Probability, Statistics, and Information Theory (SC224)

Madhukant Sharma

DAIICT Gandhinagar

Email: madhukant[underscore]sharma@daiict.ac.in

Team

- **Main Lectures:** Dr. Madhukant Sharma and Dr. Manish Kumar
- **Tutorials**
  - 202221008 Parth Varasani (PhD)
  - 202321014 Shivam Sonawane (PhD)

Reference Books

Text Book

- Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, Elsevier, Fifth Edition 2016.

Reference Books

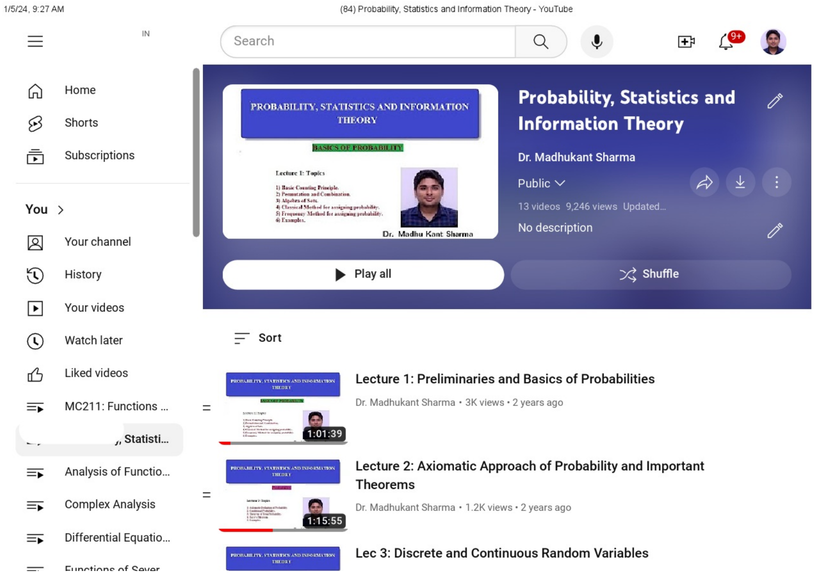
Text Book

- Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, Elsevier, Fifth Edition 2016.

Reference Books

- Sheldon M. Ross, **Introduction to Probability Models: 11th Edition**, Academic Press Elsevier, 2015.
- Mario Lefebvre, **Applied Probability and Statistics**, Springer, 2006.
- Jean Jacod and Philip Protter, **Probability Essentials**, Springer, 2004.
- Hogg, Tanis and Rao, **Probability and Statistical Inference: 7th Edition**, Pearson, 2006.

YouTube Channel



Review of Selected Mathematical Techniques

- Basic Counting Principles.

Review of Selected Mathematical Techniques

- Basic Counting Principles.
- Permutation and Combination.

Review of Selected Mathematical Techniques

- Basic Counting Principles.
- Permutation and Combination.
- Algebra of Sets.

Basic Counting Principle

Example

A small community consists of 12 women, each of whom has 2 children. If one women and one of her children are to be chosen as mother and child of the year, how many different choices are possible?

Basic Counting Principle

Example

A small community consists of 12 women, each of whom has 2 children. If one women and one of her children are to be chosen as mother and child of the year, how many different choices are possible?

Example

Suppose a person has 3 shirts and 5 ties. How many ways he can choose a shirt with a tie?

Basic Counting Principle

Suppose two experiments E_1 and E_2 are to be performed in order/simultaneously, with N_1 possible outcomes for E_1 and N_2 possible outcomes for E_2. Then, together there are

$$N_1 . N_2$$

possible outcomes of two experiments.

Basic Counting Principle

In general, if n experiments $E_1, E_2, ..., E_n$ are to be performed in order/simultaneously, with possible number of outcomes $N_1, N_2, ..., N_n$ respectively. Then, together there are

$$N_1.N_2...N_n$$

possible outcomes of n experiments in given order.

Basic Counting Principle

Example

How many 8 place License plates are possible if first four places are to be occupied by alphabets and last four by numbers?

Repitition is not allowed

$\underbrace{26 \times 25 \times 24 \times 23}$ $\times \underbrace{10 \times 9 \times 8 \times 7}$

Basic Counting Principle

Example

How many 8 place License plates are possible if first four places are to be occupied by alphabets and last four by numbers?

Example

Repetition of letters and numbers is not permitted.

Permutation and Combination

Example

Consider collection of 3 novels by authors A, B and C, 2 mathematics books by authors D and E, and 1 physics book by author P. How many arrangements are possible if the books are to be distinguished by the authors.

3!

$3! \times 3! \times 2! \times 1!$

$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6!$

Permutation and Combination

Example

Consider collection of 3 novels by authors A, B and C, 2 mathematics books by authors D and E, and 1 physics book by author P. How many arrangements are possible if the books are to be distinguished by the authors.

Example

How many arrangements are possible if the books are to be distinguished by the subjects?

Permutation and Combination Contd...

(If the order is important)

Number of arrangements of n objects $= n!$

Permutation and Combination Contd...

(If the order is important)

Number of arrangements of n objects $= n!$

(If the order is not important)

Number of arrangements of n objects where n_1 are alike, n_2 are alike,...,n_r are alike such that $n_1 + n_2 + ... + n_r = n$

$$= \frac{n!}{n_1!n_2!...n_r!}$$

Permutation and Combination Contd...

Example

A tennis tournament has 9 competitions, 3 from India, 2 from Japan and 4 from Australia. Results of the tournament are announced by nationalities of the players in the order in which they are placed. How many such lists are possible?

Permutation and Combination Contd...

In a collection of n objects, in how many way can we select r objects?

Permutation and Combination Contd...

In a collection of n objects, in how many way can we select r objects?

If the order is important:

$$\frac{n!}{(n-r)!} = {}^{n}P_{r}.$$

Permutation and Combination Contd...

In a collection of n objects, in how many way can we select r objects?

If the order is important:

$$\frac{n!}{(n-r)!} = {}^{n}P_{r}.$$

If the order is immaterial:

$$\frac{n!}{r!(n-r)!} = {}^{n}C_{r}.$$

Permutation and Combination Contd...

Example

A jury of 7 is to be formed from a group of 30 people. How many different juries can be formed?

Permutation and Combination Contd...

Example

A jury of 7 is to be formed from a group of 30 people. How many different juries can be formed?

Example

In case the group of 30 people consists of 10 women and 20 men and it is required that 2 women and 5 men should form the jury. How many juries can be forme?

$^{10}C_2 \times {}^{20}C_5$

Algebra of Sets

Definition (Set)

A set is a collection of well defined objects, called elements.

Ω

Algebra of Sets

Definition (Set)

A set is a collection of well defined objects, called elements.

Set Operations: Let Ω is an abstract/universal set.

(i) **Subset:** A set A is called a subset of B, denoted by $A \subset B$ if every element of A is also an element of B.

Algebra of Sets

Definition (Set)

A set is a collection of well defined objects, called elements.

Set Operations: Let Ω is an abstract/universal set.

(i) **Subset:** A set A is called a subset of B, denoted by $A \subset B$ if every element of A is also an element of B.

(ii) **Equality:** Two sets A and B are equal, denoted by $A = B$, iff $A \subset B$ and $B \subset A$.

Algebra of Sets

Definition (Set)

A set is a collection of well defined objects, called elements.

Set Operations: Let Ω is an abstract/universal set.

(i) **Subset:** A set A is called a subset of B, denoted by $A \subset B$ if every element of A is also an element of B.

(ii) **Equality:** Two sets A and B are equal, denoted by $A = B$, iff $A \subset B$ and $B \subset A$.

(iii) **Complementation:** Let $A \subset \Omega$. The complement of set A, denoted by $\overline{A}$ or A^c, is the set containing all elements in Ω but not in A.

Set Operations Contd...

(iv) **Union:** The union of sets A and B, denoted by $A \cup B$, is the set containing all elements in either A or B or both.

Set Operations Contd...

(iv) **Union:** The union of sets A and B, denoted by $A \cup B$, is the set containing all elements in either A or B or both.

(v) **Intersection:** The intersection of sets A and B, denoted by $A \cap B$, is the set containing all elements in both A and B.

Set Operations Contd...

(iv) **Union:** The union of sets A and B, denoted by $A \cup B$, is the set containing all elements in either A or B or both.

(v) **Intersection:** The intersection of sets A and B, denoted by $A \cap B$, is the set containing all elements in both A and B.

(vi) **Null/Empty Set:** The set containing no elements is called the null set, denoted by ϕ.

Set Operations Contd...

(iv) **Union:** The union of sets A and B, denoted by $A \cup B$, is the set containing all elements in either A or B or both.

(v) **Intersection:** The intersection of sets A and B, denoted by $A \cap B$, is the set containing all elements in both A and B.

(vi) **Null/Empty Set:** The set containing no elements is called the null set, denoted by ϕ.

(vii) **Disjoint/Mutually Exclusive Sets:** Two sets A and B are called disjoint or mutually exclusive if $A \cap B = \phi$.

Set Operations Contd...

(viii) **De Morgan's Laws:**

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$.
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

Set Operations Contd...

(viii) **De Morgan's Laws:**

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$.
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

(ix) **Distributive Laws:**

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Random Experiment and Sample Space

Definition (Random Experiment)

A random experiment is an experiment in which:

(a) the set of all possible outcomes of the experiment is known in advance;

(b) the outcome of a particular trial of the experiment can not be predicted in advance;

(c) the experiment can be repeated under identical conditions.

$\Omega = \{ \qquad \qquad \qquad \qquad \}$, $|\Omega|$

$\Omega = \{1, 2, \ldots, 6\}$

Classical method

$E \subseteq \Omega$

$$P(E) = \frac{|E|}{|\Omega|}$$

Relative frequency:

Trials happened N- times

Event 'E' occured $f_N(E)$ - times

$$0 \leq \frac{f_N(E)}{N} \leq 1$$

$$P(E) = \lim_{N \to \infty} \frac{f_N(E)}{N}$$

Probability (Axiomatic Approach)

$$P : 2^{\Omega} \to [0,1]$$

is said to be probability function if it satisfies the following three axioms.

(i) $P(E) \geq 0 \ \forall E \in 2^{\Omega}$ (non-negativity)

(ii) Let $\{E_1, E_2, \ldots \ldots\}$ be a countable collection of mutually exclusive events

* $\Omega \to$ Samplespace
* $E \subseteq \Omega \to E$ is called event
* E_1, E_2 are events. If $E_1 \cap E_2 = \phi$, then E_1 & E_2 are mutually exclusive events.

then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$ (Countable infinite additivity)

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

(ii) $E_1 = E_2 = \cdots = \phi$.

$$\bigcup_{i=1}^{\infty} E_i = \phi$$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

$$\Rightarrow P(\phi) = \sum_{i=1}^{\infty} P(\phi) \Rightarrow \underline{P(\phi) = 0}$$

(i) $A \subseteq B$, $\quad B = (B-A) \cup A$

$$P(B) = P(B-A) + P(A)$$

$$\geq P(A)$$

$$\boxed{A \subseteq B \Rightarrow P(A) \leq P(B)}$$

(ii) $P(A^c) = 1 - P(A)$?

$$\Omega = (\Omega - A) \cup A \Rightarrow P(\Omega) = P(A^c) + P(A)$$

$$\Rightarrow P(A) = 1 - P(A^c)$$

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

\# $\Omega = \{ T, FT, FFT, FFFT, \ldots \}$

$\frac{1}{6} \quad \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \frac{1}{6}, \left(\frac{5}{6}\right)^3 \frac{1}{6}, \quad \cdot$

$\{6\} \to T$

$\{1, \ldots, 5\} \to F$

$$\frac{1}{6} \sum_{i=1}^{\infty} \left(\frac{5}{6}\right)^{i-1} = 1$$

\# $\Omega = \mathbb{N} \quad \Big| \quad i \in \mathbb{N}, \quad P(\{i\}) = \frac{1}{6}\left(\frac{5}{6}\right)^{i-1}$

Random Experiment and Sample Space

Definition (Random Experiment)

A random experiment is an experiment in which:

(a) the set of all possible outcomes of the experiment is known in advance;

(b) the outcome of a particular trial of the experiment can not be predicted in advance;

(c) the experiment can be repeated under identical conditions.

Definition (Sample Space: Ω)

The collection of all possible outcomes of a random experiment is called the sample space.

Random Experiment and Sample Space: Example

Example

Write down the sample space for each of the following random experiments:

(i) Rolling a die.

(ii) Simultaneously flipping a coin and rolling a die.

(iii) Coin is tossed repeatedly until a head is observed.

Events and Mutually Exclusive

Definition (Event)

An event E is a set of outcomes of the random experiment, i.e., $E \subseteq \Omega$.

Events and Mutually Exclusive

Definition (Event)

An event E is a set of outcomes of the random experiment, i.e., $E \subseteq \Omega$.

Definition (Mutually Exclusive)

Two events E_1 and E_2 are said to be mutually exclusive if they can not occur simultaneously, i.e., if $E_1 \cap E_2 = \phi$.

Assigning Probabilities: Classical Method

- It is used for random experiments which result in a finite number of equally likely outcomes.
- For an instance, let $\Omega = \{w_1, w_2, ..., w_n\}$ with $n \in \mathbb{N}$.
- For $E \subseteq \Omega$, let $\mid E \mid$ denotes the number of elements in E.

Assigning Probabilities: Classical Method

- It is used for random experiments which result in a finite number of equally likely outcomes.
- For an instance, let $\Omega = \{w_1, w_2, ..., w_n\}$ with $n \in \mathbb{N}$.
- For $E \subseteq \Omega$, let $\mid E \mid$ denotes the number of elements in E.
- The probability of an event E is given by

$$P(E) = \frac{\text{number of outcomes favourable to E}}{\text{total number of outcomes}} = \frac{\mid E \mid}{\mid \Omega \mid} = \frac{\mid E \mid}{n}.$$

Note: An outcome $w \in \Omega$ is said to be favorable to an event E if $w \in E$.

Example

Example

Suppose that in a classroom we have 25 students (with registration number 1,2,...,25) born in the same year having 365 days. Suppose that we want to find the probability of the event that they all are born on different days of the year. Assume that outcomes are equally likely.

Assigning Probabilities: Relative Frequency Method

- Suppose that we have independent repetitions of a random experiments under identical conditions.

Assigning Probabilities: Relative Frequency Method

- Suppose that we have independent repetitions of a random experiments under identical conditions.
- Let $f_N(E)$ denote the number of times an event E occurs in the first N trials. Then, the corresponding relative frequency is

$$r_N(E) = \frac{f_N(E)}{N}.$$

Assigning Probabilities: Relative Frequency Method

- Suppose that we have independent repetitions of a random experiments under identical conditions.
- Let $f_N(E)$ denote the number of times an event E occurs in the first N trials. Then, the corresponding relative frequency is

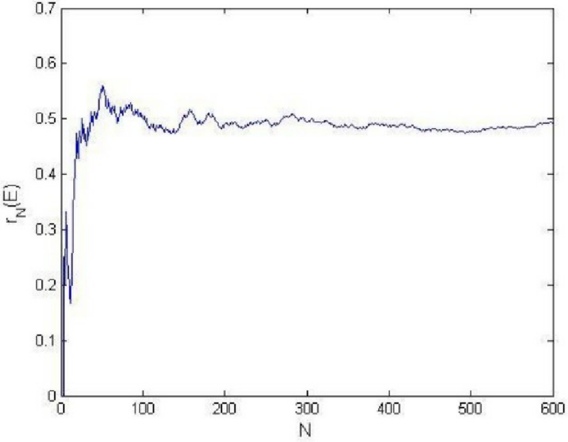
$$r_N(E) = \frac{f_N(E)}{N}.$$

- The probability of an event E is given by

$$P(E) = \lim_{N \to \infty} \frac{f_N(E)}{N}.$$

Example

Plot of relative frequencies of number of heads against number of trials in the random experiment of tossing a fair coin.



Assigning Probabilities: Axiomatic Approach

Definition

Let Ω be a sample space associated with a random experiment. Let $\mathfrak{F}$ be a $\sigma-$algebra of subsets of Ω. A probability measure (or function) is a set function P, defined on $\mathfrak{F}$, satisfying the following three axioms:

(i) $P(E) \geq 0$ for all $E \in \mathfrak{F}$. (Axiom 1: Non-negativity);

(ii) If $E_1, E_2, \ldots$ is a countable infinite collection of mutually exclusive events, then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i); \text{ (Countable infinite additivity)};$$

Assigning Probabilities: Axiomatic Approach

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

Assigning Probabilities: Axiomatic Approach

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

Note the followings:

(1) $\mathfrak{F}$ is also called the event space.

Assigning Probabilities: Axiomatic Approach

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

Note the followings:

(1) $\mathfrak{F}$ is also called the event space.

(2) The members of $\mathfrak{F}$ are called events.

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

Note the followings:

(1) $\mathfrak{F}$ is also called the event space.

(2) The members of $\mathfrak{F}$ are called events.

(3) A countable collection $\{E_i : i \in \Lambda\}$ of events is said to be exhaustive if $P\left(\bigcup_{i \in \Lambda} E_i\right) = 1$

Questions

(1) Is it possible to assign probabilities to all subsets of Ω, when Ω is countable?

Questions

(1) Is it possible to assign probabilities to all subsets of Ω, when Ω is countable?

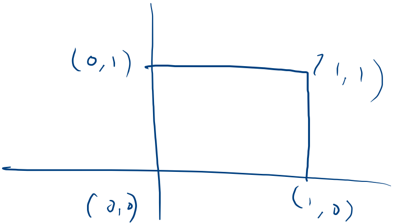
(2) In any probability space $(\Omega, \mathcal{F}, P)$, we have $P(\Omega) = 1$. Whether $P(A) = 1$ implies $A = \Omega$ or not?

$\Omega \quad \underline{\mathbb{N}}$

$P(\{i\}) = \frac{1}{2^i}$

$$P(\mathbb{N} - \{1\}) = P\left(\bigcup_{i=2}^{\infty} \{i\}\right) = \sum_{i=2}^{\infty} \frac{1}{2^i} = \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \cdots$$

$$\frac{\frac{1}{2^2}}{1 - \frac{1}{2}} = \frac{1}{2}$$



$$P(A) = \frac{\text{area of } A}{\text{area of } \square} = \text{area of } A$$

$$A - \{(a,b)\}$$

$$\text{area}\left[\square - \left\{\left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\}\right] = 1$$

Problem

Suppose that five cards are drawn at random and without replacement from a deck of 52 cards. Assuming that the outcomes are equally likely, find the probabilities for the following events:

(1) E_1 : the event that each card is spade. $\longrightarrow$ ${}^{13}C_5$

(2) E_2 : the event that at least one of the drawn cards is spade. , E_2^c : ${}^{39}C_5$

$P(E_2) = 1 - P(E_2^c)$

(3) E_3 : the event that among the drawn cards three are kings and two are queens. ${}^{4}C_3 \times {}^{4}C_2$

(4) E_4 : the event that among the drawn cards two are kings, two are queens and one is Jack. ${}^{4}C_2 \times {}^{4}C_2 \times {}^{4}C_1$

Rolling a dice two-times

$$\Omega = \{ (i,j) : i=1,\dots,6,\ j=1,\dots,6 \}$$

A : The sum is 7. $= \{ (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) \}$

B : The first dice shows 4. $= \{ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \}$.

$$P(B \mid A) = \frac{1}{6}$$

$$P(B) = \frac{1}{6}, \quad P(A) = \frac{1}{6}, \quad P(A \cap B) = \frac{1}{36}$$

$$\frac{P(A\cap B)}{P(A)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} = P(B|A)$$

C : The sum is 9 $= \{(3,6), (4,5), (5,4), (6,3)\}$

$P(B|C) = \frac{1}{4}$

$P(B\cap C) = \frac{1}{36}$

$P(C) = \frac{1}{9}$

$$\frac{P(B\cap C)}{P(C)} = \frac{\frac{1}{36}}{\frac{1}{9}} = \frac{1}{4} = P(B|C)$$

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

& $P(B|C) > P(B)$

Conditional Prob.

$$P(A|B) = \frac{P(A\cap B)}{P(B)}, \quad P(B) > 0$$

$A_i \cap B \subseteq A_i$

A_j

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

$$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right)\cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)}$$

$$= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B)$$

#

$E_1 = \{5,6\}$ $\qquad$ $\{1,2,3,4\} = E_2$

A = The drawn ball is white

$$P(A) = P(E_1)\,P(A|E_1) + P(E_2)\,P(A|E_2) = \frac{2}{6}\times\frac{4}{10} + \frac{4}{6}\times\frac{3}{11}$$

$$P(E_1|A) = \frac{P(E_1\cap A)}{P(A)} = \frac{P(A|E_1)\,P(E_1)}{P(A)} = \frac{\frac{4}{10}\times\frac{2}{6}}{P(A)}$$

$$P(E_2|A) = \frac{P(A|E_2)\,P(E_2)}{P(A)}$$

$E_1, E_2, \ldots$

$\{E_i\}_{i\in\Lambda}$ is s.t. $E_i \cap E_j = \phi \quad \forall\, i \neq j$,

$E_i \neq \phi \ \forall\, i\in\Lambda$ & $\bigcup_{i\in\Lambda} E_i = \Omega$, $P(E_i) > 0 \ \forall\, i\in\Lambda$.

$$P(A) = P(A\cap\Omega) = P\left(A \cap \left(\bigcup_{i\in\Lambda} E_i\right)\right)$$

$$= P\left(\bigcup_{i\in\Lambda} (A\cap E_i)\right) = \sum_{i\in\Lambda} P(A\cap E_i)$$

$$P(A\cap E_i) = P(A|E_i)\,P(E_i)$$

$$P(A) = \sum_{i \in \Lambda} P(A|E_i)\, P(E_i)$$

Conditional Probability

Conditional Probability

Theorem

Let $\mathfrak{F}$ be a $\sigma-$algebra of subsets of Ω and let $B \in \mathfrak{F}$. Define $\mathfrak{F}_B = \{A \cap B : \ A \in \mathfrak{F}\}$. Then $\mathfrak{F}_B$ is a $\sigma-$algebra of subsets of B and $\mathfrak{F}_B \subseteq \mathfrak{F}$.

Conditional Probability

Theorem

Let $\mathfrak{F}$ be a $\sigma-$algebra of subsets of Ω and let $B \in \mathfrak{F}$. Define $\mathfrak{F}_B = \{A \cap B : \ A \in \mathfrak{F}\}$. Then $\mathfrak{F}_B$ is a $\sigma-$algebra of subsets of B and $\mathfrak{F}_B \subseteq \mathfrak{F}$.

Theorem

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let $B \in \mathfrak{F}$ be such that $P(B) > 0$. Then $(B, \mathfrak{F}_B, P_B)$ and $(\Omega, \mathfrak{F}, P(.|B))$ are probability spaces, where the set functions $P(.|B) : \mathfrak{F} \to \mathbb{R}$ and $P_B : \mathfrak{F}_B \to \mathbb{R}$ are defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \ A \in \mathfrak{F} \quad \text{and} \quad P_B(C) = \frac{P(C)}{P(B)}, \ C \in \mathfrak{F}_B.$$

Conditional Probability Contd...

Definition

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let $B \in \mathfrak{F}$ be a fixed event such that $P(B) > 0$. Define the set function $P(.|B) : \mathfrak{F} \to \mathbb{R}$ by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \; A \in \mathfrak{F}.$$

We call $P(A|B)$ the conditional probability of event A given B.

Conditional Probability Contd...

Example

Six cards are dealt at random (without replacement) from a deck of 52 cards. Find the probability of getting all cards of heart in a hand given that there are at least 5 cards of heart in the hand.

Conditional Probability Contd...

Example

Six cards are dealt at random (without replacement) from a deck of 52 cards. Find the probability of getting all cards of heart in a hand given that there are at least 5 cards of heart in the hand.

Example

An urn contains 4 red and 6 black balls. 2 balls are drawn successively, at random and without replacement, from the urn. Find the probability that the first draw resulted in a red ball and the second draw resulted in a black ball.

Theorem of Total Probability

Theorem

Let $(\Omega,$ [illegible]$, P)$ *be a probability space and let* $\{E_i : i \in \Lambda\}$ *be a countable collection of mutually exclusive and exhaustive events such that* $P(E_i) > 0$ *for all* $i \in \Lambda$*. Then, for any event* $E \in$ [illegible]*, we have* 2^{Ω}

$$P(E) = \sum_{i \in \Lambda} P(E \cap E_i) = \sum_{i \in \Lambda} P(E_i) P(E|E_i).$$

Theorem of Total Probability: Example

Example

Urn U_1 contains 4 white and 6 black balls and urn U_2 contains 6 white and 4 black balls. A fair die is cast and urn U_1 is selected if the upper face of die shows 5 or 6 dots, otherwise urn U_2 is selected. If a ball is drawn at random from the selected urn, find the probability that the drawn ball is white.

Baye's Theorem

Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{E_i : i \in \Lambda\}$ be a countable collection of mutually exclusive and exhaustive events such that $P(E_i) > 0$ for all $i \in \Lambda$. Then, for any event $E \in 2^{\Omega}$ with $P(E) > 0$, we have

$$P(E_j|E) = \frac{P(E|E_j)P(E_j)}{\sum_{i \in \Lambda} P(E_i)P(E|E_i)}, \quad \cancel{i \in \Lambda} \quad = \frac{P(E|E_j)\,P(E_j)}{P(E)}$$

Baye's Theorem: Example

Example

Urn U_1 contains 4 white and 6 black balls and urn U_2 contains 6 white and 4 black balls. A fair die is cast and urn U_1 is selected if the upper face of die shows 5 or 6 dots, otherwise urn U_2 is selected. A ball is drawn at random from the selected urn.

(i) Given that the drawn ball is white, find the conditional probability that it came from urn U_1;

(i) Given that the drawn ball is white, find the conditional probability that it came from urn U_2;

Independence of Two Events

Definition

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let A and B be two events. Events A and B are said to be

(i) **negatively associated** if $P(A \cap B) < P(A)P(B)$;

(ii) **positively associated** if $P(A \cap B) > P(A)P(B)$;

(iii) **independent** if $P(A \cap B) = P(A)P(B)$;

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

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

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

Roll a dice

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

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

$$C = \{2\}$$

$$P(C) = \frac{1}{6}$$

$$P(C \mid A) = 0 < \frac{1}{6} = P(C)$$

Independence of Arbitrary Collection of Events

Definition

Let $(\Omega, \mathfrak{F}, P)$ be a probability space. Let $\Lambda \subseteq \mathbb{R}$ be an index set and let $\{E_i : i \in \Lambda\}$ be a collection of events in $\mathfrak{F}$.

(i) Events $\{E_i : i \in \Lambda\}$ are said to be pairwise independent if any pair of events E_i and E_j with $i \neq j$ in the collection $\{E_i : i \in \Lambda\}$ are independent.

Independence of Arbitrary Collection of Events

Definition

Let $(\Omega, \mathfrak{F}, P)$ be a probability space. Let $\Lambda \subseteq \mathbb{R}$ be an index set and let $\{E_i : i \in \Lambda\}$ be a collection of events in $\mathfrak{F}$.

(i) Events $\{E_i : i \in \Lambda\}$ are said to be pairwise independent if any pair of events E_i and E_j with $i \neq j$ in the collection $\{E_i : i \in \Lambda\}$ are independent.

(ii) Let $\Lambda = \{1, 2, ..., n\}$, for some $n \in \mathbb{N}$, so that $\{E_i : i \in \Lambda\} = \{E_1, E_2, ..., E_n\}$ is a finite collection of events in $\mathfrak{F}$. Events $E_1, E_2, ..., E_n$ are said to be independent if for any sub collection $\{E_{\alpha_1}, E_{\alpha_2}, ..., E_{\alpha_k}\}$ of $\{E_1, E_2, ..., E_n\}$

Independence of Arbitrary Collection of Events

$$P\left(\bigcap_{j=1}^{k} E_{\alpha_j}\right) = \prod_{j=1}^{k} P\left(E_{\alpha_j}\right) \quad (k = 2, 3, ..., n).$$

(iii) Let $\Lambda \subseteq \mathbb{R}$ be an arbitrary index set. Events $\{E_i : i \in \Lambda\}$ are said to be independent if any finite sub collection of events in $\{E_i : i \in \Lambda\}$ forms a collection of independent events.

Questions

- Is it true that independence of events $\{E_i : i \in \Lambda\}$ implies pairwise independence?

Questions

- Is it true that independence of events $\{E_i : i \in \Lambda\}$ implies pairwise independence?
- What can you say about the converse?

Random Variable: Motivation

Let $(\Omega, \mathfrak{F}, P)$ be a probability space.

- We are interested in the study of a real-valued function

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

Random Variable: Motivation

Let $(\Omega, \mathfrak{F}, P)$ be a probability space.

- We are interested in the study of a real-valued function

$$X : \Omega \rightarrow \mathbb{R}.$$

- It may be of interest to compute the probability that the random experiment results in a value of X in a given set $B \subseteq \mathbb{R}$.

Random Variable: Motivation

Let $(\Omega, \mathfrak{F}, P)$ be a probability space.

- We are interested in the study of a real-valued function

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

- It may be of interest to compute the probability that the random experiment results in a value of X in a given set $B \subseteq \mathbb{R}$.
- This amounts to assigning probabilities,

$$P_X(B) = P(\{w \in \Omega : X(w) \in B\}), \quad B \subseteq \mathbb{R}.$$

Random Variable: Motivation Contd...

Define $X^{-1} : 2^{\mathbb{R}} \to 2^{\Omega}$ by

$$X^{-1}(B) = \{w \in \Omega : X(w) \in B\}, \quad B \in 2^{\mathbb{R}}.$$

Lemma

Let $A, B \in 2^{\mathbb{R}}$ and let $A_\alpha \in 2^{\mathbb{R}}$, $\alpha \in \Lambda$, where $\Lambda \subseteq \mathbb{R}$ is an arbitrary index set. Then,

(i) $X^{-1}(A - B) = X^{-1}(A) - X^{-1}(B)$.

(ii) $X^{-1}\left(\bigcup_{\alpha \in \Lambda} A_\alpha\right) = \bigcup_{\alpha \in \Lambda} X^{-1}(A_\alpha)$ *and*

$X^{-1}\left(\bigcap_{\alpha \in \Lambda} A_\alpha\right) = \bigcap_{\alpha \in \Lambda} X^{-1}(A_\alpha)$.

(iii) $A \cap B = \phi$ *implies* $X^{-1}(A) \cap X^{-1}(B) = \phi$.

Random Variable

Let Ω be a sample space. A r.v. is a function, say X, defined on Ω whose range is in $\mathbb{R}$.

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

Ex 1 ~~(i)~~ $\Omega = \{HT, TH, TT, HH\}$

For $\omega \in \Omega$, $X_1(\omega) =$ no. of Heads $= \{0, 1, 2\}$

$X_2(\omega) =$ no. of Tails $= \{0, 1, 2\}$

$$X_3(\omega) := \begin{cases} 0 & \text{if no head} \\ 1 & \text{if atleast one head.} \end{cases}$$

$X_1 \in \{0, 1, 2\}$

$P(X_1 = 0) = P(TT) = \frac{1}{4}$

$P(X_1 = 1) = \frac{1}{2}$

$P(X_1 = 2) = \frac{1}{4}$

$$\boxed{P(X_1 = s) = 0 \quad \forall s \notin \{0, 1, 2\}}$$

$$\boxed{P(X_1=0) + P(X_1=1) + P(X_1=2) = 1}$$

$$\boxed{\{X = 3\} := \{\omega \in \Omega \mid X(\omega) = 3\}}$$

(i) F_{X_1} ? (ii) F_{X_3} ?

(i) $F_{X_1}(x) = P[X_1 \le x]$

$$= \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \le x < 1 \\ \frac{1}{4} + \frac{1}{2}, & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

$P[X_1 \le 1.5]$

$= P[X_1 = 0]$

$+ P[X_1 = 1]$

$$\boxed{\begin{aligned} P[X_1 = 0] &= \frac{1}{4} \\ P[X_1 = 1] &= \frac{1}{2} \\ P[X_1 = 2] &= \frac{1}{4} \\ X_1 &\in \{0, 1, 2\} \end{aligned}}$$

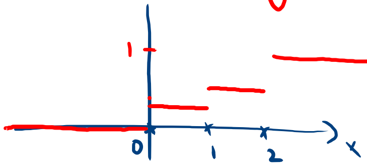
(i) $F_X(0) = F_X(0^+)$, $F_X(0) \neq F_X(0^-)$

$F_{X_1}(a) = F_{X_1}(a^+) \ni F_X$ is right-continuous at $a \in X(\Omega)$.

(ii) F_{X_1} is discont. at each point of $X(\Omega)$.

(iii) F_{X_1} is monotonically increasing fun^n.

(iv)

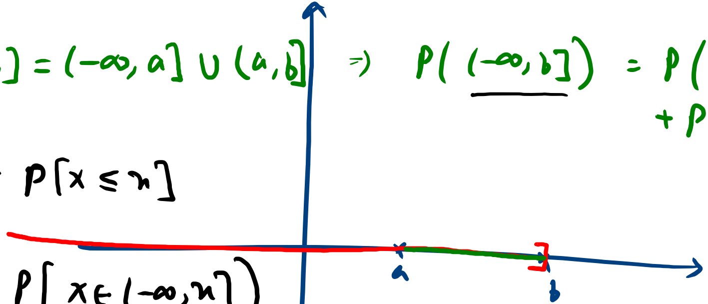


F_X behaves like a step-fun^n whenever X is discrete r.v.

(v) $F_{X_1}(a < X \leq b) = \underline{F_{X_1}(b) - F_{X_1}(a)}$

$[-\infty, b] = (-\infty, a] \cup (a, b] \Rightarrow P(\underline{(-\infty, b]}) = P(\underline{(-\infty, a]}) + P(\underline{(a, b]})$

$F_X(x) = P[X \leq x]$



$= P[X \in (-\infty, x]]$

(vi) $F_{X_1}(a \leq X \leq b) = F_{X_1}(b) - F_{X_1}(a^-)$

F_{X_3} ?

$X_3 \in \{0,1\}$

$P[X_3 = 0] = \frac{1}{4}$

$P[X_3 = 1] = \frac{3}{4}$

$$F_{X_3}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{4} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

$X_2 \in \{0, 1, 2\}$,

$P(X_2 = 0) = \frac{1}{4}$

$P(X_2 = 1) = \frac{1}{2}$

$P(X_2 = 2) = \frac{1}{4}$.

$P(X_2 = 0) + P(X_2 = 1) + P(X_2 = 2) = \underline{1}$

$P(X_2 = s) = 0 \quad \forall s \notin \{0, 1, 2\}$

$X_3 \in \{0, 1\}$

$P(X_3 = 0) = \frac{1}{4}$

$P(X_3 = 1) = \frac{3}{4}$

$P(X_3 = s) = 0$
$\forall s \notin \{0, 1\}$

$P(X_3 = 0) + P(X_3 = 1) = \underline{1}$

We say that the r.v. X is discrete if its range is either finite or countably infinite.

Ex 2

$$\Omega = \{ H, TH, TTH, TTTH, \ldots \}$$

$$X(\omega) = \text{no. of tails.}$$

$$\in \{0, 1, 2, 3, \ldots\} = \mathbb{N} \cup \{0\}.$$

$$P(X=0) = \frac{1}{2} \qquad P(X=i) = \frac{1}{2^{i+1}}$$

$$P(X=1) = \frac{1}{2^2} \quad \ldots$$

$$F_X(x) = P[X \le x],$$

$$= \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{2}, & \text{if } 0 \le x < 1 \\ \frac{1}{2} + \frac{1}{2^2}, & \text{if } 1 \le x < 2 \\ \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3}, & \text{if } 2 \le x < 3 \\ \vdots & \\ \sum_{s=0}^{i} \frac{1}{2^{s+1}} & \text{if } i \le x < i+1 \\ \vdots & \end{cases}$$

$$P[X = s] = \frac{1}{2^{s+1}}$$

if $s \in \mathbb{N} \cup \{0\}$

Ex 3 Rolling a dice two-times

$$\Omega = \{ (i,j) : i = 1,..,6;\ j = 1,..,6 \}$$

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

$$X((i,j)) = i+j \in \{2,3,4,5,6,7,8,9,10,11,12\}$$

$$P(X=7) = P(\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\})$$

$$= \frac{6}{36}$$

X	$P[X=x]$
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$

X	$P[X=x]$
9	$4/36$
10	$3/36$
11	$2/36$
12	$1/36$

$P[X=x] \geq 0 \quad \forall x \in X(\Omega)$

& $\sum_{x \in X(\Omega)} P[X=x] = 1$.

Prob. Mass function (p.m.f.)

A real-valued function p_X, defined on $\mathbb{R}$, for a given r.v. X, is said to be p.m.f. of X if

(i) $p_X(x) \geq 0 \quad \forall\, x \in \mathbb{R}$

(ii) $$\sum_{x \in \mathbb{R}} p_X(x) = 1$$

Ex: Let X be a discrete r.v. that takes the values 1, 2 or 3. Assume that we know the following.

$$p_X(2) = \frac{1}{6}, \quad p_X(3) = \frac{1}{3}$$

Then, find $p_X(1)$?

$$p_X(x) = P[X = x]$$

Sol: $p_X(x) = 0 \ \forall \ x \notin X(\Omega)$

$$\sum_{x \in \mathbb{R}} p_X(x) = \sum_{x \in X(\Omega)} p_X(x) = 1$$

$$\Rightarrow p_X(1) + p_X(2) + p_X(3) = 1.$$

Cummulative Distribution function

$$F_X(x) := P[X \le x] \, , \ \forall x \in \mathbb{R}.$$

(i) $0 \le F_X(x) \le 1 \quad \forall x \in \mathbb{R}.$

(ii) $F_X(\infty) = P[X < \infty] = P[X \in X(\Omega)] = 1$

(iii) $F_X(-\infty) = 0$

(iv) F_X is monotonically increasing.

(v) $F_X(x) = F_X(x^+)$.

Continuous r.v.

A r.v. X is said to be continuous if $\exists\ a, b \in \mathbb{R}$ s.t. $X(\Omega) = (a,b)$ or $(a,b]$ or $[a,b)$ or $[a,b]$.

Also, $\exists$ a nonnegative function f_X defined on $\mathbb{R}$ such that

$$* \quad P[X \in B] = \int_B f_X(x)\,dx, \quad B \subseteq \mathbb{R}.$$

We say that f_X is a probability density function of X.

Clearly,

$$1 = P[X \in \mathbb{R}] = \int_{\mathbb{R}} f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(x)\,dx$$

$$B = (a, b)$$

$$\boxed{P[X \in B] = \int_a^b f_X(x)\,dx}$$

Probability Density Function (p.d.f.)

A real-valued funⁿ f_X defined on $\mathbb{R}$ is said to be a p.d.f. for a continuous r.v. X if

(i) $f_X(x) \geq 0 \quad \forall\, x \in \mathbb{R}$

(ii) $$\int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

Ex: Let X be a cont. r.v. with p.d.f.

$$f_X(x) = \begin{cases} c(4x - 3x^3), & \text{if } 0 < x < 2 \\ 0 & \text{o.w.} \end{cases}$$

Find the value of c & $P[X > 1]$.

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

$$\Leftrightarrow \int_0^2 c\,(4x-3x^3)\,dx = \underline{1}$$

$$\Leftrightarrow c\left[2x^2 - \frac{3x^4}{4}\right]_0^2 = \underline{1}$$

$$\Rightarrow c\,[8-12] = 1 \Rightarrow \underline{c = -\frac{1}{4}}$$

$$g_X(x) = \begin{cases} c\,(4x-3x^2) & 0<x<2 \\ 0 & ow \end{cases}$$

$$c\left[2x^2 - x^3\right]_0^2$$

$$= c\,[8-8]$$

$$= 0$$

$$h_X(x) = \begin{cases} c\,(4x - 2x^2), & 0 < x < 2 \\ 0 & \text{ow} \end{cases}$$

$$\int_0^2 h_X(x)\,dx = 1$$

$$\Rightarrow \quad c\left[2x^2 - \frac{2}{3}x^3\right]_0^2 = 1$$

$$\Rightarrow \quad c\left[8 - \frac{16}{3}\right] = 1 \quad \Rightarrow \quad \frac{8}{3}c = 1$$

$$\Rightarrow \quad c = 3/8$$

Ex: 1 Suppose that 3 batteries are chosen from a group of 3 new, 4 used but still working and 5 defective batteries.

Let X = no. of new batteries that are chosen

& Y = no. of used batteries that are chosen

$X \in \{0,1,2,3\}$, $Y \in \{0,1,2,3\}$.

$$p_{X,Y}(x,y) = P[X=x, Y=y]$$

$p_X(x) = P[X=x]$

$p_Y(y) = P[Y=y]$

$\cancel{P[X = x}$

$$\{X = x\} = \{X = x\} \cap \Omega$$

$$= \{X = x\} \cap \left(\bigcup_{y \in Y} y \right)$$

$$= \bigcup_{y \in Y} \left(\{X = x\} \cap \{Y = y\} \right)$$

$$\Rightarrow P[\{X = x\}] = \sum_{y \in Y} p_{X,Y}(x, y)$$

$$\Rightarrow p_X(x) = \sum_{y \in Y} p_{X,Y}(x, y)$$

$X(\Omega) = x_1, x_2, \ldots \}$

$Y(\Omega) = y_1, y_2, \ldots$

$||^{ly}$ $\quad p_Y(y) = \sum_{x \in X} p_{X,Y}(x,y)$

$X \in \{0,1,2,3\}, \quad Y \in \{0,1,2,3\}$

X \ Y	0	1	2	3	
0	10/220	40/220	30/220	4/220	84/220
1	30/220	60/220	18/220	0	108/220
2	15/220	12/220	0	0	27/220
3	1/220	0	0	0	1/220
	56/220	112/220	48/220	4/220	

$$p_{X,Y}(0,0) = P[X=0, Y=0]$$

$$= {}^5C_3 \big/ {}^{12}C_3$$

$$= \frac{5 \cdot \cancel{4}^{2} \cdot 3}{\cancel{12}_{2} \cdot 11 \cdot \cancel{10}_{2}}$$

$$= \frac{60}{1320} = \frac{1}{22}$$

$$p_{X,Y}(0,1) = P[X=0, Y=1]$$

$$= \frac{{}^{4}C_1 \times {}^{5}C_2}{{}^{12}C_3} = \frac{\cancel{4}\,2 \times \frac{5\cdot 4}{\cancel{2}}}{\frac{\cancel{12}\,2\cdot 11\cdot 10}{\cancel{3}\cdot\cancel{2}\cdot 1}} = \frac{40}{220} = \frac{2}{11}$$

$$p_{X,Y}(1,2) = P[X=1, Y=2]$$

$$= \frac{{}^{3}C_1 \times {}^{4}C_2}{{}^{12}C_3} = ?$$

Ex②: Suppose that 15% of the families in a certain community have no children, 20% have 1, 35% have 2, and 30% have 3. Suppose that each child is equally likely to be a boy or a girl. If a family is chosen at random from this community, then B, the number of boys, and G, the number of girls in this family. Write their joint p.m.f.

$B \in \{0, 1, 2, 3\}$, $\quad G = \{0, 1, 2, 3\}$,

$$p_{B,G}(0,0) = P[B=0, G=0] = 0.15$$

$$p_{B,G}(1,2) = P[B=1, G=2]$$

$$= P[3\text{ child}]\ P[B=1, G=2 \mid 3\text{ child}]$$

$$= (0.3)\ {}^3C_1 \left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)^2$$

$$p_{B,G}(1,3) = 0$$

B \ G	0	1	2	3	
0	0.15	0.10	0.0875	0.0375	
1	0.10	0.175	0.1125	0	
2	0.0875	0.1125	0	0	
3	0.0375	0	0	0	

Joint cummulative Distribution Funct

$$F_{X,Y}(x,y) = P[X \le x, Y \le y]$$

$F_{X,Y}(x,\infty) = F_X(x) \longrightarrow$ marginal c.d.f. of X

& $F_{X,Y}(\infty,y) = F_Y(y) \longrightarrow$ marginal c.d.f. of Y.

$$F_{X,Y}(x,y) = \begin{cases} \displaystyle\sum_{a \le x} \sum_{b \le y} p_{X,Y}(a,b), & X, Y \text{ are discrete r.v.s} \\ \displaystyle\int_{-\infty}^{x} \int_{-\infty}^{y} \boxed{f_{X,Y}(x,y)}\, dy\, dx, & X \,\&\, Y \text{ are jointly cont.} \end{cases}$$

$$\boxed{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx = 1}$$

Ex: The joint p.d.f. of X & Y is

$$f_{X,Y}(x,y) = \begin{cases} 2\, e^{-x} e^{-2y}, & 0<x<\infty,\ 0<y<\infty \\ 0, & \text{otherwise} \end{cases}$$

Find (i) $P[X>1,\ Y>1]$

(ii) $P[X<a]$

(iii) $P[X<Y]$

(i) $P[X > 1, Y > 1] = P[X \in (1, \infty), Y \in (1, \infty)]$

$$= P[(X, Y) \in \underline{(1, \infty) \times (1, \infty)}]$$

$$= \int_1^\infty \int_1^\infty f_{X,Y}(x, y)\, dx\, dy$$

$$= \int_1^\infty \int_1^\infty 2e^{-x} e^{-2y}\, dx\, dy$$

$$= 2\left(\int_1^\infty e^{-x}\,dx\right)\left(\int_1^\infty e^{-2y}\,dy\right)$$

$$\int_1^\infty e^{-x}\,dx = \lim_{s\to\infty}\int_1^s e^{-x}\,dx = \lim_{s\to\infty}\left[-e^{-x}\right]_1^s$$

$$= \lim_{s\to\infty}\left[e^{-1} - e^{-s}\right] = e^{-1}$$

$$\text{||}^{ly} \quad \int_1^\infty e^{-2y}\,dy = \lim_{s\to\infty}\int_1^s e^{-2y}\,dy = -\frac{1}{2}\lim_{s\to\infty}\left[e^{-2y}\right]_1^s$$

$$= \frac{1}{2}e^{-2}.$$

$$\Rightarrow \quad P[X>1, Y>1] = 2\left(e^{-1}\right)\left(\frac{1}{2}e^{-2}\right)$$

$$= e^{-3}.$$

(ii) $P[X<a] = P[\,X \in (-\infty, a)\,]$

$$= P[\,X \in (-\infty, a),\ Y \in (-\infty, \infty)\,]$$

$$= P[\,(X,Y) \in (-\infty, a) \times (-\infty, \infty)\,]$$

$$= \int_{-\infty}^{a} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx.$$

$$= \begin{cases} 0 & \text{if } a \le 0 \\ \int_0^a \int_0^\infty f_{X,Y}(x,y)\,dy\,dx & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } a \le 0 \\ 2\int_0^a \int_0^\infty e^{-x} e^{-2y}\,dy\,dx & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if} \quad a \leq 0 \\ 2 \times \frac{1}{2} \int_0^a e^{-x}\,dx & \text{if} \quad a > 0 \end{cases}$$

$$\boxed{\begin{aligned} \int_0^\infty e^{-2y}\,dy &= \lim_{s\to\infty} \int_0^s e^{-2y}\,dy = -\frac{1}{2} \lim_{s\to\infty} \left[e^{-2y} \right]_0^s \\ &= -\frac{1}{2} \lim_{s\to\infty} \left[e^{-2s} - 1 \right] = \frac{1}{2} \end{aligned}}$$

$$\Rightarrow P[X < a] = \begin{cases} 0 & \text{if } a \leq 0 \\ \left[-e^{-x}\right]_0^a & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } a \leq 0 \\ 1 - e^{-a} & \text{if } a > 0 \end{cases}$$

(iii) $P[X < Y] = P[X \in (-\infty, y); Y \in (-\infty, \infty)]$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{X,Y}(x, y)\, dx\, dy$$

$$= \int_{0}^{\infty} \int_{0}^{y} f_{X,Y}(x, y)\, dx\, dy$$

$$= 2 \int_{0}^{\infty} \int_{0}^{y} e^{-x} e^{-2y}\, dx\, dy.$$

$$= 2\int_0^{\infty} e^{-2y}\left[\left\{-e^{-x}\right\}_0^{y}\right]dy$$

$$= 2\int_0^{\infty} e^{-2y}\left[1-e^{-y}\right]dy$$

$$= 2\int_0^{\infty}\left[e^{-2y}-e^{-3y}\right]dy$$

$$= 2\lim_{s\to\infty}\int_0^{s}\left[e^{-2y}-e^{-3y}\right]dy$$

$$= 2 \lim_{s \to \infty} \left[-\frac{1}{2} e^{-2y} + \frac{1}{3} e^{-3y} \right]_0^s$$

$$= 2 \lim_{s \to \infty} \left[-\frac{1}{2} e^{-2s} + \frac{1}{3} e^{-3s} + \frac{1}{2} - \frac{1}{3} \right]$$

$$= 2 \left[\frac{1}{2} - \frac{1}{3} \right] = 2 \times \frac{1}{6} = \frac{1}{3}$$

Let X, Y are given discrete r.vs.

If,
$$P[X = s \mid Y = r] = P[X = s]$$

$($ $\forall\ s \in X(\Omega)$

$\&\ r \in Y(\Omega)$

Joint p.m.f. X & Y are independent r.vs.

iff $p_{X,Y}(x,y) = P[X = x, Y = y] = p_X(x)\, p_Y(y)$.

Joint c.d.f.: We say that r.v.s. X & Y are independent if

$$F_{X,Y}(x,y) = P[X \le x, Y \le y]$$
$$= F_X(x) F_Y(y)$$
$$\forall x \in X(\Omega)$$
$$\& \ y \in Y(\Omega).$$

Joint p.d.f.: We say that the r.v.s X & Y are independent if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ $\forall x \in X(\Omega)$ & $y \in Y(\Omega)$

Ex: Let X & Y be independent r.v. that have common density function.

$$f_X(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{ow.} \end{cases}$$

Find p.d.f. of X/Y :

$$f_{X/Y}(x) = ?$$

$$F_{X/Y}(a) = P\left[X/Y \leq a\right] = P\left[X \leq aY\right]$$

$$= P\left[X \in (-\infty, ay),\ Y \in (-\infty, \infty)\right]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{ay} \underline{\underline{f_{X,Y}(x,y)}}\, dx\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{ay} f_X(x) f_Y(y)\, dx\, dy$$

$$= \begin{cases} 0 & \text{if } a \le 0 \\ \int_0^\infty \int_0^{ay} f_X(x) f_Y(y)\, dx\, dy & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } a \le 0 \\ \int_0^\infty e^{-y} \left(\underline{\int_0^{ay} e^{-x}\, dx} \right) dy & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } a \leq 0 \\ \int_0^\infty e^{-y}\left(1-e^{-ay}\right) dy & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } a \leq 0 \\ \int_0^\infty \left[e^{-y} - e^{-(1+a)y}\right] & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } a \le 0 \\ \displaystyle\lim_{s \to \infty} \left[-e^{-y} + \frac{1}{1+a} e^{-(1+a)y} \right]_0^s & \text{if } a > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } a \le 0 \\ \left[1 - \dfrac{1}{1+a} \right] & \text{if } a > 0 \end{cases}$$

$$f_{X/Y}(a) = \begin{cases} 0 & \text{if } a < 0 \\ \dfrac{1}{(1+a)^2} & \text{if } a > 0 \\ k & \text{if } a = 0 \end{cases}$$

$(k \geq 0)$

$\omega \in \Omega : \dfrac{X(\omega)}{Y(\omega)} = \left(\dfrac{X}{Y}\right)(\omega)$

$$\int_0^\infty f_{X/Y}(a)\, da = 1$$

Expectation of random variable

Let X be a discrete r.v. that assumes the values $\{x_1, x_2, x_3, \dots\}$.
Then, the expected value of X is the weighted average of the values of X, defined as,

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

Ex: 1 Let $X \in \{a, b\}$, $P[X=a] = P[X=b] = \frac{1}{2} \Rightarrow E[X] = \frac{a+b}{2}$

Ex ② $X \in \{a, b\}$, $P[X=a] = \frac{3}{4}$, $P[X=b] = \frac{1}{4}$, then

$$E[X] = a \times P[X=a] + b\, P[X=b]$$

$$= \frac{3a+b}{4}$$

Let X be a continuous r.v. with p.d.f. $f_X(x)$.

Then,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

$f_X(x) \approx P[x < X < x + \Delta x]$

Ex

$X \in \{1, 2, 3, 4, 5, 6\}$

$P[X = i] = \frac{1}{6}$. for $i = 1, \ldots, 6$.

$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$

$= \frac{1}{6}\left[\frac{6(6+1)}{2}\right] = \frac{7}{2}$.

$Y = X^2 \in \{1, 4, 9, 16, 25, 36\}$

$p_Y(y) = P[Y = y]$

$p_Y(1) = P[Y = 1] = P[X^2 = 1]$

$= P[X = 1] = p_X(1)$

$p_Y(4) = P(Y = 4) = P[X^2 = 4]$

$= p_X(2)$

$\Rightarrow$

$$p_Y(x^2) = p_X(x),$$

$$E[Y] = \sum y\, p_Y(y) = \sum x^2\, p_Y(x^2)$$

$$= \sum x^2\, p_X(x)$$

$$= \frac{1}{6}\left[1^2+2^2+3^2+4^2+5^2+6^2\right]$$

||ly $E[Z] = \sum x^3\, p_X(x)$, for $Z = X^3$

$$\boxed{E[X^n] = \sum_{x \in X(\Omega)} x^n p_X(x)}$$

<u>Ex:</u> $X \in \{-1, 1\}$, $P[X=1] = P[X=-1] = \frac{1}{2}$.

$Y = X^2 \in \{1\}$.

$$p_Y(y) = p_Y(x^2) = P[Y = x^2]$$

$$p_Y(1) = P[Y=1] = P[X^2=1] = 1$$

$E[Y] = 1 \times 1 = 1$

$\rightarrow E[X^2] = 1$.

$$\sum_{x \in X(\Omega)} x^2 p_X(x) = (-1)^2 \frac{1}{2} + (1)^2 \times \frac{1}{2} = 1.$$

\# $E[X^n] = \sum_{x \in X(\Omega)} x^n p_X(x)$ for a discrete r.v. X

& $n \in \mathbb{N}$.

Ex: Suppose that you are expecting a message at some time past 5 P.M. From experience, you know that X, the number of hours after 5 PM, until the message arrives, is a r.v. with p.d.f.

$$f_X(x) = \begin{cases} \frac{1}{1.5}, & 0 < x < 1.5 \\ 0 & \text{ow} \end{cases}$$

Find the expected amount of time past 5 PM, until the message arrives.

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$\Rightarrow E[X] = \int_0^{1.5} x \frac{1}{1.5}\, dx = \frac{1}{1.5}\left[\frac{x^2}{2}\right]_0^{1.5}$$

$$= \frac{1.5}{2} = \frac{3}{4}.$$

$$\boxed{E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx}$$

$$\int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \frac{1}{1.5}\int_0^{1.5} x^2\, dx = \frac{(1.5)^2}{3}. = \frac{\cancel{9}^{3}}{4} \times \frac{1}{\cancel{3}}$$

$$= 3/4.$$

Let $Y = X^2$,

$$F_Y(y) = P[Y \leq y] = \begin{cases} 0 & \text{if } y \leq 0 \\ P[X^2 \leq y] & \text{if } y > 0 \end{cases}$$

For $y > 0$,

$$F_Y(y) = P[X^2 \leq y] = P[X \leq \sqrt{y}]$$

$$= \int_{-\infty}^{\sqrt{y}} f_X(x)\,dx = 1 \quad \text{if } \sqrt{y} > 1.5$$

$$= \frac{\sqrt{y}}{1.5} \quad \text{if} \quad \sqrt{y} < 1.5$$

$$F_Y(y) = \begin{cases} 0 & \text{if} \quad y \leq 0 \\ \frac{\sqrt{y}}{1.5} & \text{if} \quad 0 < \sqrt{y} < 1.5 \\ 1 & \text{if} \quad \sqrt{y} \geq 1.5 . \end{cases}$$

$$\Rightarrow f_Y(y) = \begin{cases} \frac{1}{3\sqrt{y}} & \text{if} \quad 0 < \sqrt{y} < 1.5 \\ 0 & \text{ow} . \end{cases}$$

$$E[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_0^{(1.5)^2} y \frac{1}{3\sqrt{y}}\, dy$$

$$= \frac{1}{3} \int_0^{(1.5)^2} \sqrt{y}\, dy$$

$$= \frac{1}{3} \left[y^{3/2} \right]_0^{1.5} \times \frac{2}{3}$$

$$= \frac{2}{9} (1.5^2)^{3/2} = \frac{2}{9} \times (1.5)^3 = \frac{2}{9} \times \left(\frac{3}{2}\right)^3 = \frac{3}{4}$$

$$E[X^n] = \begin{cases} \displaystyle\sum_{x \in X(\Omega)} x^n \, p_X(x), & X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} x^n f_X(x)\,dx, & X \text{ is continuous} \end{cases}$$

and known as nth moment of X about the origin

Ex: The time, in hours, it takes to locate and repair an electrical breakdown in a certain factory is a r.v. with p.d.f.

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{o.w.} \end{cases}$$

If the cost involved in breakdown of duration x is x^3, what is the expected cost of such breakdown.

$$E[X^3] = \int_{-\infty}^{\infty} x^3 f_X(x)\,dx \quad \text{(verify it)}$$

Find p.d.f. of X^3 and then find its expectation.

Let $Y = X^3$.

$$F_Y(y) = P[Y \le y] = P[X^3 \le y]$$
$$= P[X \le y^{1/3}] \quad \text{for } y > 0.$$

$$= \int_0^{y^{1/3}} f_X(x)\,dx \quad \text{if} \quad y>0$$

$$= \begin{cases} 0 & \text{ow} \\ y^{1/3} & \text{if} \quad 0<y<1 \\ 1 & \text{if} \quad y \geq 1 \end{cases}$$

$$f_Y(y) = \begin{cases} F_Y'(y) & \text{if} \quad 0<y<1 \\ 0 & \text{ow.} \end{cases} = \begin{cases} \frac{1}{3} y^{-2/3} & \text{if } 0<y<1 \\ 0 & \text{ow.} \end{cases}$$

$$E[Y] = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy$$

$$= \int_0^1 y \left(\frac{1}{3} y^{-2/3} \right) dy$$

$$= \frac{1}{3} \int_0^1 y^{1/3} \, dy = \frac{1}{3} \frac{3}{4} \left[y^{4/3} \right]_0^1$$

$$= \frac{1}{4}.$$

$$\int_{-\infty}^{\infty} x^3 f_X(x)\,dx = \int_0^1 x^3\,dx = \left[\frac{x^4}{4}\right]_0^1 = \frac{1}{4}.$$

$$E[Y] = E[X^3] = \int_{-\infty}^{\infty} x^3 f_X(x)\,dx$$

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx.$$

Let g be a fun^n of r.v. X. Then,

$$E[g(X)] = \begin{cases} \displaystyle\sum_{x \in X(\Omega)} g(x)\, p_X(x), & X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x)\, dx, & X \text{ is cont.} \end{cases}$$

Properties

(i) Let X be a constant r.v., i.e., $X = b$, then

$$E[X] = E[b] = b.$$

(ii) Let X be a r.v. and $a, b \in \mathbb{R}$, then

$$E[aX+b] = aE[X] + b.$$

$$E[aX+b] = \begin{cases} \displaystyle\sum_{x \in X(\Omega)} (ax+b)\,p_X(x), & X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} (ax+b)\,f_X(x)\,dx, & X \text{ is cont.} \end{cases}$$

Let X & Y be r.vs. and $g(X,Y)$ be a real-valued fun^n of X & Y. Then,

$$E[g(X,Y)] = \begin{cases} \displaystyle\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x,y)\, p_{X,Y}(x,y), & X \text{ \& } Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy, & X \text{ \& } Y \text{ are jointly cont. r.vs.} \end{cases}$$

$E[aX \pm bY] = aE[X] \pm bE[Y]$.

$E[X_1 + X_2 + \cdots + X_n] = \sum_{i=1}^{n} E[X_i]$.

Ex: A construction firm has recently sent in bids for 3 jobs worth (in profit) 10, 20, and 40 (Thousand) dollars. If its prob. of winning the jobs are respectively 0.2, 0.8, and 0.3, what is the firm's expected total profit.

Let X_i denotes the firm profit from ith job

$$X_1 = \begin{cases} 10 & \text{with prob. } 0.2 \\ 0 & \quad 0.8 \end{cases}$$

$$X_2 = \begin{cases} 20 & \quad 0.8 \\ 0 & \quad 0.2 \end{cases}$$

$$X_3 = \begin{cases} 40 & \quad 0.3 \\ 0 & \quad 0.7 \end{cases}$$

$$X = X_1 + X_2 + X_3$$

$$E[X] = E[X_1] + E[X_2] + E[X_3]$$

$$= (10)(0.2) + 20(0.8) + 40(0.3)$$

$$=$$

Ex: Suppose there are 20 coupons of different type. Assume that each time one obtains a coupon, it is equally likely to be any one of the types. Compute the expected number of different types that are obtained in a set of 10 coupons.

Sol: Let X denotes the number of different types that are obtained in a set of 10 coupons.

$$X_i = \begin{cases} 1 & \text{if atleast one type } i \text{ contains in a set of 10 coupons} \\ 0 & \text{otherwise.} \end{cases}$$

$$X = \sum_{i=1}^{20} x_i$$

$$E[X] = \sum_{i=1}^{20} E[x_i]$$

$$= 20\left(1-\left(\frac{19}{20}\right)^{10}\right)$$

$$E[x_i] = P[x_i = 1]$$

$$= 1-\left(\frac{19}{20}\right)^{10}$$

$\forall\, i=1, \ldots, 20.$

#

$X \approx c$, $\mu_x = E[X] \Rightarrow E[X-\mu_x] = 0$

$$E[(X-c)^2] = E[(X-\mu_x+\mu_x-c)^2]$$

$$= E[(X-\mu_x)^2 + \underline{(\mu_x-c)^2} + 2(\mu_x-c)(X-\mu_x)]$$

$$= E[(X-\mu_x)^2] + (\mu_x-c)^2 + 2(\mu_x-c)\underbrace{E[\cancel{(X-\mu_x)}]}_{0}$$

$$= E[(X-\mu_x)^2] + \underline{(\mu_x-c)^2}$$

$$\geq E[(X-\mu_x)^2] = \underline{\mathrm{Var}(X)}$$

$Var(X) = E[(X-\mu_X)^2]$

(i) $Var(X) \geq 0$, $Var(X) = \sigma_X^2$

& $\sigma_X = \sqrt{Var(X)}$ is known as standard deviation of r. v. X.

(ii) $Var(X) = E[(X-\mu_X)^2] = E[X^2 + \mu_X^2 - 2\mu_X X]$

$= E[X^2] + \mu_X^2 - 2\mu_X E[X]$

$= E[X^2] + \mu_X^2 - 2\mu_X^2 = E[X^2] - (E[X])^2$

$\because \ Var(X) \geq 0$

$\Rightarrow \ E[X^2] \geq (E[X])^2.$

(iii) $Var(b) = E[b^2] - (E[b])^2 = b^2 - b^2 = 0.$

(iv) $Var(aX+b) = E\left[\left((aX+\cancel{b}) - E[aX+\cancel{b}]\right)^2\right]$

$= E\left[\{a(X-\mu_X)\}^2\right]$

$= a^2 E[(X-\mu_X)^2]$

$= a^2 Var(X)$

(v) $\mathrm{Var}(X+X) \neq 2\,\mathrm{Var}(X)$

(vi) $\mathrm{Var}(X+Y) = ?$

Var(X)

$= E[(X-\mu_X)^2]$

$= E[(X-\mu_X)(X-\mu_X)]$

Covariance of X & Y

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

$$= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]$$

$$= E[XY] - \cancel{\mu_X \mu_Y} - \mu_Y \mu_X \cancel{+ \mu_X \mu_Y}$$

$$= E[XY] - E[X]\,E[Y].$$

Ex: $\mathrm{cov}(X,Y)=0$ iff $E[XY]=E[X]\,E[Y]$.

Ex: Let X & Y are independent r.vs, then
$$E[XY]=E[X]\,E[Y].$$

Sol: $$E[XY]=\sum_x\sum_y (xy)\,p_{X,Y}(x,y)$$

$$=\sum_x\sum_y xy\,p_X(x)\,p_Y(y) \qquad [\because X, Y \text{ are independent}]$$

$$=\sum_x x\,p_X(x)\Big(\sum_y y\,p_Y(y)\Big)=E[X]\,E[Y].$$

Ex: Let X be a cont. r.v. with p.d.f.

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } -1 < x < 1 \\ 0 & \text{o.w.} \end{cases}$$

Define $Y = X^2$. Find $E[XY]$ & $E[X]E[Y]$.

$$E[XY] = E[X^3] = \int_{-1}^{1} x^3 \frac{1}{2} dx = 0$$

$$E[X] = \int_{-1}^{1} x \frac{1}{2} dx = 0$$

$$E[Y] = \int_{-1}^{1} x^2 \frac{1}{2} dx = \frac{1}{3}.$$

Properties of Covariance

$$\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E[XY] - E[X]E[Y].$$

(i) $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

(ii) $\text{Cov}(aX,Y) = E[(aX)Y] - E[(aX)]E[Y]$
$$= a\,\text{Cov}(X,Y).$$

(iii) $\text{Cov}(X+Y,Z) = E[(X+Y)Z] - E[(X+Y)]E[Z]$
$$= \underline{E[XZ]} + E[YZ] - \underline{E[X]E[Z]} - E[Y]E[Z]$$
$$= \text{Cov}(X,Z) + \text{Cov}(Y,Z)$$

(iv) $\text{Cov}\left(\sum_{i=1}^{n} X_i, Y\right) = \sum_{i=1}^{n} \text{Cov}(X_i, Y)$

(v) $\operatorname{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{Cov}(X_i, Y_j)$

(vi) $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$.

$\operatorname{Var}(X+Y) = \operatorname{Cov}(X+Y, X+Y)$

$$= \operatorname{Cov}(X, X+Y) + \operatorname{Cov}(Y, X+Y)$$

$$= \operatorname{Cov}(X,X) + \operatorname{Cov}(X,Y) + \operatorname{Cov}(Y,X) + \operatorname{Cov}(Y,Y)$$

$$= \operatorname{Var}(X) + \operatorname{Var}(Y) + \operatorname{Cov}(X,Y) + \operatorname{Cov}(Y,X)$$

$$= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$$

Ex: Let X & Y be independent r.vs., then

$$Var(X+Y) = Var(X) + Var(Y).$$

$Var(\underbrace{X_1+X_2}_{X}+\underbrace{X_3}_{Y})$

$$= \underline{Var(X_1+X_2)} + Var(X_3) + Cov(X_1+X_2, X_3) + Cov(X_3, X_1+X_2)$$

$$= Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) + Var(X_3) + 2\left[Cov(X_1, X_3) + Cov(X_2, X_3)\right]$$

$$= \sum_{i=1}^{3} Var(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{3} Cov(X_i, X_j)$$

Ex: Let $X_1, X_2, \dots, X_n$ be r.v.s., then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \mathrm{Cov}(X_i, X_j).$$

Ex: Compute the variance of the sum obtained when 10 independent rolls of a fair dice are made.

$X_i \sim$ ith rolls of a fair dice.

$$X = \sum_{i=1}^{10} X_i \Rightarrow \mathrm{Var}(X) = \sum_{i=1}^{10} \mathrm{Var}(X_i) = 10\,\mathrm{Var}(X_i)$$

$X_i \in \{1, \ldots, 6\}$

$$P[X_1 = k] = \frac{1}{6} \quad \forall \; k \in \{1, \ldots, 6\}$$

$$Var(X_1) = E[X_1^2] - (E[X_1])^2$$

$$E[X_1^2] = \sum_x x^2 p_X(x) = \frac{1}{6}(1^2 + 2^2 + \ldots + 6^2)$$

$$= \frac{1}{6} \times \frac{\cancel{6} \times 7 \times 13}{\cancel{6}} = \frac{91}{6}$$

$$E[X_1] = \sum_x x p_X(x) = \frac{1}{6}(1 + \ldots + 6) = 7/2$$

$$Var(X_1) = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12}$$

$$\Rightarrow Var(X) = \frac{35}{12}.$$

Ex: Compute the variance of the number of heads resulting from 10 independent tosses of a fair coin.

Sol:

$$X_i = \begin{cases} 1 & \text{if } i\text{th toss shows head} \\ 0 & \text{o.w.} \end{cases}$$

$X = \sum_{i=1}^{10} X_i$. & X_i's are independent

$\Rightarrow Var(X) = \sum_{i=1}^{10} Var(X_i)$

$= 10 \times \frac{1}{4} = \frac{5}{2}$.

$E[X_i^2] = P[X_i = 1] = \frac{1}{2}$

$E[X_i] = P(X_i = 1) = \frac{1}{2}$

$\Rightarrow Var(X_i) = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$

#

$$X = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{o.w.} \end{cases}$$

$$Y = \begin{cases} 1 & \text{if event } B \text{ occurs} \\ 0 & \text{o.w.} \end{cases}$$

$$XY = \begin{cases} 1 & \text{if both events } A \,\&\, B \text{ occur} \\ 0 & \text{o.w.} \end{cases}$$

Note that $E[X] = P[X=1]$

$$E[Y] = P[Y=1]$$

$$E[XY] = P[X=1, Y=1]$$

$$\Rightarrow \quad Cov(X,Y) = P[X=1, Y=1] - P[X=1]\,P[Y=1].$$

~~Now~~ Let $Cov(X,Y) \underset{(<)}{>} 0$

$$\Leftrightarrow \quad P[X=1, Y=1] \underset{(<)}{>} P[X=1]\,P[Y=1]$$

$$\Leftrightarrow \quad \frac{P[X=1, Y=1]}{P[X=1]} \underset{(<)}{>} P[Y=1], \quad \boxed{P[X=1] > 0}$$

$$\Longleftrightarrow \quad P[Y=1 \mid X=1] \underset{(<)}{>} P[Y=1]$$

The correlation co-efficient:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \, \sigma_Y}$$

where $\sigma_X = \sqrt{Var(X)}$

$\sigma_Y = \sqrt{Var(Y)}$, are standard deviations of X & Y respectively.

Ex: $\rho_{X,Y} \in [-1, 1]$.

Markov's Inequality

If X is a r.v. that takes non-negative values, then for any value $a>0$,

$$P[X \geq a] \leq \frac{E[X]}{a}$$

Proof: X is discrete and $a>0$. Let $p_X(x)$ denotes its p.m.f.

$$P[X \geq a] = \sum_{\substack{x \in X(\Omega) \\ x \geq a}} p_X(x) \quad - \text{L.H.S.}$$

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

$$= \sum_{\substack{x \in X(\Omega) \\ x < a}} x\, p_X(x) + \sum_{\substack{x \in X(\Omega) \\ x \geq a}} x\, p_X(x)$$

$$\geq \sum_{\substack{x \in X(\Omega) \\ x \geq a}} x\, p_X(x) \geq a \sum_{\substack{x \in X(\Omega) \\ x \geq a}} p_X(x)$$

$$= a\, P[X \geq a].$$

Chebyshev's Inequality

If X is a r.v. with mean μ_x and variance σ_x^2, then for any $k>0$

$$0 \leq P\left[\,|X-\mu_x| \geq k\right] \leq \frac{\sigma_x^2}{k^2}.$$

Pf: $\{|X-\mu_x| \geq k\} \Longleftrightarrow \{(X-\mu_x)^2 \geq k^2\}$

Then, $P\left[(X-\mu_x)^2 \geq k^2\right] \leq \dfrac{E\left[(X-\mu_x)^2\right]}{k^2} = \dfrac{\sigma_x^2}{k^2}.$

The weak law of large numbers

Let $X_1, X_2, X_3, \ldots$ be a seq^n of identically independently distributed (i.i.d.) r.v's with mean μ, then for any $\varepsilon > 0$

$$P\left[\left|\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right|>\varepsilon\right] \to 0$$

as $n \to \infty$

$$\bar{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$\Rightarrow \quad E[\bar{X}] = \mu$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$P\left[\,|\bar{X} - \mu| \geq \varepsilon\,\right] \leq \frac{\mathrm{Var}(\bar{X})}{\varepsilon^2}$$

$$= \frac{\sigma^2}{n\varepsilon^2}$$

$$\to 0 \text{ as } n \to \infty$$

Ex: Suppose that it is known that the number of items produced in a factory during a weak is a r.v. with mean 50.

(a) What can be said about the prob. that this week's production will exceed 75? $P[X \geq 75] \leq \frac{50}{75}$

(b) If the variance of a week's production is known to equal 25, then what can be said about the prob. that this week's production will be between 40 & 60?

$$P[40 \leq X \leq 60] = P[40-50 \leq X-50 \leq 60-50]$$

$$= P[-10 \le X-50 \le 10]$$

$$= P[\,|X-50| \le 10\,]$$

$$= 1 - P[\,|X-50| \ge 11\,]$$

$$\ge 1 - \frac{25}{121}.$$

$$P[\,|X-50| \ge 11\,]$$

$$\le \frac{\sigma_X^2}{11^2} = \frac{25}{121}$$

$$\Rightarrow -P[\,|X-50| \ge 10\,]$$

$$\ge -\frac{25}{121}$$

Bernoulli R.v.

A r.v. X is said to be Bernoulli if it assumes the value 0 & 1 with probability $1-p$ and p, respectively, for $p \in (0,1)$.

i.e., $X \in \{0,1\}$,

$$p_X(x) = \begin{cases} 1-p & \text{if } x=0 \\ p & \text{if } x=1 \\ 0 & \text{ow} \end{cases}$$

$E[X] = P[X=1] = p$

$E[X^2] = P[X=1] = p$

$\Rightarrow \operatorname{var}(X) = E[X^2] - (E[X])^2$

$= p - p^2 = p(1-p)$

Binomial R.V.

Let $X_1, X_2, \ldots, X_n$ be n-independent Bernoulli r.vs. with prob. of success p in each case.

Define $X = X_1 + X_2 + \cdots + X_n$.

$\in \{0, 1, 2, \ldots, n\}$.

$$p_X(i) = P[X=i] = i\text{-success} + (n-i)\text{-failure}.$$

$$= {}^nC_i \, p^i (1-p)^{n-i}$$

$$1 = [p+(1-p)]^n = \sum_{i=0}^{n} {}^nC_i\, p^i (1-p)^{n-i} = \sum_{i=0}^{n} p_X(i)$$

$$E[X] = \sum_{x \in X(\Omega)} x\, p_X(x) = \sum_{i=0}^{n} i\, p_X(i)$$

$$= \sum_{i=0}^{n} i\, {}^nC_i\, p^i (1-p)^{n-i} = \sum_{i=1}^{n} \frac{n!}{(i-1)!\,(n-i)!} p^i (1-p)^{n-i}$$

$$= np \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!\,((n-1)-(i-1))!} p^{i-1} (1-p)^{(n-1)-(i-1)}$$

$$= np \sum_{i=1}^{n} {}^{n-1}C_{i-1}\, p^{i-1} (1-p)^{(n-1)-(i-1)} = np\, [p+(1-p)]^{n-1} = np$$

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = np.$$

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = np(1-p).$$

$$E[X^2] = Var(X) + (E[X])^2$$

$$= np(1-p) + n^2p^2 = np - np^2 + n^2p^2$$

Ex: It is known that disks produced by a certain company will be defective with prob. 0.01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that atmost 1 of the 10 disks is defective. What proportion of package is returned.

$$P[X>1] = 1 - P[X \le 1] = 1 - P[X=0] - P[X=1]$$

$$= 1 - {}^{10}C_0\, p^0 (1-p)^{10} - {}^{10}C_1\, p^1 (1-p)^9$$

$$\approx 0.005$$

Ex: If X_1 & X_2 are independent Binomial r.v.s such that

$$X_i \sim B(n_i, p), \quad i = 1, 2.$$

then $$X_1 + X_2 \sim B(n_1 + n_2, p)$$

#

$$X_1 = X_1^{(1)} + \cdots + X_1^{(n_1)}, \quad X_1^{(i)} \sim Ber(p)$$

$$X_2 = X_2^{(1)} + \cdots + X_2^{(n_2)}, \quad X_2^{(i)} \sim Ber(p)$$

$$\Rightarrow \quad X_1 + X_2 = \sum_{i=1}^{n_1} X_1^{(i)} + \sum_{j=1}^{n_2} X_2^{(j)} \sim B(n_1 + n_2, p)$$

$X_1 \sim B(n_1, p)$, $\quad X_2 \sim B(n_2, p)$

$\Rightarrow M_{X_1}(t) = [pe^t + (1-p)]^{n_1}$, $\quad M_{X_2}(t) = [pe^t + (1-p)]^{n_2}$

$$M_X(t) = E[e^{t(X_1+X_2)}] = E[e^{tX_1} e^{tX_2}]$$

$$= \prod_{i=1}^{2} E[e^{tX_i}] = [pe^t + (1-p)]^{n_1+n_2}$$

$\Rightarrow \quad X \sim B(n_1+n_2, p)$.

Moment Generating Function (m.g.f.)

$$M_X(t) = E\left[e^{tX}\right]$$

$$= \begin{cases} \sum\limits_{x \in X(\Omega)} e^{tx} p_X(x), & X \text{ is discrete.} \\ \int\limits_{-\infty}^{\infty} e^{tx} f_X(x)\,dx, & X \text{ is cont.} \end{cases}$$

Remark: m.g.f defines the distribution of X uniquely

Ex: Find m.g.f. of ~~$\;$~~ $X \sim Ber(p)$.

Sol: $M_X(t) = E[e^{tX}] = \sum_{x \in X(\Omega)} e^{tx} p_X(x)$

$= 1 \times (1-p) + e^t p$

$= [p e^t + (1-p)]$

$M_X(0) = 1$

$M_X'(0) = p = E[X]$

$M_X''(0) = p = E[X^2]$

$\vdots$

$M_X^{(n)}(0) = E[X^n]$

Ex: Find m.g.f. of $X \sim B(n,p)$.

Sol: $M_X(t) = E[e^{tX}] = \sum_{x \in X(\Omega)} e^{tx} p_X(x) = \sum_{i=0}^{n} e^{it} p_X(i)$

$$X = \sum_{i=1}^{n} X_i \quad , \quad X_i \sim Ber(p)$$

$$\Rightarrow E[e^{tX}] = E\left[e^{t\sum_{i=1}^{n} X_i}\right] = E\left[e^{\sum_{i=1}^{n} tX_i}\right]$$

$$= E\left[\prod_{i=1}^{n} e^{tX_i}\right]$$

$$= \prod_{i=1}^{n} E[e^{tX_i}] = [pe^t + (1-p)]^n$$

$$\Rightarrow M_X(t) = [pe^t + (1-p)]^n \quad ,$$

$$M_X'(t) = n[pe^t + (1-p)]^{n-1} pe^t$$

$$M_X'(0) = np = E[X].$$

$$M_X'(t) = np\left[pe^t + (1-p)\right]^{n-1} e^t$$

$$M_X''(t) = np\left[\left(pe^t+(1-p)\right)^{n-1} e^t + (n-1)\left(pe^t+(1-p)\right)^{n-2} p e^{2t}\right]$$

$$M_X''(0) = np\left[1 + (n-1)p\right] = np + np^2(n-1)$$

$$= np + n^2p^2 - np^2 = n^2p^2 + np(1-p) = E[X^2].$$

Poisson R.V.

A r.v X, taking on one of the values from $\mathbb{N} \cup \{0\}$ is said to be a Poisson r.v. with parameter "$\lambda > 0$", if its p.m.f. is given by

$$P[X = i] = \begin{cases} e^{-\lambda} \dfrac{\lambda^i}{i!}, & i = 0, 1, 2, 3, \ldots \\ 0 & \text{o.w.} \end{cases}$$

$X \sim B(n, p)$

$\Rightarrow \mu = np$

$\sigma^2 = np(1-p)$

$= np - np^2$

We write $X \sim P(\lambda) = Pois(\lambda)$

$$E[X] = \sum_{x \in X(\Omega)} x\, p_X(x) = \sum_{i=0}^{\infty} i\, p_X(i)$$

$$= \sum_{i=1}^{\infty} i\, e^{-\lambda} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}$$

$$= \lambda\, e^{-\lambda} e^{\lambda} = \lambda .$$

Assume ~~that~~ 'n' is large in $B(n, p)$.

Define $\underset{< \infty}{\lambda = np} \Rightarrow p = \left(\frac{\lambda}{n}\right)$

$$P[X=i] = {}^nC_i \, p^i (1-p)^{n-i}$$

$$= \frac{n!}{i!\,(n-i)!} \left(\frac{\lambda}{n}\right)^i \left[1-\frac{\lambda}{n}\right]^{n-i}$$

$$= \frac{\lambda^i}{i!} \left[1-\frac{\lambda}{n}\right]^n \left[\frac{n!}{(n-i)!\,n^i}\right] \left[1-\frac{\lambda}{n}\right)^{-i}$$

$$= \frac{\lambda^i}{i!} \left[1-\frac{\lambda}{n}\right]^n \left[\frac{n(n-1)\ldots(n-i+1)}{n^i}\right] \left[1-\frac{\lambda}{n}\right)^{-i}$$

$$= \frac{\lambda^i}{i!} \left[1-\frac{\lambda}{n}\right]^n \left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{i-1}{n}\right)\right] \left[1-\frac{\lambda}{n}\right]^{-i}$$

$E[X^2] = \sum_{x \in X(\Omega)} x^2 p_X(x)$

$\because X \sim Pois(\lambda)$

$\Rightarrow E[X^2] = \sum_{i=0}^{\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!}$

$= e^{-\lambda} \sum_{i=1}^{\infty} \frac{i \lambda^i}{(i-1)!}$

$= e^{-\lambda} \sum_{i=1}^{\infty} \frac{(i-1+1)\lambda^i}{(i-1)!}$

$\Rightarrow E[X^2] = e^{-\lambda}\left[\sum_{i=2}^{\infty} \frac{\lambda^i}{(i-2)!} + \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} \right]$

$= e^{-\lambda}\left[\lambda^2 e^{\lambda} + \lambda e^{\lambda}\right]$

$= \lambda(\lambda+1)$

$Var(X) = E[X^2] - (E[X])^2$

$= \lambda^2 + \lambda - \lambda^2 = \lambda$

Ex: Let $X \sim Pois(\lambda)$. Find its m.g.f.

$$M_X(t) = E[e^{tX}] = \sum_{x \in X(\Omega)} e^{tx} p_X(x)$$

$$= \sum_{i=0}^{\infty} e^{it} e^{-\lambda} \frac{\lambda^i}{i!}$$

$$= e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!}$$

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Ex: Let $X_1 \sim Pois(\lambda_1)$ & $X_2 \sim Pois(\lambda_2)$. Then, $X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$ whenever X_1 & X_2 are independent r.vs.

$M_{X_1}(t) = e^{\lambda_1(e^t - 1)}$

$Y = X_1 + X_2$

$$M_Y(t) = E[e^{tY}] = E[e^{tX_1} e^{tX_2}]$$

$$= E[e^{tX_1}] E[e^{tX_2}]$$

$$= M_{X_1}(t) M_{X_2}(t) = e^{(\lambda_1 + \lambda_2)(e^t - 1)}$$

Ex: Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the prob. that there is at least one accident ~~dis~~ this week.

$$\boxed{\begin{aligned} P[X \geq 1] &= 1 - P[X=0] \\ &= 1 - e^{-3} \end{aligned}}$$

, $X \sim Pois(\lambda = 3)$

$$P[X=i] = \begin{cases} e^{-3} \dfrac{3^i}{i!} & , i = 0, 1, 2, 3, \ldots \\ 0 & o/w \end{cases}$$

Ex: If the average number of claims handled daily by an insurance company is 5, what proportion of days have less than 3 claims?

$$\boxed{P[X<3] = \sum_{i=0}^{2} P[X=i]}$$

$$X \sim \text{Poi}(\lambda = 5)$$

$$P[X=i] = \begin{cases} e^{-5}\dfrac{5^i}{i!}, & i=0,1,2,3,\ldots \\ 0 & \text{ow} \end{cases}$$

Ex: It has been established that the number of defective stereos produced daily at a certain plant is Poisson distributed with mean 4. Over a 2-day span, what is the prob. that the number of defective stereos does not exceed 3.

2-day span ~ Pois (4+4 = 8)

$P[X \le 3]$ ________ , $P[X=i] = \begin{cases} \dfrac{e^{-8}\, 8^{i}}{i!}, & i=0,1,2,\ldots \\ 0 & \text{ow} \end{cases}$

$X \sim Geo(p)$, $X \in \{1, 2, 3, \dots\}$

$$P[X=i] = \begin{cases} (1-p)^{i-1}\, p & i = 1, 2, 3, 4, \dots \\ 0 & \text{ow} \end{cases}$$

Ex: An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is the prob. that

(a) exactly 'n' draws are needed?

(b) at least 'k' draws are needed?

$$\boxed{p = \frac{M}{M+N}}$$

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

(b) $P[X \geq k] = (1-p)^{k-1} \left[\sum_{k=1}^{\infty} P[X = k \qquad \right]$

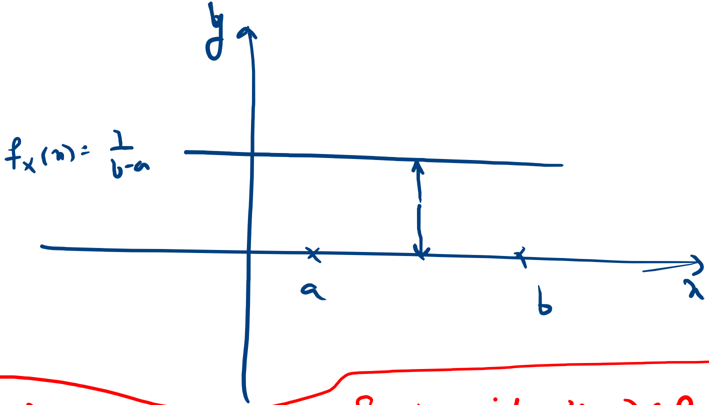
X is uniformly distributed over an interval $[a, b]$

$$f_X(x) = \begin{cases} k, & a \le x \le b \\ 0 & ow \end{cases} = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & ow \end{cases}$$

$$\boxed{\int_{-\infty}^{\infty} f_X(x)\,dx = 1}$$

$$\boxed{X \sim U[a, b]}$$

$$\Rightarrow \quad k = \frac{1}{b-a}$$



(X, Y) , $f_{X,Y}(x,y) = \begin{cases} k & \text{if } (x,y) \in R \\ 0 & \text{o.w.} \end{cases}$

where R is region covered by values of X & Y.

$$\iint_R f_{X,Y}(x,y)\, dA = 1$$

$$\Rightarrow \quad k \, \text{area}(R) = 1$$

$$\Rightarrow \quad k = \frac{1}{\text{area}(R)}$$

<u>Def</u>

Def^n: (1) A r.v. X is said to be uniformly distributed over the interval $[a,b]$ if its p.d.f. is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{ow} \end{cases}$$

We write $X \sim U[a,b]$.

Def^n ② The random vector (X,Y) is said to ~~be~~ have a uniform distribution over the 2-D region R if its joint p.d.f. is given by

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{area of } R}, & \text{if } (x,y) \in R \\ 0 & \text{ow} \end{cases}$$

Ex: Suppose that (X,Y) is uniformly distributed over the following rectangle R

$$R = \{(x,y) : 0 \le x \le a,\ 0 \le y \le b\}.$$

Then, X and Y are independent r.v.s s.t. $X \sim U[0,a]$, $Y \sim U[0,b]$.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^b \frac{1}{ab}\,dy = \frac{1}{a}$$

for $0 \le x \le a$.

||ly

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \frac{1}{b}, \text{ for } 0 \le y \le b.$$

Ex: The current in a semiconductor diode is often measured by the Schockley equation:

$$I = I_0 \left[e^{aV} - 1 \right] = g(v)$$

where V is the voltage across the diode, I_0 is the reverse current; 'a' is a constant; and I is the resulting diode current. Find $E[I]$ if $a = 5$, $I_0 = 10^{-6}$ and $V \sim U[1, 3]$.

$$f_V(v) = \begin{cases} \frac{1}{2}, & 1 \le v \le 3 \\ 0 & \text{o.w.} \end{cases}$$

$$E[g(v)] = \int_{-\infty}^{\infty} g(v) f_V(v) \, dv$$

$$\approx 0.3269.$$

Normal Random Variable

A r.v. X is said to be normally distributed with parameters μ and σ^2, and we write $X \sim N(\mu, \sigma^2)$, if its p.d.f. is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad -\infty < x < \infty.$$

$\sigma > 0$.

$\boxed{\mu = 0, \sigma = 1}$ $\quad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$\int_{-\infty}^{\infty} f_X(x)\,dx \overset{?}{=} 1 \quad , \quad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{0} + \int_{0}^{\infty} \right]$$

$$= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-x^2/2}\,dx = \frac{2}{\sqrt{2}\sqrt{2\pi}} \int_{0}^{\infty} \frac{1}{\sqrt{t}} e^{-t}\,dt$$

$$x^2/2 = t$$

$$\Rightarrow x = \sqrt{2t}$$

$$\Rightarrow dx = \frac{1}{2\sqrt{2t}} \times 2\,dt = \frac{1}{\sqrt{2t}}\,dt$$

$$= \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} t^{\frac{1}{2}-1} e^{-t}\,dt$$

$$= \frac{1}{\sqrt{\pi}} \Gamma\left(\tfrac{1}{2}\right) = \frac{1}{\sqrt{\pi}} \times \sqrt{\pi} = 1.$$

Claim: $E[X] = \mu$ & $Var(X) = \sigma^2$, $X \sim N(\mu, \sigma^2)$

(i) $E[X] = \mu \iff E[X-\mu] = 0$

$\because X \sim N(\mu, \sigma^2)$

$$\Rightarrow E[(X-\mu)] = \int_{-\infty}^{\infty} (x-\mu) f_X(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$

$\boxed{\frac{x-\mu}{\sigma} = y}$ $\Rightarrow dx = \sigma\, dy.$

$$\Rightarrow E[X-\mu] = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y\, e^{-y^2/2}\, dy = \frac{\sigma}{\sqrt{2\pi}} \left[\int_{-\infty}^{0} + \int_{0}^{\infty} \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \left[\lim_{t_1 \to -\infty} \int_{t_1}^{0} y e^{-y^2/2} dy + \lim_{t_2 \to \infty} \int_{0}^{t_2} y e^{-y^2/2} dy \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \left[\lim_{t_1 \to -\infty} \left\{ -e^{-y^2/2} \right\}_{t_1}^{0} + \lim_{t_2 \to \infty} \left\{ -e^{-y^2/2} \right\}_{0}^{t_2} \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \left[\lim_{t_1 \to -\infty} \left(-1 + e^{-t_1^2/2} \right) + \lim_{t_2 \to \infty} \left(1 - e^{-t_2^2/2} \right) \right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \left[-1 + 1 \right] = 0. \Rightarrow E[X-\mu] = 0$$

$$\Rightarrow E[X] = \mu.$$

(ii) $Var(X) = E[(X-\mu)^2] = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 y^2 \exp\left(-y^2/2\right) \sigma \, dy \qquad \left[\frac{x-\mu}{\sigma} = y\right]$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{y}_{\text{I}^{st}} \underbrace{\left(y e^{-y^2/2}\right)}_{\text{II}^{nd}} dy = \sigma^2$$

$$\boxed{\int y^2 e^{-y^2/2} dy = -y e^{-y^2/2} + \int e^{-y^2/2} dy}$$

$$\boxed{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1}$$

Ex: Let $X \sim N(\mu, \sigma^2)$. Then, $Y = a + bX \sim N(a+b\mu, b^2\sigma^2)$
$(b \neq 0)$

Sol: $E[Y] = a + b\,E[X] = a + b\mu$

$Var(Y) = Var(a+bX) = b^2\sigma^2.$

$$F_Y(y) = P[Y \le y] = P[a + bX \le y]$$

$$= \begin{cases} P\left[X \le \frac{y-a}{b}\right] & \text{if } b > 0 \\ P\left[X \geq \frac{y-a}{b}\right] & \text{if } b < 0 \end{cases}$$

$$\Rightarrow F_Y(y) = \begin{cases} F_X\left(\frac{y-a}{b}\right) & \text{if } b > 0 \\ 1 - F_X\left(\frac{y-a}{b}\right) & \text{if } b < 0 \end{cases}$$

$$\Rightarrow f_Y(y) = \begin{cases} \frac{1}{b} f_X\left(\frac{y-a}{b}\right) & \text{if } b > 0 \\ -\frac{1}{b} f_X\left(\frac{y-a}{b}\right) & \text{if } b < 0 \end{cases} = \frac{1}{|b|} f_X\left(\frac{y-a}{b}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,|b|\sigma} \exp\left(-\frac{1}{2}\left(\frac{\frac{y-a}{b} - \mu}{\sigma}\right)^2\right)$$

$$= \frac{1}{\sqrt{2\pi}\,|b|\,\sigma} \exp\left\{ -\frac{1}{2} \left(\frac{y-(a+b\mu)}{|b|\sigma} \right)^2 \right\}$$

$\Rightarrow Y \sim N(a+b\mu,\ b^2\sigma^2)$.

Let $X \sim N(\mu, \sigma^2)$. Then,

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

$a = -\frac{\mu}{\sigma}$

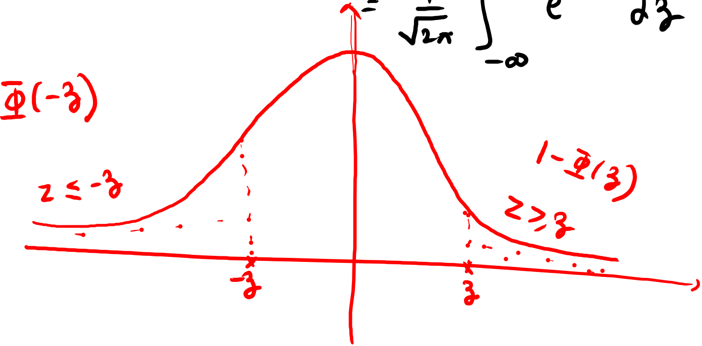
$b = \frac{1}{\sigma}$

(Z is known as Standard Normal variate).

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

$$\Phi(z) = P[Z \le z] = \int_{-\infty}^{z} \phi(z)\,dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2}\,dz$$



$$\boxed{\Phi(-z) = 1 - \Phi(z)}$$

$X \sim N(\mu, \sigma^2)$,

$$P[X < b] = P\left[\frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right]$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Ex: If $X \sim N(3, 16)$. Find $(\mu = 3, \sigma = 4)$

(a) $P[X < 11]$, (b) $P[X > -1]$, (c) $P[2 < X < 7]$

(a) $P[X < 11] = P\left[\frac{X-3}{4} < \frac{11-3}{4}\right] = P\left[\frac{X-3}{4} < 2\right] = \Phi(2)$.

(b) $P[X > -1] = P\left[\frac{X-3}{4} > \frac{-1-3}{4}\right] = P\left[\frac{X-3}{4} > -1\right] = 1 - \Phi(-1) = \Phi(1)$

(c) $P[2 < X < 7] = P\left[\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right]$

$= P\left[-0.25 < \frac{X-3}{4} < 1\right]$

$= \Phi(1) - \Phi(-0.25)$

$= \Phi(1) - [1 - \Phi(0.25)]$

$= \Phi(1) + \Phi(0.25) - 1$

$= 0.8413 + 0.5987 - 1$

$= 0.44.$

Ex: The power W dissipated in a resistor is proportional to the square of the voltage V. That is,

$$W = rV^2,$$

where 'r' is a constant. If $r=3$ and $V \sim N(6,1)$. Find

(a) $E[W]$; (b) $P[W>120]$.

(a) $E[W] = 3E[V^2]$

$= 3 \times 37 = 111$

, $Var(V) = 1$

$\Rightarrow E[V^2] - (E[V])^2 = 1$

$\Rightarrow E[V^2] = 1 + 6^2 = 37$

(b) $P[W > 120] = P[3V^2 > 120]$

$$= P[V^2 > 40]$$

$$= P[V > \sqrt{40} \text{ or } V < -\sqrt{40}]$$

$$= P[V > \sqrt{40}] + P[V < -\sqrt{40}]$$

$$= P\left[\frac{V-6}{1} > \frac{\sqrt{40}-6}{1}\right] + P\left[\frac{V-6}{1} < \frac{-\sqrt{40}-6}{1}\right]$$

$$= P\left[\frac{V-6}{1} > 0.3246\right] + P\left[\frac{V-6}{1} < -12.32\right]$$

$$= 1 - \Phi(0.3246) + \Phi(-12.3246)$$

$$= 2 - \Phi(0.3246) - \Phi(12.3246)$$

$$= 2 - 0.6255 - 1$$

$$= 1 - 0.6255$$

$$= 0.3745$$

Ex: Let X be the number of times that a fair coin that is flipped 40 times lands on heads. Find the prob. that $X = 20$.

Sol: $X \sim B(40, 0.5)$

$$P[X=20] = {}^{40}C_{20} (0.5)^{20} (0.5)^{20} \approx 0.1254.$$

Poisson $\lambda = np = 40 \times 0.5 = 20$

$$P[X=20] = e^{-20} \frac{20^{20}}{20!}$$

$$P[X=20] = \underbrace{P\left[20-\tfrac{1}{2} < X < 20+\tfrac{1}{2}\right]}_{\text{continuity correction}}$$

$$= P[19.5 < X < 20.5]$$

$$\boxed{\begin{aligned} \mu &= E[X] = np = 20 \\ \sigma^2 &= np(1-p) = 10 \end{aligned}}$$

$$P[X=20] \approx P\left[\frac{19.5-20}{\sqrt{10}} < \frac{X-20}{\sqrt{10}} < \frac{20.5-20}{\sqrt{10}}\right]$$

$$= P\left[-0.16 < \frac{X-20}{\sqrt{10}} < 0.16\right] = \Phi(0.16) - \Phi(-0.16)$$

$$= 2\Phi(0.16) - 1$$

$$= 2 \times 0.5636 - 1$$

$$= 1.1272 - 1$$

$$= 0.1272.$$

Ex: The ideal size of a 1st year class at a particular college is 150 students. The college knowing from past experience that, on the average, only 30% of those accepted for admission will actually attend, uses a policy of approving the application of 450 students. Compute the prob. that more than 150 students attend this college.

$$X \sim B(450, 0.3)$$

$$P[X > 150] = P[X > 150.5]$$

$$\mu = 450 \times 0.3 = 135$$

$$\sigma^2 = 450 \times 0.3 \times 0.7$$

$$= 94.5$$

$$P[X > 150.5] = P\left[\frac{X-135}{\sqrt{94.5}} > \frac{150.5-135}{\sqrt{94.5}}\right]$$

$$\approx P[Z > 1.5945]$$

$$= 1 - \Phi(1.5945)$$

$$= 1 - 0.9452$$

$$= 0.0548$$

<u>Ex:</u> Let $X \sim N(\mu, \sigma^2)$. Then, show that

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Ex: Sum of independent normal r.vs is also a normal r.v. That is, if $X_i \sim N(\mu_i, \sigma_i^2)$, $i=1,\dots,n$, and X_i's are independent, then

$$X = \sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2)$$

where $\mu = \sum_{i=1}^{n} \mu_i$ & $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

Ex: Data from the National Oceanic and Atmospheric Administration (NOAA) indicate that the yearly precipitation in Los Angles is a normal r.v. with a mean of 12.08 inches and a S.D. of 3.1 inches.

(a) Find the prob. that the total precipitation during the next 2 years will exceed 25 inches.

(b) Find the prob. that next years precipitation will exceed that of the following year by more than 3 inches.

Assume that the precipitation totals for the next 2 years are independent.

(a) $X = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$N(24.16, 2(3.1)^2) = N(24.16, 19.22)$

$P[X > 25] = P\left[\frac{X - 24.16}{\sqrt{19.22}} > \frac{25 - 24.16}{\sqrt{19.22}} = 0.1916\right]$

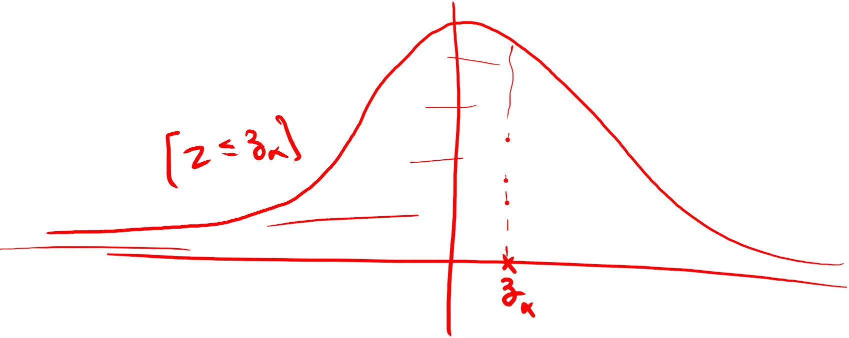
$= 1 - \Phi(0.1916) = 1 - 0.5753 = 0.4247$

(b) $Y = X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

$= N(0, 19.22)$

$P[Y > 3] = P\left[\frac{Y - 0}{\sqrt{19.22}} > \frac{3}{\sqrt{19.22}} = 0.684\right] = 1 - \Phi(0.684) = 1 - 0.7517$

$= 0.2483$



$\alpha \in (0,1)$

$\Phi(z_\alpha) = 1 - \alpha$

Memoryless Property

We say that a non-negative r.v. X is memoryless if

$$P[X>t+s \mid X>s] = P[X>t] \quad \forall\ t,s\geq 0.$$

$$\iff \quad \frac{P[X>t+s,\ X>s]}{P[X>s]} = P[X>t] \quad \forall\ t,s\geq 0$$

$$\iff \quad P[X>t+s] = P[X>t]\,P[X>s] \quad \forall\ t,s\geq 0.$$

Ex: Let $X \sim \text{Geo}(p)$. Show that X is memoryless.

$$P[X > t+s] = (1-p)^{t+s}$$

$$= (1-p)^t (1-p)^s$$

$$= P[X>t]\, P[X>s]$$

$$\forall\ t, s \in \{1, 2, 3, \ldots\}$$

$$P[X>k] = (1-p)^k$$

Exponential r.v.

A cont. r.v. X whose p.d.f. is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 . \end{cases}$$

for some $\lambda > 0$, is called exponential r.v. with parameter λ

(we write $X \sim Exp(\lambda)$)

Ex: Let $X \sim Exp(\lambda)$. Show that X is memoryless.

Sol:
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F_X(x) = P[X \leq x] = \begin{cases} 0 & \text{if } x \leq 0 \\ \int_{-\infty}^{x} f_X(x)\,dx & \text{if } x > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } x \leq 0 \\ \lambda \int_0^x e^{-\lambda x}\,dx & \text{if } x > 0 \end{cases} = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0. \end{cases}$$

$$P[X > t+s] = 1 - F_X(t+s) = e^{-\lambda(t+s)}$$

$\boxed{t, s \geq 0}$

$$= e^{-\lambda t} e^{-\lambda s}$$

$$= P[X > t]\, P[X > s] \quad \forall \; t, s \geq 0.$$

Ex: Let $X \sim Exp(\lambda)$. Find $E[X]$, $Var(X)$, and $M_X(t)$.

Sol: $M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx = \lambda \int_0^{\infty} e^{tx} e^{-\lambda x}\,dx$

$$= \lambda \int_0^{\infty} e^{-(\lambda - t)x}\,dx = \begin{cases} \frac{\lambda}{\lambda - t} & \text{if } \lambda > t \\ \text{Does not exist} & \text{if } \lambda \leq t. \end{cases}$$

$\lambda > t \geq 0$,

$$M_x(t) = \frac{\lambda}{\lambda - t}$$

$$M_x'(t) = \frac{\lambda}{(\lambda - t)^2} \Rightarrow M_x'(0) = \frac{1}{\lambda}$$

$$\Rightarrow \mu = \frac{1}{\lambda} .$$

$$M_x''(t) = \frac{2\lambda}{(\lambda - t)^3} \Rightarrow M_x''(0) = \frac{2}{\lambda^2}$$

$$\Rightarrow E[x^2] = \frac{2}{\lambda^2}$$

$$\Rightarrow Var(x) = E[x^2] - (E[x])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} .$$

Ex: If $X \sim Exp(\lambda)$, then $cX \sim Exp\left(\frac{\lambda}{c}\right)$ for any constant $c>0$.

Sol: $X \sim Exp(\lambda)$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x>0 \\ 0 & \text{ow} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{ow} \\ 1-e^{-\lambda x} & \text{if } x>0 \end{cases}$$

$Y = cX$, $c>0$.

$$F_Y(y) = P[Y \le y] = P[cX \le y] = P[X \le y/c] = F_X\left(\frac{y}{c}\right)$$

$$f_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right)$$

$$= \begin{cases} \frac{\lambda}{c} e^{-\lambda y/c} & \text{if } y > 0 \\ 0 & \text{ow.} \end{cases}$$

$$\Rightarrow Y \sim \text{Exp}\left(\frac{\lambda}{c}\right)$$

Ex: Suppose that a number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10000 miles. If a person desires to take a 5000-miles trip, what is the prob. that she will be able to complete her trip without having to replace her car battery?

$$P[X > 5] = e^{-5\lambda} = e^{-5/10} = e^{-0.5}$$

Ex: If $X_1, X_2, \ldots, X_n$ are independent exponentially distributed r.v. with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, respect-ively, then

$$Y = \min\{X_1, X_2, \ldots, X_n\} \sim Exp\left(\sum_{i=1}^{n} \lambda_i\right)$$

Sol:
$$P[Y > y] = P[\min\{X_1, \ldots, X_n\} > y]$$
$$= P[X_1 > y, X_2 > y, \ldots, X_n > y]$$
$$= \prod_{i=1}^{n} P[X_i > y] = \prod_{i=1}^{n} e^{-\lambda_i y} = e^{-\lambda y}, \quad \lambda = \sum_{i=1}^{n} \lambda_i$$

Ex: A series system is one that needs all of its components to function in order for the system itself to be functional. For an n-component series system in which the component lifetimes are independent exponential r.vs with respective parameters $\lambda_1, \ldots, \lambda_n$, what is the prob. that the system survives for a time t.

$$Y = \min\{X_1, \ldots, X_n\}$$

$$P[Y > t] = e^{-\lambda t}, \quad \lambda = \sum_{i=1}^{n} \lambda_i .$$

Gamma Distribution:

A r.v. X is said to have a Gamma Distribution with parameters (α, λ), $\alpha > 0$, $\lambda > 0$, if its p.d.f is given by

$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x} (\lambda x)^{\alpha - 1}}{\Gamma \alpha} & \text{if } x > 0 \\ 0 & \text{ow} \end{cases}$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx = \int_0^\infty \lambda (\lambda y)^{\alpha-1} e^{-\lambda y} \, dy \Rightarrow \int_{-\infty}^{\infty} f_X(x) \, dx = 1.$$

$\underline{\underline{\alpha = 1}} \Rightarrow \text{Gamm}(1, \lambda) \sim \text{Exp}(\lambda).$

<u>Ex:</u> Let $X \sim \text{Gamma}(\alpha, \lambda)$. Find its m.g.f, mean, and variance.

$$f_X(x) = \begin{cases} \dfrac{\lambda (\lambda x)^{\alpha - 1} e^{-\lambda x}}{\Gamma \alpha}, & x > 0 \\ 0 & \text{ow} \end{cases}$$

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}, \quad \lambda > t$$

~~$M_X(t)$~~ :

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$$

$$= \frac{1}{\Gamma\alpha} \int_0^{\infty} e^{tx} \left(\lambda (\lambda x)^{\alpha-1} e^{-\lambda x}\right) dx$$

$$= \frac{\lambda}{\Gamma\alpha} \int_0^{\infty} (\lambda x)^{\alpha-1} e^{-(\lambda-t)x}\,dx = \frac{\lambda^\alpha}{\Gamma\alpha} \int_0^{\infty} x^{\alpha-1} e^{-(\lambda-t)x}\,dx$$

$$(\lambda-t)x = y \Rightarrow dx = \frac{dy}{\lambda-t}$$

$$= \frac{\lambda^\alpha}{\Gamma\alpha\,(\lambda-t)^\alpha} \int_0^{\infty} y^{\alpha-1} e^{-y}\,dy = \frac{\lambda^\alpha}{(\lambda-t)^\alpha}\left[\frac{1}{\Gamma\alpha} \overbrace{\int_0^{\infty} y^{\alpha-1} e^{-y}\,dy}^{\Gamma\alpha}\right]$$

$$M_X(t) = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha}, \quad \lambda > t$$

$$E[X] = M_X'(0) = \frac{\alpha}{\lambda}$$

$$E[X^2] = \frac{\alpha(\alpha+1)}{\lambda^2}$$

$$Var(X) = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2}$$

$$= \frac{\alpha}{\lambda^2}$$

$$M_X'(t) = \alpha\left(\frac{\lambda}{\lambda-t}\right)^{\alpha-1}\frac{\lambda}{(\lambda-t)^2}$$

$$= \frac{\alpha\lambda^{\alpha}}{(\lambda-t)^{\alpha+1}}$$

$$M_X''(t) = \frac{\alpha(\alpha+1)\lambda^{\alpha}}{(\lambda-t)^{\alpha+2}}$$

Ex: Let $X_i \sim \text{Gamma}(\alpha_i, \lambda)$, $i=1,2$.

Further assume that X_1 & X_2 are independent. Then $Y = X_1 + X_2 \sim \text{Gamma}(\alpha_1+\alpha_2, \lambda)$.

$$M_Y(t) = E[e^{tY}] = E[e^{tX_1} e^{tX_2}]$$

$$= \underline{M_{X_1}(t)}\, M_{X_2}(t)$$

$$= \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1} \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_2} = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1+\alpha_2}$$

$\Rightarrow Y \sim \text{Gamma}(\alpha_1+\alpha_2, \lambda)$.

Ex: let $X_i \sim Exp(\lambda)$ $\forall$ $i = 1, .., n$. and let X_i's are independent r.vs Then,

$$X = \sum_{i=1}^{n} X_i \sim Gamma(n, \lambda).$$

The Chi-Square Distribution (χ^2_n) with n degree of freedom.

If $Z_1, .., Z_n$ are independent standard normal r.vs. then,

$$X = \sum_{i=1}^{n} Z_i^2$$

is said to have a Chi-square distribution with n. degree of freedom. We write $X \sim \chi^2_n$.

Since $Z_i \sim N(0,1)$, therefore, $E[Z_i^2] = 1 \ \forall i$.

$\Rightarrow \quad E[\chi_n^2] = E\left[\sum_{i=1}^{n} Z_i^2\right] = \sum_{i=1}^{n} E[Z_i^2] = n.$

Since Z_i's are independent r.v., then, Z_i^2 are also independent r.vs.

$\Rightarrow \quad Var(\chi_n^2) = Var\left(\sum_{i=1}^{n} Z_i^2\right) = \sum_{i=1}^{n} Var(Z_i^2)$

Let $Z \sim N(0,1)$. Find m.g.f. of $\underline{Z^2}$.

$$M_{Z^2}(t) = E\left[e^{tZ^2}\right] = \int_{-\infty}^{\infty} e^{tz^2} \phi_Z(z)\,dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz^2} e^{-z^2/2}\,dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(1-2t)z^2}{2}}\,dz$$

$$\bar{\sigma}^2 = \left(\frac{1}{1-2t}\right)$$

$$= \frac{\bar{\sigma}}{\sqrt{2\pi}\,\bar{\sigma}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2\bar{\sigma}^2}}\,dz = \bar{\sigma} = (1-2t)^{-1/2}, \quad \boxed{t < \frac{1}{2}}$$

$$\Rightarrow M_{Z^2}(t) = \frac{1}{(1-2t)^{1/2}} = \left(\frac{1/2}{\frac{1}{2}-t}\right)^{1/2} \Rightarrow Z^2 \sim \text{Gamma}\left(\alpha=\tfrac{1}{2}, \lambda=\tfrac{1}{2}\right), \; t < \frac{1}{2}.$$

Ex: Find m.g.f. of χ_n^2.

$$M_{\chi_n^2}(t) = E\left[e^{t\chi_n^2}\right] = E\left[e^{t\sum_{i=1}^n z_i^2}\right]$$

$$= \prod_{i=1}^n E\left[e^{t z_i^2}\right] = \prod_{i=1}^n M_{z_i^2}(t)$$

$$= (1-2t)^{-n/2} = \left(\frac{\frac{1}{2}}{\frac{1}{2}-t}\right)^{n/2}$$

$\Rightarrow \chi_n^2 \sim \text{Gamma}\left(\alpha = \frac{n}{2}, \lambda = \frac{1}{2}\right).$

$\Rightarrow Var(\chi_n^2) = \frac{\alpha}{\lambda^2} = \frac{n/2}{1/4} = 2n$. , $E[\chi_n^2] = n$.

Ex: Suppose that we are attempting to locate a target in 3D-space, and that the three co-ordinate errors (in meters) of the point chosen are independent normal r.vs with mean 0 and S.D. 2. Find the prob. that the distance b/w the point chosen and the target exceeds 3 meters.

$X_i \sim N(0,4)$, i^{th} co-ordinate error.

$$D^2 = \sum_{i=1}^{3} X_i^2 \Rightarrow \frac{D^2}{4} = \sum_{i=1}^{3} \left(\frac{X_i}{2}\right)^2 = \sum_{i=1}^{3} Z_i^2 \sim \chi_3^2.$$

$$P[D>3] = P\left[\frac{D^2}{4} > 2.25\right] = P\left[\chi_3^2 > 2.25\right]$$

In 2D-space

$$D^2 = \sum_{i=1}^{2} x_i^2 \Rightarrow \frac{D^2}{4} = \sum_{i=1}^{2} \left(\frac{x_i}{2}\right)^2$$

$$= \sum_{i=1}^{2} z_i^2 \sim \chi_2^2$$

$$\sim \text{Gamma}\left(\frac{2}{2}, \frac{1}{2}\right)$$

$$\sim \text{Exp}\left(\frac{1}{2}\right)$$

$$\Rightarrow P[D > 3] = P\left[\frac{D^2}{4} > 2.25\right]$$

$$= e^{-\frac{1}{2}(2.25)}$$

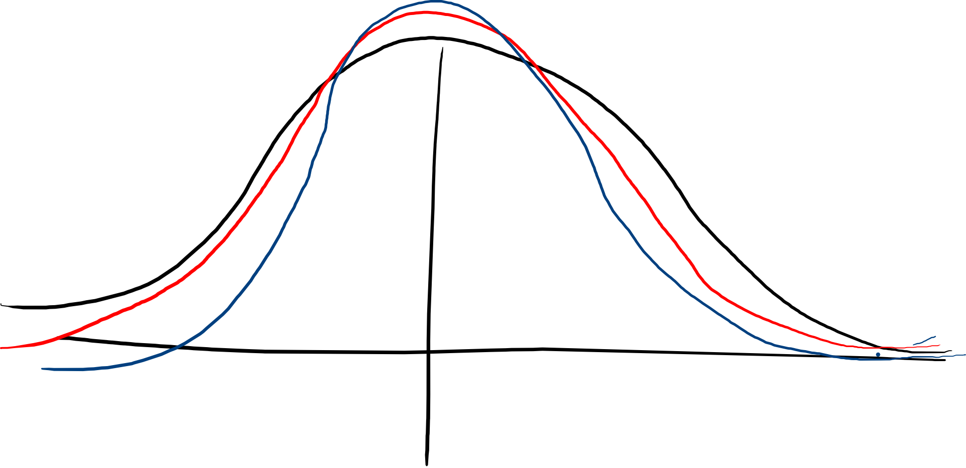
The t-distribution:

The r.v. $T_n = \dfrac{Z}{\sqrt{\frac{\chi_n^2}{n}}}$

where $Z \sim N(0,1)$ and χ_n^2 is chi-square distribution is said to have t-distribution with n-degrees of freedom.

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right] \longrightarrow 0 \text{ as } n \to \infty$$

$$\frac{\chi_n^2}{n} = \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \Rightarrow E\left[\frac{\chi_n^2}{n}\right] = 1.$$



Def: $(\chi^2_{\alpha,n})$

If $X \sim \chi^2_n$, then for any $\alpha \in (0,1)$, the quantity $\chi^2_{\alpha,n}$ is defined to be s.t.

$$P[X \geq \chi^2_{\alpha,n}] = \alpha.$$

or $$P[X \leq \chi^2_{\alpha,n}] = 1-\alpha.$$

Def: $(t_{\alpha,n})$

For $\alpha \in (0,1)$, the quantity $t_{\alpha,n}$ be such that

$$P[T_n \geq t_{\alpha,n}] = \alpha$$

$\because$ T_n & $-T_n$ shares the same distributions due to symmetry.

$$\Rightarrow \alpha = P[T_n \geq t_{\alpha,n}] = P[-T_n \geq t_{\alpha,n}]$$

$$= P[T_n \leq -t_{\alpha,n}] = 1 - P[T_n \geq -t_{\alpha,n}]$$

$$\Rightarrow \quad \alpha = 1 - P[T_n \geq -t_{\alpha,n}]$$

$$\Rightarrow \quad P[T_n \geq -t_{\alpha,n}] = 1-\alpha = P[T_n \geq t_{1-\alpha,n}]$$

$$\Rightarrow \quad \boxed{-t_{\alpha,n} = t_{1-\alpha,n}}$$

The Sample Mean ($\overline{X}$)

Let $X_1, X_2, \ldots, X_n$ be a sample of values from a certain population. The sample mean, denoted by $\overline{X}$, is defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

where X_i's are i.i.d. r.vs with mean μ & variance σ^2.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\Rightarrow E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$

$$= \frac{1}{n} \times n\mu = \mu$$

$$Var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2} \times n\sigma^2$$

$$= \frac{\sigma^2}{n}.$$

$\Rightarrow$ <u>Weak Law of Large numbers</u>

$$\Rightarrow P\big[\,|\underbrace{\bar{X}}_{\frac{1}{n}\sum_{i=1}^{n} X_i} - \mu| > \varepsilon\big] \to 0 \text{ as } n \to \infty$$

$\forall \varepsilon > 0.$

Central Limit Thm

Let $X_1, X_2, \ldots X_n, \ldots$ be a sequence of i.i.d. r.vs each having mean μ and variance σ^2. Then, for n large

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \quad \text{approximately}.$$

OR

$$P\left[\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} < x\right] \approx P[Z < x]$$

where $Z \sim N(0,1)$.

$$E[n\bar{x} = \sum_{i=1}^{n} x_i] = n\,E[\bar{x}] = n\mu.$$

$$\mathrm{Var}(n\bar{x}) = n^2\,\mathrm{Var}(\bar{x}) = \cancel{n^2} \times \frac{\sigma^2}{\cancel{n}} = n\sigma^2$$

Ex: An insurance company has 25000 automobile policy holders. If the yearly claim of a policy holder is a r.v. with mean 320 and S.D. 540, approximate the prob. that the total yearly claim exceeds 8.3 million.

$X_i \sim$ the yearly claim of a policy holder i, $i = 1, \dots, 25000$.

$\mu = 320$, $\sigma = 540$.

Total yearly claim $= X = \sum_{i=1}^{25000} X_i$

$\because$ X_i's are i.i.d. r.vs. each having mean 320 & S.D. 540,

$\therefore$ CLT $\Rightarrow$ $X \sim N(n\mu, n\sigma^2)$

$n\mu = 25000 \times 320 = 8000000 = 8 \times 10^6$

$n\sigma^2 = 25000 \times (540)^2$. $\Rightarrow$ $\sqrt{n}\,\sigma = \sqrt{25000} \times 540$

$\approx 8.5381 \times 10^4$

$\Rightarrow$ $P[X > 8.3 \times 10^6] = P\left[\frac{X - n\mu}{\sqrt{n}\,\sigma} > \frac{8.3 \times 10^6 - 8 \times 10^6}{8.5381 \times 10^4}\right]$

$\approx P[Z > 3.51] = 1 - P[Z \le 3.51] \approx 0.00023$

Ex: Civil Engineers believe that W, the amount of weight (in units of 1000 pounds) that a certain span of bridge can withstand without structural damage resulting, is normally distributed with mean 400 & S.D. 40.
Suppose that the weight of a car is a r.v. with mean 3 and S.D. 0.3. How many cars would have to be on the bridge span for the prob. of structural damage to exceed 0.1?

X_i ~ weight of ith car on the bridge span.

$X = \sum_{i=1}^{n} X_i$ ~ Total weight of ith car on the bridge span.

$P[X > W] \approx 0.1 \iff P[X - W > 0] \approx 0.1$

Clearly, X_i's are i.i.d. r.vs each having mean 3 & S.D. 0.3

CLT $\Rightarrow$ $X \sim N(3n, 0.09n)$

Given that $W \sim N(400, 1600)$

Assuming that X & W are independent r.vs.

$\Rightarrow X - W \sim N(3n - 400, 0.09n + 1600)$

$$P[X - W > 0] = 0.1$$

$$\Leftrightarrow P\left[\frac{X-W-3n+400}{\sqrt{0.09n+1600}} > \frac{400-3n}{\sqrt{0.09n+1600}}\right] = 0.1$$

$$\Leftrightarrow P\left[Z > \frac{400-3n}{\sqrt{0.09n+1600}}\right] = 0.1$$

$$\Leftrightarrow P\left[Z \le \frac{400-3n}{\sqrt{0.09n+1600}}\right] = 0.9 = P[Z \le 1.29]$$

$$\Leftrightarrow \boxed{\frac{400-3n}{\sqrt{0.09n+1600}} = 1.29}$$

Random Variable

Definition (Borel $\sigma-$Algebra)

Let J denote the class of all open intervals in $\mathbb{R}$, i.e.,

$J = \{(a, b) : -\infty \leq a < b \leq \infty\}$. The Borel $\sigma-$algebra $\mathcal{B}_1$ is the smallest $\sigma-$algebra of subsets of $\mathbb{R}$ containing J.

Random Variable

Definition (Borel $\sigma-$Algebra)

Let J denote the class of all open intervals in $\mathbb{R}$, i.e.,

$J = \{(a, b) : -\infty \leq a < b \leq \infty\}$. The Borel $\sigma-$algebra $\mathcal{B}_1$ is the smallest $\sigma-$algebra of subsets of $\mathbb{R}$ containing J.

Definition (Random Variable)

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a given function. We say that X is a random variable (r.v.) if

$X^{-1}(B) \in \mathfrak{F}, \quad \text{for all } B \in \mathcal{B}_1.$

Random Variable

Definition (Borel $\sigma-$Algebra)

Let J denote the class of all open intervals in $\mathbb{R}$, i.e., $J = \{(a, b) : -\infty \leq a < b \leq \infty\}$. The Borel $\sigma-$algebra $\mathcal{B}_1$ is the smallest $\sigma-$algebra of subsets of $\mathbb{R}$ containing J.

Definition (Random Variable)

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a given function. We say that X is a random variable (r.v.) if

$X^{-1}(B) \in \mathfrak{F}, \quad \text{for all } B \in \mathcal{B}_1.$

Note: If $\mathfrak{F} = 2^{\Omega}$ then any function $X : \Omega \to \mathbb{R}$ is a r.v.

Random Variable: Result

Theorem

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a given function. Then X is a r.v. if and only if one of the following conditions is satisfied:

(i) $X^{-1}((-\infty, a)) \in \mathfrak{F}$, *for all* $a \in \mathbb{R}$.

(ii) $X^{-1}((a, \infty)) \in \mathfrak{F}$, *for all* $a \in \mathbb{R}$.

(iii) $X^{-1}([a, \infty)) \in \mathfrak{F}$, *for all* $a \in \mathbb{R}$.

(iv) $X^{-1}((a, b]) \in \mathfrak{F}$, *whenever* $-\infty \leq a < b < \infty$.

(v) $X^{-1}([a, b)) \in \mathfrak{F}$, *whenever* $-\infty < a < b \leq \infty$.

(vi) $X^{-1}((a, b)) \in \mathfrak{F}$, *whenever* $-\infty \leq a < b \leq \infty$.

Induced Probability Measure

Definition (Induced Probability Measure)

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a r.v..

Let the set function $P_X : \mathcal{B}_1 \to \mathbb{R}$ be defined by

$$P_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}_1.$$

The probability space $(\mathbb{R}, \mathcal{B}_1, P_X)$ (prove it) is called the probability space induced by X and P_X is called the probability measure induced by X.

Example

Example

Suppose that a fair coin is independently flipped thrice. Write the sample space. Let $\mathfrak{F} = 2^{\Omega}$ and the relevant probability measure $P : \mathfrak{F} \to \mathbb{R}$ is given by

$$P(A) = \frac{\mid A \mid}{8}, \quad A \in \mathfrak{F},$$

where $\mid A \mid$ denotes the number of elements in A. Suppose that we are primarily interested in the number of times a head is observed in three flips. Write down the corresponding r.v. X and find the induced probability space $(\mathbb{R}, \mathcal{B}_1, P_X)$.

Cumulative Distribution Function

Definition (Cumulative Distribution Function (c.d.f))

The function $F_X : \mathbb{R} \to \mathbb{R}$, defined by,

$$F_X(x) = P(\{X \leq x\}) = P_X((-\infty, x]), \quad x \in \mathbb{R},$$

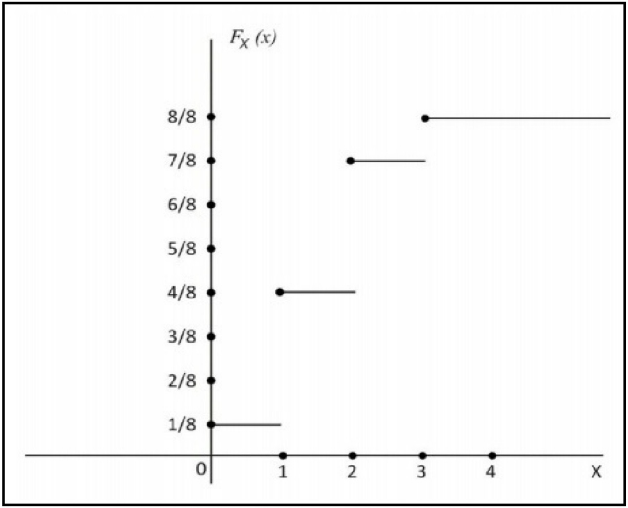
is called the distribution function of random variable X.

Example: Cumulative distribution function

The c.d.f. of the r.v. X in above example is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{8}, & \text{if } 0 \leq x < 1 \\ \frac{1}{2}, & \text{if } 1 \leq x < 2 \\ \frac{7}{8}, & \text{if } 2 \leq x < 3 \\ 1, & \text{if } x \geq 3. \end{cases}$$

Example: Plot of distribution function



Plot of distribution function $F_X(x)$

Cumulative Distribution Function: Properties

Theorem

Let F_X be the c.d.f. of a r.v. X. Then,

(i) F_X *is non-decreasing.*

(ii) F_X *is right continuous.*

(iii) $F_X(-\infty) = 0$ *and* $F_X(\infty) = 1$.

(iv) F_X *has at most countable number of discontinuities and the size of the jump at a point* $x \in \mathbb{R}$ *of discontinuity is* $F_X(x) - F_X(x-) = P(\{X = x\})$.

Discrete Random Variable

Definition

A r.v. X is said to be of **discrete type** if there exists a non-empty and countable set S_X such that

$$P(\{X = x\}) = F_X(x) - F_X(x-) > 0, \text{ for all } x \in S_X$$

and $P_X(S_X) = 1$. The set S_X is called the support of the discrete r.v. X.

Discrete Random Variable

Definition

A r.v. X is said to be of **discrete type** if there exists a non-empty and countable set S_X such that

$$P(\{X = x\}) = F_X(x) - F_X(x-) > 0, \text{ for all } x \in S_X$$

and $P_X(S_X) = 1$. The set S_X is called the support of the discrete r.v. X.

- If a r.v. X is of discrete type then $P_X(S_X^C) = 0$ and F_X is continuous at every point of S_X^C.

Discrete Random Variable

Definition

A r.v. X is said to be of **discrete type** if there exists a non-empty and countable set S_X such that

$$P(\{X = x\}) = F_X(x) - F_X(x-) > 0, \text{ for all } x \in S_X$$

and $P_X(S_X) = 1$. The set S_X is called the support of the discrete r.v. X.

- If a r.v. X is of discrete type then $P_X(S_X^C) = 0$ and F_X is continuous at every point of S_X^C.
- S_X is the set of discontinuity points of the d.f. F_X.

Probability Mass Function (p.m.f.)

Definition

Let X be a discrete r.v. with support S_X. The function

$p_X : \mathbb{R} \to \mathbb{R}$, defined by,

$$p_X(x) = \begin{cases} P(\{X = x\}), & \text{if } x \in S_X \\ 0, & \text{otherwise.} \end{cases}$$

is called the probability mass function (p.m.f.) of X.

Probability Mass Function (p.m.f.)

Definition

Let X be a discrete r.v. with support S_X. The function $p_X : \mathbb{R} \to \mathbb{R}$, defined by,

$$p_X(x) = \begin{cases} P(\{X = x\}), & \text{if } x \in S_X \\ 0, & \text{otherwise.} \end{cases}$$

is called the probability mass function (p.m.f.) of X.

Properties of p.m.f.

- $p_X(x) > 0$, for all $x \in S_X$ and $\sum_{x \in S_X} p_X(x) = 1$.

Example

Consider a r.v. X having the d.f. F_X considered in above example.

The set of discontinuity points of F_X is

$$S_X = \{0, 1, 2, 3\} \quad \text{and} \quad P_X(S_X) = 1.$$

Therefore, the r.v. X is of discrete type with support S_X and p.m.f. is

$$p_X(x) = \begin{cases} \frac{1}{8}, & \text{if } x \in \{0, 3\} \\ \frac{3}{8}, & \text{if } x \in \{1, 2\} \\ 0, & \text{otherwise.} \end{cases}$$

Continuous Random Variable

Definition

A r.v. X is said to be of **continuous type** if there exists a non-negative and integrable function $f_X : \mathbb{R} \to \mathbb{R}$, having the property that for any set $B \subseteq \mathbb{R}$

$$P(\{X \in B\}) = \int_B f_X(x)dx.$$

Continuous Random Variable

Definition

A r.v. X is said to be of **continuous type** if there exists a non-negative and integrable function $f_X : \mathbb{R} \to \mathbb{R}$, having the property that for any set $B \subseteq \mathbb{R}$

$$P(\{X \in B\}) = \int_B f_X(x)dx.$$

- The set $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ is called the support of the continuous r.v. X and f_X is called the probability density function of X.

Continuous Random Variable: Example

Example

Suppose that X is a continuous r.v. whose p.d.f. is given by

$$f_X(x) = \begin{cases} c(4x - 2x^2), & \text{if } 0 < x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) What is the value of c? (b) Find $P\{X > 1\}$.

Important Remarks

(i) If $h : \mathbb{R} \to \mathbb{R}$ is a non-negative integrable function then, for any non-empty countable set D in $\mathbb{R}$, and for $-\infty \leq a < b \leq \infty$,

$$\int_a^b h(t) 1_D(t) dt = 0.$$

Important Remarks

(i) If $h : \mathbb{R} \to \mathbb{R}$ is a non-negative integrable function then, for any non-empty countable set D in $\mathbb{R}$, and for $-\infty \leq a < b \leq \infty$,

$$\int_a^b h(t) 1_D(t) dt = 0.$$

(ii) Let E be any countable set in $\mathbb{R}$ and let $g : \mathbb{R} \to [0, \infty)$ be any non-negative function such that

$$g(x) = f_X(x), \text{ for all } x \in \overline{E} \quad \text{and} \quad g(x) \neq f_X(x), \text{ for all } x \in \overline{E}.$$

Then, g is also a p.d.f. of r.v. X.

Important Remarks

(iii) Suppose that the c.d.f. F_X of a r.v. X is differentiable at every $x \in \mathbb{R}$. Then, $f_X(x) = F'_X(x), \ x \in \mathbb{R}$.

Important Remarks

(iii) Suppose that the c.d.f. F_X of a r.v. X is differentiable at every $x \in \mathbb{R}$. Then, $f_X(x) = F'_X(x),\ x \in \mathbb{R}$.

(iv) Suppose that the c.d.f. F_X of a r.v. X is differentiable everywhere except on countable set D and

$$\int_{-\infty}^{\infty} F'_X(t) 1_{\overline{D}}(t) dt = 1.$$

Then, one may take a p.d.f. of X as

$$f_X(x) = \begin{cases} F'_X(x), & \text{if } x \in \overline{D} \\ a_x, & \text{if } x \in D, \end{cases}$$

where a_x, $x \in D$ are arbitrary nonnegative constants.

Example

Example

Consider a r.v. X having the c.d.f. F_X given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{x}{4}, & \text{if } 0 \leq x < 1 \\ \frac{x}{3}, & \text{if } 1 \leq x < 2 \\ \frac{3x}{8}, & \text{if } 2 \leq x < \frac{5}{2} \\ 1, & \text{if } x \geq \frac{5}{2}. \end{cases}$$

Show that the r.v. X is neither discrete nor continuous type.

Example

Example

Consider a r.v. X having the c.d.f. F_X given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{x^2}{2}, & \text{if } 0 \leq x < 1 \\ \frac{x}{2}, & \text{if } 1 \leq x < 2 \\ 1, & \text{if } x \geq 2. \end{cases}$$

Find a p.d.f. of X.

Example

Example

Consider a r.v. X having the c.d.f. F_X given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-x}, & \text{if } x \geq 0. \end{cases}$$

Find a p.d.f. of X.

Applications

Discrete r.v.

(i) Number of heads in 4 flips of a coin (possible outcomes are 0, 1, 2, 3, 4).

(ii) Number of classes missed last week (possible outcomes are 0, 1, 2, 3, ..., up to the maximum number of classes).

(iii) Amount won or lost when betting $1 on some Daily number lottery.

Applications

Discrete r.v.

(i) Number of heads in 4 flips of a coin (possible outcomes are 0, 1, 2, 3, 4).

(ii) Number of classes missed last week (possible outcomes are 0, 1, 2, 3, ..., up to the maximum number of classes).

(iii) Amount won or lost when betting $1 on some Daily number lottery.

Continuous r.v.

- Heights of individuals, Time to finish a test, Hours spent exercising last week.

Function of a Random Variable

Function of a Random Variable

Definition (Borel Function)

A function $h : \mathbb{R} \to \mathbb{R}$ is said to be a Borel function if

$h^{-1}(B) \in \mathcal{B}_1$, for all $B \in \mathcal{B}_1$.

Function of a Random Variable

Definition (Borel Function)

A function $h : \mathbb{R} \to \mathbb{R}$ is said to be a Borel function if

$h^{-1}(B) \in \mathcal{B}_1$, for all $B \in \mathcal{B}_1$.

Note: Every continuous real-valued function is a Borel function.

Function of a Random Variable

Definition (Borel Function)

A function $h : \mathbb{R} \to \mathbb{R}$ is said to be a Borel function if

$h^{-1}(B) \in \mathcal{B}_1$, for all $B \in \mathcal{B}_1$.

Note: Every continuous real-valued function is a Borel function.

Theorem

Let X be a r.v. defined on a probability space $(\Omega, \mathfrak{F}, P)$ and let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function. Then the function $Z : \Omega \to \mathbb{R}$, defined by $Z(w) = h(X(w))$, $w \in \Omega$, is a r.v.

Example

Example

Let X be a r.v. with p.m.f.

$$p_X(x) = \begin{cases} \frac{1}{7}, & \text{if } x \in \{-2, -1, 0, 1\} \\ \frac{3}{14}, & \text{if } x \in \{2, 3\} \\ 0, & \text{otherwise.} \end{cases}$$

Show that $Z = X^2$ is a r.v.. Find its p.m.f. and c.d.f.

Example

Example

Let X be a r.v. with p.d.f.

$$f_X(x) = \begin{cases} e^{-x}, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Show that $T = X^2$ is a r.v.. Find its p.d.f. and c.d.f.

Example

Example

Let X be a r.v. with p.d.f.

$$f_X(x) = \begin{cases} \frac{|x|}{2}, & \text{if } -1 < x < 1 \\ \frac{x}{3}, & \text{if } 1 \le x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

Show that $T = \mid X \mid$ is a r.v.. Find its p.d.f. and c.d.f.

Expectation of a r.v.

Expectation of a r.v.

(i) Let X be a discrete r.v. with p.m.f. p_X and support S_X. We say that the expected value of X (denoted by $E(X)$) is finite and equals

$$E(X) = \sum_{x \in S_X} x p_X(x), \quad \text{provided} \quad \sum_{x \in S_X} \mid x \mid p_X(x) < \infty.$$

(ii) Let X be a continuous r.v. with p.d.f. f_X. We say that the expected value of X (denoted by $E(X)$) is finite and equals

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx, \quad \text{provided} \quad \int_{-\infty}^{\infty} \mid x \mid f_X(x) dx < \infty.$$

Example

Example

Let X be a r.v. with p.m.f.

$$p_X(x) = \begin{cases} \left(\frac{1}{2}\right)^x, & \text{if } x \in \{1, 2, 3, ...\} \\ 0, & \text{otherwise.} \end{cases}$$

Show that the expected value of X is finite and find its value.

Example

Example

Let X be a r.v. with p.m.f.

$$p_X(x) = \begin{cases} \frac{3}{\pi^2 x^2}, & \text{if } x \in \{\pm 1, \pm 2, \pm 3, ...\} \\ 0, & \text{otherwise.} \end{cases}$$

Show that the expected value of X is not finite.

Example

Example

Let X be a r.v. with p.d.f.

$$f_X(x) = \frac{e^{-|x|}}{2}, \quad -\infty < x < \infty.$$

Show that the expected value of X is finite and find its value.

Example

Example

The time, in hours, it takes to locate and repair an electrical breakdown in a certain factory is a r.v., call it X, whose density function is given by

$$f_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

If the cost involved in a breakdown of duration x is x^3, what is the expected cost of such a breakdown?

Expectation of a function of r.v.

(i) Let X be a discrete r.v. with p.m.f. p_X and support S_X. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function and let $T = h(X)$. Then,

$$E(T) = \sum_{x \in S_X} h(x) p_X(x), \quad \text{provided it is finite.}$$

Expectation of a function of r.v.

(i) Let X be a discrete r.v. with p.m.f. p_X and support S_X. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function and let $T = h(X)$. Then,

$$E(T) = \sum_{x \in S_X} h(x) p_X(x), \quad \text{provided it is finite.}$$

(ii) Let X be a continuous r.v. with p.d.f. f_X. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function and let $T = h(X)$. Then,

$$E(T) = \int_{-\infty}^{\infty} h(x) f_X(x) dx, \quad \text{provided it is finite.}$$

Properties of Expectation

(i) If $E(\mid X \mid) < \infty$, then $\mid E(X) \mid \leq E(\mid X \mid)$.

Properties of Expectation

(i) If $E(|X|) < \infty$, then $|E(X)| \leq E(|X|)$.

(ii) For real constants a and b,

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

provided the involved expectations are finite.

Properties of Expectation

(i) If $E(\mid X \mid) < \infty$, then $\mid E(X) \mid \leq E(\mid X \mid)$.

(ii) For real constants a and b,

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

provided the involved expectations are finite.

(iii) If $h_1, h_2, ..., h_m$ are Borel functions then

$$E\left(\sum_{i=1}^{m} h_i(X)\right) = \sum_{i=1}^{m} E(h_i(X)),$$

provided the involved expectations are finite.

Special Kinds of Expectations

(i) $\mu = \mu_1' = E(X)$, provided it is finite, is called the mean of the r.v. X.

Special Kinds of Expectations

(i) $\mu = \mu_1^{'} = E(X)$, provided it is finite, is called the mean of the r.v. X.

(ii) For $r \in \{1, 2, ...\}$, $\mu_r^{'} = E(X^r)$, provided it is finite, is called the $r-$th moment of the r.v. X.

Special Kinds of Expectations

(i) $\mu = \mu_1^{'} = E(X)$, provided it is finite, is called the mean of the r.v. X.

(ii) For $r \in \{1, 2, ...\}$, $\mu_r^{'} = E(X^r)$, provided it is finite, is called the $r-$th moment of the r.v. X.

(iii) For $r \in \{1, 2, ...\}$, $\mu_r = E\left(\left(X - \mu_1^{'}\right)^r\right)$, provided it is finite, is called the $r-$th central moment of the r.v. X.

Special Kinds of Expectations

(i) $\mu = \mu_1^{'} = E(X)$, provided it is finite, is called the mean of the r.v. X.

(ii) For $r \in \{1, 2, ...\}$, $\mu_r^{'} = E(X^r)$, provided it is finite, is called the $r-$th moment of the r.v. X.

(iii) For $r \in \{1, 2, ...\}$, $\mu_r = E\left(\left(X - \mu_1^{'}\right)^r\right)$, provided it is finite, is called the $r-$th central moment of the r.v. X.

(iv) $\mu_2 = Var(X) = E\left(\left(X - \mu_1^{'}\right)^2\right)$, provided it is finite, is called the variance of the r.v. X.

Special Kinds of Expectations

(i) $\mu = \mu_1^{'} = E(X)$, provided it is finite, is called the mean of the r.v. X.

(ii) For $r \in \{1, 2, ...\}$, $\mu_r^{'} = E(X^r)$, provided it is finite, is called the $r-$th moment of the r.v. X.

(iii) For $r \in \{1, 2, ...\}$, $\mu_r = E\left(\left(X - \mu_1^{'}\right)^r\right)$, provided it is finite, is called the $r-$th central moment of the r.v. X.

(iv) $\mu_2 = Var(X) = E\left(\left(X - \mu_1^{'}\right)^2\right)$, provided it is finite, is called the variance of the r.v. X.

(v) The quantity $\sigma = \sqrt{\mu_2} = \sqrt{E((X-\mu)^2)}$ is called the standard deviation of the r.v. X.

Remarks

- $Var(X) = E(X^2) - (E(X))^2$ and $Var(X) \geq 0$.
- $E(X^2) \geq (E(X))^2$ (Cauchy-Schwarz inequality).
- $Var(aX + b) = a^2 Var(X)$.

Example

Example

Let X be a r.v. with p.d.f.

$$f_X(x) = \begin{cases} \frac{1}{2}, & \text{if } -2 < x < -1 \\ \frac{x}{9}, & \text{if } 0 < x < 3 \\ 0, & \text{otherwise.} \end{cases}$$

Find the mean and variance of the r.v. $Y = \max\{X, 0\}$.

Moment Generating Function (m.g.f.)

Definition

Let X be a r.v. and let $A = \{t \in \mathbb{R} : \; E(e^{tX}) \text{ is finite}\}$. Define $M_X : A \to \mathbb{R}$ by

$$M_X(t) = E(e^{tX}), \; t \in A.$$

(i) We call the function $M_X(.)$, the moment generating function of r.v. X.

(ii) We say that the m.g.f. of a r.v. X exists if there exists a positive real number a such that $(-a, a) \subseteq A$ (i.e., if $M_X(.)$ is finite in an interval containing 0).

m.g.f. versus moments

Theorem

Let X be a r.v. with m.g.f. M_X that is finite on an interval $(-a, a)$, for some $a > 0$. Then,

(i) for each $r \in \{1, 2, ...\}$, $\mu_r^{'} = E(X^r)$ is finite;

(ii) for each $r \in \{1, 2, ...\}$, $\mu_r^{'} = E(X^r) = M_X^{(r)}(0)$, where $M_X^{(r)}(0)$ is the $r-$th derivative of $M_X(t)$ at the point 0;

(iii) $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu_r^{'}$, $t \in (-a, a)$.

Example

Example

Let X be a r.v. with p.m.f.

$$p_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & \text{if } x \in \{1, 2, \dots\} \\ 0, & \text{otherwise} \end{cases}$$

where $\lambda > 0$. Find the m.g.f. M_X of the r.v. X. Show that X possesses moments of all orders. Find the mean and variance of X.

Random Variables Having the Same Distribution

Definition

Two random variables X and Y, defined on the same probability space $(\Omega, \mathfrak{F}, P)$, are said to have the same distribution (written as $X \stackrel{\mathrm{d}}{=} Y$) if they have the same distribution function, i.e., if $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$.

Random Variables Having the Same Distribution

Definition

Two random variables X and Y, defined on the same probability space $(\Omega, \mathfrak{F}, P)$, are said to have the same distribution (written as $X \stackrel{\mathrm{d}}{=} Y$) if they have the same distribution function, i.e., if $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$.

Important Remark

Let X and Y be two r.vs., of either discrete type or of absolute continuous type, having the same distribution. Then, for any Borel function h, $h(X) \stackrel{\mathrm{d}}{=} h(Y)$ and $E(h(X)) = E(h(Y))$, provided the expectations are finite.

Example 1

Example

Let X be a r.v. with p.m.f.

$$p_X(x) = \begin{cases} {}^{n}C_x \left(\frac{1}{2}\right)^n, & \text{if } x \in \{0, 1, 2, ..., n\}, \\ 0, & \text{otherwise} \end{cases}$$

where n is a given positive integer. Let $Y = n - X$. Show that $X \overset{\mathrm{d}}{=} Y$ and hence show that $E(X) = \frac{n}{2}$.

Example 2

Example

Let X be a r.v. with p.d.f.

$$f_X(x) = \frac{e^{-|x|}}{2}, \ -\infty < x < \infty,$$

and let $Y = -X$. Show that $X \overset{\mathrm{d}}{=} Y$ and hence show that $E(X^{2n+1}) = 0, \ n \in \{0, 1, 2, 3, ...\}$.

Symmetric Distribution

Definition

A r.v. X is said to have a symmetric distribution about a point $\mu \in \mathbb{R}$ if $X - \mu \overset{d}{=} \mu - X$.

Symmetric Distribution

Definition

A r.v. X is said to have a symmetric distribution about a point $\mu \in \mathbb{R}$ if $X - \mu \stackrel{\mathrm{d}}{=} \mu - X$.

Remark

- The distribution of X is symmetric about μ if and only if $f_X(\mu - x) = f_X(\mu + x)$, for all $x \in \mathbb{R}$.

Symmetric Distribution

Definition

A r.v. X is said to have a symmetric distribution about a point $\mu \in \mathbb{R}$ if $X - \mu \overset{\mathrm{d}}{=} \mu - X$.

Remark

- The distribution of X is symmetric about μ if and only if $f_X(\mu - x) = f_X(\mu + x)$, for all $x \in \mathbb{R}$.
- If the distribution of X is symmetric about μ and the expected value of X is finite, then $E(X) = \mu$.

Symmetric Distribution

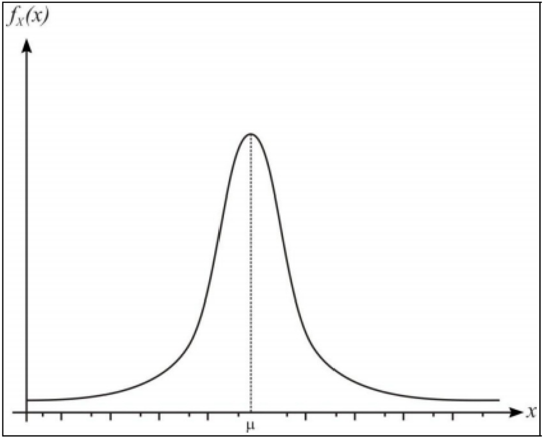
Definition

A r.v. X is said to have a symmetric distribution about a point $\mu \in \mathbb{R}$ if $X - \mu \overset{\mathrm{d}}{=} \mu - X$.

Remark

- The distribution of X is symmetric about μ if and only if $f_X(\mu - x) = f_X(\mu + x)$, for all $x \in \mathbb{R}$.
- If the distribution of X is symmetric about μ and the expected value of X is finite, then $E(X) = \mu$.
- If the distribution of X is symmetric about 0, then $E(X^{2m+1}) = 0$.

Symmetric Distribution about mean



Symmetric Distribution about mean

Probability and Moment Inequalities

Probability and Moment Inequalities

Theorem

Let X be a r.v. and let $g : [0, \infty) \to \mathbb{R}$ be a non-negative and non decreasing function such that the expected value of $g(X)$ is finite. Then, for any $c > 0$ for which $g(c) > 0$,

$$P(\{\mid X \mid \geq c\}) \leq \frac{E\left(g(\mid X \mid)\right)}{g(c)}.$$

Inequalities

Theorem (Markov Inequality)

Suppose that $E(|X|^r) < \infty$*, for some* $r > 0$*. Then, for any* $c > 0$*,*

$$P(\{|X| \geq c\}) \leq \frac{E(|X|^r)}{c^r}.$$

Inequalities

Theorem (Markov Inequality)

Suppose that $E(|X|^r) < \infty$, *for some* $r > 0$. *Then, for any* $c > 0$,

$$P(\{|X| \geq c\}) \leq \frac{E(|X|^r)}{c^r}.$$

Theorem (Chebyshev Inequality)

Suppose that X *has finite first two moments. If* $\mu = E(X)$ *and* $\sigma^2 = Var(X)(\sigma \geq 0)$. *Then, for any* $c > 0$,

$$P(\{|X - \mu| \geq c\}) \leq \frac{\sigma^2}{c^2}.$$

Example

Example

Let X be a r.v. with zero mean and $E(X^2) = \frac{1}{4}$. What can you say about the probability $P(\{| X | \geq 1\})$?

Median

Definition

A real number m satisfying

$$F_X(m-) \leq \frac{1}{2} \leq F_X(m),$$

is called the median of the r.v. X.

Median

Definition

A real number m satisfying

$$F_X(m-) \leq \frac{1}{2} \leq F_X(m),$$

is called the median of the r.v. X.

- If X is of continuous type, then the median m is given by $F_X(m) = \frac{1}{2}$.

Median

Definition

A real number m satisfying

$$F_X(m-) \leq \frac{1}{2} \leq F_X(m),$$

is called the median of the r.v. X.

- If X is of continuous type, then the median m is given by $F_X(m) = \frac{1}{2}$.
- For some distributions, it may happen that $F_X(a-) < \frac{1}{2}$ and $\{x \in \mathbb{R} : F_X(x) \geq 1/2\} = [a, b)$ for some $-\infty < a < b < \infty$, so that the median is not unique.

Median: Advantages and Disadvantages

- Unlike the mean, the median of a r.v. X is always defined.

Median: Advantages and Disadvantages

- Unlike the mean, the median of a r.v. X is always defined.
- It is not affected by a few extreme values as it takes into account only the probabilities.

Median: Advantages and Disadvantages

- Unlike the mean, the median of a r.v. X is always defined.
- It is not affected by a few extreme values as it takes into account only the probabilities.
- It is preferred over the mean if the distribution is asymmetric and a few extreme observations are assigned with positive probabilities.

Median: Advantages and Disadvantages

- Unlike the mean, the median of a r.v. X is always defined.
- It is not affected by a few extreme values as it takes into account only the probabilities.
- It is preferred over the mean if the distribution is asymmetric and a few extreme observations are assigned with positive probabilities.
- One of its demerits is that it does not at all take into account the numerical values of X.

Median: Advantages and Disadvantages

- Unlike the mean, the median of a r.v. X is always defined.
- It is not affected by a few extreme values as it takes into account only the probabilities.
- It is preferred over the mean if the distribution is asymmetric and a few extreme observations are assigned with positive probabilities.
- One of its demerits is that it does not at all take into account the numerical values of X.
- Another disadvantage with it is that for many r.vs. it is not easy to evaluate.

Mode

The mode m_0 of a probability distribution is the value that occurs with highest probability and is defined by

$$f_X(m_0) = \sup\{f_X(x) : x \in S_X\}.$$

Mode

The mode m_0 of a probability distribution is the value that occurs with highest probability and is defined by

$$f_X(m_0) = \sup\{f_X(x) : x \in S_X\}.$$

- A probability distribution may have more than one mode which may be far apart.

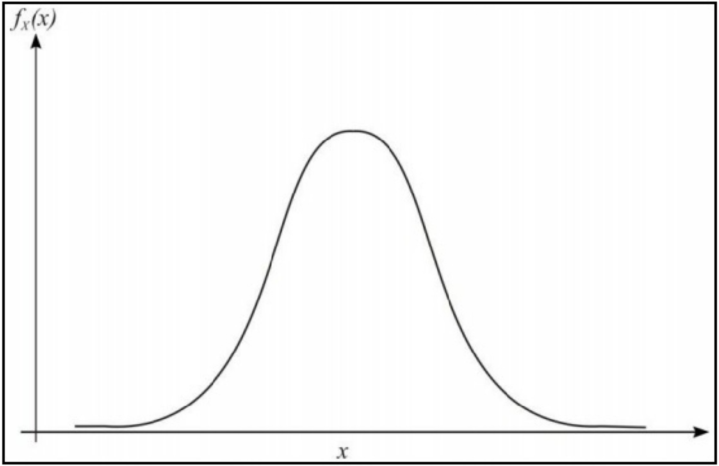
Mode

The mode m_0 of a probability distribution is the value that occurs with highest probability and is defined by

$$f_X(m_0) = \sup\{f_X(x) : x \in S_X\}.$$

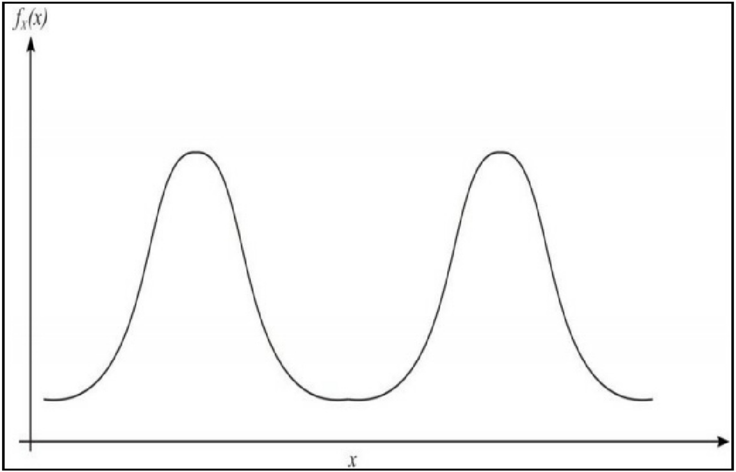
- A probability distribution may have more than one mode which may be far apart.
- It does not take into account the numerical values of X as well as their associated probabilities.

Unimodal Distribution



Unimodal distribution

Bimodal Distribution



Bimodal distribution

Skewness: A measure of asymmetry

Definition

A measure of skewness of the probability distribution of X is defined by

$$\beta_1 = \frac{E(X-\mu)^3}{\sigma^3} = \frac{\mu_3^{'}}{\sigma^3}.$$

Skewness: A measure of asymmetry

Definition

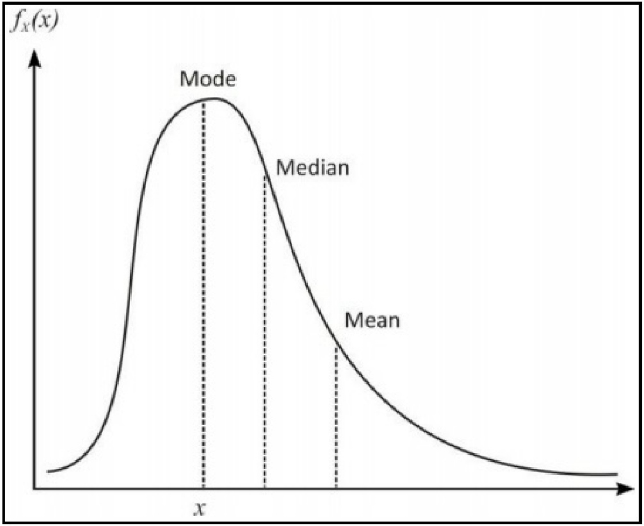
A measure of skewness of the probability distribution of X is defined by

$$\beta_1 = \frac{E(X-\mu)^3}{\sigma^3} = \frac{\mu_3^{'}}{\sigma^3}.$$

(i) A large positive value of β_1 indicates that the data is positively skewed.

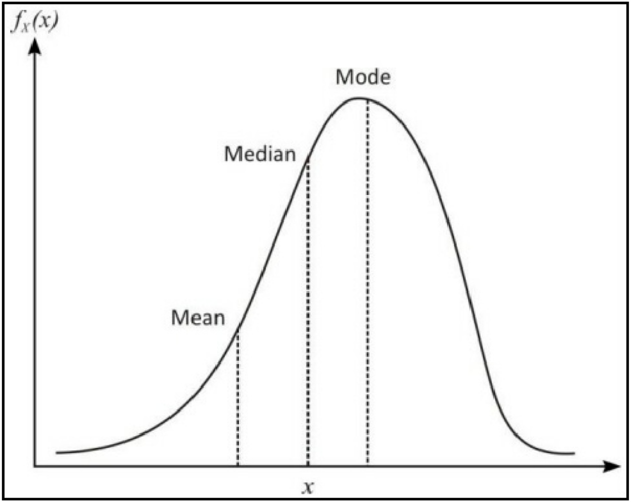
(ii) A small negative value of β_1 indicates that the data is negatively skewed.

Positively Skewed Distribution



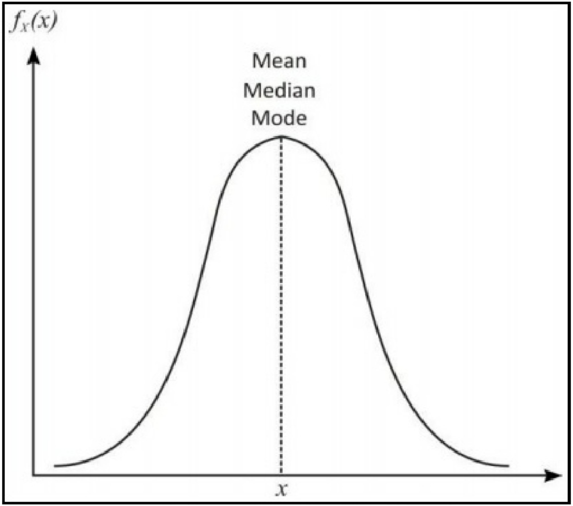
Positively skewed distribution

Negatively Skewed Distribution



Negatively Skewed Distribution

Normal (no skew) Distribution



Normal (no skew) distribution

Random Vectors

Random Vectors

Definition

Let $(\Omega, \mathfrak{F}, P)$ be a probability space. A function $\underline{X} = (X_1, X_2, ..., X_p) : \Omega \to \mathbb{R}^p$ is called a random vector if

$$\underline{X}^{-1}((-\underline{\infty}, \underline{a}]) \in \mathfrak{F}, \text{ for all } \underline{a} = (a_1, a_2, ..., a_p) \in \mathbb{R}^p.$$

Random Vectors

Definition

Let $(\Omega, \mathfrak{F}, P)$ be a probability space. A function

$\underline{X} = (X_1, X_2, ..., X_p) : \Omega \to \mathbb{R}^p$ is called a random vector if

$$\underline{X}^{-1}((-\underline{\infty}, \underline{a}]) \in \mathfrak{F}, \text{ for all } \underline{a} = (a_1, a_2, ..., a_p) \in \mathbb{R}^p.$$

Example

Let $A, B \subseteq \Omega$. Define $\underline{X} = (X_1, X_2) : \Omega \to \mathbb{R}^2$ by

$$X_1(w) = \begin{cases} 1, & \text{if } w \in A \\ 0, & \text{if } w \notin A \end{cases} \quad \text{and} \quad X_2(w) = \begin{cases} 1, & \text{if } w \in B \\ 0, & \text{if } w \notin B \end{cases}.$$

Show that $\underline{X}$ is a random vector.

Random Vectors

Theorem

Let $\underline{X} = (X_1, X_2, ..., X_p) : \Omega \to \mathbb{R}^p$ *be a given function. Then,* $\underline{X}$ *is a random vector if and only if* $X_1,\ X_2, ..., X_p$ $(X_i : \Omega \to \mathbb{R},\ i = 1, 2, ..., p)$ *are random variables.*

Random Vectors

Theorem

Let $\underline{X} = (X_1, X_2, ..., X_p) : \Omega \to \mathbb{R}^p$ *be a given function. Then,* $\underline{X}$ *is a random vector if and only if* $X_1,\ X_2, ..., X_p$ $(X_i : \Omega \to \mathbb{R},\ i = 1, 2, ..., p)$ *are random variables.*

- When Ω is countable, we have $\mathfrak{F} = 2^{\Omega}$, and therefore, any function $\underline{X} = (X_1, X_2, ..., X_p) : \Omega \to \mathbb{R}^p$ is a random vector.

Joint Distribution Function

Definition

The joint distribution function of a random vector

$\underline{X} = (X_1, X_2, ..., X_p) : \Omega \to \mathbb{R}^p$ is defined by

$$F_{\underline{X}}(x_1, ..., x_p) = P\left(\{w \in \Omega : X_1(w) \leq x_1, ..., X_p(w) \leq x_p\}\right),$$

for all $\underline{x} = (x_1, x_2, ..., x_p) \in \mathbb{R}^p$.

Joint Distribution Function

Definition

The joint distribution function of a random vector $\underline{X} = (X_1, X_2, ..., X_p) : \Omega \to \mathbb{R}^p$ is defined by

$$F_{\underline{X}}(x_1, ..., x_p) = P\left(\{w \in \Omega : X_1(w) \leq x_1, ..., X_p(w) \leq x_p\}\right),$$

for all $\underline{x} = (x_1, x_2, ..., x_p) \in \mathbb{R}^p$.

Properties

(i) Joint Distribution function is increasing and right continuous in each argument when other arguments are kept fixed.

Joint Distribution Function: Properties

(ii) Moreover, we have

$$\lim_{x_i \ (i=1,\dots,p) \to \infty} F_{\underline{X}}(x_1, \dots, x_p) = 1.$$

Joint Distribution Function: Properties

(ii) Moreover, we have

$$\lim_{x_i \ (i=1,...,p)\to\infty} F_{\underline{X}}(x_1, ..., x_p) = 1.$$

(iii) For each fixed $i \in \{1, 2, ..., p\}$ and fixed

$(x_1, ..., x_{i-1}, x_{i+1}, ..., x_p) \in \mathbb{R}^{p-1}$,

$$\lim_{y\to-\infty} F_{\underline{X}}(x_1, ..., x_{i-1}, y, x_{i+1}, ..., x_p) = 0.$$

(iv) **Marginal Distribution Function of** $\underline{Y} = (X_1, ..., X_k)$:

$$\lim_{x_i (i=k+1,...,p)\to\infty} F_{\underline{X}}(x_1, ..., x_p) = F_{\underline{Y}}(x_1, x_2, ..., x_k).$$

2-Dimensional Random Vector

2-Dimensional Random Vector

Let $(X, Y) : \Omega \to \mathbb{R}^2$ be a random vector.

2-Dimensional Random Vector

Let $(X, Y) : \Omega \to \mathbb{R}^2$ be a random vector.

(i) **Joint probability distribution function** of X and Y is given by

$$F_{X,Y}(x, y) = P\left(\{X \leq x, Y \leq y\}\right).$$

2-Dimensional Random Vector

Let $(X, Y) : \Omega \to \mathbb{R}^2$ be a random vector.

(i) **Joint probability distribution function** of X and Y is given by

$$F_{X,Y}(x, y) = P\left(\{X \leq x, Y \leq y\}\right).$$

(ii) **Marginal distribution function** F_X of X can be obtained from the joint d.f. F of X and Y by

$$F_{X_1}(x) = F_{X_1,X_2}(x_1, \infty).$$

2-Dimensional Random Vector

(iii) **Discrete type random vector**: We say that the random vector (X, Y) is of discrete type if both X_1 and X_2 are discrete r.vs..

2-Dimensional Random Vector

(iii) **Discrete type random vector**: We say that the random vector (X, Y) is of discrete type if both X_1 and X_2 are discrete r.vs..

(iv) **Joint p.m.f.**: In the case where X and Y are both discrete r.vs. whose possible values are, respectively, $x_1, x_2, ...$, and $y_1, y_2, ...$, we define the joint p.m.f. of (X, Y), $p_{X,Y}(x_i, y_j)$, by

$$p_{X,Y}(x_i, y_j) = P\left(\{X = x_i, Y = y_j\}\right).$$

Example

Example

Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, then determine the joint p.m.f. of X and Y.

Example: Solution

$P(X = i, Y = j)$

	0	1	2	3	P(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(Y=j)	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

2-Dimensional Random Vector of Continuous type

- **Continuous type random vector**: We say that the random vector (X, Y) is of continuous type if there exists a non-negative function $f(x, y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers

$$P((X, Y) \in C) = \iint_{(x,y)\in C} f(x, y)dxdy.$$

The function $f(x, y)$ is called the joint p.d.f. of X and Y.

2-Dimensional Random Vector of Continuous type

- **Continuous type random vector**: We say that the random vector (X, Y) is of continuous type if there exists a non-negative function $f(x, y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers

$$P((X, Y) \in C) = \iint_{(x,y)\in C} f(x, y)dxdy.$$

The function $f(x, y)$ is called the joint p.d.f. of X and Y.

(ii) $F_{X,Y}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)dxdy.$

Example

Example

The joint p.d.f. of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Compute (a) $P(X > 1, Y < 1)$; (b) $P(X < Y)$; and (c) $P(X < a)$.

Independence of Random Variables

Definition

The random variables X and Y are said to be independent if for any two sets of real numbers A and B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Independence of Random Variables Contd...

(i) The random variables X and Y are independent if and only if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

Independence of Random Variables Contd...

(i) The random variables X and Y are independent if and only if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

(ii) When X and Y are discrete random variables, then the condition of independence is equivalent to

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \text{for all } x, y.$$

Independence of Random Variables Contd...

(i) The random variables X and Y are independent if and only if $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

(ii) When X and Y are discrete random variables, then the condition of independence is equivalent to

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \quad \text{for all } x, y.$$

(iii) When X and Y are continuous random variables, then the condition of independence is equivalent to

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for all } x, y.$$

Example

Example

Suppose that X and Y are independent random variables having the common density function

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Find the density function of the random variable X/Y.

Conditional Distributions

(i) When X and Y are discrete random variables, it is natural to define the conditional p.m.f. of X given that $Y = y$, by

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

for all values of y such that $p_Y(y) > 0$.

Conditional Distributions

(i) When X and Y are discrete random variables, it is natural to define the conditional p.m.f. of X given that $Y = y$, by

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

for all values of y such that $p_Y(y) > 0$.

(iii) When X and Y are continuous random variables, then the conditional p.d.f. of X, given that $Y = y$, is given by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

for all values of y such that $f_Y(y) > 0$.

Conditional Distributions: Example 1

Example

Suppose that 15 percent of the families in a certain community have no children, 20 percent have 1, 35 percent have 2, and 30 percent have 3 children. Suppose further that each child is equally likely (and independently) to be a boy or a girl. Let B and G, respectively, denotes the number of boys and girls in a family. If a family is chosen at random from this community, then find the joint p.m.f. of B and G. Further, if we know that the family chosen has one girl, compute the conditional p.m.f. of the number of boys in the family.

Example: Solution

$P(B = i, G = j)$

	0	1	2	3	P(B=i)
0	.15	.10	.0875	.0375	.3750
1	.10	.175	.1125	0	.3875
2	.0875	.1125	0	0	.2000
3	.0375	0	0	0	.0375
P(G=j)	.3750	.3875	.2000	.0375	

Conditional Distributions: Example 2

Example

The joint p.d.f. of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}x(2-x-y), & \text{if } 0 < x < 1, 0 < y < 1, \\ 0 & \text{otherwise} \end{cases}$$

Compute the conditional density of X, given that $Y = y$, where $0 < y < 1$.

Expectation of Function of 2D - Random Vector

- If X and Y are random variables and g is a function of X and Y, then

$$E[g(X,Y)] = \begin{cases} \sum_y \sum_x g(x,y) p_{X,Y}(x,y) & \text{in the discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) & \text{in the continuous case} \end{cases}$$

Expectation of Function of 2D - Random Vector

- If X and Y are random variables and g is a function of X and Y, then

$$E[g(X,Y)] = \begin{cases} \sum_y \sum_x g(x,y) p_{X,Y}(x,y) & \text{in the discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) & \text{in the continuous case} \end{cases}$$

Example

A construction firm has recently sent in bids for 3 jobs worth (in profits) 10, 20, and 40 (thousand) dollars. If its probability of winning the jobs are respectively .2, .8 and .3, what is the firm's expected total profit?

Expectation contd...

Example

A secretary has typed N letters along with their respective envelopes. The envelops get mixed up when they fall on the floor. If the letters are placed in the mixed-up envelops in a completely random manner (that is, each letter is equally likely to end up in any of the envelopes), what is the expected number of letters that are placed in the correct envelopes?

Covariance of two random variables

Definition

The **covariance** of two random variables X and Y, written $\text{Cov}(X, Y)$, is defined by

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

where μ_X and μ_Y are the means of X and Y, respectively.

Covariance of two random variables

Definition

The **covariance** of two random variables X and Y, written $\mathrm{Cov}(X, Y)$, is defined by

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

where μ_X and μ_Y are the means of X and Y, respectively.

Remarks

(i) $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

Covariance of two random variables

Definition

The **covariance** of two random variables X and Y, written $\mathrm{Cov}(X, Y)$, is defined by

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

where μ_X and μ_Y are the means of X and Y, respectively.

Remarks

(i) $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

(ii) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

Covariance of two random variables: Properties

(i) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

Covariance of two random variables: Properties

(i) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

(ii) $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.

Covariance of two random variables: Properties

(i) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

(ii) $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.

(iii) $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathrm{Cov}(X_i, X_j)$.

Covariance of two random variables: Properties

(i) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

(ii) $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.

(iii) $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathrm{Cov}(X_i, X_j)$.

(iv) If X and Y are independent random variables, then

$\mathrm{Cov}(X, Y) = 0$.

Covariance and Variance: Examples

Example

Compute the variance of the number of heads resulting from 10 independent tosses of a fair coin.

Covariance and Variance: Examples

Example

Compute the variance of the number of heads resulting from 10 independent tosses of a fair coin.

Example

Compute the variance of the sum obtained when 10 independent rolls of a fair die are made.

Correlation between X and Y

(i) The correlation between random variables X and Y, a dimensionless quantity, obtained by dividing the covariance by the product of the standard deviations of X and Y. That is,

$$\rho_{X,Y} = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Correlation between X and Y

(i) The correlation between random variables X and Y, a dimensionless quantity, obtained by dividing the covariance by the product of the standard deviations of X and Y. That is,

$$\rho_{X,Y} = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

(ii) It can be shown that $-1 \leq \rho_{X,Y} \leq 1$.

Example

The joint PMF of a discrete random vector (X_1, X_2) is given by the following table

$x_2 \setminus x_1$	-1	0	1
0	$1/9$	$2/9$	$1/9$
1	$1/9$	$2/9$	$1/9$
2	0	$1/9$	0

a) Find the expectation and variance of the random variables X_1, X_2, $X_1 + X_2$ and $X_1 X_2$.

Example

b) Determine the covariance of X_1 and X_2.

Example

b) Determine the covariance of X_1 and X_2.

c) Calculate the correlation coefficient $\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$ of X and Y.

Example

b) Determine the covariance of X_1 and X_2.

c) Calculate the correlation coefficient $\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$ of X and Y.

d) Are X and Y independent random variables? Justify your answer.

Transformation of Random Vectors: Discrete Case

Example

Consider the example above. Define a new random vector

$$Y = (Y_1, Y_2) = g(X_1, X_2) = (X_1 + X_2, X_1 - X_2).$$

Find the joint p.m.f. of random vector Y.

Solution

The joint p.m.f. of random vector Y is given by the following table

$y_2 \setminus y_1$	-1	0	1	2	3
-3	0	0	0	0	0
-2	0	1/9	0	1/9	0
-1	1/9	0	2/9	0	0
0	0	2/9	0	1/9	0
1	0	0	1/9	0	0

Transformation of Random Vectors: Continuous Case

Theorem

Let X and Y be two continuous random variables, and let $V = g_1(X, Y)$ and $W = g_2(X, Y)$. Suppose that

1) *the system*

$$\begin{cases} v = g_1(x, y) \\ w = g_2(x, y) \end{cases}$$

has a unique solution $x = h_1(v, w), y = h_2(v, w)$;

Transformation of Random Vectors: Continuous Case

2) the functions g_1 and g_2 have continuous partial derivatives, for all (x,y), and the Jacobian of the transformation $J(x,y) = \frac{\partial(g_1,g_2)}{\partial(x,y)}$ is different from zero, for all (x,y). Then, we have

$$f_{V,W}(v,w) = f_{X,Y}(x,y) \mid J(x,y) \mid^{-1},$$

where $x = h_1(v,w), y = h_2(v,w)$.

Transformation of Random Vectors: Continuous Case

Example

Let $X \sim \mathrm{U}[0,1]$ and $Y \sim \mathrm{U}[0,1]$ be two random variables. Define, $Z = X + Y$ and $W = X - Y$. Find the joint p.d.f. of Z and W.

Transformation of Random Vectors: Continuous Case

Example

Let $X \sim \mathrm{U}[0,1]$ and $Y \sim \mathrm{U}[0,1]$ be two random variables. Define, $Z = X + Y$ and $W = X - Y$. Find the joint p.d.f. of Z and W. Suppose that X and Y are independent. What we can say about Z and W?

Normal (Gaussian) Distribution

Definition

We say that a random variable X follows normal distribution, written $X \sim N(\mu, \sigma^2)$, if it has p.d.f. as

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{where } -\infty < x, \mu < \infty, \sigma > 0.$$

Prove the followings:

(i) $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

(ii) $M_X(t) = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right)$.

Multinormal Distribution

Definition

We say that the random variables $X_1, X_2, ..., X_n$ have a multinormal distribution if they are a linear combination of the independent $N(0,1)$ random variables $Z_1, ..., Z_m$, that is, if

$$X_k = \mu_k + \sum_{j=1}^{m} c_{kj} Z_j \quad \text{for } k = 1, 2, ..., n,$$

where μ_k is a real constant for all values of k.

Multinormal Distribution Contd...

(i) The random variables $X_1, X_2, ..., X_n$ have a multinormal distribution, if and only if, their joint p.d.f. is of the following form

$$f_{\underline{X}}(\underline{x}) = (2\pi)^{-n/2}(\det K)^{-1/2} \exp\left[\left(-\frac{1}{2}\right)(\underline{x} - \mu_{\underline{x}})K^{-1}(\underline{x} - \mu_{\underline{x}})^T\right]$$

for $x_k \in \mathbb{R}$ for all k, provided the matrix $K = \mathrm{Cov}(\underline{X})$ is non-singular.

Multinormal Distribution Contd...

(i) The random variables $X_1, X_2, ..., X_n$ have a multinormal distribution, if and only if, their joint p.d.f. is of the following form

$$f_{\underline{X}}(\underline{x}) = (2\pi)^{-n/2}(\det K)^{-1/2} \exp\left[\left(-\frac{1}{2}\right)(\underline{x} - \mu_{\underline{x}})K^{-1}(\underline{x} - \mu_{\underline{x}})^T\right]$$

for $x_k \in \mathbb{R}$ for all k, provided the matrix $K = \mathrm{Cov}(\underline{X})$ is non-singular.

(ii) The matrix K is symmetric and positive definite.

(iii) Show that if the random variables X, Y and Z are jointly normal and independent in pairs, then they are independent?

Multinomial Distribution

- Let Ω be a sample space associated with a random experiment E, and let $B_1, B_2, .., B_n$ be a partition of Ω. Assume that we perform m independent repetitions of the experiment E and that the probability $p_k = P[B_k]$ is constant from one repetition to another. If X_k denotes the number of times that the event B_k has occurred among the m repetitions, for $k = 1, 2, .., n$, then, determine the joint PMF of the random vector $(X_1, X_2, ..., X_n)$ and $\mathrm{Cov}(X_i, X_j)$.

Multinomial Distribution: Example

Example

In a certain town, 40% of the eligible voters prefer candidate A, 10% prefer candidate B, and the remaining 50% have no preference. You randomly sample 10 eligible voters. What is the probability that 4 will prefer candidate A, 1 will prefer candidate B, and the remaining 5 will have no preference?

The Weak Law of Large Numbers

Theorem ((Recall) Chebyshev's Inequality)

Suppose that X *has finite first two moments. If* $\mu = E(X)$ *and* $\sigma^2 = Var(X)(\sigma \geq 0)$. *Then, for any* $c > 0$,

$P(\{| X - \mu |\geq c\}) \leq \frac{\sigma^2}{c^2}$.

The Weak Law of Large Numbers

Theorem ((Recall) Chebyshev's Inequality)

Suppose that X *has finite first two moments. If* $\mu = E(X)$ *and* $\sigma^2 = Var(X)(\sigma \geq 0)$*. Then, for any* $c > 0$*,* $P(\{| X - \mu | \geq c\}) \leq \frac{\sigma^2}{c^2}$.

Theorem

Let $X_1, X_2, ...$ *be a sequence of independent and identically distributed (i.i.d.) random variables, each having mean* $E[X_i] = \mu$*. Then, for any* $\epsilon > 0$*,*

$$P\left\{\left|\frac{X_1 + X_2 + ... + X_n}{n} - \mu\right| > \epsilon\right\} \to 0 \quad \text{as } n \to \infty.$$

The Strong Law of Large Numbers

Theorem

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed (i.i.d.) random variables, each having mean $E[X_i] = \mu$ and $\mathrm{Var}(X_i) < \infty$. Then,

$$P\left\{\lim_{n\to\infty} \frac{X_1 + X_2 + ... + X_n}{n} = \mu\right\} = 1.$$

The Strong Law of Large Numbers: Example

Example

Let $X_1, X_2, \ldots$ be independent random variables, all having an exponential distribution with parameter $\lambda = 1$. We define the indicator variable

$$I_k = \begin{cases} 1 & \text{if } X_k > 1, \\ 0 & \text{otherwise} \end{cases}$$

for all k. Show that $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{I_k}{n} = e^{-1}$, with probability 1.

The Central Limit Theorem

Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables, each having finite mean μ and finite variance σ^2. Let $S_n = \sum_{i=1}^{n} X_i$ and

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Then, the distribution function of Z_n tends toward that of a Standard Gaussian distribution $N(0,1)$.

The Central Limit Theorem: Example

Example

A computer, in adding numbers, rounds each number to the nearest integer. Suppose that the rounding errors are independent and have a uniform distribution on the interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$. If 1500 numbers are added, what is the probability that the total error, in absolute value, exceeds 15?

The Central Limit Theorem: Example

Example

If 20% of the diodes manufactured by a certain machine are defective, what is the probability that in a batch of 100 (independent) diodes taken at random (and without replacement) among those produced by this machine, there are exactly 15 defective?

The Central Limit Theorem: Example

Example

If 20% of the diodes manufactured by a certain machine are defective, what is the probability that in a batch of 100 (independent) diodes taken at random (and without replacement) among those produced by this machine, there are exactly 15 defective?

Theorem (De Moivre - Laplace Approximation)

Let $X \sim B(n, p)$. If n is large enough and p sufficiently close to $1/2$, then we can write that

$$p_X(k) \approx f_Z(k), \text{ where } Z \sim N(np, npq).$$

MID SEM EXAM - Syllabus

1.) Algebra of Sets, Sigma-algebra, Measures and Measurable spaces, Axiomatic definition of probability, Sample Space, Events, Conditional Probability, Independence of Events, Theorem of Total Probability, Baye's Theorem.

2.) Discrete and Continuous Random Variables, Function of a Random Variable, Probability Mass Function, Probability Density Function, Cumulative Distribution Function, Moments, Mathematical Expectation, Variance, Standard Deviation, Moment Generating Function.

MID SEM EXAM - Syllabus

3.) Binomial Distribution, Poisson Distribution, Uniform Distribution, Exponential Distribution, Normal (Gaussian) Distribution, Markov's Inequality, Chebyshev's Inequality.

Statistics

- It is an art of learning from the data.

Statistics

- It is an art of learning from the data.
- It can be used to describe, summarize and analyze the available/generated data.

Statistics

- It is an art of learning from the data.
- It can be used to describe, summarize and analyze the available/generated data.
- The description and summarization of the data is called descriptive statistics.

Statistics

- It is an art of learning from the data.
- It can be used to describe, summarize and analyze the available/generated data.
- The description and summarization of the data is called descriptive statistics.
- The drawing of conclusions from the data is called inferential statistics.

Statistics

- Totality of the assumptions about the chances of obtaining the different data values is referred as a probability model for the data.

Statistics

- Totality of the assumptions about the chances of obtaining the different data values is referred as a probability model for the data.
- Total collection of elements is referred as the population and its subgroup to learn about the population is called a sample.

Central Tendencies of a data set

Suppose that we have a data set consisting of the $n-$ numerical values $x_1, x_2, ..., x_n$.

Definition (Sample Mean)

The sample mean, designated by $\overline{x}$, is defined by

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}.$$

Central Tendencies of a data set

Note: If we have a frequency table (see below) listing the k distinct values $v_1, \ldots, v_k$ having corresponding frequencies $f_1, \ldots, f_k$ with $n = \sum_{i=1}^{k} f_i$, then the sample mean of these n data values is

$$\overline{x} = \sum_{i=1}^{n} \frac{v_i f_i}{n}.$$

v_1	v_2	$\ldots$	v_k
f_1	f_2	$\ldots$	f_k

Definition (Sample Median)

Order the values of a data set of size n from smallest to largest. If n is odd, the sample median is the value in position $(n+1)/2$; if n is even, it is the average of the values in positions $n/2$ and $n/2+1$.

Central Tendencies of a data set

Definition (Sample Median)

Order the values of a data set of size n from smallest to largest. If n is odd, the sample median is the value in position $(n+1)/2$; if n is even, it is the average of the values in positions $n/2$ and $n/2+1$.

Definition (Sample Mode)

It is defined to be the value that occurs with the greatest frequency. If no single value occurs most frequently, then all the values that occur at the highest frequency are called modal values.

Example

The following frequency table gives the values obtained in 40 rolls of a die

Value	1	2	3	4	5	6
Frequency	9	8	5	5	6	7

Find (a) the sample mean, (b) the sample median, and (c) the sample mode.

Sample Variance and Sample Standard Deviation

Suppose that we have a data set consisting of the $n-$ numerical values $x_1, x_2, ..., x_n$.

Definition (Sample Variance)

The sample variance, designated by s^2, is defined by

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{(n-1)}.$$

Sample Variance and Sample Standard Deviation

Suppose that we have a data set consisting of the $n-$ numerical values $x_1, x_2, ..., x_n$.

Definition (Sample Variance)

The sample variance, designated by s^2, is defined by

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{(n-1)}.$$

An Algebraic Identity: $\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2.$

Sample Variance and Sample Standard Deviation

Definition (Sample Standard Deviation)

The quantity s, defined by

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{(n-1)}}$$

is called the sample standard deviation.

Exercise: Let $y_i = a + bx_i,\ i = 1, 2, ..., n$. Then, $\overline{y} = a + b\overline{x}$ and $s_y^2 = b^2 s_x^2$.

Sample Variance and Sample Standard Deviation: Example

Example

The following data give the worldwide number of fatal airline accidents of commercially scheduled air transports in the years from 2000 to 2005.

Year	2000	2001	2002	2003	2004	2005
Accidents	18	13	13	7	9	18

Find the sample variance of the number of accidents in these years.

Sample Correlation Coefficient

Definition (Sample Correlation Coefficient)

Consider the data pairs (x_i, y_i), $i = 1, 2, ..., n$, and let s_x and s_y denote, respectively, the sample standard deviations of the $x-$ values and the $y-$ values. The sample correlation coefficient, call it r, of the data pairs (x_i, y_i), $i = 1, 2, ..., n$ is defined by

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(n-1)s_x s_y}.$$

When $r > 0$, we say that the sample data pairs are positively correlated, and when $r < 0$ we say that they are negatively correlated.

Properties of the Sample Correlation Coefficient r

1.) $-1 \leq r \leq 1$.

Properties of the Sample Correlation Coefficient r

1.) $-1 \leq r \leq 1$.

2.) If for constants a and b, with $b > 0$,

$y_i = a + bx_i, \;\; i = 1, 2, ..., n$, then $r = 1$.

Properties of the Sample Correlation Coefficient r

1.) $-1 \leq r \leq 1$.

2.) If for constants a and b, with $b > 0$,

$y_i = a + bx_i, \ \ i = 1, 2, ..., n$, then $r = 1$.

3.) If for constants a and b, with $b < 0$,

$y_i = a + bx_i, \ \ i = 1, 2, ..., n$, then $r = -1$.

Properties of the Sample Correlation Coefficient r

1.) $-1 \leq r \leq 1$.

2.) If for constants a and b, with $b > 0$,

$y_i = a + bx_i,\ \ i = 1, 2, ..., n$, then $r = 1$.

3.) If for constants a and b, with $b < 0$,

$y_i = a + bx_i,\ \ i = 1, 2, ..., n$, then $r = -1$.

4.) If r is the sample correlation coefficient for the data pairs

$(x_i, y_i),\ \ i = 1, 2, ..., n$, then it is also the sample correlation

coefficient for the data pairs

$$a + bx_i, \quad c + dy_i,\ \ i = 1, 2, ..., n$$

provided that b and d both have same signs.

Statistical Inferences

Definition (Random Sample)

We define a random sample $\mathbf{X_n} = (X_1, X_2, ..., X_n)$ as consisting of n independent random variables with the same distribution as X. Here X is the random variable which models the property of interest of population.

Statistical Inferences

Definition (Random Sample)

We define a random sample $\mathbf{X_n} = (X_1, X_2, ..., X_n)$ as consisting of n independent random variables with the same distribution as X. Here X is the random variable which models the property of interest of population.

Definition (Statistic)

A statistic is simply a function of the random vector X_n. That is, let $X_1, X_2, ..., X_n$ be n i.i.d. random variables, then a statistic of the common distribution is given by

$$\hat{\Theta}_n = g(X_1, X_2, ..., X_n), \quad \text{where } g : \mathbb{R}^n \to \mathbb{R}.$$

Statistical Inferences Contd...

Definition (Estimator)

A statistic $\hat{\Theta}_n$ is an estimator of a parameter θ corresponding to common distribution and $\hat{\Theta}_n(w)$, for some w is an "estimate" of θ.

Statistical Inferences Contd...

Definition (Estimator)

A statistic $\hat{\Theta}_n$ is an estimator of a parameter θ corresponding to common distribution and $\hat{\Theta}_n(w)$, for some w is an "estimate" of θ.

Example

Let X be a r.v. with uniform distribution between $[0, \theta]$, where θ is unknown. Consider the following estimator

$$\hat{\Theta}_n = \max\{X_1, X_2, ..., X_n\}.$$

Determine its sampling distribution and hence determine its sample mean and sample variance.

Parameter Estimation: Questions

1.) What properties characterize a good estimator?

2.) How do we determine that an estimator is better than another?

3.) How do we find good estimators?

Properties of Estimators

1.) We say that an estimator $\hat{\Theta}_n$ is an unbiased estimator of θ if

$$E[\hat{\Theta}_n] = \theta \quad \forall \ n.$$

The bias of any estimator $\hat{\Theta}_n$ is defined by

$$\mathrm{Bias}[\hat{\Theta}_n] = E[\hat{\Theta}_n] - \theta.$$

Properties of Estimators

1.) We say that an estimator $\hat{\Theta}_n$ is an unbiased estimator of θ if

$$E[\hat{\Theta}_n] = \theta \quad \forall\ n.$$

The bias of any estimator $\hat{\Theta}_n$ is defined by

$$\mathrm{Bias}[\hat{\Theta}_n] = E[\hat{\Theta}_n] - \theta.$$

2.) We say that $\hat{\Theta}_n$ is asymptotically unbiased if $\lim_{n\to\infty} E[\hat{\Theta}_n] = \theta$.

Properties of Estimators

1.) We say that an estimator $\hat{\Theta}_n$ is an unbiased estimator of θ if

$$E[\hat{\Theta}_n] = \theta \quad \forall\ n.$$

The bias of any estimator $\hat{\Theta}_n$ is defined by

$$\mathrm{Bias}[\hat{\Theta}_n] = E[\hat{\Theta}_n] - \theta.$$

2.) We say that $\hat{\Theta}_n$ is asymptotically unbiased if $\lim_{n\to\infty} E[\hat{\Theta}_n] = \theta$.

3.) We say that $\hat{\Theta}_n$ is "consistent" if $\hat{\Theta}_n \to \theta$ (in probability).

Examples

Example

Check the properties of the following estimators.

(i) $\hat{\Theta}_n = \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$;

(ii) $\hat{\Theta}_n = S^2 = \frac{1}{n}\sum_{j=1}^{n} (X_j - \overline{X})^2$;

Quality of Estimators

1.) (**Unbiasedness**) Ideally, a good estimator should be equal to the parameter θ, on the average.

Quality of Estimators

1.) (**Unbiasedness**) Ideally, a good estimator should be equal to the parameter θ, on the average.

2.) (**Mean Square Estimation Error**) A second measure of the quality of an estimator is the mean square estimation error given by

$$E\left[\left(\hat{\Theta}_n - \theta\right)^2\right] = \mathrm{Var}\left(\hat{\Theta}_n\right) + \mathrm{Bias}(\hat{\Theta}_n)^2.$$

Quality of Estimators

1.) (**Unbiasedness**) Ideally, a good estimator should be equal to the parameter θ, on the average.

2.) (**Mean Square Estimation Error**) A second measure of the quality of an estimator is the mean square estimation error given by

$$E\left[\left(\hat{\Theta}_n - \theta\right)^2\right] = \mathrm{Var}\left(\hat{\Theta}_n\right) + \mathrm{Bias}(\hat{\Theta}_n)^2.$$

3.) (**Consistency**) A third measure of the quality of an estimator pertains to its behavior as the sample size n is increased.

Examples

Example

The message interarrival times at a message center are exponential random variables with rate λ messages per second. Compare the following two estimators for $\theta = 1/\lambda$ the mean interarrival time:

$$\hat{\Theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\Gamma}_n = n * \min\{X_1, X_2, ..., X_n\}.$$

Examples

Example

The message interarrival times at a message center are exponential random variables with rate λ messages per second. Compare the following two estimators for $\theta = 1/\lambda$ the mean interarrival time:

$$\hat{\Theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\Gamma}_n = n * \min\{X_1, X_2, ..., X_n\}.$$

Example

Discuss the consistency of the sample mean estimator and the sample variance estimator.

Maximum Likelihood Estimators

Maximum Likelihood Estimators

Example (Poisson Distributed Typos)

Papers submitted by Bob have been found to have a Poisson distributed number of typos with mean 1 typo per page, whereas papers prepared by John have a Poisson distributed number of typos with mean 5 typos per page. Suppose that a page that was submitted by either Bob or John has 2 typos. Who is the likely author?

Likelihood Function

- Let $x_1, x_2, ..., x_n$ be the observed values of a random sample for the random variable X and θ be the parameter of our interest. Then, the **(Likelihood Function)** of the sample is a function of θ defined as follows:

$$l(x_1, x_2, ..., x_n; \theta) = \begin{cases} p_{X_1, X_2, \ldots, X_n}(x_1, x_2, ..., x_n | \theta) & X \text{ discrete} \\ f_{X_1, X_2, \ldots, X_n}(x_1, x_2, ..., x_n | \theta) & X \text{ continuous} \end{cases}$$

where $p_{X_1, X_2, \ldots, X_n}(x_1, x_2, ..., x_n | \theta)$ and $f_{X_1, X_2, \ldots, X_n}(x_1, x_2, ..., x_n | \theta)$ are the joint p.m.f. and joint p.d.f. evaluated at the observation values if the parameter values is θ.

Maximum Likelihood Method

Since the samples $X_1, X_2, ..., X_n$ are i.i.d., we have the following expressions for the likelihood function:

$$l(x_1, x_2, ..., x_n; \theta) = \begin{cases} \prod_{i=1}^{n} p_{X_i}(x_i|\theta) & \text{When } X \text{ discrete} \\ \prod_{i=1}^{n} f_{X_i}(x_i|\theta) & \text{When } X \text{ continuous.} \end{cases}$$

Maximum Likelihood Method

Since the samples $X_1, X_2, ..., X_n$ are i.i.d., we have the following expressions for the likelihood function:

$$l(x_1, x_2, ..., x_n; \theta) = \begin{cases} \prod_{i=1}^{n} p_{X_i}(x_i|\theta) & \text{When } X \text{ discrete} \\ \prod_{i=1}^{n} f_{X_i}(x_i|\theta) & \text{When } X \text{ continuous.} \end{cases}$$

- **Maximum Likelihood Method:** It selects the estimator value $\hat{\Theta} = \theta^*$ where θ^* is the parameter value that maximizes the likelihood function, that is,

$$l(x_1, x_2, ..., x_n; \theta^*) = \max_{\theta} l(x_1, x_2, ..., x_n; \theta).$$

Log Likelihood Function

- The **(log likelihood function)** of the sample is defined as follows:

$$
\begin{aligned}
L(x_1, x_2, ..., x_n; \theta) &= \log l(x_1, x_2, ..., x_n; \theta) \\
&= \begin{cases} \sum_{i=1}^{n} \log p_{X_i}(x_i|\theta) & X \text{ discrete} \\ \sum_{i=1}^{n} \log f_{X_i}(x_i|\theta) & X \text{ continuous.} \end{cases}
\end{aligned}
$$

- The maximum likelihood estimate can be obtained by finding the value θ^* for which:

$$\frac{\partial}{\partial\theta} L(x_1, x_2, .., x_n; \theta) = 0.$$

Examples

Example

Suppose we perform n independent observations of a Bernoulli random variable with probability of success p. Find the maximum likelihood estimate for $p \in (0, 1)$.

Examples

Example

Suppose we perform n independent observations of a Bernoulli random variable with probability of success p. Find the maximum likelihood estimate for $p \in (0, 1)$.

Example

Suppose we perform n independent observations of a Poisson random variable with mean λ. Find the maximum likelihood estimate for $\alpha > 0$.

Examples

Example

Suppose we perform n independent observations of a Gaussian random variable X with mean μ and variance σ_X^2. Find the maximum likelihood estimates for two parameters: the mean $\theta_1 = \mu$ and variance $\theta_2 = \sigma_X^2$.

Examples

Example

Suppose we perform n independent observations of a Gaussian random variable X with mean μ and variance σ_X^2. Find the maximum likelihood estimates for two parameters: the mean $\theta_1 = \mu$ and variance $\theta_2 = \sigma_X^2$.

Example

Show that the maximum likelihood estimator for a uniform random variable that is distributed in the interval $[0, \theta]$ is

$$\hat{\Theta} = \max\{X_1, X_2, ..., X_n\}.$$

Asymptotic Properties of MLE

(i) Maximum likelihood estimators are consistent and asymptotically unbiased.

Asymptotic Properties of MLE

(i) Maximum likelihood estimators are consistent and asymptotically unbiased.

(ii) For n large, maximum likelihood estimate θ_n^* of θ is asymptotically Gaussian distributed.

Asymptotic Properties of MLE

(i) Maximum likelihood estimators are consistent and asymptotically unbiased.

(ii) For n large, maximum likelihood estimate θ_n^* of θ is asymptotically Gaussian distributed.

(iii) MLE always tends to minimum variance.

Asymptotic Properties of MLE

(i) Maximum likelihood estimators are consistent and asymptotically unbiased.

(ii) For n large, maximum likelihood estimate θ_n^* of θ is asymptotically Gaussian distributed.

(iii) MLE always tends to minimum variance.

Example

Find the distribution of sample mean estimator for p for n large, in case of Bernoulli r.v..

Estimation by Confidence Intervals

Estimation by Confidence Intervals

Definition

An interval $[LC, UC]$ is called a (two-sided) confidence interval at $100(1-\alpha)\%$ for θ if

$$P\left[LC(X_1, X_2, ..., X_n) \leq \theta \leq UC(X_1, X_2, ..., X_n)\right] = 1 - \alpha.$$

Note:

(i) θ is a parameter, whereas LC and UC are random variables.

(ii) LC and UC are called the lower confidence and upper confidence limits, respectively.

(iii) $(1-\alpha)$ is called the confidence level.

Estimation by Confidence Intervals

Example

Let $X_1, \ldots, X_n$ be a random sample of a certain population X whose mean μ is unknown, but whose variance σ^2 is known. Find an $100(1-\alpha)\%$ confidence interval for μ.

Estimation by Confidence Intervals

Example

Let $X_1, ..., X_n$ be a random sample of a certain population X whose mean μ is unknown, but whose variance σ^2 is known. Find an $100(1-\alpha)\%$ confidence interval for μ.

Example

Let $X_1, ..., X_n$ be a random sample of a Gaussian population X whose mean μ is unknown, but whose variance σ^2 is known. Find an $100(1-\alpha)\%$ confidence interval for μ.

Random Processes

Random Processes

Definition (Stochastic/Random Process)

Let E be a random experiment and let Ω be a sample space associated with E. A **stochastic process** is a set of random variables $\{X(t, w), t \in T\}$ assigned to every $w \in \Omega$. Here, the domain of t is a set T of real numbers.

Random Processes

Definition (Stochastic/Random Process)

Let E be a random experiment and let Ω be a sample space associated with E. A **stochastic process** is a set of random variables $\{X(t, w), t \in T\}$ assigned to every $w \in \Omega$. Here, the domain of t is a set T of real numbers.

(i) The graph of $X(t, w)$ as a function of t, for a fixed w, is called a realization or a trajectory (or a sample path) of the stochastic process.

Random Processes

Definition (Stochastic/Random Process)

Let E be a random experiment and let Ω be a sample space associated with E. A **stochastic process** is a set of random variables $\{X(t, w), t \in T\}$ assigned to every $w \in \Omega$. Here, the domain of t is a set T of real numbers.

(i) The graph of $X(t, w)$ as a function of t, for a fixed w, is called a realization or a trajectory (or a sample path) of the stochastic process.

(ii) If T is countably infinite/finite set, then we say that $\{X(t), t \in T\}$ is a discrete-time stochastic process.

Random Processes

(iii) If T is an interval, then $\{X(t), t \in T\}$ is called a continuous-time stochastic process.

Random Processes

(iii) If T is an interval, then $\{X(t), t \in T\}$ is called a continuous-time stochastic process.

(iv) We say that $\{X(t), t \in T\}$ is a discrete-state random process if the set of possible values of the $X(t)'$s, called the state space, is countably infinite/finite. Otherwise, it is a continuous-state process.

Random Processes: Examples

1.) **Bernoulli Process**: It is a sequence of $X_1, X_2, \ldots$ of Bernoulli random variables.

Random Processes: Examples

1.) **Bernoulli Process**: It is a sequence of $X_1, X_2, \ldots$ of Bernoulli random variables.

2.) A physical example of stochastic process is the motion of microscopic particles in collision with the molecules in a fluid (Brownian Motion).

Random Processes: Examples

1.) **Bernoulli Process**: It is a sequence of $X_1, X_2, \ldots$ of Bernoulli random variables.

2.) A physical example of stochastic process is the motion of microscopic particles in collision with the molecules in a fluid (Brownian Motion).

3.) Another example is the voltage

$$X(t) = r\cos(wt + \phi)$$

of an AC generator with random amplitude r and phase ϕ.

Statistics of Random Processes

(i) For a specific t, $X(t)$ is a random variable with distribution

$$F(x;t) = P[X(t) \leq x].$$

The function $F(x;t)$ will be called the first-order distribution of the process $X(t)$. Its derivative (if exists) with respect to x:

$$f(x;t) = \frac{\partial F(x;t)}{\partial x}$$

is the first-order density of $X(t)$.

Statistics of Random Processes Contd...

(ii) The second-order distribution of the process $X(t)$ is the joint-distribution

$$F(x_1, x_2; t_1, t_2) = P[X(t_1) \leq x_1, X(t_2) \leq x_2]$$

of the random variables $X(t_1)$ and $X(t_2)$. The corresponding density (if the second partial derivative exists) equals

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2}.$$

Statistics of Random Processes Contd...

(ii) The second-order distribution of the process $X(t)$ is the joint-distribution

$$F(x_1, x_2; t_1, t_2) = P[X(t_1) \leq x_1, X(t_2) \leq x_2]$$

of the random variables $X(t_1)$ and $X(t_2)$. The corresponding density (if the second partial derivative exists) equals

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2}.$$

(iii) The nth-order distribution of $X(t)$ is the joint distribution $F(x_1, ..., x_n; t_1, ..., t_n)$ of the random variables $X(t_1), ..., X(t_n)$.

Statistics of Random Processes Contd...

(iv) **Mean of a stochastic process**: The mean $\eta(t)$ of $X(t)$ is the expected value of the random variable $X(t)$:

$$\eta(t) = E[X(t)] = \begin{cases} \displaystyle\sum_{x \in S_{X(t)}} x p(x;t) & \text{if discrete} - \text{state} \\ \displaystyle\int_{-\infty}^{\infty} x f(x;t) dx & \text{if continouos} - \text{state.} \end{cases}$$

Statistics of Random Processes Contd...

(iv) **Mean of a stochastic process**: The mean $\eta(t)$ of $X(t)$ is the expected value of the random variable $X(t)$:

$$\eta(t) = E[X(t)] = \begin{cases} \displaystyle\sum_{x \in S_{X(t)}} x p(x; t) & \text{if discrete} - \text{state} \\ \displaystyle\int_{-\infty}^{\infty} x f(x; t) dx & \text{if continouos} - \text{state.} \end{cases}$$

(v) **Autocorrelation function**: The autocorrelation function $R_X(t_1, t_2)$ of $X(t)$ is $E[X(t_1)X(t_2)]$.

Statistics of Random Processes Contd...

(vi) **Autocovariance function**: The autocovariance function $C_X(t_1, t_2)$ of $X(t)$ is given by

$$C_X(t_1, t_2) = R(t_1, t_2) - E[X(t_1)]E[X(t_2)].$$

Statistics of Random Processes Contd...

(vi) **Autocovariance function**: The autocovariance function $C_X(t_1, t_2)$ of $X(t)$ is given by

$$C_X(t_1, t_2) = R(t_1, t_2) - E[X(t_1)]E[X(t_2)].$$

(vii) **Correlation coefficient**: $\rho_X(t_1, t_2) = \frac{C_X(t_1,t_2)}{\sqrt{C_X(t_1,t_1)C_X(t_2,t_2)}}$.

Multiple Random Processes: Examples

- Measuring temperatures in two cities at a time.
- Let $X(t)$ be a r.p. that is the "input" to a system and $Y(t)$ be another r.p. that is the "output" of the system. We may be interested in the interplay between $X(t)$ and $Y(t)$.

Multiple Random Processes Contd...

Definition (Independent random processes)

The random processes $X(t)$ and $Y(t)$ are said to be independent random processes if the vector random variables $\widehat{X} = (X(t_1), ..., X(t_k))$ and $\widehat{Y} = (Y(t_1^{'}), ..., Y(t_j^{'}))$ are independent for all k, j and all choices of $t_1, ..., t_k$ and $t_1^{'}, ..., t_j^{'}$:

$$F_{\widehat{X},\widehat{Y}}(x_1, ..., x_k, y_1, ..., y_j) = F_{\widehat{X}}(x_1, ..., x_k)F_{\widehat{Y}}(y_1, ..., y_j).$$

Multiple Random Processes Contd...

(i) The **Cross-correlation** $R_{X,Y}(t_1, t_2)$ of $X(t)$ and $Y(t)$ is given by

$$R_{X,Y}(t_1, t_2) = E[X(t_1)Y(t_2)].$$

We say that the random processes $X(t)$ and $Y(t)$ are orthogonal if $R_{X,Y}(t_1, t_2) = 0$ for all t_1 and t_2.

Multiple Random Processes Contd...

(i) The **Cross-correlation** $R_{X,Y}(t_1, t_2)$ of $X(t)$ and $Y(t)$ is given by

$$R_{X,Y}(t_1, t_2) = E[X(t_1)Y(t_2)].$$

We say that the random processes $X(t)$ and $Y(t)$ are orthogonal if $R_{X,Y}(t_1, t_2) = 0$ for all t_1 and t_2.

(ii) The **Cross-covariance** $C_{X,Y}(t_1, t_2)$ of $X(t)$ and $Y(t)$ is given by

$$C_{X,Y}(t_1, t_2) = R_{X,Y}(t_1, t_2) - E[X(t_1)]E[Y(t_2)].$$

We say that the random processes $X(t)$ and $Y(t)$ are uncorrelated if $C_{X,Y}(t_1, t_2) = 0$ for all t_1 and t_2.

Example

Example (Signal Plus Noise)

Suppose process $Y(t)$ consists of a desired signal $X(t)$ plus noise $N(t)$:

$$Y(t) = X(t) + N(t).$$

Find the cross-correlation between the observed signal and the desired signal assuming that $X(t)$ and $N(t)$ are independent random processes.

Discrete-time Random Processes

- **i.i.d. Random Process**: Let X_n be a discrete-time random process consisting of a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2. The sequence X_n is called the **i.i.d. random process**. It is easy to verify the followings:

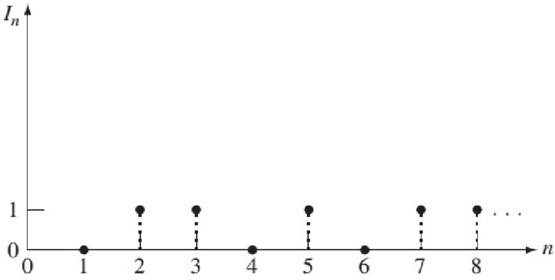
  1.) $F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \prod_{i=1}^{k} F_{X_i}(x_i)$.

  2.) $E[X_n] = \mu$ for all n.

  3.) $C_X(n_1, n_2) = \sigma^2 \delta_{n_1 n_2}$ and $R_X(n_1, n_2) = C_X(n_1, n_2) + \mu^2$.

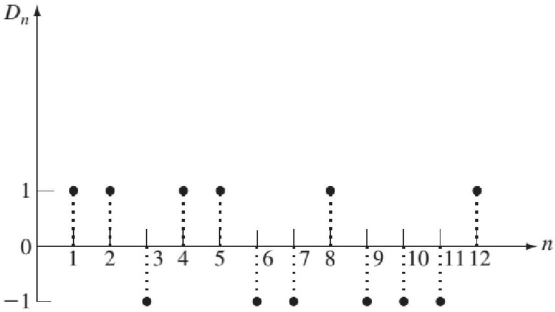
Example: Bernoulli Random Process

Let I_n be an i.i.d. random process of Bernoulli random variables taking on values from the set $\{0, 1\}$. A realization of such a process can be seen in the following figure:



Example: Random-step Process

An up-down counter is driven by $+1$ or -1 pulses. Let the input to the counter be given by $D_n = 2I_n - 1$, where I_n is the Bernoulli r.p.. A realization of such a process can be seen in the following figure:



Independent and Stationary Increments of R. P.

Independent and Stationary Increments of R. P.

Definition (Independent Increments)

Let $X(t)$ be a random process and consider two instants $t_1 < t_2$. The increment of the random process in the interval $t_1 < t \leq t_2$ is defined as $X(t_2) - X(t_1)$. A random process $X(t)$ is said to have **independent increments** if the increments in disjoint intervals are independent random variables.

Independent and Stationary Increments of R. P.

Definition (Independent Increments)

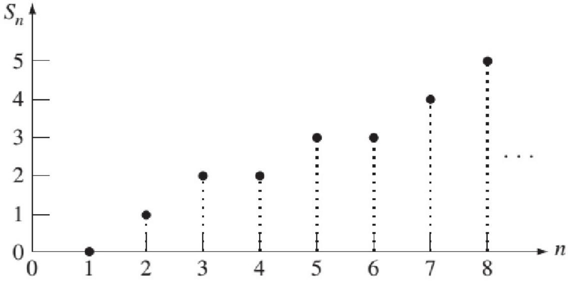
Let $X(t)$ be a random process and consider two instants $t_1 < t_2$. The increment of the random process in the interval $t_1 < t \leq t_2$ is defined as $X(t_2) - X(t_1)$. A random process $X(t)$ is said to have **independent increments** if the increments in disjoint intervals are independent random variables.

Definition (Stationary Increments)

If the distribution of $X(t_2 + \tau) - X(t_1 + \tau)$ and that of $X(t_2) - X(t_1)$ are identical for all τ, we say that $\{X(t), t \in T\}$ is a random process with **stationary increments**.

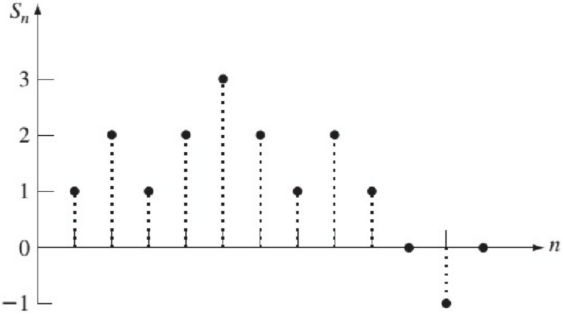
Example: Bernoulli Counting Process

Let I_n be an i.i.d. random process of Random-step variables and let $S_n = \sum_{i=1}^{n} I_i$ be the corresponding sum process. A realization of such a process can be seen in the following figure:



Example: 1D - Random Walk

Let D_n be an i.i.d. random process of Bernoulli random variables and let $S_n = \sum_{i=1}^{n} D_i$ be the corresponding sum process. A realization of such a process can be seen in the following figure:



Poisson Process

Poisson Process

(i) Let $N(t)$ be the number of event occurrences in the time interval $[0, t]$.

Poisson Process

(i) Let $N(t)$ be the number of event occurrences in the time interval $[0, t]$.

(ii) Suppose that the interval $[0, t]$ is divided into n subintervals of very short duration $\delta = t/n$. Assume that the following two condition hold:

  (a) The probability of more than one event occurrence in a subinterval is negligible compared to the probability of observing one or zero events.

  (b) Whether or not an event occurs in a subinterval is independent of the outcomes in other subintervals.

Poisson Process: Example

Example

Inquiries arrive at a recorded message device according to a Poisson process of rate 15 inquiries per minute. Find the probability that in a 1-minute period, 3 inquiries arrive during the first 10 seconds and 2 inquiries arrive during the last 15 seconds.

Markov Processes

Definition (Markov Processes)

A random process $X(t)$ is said to be **Markov** if the future of the process given the present is independent of the past; i.e., for any k and any choices of sampling instants $t_1 < t_2 < ... < t_k$ and for any $x_1, ..., x_k$,

$$f_{X(t_k)}\left(x_k | X(t_{k-1}) = x_{k-1}, ..., X(t_1) = x_1\right) = f_{X(t_k)}\left(x_k | X(t_{k-1}) = x_{k-1}\right)$$

if $X(t)$ is continuous-valued, and

$$P\left[X(t_k) = x_k | X(t_{k-1}) = x_{k-1}, ..., X(t_1) = x_1\right] = P\left[X(t_k) = x_k | X(t_{k-1}) = x_{k-1}\right]$$

if $X(t)$ is discrete-valued.

Examples:

1.) Consider the discrete-time sum process

$$S_n = \sum_{i=1}^{n} X_i$$

of i.i.d. random process $\{X_n : n = 0, 1, 2, ...\}$ with $S_0 = 0$.

Then, S_n is a Markov process.

Examples:

1.) Consider the discrete-time sum process

$$S_n = \sum_{i=1}^{n} X_i$$

of i.i.d. random process $\{X_n : n = 0, 1, 2, ...\}$ with $S_0 = 0$.

Then, S_n is a Markov process.

2.) (**Moving Average**) Consider the moving average of a i.i.d. random process X_n of Bernoulli random variables:

$$Y_n = \frac{X_n + X_{n-1}}{2}.$$

Show that Y_n is not a Markov process.

Discrete-time Markov Chain

Definition (Markov chain)

A random process $\{X_n, n = 0, 1, ...\}$ whose state space is either finite or countably infinite is called a **Markov chain** if

$$P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0] = P[X_{n+1} = j | X_n = i] = p_{ij}$$

for all states $i_0, ..., i_{n-1}, i, j$ and for all $n \geq 0$.

Assumption: The **one-step transition probabilities** are fixed and do not change with time, that is,

$$P[X_{n+1} = j | X_n = i] = p_{ij} \quad \text{for all } n.$$

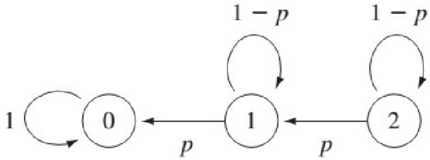
Then, X_n is completely specified by the initial p.m.f. and the

Discrete-time Markov Chain: Light Bulb Inventory

Example

On day 0 a house has two new light bulbs in reserve. The probability that the house will need a single new light bulb during day n is p, and the probability that it will not need any is 1-p. Let Y_n be the number of new light bulbs left in the house at the end of day n. Is Y_n a Markov chain? If so, draw its state transition diagram and determine its transition probability matrix.

Light Bulb Inventory



Transition Probability Matrix for the above example is

$$\begin{bmatrix} 1 & 0 & 0 \\ p & q & 0 \\ 0 & p & q \end{bmatrix}$$

Markov Chain Contd...

Notations: Let S denotes the state space.

1.) $P^{(m,n)} = [p_{ij}^{(m,n)}]$, where

$$p_{ij}^{(m,n)} = P[X_n = j | X_m = i],$$

$i, j \in S$ and $m, n \in \{0, 1, 2, ...\}$ with $m \leq n$.

2.) $\Pi_i^{(n)} \equiv P[X_n = i]$, $i \in S$, $n \in \{0, 1, 2, 3, ...\}$.

Markov Chain Contd...

Lemma

Let X_n denote a Markov chain on a (finite) state-space S. Then,

(i) $P^{(m,n)} = \prod_{k=m}^{n-1} P^{(k,k+1)}$.

(ii) $\Pi^{(n)} = \Pi^{(0)} P^n = \Pi^{(m)} P^{(m,n)}$.

Markov Chain Contd...

Lemma

Let X_n denote a Markov chain on a (finite) state-space S. Then,

(i) $P^{(m,n)} = \prod_{k=m}^{n-1} P^{(k,k+1)}$.

(ii) $\Pi^{(n)} = \Pi^{(0)} P^n = \Pi^{(m)} P^{(m,n)}$.

Find the $n-$step transition matrix in the example of "light bulb inventory". What can you conclude in the long run?

Classification of States

1.) **State j is accessible from state i** if for some $p_{ij}^{(n)} > 0$.

2.) **States i and j communicate ($i \leftrightarrow j$)** if they are accessible to each other.

3.) Two states belong to the same **class** if they communicate with each other.

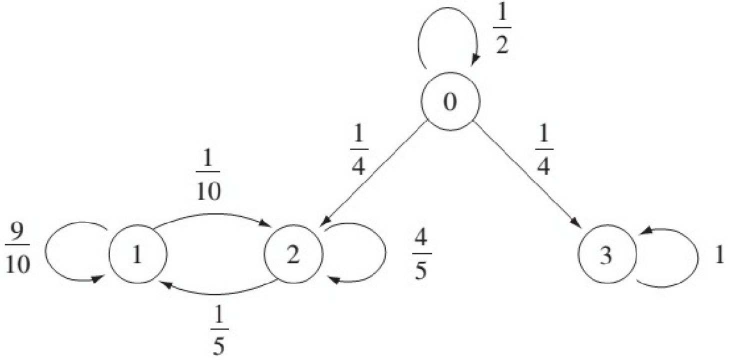
4.) A Markov chain that consists of a single class is said to be **irreducible**.

5.) A subset A of the state-space S is **closed** if none of the states in $S - A$ are accessible from A.

6.) If $\{i\}$ is closed, then i is an **absorbing state**.

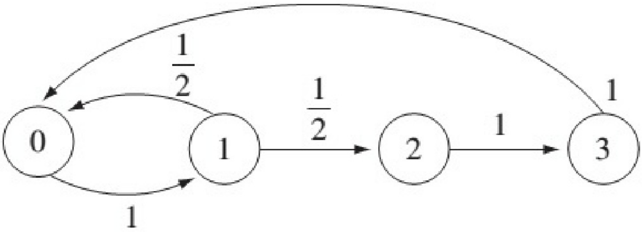
Classification of States

Example



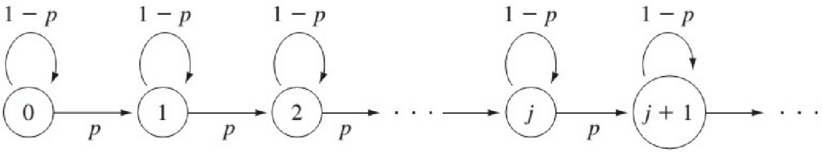
Classification of States

Example



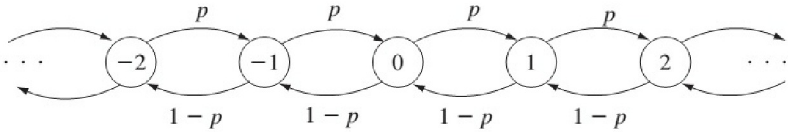
Classification of States

Example (Binomial Counting Process)



Classification of States

Example (Random Walk)



Recurrence Properties

Lemma

State i is said to be **recurrent** *if the process returns to the state with probability one, that is,*

$$f_i = P[\text{ever returning to state i}] = 1.$$

Otherwise, i is **transient**.

Recurrence Properties

Lemma

State i is said to be **recurrent** *if the process returns to the state with probability one, that is,*

$$f_i = P[\text{ever returning to state i}] = 1.$$

Otherwise, i is **transient**.

Mathematically, a state i is recurrent if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

Example In the Binomial Counting Process all the states are transient.

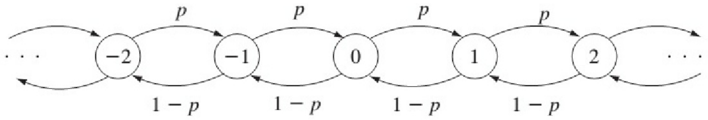
Random Walk

Example

Consider a Markov chain whose state space consists of the integers $i = 0, \pm 1, \pm 2, \ldots$, and has transition probabilities given by

$$p_{i,i+1} = p = 1 - p_{i,i-1}, \; i = 0, \pm 1, \pm 2, \ldots$$

where $0 < p < 1$. Show that all states communicate each other. Determine the value/values of p for which 0 state is recurrent.



Example

Example

Let the Markov chain consisting of the states 0, 1, 2, 3 have the transition probability matrix

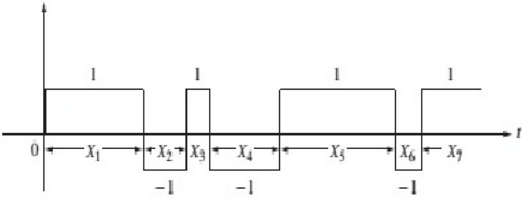
$$P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Classify the states and specify all the possible classes. Determine which states are transient and which are recurrent.

Random Telegraph Signal

Example

Consider a random process $X(t)$ that assumes the values ± 1. Suppose that $X(0) = +1$ or -1 with probability 0.5, and suppose that $X(t)$ changes polarity with each occurrence of an event in a Poisson process of rate α. Is $X(t)$ a Markov-process? Determine the state probabilities $P[X(t) = \pm 1]$.



Long-Run Proportions

- Let $i \neq j$ and $f_{i,j} = P[X_n = j \text{ for some } n > 0 | X_0 = i]$.

  Question: If i is recurrent and $i \leftrightarrow j$, then $f_{i,j} = ?$.

Long-Run Proportions

- Let $i \neq j$ and $f_{i,j} = P[X_n = j \text{ for some } n > 0 | X_0 = i]$.

  Question: If i is recurrent and $i \leftrightarrow j$, then $f_{i,j} =?$.

- Let state j is recurrent, and let

$$m_j = E[N_j | X_0 = j], \text{ where } N_j = \min\{n > 0 : X_n = j\}.$$

Definition

We say that the recurrent state j is positive recurrent if $m_j < \infty$ and say that it is null recurrent if $m_j = \infty$.

Long-Run Proportions

Lemma

Consider an irreducible Markov chain. If the chain is positive recurrent then the long-run proportions are the unique solution of the equations:

$$\Pi_j = \sum_i \Pi_i p_{ij}, \; j \geq 1 \text{ and } \sum_j \Pi_j = 1.$$

Long-Run Proportions

Lemma

Consider an irreducible Markov chain. If the chain is positive recurrent then the long-run proportions are the unique solution of the equations:

$$\Pi_j = \sum_i \Pi_i p_{ij}, \; j \geq 1 \text{ and } \sum_j \Pi_j = 1.$$

Note: The long-run proportions $\Pi_j, \; j \geq 0$ are often called stationary probabilities.

Long-Run Proportions

Example (Forecasting the Weather)

Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether of not it is raining today and not on past weather conditions. Suppose also that if it rains today, then it will rain tomorrow with probability α; and if it does not rain today, then it will rain tomorrow with probability β. Assuming that the process is in state 0 when it rains and state 1 when it does not rain, then write the one step transition probability matrix P. Calculate the long-run proportions Π_0 and Π_1, if they exists.

Limiting Probabilities

- Let $\alpha = 0.7$ and $\beta = 0.4$ in the above example. Then,

$$P^4 = \begin{bmatrix} 0.5749 & 0.428 \\ 0.5638 & 0.4332 \end{bmatrix} \text{ and } P^8 = \begin{bmatrix} 0.572 & 0.428 \\ 0.570 & 0.430 \end{bmatrix}.$$

- What can you conclude about $\lim_{n\to\infty} P[X_n = j]$?

Limiting Probabilities

- Let $\alpha = 0.7$ and $\beta = 0.4$ in the above example. Then,

$$P^4 = \begin{bmatrix} 0.5749 & 0.428 \\ 0.5638 & 0.4332 \end{bmatrix} \text{ and } P^8 = \begin{bmatrix} 0.572 & 0.428 \\ 0.570 & 0.430 \end{bmatrix}.$$

- What can you conclude about $\lim_{n\to\infty} P[X_n = j]$?
- Whether limiting probabilities and long-run proportions are same.

Limiting Probabilities

- Let $\alpha = 0.7$ and $\beta = 0.4$ in the above example. Then,

$$P^4 = \begin{bmatrix} 0.5749 & 0.428 \\ 0.5638 & 0.4332 \end{bmatrix} \text{ and } P^8 = \begin{bmatrix} 0.572 & 0.428 \\ 0.570 & 0.430 \end{bmatrix}.$$

- What can you conclude about $\lim_{n \to \infty} P[X_n = j]$?
- Whether limiting probabilities and long-run proportions are same.
- Consider a two-state Markov process with transition probability matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

The Gambler's Ruin Problem

Example

Consider a gambler who at each play of the game has probability p of winning one unit and probability $q = 1 - p$ of losing one unit. Assuming that successive plays of the game are independent, what is the probability that, starting with i units, the gambler's fortune will reach N before reaching 0?