Probability and Statistics for Data Science

Data Science Course

Dr. Ariel Mantzura

2025-01-30

Topics to be Covered in this Lecture

- Probability space
- Probability basics
- Random Variables
- Discrete Random variables
- Expectation (discrete)
- Variance (discrete)

Sources: **Practical statistics for Data Scientist** Peter Bruce, Andrew Bruce & Peter Gedeck **Probability and Statistics for Data Science** Carlos Fernandez-Granda

Probability Space

- A **probability space** or a probability triple:

$$(\Omega, \mathcal{F}, \mathcal{P})$$

  is a mathematical framework that provides a formal model of a random experiment.
- For example, one can define a probability space which models mathematically the random experiment of throwing a fair die.

Probability Space

- A Probability Space consists of 3 elements:
  - A **Sample Space** Ω which is the set of all possible outcomes of an experiment with ω denoting a one sample point.
  - An **Event Space**, which is a set of events $\mathcal{F}$ where an event is a set of outcomes in the sample space.
  - A **Probabiity function** $\mathcal{P}$ which assigns to each event in the event space, a probability which is a number in the set $[0, 1]$.

Probability space

- Sample spaces may be **discrete** or **continuous**.
- Examples of **discrete** sample spaces include the possible outcomes of a coin toss, the score of a basketball game, the number of people that show up at a party, etc.
- **Continuous** sample spaces are usually intervals of $\mathcal{R}$ or $\mathcal{R}^n$ used to model time, position, temperature, etc.
- If our sample space is discrete, a possible choice for $\mathcal{F}$ is the power set of the sample space denoted $2^{|\Omega|}$, which consists of all possible sets of elements in the sample space.

Probability space - example

- If we are tossing a fair coin then the sample space is:

$$\Omega = \{heads, tails\}$$

.

- The event space in this case can be defined as the set of all events:

$$\{\Omega = \text{heads or tails, tails, heads}, \phi\}$$

where ϕ is the empty set.

- $2^{|\Omega|} = 2^2 = 4$ because the sample space consists of 2 points: heads and tails.

Probability function - definition

- A probability function defined over the sets in $\mathcal{F}$ such that:
  1. $P(S) \geq 0$ for any event(set) $S \in \mathcal{F}$.
  2. If the sets $S_1, S_2, \ldots, S_n \in \mathcal{F}$ are disjoint (i.e. $S_i \cap S_j = \phi$ for $i \neq j$ where ϕ is the empty set.) then

$$P(\cup_{i=1}^{n} S_i) = \sum_{i=1}^{n} P(S_i)$$

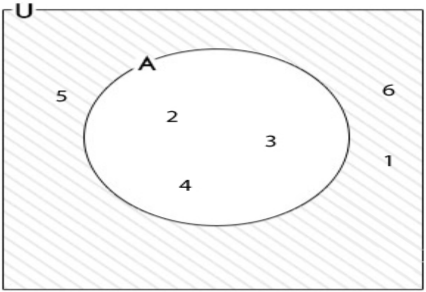
  3. $P(\Omega) = 1$.

Probability function - Exercise

1. Consider the tossing of 2 fair dice.
2. Define the sample space Ω.
3. What is $|\Omega|$?
4. How would you define $\mathcal{F}$ in this case?
5. Consider the events $A = \{\textbf{Odd number double}\}$ and $B = \{\textbf{Even number double}\}$. What is the probability of A and B?
6. IS $A \cap B = \phi$ i.e. are A and B disjoint events?
7. Calculate $P(A \cup B)$.

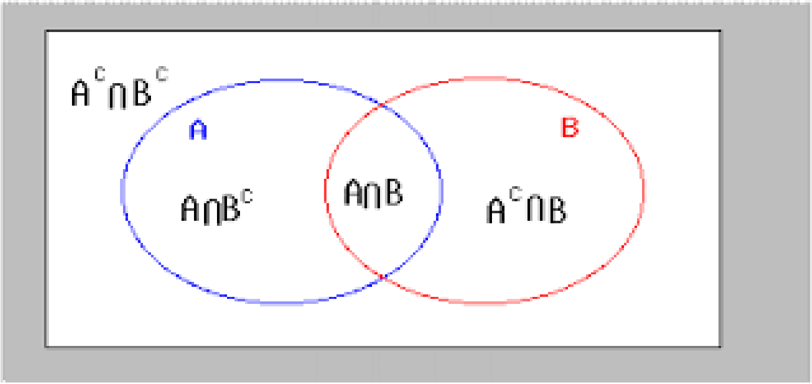
Probabilty function - definition and Ven diagram

- Define the complementary event of event S denoted S^C such that $S \cup S^C = \Omega$ and $S \cap S^C = \phi$
- In words: the complementary set includes all the sample points in Ω that are not in S, i.e $\Omega \backslash S$
- $B \backslash A$ means the elements in B minus the elements in A (included in B).
- It is easy to see that $P(S^C) = 1 - P(S)$.



$A^c = \{1, 5, 6\}$

Probability definition -Ven diagram



Probability definition

- The probability definition implies the following:
  1. $P(\phi) = 0$
  2. $P(S) \leq 1$ for any event(set) $S \in \mathcal{F}$.
  3. $A \subseteq B$ then $P(A) \leq P(B)$.
  4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

  - Exercise: prove 1-4.

Conditional Probability

- Conditional probability is an important concept in probability.
- It allows us to update probabilities of events when additional information is revealed.
- Consider a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where we find out that the outcome of the experiment belongs to a certain event $S \in F$.

Conditional Probability

- This affects how likely it is for any other event $S' \in \mathcal{F}$ to have occurred: we can rule out any outcome not belonging to S.
- The updated probability of each event is known as the conditional probability of S' given S.
- Intuitively, the conditional probability can be interpreted as the fraction of outcomes in S that are also in S'.

Conditional Probability

- The formula for the conditional probability of an event S is simply:
$$P(S'|S) = \frac{P(S \cap S')}{P(S)}$$
- This probability can be perceived as an update to $P(S)$ after finding out that event S has occured.
- Intuitively, the conditional probability can be interpreted as the fraction of outcomes in S that are also in S'.

Conditional Probability - simple example

- Consider a fair die, i.e each number has probability $\frac{1}{6}$.
- The probability to get an odd number in a toss is

$$P(\text{odd number}) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

- Let S be the event that the number is odd, i.e $\{1, 3, 5\}$
- What is the probability that we get the number 3 given that the number is odd?
- The previous probability is now updated as follows:

$$P(S'|S) = \frac{P(3 \cap \{1,3,5\})}{P(odd)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Intersection of Events

- Conditional probabilities can be used to compute the intersection of several events in a structured way.
- By definition, we can express the probability of the intersection of two events A and B as follows:

$$P(A \cap B) = P(A)P(B|A)$$

- In this formula $P(A)$ is known as the **prior probability** of A, as it captures the information we have about A before anything else is revealed.
- $P(B|A)$ is known as the **posterior probability**.
- These are basic terms in Bayesian models which will be discussed later in the course.

Intersection of Events - Example

- In a fair die toss, Let $A = \{\textbf{Odd number}\}$ and let $B = \{\textbf{number greater or equal than 3}\}$
- Calculate $P(A \cap B)$ using the previous formula.
- $P(A) = 0.5$ and $P(B|A) = P(\{3,5\}|\{1,3,5\}) = \frac{2}{3}$
- $$P(A \cap B) = 0.5 * \frac{2}{3} = \frac{1}{3}$$

Intersection of Events - The Chain Rule

- Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $S_1, S_2, \ldots, S_n$ a collection of events in F, then

$$P(\cap_{i=1}^{n} S_i) = P(S_1 \cap S_2 \cdots \cap S_n) =$$

$$= P(S_1)P(S_2|S_1)P(S_3|S_1 \cap S_2) \ldots P(S_n|S_1 \cap S_2 \cdots \cap S_{n-1}) =$$

$$= P(S_1) \prod_{i=2}^{n} P(S_i| \cap_{j=1}^{i-1} S_j)$$

Law of Total Probability

- $A_1, A_2, \ldots, A_n \in F$ is a partition of Ω if $\cup_{i=1}^{n} A_i = \Omega$ and $A_1, A_2, \ldots, A_n$ are disjoint.
- Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let the collection of disjoint sets $A_1, A_2, \ldots A_n \in F$ be any partition of Ω.
- For any set $S \in F$

$$P(S) = P(S|A_1)P(A_1) + P(S|A_2)P(A_2) \ldots P(S|A_n)P(A_n)$$

$$=$$

$$\sum_{i=1}^{n} P(S|A_i)P(A_i) =$$

$$= \sum_{i=1}^{n} P(S \cap A_i)$$

Law of Total probability - Example

- Imagine you have three bags of marbles:
- Bag 1 has 2 red and 3 blue marbles.
- Bag 2 has 1 red and 4 blue marbles.
- Bag 3 has 3 red and 2 blue marbles.
- You randomly pick one of the three bags and then draw a marble from it.
- We want to find the probability of drawing a red marble (event A).

Law of Total probability - Example

- Define B_1 the event of choosing bag 1.
- Define B_2 the event of choosing bag 2.
- Define B_3 the event of choosing bag 3.
- $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$
- $P(A|B_1) = \frac{2}{5}$
- $P(A|B_2) = \frac{1}{5}$
- $P(A|B_3) = \frac{3}{5}$
- Finally,

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) = \frac{2}{5}$$

Bayes law

- For any events A and B in a probability space $(\Omega, \mathcal{F}, \mathcal{P})$

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

- as long as $P(A) > 0$.
- Exercise - In a fair die throw let $B = \{2, 4, 6\}$ and $A = \{4\}$.
- Calculate $P(B\backslash A)$ using the Bayes equation.
- Prove the Bayes law.

Independence

- Events A and B are independent if and only if:

$$P(A|B) = P(A)$$

- Alternatively,

$$P(A \cap B) = P(A)P(B)$$

Independence - example

- Consider the following events:

$$A = \{1, 3, 5\}$$

and

$$B = \{2, 4, 6\}$$

.

- Are these 2 events disjoint? answer: YES
- Are these two events independent? answer: NO

$$P(A \cap B) = P(\phi) = 0 \neq P(A)P(B) = \frac{1}{2}\frac{1}{2}$$

Random Variables - definition

- A random variable X is a function that maps each outcome in a probability space to a real number.
- Mathematically,

$$X(\omega) : \Omega \to \mathcal{R}$$

Formal Definition:

- Given a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ a random variable X is a function from the sample space Ω to the real numbers $\mathcal{R}$.
- Once the outcome $\omega \in \Omega$ of the experiment is revealed, the corresponding $X(\omega)$ is known as a realization of the random variable.

Random Variables - fair coin example

- Consider a fair coin where the sample space is Ω={heads,tails}.
- Define the random variable X as follows:

$$X(\omega) = \begin{cases} 1 & \texttt{if} \ \ \omega = heads \\ 0 & \texttt{if} \ \ \omega = tails \end{cases}$$

Random Variables - fair die example

- Consider a fair die.
- The sample space is Ω={1,2,3,4,5,6}.
- Define the random variable X as the identitiy function as follows:

$$X(\omega) = \begin{cases} 1 & \texttt{if } \omega = 1 \\ 2 & \texttt{if } \omega = 2 \\ 3 & \texttt{if } \omega = 3 \\ 4 & \texttt{if } \omega = 4 \\ 5 & \texttt{if } \omega = 5 \\ 6 & \texttt{if } \omega = 6 \end{cases}$$

Random Variables - notation of events

- We often denote events of the form:

$$\{\omega \in \Omega : X(\omega) \in A\}$$

as:

$$\{X \in A\}$$

- For example the event S:

$$S = \{\omega : X(\omega) = 1\} = \{X = 1\} = \{heads\}$$

in the fair coin toss.
- This shows that each event $S \in \mathcal{F}$ is always a set of sample points.

Random Variables

- A random variable is defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$.
- $$P(X \in A) = P(\{\omega | X(\omega) \in A\})$$

The Probability Mass function of a discrete random variable - Definition

- Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $X : \Omega \to \mathcal{Z}$ a random variable.
- The probability Mass Function of X is defined as:

$$p_X(x) = P(\{\omega : X(\omega) = x\})$$

- X has a discrete range of values denoted by D which satisfies

$$P(\{\omega : X(\omega) = x\}) \geq 0$$

for all x and:

$$\sum_{x \in D} P(\{\omega : X(\omega) = x\}) = 1$$

The Probability Mass function (pmf) of a discrete random variable

- To compute the probability that a random variable X is in a certain set A we take the sum of the pmf over all the values contained in A:

$$P(X \in A) = \sum_{x \in A} p_X(x)$$

- An example is the fair die.
- X can take values in a discrete set A where $A = \{1, 2, 5\}$
- Let A be the following event: $A = \{1, 2, 5\}$.

$$P(X \in A) = P(1) + P(2) + P(5) = \frac{3}{6}$$

Discrete Random Variables - Bernouli

- Bernoulli random variables are used to model experiments that have two possible outcomes.
- We usually represent an outcome by 0 and the other outcome by 1.
- An example is flipping a biased coin where the probability of landing on heads is p.
- The pmf of a Bernoulli random variable X with parameter $p \in [0, 1]$ is given by:

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

Discrete Random Variables - Indicator function

- Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space.
- The indicator random variable of an event $S \in \mathcal{F}$ is defined as

$$\mathcal{I}_S(\omega) = \begin{cases} 1 \text{ if } \ \omega \in S \\ 0 \ \ otherwise \end{cases}$$

- By definition the distribution of an indicator random variable is Bernoulli with parameter $P(S)$.
- This is simply because

$$P(\mathcal{I}_S(\omega) = 1) = P(\omega \in S) = P(S)$$

$$P(\mathcal{I}_S(\omega) = 0) = 1 - P(S)$$

Discrete Random Variables - Binomial

- Binomial random variables are used to model the number of successes (arbitrary) of n trials modeled as independent Bernoulli random variables with the same parameter.
- The Binomial random variable sums the number of $1's$ in n independent Bernoulli random variables with the same parameter .
- The pmf of the Binomial Random variable is:

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

- where p is the probability to get 1 in one Bernoulli trial.

Discrete Random Variables - Binomial example

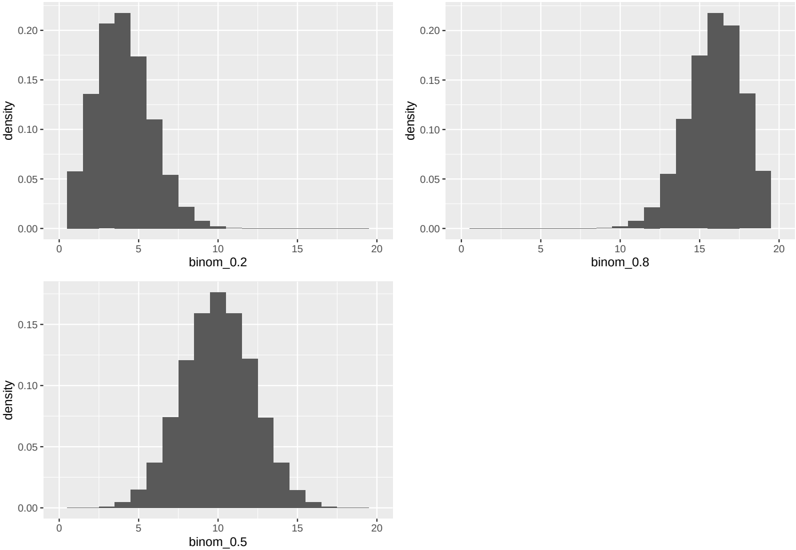
- Consider the following experiment: toss a biased coin 5 times where the probability to fall on heads is 0.2.
- What is probability to get 3 times heads?
- p=P(heads)=0.2

$$P(X = 3) = \binom{5}{3} 0.2^3 0.8^2$$

```
from scipy.stats import binom
result = binom.pmf(k=3, n=5, p=0.2)
print("Probability:", round(result,4))
```

```
## Probability: 0.0512
```

Discrete Random Variables - Binomial



Discrete Random Variables - Multinomial

- A multinomial random variable counts the number of occurrences of each possible outcome in a sequence of n trials, where each trial can result in one of k possible outcomes.
- Consider an experiment that consists of n independent trials, where each trial can result in one of k different outcomes.
- The probability of the i_{th} outcome occurring on any given trial is P_i where $\sum_{i=1}^{k} p_i = 1$
- X_i represents the number of times the i_{th} outcome occurs in n trials.
- $X = \{X_1, X_2, \dots X_k\}$ follows a multinomial distribution with parameters n and $p_1, p_2, \dots, p_k$

Discrete Random Variables - Multinomial

- The pmf of the Binomial Random variable is:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \binom{n}{x_1 x_2 x_3 \ldots x_k} \Pi_{i=1}^{k} p_i^{x_i}$$

- Where

$$\binom{n}{x_1 x_2 x_3 \ldots x_k} = \frac{n!}{x_1! x_2! x_3! \ldots, x_k!}$$

Multinomial - example

- Suppose we roll a six-sided die 10 times. Each roll has 6 possible outcomes (1 through 6), each with a probability $\frac{1}{6}$.
- X_i is the number of times the face i appears in the 10 rolls.
- In this example we have a multinomial with $n = 10$ and $p_1, p_2, \ldots, p_n = \{\frac{1}{6}, \frac{1}{6}, \ldots, \frac{1}{6}\}$

Multinomial - example

```
import numpy as np
from scipy.stats import multinomial
a = np.random.multinomial(10, [1/6] * 6)
print(a)
```

```
## [0 2 2 2 2 2]
```

```
prob=multinomial.pmf([2,2,1,4,1,0],n=10,p=[1/6]*6)
print(prob)
```

```
## 0.0006251428898033849
```

Discrete Random Variables - Geometric

- A Geometric Random Variable counts the number of Bernoulli trials till we get a success (arbitrary).
- In the coin example: The number of throws till we get heads.
- The "succesful" throw is included in the number of throws.
- The pmf of a Geometric random variable is:

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, \ldots$$

Discrete Random Variables - Geometric example

- Lets define a “success” as getting the number 6 in a fair die toss.
- The number of tosses till the first time we get 6 is a Geometric Random Variable with $p = \frac{1}{6}$:
- We may ask: what is the probability of getting 6 after 4 tosses?

$$P(X = k) = (1 - \frac{1}{6})^3)\frac{1}{6} = 0.08037551$$

```
from scipy.stats import geom
result = geom.pmf(5,1/6)
print("Probability:", result)
```

```
## Probability: 0.0803755144032921
```

Discrete Random Variables - Poisson

- The Poisson random variable counts the number of event occurrences in 1 unit of time.
- For instance, imagine a call center which receives, on average 3 calls per minute.
- The Poisson random variable counts the number of calls in 1 time unit.
- The pmf of the Poisson random variable is

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- λ is a parameter of the pmf which is the expected (see below) number of occurrences per unit of time.

Discrete Random Variables - Poisson example

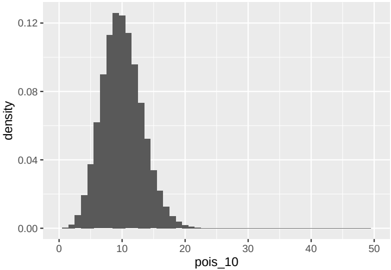
- In the call center example assume that $\lambda = 3$
- We may ask: what is the probability that

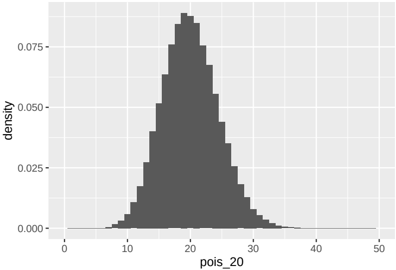
$$P(x = 5) = \frac{3^5 e^{-3}}{5!} = 0.1008188$$

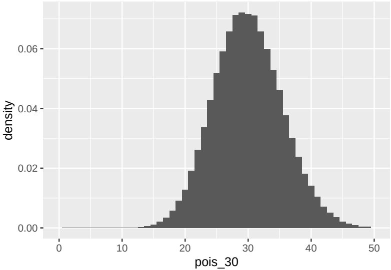
```
from scipy.stats import poisson
result = poisson.pmf(k=5, mu=3)
print("Probability:", result)
```

```
## Probability: 0.10081881344492458
```

Discrete Random Variables - Poisson







Expectation or mean of discrete random variable

- The expectation or (theoretical) mean of a discrete random variable X is defined as:

$$E(X) = \sum_{x \in \mathcal{Z}} x P(X = x)$$

- The expectation is the weighted average of all possible incomes by their probability.
- The expected value is a parameter and is not random.
- The expected value does not depend on a certain sample like the average but alternatively depends on the theoretical distribution.

Expectation or mean of discrete random variables

- Expectation of Bernouli random variable:

$$E(X) = p$$

- Expectation of Binomial random variable:

$$E(X) = np$$

- Expectation of Geometric random variable:

$$E(X) = \frac{1}{p}$$

- Expectation Poisson Random variable:

$$E(X) = \lambda$$

Expectation or mean of discrete random variable - Linearity

- For any constant a and b:

$$E(aX + b) = aE(X) + b$$

.

- For any 2 random variables:

$$E(X + Y) = E(X) + E(Y)$$

and hence:

$$E\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} E(X_i)$$

Variance and standard deviation

- The variance of X is the mean square deviation from the mean:
$$Var(X) = E\left[(X - EX)^2\right] =$$
$$= E(X^2) - (EX)^2$$
- The standard deviation of X is simply the square root of the variance:
$$\sigma_X = \sqrt{Var(X)}$$

Variance of linear function

- For any constants a and b:

$$Var(aX + b) = a^2 Var(X)$$

- Exercise - prove the equation

Variance of discrete random variables

- Variance of Bernoulli random variable:

$$V(X) = p(1-p)$$

- Variance of Binomial random variable:

$$V(X) = np(1-p)$$

- Variance of Geometric random variable:

$$V(X) = \frac{1}{p} = \frac{1-p}{p^2}$$

- Variance of Poisson random variable:

$$V(X) = \lambda$$

Exercise

- Prove that the expected value of a Bernoulli random variable is p.
- Let X be a random variable that counts the number of people entering 7- eleven on a given day. Assume that the expexted number of people entering on any given day is 20. Calculate the probability that 12 people enter on a given day.
- Let X denote a random variable that denotes the number of times that a person will win in 10 games of a slot machine. Assume that in each game there is a probability of 0.2 to win. Calculate the probability that a person will not win even one game.
- Assume the same mechanism as before. Calculate the probabilty that some person will win the first time after 8 games.