Probability & Statistics for DS & AI

Random Variables

Michele Guindani

Summer

Mathematicians are lazy!

Mathematicians are lazy!

Example (Flipping one coin)

$$\text{"probability of obtaining a head"} = \mathbb{P}[\text{ "Head"}]$$

Mathematicians are lazy!

Example (Flipping one coin)

$$\text{“probability of obtaining a head”} = \mathbb{P}[\text{ “Head”}]$$

... Easy...! 😜

Mathematicians are lazy!

Example (Flipping one coin)

$$\text{“probability of obtaining a head”} = \mathbb{P}[\text{ “Head”}]$$

... Easy...! 😜

Example (Flipping three coins)

$$\text{“probability of obtaining 3 heads”} = \mathbb{P}[\text{ “Head”} \cap \text{“Head”} \cap \text{“Head” }]$$

Mathematicians are lazy!

Example (Flipping one coin)

"probability of obtaining a head" $= \mathbb{P}[$ "Head"$]$

... Easy...! 😜

Example (Flipping three coins)

"probability of obtaining 3 heads" $= \mathbb{P}[$ "Head" $\cap$ "Head" $\cap$ "Head" $]$

...😒...🤒...😩...

Mathematicians are lazy!

Example (Flipping one coin)

"probability of obtaining a head" $= \mathbb{P}[\text{ "Head"}]$

... Easy...! 😜

Example (Flipping three coins)

"probability of obtaining 3 heads" $= \mathbb{P}[\text{ "Head"} \cap \text{"Head"} \cap \text{"Head" }]$

...😒...🤒...😩...

- Mathematicians say: Call "Head" $= 1$ and "Tail" $= 0$.

Mathematicians are lazy!

Example (Flipping one coin)

$$\text{“probability of obtaining a head”} = \mathbb{P}[\text{ “Head”}]$$

... Easy...! 😜

Example (Flipping three coins)

$$\text{“probability of obtaining 3 heads”} = \mathbb{P}[\text{ “Head”} \cap \text{“Head”} \cap \text{“Head” }]$$

...😒...🤒...😩...

- Mathematicians say: Call “Head” $= 1$ and “Tail” $= 0$. Let X be the sum of heads. Then

$$\text{“probability of obtaining a head”} = \mathbb{P}[X = 1]$$

$$\text{“probability of obtaining 3 heads”} = \mathbb{P}[X = 3]$$

🥰 🥰 🥰 🥰

Everything you need to know about a random variable

- What are random variables?
- Random variables are functions that translate words to numbers!
- **Example:** "Head" is a word description ⇨ "$X = 1$" is a numerical description

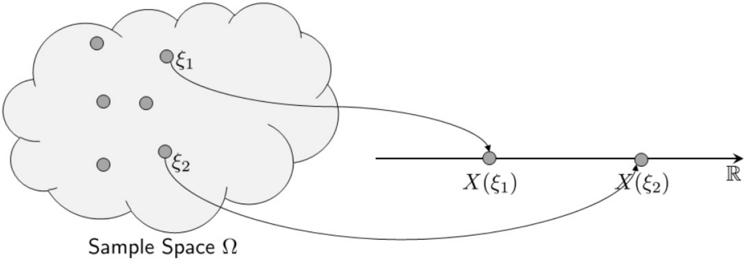


Figure: Illustration of a random variable.

The formal definition

Random variable

A random variable X is a function $X : \Omega \to \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

Example (Flipping a coin twice)

- The sample space Ω is

$$\Omega = \{(\mathrm{HH}), (\mathrm{HT}), (\mathrm{TH}), (\mathrm{TT})\}$$

Four events $\xi_1 = \mathrm{HH}, \xi_2 = \mathrm{HT}, \xi_3 = \mathrm{TH}, \xi_4 = \mathrm{TT}$

The formal definition

Random variable

A random variable X is a function $X : \Omega \to \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

Example (Flipping a coin twice)

- The sample space Ω is

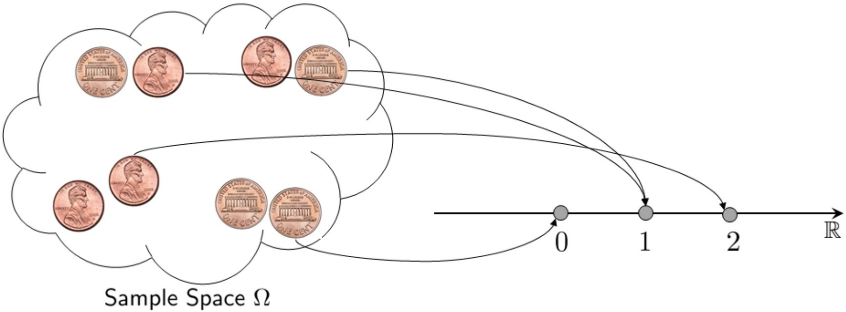
$$\Omega = \{(\mathrm{HH}), (\mathrm{HT}), (\mathrm{TH}), (\mathrm{TT})\}$$

  Four events $\xi_1 = \mathrm{HH}, \xi_2 = \mathrm{HT}, \xi_3 = \mathrm{TH}, \xi_4 = \mathrm{TT}$

  Then, let $X =$ number of H.

- $X(\xi_1) = \#$ of heads in $\{HH\} = 2$
- $X(\xi_2) = \#$ of heads in $\{HT\} = 1$
- $X(\xi_3) = \#$ of heads in $\{TH\} = 1$
- $X(\xi_4) = \#$ of heads in $\{TT\} = 0$

Map from Ω to X



A random variable that maps a pair of coins to a number, where the number represents the number of heads.

Calculating the probability

Example

🤔 How to "calculate" $\mathbb{P}[X = 1]$?

Calculating the probability

Example

- 🤔 How to "calculate" $\mathbb{P}[X = 1]$?
  - 🤔 Is $\{X = 1\}$ an event in the event space?

Calculating the probability

Example

🤔 How to "calculate" $\mathbb{P}[X = 1]$?

- 🤔 Is $\{X = 1\}$ an event in the event space?
- ❌ No. The sample space is

$$\Omega = \{(\mathrm{HH}), (\mathrm{HT}), (\mathrm{TH}), (\mathrm{TT})\}$$

The event space is the combination of these outcomes.

Calculating the probability

Example

🤔 How to "calculate" $\mathbb{P}[X = 1]$?

- 🤔 Is $\{X = 1\}$ an event in the event space?
- ❌ No. The sample space is

$$\Omega = \{(\mathrm{HH}), (\mathrm{HT}), (\mathrm{TH}), (\mathrm{TT})\}$$

The event space is the combination of these outcomes.

- ⚠️ "$\{X = 1\}$" is not the same as HT or TH.

"$\{X = 1\}$" lives in the translated space.

Calculating the probability

Example

- How to "calculate" $\mathbb{P}[X = 1]$?
  - Is $\{X = 1\}$ an event in the event space?
  - No. The sample space is

$$\Omega = \{(\mathrm{HH}), (\mathrm{HT}), (\mathrm{TH}), (\mathrm{TT})\}$$

    The event space is the combination of these outcomes.

  - "$\{X = 1\}$" is not the same as HT or TH.

    "$\{X = 1\}$" lives in the translated space.

  - ⇨ So how to measure the probability?

Calculating the probability

Example

⇨ So how to measure the probability?

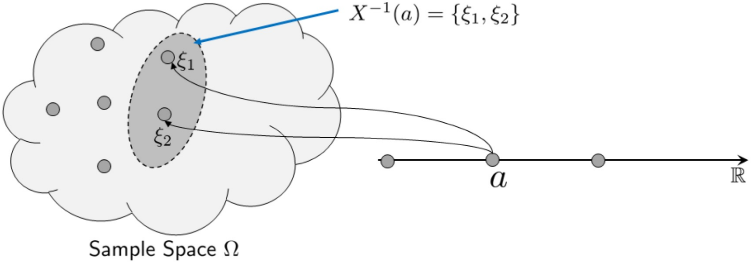
- Map $\{X = 1\}$ back to the sample space!

$$\mathbb{P}[\{X = 1\}] = \mathbb{P}[\{(\mathrm{HT}) \cup (\mathrm{TH})\}] = \frac{2}{4}$$

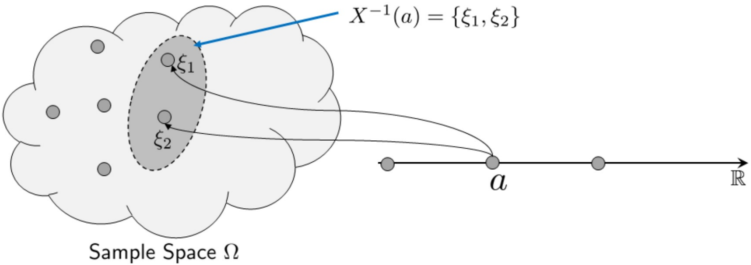
- Essentially, you are finding ξ such that $X(\xi) = 1$. So,

$$\mathbb{P}[X = 1] = \mathbb{P}[X(\xi) = 1] = \mathbb{P}\left[\xi = X^{-1}(1)\right] = \mathbb{P}[\{(\mathrm{HT}) \cup (\mathrm{TH})\}] = \frac{2}{4}$$

When calculating the probability, go backward to the sample space!



When calculating the probability, go backward to the sample space!



- In practice, this is just the intuition, you do not need to worry about these.

⇨ You never do this translation per se.

- See also Ex. 3.3 and 3.4 in the textbook (throwing dice)

Probability & Statistics for DS & AI

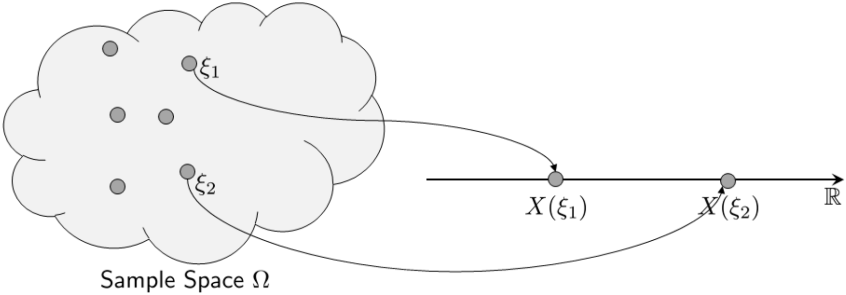
Probability Mass Functions

Michele Guindani

Summer

Problem

Random variables are functions that translate words to numbers!



- So many numbers
- So many events
- How to systematically describe them?

Probability mass function (PMF)

The probability mass function (PMF) of a random variable X is a function which specifies the probability of obtaining a number $X(\xi) = a$. We denote a PMF as

$$p_X(a) = \mathbb{P}[X = a]$$

- There are two functions here:
  - Function X : the random variable which translates words to numbers
  - Function p_X : the mapping from event $\{X = a\}$ to a probability

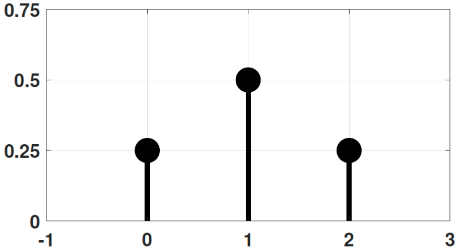
⚠ Difference between X and a :

- X is the random variable. Technically it should be $X(\xi)$
- a is a state. So $X = a$ means X is taking the state a.

Example (Flipping a coin twice)

Define X = number of heads. Then the probability mass function is

$$p_X(0) = \mathbb{P}[X = 0] = \mathbb{P}[\{\ \text{"TT"}\ \}] = \frac{1}{4}$$
$$p_X(1) = \mathbb{P}[X = 1] = \mathbb{P}[\{\ \text{"TH", "HT"}\ \}] = \frac{1}{2}$$
$$p_X(2) = \mathbb{P}[X = 2] = \mathbb{P}\left[\{\text{HH}''\}\right] = \frac{1}{4}$$

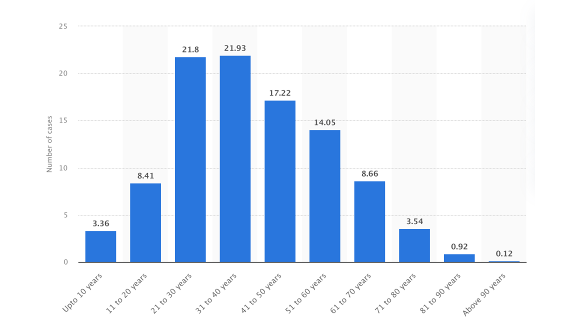


Theoretical distribution of the number X of heads

Ideal (Theoretical) vs practical (empirical; data-based)

- When we look at data, one way to summarize data is through a histogram

Example (# of Covid-19 cases in India (October, 21st))

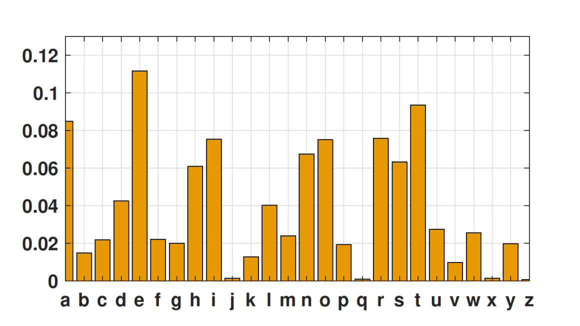


- The histogram is constructed from the data (empirical). However, data are often the result of a selection (e.g., a random sample from a population)

Ideal (Theoretical) vs practical (empirical; data)

- When we look at data, one way to summarize data is through a histogram

Example (Frequency of 26 English Letters)



- The histogram is constructed from the data (empirical). However, data are often the result of a selection (e.g., a random sample from a population)

Histogram vs PMF

Example (Die is thrown N times)

☞ Assuming that the die is fair, the PMF is simply

$$p_X(k) = 1/6$$

for $k = 1, \ldots, 6$, which is a uniform distribution across the 6 states (theoretical, ideal)

Histogram vs PMF

Example (Die is thrown N times)

☞ Assuming that the die is fair, the PMF is simply

$$p_X(k) = 1/6$$

for $k = 1, \ldots, 6$, which is a uniform distribution across the 6 states (theoretical, ideal)

- Let's investigate the frequency of the different faces when we generate data with increasing values of N (practical)

Histogram vs PMF

Example (Die is thrown N times)

- Assuming that the die is fair, the PMF is simply

$$p_X(k) = 1/6$$

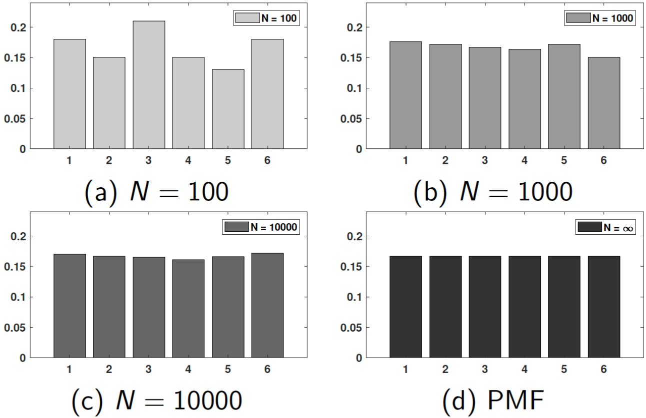
  for $k = 1, \ldots, 6$, which is a uniform distribution across the 6 states (theoretical, ideal)

- Let's investigate the frequency of the different faces when we generate data with increasing values of N (practical)

Python code to generate the histogram

```
import numpy as np
import matplotlib.pyplot as plt
q = np.random.randint(low=1, high=7, size=100)
plt.hist(q+0.5,bins=6)
plt.show()
```

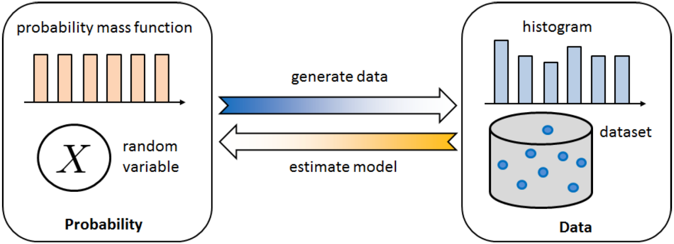
PMF is the **ideal** histogram!



Histogram and PMF, when throwing a fair dice N times. As N increases, the histograms are becoming more like the PMF.

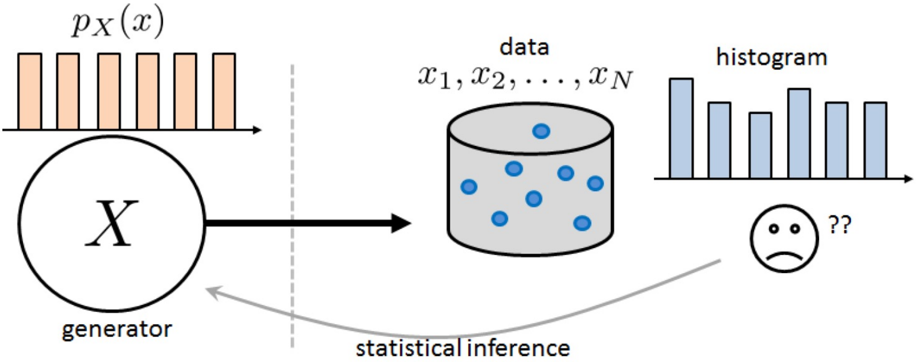
Why bother to study the PMF?

- Ideal vs empirical. You always want something ideal.
- Modeling the data. We can't model the data (understand where the data we observe come from) from a simple histogram!



When analyzing a dataset, one can treat the data points are samples drawn according to a latent random variable with certain a PMF. The dataset we observe is often finite, and so the histogram we obtain is empirical. A major task in data analysis is statistical inference, which tries to retrieve the model information from the available measurements.

Why bother to study the PMF?



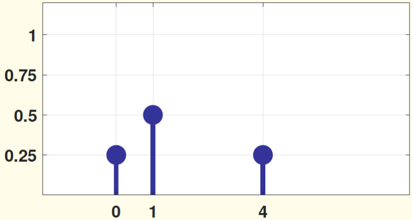
Cumulative distribution function (CDF)

- In some cases, instead of the PMF, it may be useful to look (equivalently) at the CDF

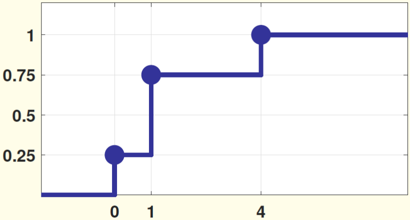
Example

Consider a random variable X with PMF

$$p_X(0) = \frac{1}{4}, \quad p_X(1) = \frac{1}{2}, \quad p_X(4) = \frac{1}{4}$$



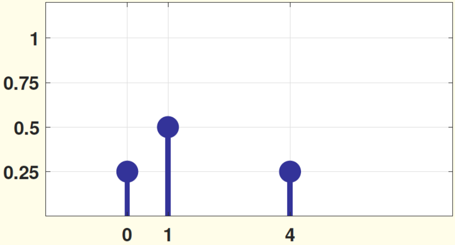
(a) PMF $p_X(k)$



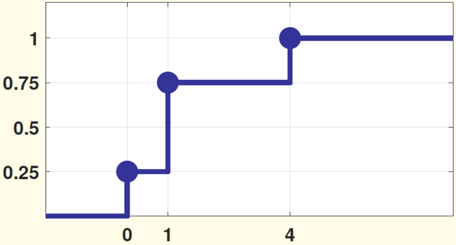
(b) CDF $F_X(k)$

Cumulative distribution function (CDF)

Example



(a) PMF $p_X(k)$



(b) CDF $F_X(k)$

The CDF of X can be computed as

$$F_X(0) = \mathbb{P}[X \leq 0] = p_X(0) = \frac{1}{4},$$

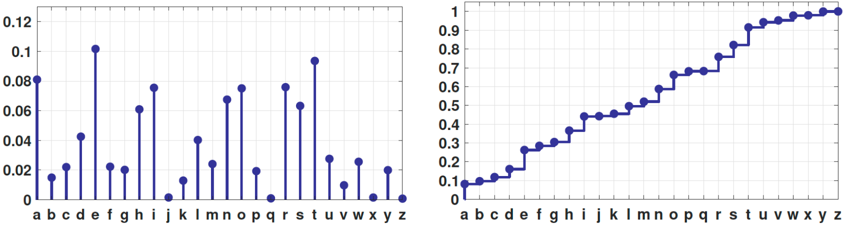
$$F_X(1) = \mathbb{P}[X \leq 1] = p_X(0) + p_X(1) = \frac{3}{4},$$

$$F_X(4) = \mathbb{P}[X \leq 4] = p_X(0) + p_X(1) + p_X(4) = 1.$$

Python code to generate a PMF and a CDF

```
import numpy as np
import matplotlib.pyplot as plt
p = np.array([0.25, 0.5, 0.25])
x = np.array([0, 1, 4])
F = np.cumsum(p)
plt.stem(x,p,use_line_collection=True); plt.show()
plt.step(x,F); plt.show()
```

Example (English letters)



(a) PMF $p_X(k)$ (b) CDF $F_X(k)$

The **cumulative distribution function** (CDF) of a discrete random variable X is

$$F_X(x) \stackrel{\text{def}}{=} \mathbb{P}[X \leq x] = \sum_{x' \leq x} p_X(x').$$

Expectation

- It is often useful to summarize a PMF through some simple quantities

Expectation

- It is often useful to summarize a PMF through some simple quantities
- The expectation of a random variable provides a measure of central tendency, or average, of a PMF

Expectation

- It is often useful to summarize a PMF through some simple quantities
- The expectation of a random variable provides a measure of central tendency, or average, of a PMF

Key Concept : What is expectation?

Expectation = Mean = Average computed from a PMF.

!! The expectation characterizes a !! theoretical !! property of the probability mass function.

💪 In DS, it can be seen as the true average of the population that we are trying to estimate with our data.

Expectation

The expectation of a random variable X is

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \, p_X(x)$$

Example

Let X be a random variable with PMF $p_X(0) = 1/4, p_X(1) = 1/2$ and $p_X(2) = 1/4$. The expectation is

$$\mathbb{E}[X] = (0) \underbrace{\left(\frac{1}{4}\right)}_{p_X(0)} + (1) \underbrace{\left(\frac{1}{2}\right)}_{p_X(1)} + (2) \underbrace{\left(\frac{1}{4}\right)}_{p_X(2)} = 1$$

Expectation

The expectation of a random variable X is

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \, p_X(x)$$

Example

Let X be a random variable with PMF $p_X(0) = 1/4, p_X(1) = 1/2$ and $p_X(2) = 1/4$. The expectation is

$$\mathbb{E}[X] = (0) \underbrace{\left(\frac{1}{4}\right)}_{p_X(0)} + (1) \underbrace{\left(\frac{1}{2}\right)}_{p_X(1)} + (2) \underbrace{\left(\frac{1}{4}\right)}_{p_X(2)} = 1$$

Python code to compute the expectation

```
import numpy as np
p = np.array([0.25, 0.5, 0.25])
x = np.array([0, 1, 2])
EX = np.sum(p*x)
```

Example (Throwing a dice once)

In this case the expectation is

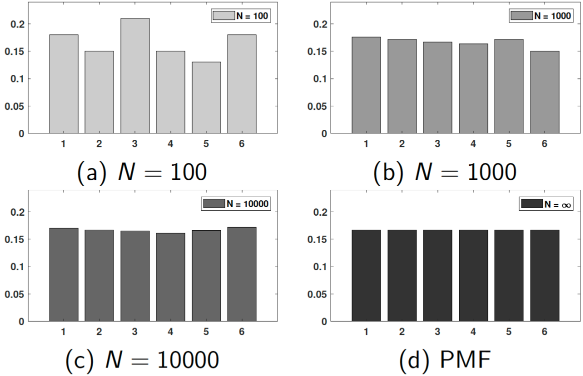
$$E(X) = \sum_{x \in X(\Omega)} x\, p_X(x) = \sum_{i=1}^{6} i\, \frac{1}{6} = 3.5$$

Python code

```
import numpy as np
p = np.array([1/6, 1/6, 1/6, 1/6, 1/6, 1/6])
x = np.array([1, 2, 3, 4, 5, 6])
EX = np.sum(p*x)
```

Expectation = "Average"?

No. Expectation is computed from PMF. Average is computed from histogram.



Histogram and PMF, when throwing a fair dice N times. As N increases, the histograms are becoming more like the PMF.

Expectation = "Average"?

Try the following in Python to generate data and compute expectations

```
import numpy as np
np.random.seed(12345)
#try different random seeds to generate other values
q_10 = np.random.randint(low=1, high=7, size=10)
average_q_10 = np.mean(q_10); print(average_q_10)
#4.1
q_100 = np.random.randint(low=1, high=7, size=100)
average_q_100 = np.mean(q_100); print(average_q_100)
#3.57
q_1000000 = np.random.randint(low=1, high=7, size=10000)
average_q_1000000 = np.mean(q_1000000); print(average_q_1000000)
#3.506
```

Expectation = “Average”?

Try the following in Python to generate data and compute expectations

```
import numpy as np
np.random.seed(12345)
#try different random seeds to generate other values
q_10 = np.random.randint(low=1, high=7, size=10)
average_q_10 = np.mean(q_10); print(average_q_10)
#4.1
q_100 = np.random.randint(low=1, high=7, size=100)
average_q_100 = np.mean(q_100); print(average_q_100)
#3.57
q_1000000 = np.random.randint(low=1, high=7, size=10000)
average_q_1000000 = np.mean(q_1000000); print(average_q_1000000)
#3.506
```

- Check out Examples 3.11 and 3.12 in the textbook. Build simulations where you generate data and approximate the theoretical expectations.

True Mean and Sample Mean

True Mean $\mathbb{E}[X]$

- A statistical property of a random variable.
- A deterministic number.
- Often unknown, or is the center question of estimation.
- You have to know X in order to find $\mathbb{E}[X]$; Top down.

Sample Mean $\bar{X}$

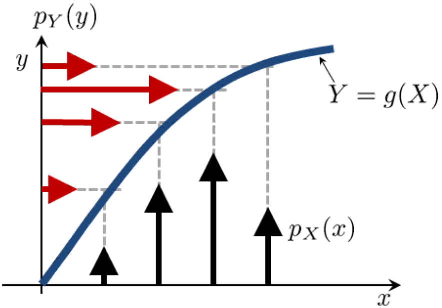
- A numerical value. Calculated from data.
- Itself is a random variable.
- It has uncertainty.
- Uncertainty reduces as more samples are used.
- We use sample mean to estimate the true mean.
- You do not need to know X in order to find $\bar{X}$; Bottom up.

Properties of $\mathbb{E}[X]$

Property (1. Function of X)

For any function g,

$$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x).$$



Properties of $\mathbb{E}[X]$

Property (2. Linearity)

For any function g and h,

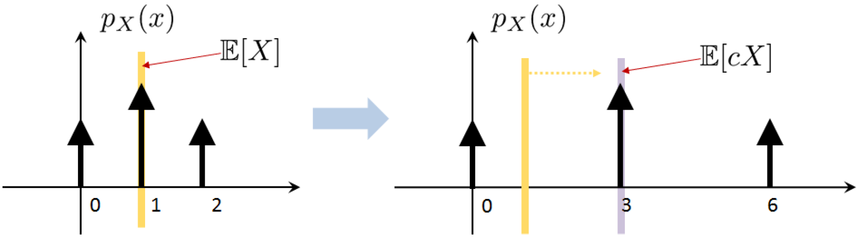
$$\mathbb{E}[g(X) + h(X)] = \mathbb{E}[g(X)] + \mathbb{E}[h(X)].$$

Properties of $\mathbb{E}[X]$

Property (3. Scale)

For any constant c,

$$\mathbb{E}[cX] = c\mathbb{E}[X].$$

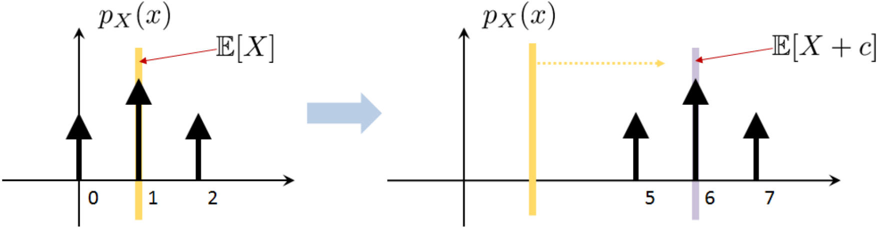


Properties of $\mathbb{E}[X]$

Property (4. DC Shift)

For any constant c,

$$\mathbb{E}[X + c] = \mathbb{E}[X] + c.$$



Variance

⚠ An other important quantity that summarizes a distribution

Variance

⚠ An other important quantity that summarizes a distribution

Variance of a random variable X

$$\sigma^2 = \mathrm{Var}[X] = \mathbb{E}\left[(X-\mu)^2\right]$$

The square root of the variance, $\sigma = \sqrt{\mathrm{Var}[X]}$, is called the standard deviation.

Variance

⚠ An other important quantity that summarizes a distribution

Variance of a random variable X

$$\sigma^2 = \mathrm{Var}[X] = \mathbb{E}\left[(X-\mu)^2\right]$$

The square root of the variance, $\sigma = \sqrt{\mathrm{Var}[X]}$, is called the standard deviation.

🤔 What does the variance mean?

Variance

⚠ An other important quantity that summarizes a distribution

Variance of a random variable X

$$\sigma^2 = \mathrm{Var}[X] = \mathbb{E}\left[(X-\mu)^2\right]$$

The square root of the variance, $\sigma = \sqrt{\mathrm{Var}[X]}$, is called the standard deviation.

🤔 What does the variance mean?

It is a measure of the deviation (spread) of the random variable X relative to its mean.

☞ This deviation is quantified by the squared difference $(X-\mu)^2$.

The expectation operator takes the average of the deviation, giving us a deterministic number $\mathbb{E}\left[(X-\mu)^2\right]$.

Variance

- It is often convenient to use the following equivalent definition of Variance:

Variance - alternative definition (Theorem)

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Variance

- It is often convenient to use the following equivalent definition of Variance:

Variance - alternative definition (Theorem)

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Example (Coin flipping)

Flip a coin with probability p to get a head. Let X be a random variable denoting the outcome. The PMF of X is

$$p_X(0) = 1 - p, \quad p_X(1) = p$$

Find $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

Example (Coin flipping - ctd)

The expectation of X is

$$\mathbb{E}[X] = (0)p_X(0) + (1)p_X(1)$$

Example (Coin flipping - ctd)

The expectation of X is

$$\mathbb{E}[X] = (0)p_X(0) + (1)p_X(1) = (0)(1-p) + (1)(p) = p$$

Example (Coin flipping - ctd)

The expectation of X is

$$\mathbb{E}[X] = (0)p_X(0) + (1)p_X(1) = (0)(1-p) + (1)(p) = p$$

In order to compute the variance we need to compute $\mathbb{E}\left[X^2\right]$ first:

$$\mathbb{E}\left[X^2\right] = (0)^2 p_X(0) + (1)^2 p_X(1) = p$$

Example (Coin flipping - ctd)

The expectation of X is

$$\mathbb{E}[X] = (0)p_X(0) + (1)p_X(1) = (0)(1-p) + (1)(p) = p$$

In order to compute the variance we need to compute $\mathbb{E}\left[X^2\right]$ first:

$$\mathbb{E}\left[X^2\right] = (0)^2 p_X(0) + (1)^2 p_X(1) = p$$

Then, the Variance is obtained as

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Example (Coin flipping - ctd)

The expectation of X is

$$\mathbb{E}[X] = (0)p_X(0) + (1)p_X(1) = (0)(1-p) + (1)(p) = p$$

In order to compute the variance we need to compute $\mathbb{E}\left[X^2\right]$ first:

$$\mathbb{E}\left[X^2\right] = (0)^2 p_X(0) + (1)^2 p_X(1) = p$$

Then, the Variance is obtained as

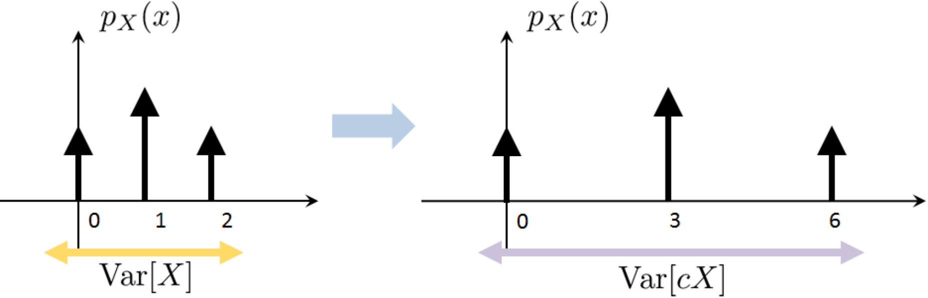
$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = p - p^2 = p\,(1-p)$$

The standard deviation $\sigma = \sqrt{\mathrm{Var}[X]} = \sqrt{p\,(1-p)}$.

Properties of the Variance

(**Scale**) For any constant c

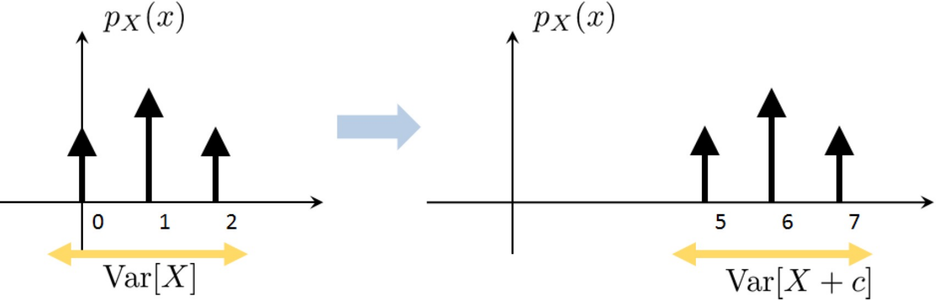
$$\mathrm{Var}[cX] = c^2\,\mathrm{Var}[X]$$



Properties of the Variance

DC Shift For any constant c,

$$\text{Var}[X + c] = \text{Var}[X]$$



Probability & Statistics for DS & AI

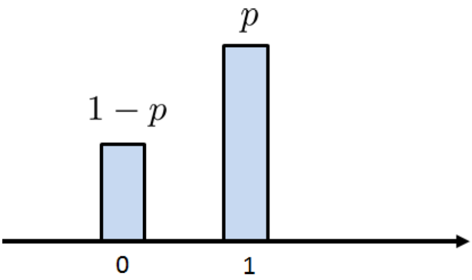
Common models used to describe Discrete Random Variables

Michele Guindani

Summer

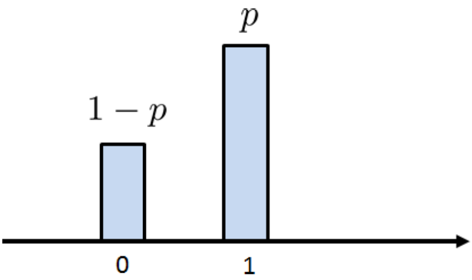
Bernoulli model (Bernoulli Random Variables)

- A Bernoulli random variable is a coin-flip-type of random variable.
- The random variable has two states: either 1 or 0 .
- The probability of getting 1 is p, and the probability of getting 0 is $1-p$.



Bernoulli model (Bernoulli Random Variables)

- A Bernoulli random variable is a coin-flip-type of random variable.
- The random variable has two states: either 1 or 0 .
- The probability of getting 1 is p, and the probability of getting 0 is $1 - p$.



💪 Bernoulli random variables are useful for all kinds of binary state events: coin flip (H or T), binary bit (1 or 0), true or false, yes or no, present or absent, Democrat or Republican, choices, friendships/links in social networks, etc.

Bernoulli r.v. - definition

Let X be a Bernoulli random variable. Then, the PMF of X is

$$p_X(0) = 1 - p, \quad p_X(1) = p$$

where $0 < p < 1$ is called the Bernoulli parameter. We write

$$X \sim \text{ Bernoulli } (p)$$

to say that X is drawn from a Bernoulli distribution with a parameter p. ,
The parameter p controls the probability of obtaining 1.

Example (Coin flip)

In a coin flipping event, p is usually $\frac{1}{2}$, meaning that the coin is fair. However, for biased coins p is not necessarily $\frac{1}{2}$.

!! If $X \sim \text{Bernoulli}(p)$, then

$$\mathbb{E}[X] = p, \qquad \text{Var}[X] = p(1-p)$$

!! If $X \sim \text{Bernoulli}(p)$, then

$$\mathbb{E}[X] = p, \qquad \text{Var}[X] = p(1-p)$$

Example (Simulation of a coin flipping)

In a coin flipping case with $p = 1/2$, then $E(X) = 1/2$ and $Var(X) = 1/4$

Let's verify the results above by doing an approximation via simulation 💻.

We generate 1,000 Bernoulli random variables and empirically verify the values.

Approximation via simulation

```
import numpy as np
np.random.seed(12345)
import matplotlib.pyplot as plt
p = 0.5
n = 1 #note the n=1
X = np.random.binomial(n,p,size=1000)
# average of ones approximates EX
p_hat=np.mean(X); print(p_hat)
# approximation of the variance
v_hat=p_hat*(1-p_hat); print(v_hat)
plt.hist(X,bins='auto')
```

Running the code above we get $\hat{p} = \frac{1}{1{,}000} \sum_{1}^{1{,}000} x_i = 0.506$ and $\hat{V}(X) = \hat{p} \times (1 - \hat{p}) = 0.249964$.

The "hat" 🎩 means we are considering an approximation (or an "estimate" based on the data that we have generated)

Probability & Statistics for DS & AI

Application
Modeling social networks via Bernoulli r.v.'s

Michele Guindani

Summer

Erdös-Rényi graph

- The study of networks is a big branch of modern data science.
- The history of network science is very long, but a famous (and simple!) model of a network is the Erdös-Rényi graph.
- The underlying probabilistic model of the Erdös-Rényi graph is the Bernoulli random variable.
- Very used nowadays to model social networks (Facebook, Google, etc), computer networks, traffic networks, financial contagion, etc.

What's a graph?

- A graph contains two elements: nodes and edges.
- For node i and node j, we denote the edge connecting i and j as A_{ij}.
- Therefore, if we have N nodes, then we can construct a matrix $\boldsymbol{A}$ of size $N \times N$ (adjacency matrix)

Example (Binary Graph)

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Edges for node pairs $(1,2),(1,3)$ and $(3,4)$.

Erdos-Renyi graph model

The probability of getting an edge is an independent Bernoulli random variable,

$$A_{ij} \sim \text{Bernoulli}(p)$$

for $i < j$. e parameter p controls the density of the graph. High values of p mean that there is a higher chance for an edge to be present. ($\uparrow p$, more dense graph)

Erdos-Renyi graph model

The probability of getting an edge is an independent Bernoulli random variable,

$$A_{ij} \sim \text{Bernoulli}(p)$$

for $i < j$. e parameter p controls the density of the graph. High values of p mean that there is a higher chance for an edge to be present. ($\uparrow p$, more dense graph)

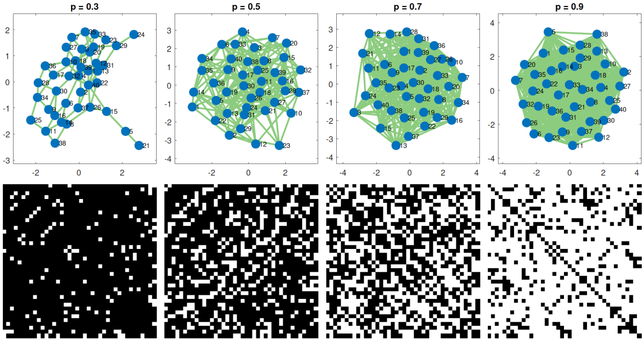
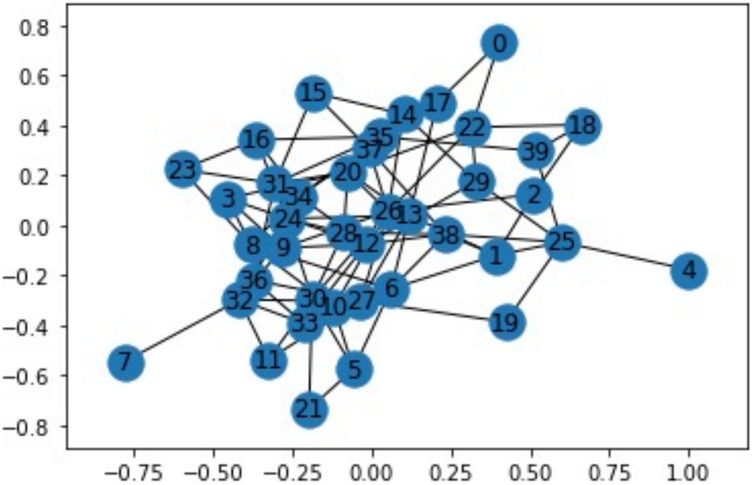


Figure 3.25: The Erdős-Rényi graph. [Top] The graphs. [Bottom] The adjacency matrices.

```python
# Python code to generate Erdos Renyi Graph
import numpy as np
import matplotlib.pyplot as plt
import networkx as nx
np.random.seed(1234)
A = np.random.rand(40, 40) < 0.1 #probability of a connection 0.1
A = np.triu(A, 1)
A = A + A.T
G = nx.convert_matrix.from_numpy_matrix(A)
fig, ax = plt.subplots()
nx.draw(G, with_labels=True, ax=ax)
limits = plt.axis("on")
ax.tick_params(left=True, bottom=True, labelleft=True, labelbottom=True)
plt.show()
```


0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

Stochastic Block Models

In Stochastic blockmodels (SBMs) nodes form small communities within a large network.

Example: Students within the same major in a University tend to have more interactions than with students of another major.

SBMs partitions nodes into communities.

Within each community, the nodes can have a high degree of connectivity.

Across different communities, the connectivity will be much lower.

Stochastic Block Models

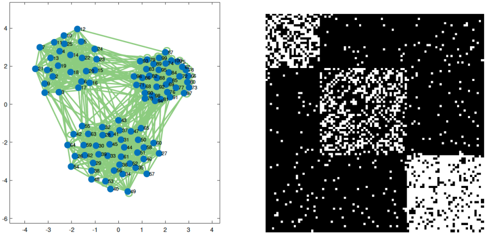
Each node belongs to one of K latent communities, where the probability of an edge depends on how strongly their communities are connected.

Let $c_i \in \{1, \ldots, K\}$ indicate the community of node i

☞ Edges are then sampled independently as

$$Y_{ij} \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\eta_{c_i c_j}\right)$$

where $\eta_{k\ell} = \eta_{\ell k}$ is the probability of an edge between a node in community k and a node in community ℓ.



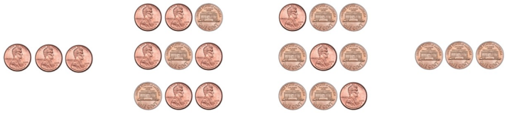
A stochastic block model containing three communities. [Left] The graph. [Right] The adjacency matrix.

Binomial model

Binomial model

Example (Coin Flipping)

- Suppose we flip the coin n times count the number of heads. Since each coin flip is a random variable (Bernoulli), the sum is also a random variable.
- Let's consider the probability of obtaining $\{H\}$ three times when flipping a coin 3 times



p^3 $\quad$ $p^2(1-p)$ $\quad$ $p(1-p)^2$ $\quad$ $(1-p)^3$

The probability of getting k heads out of $n = 3$ coins.

Here are the detailed calculations. Let us start with $X = 3$.

$$\begin{aligned} p_X(3) &= \mathbb{P}[\{\text{HHH}\}] \\ &= \mathbb{P}[\{\text{H}\} \cap \{\text{H}\} \cap \{\text{H}\}] \\ &\overset{(a)}{=} \mathbb{P}[\{\text{H}\}]\mathbb{P}[\{\text{H}\}]\mathbb{P}[\{\text{H}\}] \\ &\overset{(b)}{=} p^3, \end{aligned}$$

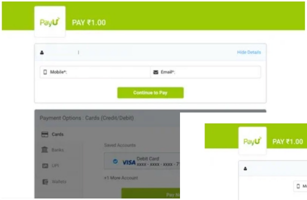
Example (A/B Testing)

PayU is an Indian-origin fin-tech payment facilitator. PayU finds it important to maintain a simple, intuitive, and convenient checkout process and eliminate all plausible elements which may cause drop-offs.

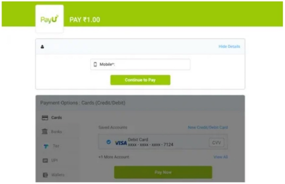
But their Checkout page statistics spoke an altogether different story. The company found that a lot of people were dropping off from the page, which was significantly impacting their sales and revenue graph.

To streamline everything, PayU decided to make minor changes on its Checkout page form basis the data gathered and run an A/B test to validate their hypothesis.

While the old PayU Checkout page asked users to enter their mobile number and email address to complete the purchase process, the new Checkout page, on the other hand, only asked for a user's mobile number.



Control



Variation

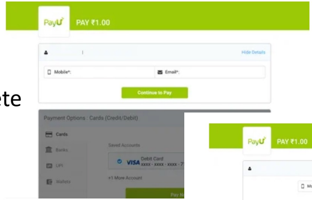
Example (A/B Testing)

PayU is an Indian-origin fin-tech payment facilitator. PayU finds it important to maintain a simple, intuitive, and convenient checkout process and eliminate all plausible elements which may cause drop-offs.

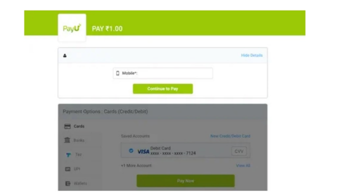
Show old page A to N_A customers
Show new page B to N_B customers

Let X_i denote how many customers complete check out out of N_i with site i=A,B

Will the new site lead to more customers finishing the check out process than the old site?



Control



Variation

Binomial random variable

Definition

Let X be a Binomial random variable. Then, the PMF of X is

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

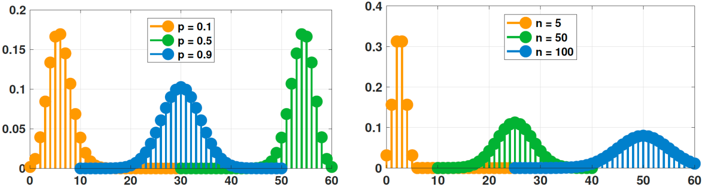
where $0 < p < 1$ is the Binomial parameter, and n is the total number of states. We write

$$X \sim \text{Binomial}(n, p)$$

Origin of Binomial Random Variables

- Flip a coin n times. Find the probability of obtaining k heads.

$$p_X(k) = \underbrace{\binom{n}{k}}_{\text{number of combinations}} \underbrace{p^k}_{\text{prob obtain k H's}} \underbrace{(1-p)^{n-k}}_{\text{prob obtain } n-k \text{ T's}}$$



(a) $n = 60$ (b) $p = 0.5$

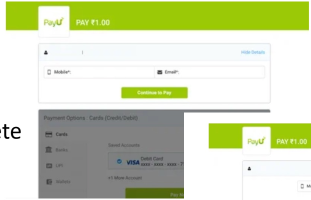
PMFs of a binomial random variable $X \sim \text{Binomial}(n, p)$.

Example (A/B Testing)

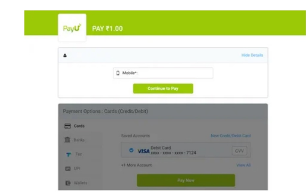
PayU is an Indian-origin fin-tech payment facilitator. PayU finds it important to maintain a simple, intuitive, and convenient checkout process and eliminate all plausible elements which may cause drop-offs.

Show old page A to N_A customers
Show new page B to N_B customers

Let X_i denote how many customers complete check out out of N_i with site i=A,B



Control



Variation

$$X_A \sim \text{Binomial}(N_A, p_A)$$

$$X_B \sim \text{Binomial}(N_B, p_B)$$

Example (Basketball)

LeBron James' three-point percentage scores (2018-2019 season)

Suppose Y_i are the number of 3-pointers LeBron James has scored in game i ($2018-19$ season).

	game	ThreeP	ThreePA
1	1	0	4
2	2	1	7
3	3	2	8
4	4	1	5
5	5	1	5
6	6	2	6
7	7	3	6
8	8	2	3
9	9	1	5
10	10	2	3
11	11	3	9
12	12	3	6
13	13	3	9
14	14	5	6
15	15	3	6
16	16	6	8
17	17	2	6
18	18	2	5
19	19	2	7
20	20	0	4

We can assume that each game

$$Y_i \sim \text{Bin}\left(n_i, \theta\right)$$

where
n_i are the number of 3-pointers attempted in game i and
θ is the probability of making a 3-pointer.

Example (Clinical Trials)

The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy

Senn S. (2021), *Pharmaceutical Statistics*,
https://onlinelibrary.wiley.com/doi/10.1002/pst.2226

TABLE 1 Numbers of subjects and cases for five large trials.

	Subjects		**Cases**	
Sponsor	**Vaccine**	**Placebo**	**Vaccine**	**Placebo**
Pfizer/BioNTech[9]	20,712	21,096	77	850
AZ/Oxford[10]	17,662	8550	73	130
Moderna[11]	14,134	14,073	11	185
Novavax[12]	7020	7019	10	96
J&J Janssen[13]	19,630	19,691	116	348

Conditionally on the total number of cases, the number of cases under vaccine

$$Y_V \sim \text{Bin}(\theta, n)$$

with θ equal to the probability of cases under vaccine.
Similarly for the controls.

Moments of Binomial

- If $X \sim \text{Binomial}(n, p)$, then

$$\mathbb{E}[X] = np$$
$$\text{Var}[X] = np(1-p)$$

Python code to compute the mean and var of a binomial r.v.

```
import scipy.stats as stats
p = 0.5
n = 10
rv = stats.binom(n,p)
M, V = rv.stats(moments='mv')
print(M, V)
#5.0 2.5
```

Geometric random variable

Geometric random variable

Example (Flipping a Coin)

Suppose you flip a coin until you see a head. This requires you to have $N-1$ tails, and then followed by a head in the N-th coin flip. The probability of this sequence of events are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\ldots$, which forms an infinite series.

Geometric random variable

Example

- IT: The probability that a a certain company experiences a network failure in a given week is 10%. What's the probability that the company can go 5 weeks or longer without experiencing a network failure?
- Sports: The probability that a batter is able to make a successful hit before three strikes
- Marketing: model how many attempts to make a sale will end in a success

Geometric Random Variable

Definition

Let X be a Geometric random variable. Then, the PMF of X is

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

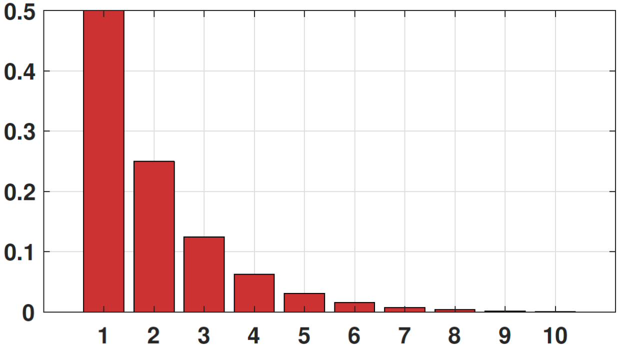
where $0 < p < 1$ is the Geometric parameter. We write

$$X \sim \text{Geometric}(p)$$

- Interpretation: Flip a coin until you get a head.

$$p_X(k) = \underbrace{(1-p)^{k-1}}_{k-1 \text{ failures}} \quad \underbrace{p}_{\text{final success}}$$

Probability mass function



The histogram of flipping a coin until we see a Head. The x-axis denotes the number of coins we need to flip, and the y-axis denotes the probability.

Probability of success

- If $p = \frac{1}{2}$, then

$$\begin{aligned}
\mathbb{P}[\text{ success after 1 attempt }] &= \frac{1}{2} = 0.5 \\
\mathbb{P}[\text{ success after 2 attempts }] &= \frac{1}{2} + \frac{1}{4} = 0.75 \\
\mathbb{P}[\text{ success after 3 attempts }] &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = 0.875 \\
\mathbb{P}[\text{ success after 4 attempts }] &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = 0.9375
\end{aligned}$$

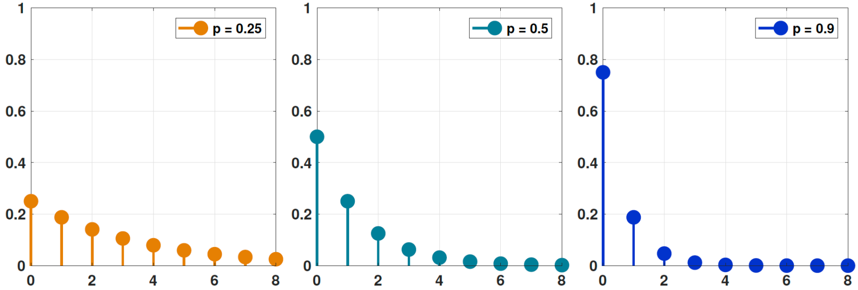


Table: PMFs of a geometric random variable $X \sim \text{Geometric}(p)$.

If $X \sim \text{Geometric}(p)$, then

$$\mathbb{E}[X] = \frac{1}{p} \quad \text{and} \quad \text{Var}[X] = \frac{1-p}{p^2}$$

Probability & Statistics for DS & AI

Poisson Model

Michele Guindani

Summer

A ridepooling platform requires to dispatch for-hire vehicles to serve real-time ride-pooling demand on a general road network

RESEARCH ARTICLE

WILEY

Dynamic pricing and matching in ride-hailing platforms

Chiwei Yan | Helin Zhu | Nikita Korolko | Dawn Woodard

Uber Technologies, San Francisco, California,

Correspondence
Chiwei Yan, Uber Technologies, San Francisco, CA.
Email: chiwei@mit.edu

Abstract

Ride-hailing platforms such as Uber, Lyft, and DiDi have achieved explosive growth and reshaped urban transportation. The theory and technologies behind these platforms have become one of the most active research topics in the fields of economics, operations research, computer science, and transportation engineering. In particu- dynamic pricing (DP) algorithms—the two key levers in d tremendous attention from the research community and gned and implemented at industrial scales by ride-hailing view of matching and DP techniques in ride-hailing, and or providing an experience with low waiting time for both link the two levers together by studying a pool-matching c *waiting* (DW) that varies rider waiting and walking spired by a recent carpooling product Express Pool from from Uber that by jointly optimizing DP and DW, price d, while increasing capacity utilization, trip throughput, light several key practical challenges and directions of titioner's perspective.

Two key technologies employed by ride-hailing platforms to provide high service reliability and capacity utilization rate and low rider waiting time are *matching* and *dynamic pricing* (DP). Matching means the process of dispatching available drivers to pick up riders, and DP means dynamic adjustment of prices for rides based on real-time demand and supply conditions. DP is called "surge pricing" by Uber and "prime time" by Lyft.

DiDi

OLA

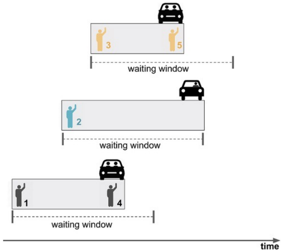
Uber

lyft

Grab

gojek

A ridepooling platform requires to dispatch for-hire vehicles to serve real-time ride-pooling demand on a general road network



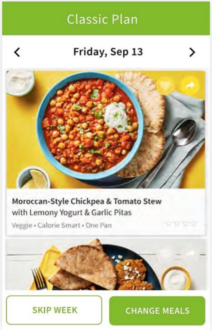
What's the probability that more than 10 cars are needed at the P4 parking lot of the Mumbai Airport at 2:10am?

HelloFresh is a meal kit subscription service that delivers recipes to global users. As its user base grew, its recipe count grew, but it became more difficult for users to navigate through the app and find what they needed.

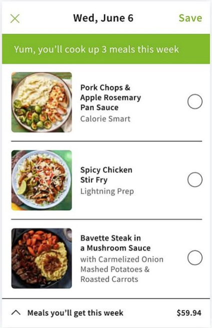
The business set out to redesign its menu pages for a seamless user experience while also drawing attention to upselling opportunities.

HelloFresh ran an experiment that compared the impact of the original control menu display to a new version

A



B



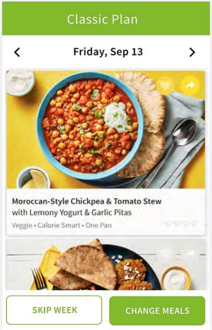
HelloFresh is a meal kit subscription service that delivers recipes to global users. As its user base grew, its recipe count grew, but it became more difficult for users to navigate through the app and find what they needed.

Show menu A to N_A people
Show menu B to N_B people

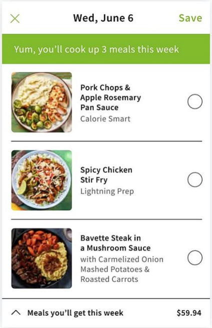
Let X_A denote the number of items chosen by people using the A menu

Let X_B denote the number of items chosen by people using the B menu

A



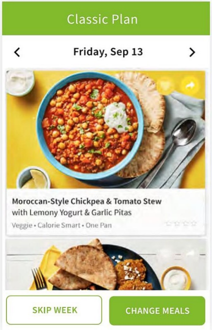
B



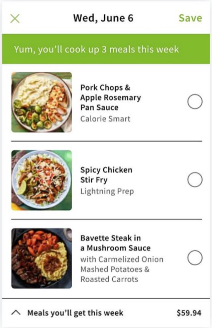
HelloFresh is a meal kit subscription service that delivers recipes to global users. As its user base grew, its recipe count grew, but it became more difficult for users to navigate through the app and find what they needed.

What's the Probability that a person will choose more than 2 items under the two menus?

A



B



Poisson random variable

Let X be a Poisson random variable. Then, the PMF of X is

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 1, 2, \ldots$$

where $\lambda > 0$ is the Poisson rate. We write

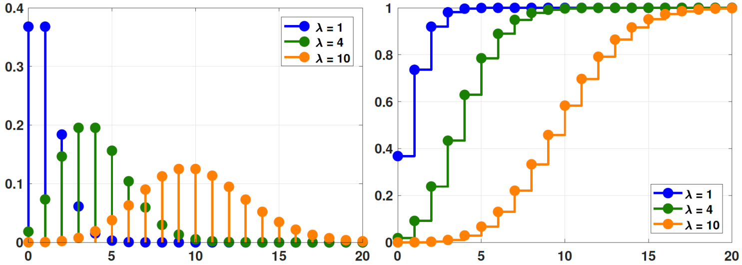
$$X \sim \text{Poisson}(\lambda)$$

- X = number of arrivals
- α = arrival rate = number per unit time
- t = time
- So, $\lambda = \alpha t$ = average number within t unit time.

Shape of the Poisson PMF and CDF

The CDF of Poisson is

$$F_X(k) = \mathbb{P}[X \leq k] = \sum_{\ell=0}^{k} \frac{\lambda^\ell}{\ell!} e^{-\lambda}$$



Example (Cars needed at Mumbai Airport)

Suppose that the average number of cars over a certain time period (say, 2-2:30am) at a given platform is 6.
Let X be the actual number of cars one may need at that time. Then, find $\mathbb{P}[X > 10]$ and $\mathbb{P}[X \leq 8]$

Example (Cars needed at Mumbai Airport)

Suppose that the average number of cars over a certain time period (say, 2-2:30am) at a given platform is 6.
Let X be the actual number of cars one may need at that time. Then, find $\mathbb{P}[X > 10]$ and $\mathbb{P}[X \leq 8]$
Solution. By using a Poisson random variable with parameter λ, we can show that the probabilities are

$$\mathbb{P}[X > 10] = 1 - \mathbb{P}[X \leq 10] = 1 - \sum_{k=0}^{10} \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\mathbb{P}[X \leq 8] = \sum_{k=0}^{8} \frac{\lambda^k}{k!} e^{-\lambda}$$

using $\lambda = 6$.

First, let's plot the pdf and see what events we are interested in:

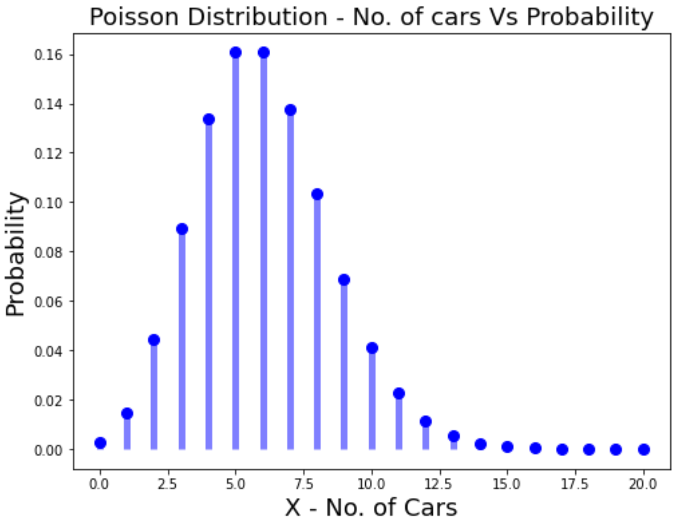
```
from scipy.stats import poisson
import matplotlib.pyplot as plt

# number of cars for the plot (support)

X = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
16, 17, 18, 19, 20]
lmbda = 6

# Compute pdf values
poisson_pd = poisson.pmf(X, lmbda)

# Plot the pdf
fig, ax = plt.subplots(1, 1, figsize=(8, 6))
ax.plot(X, poisson_pd, 'bo', ms=8, label='poisson pmf')
plt.ylabel("Probability", fontsize="18")
plt.xlabel("X - No. of Cars", fontsize="18")
plt.title("Poisson Distribution - No. of cars Vs Probability",
fontsize="18")
ax.vlines(X, 0, poisson_pd, colors='b', lw=5, alpha=0.5)
```


Poisson Distribution - No. of cars Vs Probability
Probability
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
X - No. of Cars

```
from scipy.stats import poisson

#probability of  less or equal than 8 cars
poisson.cdf(k=8, mu=6)
0.847

#probability greater than 10
1-poisson.cdf(k=10, mu=6)
0.043
```

If $X \sim \text{Poisson}(\lambda)$, then

$$\mathbb{E}[X] = \lambda, \quad \text{Var}[X] = \lambda$$

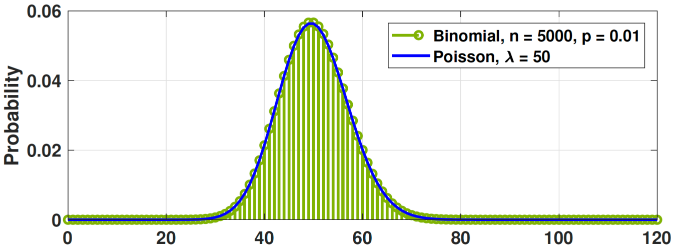
Poisson approximation to Binomial

- When N is large, binomial is approximately Poisson

Poisson Approximation to Binomial. *For small p and large n,*

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda},$$

where $\lambda \stackrel{def}{=} np$.



Example

- Data arrival rate: $n = 10^9$ bits per second.
- Probability of having one error bit: $p = 10^{-9}$.
- In one second, how likely will we get $k = 5$ error bits?

Example

- Data arrival rate: $n = 10^9$ bits per second.
- Probability of having one error bit: $p = 10^{-9}$.
- In one second, how likely will we get $k = 5$ error bits?
- **Solution**

① If you stick to binomial:

- Binomial: Flip coin 10^9 times. Get 5 heads:

$$\binom{10^9}{5}\left(10^{-9}\right)^5\left(1-10^{-9}\right)^{10^9-5}$$

- Numerical Problem!!

Example

- Data arrival rate: $n = 10^9$ bits per second.
- Probability of having one error bit: $p = 10^{-9}$.
- In one second, how likely will we get $k = 5$ error bits?
- **Solution**

① If you stick to binomial:

- Binomial: Flip coin 10^9 times. Get 5 heads:

$$\binom{10^9}{5}\left(10^{-9}\right)^5\left(1-10^{-9}\right)^{10^9-5}$$

- Numerical Problem!!

② If we use Poisson approximation:

- $\lambda = np = \left(10^9\right)\left(10^{-9}\right) = 1$
- $\frac{1^5}{5!}e^{-}$
- Much easier!!