Mathematics for Artificial Intelligence

Introduction to Probability

Andres Mendez-Vazquez

April 9, 2020

Outline

Cinvestav

Outline

Cinvestav

Gerolamo Cardano: Gambling out of Darkness

Gambling

Gambling shows our interest in quantifying the ideas of probability for millennia, but exact mathematical descriptions arose much later.

Gerolamo Cardano (16th century)

While gambling he developed the following rule!!!

Equal conditions

"The most fundamental principle of all in gambling is simply equal conditions, e.g. of opponents, of bystanders, of money, of situation, of the dice box and of the dice itself. To the extent to which you depart from that equity, if it is in your opponent's favour, you are a fool, and if in your own, you are unjust."

Cinvestav

Gerolamo Cardano: Gambling out of Darkness

Gambling

Gambling shows our interest in quantifying the ideas of probability for millennia, but exact mathematical descriptions arose much later.

Gerolamo Cardano (16th century)

While gambling he developed the following rule!!!

Equal conditions

"The most fundamental principle of all in gambling is simply equal conditions, e.g. of opponents, of bystanders, of money, of situation, of the dice box and of the dice itself. To the extent to which you depart from that equity, if it is in your opponent's favour, you are a fool, and if in your own, you are unjust."

Gerolamo Cardano: Gambling out of Darkness

Gambling

Gambling shows our interest in quantifying the ideas of probability for millennia, but exact mathematical descriptions arose much later.

Gerolamo Cardano (16th century)

While gambling he developed the following rule!!!

Equal conditions

"The most fundamental principle of all in gambling is simply equal conditions, e.g. of opponents, of bystanders, of money, of situation, of the dice box and of the dice itself. To the extent to which you depart from that equity, if it is in your opponent's favour, you are a fool, and if in your own, you are unjust."

Gerolamo Cardano's Definition

Probability

"If therefore, someone should say, I want an ace, a deuce, or a trey, you know that there are 27 favorable throws, and since the circuit is 36, the rest of the throws in which these points will not turn up will be 9; the odds will therefore be 3 to 1."

Meaning

Probability as a ratio of favorable to all possible outcomes!!! As long all events are equiprobable...

Thus, we get

$$P(\text{All favourable throws}) = \frac{\text{Number All favourable throws}}{\text{Number of All throws}} \tag{1}$$

Gerolamo Cardano's Definition

Probability

"If therefore, someone should say, I want an ace, a deuce, or a trey, you know that there are 27 favorable throws, and since the circuit is 36, the rest of the throws in which these points will not turn up will be 9; the odds will therefore be 3 to 1."

Meaning

Probability as a ratio of favorable to all possible outcomes!!! As long all events are equiprobable...

Thus, we get

$$P(\text{All favourable throws}) = \frac{\text{Number All favourable throws}}{\text{Number of All throws}} \quad (1)$$

Gerolamo Cardano's Definition

Probability

"If therefore, someone should say, I want an ace, a deuce, or a trey, you know that there are 27 favorable throws, and since the circuit is 36, the rest of the throws in which these points will not turn up will be 9; the odds will therefore be 3 to 1."

Meaning

Probability as a ratio of favorable to all possible outcomes!!! As long all events are equiprobable...

Thus, we get

$$P(\text{All favourable throws}) = \frac{\text{Number All favourable throws}}{\text{Number of All throws}} \tag{1}$$

Cinvestav

Intuitive Formulation

Empiric Definition

Intuitively, the probability of an event A could be defined as:

$$P(A) = \lim_{n \to \infty} \frac{N(A)}{n}$$

Where $N(A)$ is the number that event a happens in n trials.

Intuitive Formulation

Empiric Definition

Intuitively, the probability of an event A could be defined as:

$$P(A) = \lim_{n \to \infty} \frac{N(A)}{n}$$

Where $N(A)$ is the number that event a happens in n trials.

Example

Imagine you have three dices, then

- The total number of outcomes is 6^3
- If we have event A = all numbers are equal, $|A| = 6$
- Then, we have that $P(A) = \frac{6}{6^3} = \frac{1}{36}$

Intuitive Formulation

Empiric Definition

Intuitively, the probability of an event A could be defined as:

$$P(A) = \lim_{n \to \infty} \frac{N(A)}{n}$$

Where $N(A)$ is the number that event a happens in n trials.

Example

Imagine you have three dices, then

- The total number of outcomes is 6^3
- If we have event A = all numbers are equal, $|A| = 6$
- Then, we have that $P(A) = \frac{6}{6^3} = \frac{1}{36}$

Intuitive Formulation

Empiric Definition

Intuitively, the probability of an event A could be defined as:

$$P(A) = \lim_{n\to\infty} \frac{N(A)}{n}$$

Where $N(A)$ is the number that event a happens in n trials.

Example

Imagine you have three dices, then

- The total number of outcomes is 6^3
- If we have event $A =$ all numbers are equal, $|A| = 6$
- Then, we have that $P(A) = \frac{6}{6^3} = \frac{1}{36}$

Intuitive Formulation

Empiric Definition

Intuitively, the probability of an event A could be defined as:

$$P(A) = \lim_{n \to \infty} \frac{N(A)}{n}$$

Where $N(A)$ is the number that event a happens in n trials.

Example

Imagine you have three dices, then

- The total number of outcomes is 6^3
- If we have event $A =$ all numbers are equal, $|A| = 6$
- Then, we have that $P(A) = \frac{6}{6^3} = \frac{1}{36}$

Outline

Cinvestav

Some Famous Examples

Famous Coin Tosses

- Count of Buffon tossed a coin 4040 times. Heads appeared 2048 times.
- K. Pearson tossed a coin 12000 times and 24000 times.
  - The heads appeared 6019 times and 12012, respectively.

Something Notable

- For these three tosses the relative frequencies of heads are 0.5049, 0.5016,and 0.5005.

Some Famous Examples

Famous Coin Tosses

- Count of Buffon tossed a coin 4040 times. Heads appeared 2048 times.
- K. Pearson tossed a coin 12000 times and 24000 times.
  - The heads appeared 6019 times and 12012, respectively.

Something Notable

- For these three tosses the relative frequencies of heads are 0.5049, 0.5016,and 0.5005.

Outline

Cinvestav

Axioms of Probability

Axioms

Given a sample space S of events, we have that

1. $0 \leq P(A)$ for $A \subseteq S$
2. $P(S) = 1$
3. If A_1 and A_2 are mutually exclusive events (i.e. $P(A_1 \cap A_2) = 0$), then:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Axioms of Probability

Axioms

Given a sample space S of events, we have that

1. $0 \leq P(A)$ for $A \subseteq S$
2. $P(S) = 1$
3. If A_1 and A_2 are mutually exclusive events (i.e. $P(A_1 \cap A_2) = 0$), then:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Axioms of Probability

Axioms

Given a sample space S of events, we have that

1. $0 \leq P(A)$ for $A \subseteq S$
2. $P(S) = 1$
3. If A_1 and A_2 are mutually exclusive events (i.e. $P(A_1 \cap A_2) = 0$), then:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Axioms of Probability

Axioms

Given a sample space S of events, we have that

1. $0 \leq P(A)$ for $A \subseteq S$
2. $P(S) = 1$
3. If A_1 and A_2 are mutually exclusive events (i.e. $P(A_1 \cap A_2) = 0$), then:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Outline

Cinvestav

Events as Sets

For example, in a dice experiment

$$A = \{i|\text{with } i \text{ an even number}\}$$

Events as Sets

For example, in a dice experiment

$$A = \{i | \text{with } i \text{ an even number}\}$$

Thus, we have the following set operations

1. $A \cup B = \{x | x \in A \text{ or } x \in B\}$
2. $A \cap B = \{x | x \in A \text{ and } x \in B\}$
3. $A^C = \{x | x \notin A\}$

Events as Sets

For example, in a dice experiment

$$A = \{i | \text{with } i \text{ an even number}\}$$

Thus, we have the following set operations

1. $A \cup B = \{x | x \in A \text{ or } x \in B\}$
2. $A \cap B = \{x | x \in A \text{ and } x \in B\}$
3. $A^C = \{x | x \notin A\}$

Events as Sets

For example, in a dice experiment

$$A = \{i | \text{with } i \text{ an even number}\}$$

Thus, we have the following set operations

1. $A \cup B = \{x | x \in A \text{ or } x \in B\}$
2. $A \cap B = \{x | x \in A \text{ and } x \in B\}$
3. $A^C = \{x | x \notin A\}$

Therefore

We can use combinations

Of such events with the previous operations to describe random phenomenas

Set of all throws even and greater than 3

- $A = \{i | i \text{ is even}\}$
- $B = \{i | i > 3\}$

Then

$$A \cap B = \{i | i \text{ is even and } i > 3\}$$

Therefore

We can use combinations

Of such events with the previous operations to describe random phenomenas

Set of all throws even and greater than 3

- $A = \{i|i \text{ is even}\}$
- $B = \{i|i > 3\}$

Then

$$A \cap B = \{i|i \text{ is even and } i > 3\}$$

Cinvestav

Therefore

We can use combinations

Of such events with the previous operations to describe random phenomenas

Set of all throws even and greater than 3

- $A = \{i | i \text{ is even}\}$
- $B = \{i | i > 3\}$

Then

$$A \cap B = \{i | i \text{ is even and } i > 3\}$$

Cinvestav

Example

The Probability of the empty set is

$$P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$$

Given that $\overline{S} = \emptyset$, therefore

$$P(\emptyset) = 0$$

Example

The Probability of the empty set is

$$P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$$

Given that $\overline{S} = \emptyset$, therefore

$$P(\emptyset) = 0$$

Examples

The union $A \cup B$ of two events A and B

It is an event that occurs if at least one of the events A or B occur

For mutually exclusive events

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

Cinvestav

Examples

The union $A \cup B$ of two events A and B

It is an event that occurs if at least one of the events A or B occur

For mutually exclusive events

$$P\left(A \cup B\right) = P\left(A\right) + P\left(B\right)$$

Further

In the General Case

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

In the case of the complement

$$P\left(A^C\right) = 1 - P(A)$$

Given that

$$P(S) = P\left(A^C\right) + P(A)$$

Cinvestav

Further

In the General Case

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

In the case of the complement

$$P\left(A^C\right) = 1 - P(A)$$

Given that

$$P(S) = P\left(A^C\right) + P(A)$$

Further

In the General Case

$$P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right)$$

In the case of the complement

$$P\left(A^{C}\right) = 1 - P\left(A\right)$$

Given that

$$P\left(S\right) = P\left(A^{C}\right) + P\left(A\right)$$

Cinvestav

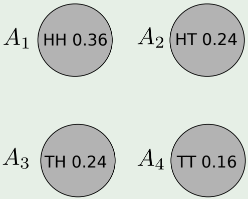
Outline

Cinvestav

Example

Setup

Throw a biased coin twice



We have the following event

At least one head!!! Can you tell me which events are part of it?

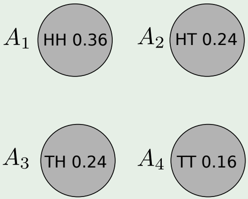
What about this one?

Tail on first toss.

Example

Setup

Throw a biased coin twice



We have the following event

At least one head!!! Can you tell me which events are part of it?

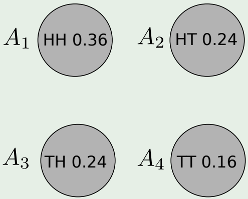
What about this one?

Tail on first toss.

Example

Setup

Throw a biased coin twice



We have the following event

At least one head!!! Can you tell me which events are part of it?

What about this one?

Tail on first toss.

Outline

Cinvestav

We have that experiments in Probability are Defined as

We have

1. The Set $\mathcal{B}$ of all experimental outcomes
2. The Borel Field of all events of $\mathcal{B}$
3. The Probability of Such Events

Remark about the Borel Field

- We us this fields because we are given a way to measure infinite phenomenas but Bounded.

Therefore

- If you have a measure over a set $\mathcal{B}$, we would love to be able to measure:
  - The Union of such events
  - The Measure should be bounded.

We have that experiments in Probability are Defined as

We have

1. The Set $\mathcal{B}$ of all experimental outcomes
2. The Borel Field of all events of $\mathcal{B}$
3. The Probability of Such Events

Remark about the Borel Field

- We us this fields because we are given a way to measure infinite phenomenas but Bounded.

Therefore

- If you have a measure over a set $\mathcal{B}$, we would love to be able to measure:
  - The Union of such events
  - The Measure should be bounded.

We have that experiments in Probability are Defined as

We have

1. The Set $\mathcal{B}$ of all experimental outcomes
2. The Borel Field of all events of $\mathcal{B}$
3. The Probability of Such Events

Remark about the Borel Field

- We us this fields because we are given a way to measure infinite phenomenas but Bounded.

Therefore

- If you have a measure over a set $\mathcal{B}$, we would love to be able to measure:
  - The Union of such events
  - The Measure should be bounded.

Measuring Countable Spaces

If $\mathcal{B} = \{A_1, A_2, ..., A_N\}$

$$P(A_i) = p_i$$

Where

$$p_1 + p_2 + ... + p_N = 1$$

Then, if you have $B = A_1 \cup ... \cup A_k$ and $k \leq N$

$$P(B) = \sum_{i=1}^{k} P(A_i)$$

Measuring Countable Spaces

If $\mathcal{B} = \{A_1, A_2, ..., A_N\}$

$$P(A_i) = p_i$$

Where

$$p_1 + p_2 + ... + p_N = 1$$

Then, if you have $B = A_1 \cup ... \cup A_k$ and $k \leq N$

$$P(B) = \sum_{i=1}^{k} P(A_i)$$

Measuring Countable Spaces

If $\mathcal{B} = \{A_1, A_2, ..., A_N\}$

$$P(A_i) = p_i$$

Where

$$p_1 + p_2 + ... + p_N = 1$$

Then, if you have $B = A_1 \cup ... \cup A_k$ and $k \leq N$

$$P(B) = \sum_{i=1}^{k} P(A_i)$$

Cinvestav

In the Case of Equally Likely Events

We have that

$$p_i = \frac{1}{N}$$

Therefore

$$P(B) = \sum_{i=1}^{k} P(A_i) = \sum_{i=1}^{k} \frac{1}{N} = \frac{k}{N}$$

In the Case of Equally Likely Events

We have that

$$p_i = \frac{1}{N}$$

Therefore

$$P(B) = \sum_{i=1}^{k} P(A_i) = \sum_{i=1}^{k} \frac{1}{N} = \frac{k}{N}$$

The Real Line

Here the Borel Sets

- It comes to save us...

Something Notable

- In this case we are using events as intervals $x_1 \leq x \leq x_2$
- And their finite Unions and Intersections

For this, we define $\mathcal{B}$

The smallest Borel Field that includes half lines $x \leq x_1$ with $x_i \in \mathbb{R}$.

The Real Line

Here the Borel Sets

- It comes to save us...

Something Notable

- In this case we are using events as intervals $x_1 \leq x \leq x_2$
- And their finite Unions and Intersections

For this, we define $\mathcal{B}$

The smallest Borel Field that includes half lines $x \leq x_1$ with $x_i \in \mathbb{R}$.

The Real Line

Here the Borel Sets

- It comes to save us...

Something Notable

- In this case we are using events as intervals $x_1 \leq x \leq x_2$
- And their finite Unions and Intersections

For this, we define $\mathcal{B}$

The smallest Borel Field that includes half lines $x \leq x_1$ with $x_i \in \mathbb{R}$.

Important

This contains all the open and closed intervals, and all points

- This is not all possible subsets

Those sets are not result of countable unions and intersections of intervals

- A Vitali set is a subset V of the interval $[0, 1]$ of real numbers such that, for each real number r:
  - There is exactly one number $v \in V$ such that $v - r$ is a rational number

They do not describe experiments of interest

- These are of no interest for Probability

Important

This contains all the open and closed intervals, and all points

- This is not all possible subsets

Those sets are not result of countable unions and intersections of intervals

- A Vitali set is a subset V of the interval $[0,1]$ of real numbers such that, for each real number r:
  - There is exactly one number $v \in V$ such that $v - r$ is a rational number

They do not describe experiments of interest

- These are of no interest for Probability

Cinvestav

Important

This contains all the open and closed intervals, and all points

- This is not all possible subsets

Those sets are not result of countable unions and intersections of intervals

- A Vitali set is a subset V of the interval $[0,1]$ of real numbers such that, for each real number r:
  - There is exactly one number $v \in V$ such that $v - r$ is a rational number

They do not describe experiments of interest

- These are of no interest for Probability

Therefore, we have

Assume that we have a function $\alpha\left(x\right)$ such that

$$\int_{-\infty}^{\infty} \alpha\left(x\right) dx = 1 \text{ and } \alpha\left(x\right) \geq 0$$

We define that

$$P\left(x \leq x_1\right) = \int_{-\infty}^{x_1} \alpha\left(x\right) dx$$

Further, $x_1 \leq x \leq x_2$ is defined as

$$P\left(x_1 \leq x \leq x_2\right) = \int_{x_1}^{x_2} \alpha\left(x\right) dx$$

Cinvestav

Therefore, we have

Assume that we have a function $\alpha(x)$ such that

$$\int_{-\infty}^{\infty} \alpha(x)\, dx = 1 \text{ and } \alpha(x) \geq 0$$

We define that

$$P(x \leq x_1) = \int_{-\infty}^{x_1} \alpha(x)\, dx$$

Further, $x_1 \leq x \leq x_2$ is defined as

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \alpha(x)\, dx$$

Cinvestav

Therefore, we have

Assume that we have a function $\alpha(x)$ such that

$$\int_{-\infty}^{\infty} \alpha(x)\, dx = 1 \text{ and } \alpha(x) \geq 0$$

We define that

$$P(x \leq x_1) = \int_{-\infty}^{x_1} \alpha(x)\, dx$$

Further, $x_1 \leq x \leq x_2$ is defined as

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \alpha(x)\, dx$$

Cinvestav

Example

We have the following probability of emission of radioactive probabilities

$$\alpha(t) = ce^{-ct} I[t \geq 0] \text{ and } t \in \mathbb{R}$$

Therefore, the probability ob being emitted in the interval $(0, t_0)$

$$\int_0^{t_0} ce^{ct} dt = 1 - e^{-ct_0}$$

Example

We have the following probability of emission of radioactive probabilities

$$\alpha(t) = ce^{-ct} I[t \geq 0] \text{ and } t \in \mathbb{R}$$

Therefore, the probability ob being emitted in the interval $(0, t_0)$

$$\int_0^{t_0} ce^{ct} dt = 1 - e^{-ct_0}$$

Outline

Cinvestav

We need to count!!!

We have four main methods of counting

1. Ordered samples of size r with replacement
2. Ordered samples of size r without replacement
3. Unordered samples of size r without replacement
4. Unordered samples of size r with replacement

We need to count!!!

We have four main methods of counting

1. Ordered samples of size r with replacement
2. Ordered samples of size r without replacement
3. Unordered samples of size r without replacement
4. Unordered samples of size r with replacement

We need to count!!!

We have four main methods of counting

1. Ordered samples of size r with replacement
2. Ordered samples of size r without replacement
3. Unordered samples of size r without replacement
4. Unordered samples of size r with replacement

We need to count!!!

We have four main methods of counting

1. Ordered samples of size r with replacement
2. Ordered samples of size r without replacement
3. Unordered samples of size r without replacement
4. Unordered samples of size r with replacement

Ordered samples of size r with replacement

Definition

The number of possible sequences $(a_{i_1}, ..., a_{i_r})$ for n different numbers is $n \times n \times ... \times n = n^r$

Example

If you throw three dices you have $6 \times 6 \times 6 = 216$

Ordered samples of size r with replacement

Definition

The number of possible sequences $(a_{i_1}, ..., a_{i_r})$ for n different numbers is $n \times n \times ... \times n = n^r$

Example

If you throw three dices you have $6 \times 6 \times 6 = 216$

Ordered samples of size r without replacement

Definition

The number of possible sequences $(a_{i_1}, ..., a_{i_r})$ for n different numbers is $n \times n - 1 \times ... \times (n - (r - 1)) = \frac{n!}{(n-r)!}$

Example

The number of different numbers that can be formed if no digit can be repeated. For example, if you have 4 digits and you want numbers of size 3.

Ordered samples of size r without replacement

Definition

The number of possible sequences $(a_{i_1}, ..., a_{i_r})$ for n different numbers is $n \times n - 1 \times ... \times (n - (r - 1)) = \frac{n!}{(n-r)!}$

Example

The number of different numbers that can be formed if no digit can be repeated. For example, if you have 4 digits and you want numbers of size 3.

Unordered samples of size r without replacement

Definition

Actually, we want the number of possible unordered sets.

However

We have $\frac{n!}{(n-r)!}$ collections where we care about the order. Thus

$$\frac{\frac{n!}{(n-r)!}}{r!} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r} \quad (2)$$

Unordered samples of size r without replacement

Definition

Actually, we want the number of possible unordered sets.

However

We have $\frac{n!}{(n-r)!}$ collections where we care about the order. Thus

$$\frac{\frac{n!}{(n-r)!}}{r!} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r} \tag{2}$$

Unordered samples of size r with replacement

Definition

We want to find an unordered set $\{a_{i_1}, ..., a_{i_r}\}$ with replacement

Thus

$$\binom{n+r-1}{r} \tag{3}$$

Unordered samples of size r with replacement

Definition

We want to find an unordered set $\{a_{i_1}, ..., a_{i_r}\}$ with replacement

Thus

$$\binom{n+r-1}{r} \tag{3}$$

How? Use a digit trick for that

Change encoding by adding more signs

Imagine all the strings of three numbers with $\{1, 2, 3\}$

We have

Old String	New String
111	1+0,1+1,1+2=123
112	1+0,1+1,2+2=124
113	1+0,1+1,3+2=125
122	1+0,2+1,2+2=134
123	1+0,2+1,3+2=135
133	1+0,3+1,3+2=145
222	2+0,2+1,2+2=234
223	2+0,2+1,3+2=235
233	2+0,3+1,3+2=245
333	3+0,3+1,3+2=345

How? Use a digit trick for that

Change encoding by adding more signs

Imagine all the strings of three numbers with $\{1, 2, 3\}$

We have

Old String	New String
111	1+0,1+1,1+2=123
112	1+0,1+1,2+2=124
113	1+0,1+1,3+2=125
122	1+0,2+1,2+2=134
123	1+0,2+1,3+2=135
133	1+0,3+1,3+2=145
222	2+0,2+1,2+2=234
223	2+0,2+1,3+2=235
233	2+0,3+1,3+2=245
333	3+0,3+1,3+2=345

Outline

Cinvestav

Sometimes

We would like to model certain phenomena like

$$P(A_1, A_2, ..., A_K)$$

The Problem is the complexity of calculating the joint distribution

We would like something simpler

Something like

$$P(A_1, A_2, ..., A_K) = Operation_{i=1}^{k} P(A_i)$$

Sometimes

We would like to model certain phenomena like

$$P\left(A_1, A_2, ..., A_K\right)$$

The Problem is the complexity of calculating the joint distribution

We would like something simpler

Something like

$$P\left(A_1, A_2, ..., A_K\right) = Operation_{i=1}^{k} P\left(A_1\right)$$

Sometimes

We would like to model certain phenomena like

$$P\left(A_1, A_2, ..., A_K\right)$$

The Problem is the complexity of calculating the joint distribution

We would like something simpler

Something like

$$P\left(A_1, A_2, ..., A_K\right) = Operation_{i=1}^{k} P\left(A_1\right)$$

Cinvestav

Independence

Definition

Two events A and B are independent if and only if $P(A, B) = P(A \cap B) = P(A)P(B)$

Example

We have two dices

Thus, we have all pairs (i, j) such that $i, j = 1, 2, 3, ..., 6$

Example

We have two dices

Thus, we have all pairs (i, j) such that $i, j = 1, 2, 3, ..., 6$

We have the following events

- $A =${First dice 1,2 or 3}
- $B =$ {First dice 3, 4 or 5}
- $C =$ {The sum of two faces is 9}

Example

We have two dices

Thus, we have all pairs (i, j) such that $i, j = 1, 2, 3, ..., 6$

We have the following events

- $A =$ {First dice 1,2 or 3}
- $B =$ {First dice 3, 4 or 5}
- $C =$ {The sum of two faces is 9}

So, we can do

Look at the board!!! Independence between A, B, C

Example

We have two dices

Thus, we have all pairs (i,j) such that $i,j = 1,2,3,...,6$

We have the following events

- A ={First dice 1,2 or 3}
- B = {First dice 3, 4 or 5}
- C = {The sum of two faces is 9}

So, we can do

Look at the board!!! Independence between A, B, C

Example

We have two dices

Thus, we have all pairs (i,j) such that $i,j = 1,2,3,...,6$

We have the following events

- A ={First dice 1,2 or 3}
- B = {First dice 3, 4 or 5}
- C = {The sum of two faces is 9}

So, we can do

Look at the board!!! Independence between A, B, C

Outline

Cinvestav

Outline

Cinvestav

We have that

Given two sets $\mathcal{A}$ and $\mathcal{B}$

$$\mathcal{A} \times \mathcal{B} = \{(a, b) \,|a \in \mathcal{A} \text{ and } b \in \mathcal{B}\}$$

Example $\mathcal{A} = \{a_1, a_2, a_3\}$ and $\mathcal{B} = \{b_1, b_2\}$

$$\mathcal{A} \times \mathcal{B} = \{(a_1, b_1), (a_2, b_1), (a_3, b_1), (a_1, b_2), (a_2, b_2), (a_3, b_2)\}$$

We have that

Given two sets $\mathcal{A}$ and $\mathcal{B}$

$$\mathcal{A} \times \mathcal{B} = \{(a, b) \,| a \in \mathcal{A} \text{ and } b \in \mathcal{B}\}$$

Example $\mathcal{A} = \{a_1, a_2, a_3\}$ and $\mathcal{B} = \{b_1, b_2\}$

$$\mathcal{A} \times \mathcal{B} = \{(a_1, b_1), (a_2, b_1), (a_3, b_1), (a_1, b_2), (a_2, b_2), (a_3, b_2)\}$$

Cinvestav

Furthermore

If $A \subseteq \mathcal{A}$ and $B \subseteq \mathcal{B}$

$$C = A \times B$$

Look At the Board

- It is interesting!!!

Therefore, $A \times \mathcal{B}$ and $\mathcal{A} \times B$

$$A \times B = A \times \mathcal{B} \cap \mathcal{A} \times B$$

Furthermore

If $A \subseteq \mathcal{A}$ and $B \subseteq \mathcal{B}$

$$C = A \times B$$

Look At the Board

- It is interesting!!!

Therefore, $A \times \mathcal{B}$ and $\mathcal{A} \times B$

$$A \times B = A \times \mathcal{B} \cap \mathcal{A} \times B$$

Furthermore

If $A \subseteq \mathcal{A}$ and $B \subseteq \mathcal{B}$

$$C = A \times B$$

Look At the Board

- It is interesting!!!

Therefore, $A \times \mathcal{B}$ and $\mathcal{A} \times B$

$$A \times B = A \times \mathcal{B} \cap \mathcal{A} \times B$$

Cinvestav

Re-framing Independence

We have

- $P(A \times \mathcal{B}) = P((a,b)\,|a \in A \text{ and } b \in \mathcal{B}) = P(A)$
- $P(\mathcal{A} \times B) = P((a,b)\,|a \in \mathcal{A} \text{ and } b \in B) = P(B)$

Therefore, we can use our previous relation and assuming $A \times \mathcal{B}$ and $\mathcal{A} \times B$ independent events

$$P(A \times B) = P(A \times \mathcal{B} \cap \mathcal{A} \times B) = P(A)\,P(B)$$

Re-framing Independence

We have

- $P(A \times \mathcal{B}) = P((a,b)\,|a \in A \text{ and } b \in \mathcal{B}) = P(A)$
- $P(\mathcal{A} \times B) = P((a,b)\,|a \in \mathcal{A} \text{ and } b \in B) = P(B)$

Therefore, we can use our previous relation and assuming $A \times \mathcal{B}$ and $\mathcal{A} \times B$ independent events

$$P(A \times B) = P(A \times \mathcal{B} \cap \mathcal{A} \times B) = P(A)\,P(B)$$

We can use this to derive the Binomial Distribution

What???

We can do something quite interesting

First, we use a sequence of n Bernoulli Trials

We have this

- "Success" has a probability p.
- "Failure" has a probability $1 - p$.

First, we use a sequence of n Bernoulli Trials

We have this

- "Success" has a probability p.
- "Failure" has a probability $1 - p$.

Examples

- Toss a coin independently n times.
- Examine components produced on an assembly line.

Cinvestav

First, we use a sequence of n Bernoulli Trials

We have this

- "Success" has a probability p.
- "Failure" has a probability $1-p$.

Examples

- Toss a coin independently n times.
- Examine components produced on an assembly line.

Now

We take S =all 2^n ordered sequences of length n, with components **0** **(failure)** and **1** **(success)**.

First, we use a sequence of n Bernoulli Trials

We have this

- "Success" has a probability p.
- "Failure" has a probability $1 - p$.

Examples

- Toss a coin independently n times.
- Examine components produced on an assembly line.

Now

We take S =all 2^n ordered sequences of length n, with components **0 (failure)** and **1 (success)**.

First, we use a sequence of n Bernoulli Trials

We have this

- "Success" has a probability p.
- "Failure" has a probability $1 - p$.

Examples

- Toss a coin independently n times.
- Examine components produced on an assembly line.

Now

We take S =all 2^n ordered sequences of length n, with components **0 (failure)** and **1 (success)**.

Cinvestav

First

How do we represent such events?

We can use a sequence as

$$\langle a_1, a_2, ..., a_n \rangle$$

With the following features

$$a_i \in S = \{0, 1\}$$

First

How do we represent such events?

We can use a sequence as

$$\langle a_1, a_2, ..., a_n \rangle$$

With the following features

$$a_i \in S = \{0, 1\}$$

Meaning

We have one event A

$A = Success = 1$

The Other Event A^C

$A^C = Failure = 0$

Meaning

We have one event A

$A = Success = 1$

The Other Event A^C

$A^C = Failure = 0$

Thus, taking a sample ω

$\omega = 11\cdots10\cdots0 = \{0,1\}\times\cdots\{0,1\}$

k 1's followed by $n-k$ 0's.

Thus, taking a sample ω

$\omega = 11\cdots 10\cdots 0 = \{0,1\} \times \cdots \{0,1\}$

k 1's followed by $n-k$ 0's.

We have then

$$\begin{aligned} P(\omega) &= P\left(A_1 \cap A_2 \cap \ldots \cap A_k \cap A_{k+1}^c \cap \ldots \cap A_n^c\right) \\ &= P(A_1) P(A_2) \cdots P(A_k) P\left(A_{k+1}^c\right) \cdots P(A_n^c) \\ &= p^k (1-p)^{n-k} \end{aligned}$$

Cinvestav

Did you notice the following?

After mapping the events through the probability

- We are loosing the internal event structure

Which is not important because

Events are mutually independent!!!

Important

The number of such sample is the number of sets with k elements.... or...

$$\binom{n}{k}$$

Did you notice the following?

After mapping the events through the probability

- We are loosing the internal event structure

Which is not important because

Events are mutually independent!!!

Important

The number of such sample is the number of sets with k elements.... or...

$$\binom{n}{k}$$

Cinvestav

Did you notice the following?

After mapping the events through the probability

- We are loosing the internal event structure

Which is not important because

Events are mutually independent!!!

Important

The number of such sample is the number of sets with k elements.... or...

$$\binom{n}{k}$$

Cinvestav

Therefore

We do not care where the 1's and 0's are

Thus all the probabilities are equal to $p^k (1-p)^k$

Thus, we are looking to sum all those probabilities of all those combinations of 1's and 0's

$$\sum_{k\ 1\text{'s}} p\left(\omega^k\right)$$

Then

$$\sum_{k\ 1\text{'s}} p\left(\omega^k\right) = \begin{pmatrix} n \\ k \end{pmatrix} p\,(1-p)^{n-k}$$

Therefore

We do not care where the 1's and 0's are

Thus all the probabilities are equal to $p^k (1-p)^k$

Thus, we are looking to sum all those probabilities of all those combinations of 1's and 0's

$$\sum_{k \text{ 1's}} p\left(\omega^k\right)$$

Then

$$\sum_{k \text{ 1's}} p\left(\omega^k\right) = \binom{n}{k} p\left(1-p\right)^{n-k}$$

Therefore

We do not care where the 1's and 0's are

Thus all the probabilities are equal to $p^k (1-p)^k$

Thus, we are looking to sum all those probabilities of all those combinations of 1's and 0's

$$\sum_{k \text{ 1's}} p\left(\omega^k\right)$$

Then

$$\sum_{k \text{ 1's}} p\left(\omega^k\right) = \begin{pmatrix} n \\ k \end{pmatrix} p\,(1-p)^{n-k}$$

Proving this is a probability

Sum of these probabilities is equal to 1

$$\sum_{k=0}^{n} \binom{n}{k} p\left(1-p\right)^{n-k} = \left(p+\left(1-p\right)\right)^{n} = 1$$

The other is simple

$$0 \leq \binom{n}{k} p\left(1-p\right)^{n-k} \leq 1 \ \forall k$$

This is know as

The Binomial probability function!!!

Proving this is a probability

Sum of these probabilities is equal to 1

$$\sum_{k=0}^{n}\begin{pmatrix} n \\ k \end{pmatrix} p\left(1-p\right)^{n-k} = \left(p+\left(1-p\right)\right)^{n} = 1$$

The other is simple

$$0 \leq \begin{pmatrix} n \\ k \end{pmatrix} p\left(1-p\right)^{n-k} \leq 1 \ \forall k$$

This is know as

The Binomial probability function!!!

Proving this is a probability

Sum of these probabilities is equal to 1

$$\sum_{k=0}^{n} \binom{n}{k} p\,(1-p)^{n-k} = (p + (1-p))^{n} = 1$$

The other is simple

$$0 \leq \binom{n}{k} p\,(1-p)^{n-k} \leq 1 \;\forall k$$

This is know as

The Binomial probability function!!!

Outline

Cinvestav

Unconditional Probability

Definition

An **unconditional probability** is the probability of an event A prior to arrival of any evidence.

- "There is a 10% chance that the patient is having a cavity"

Unconditional Probability

Definition

An **unconditional probability** is the probability of an event A prior to arrival of any evidence.

For Example

- $P(Cavity) = 0.1$means that in the absence of any other information.
    - "There is a 10% chance that the patient is having a cavity"

Unconditional Probability

Definition

An **unconditional probability** is the probability of an event A prior to arrival of any evidence.

For Example

- $P(Cavity) = 0.1$means that in the absence of any other information.
  - "**There is a 10% chance that the patient is having a cavity**"

Conditional Probability

Definition

A **conditional probability** is the probability of one event if another event occurred.

- "there is an 80% chance that the patient is having a cavity given that he is having a toothache"

Conditional Probability

Definition

A **conditional probability** is the probability of one event if another event occurred.

For Example

- $P(Cavity/Toothache) = 0.8$ means that
  - "there is an 80% chance that the patient is having a cavity given that he is having a toothache"

Conditional Probability

Definition

A **conditional probability** is the probability of one event if another event occurred.

For Example

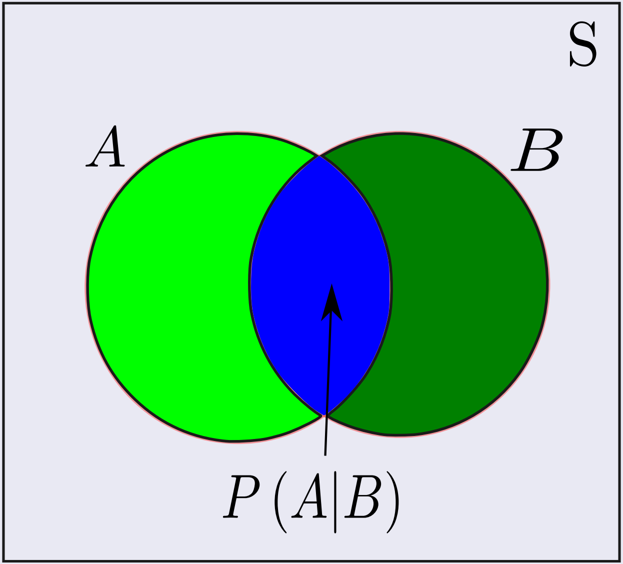
- $P(Cavity/Toothache) = 0.8$ means that
  - "there is an 80% chance that the patient is having a cavity given that he is having a toothache"

Outline

Cinvestav

Basically

Using Set Theory



However

We need a distribution!!!

$$\sum_{A \subseteq S} P(A) = 1$$

We then do the following

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

However

We need a distribution!!!

$$\sum_{A \subseteq S} P(A) = 1$$

We then do the following

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

Therefore

The conditional probability of A given B is written $P(A|B)$

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

with $P(B) > 0$

We have that this are probabilities

First given $0 < P(B)$ and $0 \leq P(A \cap B)$

Then,

$$\frac{P(A, B)}{P(B)} \geq 0$$

Second, given if $B \subseteq A$

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

If $A \subseteq B$

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(A)}{P(B)} \geq P(A) \geq 0$$

We have that this are probabilities

First given $0 < P(B)$ and $0 \leq P(A \cap B)$

Then,

$$\frac{P(A,B)}{P(B)} \geq 0$$

Second, given if $B \subseteq A$

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

If $A \subseteq B$

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A)}{P(B)} \geq P(A) \geq 0$$

Cinvestav

We have that this are probabilities

First given $0 < P(B)$ and $0 \leq P(A \cap B)$

Then,

$$\frac{P(A,B)}{P(B)} \geq 0$$

Second, given if $B \subseteq A$

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

If $A \subseteq B$

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A)}{P(B)} \geq P(A) \geq 0$$

Finally

We have that for $A \cap B = \emptyset$

$$P(A \cup B|C) = \frac{P([A \cup B] \cap C)}{P(C)} = \frac{P([A \cap C] \cup [B \cap C])}{P(C)}$$

Then

$$P(A \cup B|C) = \frac{P(A \cap C) + P(B \cap C)}{P(C)} = \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C)}{P(C)}$$

Finally

We have that for $A \cap B = \emptyset$

$$P(A \cup B|C) = \frac{P([A \cup B] \cap C)}{P(C)} = \frac{P([A \cap C] \cup [B \cap C])}{P(C)}$$

Then

$$P(A \cup B|C) = \frac{P(A \cap C) + P(B \cap C)}{P(C)} = \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C)}{P(C)}$$

Chain Rule

The probability that two events A and B will both occur is

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

How?

Any Ideas?

Chain Rule

The probability that two events A and B will both occur is

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

How?

Any Ideas?

Chain Rule

The probability that two events A and B will both occur is

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

How?

Any Ideas?

Therefore

This is also know

As the chain rule

Prove by induction

$$P(A_1, ..., A_n) =$$
$$P(A_n|A_{n-1}...A_1)\, P(A_{n-1}|A_{n-2}...A_1) \cdots P(A_2|A_1)\, P(A_1)$$

Proof

Any idea?

Therefore

This is also know

As the chain rule

Prove by induction

$$P\left(A_1, \ldots, A_n\right) =$$
$$P\left(A_n | A_{n-1} \ldots A_1\right) P\left(A_{n-1} | A_{n-2} \ldots A_1\right) \cdots P\left(A_2 | A_1\right) P\left(A_1\right)$$

Proof

Any idea?

Cinvestav

Therefore

This is also know

As the chain rule

Prove by induction

$$P\left(A_1, \ldots, A_n\right) =$$
$$P\left(A_n | A_{n-1} \ldots A_1\right) P\left(A_{n-1} | A_{n-2} \ldots A_1\right) \cdots P\left(A_2 | A_1\right) P\left(A_1\right)$$

Proof

Any idea?

Cinvestav

Outline

Independence

If two events are independent

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

Therefore, two events A and B are independent if

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

Independence

If two events are independent

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

Therefore, two events A and B are independent if

$$P(A, B) = P(A) P(B)$$

Example

Experiment

It involves a random draw from a standard deck of 52 playing cards.

Define events A and B to be

A =The card is heart and B =The card is queen

Are the events independent?

How do we do it?

Example

Experiment

It involves a random draw from a standard deck of 52 playing cards.

Define events A and B to be

A =The card is heart and B =The card is queen

Are the events independent?

How do we do it?

Example

Experiment

It involves a random draw from a standard deck of 52 playing cards.

Define events A and B to be

$A =$The card is heart and $B =$The card is queen

Are the events independent?

How do we do it?

Example

We have that

$$P(A, B) = \frac{1}{52}$$

But

$$P(A)\,P(B) = \frac{13}{52} \times \frac{4}{52}$$

Example

We have that

$$P(A, B) = \frac{1}{52}$$

But

$$P(A)\,P(B) = \frac{13}{52} \times \frac{4}{52}$$

What happen when you have independence in conditional setups?

Conditional independence

A and B are conditionally independent given C if and only if

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

Example

$P(WetGrass|Season, Rain) = P(WetGrass|Rain)$.

What happen when you have independence in conditional setups?

Conditional independence

A and B are conditionally independent given C if and only if

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

Example

$P(WetGrass|Season, Rain) = P(WetGrass|Rain)$.

Example

Three cards are drawn from a deck

Find the probability of no obtaining a heart

We have

- 52 cards
- 39 of them not a heart

Define each of the draws

$A_i = \{\text{Card } i \text{ is not a heart}\}$ Then?

Example

Three cards are drawn from a deck

Find the probability of no obtaining a heart

We have

- 52 cards
- 39 of them not a heart

Define each of the draws

$A_i = \{\text{Card } i \text{ is not a heart}\}$ Then?

Example

Three cards are drawn from a deck

Find the probability of no obtaining a heart

We have

- 52 cards
- 39 of them not a heart

Define each of the draws

$A_i = \{\text{Card } i \text{ is not a heart}\}$ Then?

Outline

We have

Definition

Events $H_1, H_2, ..., H_n$ form a partition of the sample space S if

1. They are mutually exclusive $H_i \cap H_j = \emptyset$ and $i \neq j$
2. Their union is the sample space S, $\cup_{i=1}^{n} H_i = S$

We have

Definition

Events $H_1, H_2, ..., H_n$ form a partition of the sample space S if

1. They are mutually exclusive $H_i \cap H_j = \emptyset$ and $i \neq j$
2. Their union is the sample space S, $\cup_{i=1}^{n} H_i = S$

The events $H_1, H_2, ..., H_n$ are usually called hypotheses

$$P(S) = P(H_1) + P(H_2) + \cdots + P(H_n)$$

We have

Definition

Events $H_1, H_2, ..., H_n$ form a partition of the sample space S if

1. They are mutually exclusive $H_i \cap H_j = \emptyset$ and $i \neq j$
2. Their union is the sample space S, $\cup_{i=1}^{n} H_i = S$

The events $H_1, H_2, ..., H_n$ are usually called hypotheses

$$P(S) = P(H_1) + P(H_2) + \cdots + P(H_n)$$

We have

Definition

Events $H_1, H_2, ..., H_n$ form a partition of the sample space S if

1. They are mutually exclusive $H_i \cap H_j = \emptyset$ and $i \neq j$
2. Their union is the sample space S, $\cup_{i=1}^{n} H_i = S$

The events $H_1, H_2, ..., H_n$ are usually called hypotheses

$$P(S) = P(H_1) + P(H_2) + \cdots + P(H_n)$$

Now

Let the event of interest A happens under any of the hypotheses H_i

- With a know conditional probability $P(A|H_i)$

Assume

- The probabilities of hypotheses $H_1, \ldots, H_n$ are known.

Total Probability Formula

$$P(A) = P(A|H_1)P(H_1) + \cdots + P(A|H_n)P(H_n)$$

Now

Let the event of interest A happens under any of the hypotheses H_i

- With a know conditional probability $P(A|H_i)$

Assume

- The probabilities of hypotheses $H_1, ..., H_n$ are known.

Total Probability Formula

$$P(A) = P(A|H_1)P(H_1) + \cdots + P(A|H_n)P(H_n)$$

Cinvestav

Now

Let the event of interest A happens under any of the hypotheses H_i

- With a know conditional probability $P(A|H_i)$

Assume

- The probabilities of hypotheses $H_1, ..., H_n$ are known.

Total Probability Formula

$$P(A) = P(A|H_1)P(H_1) + \cdots + P(A|H_n)P(H_n)$$

Example

Two-headed coin

Out of 100 coins one has heads on both sides.

One coin is chosen at random and flipped two times

What is the probability to get

1. Two heads?
2. Two tails?

Cinvestav

Example

Two-headed coin

Out of 100 coins one has heads on both sides.

One coin is chosen at random and flipped two times

What is the probability to get

1. Two heads?
2. Two tails?

Example

Two-headed coin

Out of 100 coins one has heads on both sides.

One coin is chosen at random and flipped two times

What is the probability to get

1. Two heads?
2. Two tails?

Example

Let A be the event that two heads are obtained

Denote by H_1 the event (hypothesis) that a fair coin was chosen.

Example

Let A be the event that two heads are obtained

Denote by H_1 the event (hypothesis) that a fair coin was chosen.

Now

The Hypothesis $H_2 = H_1^C$ is the event that the two-headed coin was chosen.

Example

Let A be the event that two heads are obtained

Denote by H_1 the event (hypothesis) that a fair coin was chosen.

Now

The Hypothesis $H_2 = H_1^C$ is the event that the two-headed coin was chosen.

Then, we have that

$$
\begin{aligned}
P(A) &= P(A|H_1)\,P(H_1) + P(A|H_2)\,P(H_2) \\
&= \frac{1}{4} \times \frac{99}{100} + 1 \times \frac{1}{100} \\
&= \frac{103}{400} \\
&= 0.2575
\end{aligned}
$$

Example

Let A be the event that two heads are obtained

Denote by H_1 the event (hypothesis) that a fair coin was chosen.

Now

The Hypothesis $H_2 = H_1^C$ is the event that the two-headed coin was chosen.

Then, we have that

$$\begin{aligned} P(A) &= P(A|H_1)\,P(H_1) + P(A|H_2)\,P(H_2) \\ &= \frac{1}{4} \times \frac{99}{100} + 1 \times \frac{1}{100} \\ &= \frac{103}{400} \\ &= 0.2575 \end{aligned}$$

Example

Let A be the event that two heads are obtained

Denote by H_1 the event (hypothesis) that a fair coin was chosen.

Now

The Hypothesis $H_2 = H_1^C$ is the event that the two-headed coin was chosen.

Then, we have that

$$
\begin{aligned}
P(A) &= P(A|H_1)\,P(H_1) + P(A|H_2)\,P(H_2) \\
&= \frac{1}{4} \times \frac{99}{100} + 1 \times \frac{1}{100} \\
&= \frac{103}{400} \\
&= 0.2575
\end{aligned}
$$

Example

Let A be the event that two heads are obtained

Denote by H_1 the event (hypothesis) that a fair coin was chosen.

Now

The Hypothesis $H_2 = H_1^C$ is the event that the two-headed coin was chosen.

Then, we have that

$$
\begin{aligned}
P(A) &= P(A|H_1)\,P(H_1) + P(A|H_2)\,P(H_2)\\
&= \frac{1}{4} \times \frac{99}{100} + 1 \times \frac{1}{100}\\
&= \frac{103}{400}\\
&= 0.2575
\end{aligned}
$$

What about the second one

Exercise

Answer: 0.2475

Outline

Cinvestav

Bayes Theorem

First

Let the event of interest A happens under any of hypotheses H_i with a known (conditional) probability $P(A|H_i)$.

Assume

That the probabilities of hypotheses $H_1, ..., H_n$ are known (prior probabilities).

Then

The conditional (posterior) probability of the hypothesis H_i given that A happened is

$$P(H_i|A) = \frac{P(A|H_i)\,P(H_i)}{P(A)}$$

Bayes Theorem

First

Let the event of interest A happens under any of hypotheses H_i with a known (conditional) probability $P(A|H_i)$.

Assume

That the probabilities of hypotheses $H_1, ..., H_n$ are known (prior probabilities).

Then

The conditional (posterior) probability of the hypothesis H_i given that A happened is

$$P(H_i|A) = \frac{P(A|H_i)\,P(H_i)}{P(A)}$$

Bayes Theorem

First

Let the event of interest A happens under any of hypotheses H_i with a known (conditional) probability $P(A|H_i)$.

Assume

That the probabilities of hypotheses $H_1, ..., H_n$ are known (prior probabilities).

Then

The conditional (posterior) probability of the hypothesis H_i given that A happened is

$$P\left(H_i|A\right) = \frac{P\left(A|H_i\right)P\left(H_i\right)}{P\left(A\right)}$$

Given the independence of the events

$H_1, H_2, ..., H_n$ form a partition of the sample space S

- Therefore

$$A = S \cap A = (H_1 \cup H_2 \cup \cdots \cup H_n) \cap A$$

Therefore

$$A = \cup_{i=1}^{n} (H_i \cap A)$$

Given the independence of the events

$H_1, H_2, ..., H_n$ form a partition of the sample space S

- Therefore

$$A = S \cap A = (H_1 \cup H_2 \cup \cdots \cup H_n) \cap A$$

Therefore

$$A = \cup_{i=1}^{n} (H_i \cap A)$$

Where

We have

$$
\begin{aligned}
P(A) &= P(H_1 \cap A) + P(H_2 \cap A) + \cdots + P(H_n \cap A) \\
&= P(A|H_1)\, P(H_1) + \cdots + P(A|H_n)\, P(H_n)
\end{aligned}
$$

Bayes Law of Total Probability

Therefore for an event H_i

$$p\left(A, H_i\right) = P\left(A|H_i\right) P\left(H_i\right)$$

Then

$$P\left(H_i|A\right) = \frac{p\left(A, H_i\right)}{P\left(A\right)}$$

Bayes Law of Total Probability

Therefore for an event H_i

$$p\left(A, H_i\right) = P\left(A|H_i\right) P\left(H_i\right)$$

Then

$$P\left(H_i|A\right) = \frac{p\left(A, H_i\right)}{P\left(A\right)}$$

Cinvestav

Thus

We have that

$$P\left(H_i|A\right)=\frac{P\left(A|H_i\right)P\left(H_i\right)}{P\left(A\right)}$$

Finally

$$P\left(H_i|A\right)=\frac{P\left(A|H_i\right)P\left(H_i\right)}{P\left(A|H_1\right)P\left(H_1\right)+\cdots+P\left(A|H_n\right)P\left(H_n\right)}$$

Thus

We have that

$$P\left(H_i|A\right) = \frac{P\left(A|H_i\right)P\left(H_i\right)}{P\left(A\right)}$$

Finally

$$P\left(H_i|A\right) = \frac{P\left(A|H_i\right)P\left(H_i\right)}{P\left(A|H_1\right)P\left(H_1\right) + \cdots + P\left(A|H_n\right)P\left(H_n\right)}$$

Another Interpretation

One Version

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

- It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- It is also called the likelihood.

Another Interpretation

One Version

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

Where

- $P(A)$ is the **prior probability** or marginal probability of A.
  - It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B.
  - It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A.
  - It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

Another Interpretation

One Version

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

Where

- $P(A)$ is the **prior probability** or marginal probability of A.
  - It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B.
  - It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A.
  - It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

Another Interpretation

One Version

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

Where

- $P(A)$ is the **prior probability** or marginal probability of A.
  - It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B.
  - It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A.
  - It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

Another Interpretation

One Version

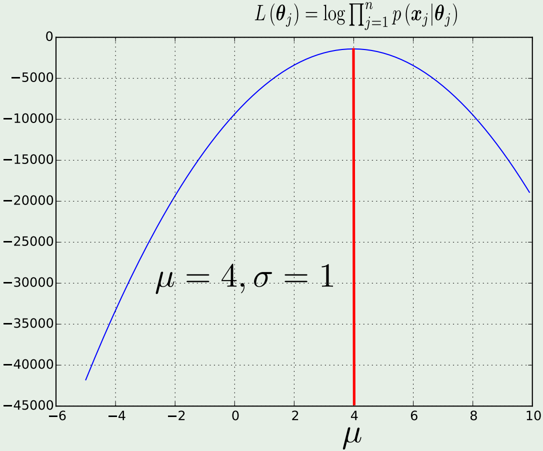
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where

- $P(A)$ is the **prior probability** or marginal probability of A.
  - It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B.
  - It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A.
  - It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

In the case of Gaussian Distributions

$L(\boldsymbol{\theta}_j) = \log \prod_{j=1}^{n} p(\boldsymbol{x}_j | \boldsymbol{\theta}_j)$



Example

Setup

Throw two unbiased dice independently.

Let

1. A ={sum of the faces =8}
2. B ={faces are equal}

Then calculate $P(B|A)$

Look at the board

Example

Setup

Throw two unbiased dice independently.

Let

1. A ={sum of the faces =8}
2. B ={faces are equal}

Then calculate $P(B|A)$

Look at the board

Example

Setup

Throw two unbiased dice independently.

Let

1. A ={sum of the faces =8}
2. B ={faces are equal}

Then calculate $P(B|A)$

Look at the board

Another Example

We have the following

Two coins are available, one unbiased and the other two headed

Assume

That you have a probability of $\frac{3}{4}$ to choose the unbiased

- $B_1 = \{\text{Unbiased coin chosen}\}$
- $B_2 = \{\text{Biased coin chosen}\}$
  - Find that if a head come up, find the probability that the two headed coin was chosen

Another Example

We have the following

Two coins are available, one unbiased and the other two headed

Assume

That you have a probability of $\frac{3}{4}$ to choose the unbiased

Events

- A= {head comes up}
- B_1= {Unbiased coin chosen}
- B_2= {Biased coin chosen}

Another Example

We have the following

Two coins are available, one unbiased and the other two headed

Assume

That you have a probability of $\frac{3}{4}$ to choose the unbiased

Events

- A= {head comes up}
- B_1= {Unbiased coin chosen}
- B_2= {Biased coin chosen}
  - Find that if a head come up, find the probability that the two headed coin was chosen

Another Example

We have the following

Two coins are available, one unbiased and the other two headed

Assume

That you have a probability of $\frac{3}{4}$ to choose the unbiased

Events

- $A=$ {head comes up}
- $B_1=$ {Unbiased coin chosen}
- $B_2=$ {Biased coin chosen}
    - Find that if a head come up, find the probability that the two headed coin was chosen

Another Example

We have the following

Two coins are available, one unbiased and the other two headed

Assume

That you have a probability of $\frac{3}{4}$ to choose the unbiased

Events

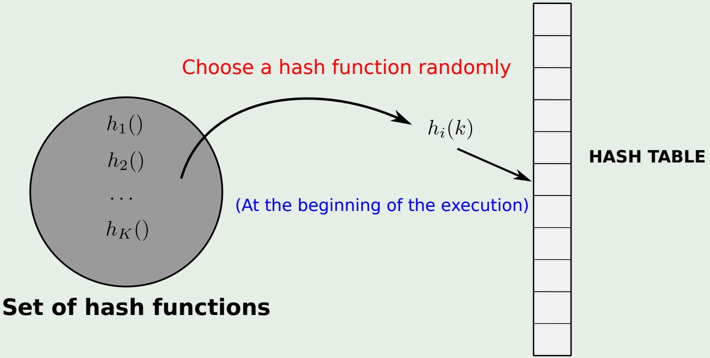
- $A=$ {head comes up}
- $B_1=$ {Unbiased coin chosen}
- $B_2=$ {Biased coin chosen}
  - Find that if a head come up, find the probability that the two headed coin was chosen

Outline

Cinvestav

Universal Hashing

Example



Definition of Universal Hash Functions

Definition

Let $H = \{h : U \rightarrow \{0, 1, ..., m-1\}\}$ be a family of hash functions. H is called a universal family if

$$\forall x, y \in U, x \neq y : \underset{h \in H}{Pr}(h(x) = h(y)) \leq \frac{1}{m} \tag{4}$$

Main result

With universal hashing the chance of collision between distinct keys k and l is no more than the $\frac{1}{m}$ chance of collision if locations $h(k)$ and $h(l)$ were randomly and independently chosen from the set $\{0, 1, ..., m-1\}$.

Cinvestav

Definition of Universal Hash Functions

Definition

Let $H = \{h : U \to \{0, 1, ..., m-1\}\}$ be a family of hash functions. H is called a universal family if

$$\forall x, y \in U, x \neq y : \Pr_{h \in H}(h(x) = h(y)) \leq \frac{1}{m} \tag{4}$$

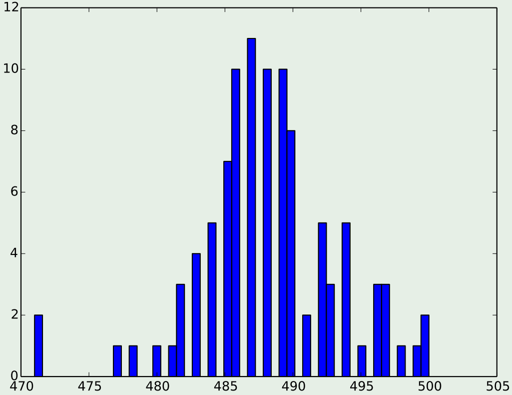
Main result

With universal hashing the chance of collision between distinct keys k and l is no more than the $\frac{1}{m}$ chance of collision if locations $h(k)$ and $h(l)$ were randomly and independently chosen from the set $\{0, 1, ..., m-1\}$.

Cinvestav

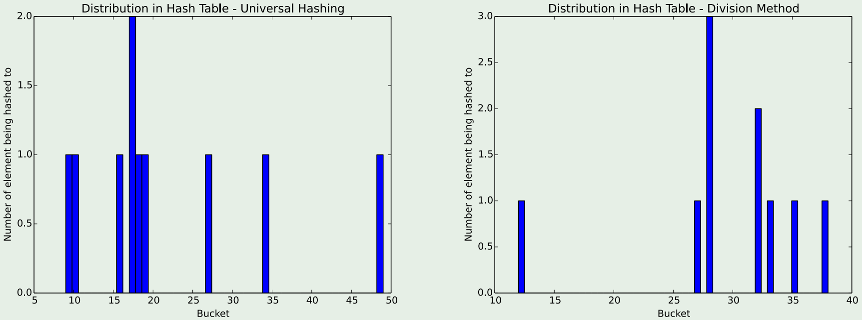
Example of key distribution

Example, mean $= 488.5$ and dispersion $= 5$



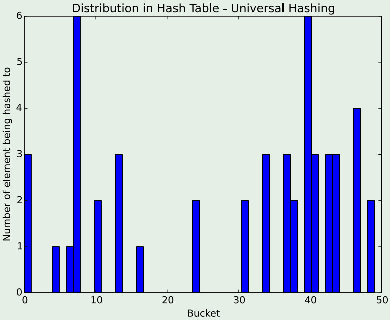
Example with 10 keys

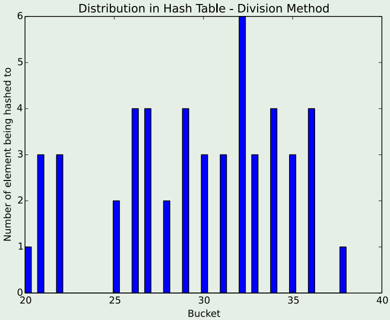
Universal Hashing Vs Division Method



Example with 50 keys

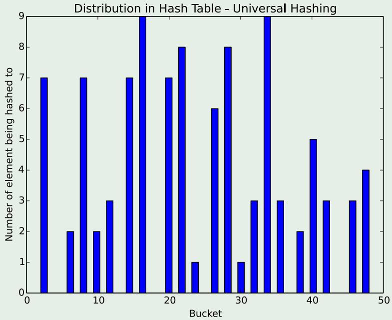
Universal Hashing Vs Division Method

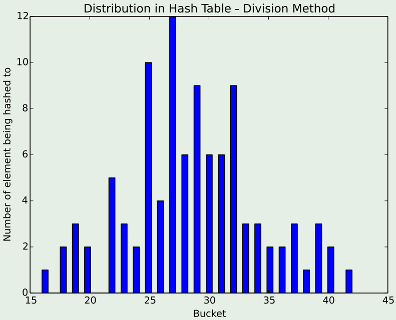




Example with 100 keys

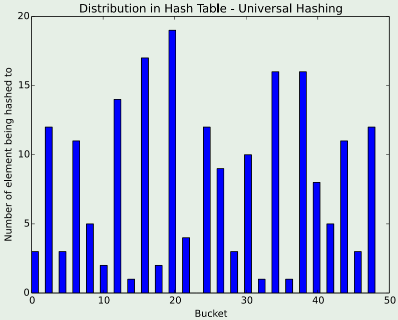
Universal Hashing Vs Division Method

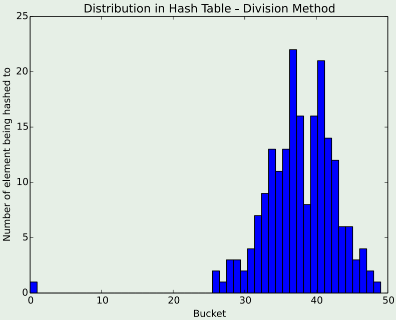




Example with 200 keys

Universal Hashing Vs Division Method





Outline

Cinvestav

Random Variables

In many experiments,

It is easier to deal with a summary variable than with the original probability structure.

Example

In an opinion poll, we ask 50 people whether agree or disagree with a certain issue

- Suppose we record a "1" for agree and "0" for disagree.

The sample space for this experiment has 2^{50} elements

- Why?

Suppose we are only interested in the number of people who agree

- Define the variable X =number of "1" 's recorded out of 50.
  - Easier to deal with this sample space (has only 51 elements).

Example

In an opinion poll, we ask 50 people whether agree or disagree with a certain issue

- Suppose we record a "1" for agree and "0" for disagree.

The sample space for this experiment has 2^{50} elements

- Why?

Suppose we are only interested in the number of people who agree

- Define the variable X =number of "1" 's recorded out of 50.
  - Easier to deal with this sample space (has only 51 elements).

Cinvestav

Example

In an opinion poll, we ask 50 people whether agree or disagree with a certain issue

- Suppose we record a "1" for agree and "0" for disagree.

The sample space for this experiment has 2^{50} elements

- Why?

Suppose we are only interested in the number of people who agree

- Define the variable X =number of "1" 's recorded out of 50.
  - Easier to deal with this sample space (has only 51 elements).

Cinvestav

Thus

It is necessary to define a function "random variable as follow"

$$X : S \to \mathbb{R}$$

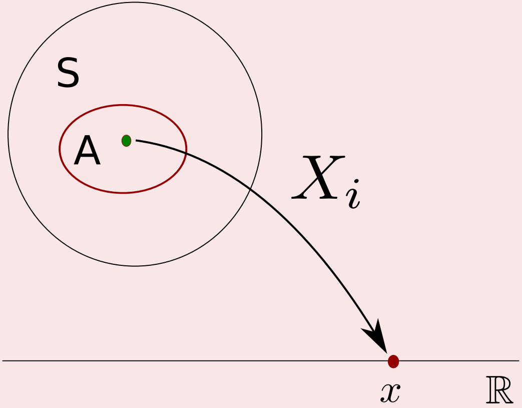
Graphically

Thus

It is necessary to define a function "random variable as follow"

$$X : S \to \mathbb{R}$$

Graphically



Definition

How?

What is the probability function of the random variable is being defined from the probability function of the original sample space?

For this

- Suppose the sample space is $S = \{s_1, s_2, ..., s_n\}$

Now

- Suppose the range of the random variable $X =< x_1, x_2, ..., x_m >$

Definition

How?

What is the probability function of the random variable is being defined from the probability function of the original sample space?

For this

- Suppose the sample space is $S = \{s_1, s_2, ..., s_n\}$

Now

- Suppose the range of the random variable $X = < x_1, x_2, ..., x_m >$

Definition

How?

What is the probability function of the random variable is being defined from the probability function of the original sample space?

For this

- Suppose the sample space is $S = \{s_1, s_2, ..., s_n\}$

Now

- Suppose the range of the random variable $X =< x_1, x_2, ..., x_m >$

Then

We have that

- We observe $X = x_i$ if and only if the outcome of the random experiment is an $s \in S$ s.t. $X(s) = x_j$

Or

$$P(X = x_j) = P(s \in S | X(s) = x_j)$$

Then

We have that

- We observe $X = x_i$ if and only if the outcome of the random experiment is an $s \in S$ s.t. $X(s) = x_j$

Or

$$P\left(X = x_j\right) = P\left(s \in S | X(s) = x_j\right)$$

Therefore

If the events in S are disjoint

$$P\left(X = x_j\right) = \sum_{s \in S} P\left(s \middle| X\left(s\right) = x_j\right)$$

Therefore if we can decompose S

We can easily see the relationship between Random Variables and The Events in S

Therefore

If the events in S are disjoint

$$P\left(X = x_j\right) = \sum_{s \in S} P\left(s | X\left(s\right) = x_j\right)$$

Therefore if we can decompose S

We can easily see the relationship between Random Variables and The Events in S

Outline

Cinvestav

We have

Definition

- A Random Variable X is a process of assigning a number $X(A)$ to every outcome A.

The resulting function must satisfy the the following two conditions

1. The set $\{X \leq x\}$ is an event for every $x \in \mathbb{R}$.
2. The probability of the events $\{X = \infty\}$ and $X = -\infty$ equal zero:

$$P\{X = \infty\} = 0 \; P\{X = -\infty\} = 0$$

We have

Definition

- A Random Variable X is a process of assigning a number $X(A)$ to every outcome A.

The resulting function must satisfy the the following two conditions

1. The set $\{X \leq x\}$ is an event for every $x \in \mathbb{R}$.
2. The probability of the events $\{X = \infty\}$ and $X = -\infty$ equal zero:

$$P\{X = \infty\} = 0 \ P\{X = -\infty\} = 0$$

Example

Setup

Throw a coin 10 times, and let R be the number of heads.

Then

S = all sequences of length 10 with components H and T

We have for

ω =HHHHTTHTTH $\Rightarrow R(\omega) = 6$

Example

Setup

Throw a coin 10 times, and let R be the number of heads.

Then

$S =$ all sequences of length 10 with components H and T

We have for

$\omega =$ HHHHTTHTTH $\Rightarrow R(\omega) = 6$

Example

Setup

Throw a coin 10 times, and let R be the number of heads.

Then

$S =$ all sequences of length 10 with components H and T

We have for

$\omega =$HHHHTTHTTH $\Rightarrow R(\omega) = 6$

Example

Setup

Let R be the number of heads in two independent tosses of a coin.

- Probability of head is .6

What are the probabilities?

Ω ={HH,HT,TH,TT}

Thus, we can calculate

$P(R=0), P(R=1), P(R=2)$

Example

Setup

Let R be the number of heads in two independent tosses of a coin.

- Probability of head is .6

What are the probabilities?

$\Omega = \{\text{HH,HT,TH,TT}\}$

Thus, we can calculate

$P(R=0), P(R=1), P(R=2)$

Example

Setup

Let R be the number of heads in two independent tosses of a coin.

- Probability of head is .6

What are the probabilities?

Ω ={HH,HT,TH,TT}

Thus, we can calculate

$P\left(R=0\right), P\left(R=1\right), P\left(R=2\right)$

Outline

Cinvestav

Note

If we are interested in a random variable X

We want to know its probabilities

Basically

Measurement of such variables leads to measurements as

$$a \leq X \leq b$$

Therefore, we are looking at the following probabilities

$$P\left(s \middle| a \leq X\left(s\right) \leq b\right)$$

Note

If we are interested in a random variable X

We want to know its probabilities

Basically

Measurement of such variables leads to measurements as

$$a \leq X \leq b$$

Therefore, we are looking at the following probabilities

$$P\left(s \middle| a \leq X\left(s\right) \leq b\right)$$

Note

If we are interested in a random variable X

We want to know its probabilities

Basically

Measurement of such variables leads to measurements as

$$a \leq X \leq b$$

Therefore, we are looking at the following probabilities

$$P\left(s | a \leq X\left(s\right) \leq b\right)$$

Then

Definition

- The distribution of a Random Variable X is the function

$$F_X(x) = P\{X \leq x\}$$

  - Defined for all $x \in \mathbb{R}$

Example

For example, if a coin is tossed independently n times

With:

1. Probability p of coming heads on a given toss.
2. And X is the number of heads

We have that

$$P\left(a \leq X\left(s\right) \leq b\right) = \sum_{k=1}^{b} \begin{pmatrix} n \\ k \end{pmatrix} p^{k} \left(1-p\right)^{n-k}$$

Example

For example, if a coin is tossed independently n times

With:

1. Probability p of coming heads on a given toss.
2. And X is the number of heads

We have that

$$P\left(a \leq X\left(s\right) \leq b\right) = \sum_{k=1}^{b} \left(\begin{array}{c} n \\ k \end{array} \right) p^{k} \left(1-p\right)^{n-k}$$

Cinvestav

Outline

Cinvestav

We have Two Types of Random Variables

Definition

The Random Variable X is said to be discrete if and only if the set of possible values of X is finite or countably infinite.

Then

If $x_1, x_2, \ldots$ are the values of X that belong to the range R of it,

$$P\left(X = x_1, X = x_2, \ldots\right) = \sum_{x \in R} p_X\left(x\right)$$

We have Two Types of Random Variables

Definition

The Random Variable X is said to be discrete if and only if the set of possible values of X is finite or countably infinite.

Then

If $x_1, x_2, ...$ are the values of X that belong to the range R of it,

$$P\left(X = x_1, X = x_2, ...\right) = \sum_{x \in R} p_X\left(x\right)$$

In the case of Continuous Random Variables

Definition

A continuous random variable can assume a continuous range of values.

However, we would use something more formal for this

Using integrals.

In the case of Continuous Random Variables

Definition

A continuous random variable can assume a continuous range of values.

However, we would use something more formal for this

Using integrals.

Examples

Random variable X has uniform $U(a,b)$ distribution if its density is given by

$$f\left(x|a,b\right)=\begin{cases}\frac{1}{b-a} & a\leq x\leq b\\ 0 & else\end{cases}$$

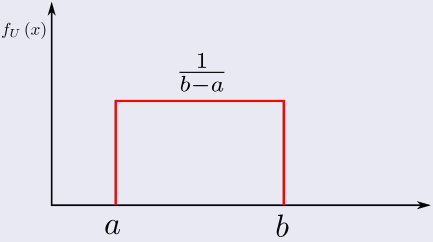
For Example

Examples

Random variable X has uniform $U(a,b)$ distribution if its density is given by

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & else \end{cases}$$

For Example



Example

Bernoulli Distribution

Random variable X has Bernoulli $\mathcal{B}er(p)$ distribution with parameter $0 \leq p \leq 1$

if its probability mass function is given by

$$f(x|p) = p^x (1-p)^{1-x}, \; x \in \{0, 1\}$$

What is the structure of the distribution

Any idea?

Cinvestav

Example

Bernoulli Distribution

Random variable X has Bernoulli $\mathcal{B}er(p)$ distribution with parameter $0 \leq p \leq 1$

if its probability mass function is given by

$$f(x|p) = p^{x}(1-p)^{1-x}, \; x \in \{0,1\}$$

What is the structure of the distribution

Any idea?

Example

Bernoulli Distribution

Random variable X has Bernoulli $\mathcal{B}er(p)$ distribution with parameter $0 \leq p \leq 1$

if its probability mass function is given by

$$f(x|p) = p^x (1-p)^{1-x}, \ x \in \{0, 1\}$$

What is the structure of the distribution

Any idea?

Basic Properties

As you can imagine

They need to follow the rules of a probability.

- $\sum_x f(x) = 1$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Basic Properties

As you can imagine

They need to follow the rules of a probability.

The Probability sums to one

For the PMF and PDF

- $\sum_x f(x) = 1$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Basic Properties

As you can imagine

They need to follow the rules of a probability.

The Probability sums to one

For the PMF and PDF

- $\sum_x f(x) = 1$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Basic Properties

As you can imagine

They need to follow the rules of a probability.

The Probability sums to one

For the PMF and PDF

- $\sum_x f(x) = 1$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

The Probability

It can be “easily” calculated

- One of my ironies.

PMF

$$F_X(a < X < b) = \sum_{k=a}^{b} f_X(k).$$

PDF

$$F_X(a < X < b) = \int_a^b f_X(t)dt$$

Cinvestav

The Probability

It can be "easily" calculated

- One of my ironies.

PMF

$$F_X(a < X < b) = \sum_{k=a}^{b} f_X(k).$$

PDF

$$F_X(a < X < b) = \int_a^b f_X(t)dt$$

The Probability

It can be "easily" calculated

- One of my ironies.

PMF

$$F_X(a < X < b) = \sum_{k=a}^{b} f_X(k).$$

PDF

$$F_X(a < X < b) = \int_a^b f_X(t)dt$$

Cinvestav

In the Continuous Case

We have

$$F_X(a < X < b) = F_X(b) - F_X(a)$$

Additionally, we have that for a single point

$$F_X(a < X < a) = F_X(a) - F_X(a) = 0$$

In the Continuous Case

We have

$$F_X(a < X < b) = F_X(b) - F_X(a)$$

Additionally, we have that for a single point

$$F_X(a < X < a) = F_X(a) - F_X(a) = 0$$

Outline

Now

We have some basic ideas about the descriptions of the Random Variables

We need to be more formal to connect our basic intuitions on continuous spaces.

- There is a unique probability measure P defined in the Borel Subsets of $\mathbb{R}$
- Such That

$$P(B) = \int_B f(x)dx$$

For all intervals $B = (a, b]$

Now

We have some basic ideas about the descriptions of the Random Variables

We need to be more formal to connect our basic intuitions on continuous spaces.

Theorem

- Let f be a nonnegative real-valued function on $\mathbb{R}$ with $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
- There is a unique probability measure P defined in the Borel Subsets of $\mathbb{R}$.
- Such That

$$P(B) = \int_B f(x)\,dx$$

For all intervals $B = (a, b]$

Now

We have some basic ideas about the descriptions of the Random Variables

We need to be more formal to connect our basic intuitions on continuous spaces.

Theorem

- Let f be a nonnegative real-valued function on $\mathbb{R}$ with $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
- There is a unique probability measure P defined in the Borel Subsets of $\mathbb{R}$.
- Such That

$$P(B) = \int_B f(x)\, dx$$

For all intervals $B = (a, b]$

Now

We have some basic ideas about the descriptions of the Random Variables

We need to be more formal to connect our basic intuitions on continuous spaces.

Theorem

- Let f be a nonnegative real-valued function on $\mathbb{R}$ with $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
- There is a unique probability measure P defined in the Borel Subsets of $\mathbb{R}$.
- Such That

$$P(B) = \int_{B} f(x)\, dx$$

For all intervals $B = (a, b]$

Now

We have some basic ideas about the descriptions of the Random Variables

We need to be more formal to connect our basic intuitions on continuous spaces.

Theorem

- Let f be a nonnegative real-valued function on $\mathbb{R}$ with $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
- There is a unique probability measure P defined in the Borel Subsets of $\mathbb{R}$.
- Such That

$$P(B) = \int_B f(x)\, dx$$

For all intervals $B = (a, b]$

Therefore

Definition

The random variable X is said to be absolutely continuous if and only if there is a non-negative function $f = f_X$ defined over $\mathbb{R}$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Here

f_X is called the Density function of X and F_X is called a Cumulative Density Function (CDF).

Therefore

Definition

The random variable X is said to be absolutely continuous if and only if there is a non-negative function $f = f_X$ defined over $\mathbb{R}$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

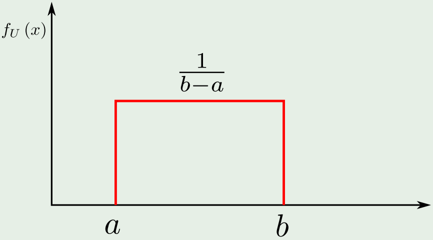
Here

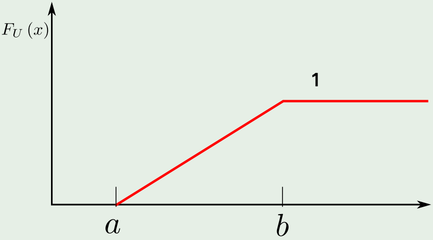
f_X is called the Density function of X and F_X is called a Cumulative Density Function (CDF).

Cinvestav

Graphically

Example uniform distribution





Properties

CDF's Properties

- $F_X(x) \geq 0$
- $F_X(x)$ in a non-decreasing function of X.

Example

- If X is discrete, its CDF can be computed as follows:

$$F_X(x) = P(f(X) \leq x) = \sum_{k=1}^{N} P(X_k = p_k).$$

Properties

CDF's Properties

- $F_X(x) \geq 0$
- $F_X(x)$ in a non-decreasing function of X.

Example

- If X is discrete, its CDF can be computed as follows:

$$F_X(x) = P(f(X) \leq x) = \sum_{k=1}^{N} P(X_k = p_k).$$

Properties

CDF's Properties

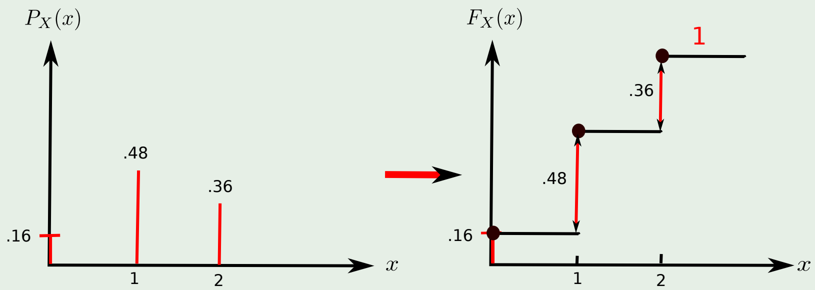
- $F_X(x) \geq 0$
- $F_X(x)$ in a non-decreasing function of X.

Example

- If X is discrete, its CDF can be computed as follows:

$$F_X(x) = P(f(X) \leq x) = \sum_{k=1}^{N} P(X_k = p_k).$$

Example on Discrete Function



Derivative of Cumulative Densitiy Function

Continuous Function

If X is continuous, its CDF can be computed as follows:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Remark

Based in the fundamental theorem of calculus, we have the following equality.

$$f(x) = \frac{dF}{dx}(x)$$

Note

This particular $p(x)$ is known as the Probability Distribution Function (PDF).

Derivative of Cumulative Densitiy Function

Continuous Function

If X is continuous, its CDF can be computed as follows:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Remark

Based in the fundamental theorem of calculus, we have the following equality.

$$f(x) = \frac{dF}{dx}(x)$$

Note

This particular $p(x)$ is known as the Probability Distribution Function (PDF).

Derivative of Cumulative Densitiy Function

Continuous Function

If X is continuous, its CDF can be computed as follows:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Remark

Based in the fundamental theorem of calculus, we have the following equality.

$$f(x) = \frac{dF}{dx}(x)$$

Note

This particular $p(x)$ is known as the Probability Distribution Function (PDF).

Some Basic Properties of These Densities

Conditional PMF/PDF

We have the conditional pdf:

$$p(y|x) = \frac{p(x,y)}{p(x)}.$$

From this, we have the general chain rule

$$p(x_1, x_2, ..., x_n) = p(x_1|x_2, ..., x_n)p(x_2|x_3, ..., x_n)...p(x_n).$$

Independence

If X and Y are independent, then:

$$p(x,y) = p(x)p(y).$$

Some Basic Properties of These Densities

Conditional PMF/PDF

We have the conditional pdf:

$$p(y|x) = \frac{p(x,y)}{p(x)}.$$

From this, we have the general chain rule

$$p(x_1, x_2, ..., x_n) = p(x_1|x_2, ..., x_n)p(x_2|x_3, ..., x_n)...p(x_n).$$

Independence

If X and Y are independent, then:

$$p(x,y) = p(x)p(y).$$

Also the Law of Total Probability

Law of Total Probability is still working correctly

$$p(y) = \sum_{x} p(y|x)p(x).$$

Outline

Cinvestav

We have a common problem

Given a function g

Describing a specific phenomena.

We can have a stochastic input

For example a Random Variable X_1

Then, we have another random variable

$$X_2 = g(X_1)$$

We have a common problem

Given a function g

Describing a specific phenomena.

We can have a stochastic input

For example a Random Variable X_1

Then, we have another random variable

$$X_2 = g(X_1)$$

We have a common problem

Given a function g

Describing a specific phenomena.

We can have a stochastic input

For example a Random Variable X_1

Then, we have another random variable

$$X_2 = g\left(X_1\right)$$

Example

Let X_1 a random variable such that $X_2 = X_1^2$

What is the density function of X_2?

For this, we need to express the event $\{X_2 \leq y\}$

In terms of the random variable X_1

First $X_2 \geq 0$

Thus, we have that for $y < 0$

$$F_2(y) = F(X_2 \leq y) = 0$$

Example

Let X_1 a random variable such that $X_2 = X_1^2$

What is the density function of X_2?

For this, we need to express the event $\{X_2 \leq y\}$

In terms of the random variable X_1

First $X_2 \geq 0$

Thus, we have that for $y < 0$

$$F_2(y) = F(X_2 \leq y) = 0$$

Example

Let X_1 a random variable such that $X_2 = X_1^2$

What is the density function of X_2?

For this, we need to express the event $\{X_2 \leq y\}$

In terms of the random variable X_1

First $X_2 \geq 0$

Thus, we have that for $y < 0$

$$F_2(y) = F(X_2 \leq y) = 0$$

Then

if $y \geq 0$ then $R_2 \leq y$

If and only if $-\sqrt{y} \leq X_1 \leq \sqrt{y}$

Then

$$F\left(X_2 \leq y\right) = F\left(-\sqrt{y} \leq X_1 \leq \sqrt{y}\right) = \int_{-\sqrt{y}}^{\sqrt{y}} f_1\left(x\right) dx$$

If

$$f_1\left(x\right) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{2} & \text{if } -1 \leq x < 0 \\ \frac{1}{2}\exp\left\{-x\right\} & \text{if } 0 \leq x \end{cases}$$

Then

if $y \geq 0$ then $R_2 \leq y$

If and only if $-\sqrt{y} \leq X_1 \leq \sqrt{y}$

Then

$$F\left(X_2 \leq y\right) = F\left(-\sqrt{y} \leq X_1 \leq \sqrt{y}\right) = \int_{-\sqrt{y}}^{\sqrt{y}} f_1\left(x\right) dx$$

If

$$f_1\left(x\right) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{2} & \text{if } -1 \leq x < 0 \\ \frac{1}{2}\exp\left\{-x\right\} & \text{if } 0 \leq x \end{cases}$$

Then

if $y \geq 0$ then $R_2 \leq y$

If and only if $-\sqrt{y} \leq X_1 \leq \sqrt{y}$

Then

$$F\left(X_2 \leq y\right) = F\left(-\sqrt{y} \leq X_1 \leq \sqrt{y}\right) = \int_{-\sqrt{y}}^{\sqrt{y}} f_1\left(x\right) dx$$

If

$$f_1\left(x\right) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{2} & \text{if } -1 \leq x < 0 \\ \frac{1}{2}\exp\left\{-x\right\} & \text{if } 0 \leq x \end{cases}$$

Cinvestav

We have then

if $0 \leq y \leq 1$

$$
\begin{aligned}
F_2(y) &= \int_{-\sqrt{y}}^{\sqrt{y}} f_1(x)\, dx \\
&= \int_{-\sqrt{y}}^{0} \frac{1}{2} dx + \int_{0}^{\sqrt{y}} \frac{1}{2} \exp\{-x\}\, dx \\
&= \frac{1}{2}\sqrt{y} + \frac{1}{2}\left(1 - \exp\{-\sqrt{y}\}\right)
\end{aligned}
$$

We have then

if $0 \leq y \leq 1$

$$
\begin{aligned}
F_2(y) &= \int_{-\sqrt{y}}^{\sqrt{y}} f_1(x)\,dx \\
&= \int_{-\sqrt{y}}^{0} \frac{1}{2} dx + \int_{0}^{\sqrt{y}} \frac{1}{2} \exp\{-x\}\,dx \\
&= \frac{1}{2}\sqrt{y} + \frac{1}{2}\left(1 - \exp\{-\sqrt{y}\}\right)
\end{aligned}
$$

If $y > 1$

What is $F_2(y)$?

We have then

if $0 \leq y \leq 1$

$$
\begin{aligned}
F_2(y) &= \int_{-\sqrt{y}}^{\sqrt{y}} f_1(x)\, dx \\
&= \int_{-\sqrt{y}}^{0} \frac{1}{2} dx + \int_{0}^{\sqrt{y}} \frac{1}{2} \exp\{-x\}\, dx \\
&= \frac{1}{2}\sqrt{y} + \frac{1}{2}\left(1 - \exp\{-\sqrt{y}\}\right)
\end{aligned}
$$

If $y > 1$

What is $F_2(y)$?

Finally

For $y < 0$

$$f_2(y) = \frac{dF_2(y)}{dy} = 0$$

For $0 < y < 1$

$$f_2(y) = \frac{dF_2(y)}{dy} = \frac{1}{4\sqrt{y}}\left(1 + \exp\left\{-\sqrt{y}\right\}\right)$$

For $y > 1$

$$f_2(y) = \frac{dF_2(y)}{dy} = \frac{1}{4\sqrt{y}}\exp\left\{-\sqrt{y}\right\}$$

Cinvestav

Finally

For $y < 0$

$$f_2(y) = \frac{dF_2(y)}{dy} = 0$$

For $0 < y < 1$

$$f_2(y) = \frac{dF_2(y)}{dy} = \frac{1}{4\sqrt{y}}\left(1 + \exp\left\{-\sqrt{y}\right\}\right)$$

For $y > 1$

$$f_2(y) = \frac{dF_2(y)}{dy} = \frac{1}{4\sqrt{y}}\exp\left\{-\sqrt{y}\right\}$$

Finally

For $y < 0$

$$f_2(y) = \frac{dF_2(y)}{dy} = 0$$

For $0 < y < 1$

$$f_2(y) = \frac{dF_2(y)}{dy} = \frac{1}{4\sqrt{y}}\left(1 + \exp\left\{-\sqrt{y}\right\}\right)$$

For $y > 1$

$$f_2(y) = \frac{dF_2(y)}{dy} = \frac{1}{4\sqrt{y}}\exp\left\{-\sqrt{y}\right\}$$

Outline

The Situation Becomes Interesting

When you take into account two or more variables

Here, we have two random variables that are defined by a density function:

$$f_{X,Y}\left(x,y\right)$$

Therefore

We need to understand how these random variables interact.

The Situation Becomes Interesting

When you take into account two or more variables

Here, we have two random variables that are defined by a density function:

$$f_{X,Y}(x, y)$$

Therefore

We need to understand how these random variables interact.

Joint Distributions

Suppose we have a non-negative function real-valued function f in $\mathbb{R}^2$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left(x,y\right)dxdy = 1$$

Now, if we define

$X_1\left(x,y\right)$ and $X_2\left(x,y\right)$, then

$$P\left(\left(X_1,X_2\right)\in B\right) = P\left(B\right) = \int\int_{B} f\left(x,y\right)dxdy$$

Joint Distributions

Suppose we have a non-negative function real-valued function f in $\mathbb{R}^2$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left(x,y\right)dxdy = 1$$

Now, if we define

$X_1\left(x,y\right)$ and $X_2\left(x,y\right)$, then

$$P\left(\left(X_1, X_2\right) \in B\right) = P\left(B\right) = \int\int_B f\left(x,y\right)dxdy$$

Therefore

The Joint Distribution Function is defined as

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du dv$$

Example

Let

$$f(x,y)=\begin{cases}1 & \text{if } 0\leq x\leq 1 \text{ and } 0\leq y\leq 1\\ 0 & elsewhere\end{cases}$$

It looks like

The Unit Square in $\mathbb{R}^2$

Example

Let

$$f(x,y) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & elsewhere \end{cases}$$

It looks like

The Unit Square in $\mathbb{R}^2$

Then

Assume the following random variables

$X_1(x,y) = x$ and $X_1(x,y) = y$.

Why don't we calculate the following probability? For

$$\frac{1}{2} \leq X_1 + X_2 \leq \frac{3}{2}$$

Therefore

$$\frac{1}{2} \leq x + y \leq \frac{3}{2}$$

Then

Assume the following random variables

$X_1\left(x,y\right)=x$ and $X_1\left(x,y\right)=y$.

Why don't we calculate the following probability? For

$$\frac{1}{2}\leq X_1+X_2\leq\frac{3}{2}$$

Therefore

$$\frac{1}{2}\leq x+y\leq\frac{3}{2}$$

Then

Assume the following random variables

$X_1(x,y) = x$ and $X_1(x,y) = y$.

Why don't we calculate the following probability? For

$$\frac{1}{2} \leq X_1 + X_2 \leq \frac{3}{2}$$

Therefore

$$\frac{1}{2} \leq x + y \leq \frac{3}{2}$$

Cinvestav

Look

We have the following

$$P\left\{\frac{1}{2} \leq x + y \leq \frac{3}{2}\right\} = \int\int_{B} 1 dx dy$$

What is B?

We can draw it!!!

Therefore

$$P\left\{\frac{1}{2} \leq x + y \leq \frac{3}{2}\right\} = 1 - 2\left(\frac{1}{8}\right)$$

Look

We have the following

$$P\left\{\frac{1}{2} \leq x + y \leq \frac{3}{2}\right\} = \int\int_B 1dxdy$$

What is B?

We can draw it!!!

Therefore

$$P\left\{\frac{1}{2} \leq x + y \leq \frac{3}{2}\right\} = 1 - 2\left(\frac{1}{8}\right)$$

Look

We have the following

$$P\left\{\frac{1}{2} \leq x + y \leq \frac{3}{2}\right\} = \int\int_{B} 1 dx dy$$

What is B?

We can draw it!!!

Therefore

$$P\left\{\frac{1}{2} \leq x + y \leq \frac{3}{2}\right\} = 1 - 2\left(\frac{1}{8}\right)$$

Cinvestav

Outline

Cinvestav

If we have a Joint Distribution

Can we get the Individual Distributions?

Actually, we have that we can integrate one of the variables.

For Example

What if we have the following age-weight distributions

X_1=Weight			
170-160	**2**	**3**	
160-150	**4**	**5**	
	20-25	25-30	X_2=Age

Cinvestav

If we have a Joint Distribution

Can we get the Individual Distributions?

Actually, we have that we can integrate one of the variables.

For Example

What if we have the following age-weight distributions

X_1=Weight			
170-160	**2**	**3**	
160-150	**4**	**5**	
	20-25	25-30	X_2=Age

Therefore

The Joint Distribution for two discrete variables

$$f(x, y) = F(X_1 = x, X_2 = y)$$

Then

$$\{X_1 = x\} = \{X_1 = x, X_2 = y_1\} \cup \{X_1 = x, X_2 = y_2\} \cup \dots$$

Remember

The events are independent!!!

Therefore

The Joint Distribution for two discrete variables

$$f(x, y) = F(X_1 = x, X_2 = y)$$

Then

$$\{X_1 = x\} = \{X_1 = x, X_2 = y_1\} \cup \{X_1 = x, X_2 = y_2\} \cup \ldots$$

Remember

The events are independent!!!

Therefore

The Joint Distribution for two discrete variables

$$f(x, y) = F(X_1 = x, X_2 = y)$$

Then

$$\{X_1 = x\} = \{X_1 = x, X_2 = y_1\} \cup \{X_1 = x, X_2 = y_2\} \cup \ldots$$

Remember

The events are independent!!!

Therefore

We have the marginal distribution for X_1

$$f_1(x) = F(X_1 = x) = \sum_{y} f(x, y)$$

Similarly

$$f_2(y) = F(X_2 = y) = \sum_{x} f(x, y)$$

Therefore

We have the marginal distribution for X_1

$$f_1(x) = F(X_1 = x) = \sum_y f(x, y)$$

Similarly

$$f_2(y) = F(X_2 = y) = \sum_x f(x, y)$$

Therefore

We have

$$F\left(x_0 \leq X_1 \leq x_0 + dx_0\right) \approx f_1\left(x_0\right) dx_0$$

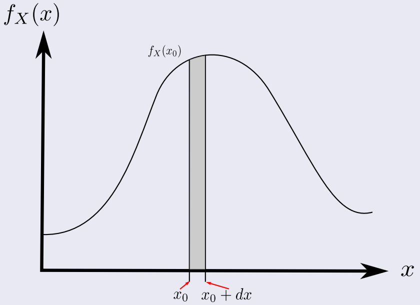
Basically

Therefore

We have

$$F\left(x_0 \leq X_1 \leq x_0 + dx_0\right) \approx f_1\left(x_0\right) dx_0$$

Basically



Then

We have

$$
\begin{aligned}
F\left(x_0 \leq X_1 \leq x_0 + dx_0\right) &= F\left(x_0 \leq X_1 \leq x_0 + dx_0, -\infty < X_2 < \infty\right) \\
&= \int_{x_0}^{x_0+dx_0} dx \int_{-\infty}^{\infty} f\left(x, y\right) dy \\
&\approx dx_0 \int_{-\infty}^{\infty} f\left(x, y\right) dy
\end{aligned}
$$

Therefore

We have if $f\left(x,y\right)$ is well behaved

$$f_1\left(x_0\right)dx_0 \approx dx_0 \int_{-\infty}^{\infty} f\left(x_0,y\right)dy$$

Then

$$f_1\left(x_0\right) \approx \int_{-\infty}^{\infty} f\left(x_0,y\right)dy$$

Therefore

We have if $f(x,y)$ is well behaved

$$f_1(x_0)\,dx_0 \approx dx_0 \int_{-\infty}^{\infty} f(x_0,y)\,dy$$

Then

$$f_1(x_0) \approx \int_{-\infty}^{\infty} f(x_0,y)\,dy$$

In this way

We have

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

Also

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

In this way

We have

$$f_1\left(x\right)=\int_{-\infty}^{\infty} f\left(x,y\right)dy$$

Also

$$f_2\left(y\right)=\int_{-\infty}^{\infty} f\left(x,y\right)dx$$

Cinvestav

Example

Given

$$f(x,y) = \begin{cases} 8xy & 0 \leq y \leq x \leq 1 \\ 0 & elsewhere \end{cases}$$

Then for $0 \leq x \leq 1$

$$f_1(x) = \int_0^x 8xy\,dy = 4x^3$$

If $y < 0$ or $y > 1$

$$f_2(y) = 0$$

Example

Given

$$f(x,y) = \begin{cases} 8xy & 0 \leq y \leq x \leq 1 \\ 0 & elsewhere \end{cases}$$

Then for $0 \leq x \leq 1$

$$f_1(x) = \int_0^x 8xy\,dy = 4x^3$$

If $y < 0$ or $y > 1$

$$f_2(y) = 0$$

Cinvestav

Example

Given

$$f(x,y) = \begin{cases} 8xy & 0 \leq y \leq x \leq 1 \\ 0 & elsewhere \end{cases}$$

Then for $0 \leq x \leq 1$

$$f_1(x) = \int_0^x 8xydy = 4x^3$$

If $y < 0$ or $y > 1$

$$f_2(y) = 0$$

Cinvestav

Therefore

We have for $0 \leq y \leq 1$

$$f_2(y) = \int_y^1 8xy\,dx = 4y\left(1 - y^2\right)$$

Outline

Cinvestav

Expectation

Imagine the following situation

You have the random variables R_1, R_2 representing how long is a call and how much you pay for an international call

$$\text{if } 0 \leq X_1 \leq 3(minute)\ \ X_2 = 10(cents)$$
$$\text{if } 3 < X_1 \leq 6(minute)\ \ X_2 = 20(cents)$$
$$\text{if } 6 < X_1 \leq 9(minute)\ \ X_2 = 30(cents)$$

Expectation

Imagine the following situation

You have the random variables R_1, R_2 representing how long is a call and how much you pay for an international call

$$\text{if } 0 \leq X_1 \leq 3(minute)\ X_2 = 10(cents)$$
$$\text{if } 3 < X_1 \leq 6(minute)\ X_2 = 20(cents)$$
$$\text{if } 6 < X_1 \leq 9(minute)\ X_2 = 30(cents)$$

Cinvestav

Then

We have then the probabilities

$$P\{R_2 = 10\} = 0.6,\ P\{R_2 = 20\} = 0.25,\ P\{R_2 = 10\} = 0.15$$

If we observe N calls and N is very large

We can say that we have $N \times 0.6$ calls and $10 \times N \times 0.6$ the cost of those calls

Then

We have then the probabilities

$$P\{R_2 = 10\} = 0.6,\ P\{R_2 = 20\} = 0.25,\ P\{R_2 = 10\} = 0.15$$

If we observe N calls and N is very large

We can say that we have $N \times 0.6$ calls and $10 \times N \times 0.6$ the cost of those calls

Expectation

Similarly

- $\{R_2 = 20\} \Longrightarrow 0.25N$ and total cost $5N$
- $\{R_2 = 20\} \Longrightarrow 0.15N$ and total cost $4.5N$

Cinvestav

Expectation

Similarly

- $\{R_2 = 20\} \implies 0.25N$ and total cost $5N$
- $\{R_2 = 20\} \implies 0.15N$ and total cost $4.5N$

We have then the probabilities

The total cost is $6N + 5N + 4.5N = 15.5N$ or in average 15.5 cents per call

Expectation

Similarly

- $\{R_2 = 20\} \Longrightarrow 0.25N$ and total cost $5N$
- $\{R_2 = 20\} \Longrightarrow 0.15N$ and total cost $4.5N$

We have then the probabilities

The total cost is $6N + 5N + 4.5N = 15.5N$ or in average 15.5 cents per call

Then

The weighted average

$$
\begin{aligned}
\frac{10\left(0.6N\right)+20\left(.25N\right)+30\left(0.15N\right)}{N} &= 10\left(0.6\right)+20\left(.25\right)+30\left(0.15\right)\\
&= \sum_{y} yP\left\{R_2=y\right\}
\end{aligned}
$$

Then

The Expected Value is a weighted average!!!

Then

The weighted average

$$\frac{10\,(0.6N) + 20\,(.25N) + 30\,(0.15N)}{N} = 10\,(0.6) + 20\,(.25) + 30\,(0.15)$$
$$= \sum_{y} yP\left\{R_2 = y\right\}$$

Then

The Expected Value is a weighted average!!!

Then

John Cage

Assume

Given X a simple random variable i.e. a discrete random variable with a finite range!

We define the expectation of as

$$E(X) = \sum_{x} xP(X = x)$$

Given that you have a simple random variable

The sum is finite and there are not convergence problems.

Cinvestav

Then

John Cage

Assume

Given X a simple random variable i.e. a discrete random variable with a finite range!

We define the expectation of as

$$E(X) = \sum_{x} xP(X = x)$$

Given that you have a simple random variable

The sum is finite and there are not convergence problems.

Then

John Cage

Assume

Given X a simple random variable i.e. a discrete random variable with a finite range!

We define the expectation of as

$$E(X) = \sum_{x} xP(X = x)$$

Given that you have a simple random variable

The sum is finite and there are not convergence problems.

Outline

Cinvestav

Now

This expected function can be extended to random functions too

$$E(X_2) = E(g(X_1)) = \sum_x g(x) f_{X_1}(x)$$

In a similar way, it is possible to define for the continuous random variables

$$E(X_3) = \int_{-\infty}^{\infty} x f_{x_3}(x)\, dx$$

Similarly

$$E(g(X_3)) = \int_{-\infty}^{\infty} g(x) f_{X_3}(x) dx$$

Cinvestav

Now

This expected function can be extended to random functions too

$$E\left(X_2\right) = E\left(g\left(X_1\right)\right) = \sum_{x} g\left(x\right) f_{X_1}\left(x\right)$$

In a similar way, it is possible to define for the continuous random variables

$$E\left(X_3\right) = \int_{-\infty}^{\infty} x f_{x_3}\left(x\right) dx$$

Similarly

$$E\left(g\left(X_3\right)\right) = \int_{-\infty}^{\infty} g(x) f_{X_3}(x) dx$$

Now

This expected function can be extended to random functions too

$$E\left(X_2\right) = E\left(g\left(X_1\right)\right) = \sum_x g\left(x\right) f_{X_1}\left(x\right)$$

In a similar way, it is possible to define for the continuous random variables

$$E\left(X_3\right) = \int_{-\infty}^{\infty} x f_{x_3}\left(x\right) dx$$

Similarly

$$E\left(g\left(X_3\right)\right) = \int_{-\infty}^{\infty} g(x) f_{X_3}(x) dx$$

Cinvestav

Example

Normal Density Function

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

Then

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left\{-\frac{x^2}{2}\right\} dx$$

Then

$$E[X] = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}\right\} d\left\{-\frac{x^2}{2}\right\}$$

Example

Normal Density Function

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

Then

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left\{-\frac{x^2}{2}\right\} dx$$

Then

$$E[X] = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}\right\} d\left\{-\frac{x^2}{2}\right\}$$

Example

Normal Density Function

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

Then

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left\{-\frac{x^2}{2}\right\} dx$$

Then

$$E[X] = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}\right\} d\left\{-\frac{x^2}{2}\right\}$$

Finally

We have

$$E[X] = -\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}\Bigg|_{-\infty}^{\infty} = 0$$

Example

Imagine the following

We have the following functions

1. $f(x) = e^{-x},\ x \geq 0$
2. $g(x) = 0,\ x < 0$

Example

Imagine the following

We have the following functions

1. $f(x) = e^{-x}$, $x \geq 0$
2. $g(x) = 0$, $x < 0$

Find

The expected Value

Example

Imagine the following

We have the following functions

1. $f(x) = e^{-x}$, $x \geq 0$
2. $g(x) = 0$, $x < 0$

Find

The expected Value

Example

Imagine the following

We have the following functions

1. $f(x) = e^{-x}$, $x \geq 0$
2. $g(x) = 0$, $x < 0$

Find

The expected Value

Outline

Then

Given a random variable X, and a, b, c constants

Then, for any functions $g_1(x)$ and $g_2(x)$ whose expectation exists

1. $E[ag_1(x) + bg_2(x) + c] = aE[g_1(x)] + bE[g_2(x)] + c$
2. If $g_1(x) \geq 0$ for all x, then $E[g_1(x)] \geq 0$
3. If $g_1(x) \geq g_2(x)$ for all x, then $E[g_1(x)] \geq E[g_2(x)]$
4. If $a \leq g_1(x) \leq b$ for all, then $a \leq E[g_1(x)] \leq b$

Then

Given a random variable X, and a, b, c constants

Then, for any functions $g_1(x)$ and $g_2(x)$ whose expectation exists

1. $E[ag_1(x) + bg_2(x) + c] = aE[g_1(x)] + bE[g_2(x)] + c$
2. If $g_1(x) \geq 0$ for all x, then $E[g_1(x)] \geq 0$
3. If $g_1(x) \geq g_2(x)$ for all x, then $E[g_1(x)] \geq E[g_2(x)]$
4. If $a \leq g_1(x) \leq b$ for all, then $a \leq E[g_1(x)] \leq b$

Then

Given a random variable X, and a, b, c constants

Then, for any functions $g_1(x)$ and $g_2(x)$ whose expectation exists

1. $E[ag_1(x) + bg_2(x) + c] = aE[g_1(x)] + bE[g_2(x)] + c$
2. If $g_1(x) \geq 0$ for all x, then $E[g_1(x)] \geq 0$
3. If $g_1(x) \geq g_2(x)$ for all x, then $E[g_1(x)] \geq E[g_2(x)]$
4. If $a \leq g_1(x) \leq b$ for all, then $a \leq E[g_1(x)] \leq b$

Then

Given a random variable X, and a, b, c constants

Then, for any functions $g_1(x)$ and $g_2(x)$ whose expectation exists

1. $E\left[ag_1(x) + bg_2(x) + c\right] = aE\left[g_1(x)\right] + bE\left[g_2(x)\right] + c$
2. If $g_1(x) \geq 0$ for all x, then $E\left[g_1(x)\right] \geq 0$
3. If $g_1(x) \geq g_2(x)$ for all x, then $E\left[g_1(x)\right] \geq E\left[g_2(x)\right]$
4. If $a \leq g_1(x) \leq b$ for all, then $a \leq E\left[g_1(x)\right] \leq b$

Then

Given a random variable X, and a, b, c constants

Then, for any functions $g_1(x)$ and $g_2(x)$ whose expectation exists

1. $E[ag_1(x) + bg_2(x) + c] = aE[g_1(x)] + bE[g_2(x)] + c$
2. If $g_1(x) \geq 0$ for all x, then $E[g_1(x)] \geq 0$
3. If $g_1(x) \geq g_2(x)$ for all x, then $E[g_1(x)] \geq E[g_2(x)]$
4. If $a \leq g_1(x) \leq b$ for all, then $a \leq E[g_1(x)] \leq b$

Cinvestav

Outline

Cinvestav

Minimizing Distances

Observation

The expected value of a Random Variable has an important property!!!

One can be seen as

The interpretation of $E[X]$ as a good guess for X

Suppose the following

We measure the distance between a random variable X and a constant b by $(X-b)^2$

- The closer the b is to X, the smaller the quantity is!!!

Minimizing Distances

Observation

The expected value of a Random Variable has an important property!!!

One can be seen as

The interpretation of $E[X]$ as a good guess for X

Suppose the following

We measure the distance between a random variable X and a constant b by $(X-b)^2$

- The closer the b is to X, the smaller the quantity is!!!

Minimizing Distances

Observation

The expected value of a Random Variable has an important property!!!

One can be seen as

The interpretation of $E[X]$ as a good guess for X

Suppose the following

We measure the distance between a random variable X and a constant b by $(X-b)^2$

- The closer the b is to X, the smaller the quantity is!!!

Cinvestav

Then

We can then determine the value of b

$$
\begin{aligned}
E\left(X-b\right)^2 =& E\left(X-EX+EX-b\right)^2 \\
=& E\left(\left(X-EX\right)+\left(EX-b\right)\right)^2 \\
=& E\left(X-EX\right)^2+\left(EX-b\right)^2+\ldots \\
=& 2E\left(\left(X-EX\right)\left(EX-b\right)\right)
\end{aligned}
$$

Then

We can then determine the value of b

$$
\begin{aligned}
E\left(X-b\right)^2 =& E\left(X-EX+EX-b\right)^2 \\
=& E\left(\left(X-EX\right)+\left(EX-b\right)\right)^2 \\
=& E\left(X-EX\right)^2+\left(EX-b\right)^2+\dots \\
=& 2E\left(\left(X-EX\right)\left(EX-b\right)\right)
\end{aligned}
$$

Then

We can then determine the value of b

$$
\begin{aligned}
E\left(X-b\right)^{2} =& E\left(X-EX+EX-b\right)^{2}\\
=& E\left(\left(X-EX\right)+\left(EX-b\right)\right)^{2}\\
=& E\left(X-EX\right)^{2}+\left(EX-b\right)^{2}+...\\
=& 2E\left(\left(X-EX\right)\left(EX-b\right)\right)
\end{aligned}
$$

Cinvestav

Then

We can then determine the value of b

$$
\begin{aligned}
E\left(X-b\right)^{2} =& E\left(X-EX+EX-b\right)^{2}\\
=& E\left(\left(X-EX\right)+\left(EX-b\right)\right)^{2}\\
=& E\left(X-EX\right)^{2}+\left(EX-b\right)^{2}+...\\
=& 2E\left(\left(X-EX\right)\left(EX-b\right)\right)
\end{aligned}
$$

Cinvestav

We notice the following

We have

$$E\left(\left(X-EX\right)\left(EX-b\right)\right)=\left(EX-b\right)E\left(X-EX\right)=0$$

Then

$$E\left(X-b\right)^{2}=E\left(X-EX\right)^{2}+\left(EX-b\right)^{2}$$

What if we choose $b=EX$

$$\min_{b}E\left(X-b\right)^{2}=E\left(X-EX\right)^{2}$$

We notice the following

We have

$$E\left(\left(X - EX\right)\left(EX - b\right)\right) = \left(EX - b\right) E\left(X - EX\right) = 0$$

Then

$$E\left(X - b\right)^2 = E\left(X - EX\right)^2 + \left(EX - b\right)^2$$

What if we choose $b = EX$

$$\min_b E\left(X - b\right)^2 = E\left(X - EX\right)^2$$

We notice the following

We have

$$E\left(\left(X-EX\right)\left(EX-b\right)\right)=\left(EX-b\right)E\left(X-EX\right)=0$$

Then

$$E\left(X-b\right)^{2}=E\left(X-EX\right)^{2}+\left(EX-b\right)^{2}$$

What if we choose $b=EX$

$$\min_{b}E\left(X-b\right)^{2}=E\left(X-EX\right)^{2}$$

Outline

Cinvestav

First, the central moments

Definition

For each integer n, the n^{th} moment of X, $\mu_n^{'}$, is

$$\mu_n^{'} = E\left[X^n\right]$$

The n^{th} central moment of X is

$$\mu_n = E\left[X - \mu\right]^2$$

Where

$$\mu = \mu_n^{'} = EX$$

First, the central moments

Definition

For each integer n, the n^{th} moment of X, $\mu_n^{'}$, is

$$\mu_n^{'} = E\left[X^n\right]$$

The n^{th} central moment of X is

$$\mu_n = E\left[X - \mu\right]^2$$

Where

$$\mu = \mu_1^{'} = EX$$

First, the central moments

Definition

For each integer n, the n^{th} moment of X, $\mu_n^{'}$, is

$$\mu_n^{'} = E\left[X^n\right]$$

The n^{th} central moment of X is

$$\mu_n = E\left[X - \mu\right]^2$$

Where

$$\mu = \mu_n^{'} = EX$$

Cinvestav

Outline

Then

Definition

The Variance of a Random Variable X is its second central moment

$$Var\ X = E\left[X - EX\right]^2$$

Then

- The standard deviation is simply $\sigma = \sqrt{Var(X)}$.

Then

Definition

The Variance of a Random Variable X is its second central moment

$$Var\ X = E\left[X - EX\right]^2$$

Then

- The standard deviation is simply $\sigma = \sqrt{Var(X)}$.

Now

The variance gives a measure of the degree of spread around its mean

Then, we have two cases

A large variance

In such case X is more variable

At the extreme

- If $Var\ X = E\left(X - EX\right)^2 = 0$, then $X = EX$ with probability 1.
  - No Variation!!!

Cinvestav

Now

The variance gives a measure of the degree of spread around its mean

Then, we have two cases

A large variance

In such case X is more variable

At the extreme

- If $Var\ X = E\left(X - EX\right)^2 = 0$, then $X = EX$ with probability 1.
  - No Variation!!!

Now

The variance gives a measure of the degree of spread around its mean

Then, we have two cases

A large variance

In such case X is more variable

At the extreme

- If $Var\ X = E\left(X - EX\right)^2 = 0$, then $X = EX$ with probability 1.
  - No Variation!!!

Example

Exponential Variance

Let X have the exponential(λ) distribution.

Example

Exponential Variance

Let X have the exponential(λ) distribution.

We know that $EX = \lambda$

$$
\begin{aligned}
Var\ X &= E\left(X-\lambda\right)^2\\
&= \int_0^{\infty} (x-\lambda)^2 \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\} dx\\
&= \int_0^{\infty} \left(x^2 - 2x\lambda + \lambda^2\right) \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\} dx
\end{aligned}
$$

Example

Exponential Variance

Let X have the exponential(λ) distribution.

We know that $EX = \lambda$

$$
\begin{aligned}
Var\ X &= E\left(X-\lambda\right)^2 \\
&= \int_0^\infty (x-\lambda)^2 \frac{1}{\lambda}\exp\left\{-\frac{x}{\lambda}\right\} dx \\
&= \int_0^\infty \left(x^2 - 2x\lambda + \lambda^2\right) \frac{1}{\lambda}\exp\left\{-\frac{x}{\lambda}\right\} dx
\end{aligned}
$$

Example

Exponential Variance

Let X have the exponential(λ) distribution.

We know that $EX = \lambda$

$$
\begin{aligned}
Var\ X &= E\left(X-\lambda\right)^2 \\
&= \int_0^{\infty} (x-\lambda)^2 \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\} dx \\
&= \int_0^{\infty} \left(x^2 - 2x\lambda + \lambda^2\right) \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\} dx
\end{aligned}
$$

Cinvestav

Example

Exponential Variance

Let X have the exponential(λ) distribution.

We know that $EX = \lambda$

$$
\begin{aligned}
Var\ X &= E\left(X-\lambda\right)^2 \\
&= \int_0^\infty (x-\lambda)^2 \frac{1}{\lambda}\exp\left\{-\frac{x}{\lambda}\right\} dx \\
&= \int_0^\infty \left(x^2 - 2x\lambda + \lambda^2\right) \frac{1}{\lambda}\exp\left\{-\frac{x}{\lambda}\right\} dx
\end{aligned}
$$

Further

We can use integration by parts to find the variance

$$\int u dv = uv - \int v du$$

Please, try to calculate it

Answer: $Var\ X = \lambda^2$

Further

We can use integration by parts to find the variance

$$\int u dv = uv - \int v du$$

Please, try to calculate it

Answer: $Var\ X = \lambda^2$

About the Possible Linearity

We have

If X is a random variable with finite variance, then for any constants a and b

$$Var\left(aX+b\right)=a^{2}Var\ X$$

Alternative formula for the variance

$$Var\ X=EX^{2}-\left(EX\right)^{2}$$

Proof

At the White Board

About the Possible Linearity

We have

If X is a random variable with finite variance, then for any constants a and b

$$Var\left(aX+b\right)=a^{2}Var\ X$$

Alternative formula for the variance

$$Var\ X=EX^{2}-\left(EX\right)^{2}$$

Proof

At the White Board

About the Possible Linearity

We have

If X is a random variable with finite variance, then for any constants a and b

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

Alternative formula for the variance

$$Var\ X = EX^2 - (EX)^2$$

Proof

At the White Board