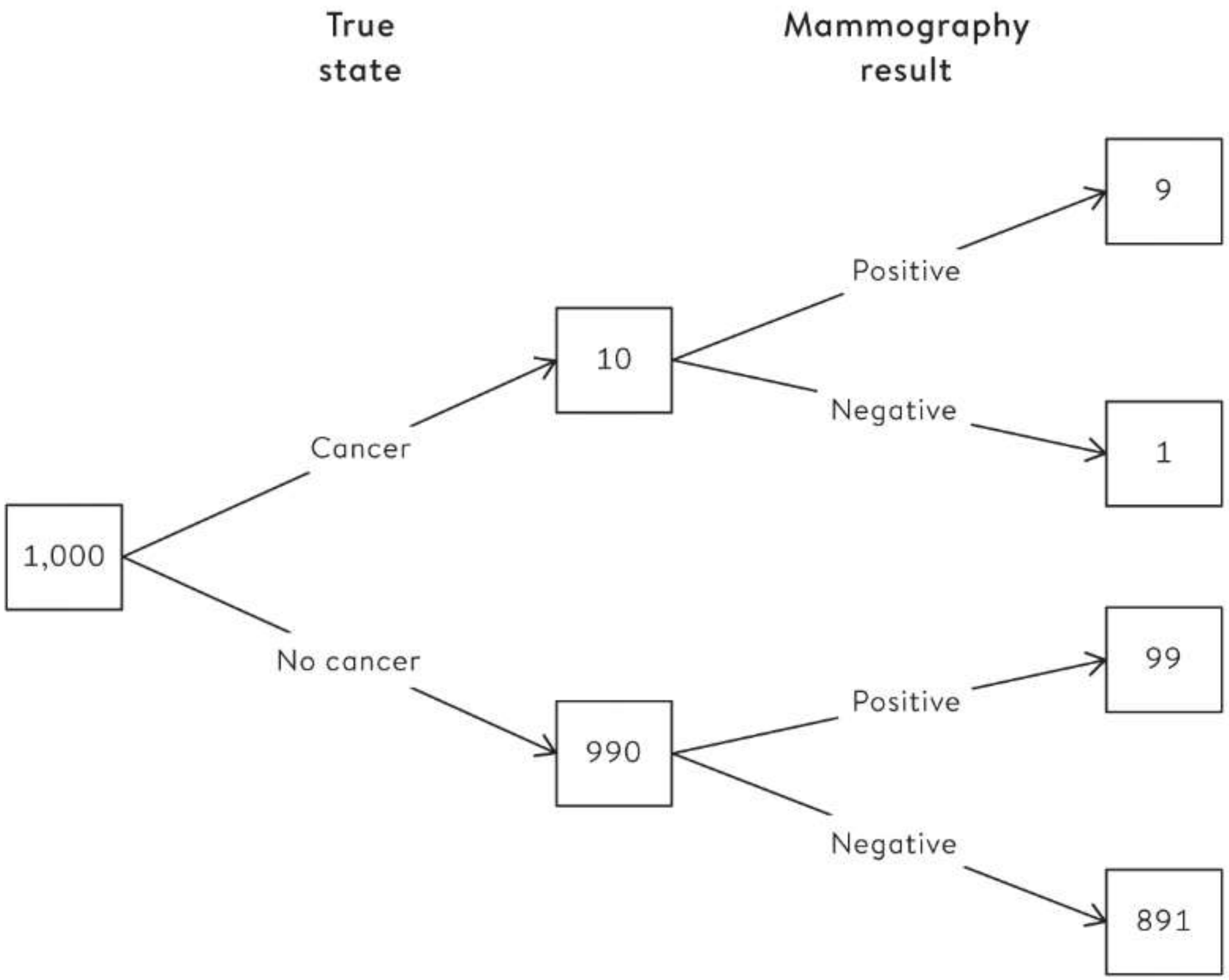
- $P(BC^+|M^+) = \frac{P(BC^+ \cap M^+)}{P(M^+)} = \frac{P(M^+|BC^+) * P(BC^+)}{P(M^+)} = \frac{0.009}{0.11} = 0.08$

Conditional Probabilities

- Suppose 1% of women being screened actually have cancer: **what is the probability that a randomly chosen woman will have a positive mammogram**, and if she does, what is the chance she really has cancer?
- …

- Another way to think about this is in terms of expected frequencies. For 1000 women, how many have breast cancer? What percentage of them will test positive?

Conditional Probabilities



Conditional Probabilities from Contingency Table

	Cancer (BC+)	No Cancer (BC-)	
Positive Test (M+)	9	99	108
Negative Test (M-)	1	891	892
	10	990	200

- $P(BC^+|M^+) = \dfrac{P(BC^+ \cap M^+)}{P(M^+)} = \dfrac{9}{108} = 0.08$

Conditional Probabilities

- Despite "*90% accuracy*", this is pretty bad overall result.
- A positive mammography here only gives an 8% chance of having breast cancer. What use is that?
- At what percentage would this test be useful?
- If accuracy is 99%, then we have:

$$P(M^+) = P(M^+BC^-) + P(M^+BC^+) = 0.01 * 0.99 + 0.99 * 0.01 = \sim 0.02$$

$$P(BC^+|M^+) = \frac{P\left(BC^+ \,\&\, M^+\right)}{P(M^+)} = \frac{0.0099}{0.02} = \sim 0.5$$

Exercise: Build both the tree and contingency table.

- The Triple Blood Test screens a pregnant woman and provides as estimated risk of her baby being born with the genetic disorder Down syndrome.
- A study of 5282 women aged 35 or over analyzed the Triple Blood Test to test its accuracy. It was reported that of the 5282 women, “48 of the 54 cases of Down syndrome would have been identified using the test and 25 percent of the unaffected pregnancies would have been identified as being at high risk (positive) for Down syndrome.”
- What is probability of Down’s if you had a positive test?

Bayes’ rule

Statistical Interpretation of Bayes’ rule

Prior probability of the hypothesis

How expected the observation is given the hypothesis

$$P(H \mid E) = \frac{P(H) \cdot P(E \mid H)}{P(E)}$$

$$\boldsymbol{P(BC^+ \mid M^+) = \frac{P(BC^+) * P(M^+ \mid BC^+)}{P(M^+)}}$$

Probability of the hypothesis given the observation

How expected the observation is without the hypothesis

$$\boldsymbol{P(BC^+ \mid M^+) = \frac{P(BC^+) * P(M^+ \mid BC^+)}{P(M^+)}}$$

Bayes' rule

Probability of Spam

How expected the email words are assuming it is spam

Prior probability that something is spam

$$P(\text{Spam} \mid \mathit{Email}) = \frac{P(\mathit{Spam}) \cdot P(\text{Email} \mid \mathit{Spam})}{P(\text{E} \mid \text{Spam}) + P(E|{\sim}\mathit{Spam})}$$

Probability of the spam given the email words

Probability of words in email given spam vs not spam

Bayes' Rule

- Writing Bayes' Rule using Odds:

$$odds(A) = \frac{P(A)}{P(A^c)}$$

$$P(A) = \frac{odds(A)}{1 + odds(A)}$$

Bayes' Rule

- Writing Bayes' Rule using Odds:

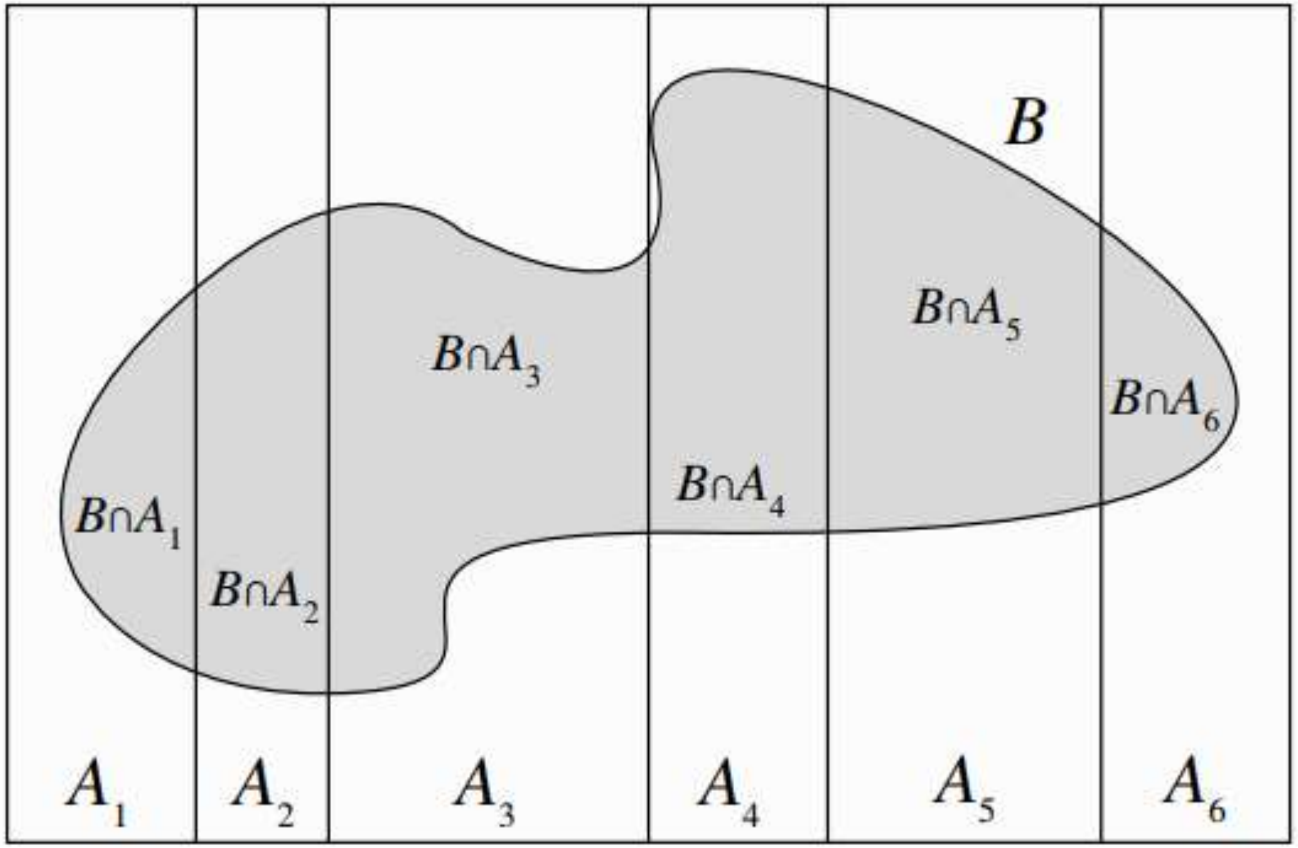
$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(B|A)}{P(B|A^c)} \frac{P(A)}{P(A^c)}$$

- Posterior odds $\frac{P(A|B)}{P(A^c|B)}$ are equal to prior odds $\frac{P(A)}{P(A^c)}$ times the factor $\frac{P(B|A)}{P(B|A^c)}$ which is called the **likelihood ratio**.

Law of Total Probability

- From Bayes' rule we end up at the law of total probability which helps link conditional to unconditional probability:
- $P(B) = \sum_{i=1}^{n} P(B|A_i) * P(A_i) = \sum_{i=1}^{n} P(B \cap A_i)$
- Where $A_1, A_2, \ldots, A_n$ is a partition of the sample space S and they are disjoint (union of A's = S)

Law of Total Probability



Bayes' rule

Example

Let C and D denote the events that a randomly chosen person has cancer and is diagnosed as having cancer, respectively.

Data: $P(C) = 0.05$, $P(C') = 0.95$, $P(D \mid C) = 0.78$, $P(D' \mid C) = 0.22$, $P(D \mid C') = 0.06$, and $P(D' \mid C') = 0.94$

Question: Calculate $P(D)$

Solution:

$P(D)$

Bayes' rule

Example

Let C and D denote the events that a randomly chosen person has cancer and is diagnosed as having cancer, respectively.

Data: $P(C) = 0.05$, $P(C') = 0.95$, $P(D \mid C) = 0.78$, $P(D' \mid C) = 0.22$, $P(D \mid C') = 0.06$, and $P(D' \mid C') = 0.94$

Derived information: $P(D) = 0.096$ and $P(D') = 0.904$

Question (2): What is the probability that someone diagnosed with cancer actually has the disease?

Solution:

$$P(C \mid D)$$

Bayes' rule

Example

Let C and D denote the events that a randomly chosen person has cancer and is diagnosed as having cancer, respectively.

Data: $P(C) = 0.05$, $P(C') = 0.95$, $P(D \mid C) = 0.78$, $P(D' \mid C) = 0.22$, $P(D \mid C') = 0.06$, and $P(D' \mid C') = 0.94$

Derived information: $P(D) = 0.096$ and $P(D') = 0.904$

Question (3): What is the probability that someone who is diagnosed as not having cancer actually has the disease?

Solution:

$$P(C \mid D') :$$

Conditional Probability

- Used when “adjusting for” confounds
- We condition on, for example, summer:
- P(Ice Cream Sales | Summer, Shark Attacks) = 0.
- Conditioning simply mean that we have altered the subset of events over which we measure probability.

Conditional Probability is Probability

- $P(S|E) = 1$
- $P(\emptyset|E) = 0$
- Conditional probabilities must lie between 0 and 1
- For A and B disjoint:

$$P(A \cup B|E) = P(A|E) + P(B|E)$$

Bayes’ Rule, now with extra conditioning

$$P(H \mid \mathrm{A}, E) = \frac{P(H|\mathrm{A}) \cdot P(E \mid H, \mathrm{A})}{P(E|\mathrm{A})}$$

Prosecutor's fallacy

- Assuming that the probability $P(A|B)$ is the same as $P(B|A)$
- In Forensic science, the statement:
- "If the accused is innocent, there is only a 1 in a billion chance that the DNA at the crime scene matches them"
- Is understood to mean:
- "Given the DNA evidence, there is only 1 in a billion chance the accused in innocent"

Prosecutor’s fallacy

- “If the accused is innocent, there is only a 1 in a billion chance that the DNA at the crime scene matches them”
- P(DNA match | innocent) = 1/1000000000
- “Given the DNA evidence, there is only 1 in a billion chance the accused in innocent”
- P(innocent | DNA match) = ?
- = P(DNA match | innocent) * P(innocent)/P(DNA match)
- = We don’t know!

Prosecutor’s fallacy

- Easier to see the error when you frame it as:
- “Given you’re a student, you don’t earn money”
- Vs
- “Given you don’t earn money, you’re a student”

Understanding the nature of probability

- Let’s philosophise!
- Probability is tricky.
- How can we understand what it means?

Understanding the nature of probability

- Classical probability: Dice, coins, packs of cards – derived based on symmetries. We assume fairness, and don't think about the conditionality on fairness.
- Enumerative or combinatorial probability: We calculate the expected frequencies of events when we have categories. Like picking a green ball out of a bag of 4 green and 5 blue balls. This too assumes something about fairness and so again, under this approach we have to pretend everything is fair.
  - Consider if the bag is shaped weirdly, or if you naturally put your hand in a certain way. Or as happened with the Vietnam war draft, the balls are badly randomized.

Understanding the nature of probability

- Long-run frequency probability: We don’t make a definition of equally likely. Instead, we do 387500 coin tosses and assess the long run probability. How do we understand the probability of rain in tomorrow’s weather?
- Propensity or “chance”: The idea of probability being out in the world. As though it is truly measureable – if you were all-knowing, then you would know it.

Understanding the nature of probability

- Subjective probability: The idea that every probability is essentially saying something about a person's judgement about that specific occasion. This can definitely be based on background knowledge, but in essence says that the probability isn't out in the world, instead it appears out of a desire to try to make sense of the world.
- For example, a weatherman has a better sense for the possibility of rain tomorrow than you or I.

Understanding the nature of probability

- Subjective probability: Calling it subjective doesn't mean it isn't useful or replicable across individuals. We can both come to the same value for the probability of dice because we both have the same knowledge about it.
- So when working with probability, even if we think it is subjective, we can pretend it represents something objective in the world.
- Indeed, *As if* thinking is how we do science!

Connecting probability, data, and statistical learning

Probability is natural in situation 1:

1. Data is *generated* by a randomizing device like flipping dice or coins or a randomized control trial.

Generally, we face the following situation:

2. The pre-existing data-point is *chosen* by a randomizing device, like when people take part in a survey.

Connecting probability, data, and statistical learning

Or we have situation 3, which is how a lot of data is generally generated:

3. There is no randomness at all but we act *as if* the data is in fact generated by a random process. Like how salty the spoon of soup was and pretending it came from a bell curve of saltiness.

Note: Under the subjective probability expectation, all three situations become situation 3.

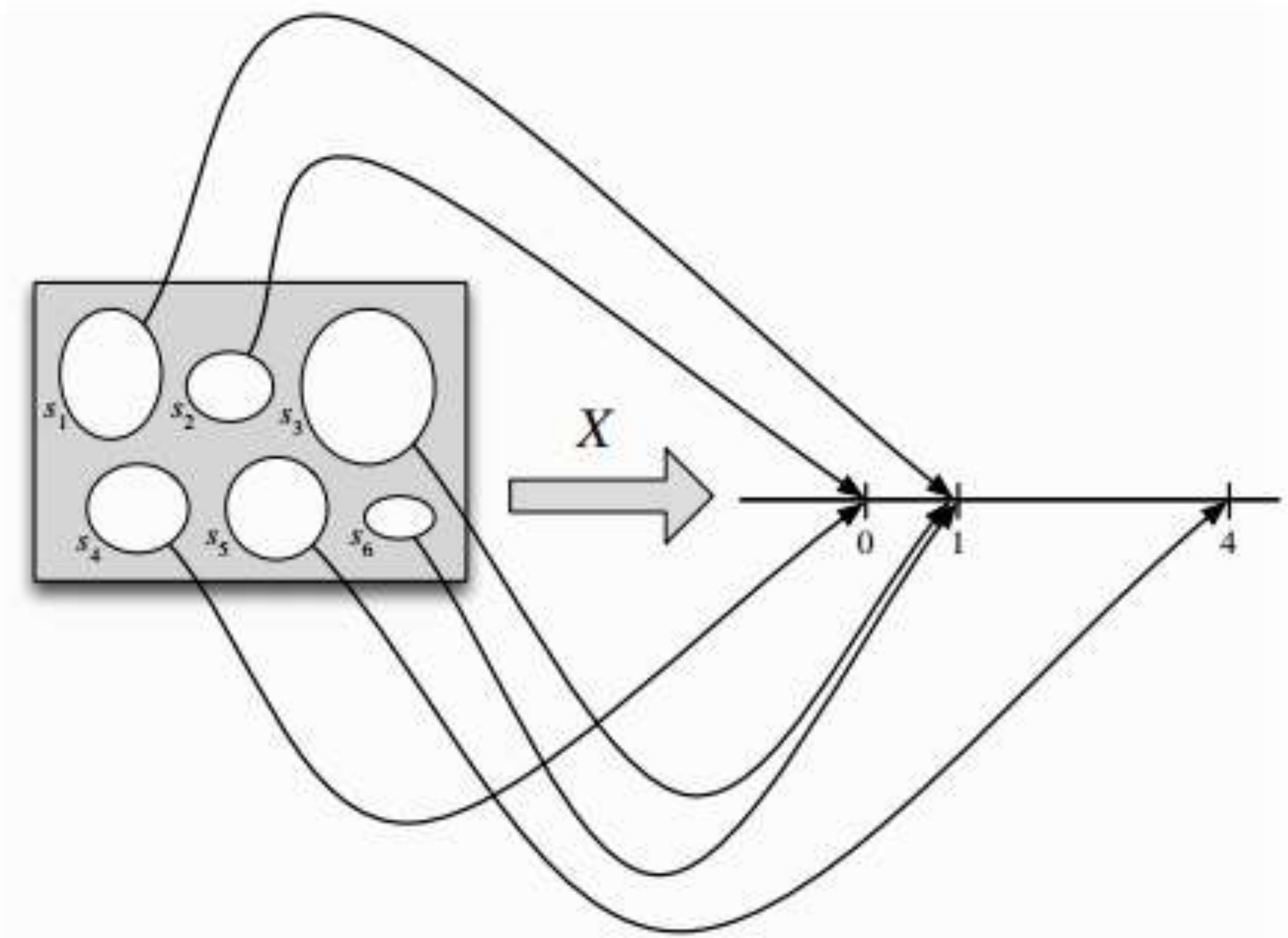
Connecting probability, data, and statistical learning

Consider situations 1 and 2: so dice and times when we have a randomizing device allowing selection, then we have a random variable.

Definition

Random variable: A mapping of events in a sample space that is believed to be associated with some probabilities to the real line.
For example, outcomes of a coin can be assumed to come from a probability distribution with weight 0.5 at heads and 0.5 at tails. Every time we flip a coin, then the probabilities collapse, and we have an outcome.

Introduction to Random Variables



Introduction to Random Variables

Random Variable (RV)

X is a random variable for the sample space S if it assigns a real number to each element of S, $X\colon\ S \to \mathbb{R}$

Example 2

Rolling a die until a 6 comes up $S = \{6, N6, NN6, NNN6, \dots\}$ where N denotes an outcome of 1,2,3,4 or 5

X: number of rolls required

$$X(6) = 1, X(N6) = 2, X(NN6) = 3, \dots$$

Introduction to Random Variables

- For two coin tosses S = {HH, HT, TT, TH}.
  - Consider X = number of Heads
  - Y = 1 when first toss is heads
  - Y = 1 when both heads and tails are tossed and 0 otherwise
- These are all valid random variables and enable some useful further algebra for us.

Discrete Random Variables

Probability Mass Function

The probability distribution of a discrete random variable (RV) is defined as $f(x) = P(X = x)$

Example: Rolling a die until a 6 comes up

$S = \{6, N6, NN6, NNN6, \dots\}$

X: number of rolls required, i.e. $X(6) = 1, X(N6) = 2, \dots$

$$f(1) = \frac{1}{6}$$

Discrete Random Variables

Probability Mass Function

The probability distribution of a discrete random variable (RV) is defined as $f(x) = P(X = x)$

Example: Rolling a die

$S = \{1,2,3,4,5,6\}$, $\mathrm{X}(s) = \left\lfloor \frac{s}{2} \right\rfloor$, where s is the outcome of the roll

$$f(0) = P(X = 0) = P(\{1\}) = \frac{1}{6}$$

Probability Mass Function

- Having a probability mass function for a discrete random variable lets us answer questions like:
  - If M is the number of earthquakes in Myanmar in the next 10 years, what is the probability of $M = 0$?
  - If K is the number of 4s rolled when 5 dice are rolls, what is the probability of $K = 2$?

Discrete Random Variables

Cumulative distribution (denoted by $F(x)$)

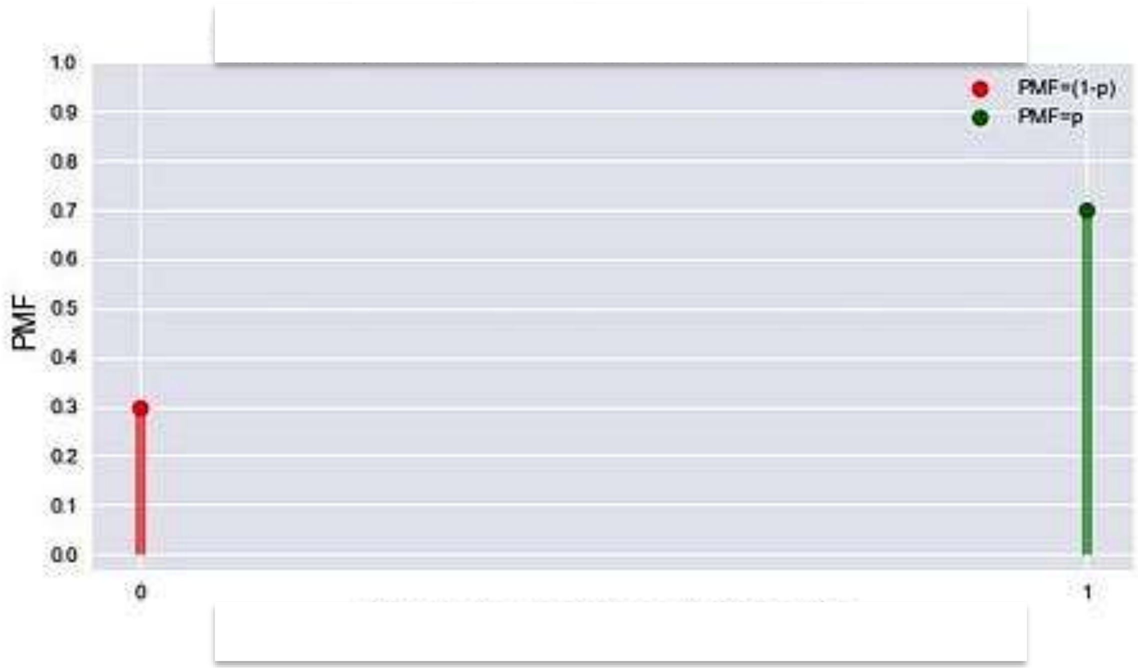
$$F(x) = P(X \leq x) = \sum_{y \leq x} f(y)$$

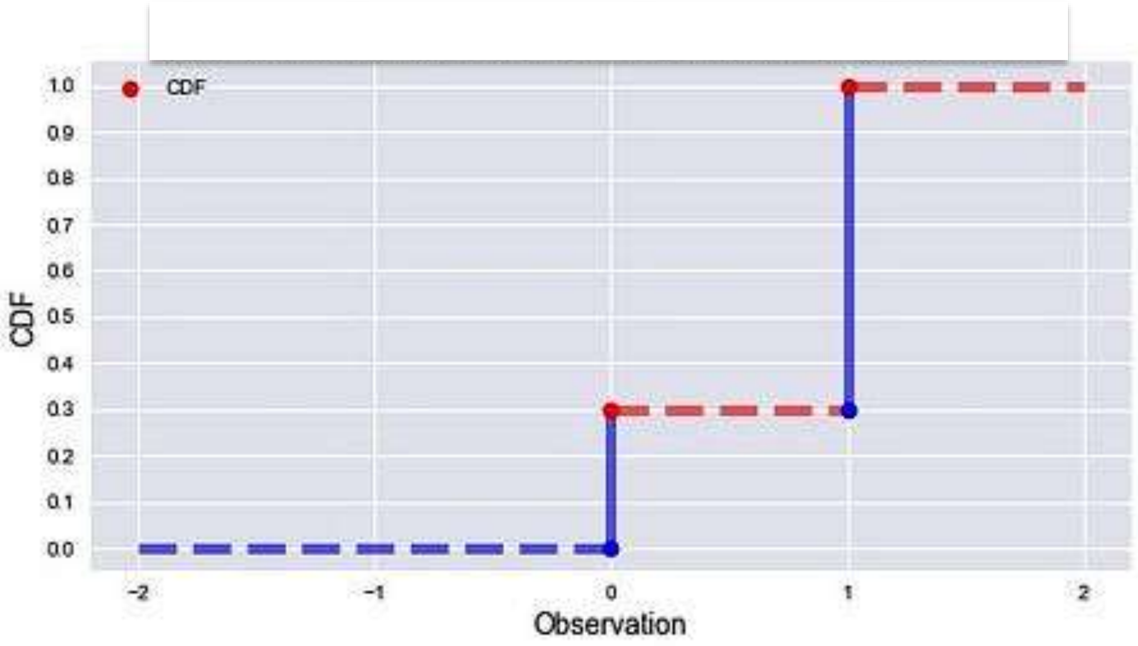
Example: Rolling a die

$S = \{1,2,3,4,5,6\}$, $X(s) = \left\lfloor \frac{s}{2} \right\rfloor$, where s is the outcome of the roll

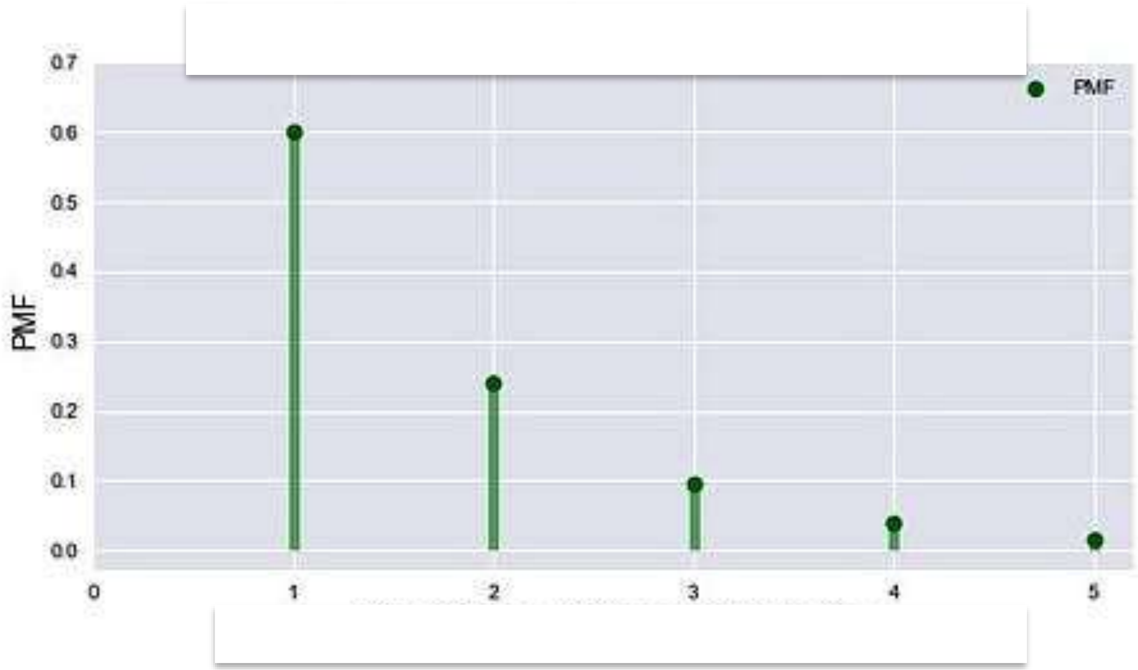
x	0	1	2	3
$f(x) = P(X = x)$	1/6	1/3	1/3	1/6
$F(x) = P(X \leq x)$				

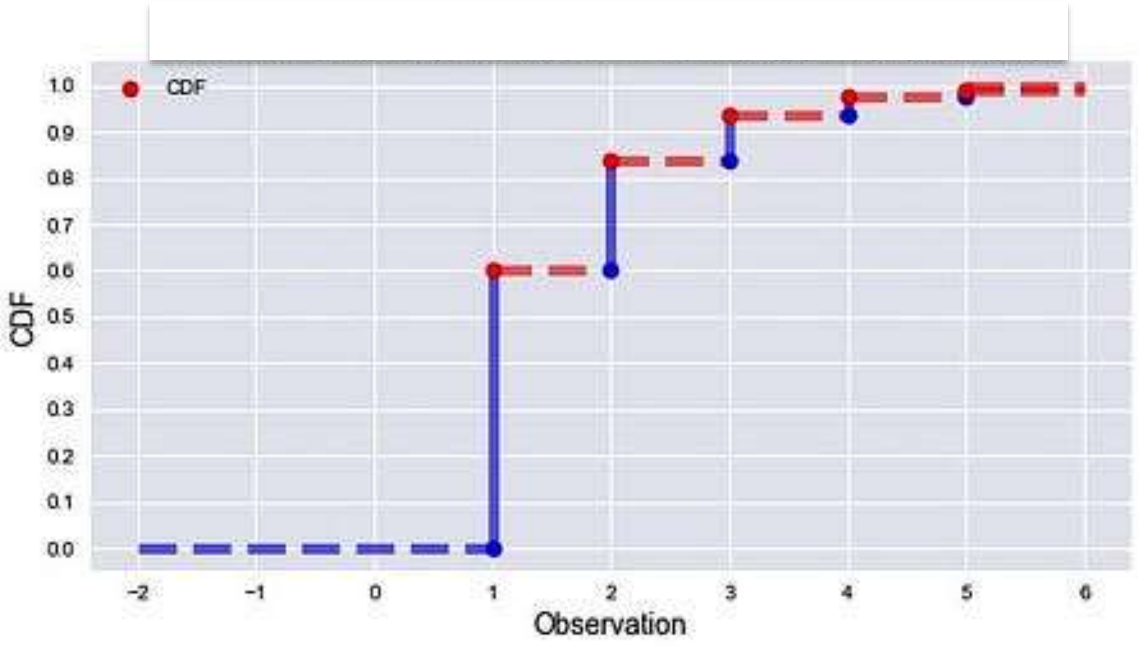
Cumulative Probability Distribution (CDF): Accumulating Probability





Cumulative Probability Distribution (CDF): Accumulating Probability





Cumulative Probability Distribution (CDF): Accumulating Probability

- Having a cumulative distribution function for a discrete random variable lets us answer questions like:
  - If M is the number of earthquakes in Myanmar in the next 10 years, what is the probability of $M > 0$?
  - If K is the number of 4s rolled when 5 dice are rolls, what is the probability of $K > 2$?

Percentiles and CDF

- The value X at a percentile of 50% asks for 50% of the possible data to be below X.
- The cumulative probability mass gives us immediate access to that.
- For example: For a die roll, with X = value of die, $P(X<4) = \frac{1}{2}$

Question

For the number of emails you get per day, which question is answered by the CDF?

A) What is the chance I get exactly 10 emails?

B) What is the chance I get 10 or fewer emails?

C) What is the chance I get more than 10 emails?

D) None of the above.

Connecting probability, data, and statistical learning

Consider this situation again:
There is no randomness at all but we act as if the data is in fact generated by a random process.

For example, consider the case where we have data about homicide events. This is not coming from a randomizing device, or any sort of random process we can immediately identify.

Can we connect this to probability theory to help answer the following question: How possible is it that tomorrow there are 7 homicide events?

Modeling Homicide Events

- What is the data? We have the number of homicide events on every data for the past year.

- We could make a prediction for tomorrow based on the mean: note that this is already connecting to probability theory. We are assuming that the highest probability for the number of homicide events tomorrow is the mean for the past year.

- Mean = 1.41/day – this seems a reasonable guess for tomorrow.

Modeling Homicide events

- Suppose we wanted to answer the question: how likely is it that there are over 7 homicide events tomorrow?

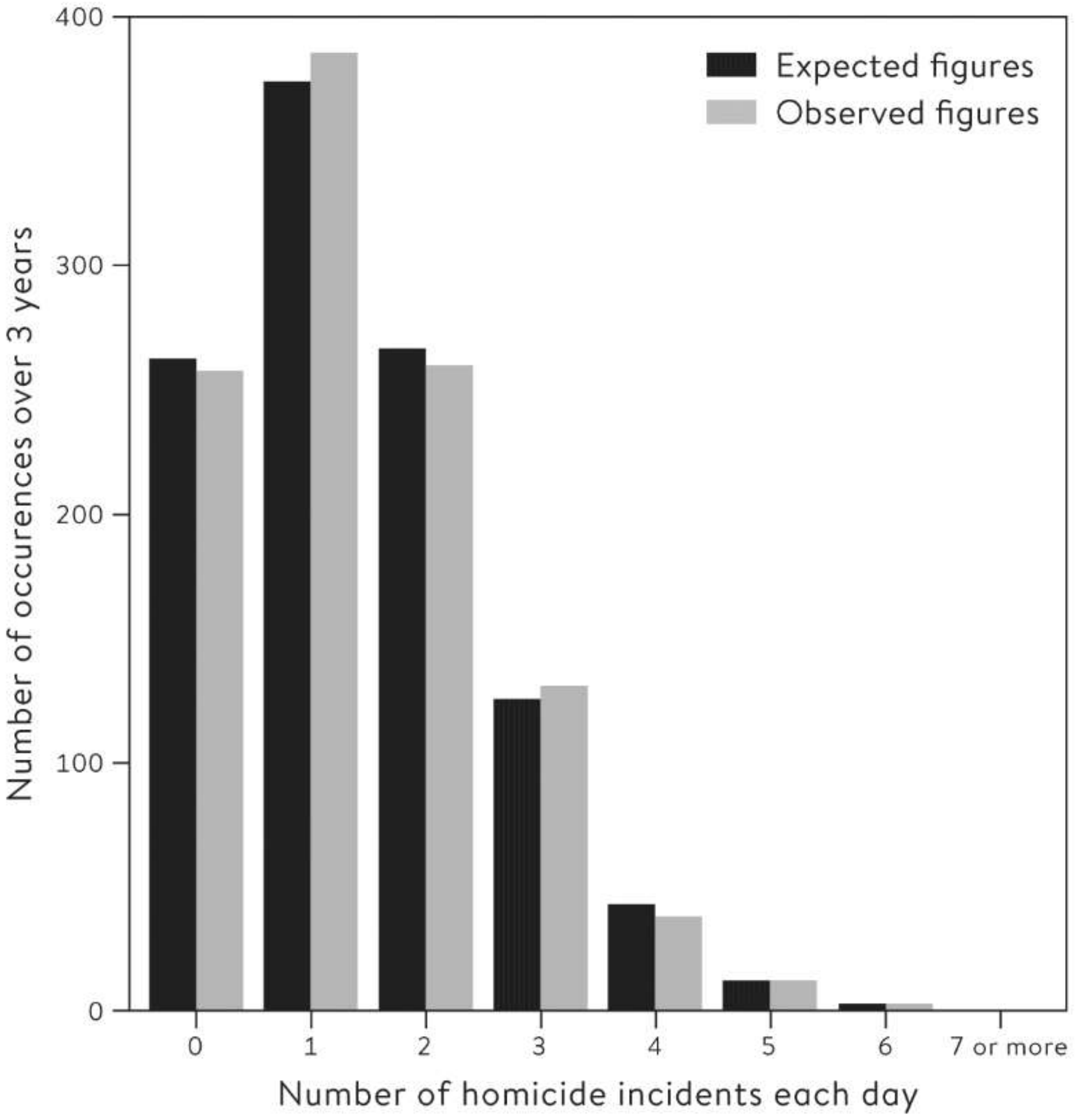
- One way to do this is to calculate how many days had more than 7 events in the past 3 years and divide by the number of days. And lean on the long-run frequency as a measure of probability.

- Alternatively, we can use a discrete probability distribution (as opposed to *continuous*): the *Poisson* distribution.
- Why use a distribution? They tend to allow us to generalize our understanding of the data and to connect it to other similar situations. And they are less susceptible to being influenced by outliers.

Modeling Homicide

A Poisson distribution is fully characterized by the mean: all we need is the mean to draw samples and make a distribution.

We should compare the real data to the expected distribution to make sure they fit (more on this in *Simulation and Statistical Analysis*).

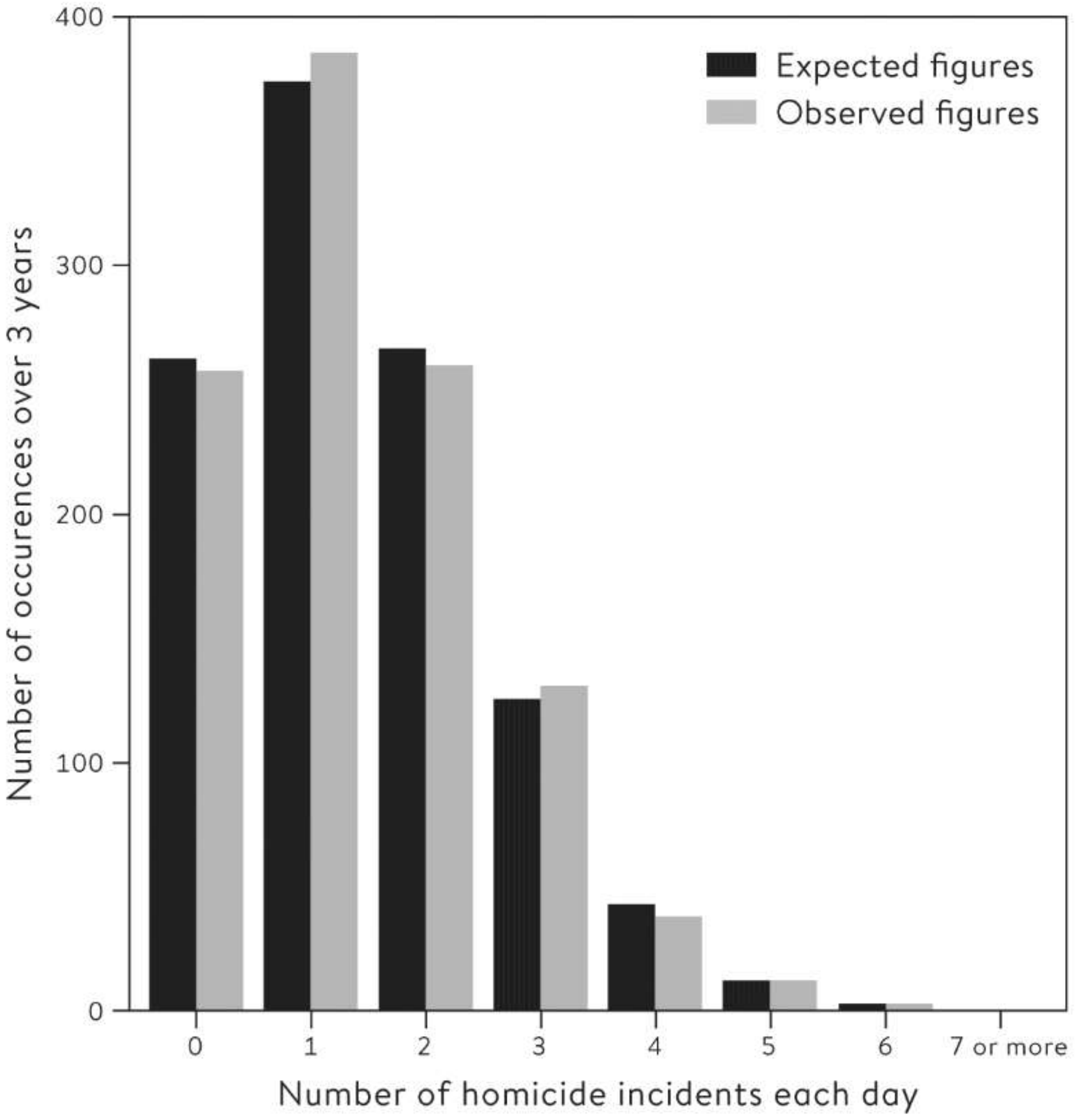


Modeling Homicide

Ok, the distribution seems reasonable.

Then we can calculate the probability based on the distribution.

This means summing over the probability that the number of homicides is > 7.

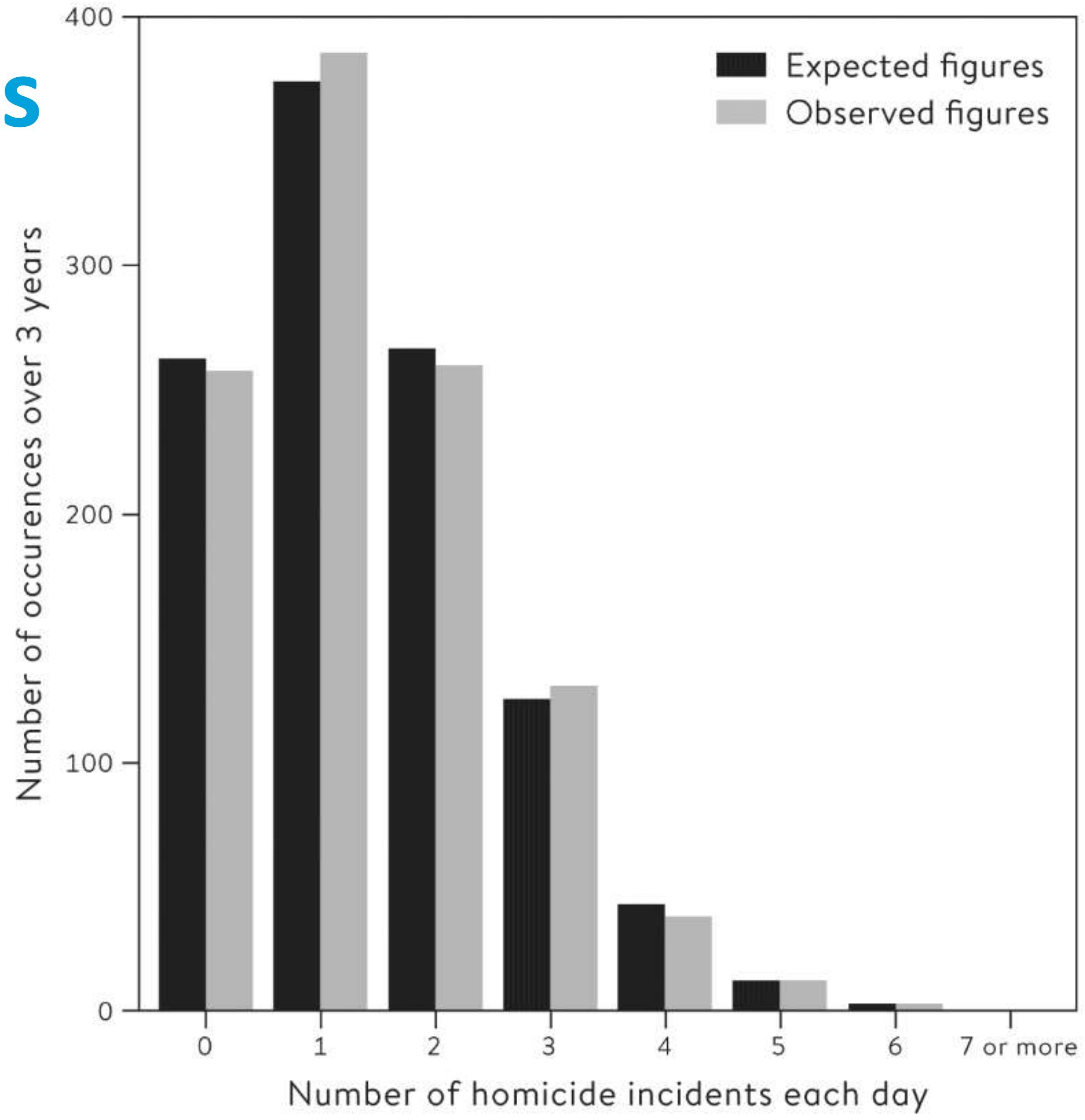


Modeling Homicide Events

P(Homicide events > 7| Poisson) = 0.07 %

Can we check this with the past data?

Yes, we can calculate the expected frequency directly.



Modeling Homicide events

- Despite having all the data, we could pretend the data came from a random distribution.
- The function for the distribution seems to capture some underlying structure to the data.
- Like the central limit theorem, there simply are distributions that can do a good job a being a way to summarize data that are otherwise very difficult to understand how they are generated.

Modeling Homicide events

- For example, another way to predict how many homicides happen tomorrow is to implement some way to read people's minds.
- Now we know how many people are feeling murderous and based on that make a prediction for how many homicide events happen tomorrow.
- This is not something we can do.
- The Poisson captures the doubt that comes from not having all the information about every individual and gives us a way to quantify that.

Summary

- Probability is a useful way to quantify uncertainty.
- Probability is most intuitively understood from the place of expected frequency.
- Can think about probability in several ways.
- Having access to probability theory allows us to do statistics in a more formal way, that is, not simply relying on long-run frequencies.