Probability

Dicrete Random Variables

Mohamad GHASSANY

EFREI PARIS

Mohamad GHASSANY

- Associate Professor at EFREI Paris, head of Data & Artificial Intelligence Master program.
- Phd in Computer Science Université Paris 13.
- Master 2 in Applied Mathematics & Statistics from Université Grenoble Alpes.
- Personal Website: mghassany.com

Plan

Introduction to probability theory

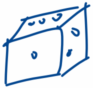
Measurable space

Randomness (Uncertainty)

hasard

Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.



Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ Action.

Random experience

experiment

Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ `Action`.
- Sample space: the set of all possible results of this random experiment

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

finite

countable

Omega ← set

$\Omega = \{$

Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ `Action`.
- Sample space: the set of all possible results of this random experiment

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

- Events: In this random experiment, one can be interested in more complex events than just a simple result of the experiment.

Randomness (Uncertainty)

$\Omega \subset \Omega$

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ `Action`.
- Sample space: the set of all possible results of this random experiment

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

subset: $\{1,2\} \subset \Omega$

element: $\{1,2\} \in \mathcal{P}(\Omega)$

- Events: In this random experiment, one can be interested in more complex events than just a simple result of the experiment.
- The The Power set of Ω, called $\mathcal{P}(\Omega)$, is the set of all subsets of Ω.

$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, ---, \{1,2\}, \{2,3\}, \_\_\_\_, \{3,4,5\}, \_\_\_\_, \Omega\}$$

$\emptyset$

$\{1\}$

$\{2,3\}$

$\{2,4,6\}$

Ω

Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ `Action`.
- Sample space: the set of all possible results of this random experiment

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

- Events: In this random experiment, one can be interested in more complex events than just a simple result of the experiment.
- The The Power set Ω, called $\mathcal{P}(\Omega)$, is the set of all subsets of Ω.
- A family of subsets $\mathcal{A}$ of Ω. These subsets are called events. We say that the event A has occured if and only if the result ω of Ω that has occurred belongs to A.

$\mathcal{A} \subseteq \mathcal{P}(\Omega)$

Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ `Action`.
- Sample space: the set of all possible results of this random experiment

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

- Events: In this random experiment, one can be interested in more complex events than just a simple result of the experiment.
- The The Power set Ω, called $\mathcal{P}(\Omega)$, is the set of all subsets of Ω.
- A family of subsets $\mathcal{A}$ of Ω. These subsets are called events. We say that the event A has occured if and only if the result ω of Ω that has occurred belongs to A.
- **σ-Algebra**: We call σ-Algebra any family $\mathcal{A}$ of subsets of Ω satisfying:
  1. $\Omega \in \mathcal{A}$.
  2. if $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$.
  3. if $(A_n)_{n\in\mathbb{N}}$ is a sequence of elements in $\mathcal{A}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

$A = \{1,2\} \in \mathcal{P}(\Omega)$

$\bar{A} = \{3,4,5,6\} \in \mathcal{P}(\Omega)$

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

$A \cup B = \{1,2,4,6\} \in \mathcal{P}(\Omega)$

Tribu

Randomness (Uncertainty)

Fundamental example: consider the game of a die throw.

- Fundamental example ε : "throw a balanced die" $\longleftarrow$ `Action`.
- Sample space: the set of all possible results of this random experiment

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

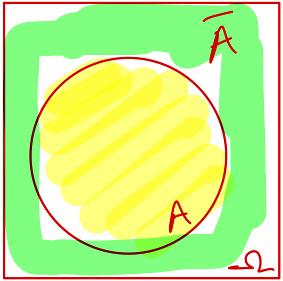
- Events: In this random experiment, one can be interested in more complex events than just a simple result of the experiment.
- The The Power set Ω, called $\mathcal{P}(\Omega)$, is the set of all subsets of Ω.
- A family of subsets $\mathcal{A}$ of Ω. These subsets are called events. We say that the event A has occured if and only if the result ω of Ω that has occurred belongs to A.
- **σ-Algebra**: We call σ-Algebra any family $\mathcal{A}$ of subsets of Ω satisfying:
  1. $\Omega \in \mathcal{A}$.
  2. if $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$.
  3. if $(A_n)_{n\in\mathbb{N}}$ is a sequence of elements in $\mathcal{A}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.
- $(\Omega, \mathcal{A})$ is a measurable space (or a Borel space).

Notions on Events

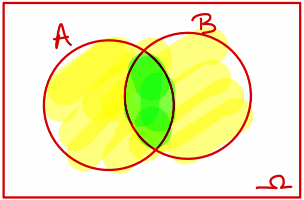
Event $\Longleftrightarrow$ (Subset of Ω) (element of $\mathcal{A}$)

- Let $(\Omega, \mathcal{A})$ be a measurable space:
  - The set $\mathcal{A}$ is called σ-Algebra of events. The elements of $\mathcal{A}$ are called the events.
  - The event Ω is called certain event. The event $\emptyset$ is called impossible event.

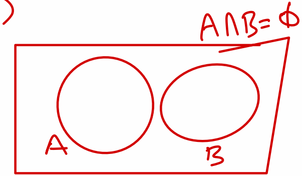
- Let $(\Omega, \mathcal{A})$ be a measurable space:
  - The set $\mathcal{A}$ is called σ-Algebra of events. The elements of $\mathcal{A}$ are called the events.
  - The event Ω is called certain event. The event $\emptyset$ is called impossible event.
- Operations on events. Let A and B be two events:
  - $\bar{A}$ is the complement event of A (we also note A^c). barA=Om $\setminus$ A.
    barA occurs if and only if A does not occur.
  - $A \cap B$ is the event «A and B».
    $A \cap B$ occurs when both events occur.
  - $A \cup B$ is the event «A or B».
    $A \cup B$ occurs when at least one of the two events occurs.



- Let $(\Omega, \mathcal{A})$ be a measurable space:
  - The set $\mathcal{A}$ is called σ-Algebra of events. The elements of $\mathcal{A}$ are called the events.
  - The event Ω is called certain event. The event $\emptyset$ is called impossible event.

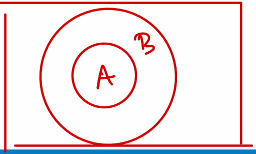


- Operations on events. Let A and B be two events:
  - $\bar{A}$ is the complement event of A (we also note A^c). $barA{=}Om \setminus A$.
    $barA$ occurs if and only if A does not occur.
  - $A \cap B$ is the event «A and B».
    $A \cap B$ occurs when both events occur.
  - $A \cup B$ is the event «A or B».
    $A \cup B$ occurs when at least one of the two events occurs.
- Mutually Exclusive Events: A and B are mutually exclusive if their simultaneous realization is impossible: $A \cap B = \emptyset$. → incompatible events (disjoint events)



- Implication: A implies B means that if A occurs, then B also occurs: $A \subset B$.

$A \Rightarrow B$



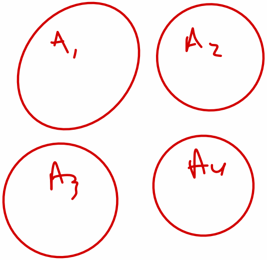
- Let $(\Omega, \mathcal{A})$ a measurable space. A probability function on $(\Omega, \mathcal{A})$, is any map

$$P : \mathcal{A} \to \mathbb{R}$$

such that:

1. $\forall A \in \mathcal{A}, P(A) \geqslant 0$.
2. $P(\Omega) = 1$.
3. $\forall (A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$, a familty of pairwise disjoint (mutually exclusive) events, we have:

$$P\Big(\bigcup_{n \in \mathbb{N}^*} A_n\Big) = \sum_{n=1}^{+\infty} P(A_n)$$



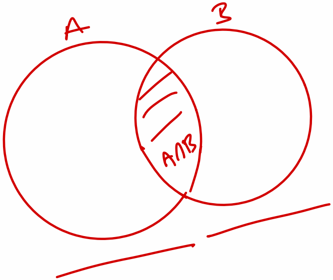
- The triplet $(\Omega, \mathcal{A}, P)$ is called a **probability space**.

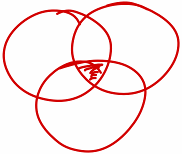
Probability: Properties

1. $P(\emptyset) = 0$.
2. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.
3. If A_1 and A_2 are mutually exclusive, $A_1 \cap A_2 = \emptyset$, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.
4. $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$.
5. $P(\bar{A}) = 1 - P(A)$.
6. $P(B \setminus A) = P(B) - P(B \cap A)$.
7. $A \subset B \Rightarrow P(A) \leqslant P(B)$.

$P(\Omega) = 1$

$P(\bar{A}) = P(\Omega \setminus A) = P(\Omega) - P(A) = 1 - P(A)$





Probability: Properties

1. $P(\emptyset) = 0$.
2. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.
3. If A_1 and A_2 are mutually exclusive, $A_1 \cap A_2 = \emptyset$, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.
4. $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$.
5. $P(\bar{A}) = 1 - P(A)$.
6. $P(B \setminus A) = P(B) - P(B \cap A)$.
7. $A \subset B \Rightarrow P(A) \leqslant P(B)$.

Uniform probability on finite Ω

- Let Ω be a finite sample space. We say that P is the **uniform probability** on the measurable space $(\Omega, P(\Omega))$ if:

$$\forall \omega, \omega' \in \Omega, \qquad P(\{\omega\}) = P(\{\omega'\})$$

One also says that there is **equiprobability** of elementary events.

- Let $(\Omega, \mathcal{P}(\Omega), P)$ be a finite probability space. If P is the uniform probability, then

$$\forall A \in \mathcal{A}, \qquad P(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}$$

Conditional probability

- Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ such that $P(B) > 0$. The map function P_B defined on $\mathcal{A}$ by:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{A}$$

is a probability function on $(\Omega, \mathcal{A})$; it is called the conditional probability given B. It is the probability of event A occurring given that event B has occurred.

- Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ such that $P(B) > 0$. The map function P_B defined on $\mathcal{A}$ by:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{A}$$

is a probability function on $(\Omega, \mathcal{A})$; it is called the conditional probability given B. It is the probability of event A occurring given that event B has occurred.

- Remark: $(A|B)$ is not an event! We use the notation $P(A|B)$ for simplicity, but the notation $P_B(A)$ is the correct one.

Conditional probability

- Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ such that $P(B) > 0$. The map function P_B defined on $\mathcal{A}$ by:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{A}$$

  is a probability function on $(\Omega, \mathcal{A})$; it is called the conditional probability given B. It is the probability of event A occurring given that event B has occurred.
- Remark: $(A|B)$ is not an event! We use the notation $P(A|B)$ for simplicity, but the notation $P_B(A)$ is the correct one.
- Chain rule:

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

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

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

Conditional probability

- Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ such that $P(B) > 0$. The map function P_B defined on $\mathcal{A}$ by:

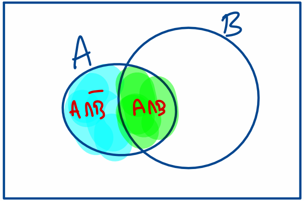
$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{A}$$

  is a probability function on $(\Omega, \mathcal{A})$; it is called the conditional probability given B. It is the probability of event A occurring given that event B has occurred.

- Remark: $(A|B)$ is not an event! We use the notation $P(A|B)$ for simplicity, but the notation $P_B(A)$ is the correct one.

- Chain rule:

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



- Law of total probability:
  - $\forall A \in \mathcal{A}, \quad P(A) = P(A \cap B) + P(A \cap \bar{B})$

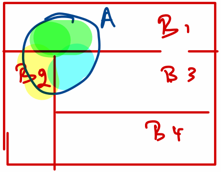
are mutually exclusive

- Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ such that $P(B) > 0$. The map function P_B defined on $\mathcal{A}$ by:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{A}$$

  is a probability function on $(\Omega, \mathcal{A})$; it is called the conditional probability given B. It is the probability of event A occurring given that event B has occurred.
- Remark: $(A|B)$ is not an event! We use the notation $P(A|B)$ for simplicity, but the notation $P_B(A)$ is the correct one.
- Chain rule:

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



- Law of total probability:
  - $\forall A \in \mathcal{A}, \quad P(A) = P(A \cap B) + P(A \cap \bar{B})$
  - We call partition of Ω, a set of events that are pairwise disjoint and whose union is the sample space Ω. The partition is said to be "countable" if its cardinality is at most equal to that of $\mathbb{N}$.
  - Let $(B_n)_{n \geqslant 0}$ a partition of Ω. We have:

$$\forall A \in \mathcal{A}, \qquad P(A) = \sum_{n \geqslant 0} P(A \cap B_n)$$

Conditional probability

- Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ such that $P(B) > 0$. The map function P_B defined on $\mathcal{A}$ by:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{A}$$

A+B are indep $\Leftrightarrow$ $P(A/B) = P(A)$

  is a probability function on $(\Omega, \mathcal{A})$; it is called the conditional probability given B. It is the probability of event A occurring given that event B has occurred.
- Remark: $(A|B)$ is not an event! We use the notation $P(A|B)$ for simplicity, but the notation $P_B(A)$ is the correct one.
- Chain rule:

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

- Law of total probability:
  - $\forall A \in \mathcal{A}, \quad P(A) = P(A \cap B) + P(A \cap \bar{B})$
  - We call partition of Ω, a set of events that are pairwise disjoint and whose union is the sample space Ω. The partition is said to be "countable" if its cardinality is at most equal to that of $\mathbb{N}$.
  - Let $(B_n)_{n \geqslant 0}$ a partition of Ω. We have:

$$\forall A \in \mathcal{A}, \qquad P(A) = \sum_{n \geqslant 0} P(A \cap B_n)$$

- Independence: Events A and B are independent iff $P(A \cap B) = P(A)P(B)$. (or $P(B/A) = P(B)$)

First Bayes' theorem

Let $(\Omega, \mathcal{A}, P)$ a probability space. For all events A and B such that $P(A) \neq 0$ and $P(B) \neq 0$, we have:

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

Second Bayes' theorem

Let $(\Omega, \mathcal{A}, P)$ a probability space and $(B_n)_{n\geqslant 0}$ a partition of Ω s.t. for all $n \geqslant 0$ $P(B_n) \neq 0$. We have for all $A \in \mathcal{A}$ s.t. $P(A) \neq 0$

chain Rule

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{n\geqslant 0} P(A|B_n)P(B_n)} \quad \forall i \geqslant 0$$

total prob. law

Break

9:10 → 9:20

Real Random Variable

SwS 01679

Concept of Real Random Variable

Definition

Let ε an experiment and $(\Omega, \mathcal{A}, \mathrm{P})$ the associated probability space. In many situations, one associates to each result $\omega \in \Omega$ a real number denoted $\mathrm{X}(\omega)$; In this way, one builds a map $\mathrm{X} : \Omega \to \mathbb{R}$. Historically, ε was a game and X représented the earning of a player.

$P : \mathcal{A} \longrightarrow \mathbb{R}$ $[0,1]$

σ-Algebra

$X : \Omega \longrightarrow \mathbb{R}$

also a map

Definition

Let ε *an experiment and* $(\Omega, \mathcal{A}, P)$ *the associated probability space. . In many situations, one associates to each result* $\omega \in \Omega$ *a real number denoted* $X(\omega)$*; In this way, one builds a map* $X : \Omega \to \mathbb{R}$*. Historically,* ε *was a game and* X *représented the earning of a player.*

Example: a die throw

A player throws a fair six faces dice and we observe the obtained number:

- If the result is 1,3 or 5, the player earns 1 euro.
- If the result is 2 or 4, the player earns 5 euros.
- If the result is 6, the player loses 10 euros.

Analysis

- ε: "throw a fair die".

Analysis

- ε: "throw a fair die".
- $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Analysis

- ε: "throw a fair die".
- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{A} = \mathcal{P}(\Omega)$.

Analysis

- ε: "throw a fair die".
- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{A} = \mathcal{P}(\Omega)$.

$(\Omega, \mathcal{A})$ measurable space

- P is the equiprobability on $(\Omega, \mathcal{A})$.

$(\Omega, \mathcal{A}, \mathbb{P})$ is the probability space

Analysis

- ε: "throw a fair die".
- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{A} = \mathcal{P}(\Omega)$.
- P is the equiprobability on $(\Omega, \mathcal{A})$.

Let X the map function from Ω to $\mathbb{R}$ that associates to each result the corresponding earning.

Analysis

- ε: "throw a fair die".
- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{A} = \mathcal{P}(\Omega)$.
- P is the equiprobability on $(\Omega, \mathcal{A})$.

Let X the map function from Ω to $\mathbb{R}$ that associates to each result the corresponding earning.

So we have

- $X(1) = X(3) = X(5) = 1$ (+1 €)
- $X(2) = X(4) = 5$ (+5 €)
- $X(6) = -10$ (−10 €)

We say that X is a **random variable** on Ω.

One can ask what is the probability for the player to win 1 euro:

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow X(\omega) = 1.$

One can ask what is the probability for the player to win 1 euro:

$P(X=1)$

$\Rightarrow X(\omega) = 1.$

- this is the case if and only if $\omega \in \{1, 3, 5\}$.

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\}$$

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow X(\omega) = 1.$

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\}$$

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\}$$

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\})$$

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

$P(X=1) = P(\text{"obtain 1 or 3 or 5"}) = P(\{1,3,5\}) = \frac{|\{1,3,5\}|}{|\Omega|} = 3/6 = 1/2.$

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\}) = \{1, 3, 5\}.$$

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\}) = \{1, 3, 5\}.$$

Similarly, we have:

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow$ $X(\omega) = 1$.

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\}) = \{1, 3, 5\}.$$

Similarly, we have:

- $P(X = 5) = 1/3$.

Concept of Real Random Variable

One can ask what is the probability for the player to win 1 euro:

$\Rightarrow X(\omega) = 1.$

- this is the case if and only if $\omega \in \{1, 3, 5\}$.
- The sought-for probability is therefore $P(\{1, 3, 5\}) = 1/2$.
- Which can also be written as $P(X = 1) = 1/2$.

Thus, we will consider the event:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\}) = \{1, 3, 5\}.$$

Similarly, we have:

- $P(X = 5) = 1/3$.
- $P(X = -10) = 1/6$.

One can present the three previous probabilities in a table:

One can present the three previous probabilities in a table:

One can present the three previous probabilities in a table:

x_i	-10	1	5

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$			

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$	$1/6$	$1/2$	$1/3$

1/6 3/6 2/6 = 1

Concept of Real Random Variable

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$	$1/6$	$1/2$	$1/3$

This is tantamount of considering a new sample space:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$	1/6	1/2	1/3

This is tantamount of considering a new sample space:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

equipped with the probability P_X defined in the table above. This new probability is called the **probability distribution** of X.

Law of prob of X

Prob law of X

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$	1/6	1/2	1/3

This is tantamount of considering a new sample space:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

equipped with the probability P_X defined in the table above. This new probability is called the **probability distribution** of X.

Notice that

$$P(\bigcup_{x_i \in \Omega_X} \{X = x_i\}) =$$

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$	1/6	1/2	1/3

This is tantamount of considering a new sample space:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

equipped with the probability P_X defined in the table above. This new probability is called the **probability distribution** of X.

Notice that

$$P(\bigcup_{x_i \in \Omega_X} \{X = x_i\}) = \sum_{x_i \in \Omega_X} P(X = x_i) =$$

One can present the three previous probabilities in a table:

x_i	-10	1	5
$p_i = P(X = x_i)$	1/6	1/2	1/3

This is tantamount of considering a new sample space:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

equipped with the probability P_X defined in the table above. This new probability is called the **probability distribution** of X.

Notice that

$$P\left(\bigcup_{x_i \in \Omega_X} \{X = x_i\}\right) = \sum_{x_i \in \Omega_X} P(X = x_i) = 1$$

In this chapter:

- We treat the case where $X(\Omega)$ is countable.
- The random variable in this case is ***discrete***.
- We define its probability law by its probability distribution.
- We will define the two main numerical characteristics of a discrete random variable:
  - Expected value: characteristic of centrality (the *mean*).
  - Variance: characteristic of dispersion.
- We will also define the couples of random variables.

Discrete Random Variables

Definition

*We say that a real random variable X is **discrete** if the set of all possible values that X can take is finite or countable.*

Definition

We say that a real random variable X is ***discrete*** *if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

Definition

*We say that a real random variable X is **discrete** if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

- *The set $X(\Omega)$ of all possible values of X, sorted in ascending order: $X(\Omega) = \{x_1, x_2, \dots, x_i, \dots\}$ with $x_1 \leqslant x_2 \leqslant \dots \leqslant x_i \leqslant \dots$.*

Definition

*We say that a real random variable X is **discrete** if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

- *The set $X(\Omega)$ of all possible values of X, sorted in ascending order: $X(\Omega) = \{x_1, x_2, \dots, x_i, \dots\}$ with $x_1 \leqslant x_2 \leqslant \dots \leqslant x_i \leqslant \dots$.*
- *The probability distribution defined on $X(\Omega)$ by*

$$p_i = p(x_i) = P(X = x_i) \quad \forall\, i = 1, 2, \dots$$

Definition

*We say that a real random variable X is **discrete** if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

- *The set $X(\Omega)$ of all possible values of X, sorted in ascending order: $X(\Omega) = \{x_1, x_2, \ldots, x_i, \ldots\}$ with $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_i \leqslant \ldots$.*
- *The probability distribution defined on $X(\Omega)$ by*

$$p_i = p(x_i) = P(X = x_i) \quad \forall\, i = 1, 2, \ldots$$

Remarks:

- $B \subset \mathbb{R}$, $P(X \in B) = \sum_{i/x_i \in B} p(x_i)$.

Probability law

Definition

*We say that a real random variable X is **discrete** if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

- *The set $X(\Omega)$ of all possible values of X, sorted in ascending order: $X(\Omega) = \{x_1, x_2, \ldots, x_i, \ldots\}$ with $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_i \leqslant \ldots$.*
- *The probability distribution defined on $X(\Omega)$ by*

$$p_i = p(x_i) = P(X = x_i) \quad \forall\, i = 1, 2, \ldots$$

Remarks:

- $B \subset \mathbb{R}$, $P(X \in B) = \sum_{i/x_i \in B} p(x_i)$.
- $P(a < X \leqslant b) = \sum_{i/a < x_i \leqslant b} p(x_i)$.

Definition

We say that a real random variable X is ***discrete*** *if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

- *The set $X(\Omega)$ of all possible values of X, sorted in ascending order: $X(\Omega) = \{x_1, x_2, \ldots, x_i, \ldots\}$ with $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_i \leqslant \ldots$.*
- *The probability distribution defined on $X(\Omega)$ by*

$$p_i = p(x_i) = P(X = x_i) \quad \forall\, i = 1, 2, \ldots$$

Remarks:

- $B \subset \mathbb{R}$, $P(X \in B) = \sum_{i/x_i \in B} p(x_i)$.
- $P(a < X \leqslant b) = \sum_{i/a < x_i \leqslant b} p(x_i)$.
- $p(x_i) \geqslant 0$ and $\sum_{i=1}^{\infty} p(x_i) = 1$.

Probability law

Definition

*We say that a real random variable X is **discrete** if the set of all possible values that X can take is finite or countable.*

If we suppose that the set $X(\Omega)$ of all possible values of X admits a smallest element x_1. Then the discrete real random variable X is completely defined by:

- *The set $X(\Omega)$ of all possible values of X, sorted in ascending order: $X(\Omega) = \{x_1, x_2, \dots, x_i, \dots\}$ with $x_1 \leqslant x_2 \leqslant \dots \leqslant x_i \leqslant \dots$.*
- *The probability distribution defined on $X(\Omega)$ by*

$$p_i = p(x_i) = P(X = x_i) \quad \forall\, i = 1, 2, \dots$$

Remarks:

- $B \subset \mathbb{R}$, $P(X \in B) = \sum_{i/x_i \in B} p(x_i)$.
- $P(a < X \leqslant b) = \sum_{i/a < x_i \leqslant b} p(x_i)$.
- $p(x_i) \geqslant 0$ and $\sum_{i=1}^{\infty} p(x_i) = 1$.
- If the number of possible values of X is small enough, the probability distribution of X is often presented as a table.

$P(X \in [0,6]) = P(X=1) + P(X=5) = 1/3 + 1/2$ in the prev. example

Definition

Given a discrete random variable X, we call cumulative distribution function of X (or simply distribution function), denoted F_X, the function defined by: for any real a,

$F_X : \mathbb{R} \to \mathbb{R}$ ([0,1])

$a \in \mathbb{R}$,

$$F(a) = P(X \leqslant a) = \sum_{i / x_i \leqslant a} P(X = x_i)$$

The value $F_X(a)$ represents the probability that X takes a value smaller or equal to a.

$$F_X(a) = P(X \leq a)$$

$P : \mathcal{A} \to [0,1]$

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

$F_X : \mathbb{R} \to [0,1]$

$F_X(0) = P(X \leq 0) = P(X = -10)$

$F_X(2) = P(X \leq 2) = P(X = -10) + P(X = 1) \ldots$

Definition

Given a discrete random variable X, we call cumulative distribution function of X (or simply distribution function), denoted F_X, the function defined by: for any real a,

$$F(a) = P(X \leqslant a) = \sum_{i/x_i \leqslant a} P(X = x_i)$$

The value $F_X(a)$ represents the probability that X takes a value smaller or equal to a.

Properties

1. It is a staircase function.

Definition

Given a discrete random variable X*, we call cumulative distribution function of* X *(or simply distribution function), denoted* F_X*, the function defined by: for any real a,*

$$F(a) = P(X \leqslant a) = \sum_{i/x_i \leqslant a} P(X = x_i)$$

The value $F_X(a)$ represents the probability that X takes a value smaller or equal to a.

Properties

1. It is a staircase function.
2. $F(a) \leqslant 1$ since it is a probability.

Definition

Given a discrete random variable X, we call cumulative distribution function of X (or simply distribution function), denoted F_X, the function defined by: for any real a,

$$F(a) = P(X \leqslant a) = \sum_{i/x_i \leqslant a} P(X = x_i)$$

The value $F_X(a)$ represents the probability that X takes a value smaller or equal to a.

Properties

1. It is a staircase function.
2. $F(a) \leqslant 1$ since it is a probability.
3. $F(a)$ is continuous from the right.

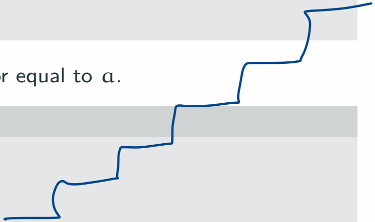
Cumulative distribution function

Definition

Given a discrete random variable X, we call cumulative distribution function of X (or simply distribution function), denoted F_X, the function defined by: for any real a,

$$F(a) = P(X \leqslant a) = \sum_{i/x_i \leqslant a} P(X = x_i)$$

The value $F_X(a)$ represents the probability that X takes a value smaller or equal to a.



Properties

1. It is a staircase function.
2. $F(a) \leqslant 1$ since it is a probability.
3. $F(a)$ is continuous from the right.
4. $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to \infty} F(a) = 1$

Definition

Given a discrete random variable X, we call cumulative distribution function of X (or simply distribution function), denoted F_X, the function defined by: for any real a,

$$F(a) = P(X \leqslant a) = \sum_{i/x_i \leqslant a} P(X = x_i)$$

The value $F_X(a)$ represents the probability that X takes a value smaller or equal to a.

Properties

1. It is a staircase function.
2. $F(a) \leqslant 1$ since it is a probability.
3. $F(a)$ is continuous from the right.
4. $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to \infty} F(a) = 1$

The distribution function characterizes the distribution of X. In other words, if X and Y are two random variables, we have $F_X = F_Y$ if and only if their probability distributions are the same.

All the computations of probabilities about X can be carried out using the distribution function.

All the computations of probabilities about X can be carried out using the distribution function. For example,

$$P(a < X \leqslant b) = F(b) - F(a) \qquad \text{pour tout } a < b$$

All the computations of probabilities about X can be carried out using the distribution function. For example,

$$P(a < X \leqslant b) = F(b) - F(a) \qquad \text{pour tout } a < b$$

This is easier to understand if one writes the event $\{X \leqslant b\}$ as a union of two incompatible events $\{X \leqslant a\}$ and $\{a < X \leqslant b\}$, Let

$$\{X \leqslant b\} = \{X \leqslant a\} \cup \{a < X \leqslant b\}$$

All the computations of probabilities about X can be carried out using the distribution function. For example,

$$P(a < X \leqslant b) = F(b) - F(a) \qquad \text{pour tout } a < b$$

This is easier to understand if one writes the event $\{X \leqslant b\}$ as a union of two incompatible events $\{X \leqslant a\}$ and $\{a < X \leqslant b\}$, Let

$$\{X \leqslant b\} = \{X \leqslant a\} \cup \{a < X \leqslant b\}$$

In this way,

$$P(X \leqslant b) = P(X \leqslant a) + P(a < X \leqslant b)$$

which proves the equality above.

All the computations of probabilities about X can be carried out using the distribution function. For example,

$$P(a < X \leqslant b) = F(b) - F(a) \qquad \text{pour tout } a < b$$

This is easier to understand if one writes the event $\{X \leqslant b\}$ as a union of two incompatible events $\{X \leqslant a\}$ and $\{a < X \leqslant b\}$, Let

$$\{X \leqslant b\} = \{X \leqslant a\} \cup \{a < X \leqslant b\}$$

In this way,

$$P(X \leqslant b) = P(X \leqslant a) + P(a < X \leqslant b)$$

which proves the equality above.

Remark

One can compute the individual probabilities by:

$$p_i = P(X = x_i) = F(x_i) - F(x_{i-1}) \qquad \text{pour } 1 \leqslant i \leqslant n$$

Example

We play three times to “heads or tails”

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{\mathrm{P}, \mathrm{F}\}^3$.

P, F, F
PPP
PPF
⋮

Example

We play three times to "heads or tails" $\Rightarrow$

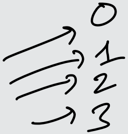
- $\Omega = \{P, F\}^3$.
- $\text{card}(\Omega) = |\Omega| = 2^3 = 8$.

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $\text{card}(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained"



Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $\mathrm{card}(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained" $\Rightarrow X(\Omega) = \{0, 1, 2, 3\}$.

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $\text{card}(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained" $\Rightarrow X(\Omega) = \{0, 1, 2, 3\}$.

- Let's calculate $P(X = 1)$.

T H H
H T H
H H T

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $\mathrm{card}(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained" $\Rightarrow$ $X(\Omega) = \{0, 1, 2, 3\}$.

- Let's calculate $P(X = 1)$.
- $X^{-1}(1) =$

Distribution function and probabilities over X

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $card(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained" $\Rightarrow X(\Omega) = \{0, 1, 2, 3\}$.

- Let's calculate $P(X = 1)$.
- $X^{-1}(1) = \{(P, F, F), (F, P, F), (F, F, P)\}$.

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $card(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained" $\Rightarrow$ $X(\Omega) = \{0, 1, 2, 3\}$.

- Let's calculate $P(X = 1)$.
- $X^{-1}(1) = \{(P, F, F), (F, P, F), (F, F, P)\}$.

$\Rightarrow$ $P(X = 1) = \frac{3}{8}$

Example

We play three times to "heads or tails" $\Rightarrow$

- $\Omega = \{P, F\}^3$.
- $\mathrm{card}(\Omega) = |\Omega| = 2^3 = 8$.

Let X the random variable "number of tails obtained" $\Rightarrow$ $X(\Omega) = \{0, 1, 2, 3\}$.

- Let's calculate $P(X = 1)$.
- $X^{-1}(1) = \{(P, F, F), (F, P, F), (F, F, P)\}$.

$\Rightarrow$ $P(X = 1) = \frac{3}{8}$

Using the same method we obtain the probability distribution of X:

k	0	1	2	3
$P(X = k)$	$1/8$	$3/8$	$3/8$	$1/8$

Distribution function and probabilities over X

The distribution function X is therefore given by:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/8 & \text{si } 0 \leqslant x < 1 \\ 1/2 & \text{si } 1 \leqslant x < 2 \\ 7/8 & \text{si } 2 \leqslant x < 3 \\ 1 & \text{si } x \geqslant 3 \end{cases}$$

The distribution function X is therefore given by:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/8 & \text{si } 0 \leqslant x < 1 \\ 1/2 & \text{si } 1 \leqslant x < 2 \\ 7/8 & \text{si } 2 \leqslant x < 3 \\ 1 & \text{si } x \geqslant 3 \end{cases}$$

One can represent both the probability distribution and the distribution function of X in the same table:

k	0	1	2	3
$P(X = k)$	1/8	3/8	3/8	1/8
$F_X(x)$	1/8	1/2	7/8	1

The graph of the distribution function is represented below:

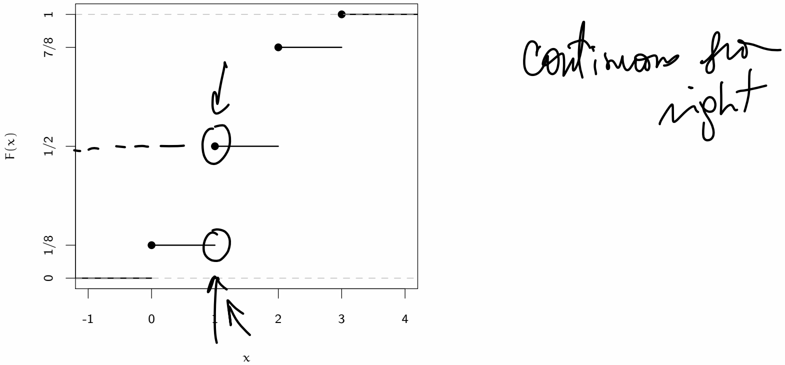


Figure 1: Distribution function

Here is another slightly different representation of the distribution function:

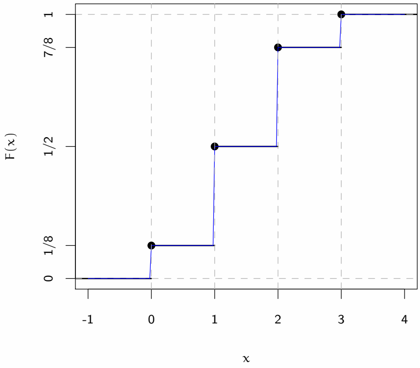


Figure 2: Distribution function

The indicator random variable

Definition

Let A an event. We call indicator random variable of the event A, the random variable $X = \mathbb{1}_A$ defined by:

$$X(\omega) = \begin{cases} 1 & \textit{si } \omega \in A \\ 0 & \textit{si } \omega \in \bar{A} \end{cases}$$

Definition

Let A an event. We call indicator random variable of the event A, the random variable $X = \mathbb{1}_A$ defined by:

$$X(\omega) = \begin{cases} 1 & \textit{si } \omega \in A \\ 0 & \textit{si } \omega \in \bar{A} \end{cases}$$

Therefore:

- $P(X = 1) = P(A) = p$
- $P(X = 0) = P(\bar{A}) = 1 - p$

The indicator random variable

Definition

Let A an event. We call indicator random variable of the event A, the random variable $X = \mathbb{1}_A$ defined by:

$$X(\omega) = \begin{cases} 1 & \textit{si } \omega \in A \\ 0 & \textit{si } \omega \in \bar{A} \end{cases}$$

Therefore:

- $P(X = 1) = P(A) = p$
- $P(X = 0) = P(\bar{A}) = 1 - p$

The distribution function of the indicator random variable is therefore:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - p & \text{si } 0 \leqslant x < 1 \\ 1 & \text{si } x \geqslant 1 \end{cases}$$

Example

- Let $\mathbb{U}$ an urn containing two white ball and three red balls.
- We randomly take one ball out of the box.
- Let A : “take one white ball out”.
- Let X be the indicator random variable of A.

Find the probability distribution and the distribution function of X.

Example

- Let U an urn containing two white ball and three red balls.
- We randomly take one ball out of the box.
- Let A : "take one white ball out".
- Let X be the indicator random variable of A.

Find the probability distribution and the distribution function of X.

The probability distribution of X is

k	0	1
$P(X=k)$	$\frac{3}{5}$	$\frac{2}{5}$

and its distribution function is:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 3/5 & \text{si } 0 \leqslant x < 1 \\ 1 & \text{si } x \geqslant 1 \end{cases}$$

Moments of a discrete random variable

Definition

For a discrete random variable X *with probability distribution* $p(.)$, *we define the expected value of* X, *called* $E(X)$, *by*

$$E(X) = \sum_{i \in \mathbb{N}} x_i p(x_i)$$

Think about Means

Expected value

Definition

For a discrete random variable X *with probability distribution* $p(.)$*, we define the expected value of* X*, called* $E(X)$*, by*

$$E(X) = \sum_{i \in \mathbb{N}} x_i p(x_i)$$

In concrete terms, the expected value of X *is the weighted mean of the values of* X*, the weights being the probabilities associated to the values of* X*.*

Expected value

Definition

For a discrete random variable X with probability distribution $p(.)$, we define the expected value of X, called $E(X)$, by

$$E(X) = \sum_{i \in \mathbb{N}} x_i p(x_i)$$

In concrete terms, the expected value of X is the weighted mean of the values of X, the weights being the probabilities associated to the values of X.

Examples

1. In the previous example where we play three times to "heads or tails", the expected value of X is:

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

2. For the indicator random variable of A:

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = P(A) = p$$

which means that the expected value of the indicator of an event A corresponds to the probability that A occurs.

Expected value of a function of a random variable

Theorem

Let X be a discrete random variable whose possible values are x_i, $i \geqslant 1$, and denote by $p(x_i)$ the probability that $X = x_i$ occurs. Then, for any real function g, we have

Transfer theorem

$$E(g(X)) = \sum_i g(x_i)p(x_i)$$

$$E(X^2) = \sum x_i^2 \, P(X = x_i)$$

Theorem

Let X be a discrete random variable whose possible values are x_i, $i \geqslant 1$, and denote by $p(x_i)$ the probability that $X = x_i$ occurs. Then, for any real function g, we have

$$E(g(X)) = \sum_i g(x_i)p(x_i)$$

Example

Let X be a random variable that can take three values $\{-1, 0, 1\}$ with the following probabilities:

$$P(X = -1) = 0.2 \qquad P(X = 0) = 0.5 \qquad P(X = 1) = 0.3$$

Calculate $E(X^2)$.

Solution

Solution

First method: Let $Y = X^2$.

Second method: Using the theorem

Solution

First method: Let $Y = X^2$. The probability distribution of Y is given by

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.5$$
$$P(Y = 0) = P(X = 0) = 0.5$$

So

$$E(X^2) = E(Y) = 1(0.5) + 0(0.5) = 0.5$$

Second method: Using the theorem

Solution

First method: Let $Y = X^2$. The probability distribution of Y is given by

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.5$$
$$P(Y = 0) = P(X = 0) = 0.5$$

So

$$E(X^2) = E(Y) = 1(0.5) + 0(0.5) = 0.5$$

Second method: Using the theorem

$$E(X^2) = (-1)^2(0.2) + 0^2(0.5) + 1^2(0.3)$$
$$= 1(0.2 + 0.3) + 0(0.5) = 0.5$$

Solution

First method: Let $Y = X^2$. The probability distribution of Y is given by

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.5$$
$$P(Y = 0) = P(X = 0) = 0.5$$

So

$$E(X^2) = E(Y) = 1(0.5) + 0(0.5) = 0.5$$

Second method: Using the theorem

$$E(X^2) = (-1)^2(0.2) + 0^2(0.5) + 1^2(0.3)$$
$$= 1(0.2 + 0.3) + 0(0.5) = 0.5$$

Remark

$$0.5 = E(X^2) \neq (E(X))^2 = 0.01$$

Linearity of the expected value

Properties

1. $E(X+a) = E(X) + a, \quad a \in \mathbb{R}$

Properties

1. $E(X + a) = E(X) + a, \quad a \in \mathbb{R}$

   results which follows from:

$$\sum_i p_i(x_i + a) = \sum_i p_i x_i + \sum_i a p_i = \sum_i p_i x_i + a \sum_i p_i = \sum_i p_i x_i + a$$

Properties

1. $E(X+a) = E(X) + a, \quad a \in \mathbb{R}$

   results which follows from:

$$\sum_i p_i(x_i + a) = \sum_i p_i x_i + \sum_i a p_i = \sum_i p_i x_i + a \sum_i p_i = \sum_i p_i x_i + a$$

2. $E(aX) = aE(X), \quad a \in \mathbb{R}$

Properties

1. $E(X+a) = E(X) + a, \quad a \in \mathbb{R}$

   results which follows from:

$$\sum_i p_i(x_i + a) = \sum_i p_i x_i + \sum_i a p_i = \sum_i p_i x_i + a \sum_i p_i = \sum_i p_i x_i + a$$

2. $E(aX) = aE(X), \quad a \in \mathbb{R}$

   to prove it, just write:

$$\sum_i p_i a x_i = a \sum_i p_i x_i$$

Linearity of the expected value

Properties

1. $E(X+a) = E(X) + a, \quad a \in \mathbb{R}$
   results which follows from:

$$\sum_i p_i(x_i + a) = \sum_i p_i x_i + \sum_i a p_i = \sum_i p_i x_i + a \sum_i p_i = \sum_i p_i x_i + a$$

2. $E(aX) = aE(X), \quad a \in \mathbb{R}$
   to prove it, just write:

$$\sum_i p_i a x_i = a \sum_i p_i x_i$$

3. $E(X+Y) = E(X) + E(Y)$, X and Y being two random variables.

Properties

1. $E(X+a) = E(X) + a, \quad a \in \mathbb{R}$
   results which follows from:

$$\sum_i p_i(x_i + a) = \sum_i p_i x_i + \sum_i a p_i = \sum_i p_i x_i + a \sum_i p_i = \sum_i p_i x_i + a$$

2. $E(aX) = aE(X), \quad a \in \mathbb{R}$
   to prove it, just write:

$$\sum_i p_i a x_i = a \sum_i p_i x_i$$

3. $E(X+Y) = E(X) + E(Y)$, X and Y being two random variables.

All these three properties are summarised in the claim that the expected value is linear:

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y), \quad \forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R}.$$

Definition

Let X be a discrete random variable. We call variance of X, denoted $V(X)$, the quantity defined by, when it exists,

$$V(X) = E\left[(X - E(X))^2\right]$$

Thus, the variance is the expected value of the square of the centered random variable $X - E(X)$. The variance can be interpreted as a measure of the dispersion of the possible values of X around its expected value.

Definition

Let X be a discrete random variable. We call variance of X, denoted $V(X)$, the quantity defined by, when it exists,

$$V(X) = E\left[(X - E(X))^2\right]$$

Thus, the variance is the expected value of the square of the centered random variable $X - E(X)$. The variance can be interpreted as a measure of the dispersion of the possible values of X around its expected value.

Remark

Equivalently, the variance might be defined by the following formula:

$$V(X) = E(X^2) - E^2(X)$$

Definition

Let X be a discrete random variable. We call variance of X, denoted $V(X)$, the quantity defined by, when it exists,

$$V(X) = E\left[(X - E(X))^2\right]$$

Thus, the variance is the expected value of the square of the centered random variable $X - E(X)$. The variance can be interpreted as a measure of the dispersion of the possible values of X around its expected value.

Remark

Equivalently, the variance might be defined by the following formula:

$$V(X) = E(X^2) - E^2(X)$$

Indeed:

$$\begin{aligned} V(X) &= E\left[X^2 - 2XE(X) + E^2(X)\right] \\ &= E(X^2) - E[2XE(X)] + E[E^2(X)] \\ &= E(X^2) - 2E^2(X) + E^2(X) \end{aligned}$$

Example

OLet us compute $V(X)$ in the case where X is the number obtained when throwing a fair die.

Example

OLet us compute $V(X)$ in the case where X is the number obtained when throwing a fair die.

Previously, we saw that $E(X) = \frac{7}{2}$. Moreover,

$$\begin{aligned} E(X^2) &= \sum_i x_i^2 p(x_i) \\ &= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) \\ &= \left(\frac{1}{6}\right)(91) = \frac{91}{6} \end{aligned}$$

Variance

Example

OLet us compute $V(X)$ in the case where X is the number obtained when throwing a fair die.

Previously, we saw that $E(X) = \frac{7}{2}$. Moreover,

$$\begin{aligned} E(X^2) &= \sum_i x_i^2 p(x_i) \\ &= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) \\ &= \left(\frac{1}{6}\right)(91) = \frac{91}{6} \end{aligned}$$

And therefore

$$\begin{aligned} V(X) &= E(X^2) - E^2(X) \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \end{aligned}$$

Properties

1. $V(X) \geqslant 0$

Properties

1. $V(X) \geqslant 0$

2. $\forall\, a \in \mathbb{R}, \quad V(X + a) = V(X)$

Properties

1. $V(X) \geqslant 0$
2. $\forall\, a \in \mathbb{R}, \quad V(X+a) = V(X)$

   en effet:

$$\begin{aligned} V(X+a) &= E\left[[X+a-E(X+a)]^2 \right] \\ &= E\left[[X+a-E(X)-a]^2 \right] \\ &= E\left[[X-E(X)]^2 \right] = V(X) \end{aligned}$$

Properties

1. $V(X) \geqslant 0$
2. $\forall\, a \in \mathbb{R}, \quad V(X+a) = V(X)$

   en effet:

$$\begin{aligned} V(X+a) &= E\left[\left[X+a-E(X+a)\right]^2\right] \\ &= E\left[\left[X+a-E(X)-a\right]^2\right] \\ &= E\left[\left[X-E(X)\right]^2\right] = V(X) \end{aligned}$$

3. $\forall\, a \in \mathbb{R}, \quad V(aX) = a^2 V(X)$

Properties

1. $V(X) \geqslant 0$

2. $\forall\, a \in \mathbb{R}, \quad V(X+a) = V(X)$

   en effet:

$$\begin{aligned} V(X+a) &= E\left[\,[X+a-E(X+a)]^2\,\right] \\ &= E\left[\,[X+a-E(X)-a]^2\,\right] \\ &= E\left[\,[X-E(X)]^2\,\right] = V(X) \end{aligned}$$

3. $\forall\, a \in \mathbb{R}, \quad V(aX) = a^2 V(X)$

   en effet:

$$\begin{aligned} V(aX) &= E\left[\,[aX-E(aX)]^2\,\right] \\ &= E\left[\,[aX-aE(X)]^2\,\right] \\ &= E\left[a^2\,[X-E(X)]^2\,\right] \\ &= a^2\left[E\,[X-E(X)]^2\,\right] = a^2 V(X) \end{aligned}$$

Definition

*Let X be a discrete random variable. The square root of the variance is called the **standard deviation** of X and is denoted*

$$\sigma_X = \sqrt{V(X)}$$

σ_X *has the same physical units as the random variable* X.

Standard deviation

Definition

*Let X be a discrete random variable. The square root of the variance is called the **standard deviation** of X and is denoted*

$$\sigma_X = \sqrt{V(X)}$$

σ_X *has the same physical units as the random variable* X.

- The standard deviation allows to measure the dispersion of a set of data.
- The smaller sigma is, the closer to each other the values of the data are.

Definition

Let X be a discrete random variable. The square root of the variance is called the ***standard deviation*** *of X and is denoted*

$$\sigma_X = \sqrt{V(X)}$$

σ_X *has the same physical units as the random variable X.*

- The standard deviation allows to measure the dispersion of a set of data.
- The smaller sigma is, the closer to each other the values of the data are.
- Example: the dispersion of the grades in an exam. The smaller sigma is, the more homogeneous the class is.

Standard deviation

Definition

Let X be a discrete random variable. The square root of the variance is called the ***standard deviation*** *of X and is denoted*

$$\sigma_X = \sqrt{V(X)}$$

σ_X has the same physical units as the random variable X.

- The standard deviation allows to measure the dispersion of a set of data.
- The smaller sigma is, the closer to each other the values of the data are.
- Example: the dispersion of the grades in an exam. The smaller sigma is, the more homogeneous the class is.
- - Expected value and standard deviation are linked through *Bienaymé-Tchebychev inequality*.

Here

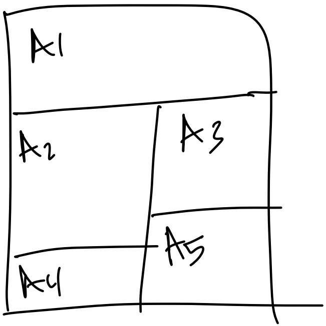
Bienaymé-Tchebychev inequality

Theorem

Let X *a random variable of expected value* μ *and variance* σ^2. *For all* $\varepsilon > 0$, *We have:*

$$P\left(|X - E(X)| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

$$P\left(\bigcup_i A_i\right) = \sum P(A_i)$$



Theorem

Let X a random variable of expected value μ and variance σ^2. For all $\varepsilon > 0$, We have:

$$P\left(|X - E(X)| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

Remark

This inequality can be written in a slightly different fashion. Let $k = \varepsilon/\sigma$.

$$P\left(|X - E(X)| \geqslant k\sigma\right) \leqslant \frac{1}{k^2}$$

Theorem

Let X *a random variable of expected value* μ *and variance* σ^2*. For all* $\varepsilon > 0$*, We have:*

$$P\left(|X - E(X)| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

Remark

This inequality can be written in a slightly different fashion. Let $k = \varepsilon / \sigma$.

$$P\left(|X - E(X)| \geqslant k\sigma\right) \leqslant \frac{1}{k^2}$$

Importance

This inequality relates the probability for X to deviate from its expected value $E(X)$ to its variance, which is precisely an indicator of the dispersion around the expected value. The inequality makes quantitatively precise the statement "the smaller the variance is, the less likely it is to find values far away from the expected value".

Definition

We call non centered moment of order $r \in \mathbb{N}^*$ *of* X *the quantity, when it exists:*

$$m_r(X) = \sum_{i \in \mathbb{N}} x_i^r p(x_i) = E(X^r).$$

Definition

We call non centered moment of order $r \in \mathbb{N}^*$ *of* X *the quantity, when it exists:*

$$m_r(X) = \sum_{i \in \mathbb{N}} x_i^r p(x_i) = E(X^r).$$

Definition

The centered moment of order $r \in \mathbb{N}^*$ *the quantity, when it exists:*

$$\mu_r(X) = \sum_{i \in \mathbb{N}} p_i \left[x_i - E(X)\right]^r = E\left[X - E(X)\right]^r.$$

Definition

We call *non centered moment* *of order* $r \in \mathbb{N}^*$ *of* X *the quantity, when it exists:*

$$m_r(X) = \sum_{i \in \mathbb{N}} x_i^r p(x_i) = E(X^r).$$

Definition

The *centered moment* *of order* $r \in \mathbb{N}^*$ *the quantity, when it exists:*

$$\mu_r(X) = \sum_{i \in \mathbb{N}} p_i \left[x_i - E(X)\right]^r = E\left[X - E(X)\right]^r.$$

Remark

The first moments are:

- $m_1(X) = E(X), \quad \mu_1(X) = 0.$
- $\mu_2(X) = V(X) = m_2(X) - m_1^2(X).$

Two Random Variables

So far, we have dealt with one random variable. However, it is often necessary to consider events related to two variables simultaneously, or even to more than two variables.

So far, we have dealt with one random variable. However, it is often necessary to consider events related to two variables simultaneously, or even to more than two variables.

Definition

Let X *and* Y *two discrete random variables, defined on probability space* $(\Omega, \mathcal{A}, P)$ *and that* $X(\Omega) = \{x_1, x_2, \dots, x_l\}$ *and* $Y(\Omega) = \{y_1, y_2, \dots, y_k\}$, l *and* $k \in \mathbb{N}$.

The ***probability law of*** (X, Y) *is defined by joint probabilities:*

$$p_{ij} = P(X = x_i; Y = y_j) = P(\{X = x_i\} \cap \{Y = y_j\})$$

We have

$$p_{ij} \geqslant 0 \quad \textit{et} \quad \sum_{i=1}^{l} \sum_{j=1}^{k} p_{ij} = 1$$

The pair (X, Y) *is called two dimensional random vector and can have* $l \times k$ *valeurs.*

The probabilities p_{ij} can be presented in a two dimensional table than we call joint probability distribution table:

The probabilities p_{ij} can be presented in a two dimensional table than we call joint probability distribution table:

Table 1: Joint probability distribution table

$X\backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_k
x_1	p_{11}	p_{12}		p_{1j}		p_{1k}
x_2	p_{21}	p_{22}		p_{2j}		p_{2k}
$\vdots$						
x_i	p_{i1}	p_{i2}		p_{ij}		p_{ik}
$\vdots$						
x_l	p_{l1}	p_{l2}		p_{lj}		p_{lk}

In the header we have the possible values of Y and in the first column the possible values of X. The probability $p_{ij} = P(X = x_i; Y = y_j)$ is at the intersection of i^{th} line and j^{th} column.

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

- ε: "draw 3 balls from an urn containing 12 balls".

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

- ε: "draw 3 balls from an urn containing 12 balls".
- $|\Omega| = C_{12}^{3} = 220$.

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

- ε: "draw 3 balls from an urn containing 12 balls".
- $|\Omega| = C_{12}^{3} = 220$.
- $X(\Omega) = \{0, 1, 2, 3\}$ and $Y(\Omega) = \{0, 1, 2, 3\}$.

Example: pair of random variables

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

- ε: "draw 3 balls from an urn containing 12 balls".
- $|\Omega| = C_{12}^3 = 220$.
- $X(\Omega) = \{0, 1, 2, 3\}$ and $Y(\Omega) = \{0, 1, 2, 3\}$.
- $p(X = 0, Y = 0) = p(0, 0) = C_5^3 / C_{12}^3 = \frac{10}{220}$.

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

- ε: "draw 3 balls from an urn containing 12 balls".
- $|\Omega| = C_{12}^{3} = 220$.
- $X(\Omega) = \{0, 1, 2, 3\}$ and $Y(\Omega) = \{0, 1, 2, 3\}$.
- $p(X = 0, Y = 0) = p(0, 0) = C_5^3 / C_{12}^3 = \frac{10}{220}$.
- $p(0, 1) = C_4^1 C_5^2 / C_{12}^3 = \frac{40}{220}$.

Example: pair of random variables

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

- ε: "draw 3 balls from an urn containing 12 balls".
- $|\Omega| = C_{12}^3 = 220$.
- $X(\Omega) = \{0, 1, 2, 3\}$ and $Y(\Omega) = \{0, 1, 2, 3\}$.
- $p(X = 0, Y = 0) = p(0, 0) = C_5^3 / C_{12}^3 = \frac{10}{220}$.
- $p(0, 1) = C_4^1 C_5^2 / C_{12}^3 = \frac{40}{220}$.
- $p(1, 0) = C_3^1 C_5^2 / C_{12}^3 = \frac{30}{220}$.

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

Table 2: Joint probability distribution table

$X\backslash Y$	0	1	2	3
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0
3	$\frac{1}{220}$	0	0	0

When we know the joint distribution of the random variables X and Y, we can also look at the probability distribution of X alone and Y alone. These are the marginal probability distributions.

When we know the joint distribution of the random variables X and Y, we can also look at the probability distribution of X alone and Y alone. These are the marginal probability distributions.

- Marginal distribution of X:

$$p_{i.} = P(X = x_i) = P[\{X = x_i\} \cap \Omega] = \sum_{j=1}^{k} p_{ij} \qquad \forall\, i = 1, 2, \ldots, l$$

When we know the joint distribution of the random variables X and Y, we can also look at the probability distribution of X alone and Y alone. These are the marginal probability distributions.

- Marginal distribution of X:

$$p_{i.} = P(X = x_i) = P[\{X = x_i\} \cap \Omega] = \sum_{j=1}^{k} p_{ij} \qquad \forall\, i = 1, 2, \dots, l$$

- Marginal distribution of Y:

$$p_{.j} = P(Y = y_j) = P[\Omega \cap \{Y = y_j\}] = \sum_{i=1}^{l} p_{ij} \qquad \forall\, j = 1, 2, \dots, k$$

We can calculate the marginal distributions directly from the table of the joint distribution.

Table 3: Joint distribution table with marginal distributions

$X\backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_k	Marginal of X
x_1	p_{11}	p_{12}		p_{1j}		p_{1k}	$p_{1.}$
x_2	p_{21}	p_{22}		p_{2j}		p_{2k}	$p_{2.}$
$\vdots$							
x_i	p_{i1}	p_{i2}		p_{ij}		p_{ik}	$p_{i.}$
$\vdots$							
x_l	p_{l1}	p_{l2}		p_{lj}		p_{lk}	$p_{l.}$
Marginal of Y	$p_{.1}$	$p_{.2}$		$p_{.l}$		$p_{.k}$	1

Example: Marginal Probability Distributions

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Example: Marginal Probability Distributions

Example

Three balls are drawn at random from an urn containing 3 red, 4 white and 5 black balls. X and Y are respectively the number of red and white balls drawn. Determine the joint probability distribution of the pair (X, Y).

Solution

Table 4: Joint distribution table

$X \backslash Y$	0	1	2	3	$p_{i.} = P(X = x_i)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$p_{.j} = P(Y = y_j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

Definition

For each value y_j *of* Y *such that* $p_{.j} = P(Y = y_j) \neq 0$ *we can define the conditional distribution of* X *given* $Y = y_j$ *by*

Definition

For each value y_j *of* Y *such that* $p_{.j} = P(Y = y_j) \neq 0$ *we can define the conditional distribution of* X *given* $Y = y_j$ *by*

$$p_{i/j} = P(X = x_i / Y = y_j) = \frac{P(X = x_i ; Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}} \qquad \forall i = 1, 2, \ldots, l$$

Definition

For each value y_j *of* Y *such that* $p_{.j} = P(Y = y_j) \neq 0$ *we can define the conditional distribution of* X *given* $Y = y_j$ *by*

$$p_{i/j} = P(X = x_i / Y = y_j) = \frac{P(X = x_i ; Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}} \qquad \forall i = 1, 2, \ldots, l$$

Same for Y *given* $X = x_i$*:*

$$p_{j/i} = P(Y = y_j / X = x_i) = \frac{P(X = x_i ; Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}} \qquad \forall j = 1, 2, \ldots, k$$

Definition

We say that two random variables are independent iff

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j) \qquad \forall i = 1, 2, \dots, l \text{ and } j = 1, 2, \dots, k$$

One demonstrates that

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}) \qquad \forall\ A \text{ and } B \in \mathcal{A}$$

Definition

We say that two random variables are independent iff

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j) \qquad \forall i = 1, 2, \ldots, l \text{ and } j = 1, 2, \ldots, k$$

One demonstrates that

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}) \qquad \forall\ A \text{ and } B \in \mathcal{A}$$

Properties

Let two random variables X and Y,

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

Independence of random variables

Definition

We say that two random variables are independent iff

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j) \qquad \forall i = 1, 2, \dots, l \text{ and } j = 1, 2, \dots, k$$

One demonstrates that

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}) \qquad \forall\ A \text{ and } B \in \mathcal{A}$$

Properties

Let two random variables X and Y,

1. $E(X + Y) = E(X) + E(Y)$
2. If X and Y are independent so $E(XY) = E(X)E(Y)$. But the reciprocal is not always true.

Definition

Let two random variables X *and* Y*. The* ***covariance*** *of* X *and* Y*, when it exists, is*

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

that we can calculate using the formula

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Definition

*Let two random variables X and Y. The **covariance** of X and Y, when it exists, is*

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

that we can calculate using the formula

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

Properties

- $Cov(X, Y) = Cov(Y, X)$

Definition

Let two random variables X and Y. The ***covariance*** *of X and Y, when it exists, is*

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

that we can calculate using the formula

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Properties

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX_1 + bX_2, Y) = aCov(X_1, Y) + bCov(X_2, Y)$

Definition

*Let two random variables X and Y. The **covariance** of X and Y, when it exists, is*

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

that we can calculate using the formula

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Properties

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX_1 + bX_2, Y) = aCov(X_1, Y) + bCov(X_2, Y)$
- $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$

Definition

*Let two random variables X and Y. The **covariance** of X and Y, when it exists, is*

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

that we can calculate using the formula

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Properties

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX_1 + bX_2, Y) = aCov(X_1, Y) + bCov(X_2, Y)$
- $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$
- If X and Y are independant so
  - $Cov(X,Y) = 0$ (the reciprocal is not always true)
  - $V(X+Y) = V(X) + V(Y)$ (the reciprocal is not always true)

Definition

The correlation between X and Y is defined by

$$\rho = \rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

We can demonstrate that

$$-1 \leqslant \rho(X, Y) \leqslant 1$$

Definition

The correlation between X and Y is defined by

$$\rho = \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

We can demonstrate that

$$-1 \leqslant \rho(X,Y) \leqslant 1$$

Interpretation of ρ

- The correlation coefficient is a measure of the degree of linearity between X and Y.
- Values of rho close to 1 or -1 indicate an almost rigorous linearity between X and Y.
- Values of rho close to 0 indicate the absence of any linear relationship.
- When $\rho(X,Y)$ is positive, Y tends to increase if X does the same.
- When $\rho(X,Y) < 0$, Y tends to decrease if X increases.

Linear correlation