

Introduction to Probability Theory 1

Shyam Gopalakrishnan, 2025

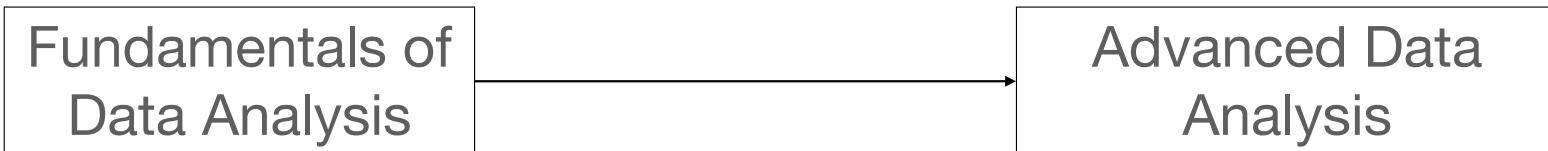
Heavily borrowed from Fernando Racimo's slides

Fundamentals of Data Analysis

Fundamentals of
Data Analysis

Advanced Data
Analysis





- Properties of estimators
- Likelihood inference
- Bayesian thinking
- Generalized models
- Model comparison and regularization
- Resampling
- Mixed models
- Unsupervised learning

The two sides of statistics

Probability
Theory



“What can we say about
the data generated by a
given process?”

The two sides of statistics

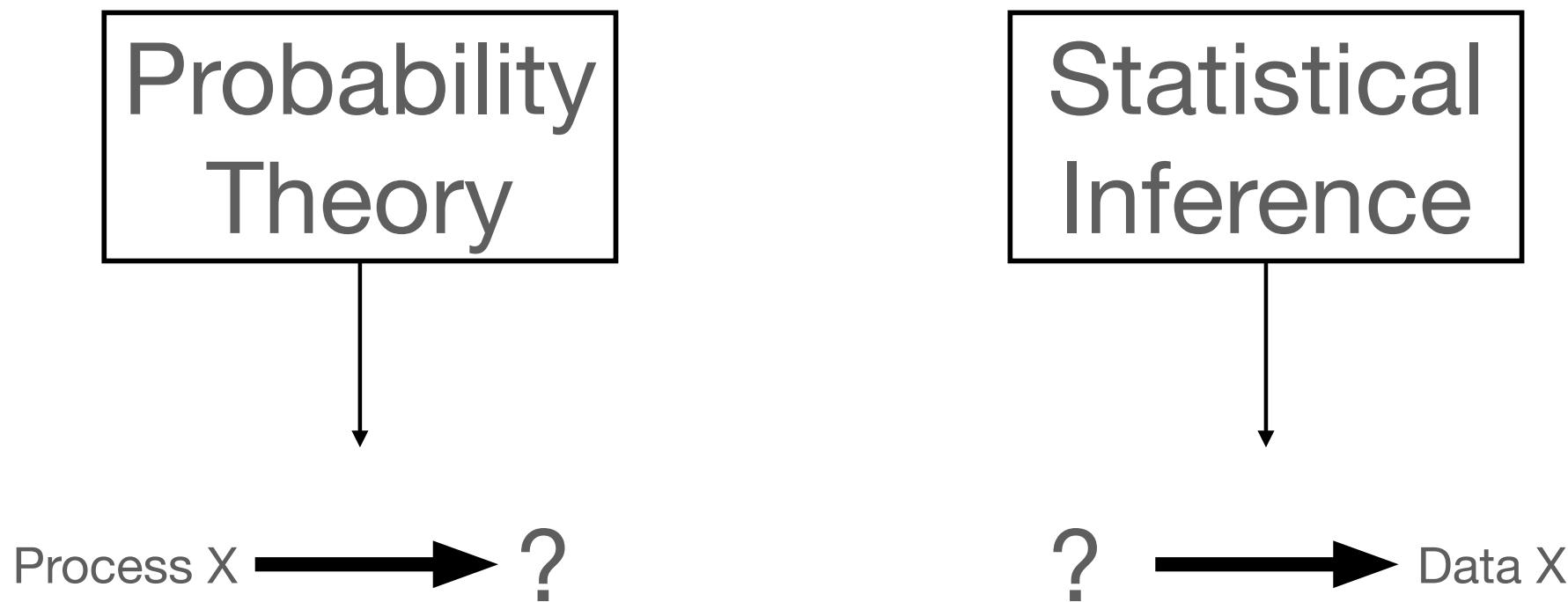
Probability
Theory

“What can we say about
the data generated by a
given process?”

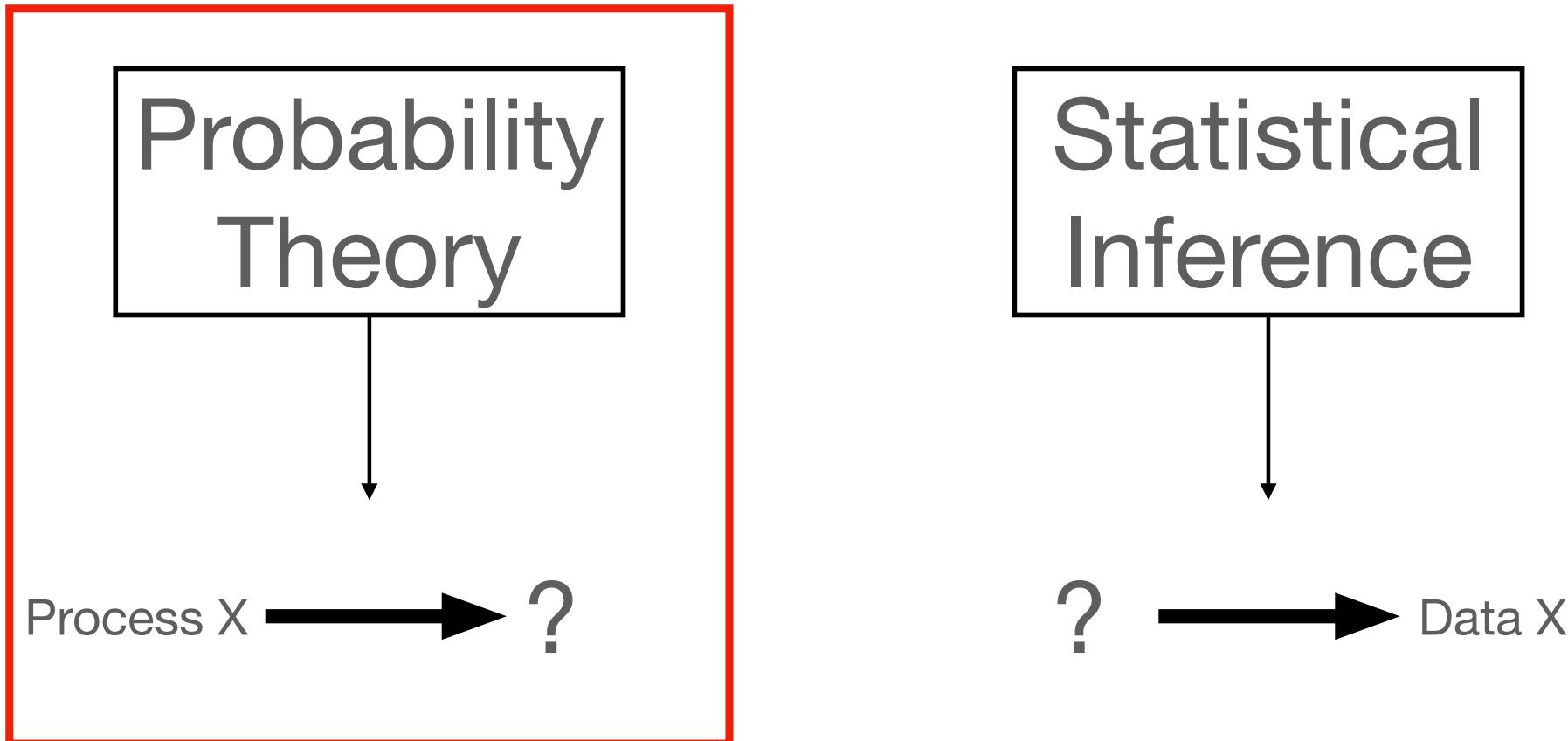
Statistical
Inference

“What can we say about
the process that
generated a given data?”

The two sides of statistics



The two sides of statistics



Some statistics lingo

Some statistics lingo

- We use the phrase **random variable** to refer to the **outcome** of a process, experiment or phenomenon we want to model

Some statistics lingo

- We use the phrase **random variable** to refer to the **outcome** of a process, experiment or phenomenon we want to model
- **Sample set** = possible range of outcomes the random variable can take.

Some statistics lingo

- We use the phrase **random variable** to refer to the **outcome** of a process, experiment or phenomenon we want to model
- **Sample set** = possible range of outcomes the random variable can take.



Some statistics lingo

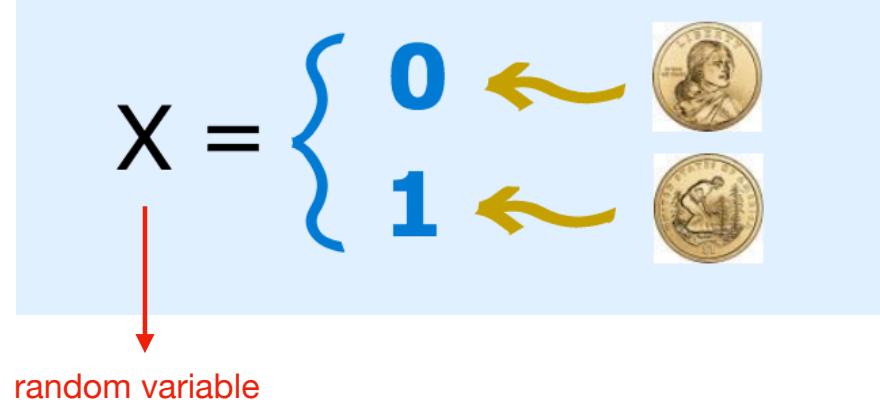
- We use the phrase **random variable** to refer to the **outcome** of a process, experiment or phenomenon we want to model
- **Sample set** = possible range of outcomes the random variable can take.



$$X = \begin{cases} 0 \\ 1 \end{cases}$$
A light blue rectangular box containing a mathematical definition of a random variable X. The variable is defined as a set with two elements: 0 and 1. To the right of the equation, there are two yellow arrows pointing from the numbers 0 and 1 to two separate US quarters. The top quarter shows the profile of George Washington, and the bottom quarter shows the reverse side of a quarter.

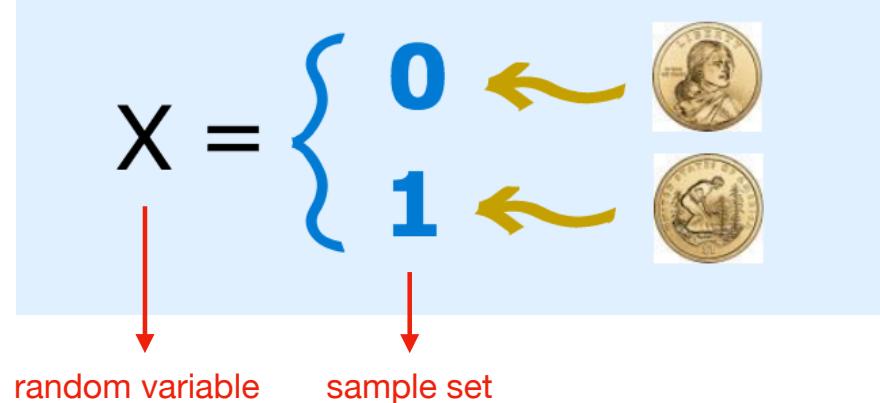
Some statistics lingo

- We use the phrase **random variable** to refer to the **outcome** of a process, experiment or phenomenon we want to model
- **Sample set** = possible range of outcomes the random variable can take.



Some statistics lingo

- We use the phrase **random variable** to refer to the **outcome** of a process, experiment or phenomenon we want to model
- **Sample set** = possible range of outcomes the random variable can take.



Basic rules of probability

- $P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$
- In other words, the probability that A or B happens is equal to the probability that A happens plus the probability that B happens, minus the probability that both A and B happen.

Basic rules of probability

- $P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$
- In other words, the probability that A or B happens is equal to the probability that A happens plus the probability that B happens, minus the probability that both A and B happen.
- Exercise: why do we have to subtract $P[A \text{ and } B]$?
- Hint:
 - $P[A] = P[A \text{ and } B] + P[A \text{ and not-}B]$

Rules of probability: conditioning

- $P[A | B]$ stands for “the probability of A conditional on B having happened”

Rules of probability: conditioning

- $P[A | B]$ stands for “the probability of A conditional on B having happened”
- $P[A \text{ and } B] = P[A | B] P[B] = P[B | A] P[A]$

Rules of probability: conditioning

- $P[A | B]$ stands for “the probability of A conditional on B having happened”
- $P[A \text{ and } B] = P[A | B] P[B] = P[B | A] P[A]$
- Example:
 - $P[\text{has covid} \text{ and tested positive}] =$
 - $= P[\text{tested positive} | \text{has covid}] P[\text{has covid}]$
 - $= P[\text{has covid} | \text{tested positive}] P[\text{tested positive}]$

Rules of probability: conditioning

- $P[A | B]$ stands for “the probability of A conditional on B having happened”
 - $P[A \text{ and } B] = P[A | B] P[B] = P[B | A] P[A]$
 - Example:
 - $P[\text{has covid and tested positive}] =$
 - $= P[\text{tested positive} | \text{has covid}] P[\text{has covid}]$ 
 - $= P[\text{has covid} | \text{tested positive}] P[\text{tested positive}]$ 
- Often, one of these is easier to obtain than the other

Rules of probability: independence

- A and B are said to be independent **if and only if** $P[A \text{ and } B] = P[A] P[B]$

Rules of probability: independence

- A and B are said to be independent **if and only if** $P[A \text{ and } B] = P[A] P[B]$
- When two events are independent, knowing that one event happens gives me no information about the other event
- If A and B are independent, then:
 - $P[A | B] = P[A]$
 - $P[B | A] = P[B]$

Rules of probability: independence

$$P\left[\begin{matrix} \text{Man with heart} \\ | \\ \text{Old man with cane} \end{matrix}\right] \neq P\left[\begin{matrix} \text{Man with heart} \end{matrix}\right]$$

Rules of probability: independence

$$P\left[\begin{array}{c} \text{Image of a young man with heart highlighted} \\ | \\ \text{Image of an elderly person with cane} \end{array}\right] \neq P\left[\begin{array}{c} \text{Image of a young man with heart highlighted} \end{array}\right]$$



and



are not independent

Rules of probability: independence

$$P\left[\begin{array}{c} \text{Image of a man with heart highlighted in red} \\ | \\ \text{Image of a calendar page showing Monday} \end{array}\right] = P\left[\begin{array}{c} \text{Image of a man with heart highlighted in red} \end{array}\right]$$

Rules of probability: independence

$$P\left[\begin{array}{c} \text{Image of a man with heart highlighted} \\ | \\ \text{Image of a calendar page labeled MONDAY} \end{array}\right] = P\left[\begin{array}{c} \text{Image of a man with heart highlighted} \end{array}\right]$$



and



are independent

Bayes Rule

- $P[A | B] = \frac{P[B | A]P[A]}{P[B]}$
- Exercise: derive this rule.
- Hint: write down what $P[A \text{ and } B]$ is equal to (from previous slides)

Law of total probability

- $P[A] = P[A|B]P[B] + P[A|C]P[C] + P[A|D]P[D] + \dots$
- In other words: $P[A] = \sum_i P[A|X_i]P[X_i]$
- Why might it be useful to know this?

Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random

Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random ≠ Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random ≠ Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



“7”

Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



“1/3”

Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



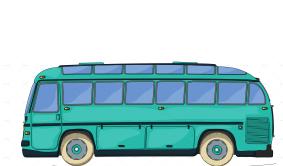
“7”



“1/3”

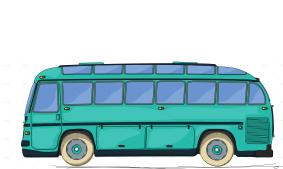
Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



Probability distribution

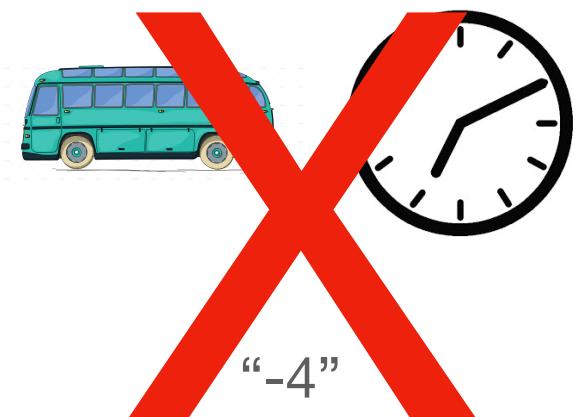
- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



“-4”

Probability distribution

- A **probability distribution** is a mathematical description of a random process, experiment or phenomenon (i.e. a random variable)
- Random \neq Aleatory
- There are still **rules** that define any phenomenon, even if the phenomenon is random



The Bernoulli distribution



\sim Bernoulli (p)

- Here, p is the parameter that determines how fair the coin is, i.e. the probability of heads.
- For example:
 - **X ~ Bernoulli (0.5)** means “The random variable X can be modeled as a fair coin toss”
 - **X ~ Bernoulli (0.9)** means “The random variable X can be modeled as a biased coin toss with 90% probability of heads”

The Bernoulli distribution



This squiggly symbol means:
“can be modeled as”



Bernoulli (p)

A red arrow points from the explanatory text above to the squiggle symbol (~) inside the red circle.

- Here, p is the parameter that determines how fair the coin is, i.e. the probability of heads.
- For example:
 - **$X \sim \text{Bernoulli (0.5)}$** means “The random variable X can be modeled as a fair coin toss”
 - **$X \sim \text{Bernoulli (0.9)}$** means “The random variable X can be modeled as a biased coin toss with 90% probability of heads”

The Bernoulli distribution



\sim Bernoulli (p)

- Let a “heads” outcome be defined as $X = 1$ and a “tails” outcome be defined as $X = 0$

The Bernoulli distribution



\sim Bernoulli (p)

- Let a “heads” outcome be defined as $X = 1$ and a “tails” outcome be defined as $X = 0$
- Assuming **$X \sim \text{Bernoulli (0.5)}$** then:
 - $P[X = 1] = 0.5$
 - $P[X = 0] = 0.5$

The Bernoulli distribution



\sim Bernoulli (p)

- Let a “heads” outcome be defined as $X = 1$ and a “tails” outcome be defined as $X = 0$
- Assuming **$X \sim \text{Bernoulli (0.5)}$** then:
 - $P[X = 1] = 0.5$
 - $P[X = 0] = 0.5$
- Assuming **$X \sim \text{Bernoulli (0.9)}$** then:
 - $P[X = 1] = 0.9$
 - $P[X = 0] = 0.1$

Probability distributions: what are they good for?

- They allow us to **simulate** data
- They allow us to find good approximations to the processes that generated our data (in other words, perform **statistical inference**)
- They allow us to choose **among competing models** to explain our data

Probability distributions: what are they good for?

- They allow us to **simulate** data
- They allow us to find good approximations to the processes that generated our data (in other words, perform **statistical inference**)
- They allow us to choose **among competing models** to explain our data
- They allow us to better **think about** our data and the processes that generate it

The binomial distribution



0

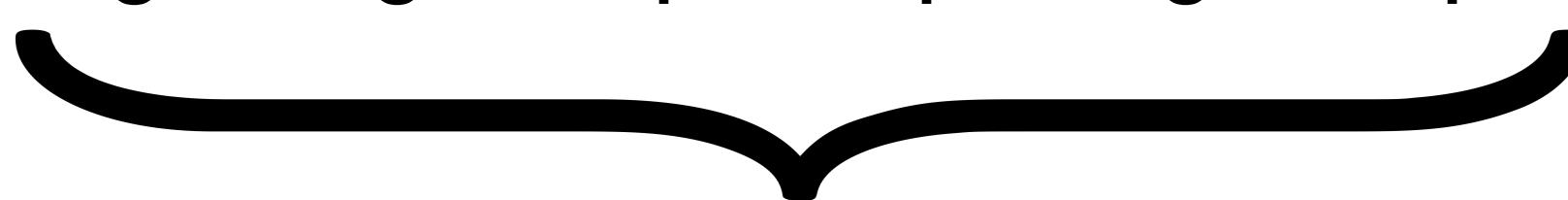
0

1

1

0

1



sum ~ Binomial(n, p)

The binomial distribution

- What is the average number of heads I expect to get?
- What is the probability that I will get exactly 4 heads?
- What is the probability that I will get 4 or more heads?
- What is the probability that I will get 3 or less heads?

Exercises in R

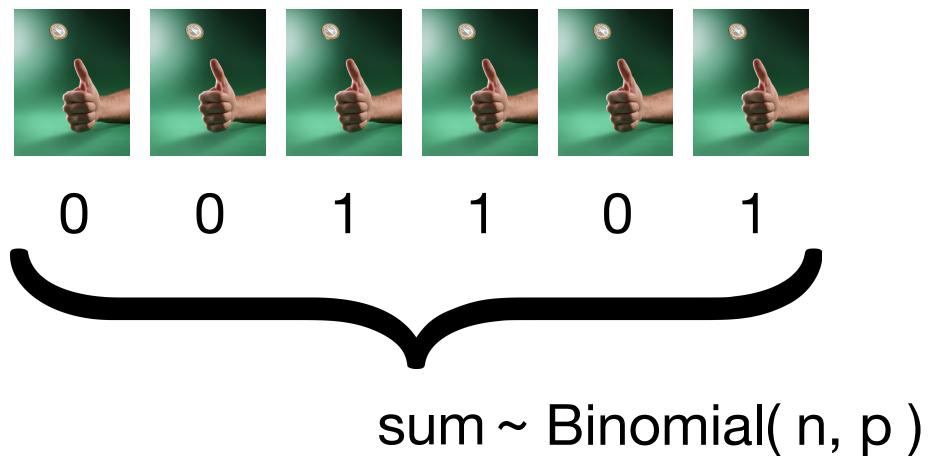
- Bernoulli distribution: tossing a coin
- Binomial distribution: adding up coin tosses



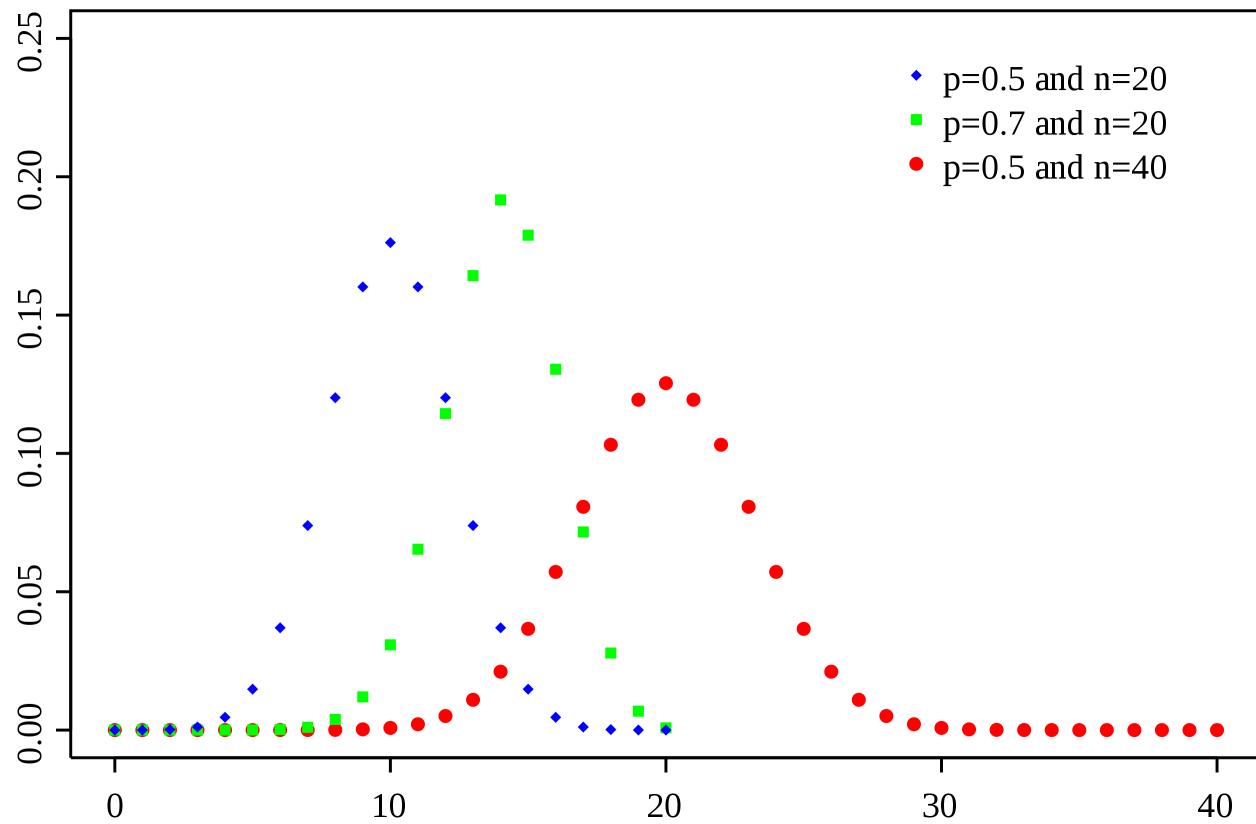
The Binomial distribution

- Two parameters: n and p

- $P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$



The binomial distribution



Two broad families of random variables

Discrete

- Examples:
 - Single coin flip (Bernoulli)
 - Number of heads in a series of flips (Binomial)
 - Number of times I roll a dice and get 4 or more
 - Number of trials until I roll a six
 - Number of rain drops in a bucket
 - Number of trees in a forest

Two broad families of random variables

Discrete

- Examples:
 - Single coin flip (Bernoulli)
 - Number of heads in a series of flips (Binomial)
 - Number of times I roll a dice and get 4 or more
 - Number of trials until I roll a six
 - Number of rain drops in a bucket
 - Number of trees in a forest

Continuous

- Examples:
 - Height in a population
 - Waiting time till the next bus arrives
 - Time until a particle decays
 - Temperature in a room
 - Lifespan of a species
 - Prevalence of a disease

Two broad families of random variables

Discrete

- Examples:
 - Single coin flip (Bernoulli)
 - Number of heads in a series of flips (Binomial)
 - Number of times I roll a dice and get 4 or more
 - Number of trials until I roll a six
 - Number of rain drops in a bucket
 - Number of trees in a forest

Continuous

- Examples:
 - Height in a population
 - Waiting time till the next bus arrives
 - Time until a particle decays
 - Temperature in a room
 - Lifespan of a species
 - Prevalence of a disease

We can model all of these (and much more) using common probability distributions

The Expectation of a Random Variable

- $E[X] = \sum_i x_i P[X = x_i]$

Properties of the expectation

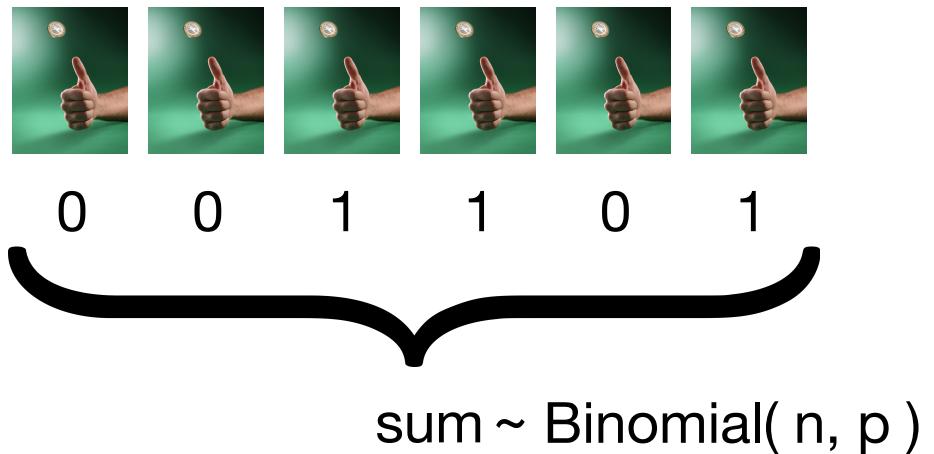
- If a is a constant: $E[aX] = aE[X]$
- If a is a constant: $E[X + a] = E[X] + a$
- For two random variables X and Y : $E[X + Y] = E[X] + E[Y]$
- More generally: $E[\sum X_i] = \sum E[X_i]$

The Average or Sample Mean

- $E[X] = \sum_i x_i P[X = x_i]$
- The average or sample mean (\bar{x}) is an approximation to $E[X]$ when one has observed N samples
 - $\bar{x} = \sum_{i=1}^N \frac{x_i}{N}$

The Binomial distribution

- $P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$
- $E[X] = np$



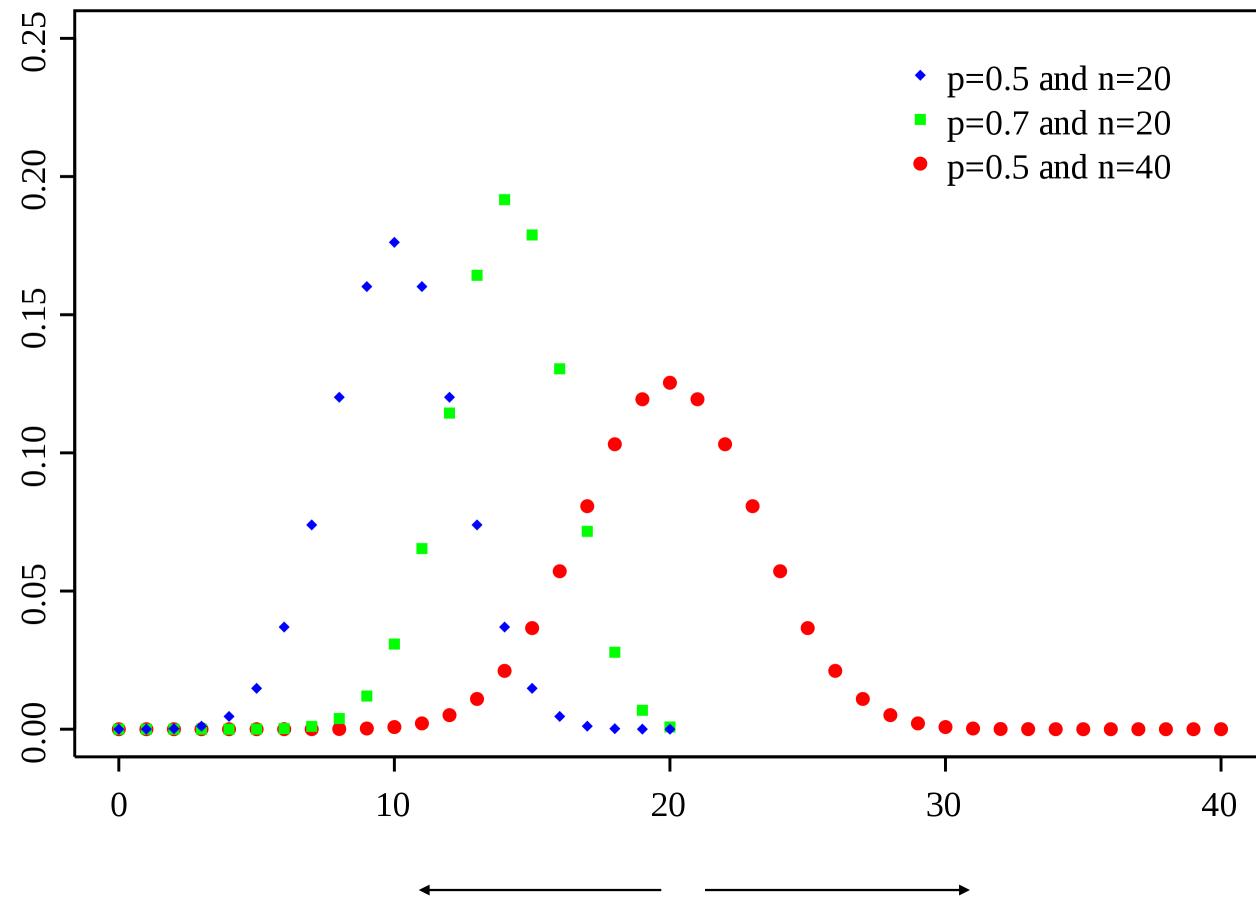
The Expectation - exercises

- In the case of the Bernoulli distribution: $E[X] = p$...LET'S PROVE THIS!
- In the case of the Binomial distribution: $E[X] = np$...LET'S PROVE THIS!

The Variance

- $Var[X] = E[(X - E[X])^2]$
- This can also be written as: $Var[X] = E[X^2] - E[X]^2$

The Variance



The Sample Variance

- $Var[X] = E[(X - E[X])^2]$
- The sample variance (s^2) is an unbiased approximation to $Var[X]$
- $$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n - 1}$$

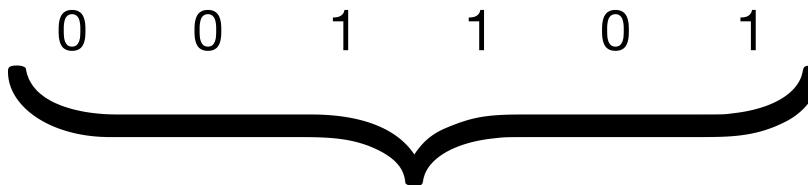
The Sample Variance

- $Var[X] = E[(X - E[X])^2]$
- The sample variance (s^2) is an unbiased approximation to $Var[X]$
- $$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n - 1}$$

Why “n-1”, and not “n”? See: <https://web.ma.utexas.edu/users/mks/M358KInstr/SampleSDPf.pdf>

The Binomial distribution

- $P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$



sum ~ Binomial(n, p)

- $E[X] = np$

- $Var[X] = np(1 - p)$

Exercises in R

- The Expectation
- Our first probability mass function
- The Variance



The Geometric distribution

The Geometric distribution



0

The Geometric distribution



0



0

The Geometric distribution



0



0



0

The Geometric distribution



0



0



0



0

The Geometric distribution



0



0



0



0



0

The Geometric distribution



0



0



0



0



0



1

The Geometric distribution



0

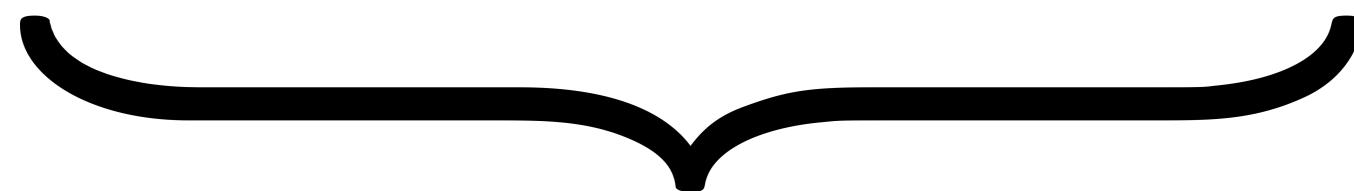
0

0

0

0

1



no. tails until 1 heads

The Geometric distribution



0

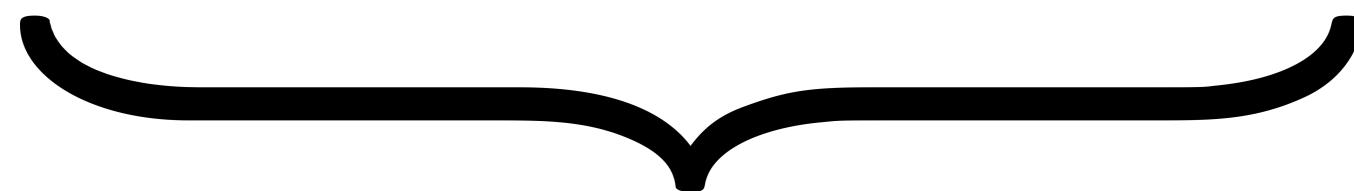
0

0

0

0

1



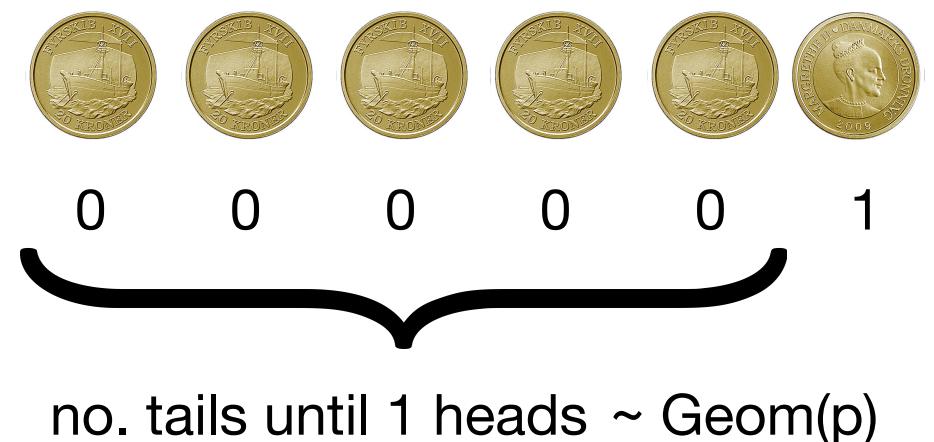
no. tails until 1 heads $\sim \text{Geom}(p)$

The Geometric distribution

$$\text{PMF: } P[T = t] = \underbrace{(1 - p)^{t-1} p}_{\substack{t-1 \text{ failures} \\ 1 \text{ success}}}$$

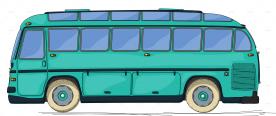
$$\text{Expectation: } E[T] = \frac{1 - p}{p}$$

$$\text{Variance: } Var[T] = \frac{(1 - p)}{p^2}$$



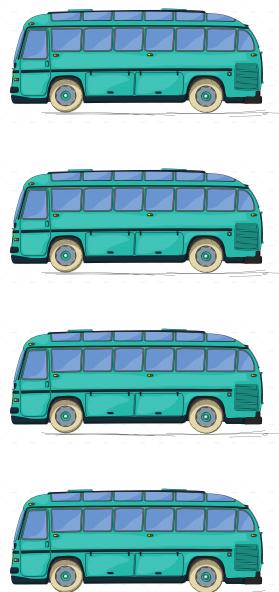
The Poisson distribution

- How many buses arrive at this station every hour?

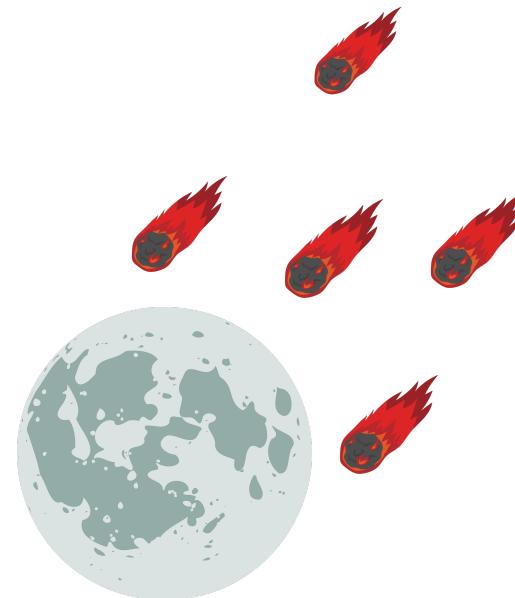


The Poisson distribution

- How many buses arrive at this station every hour?

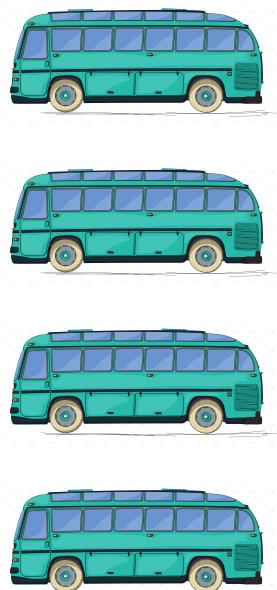


- How many meteors hit the moon every year?

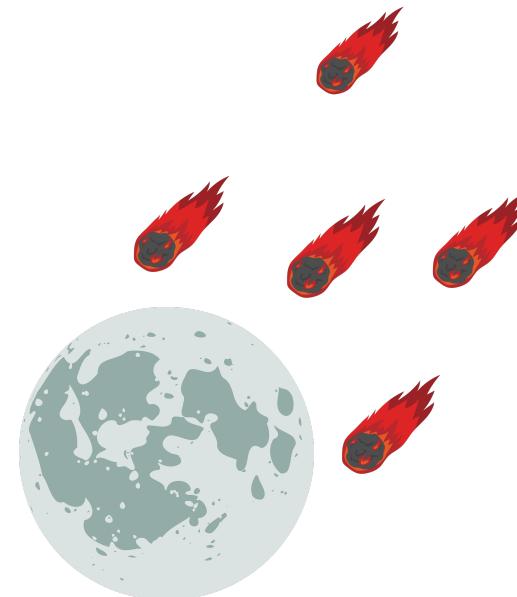


The Poisson distribution

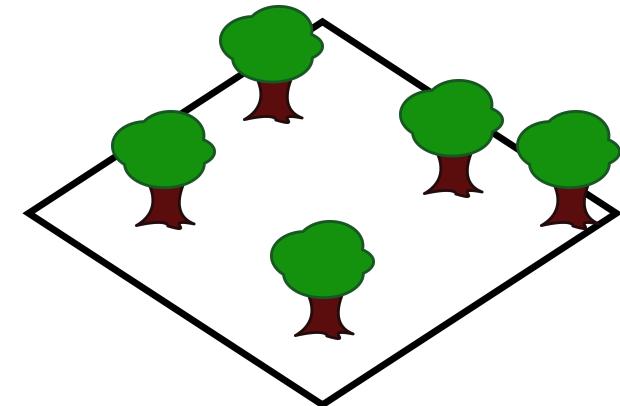
- How many buses arrive at this station every hour?



- How many meteors hit the moon every year?



- How many trees are there in a hectare of land?

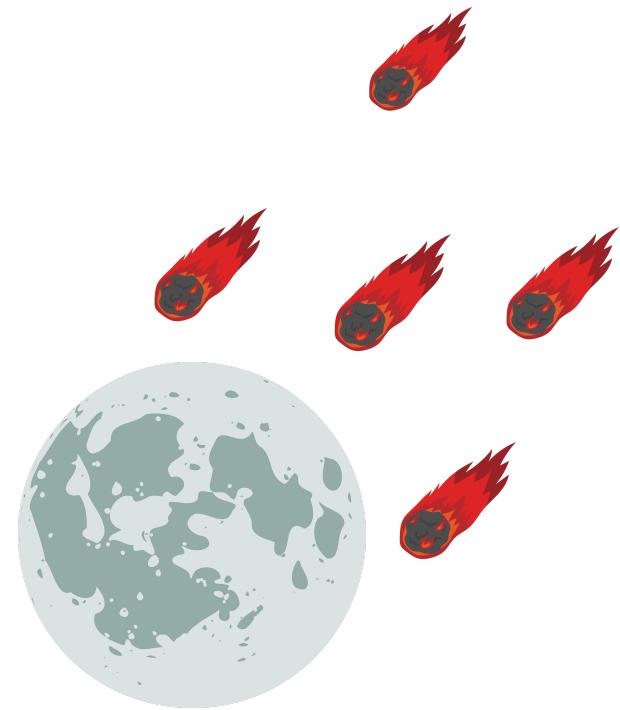


The Poisson distribution

- Events or objects that occur **over a given stretch** of time and/or space
- The **rate or intensity** of occurrence (per unit of time and/or space) is the **same for all** events or objects
- Each event or object occurs **independently** of the other events or objects

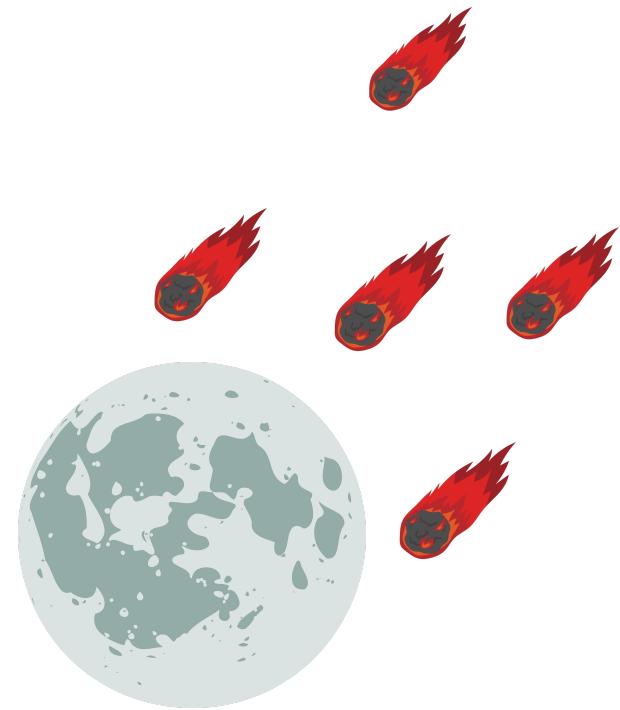
The Poisson distribution

- **Average rate** at which meteorites fall on a planet is 5 per year
- Some years, there might be 4 meteorites. Other years, there might be 6, or 3, or 5, or 7, etc...
- It is **technically possible** that 1,000,000 meteorites fall on a given year, but **extremely unlikely**

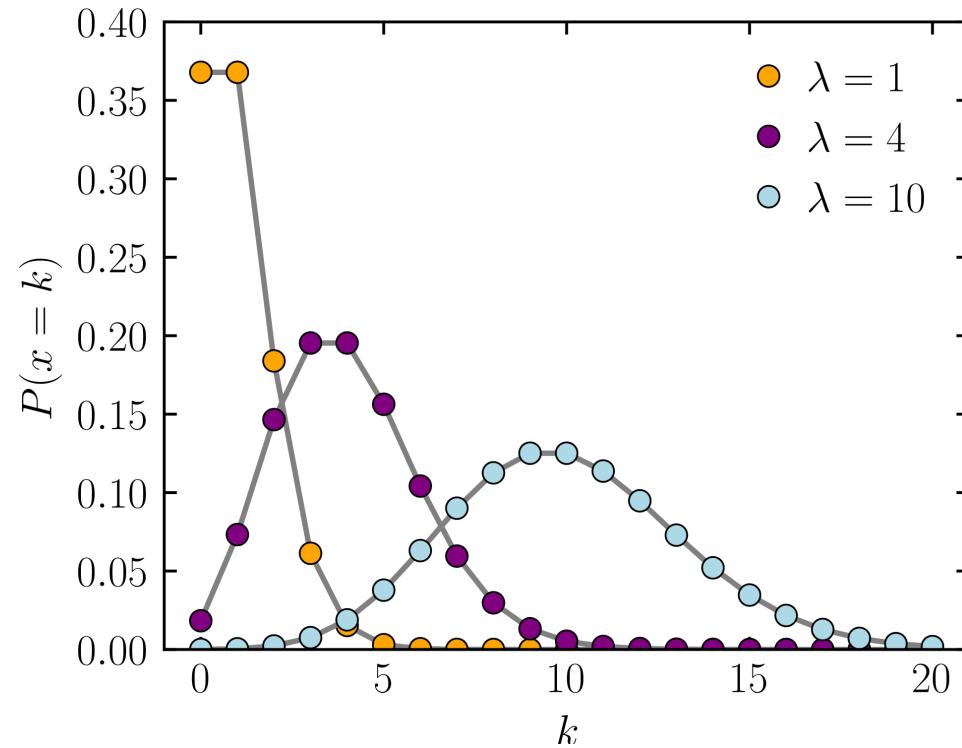


The Poisson distribution

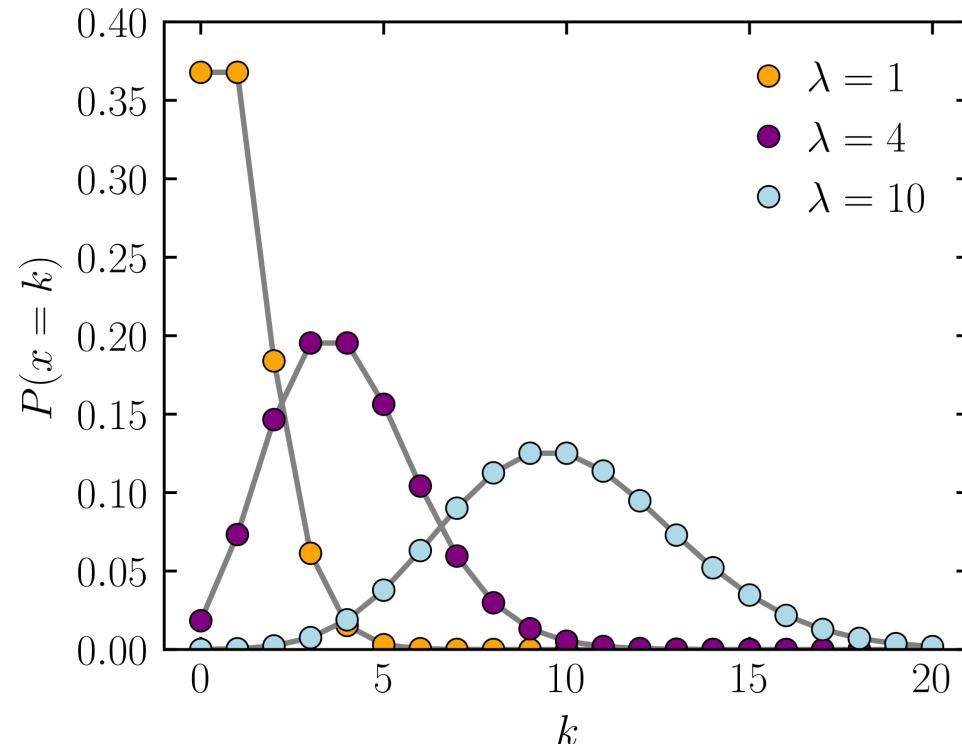
- One parameter: λ (“intensity” or “rate” of an event or object)
- $E[X] = \lambda$
- $Var[X] = \lambda$
- $P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$



The Poisson distribution: PMF



The Poisson distribution: PMF



NOTE: unlike the Binomial distribution, there is no “maximum” number of possible events

The Uniform distribution (discrete)

- Every integer in a given range is **equally likely**
- Example: throwing a fair dice
- Two parameters: a and b , with $b \geq a$
- For example, in the case of a dice, $a=1$ and $b=6$, so the sample set is 1, 2, 3, 4, 5 and 6.



The Uniform distribution (discrete)

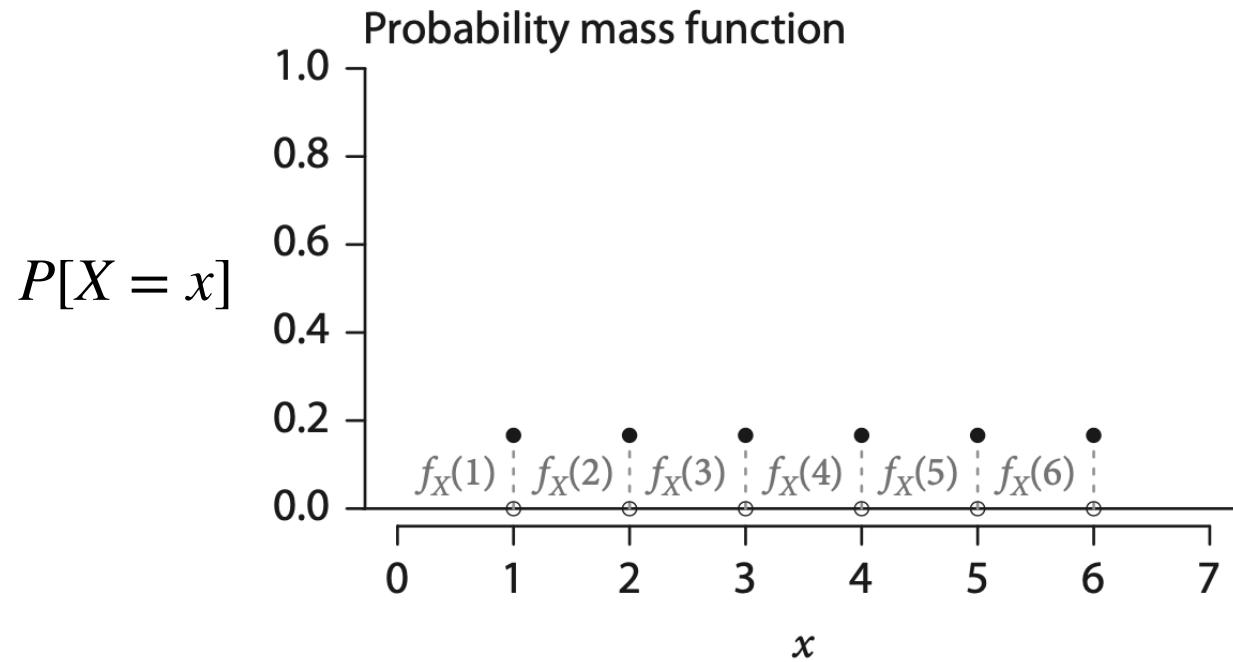
- Two parameters: a and b , with $b \geq a$

- $E[X] = \frac{a + b}{2}$

- $P[X = k] = \frac{1}{n}$, where $n = b - a + 1$



The Uniform distribution (discrete)



Discrete distributions re-cap

- Bernoulli (p)
- Binomial (n, p)
- Geometric (p)
- Poisson (λ)
- Uniform $[a,b]$

Discrete distributions re-cap

- Bernoulli (p) —————> “What is the probability I get heads?”
- Binomial (n, p)
- Geometric (p)
- Poisson (λ)
- Uniform $[a,b]$

Discrete distributions re-cap

- Bernoulli (p) —————> “What is the probability I get heads?”
- Binomial (n, p) —————> “What is the probability I get 5 heads in 20 tosses?”
- Geometric (p)
- Poisson (λ)
- Uniform $[a,b]$

Discrete distributions re-cap

- Bernoulli (p) —————> “What is the probability I get heads?”
- Binomial (n, p) —————> “What is the probability I get 5 heads in 20 tosses?”
- Geometric (p) —————> “What is the probability I get 8 tails before 1 head?”
- Poisson (λ)
- Uniform $[a,b]$

Discrete distributions re-cap

- Bernoulli (p) —————> “What is the probability I get heads?”
- Binomial (n, p) —————> “What is the probability I get 5 heads in 20 tosses?”
- Geometric (p) —————> “What is the probability I get 8 tails before 1 head?”
- Poisson (λ) —————> “What is the probability that 5 buses arrive over the course of this hour?”
- Uniform $[a,b]$

Discrete distributions re-cap

- Bernoulli (p) → “What is the probability I get heads?”
- Binomial (n, p) → “What is the probability I get 5 heads in 20 tosses?”
- Geometric (p) → “What is the probability I get 8 tails before 1 head?”
- Poisson (λ) → “What is the probability that 5 buses arrive over the course of this hour?”
- Uniform $[a,b]$ → “What is the probability I get a “3” in a dice toss?”

Exercises in R

- Playing with discrete distributions

