EFOP-3.4.3-16-2016-00009

A felsőfokú oktatás minőségének és hozzáférhetőségének együttes javítása a Pannon Egyetemen

PROBABILITY THEORY AND MATHEMATICAL STATISTICS

Mihálykóné Orbán Éva
Department of Mathematics, University of Pannonia

SZÉCHENYI 2020
MAGYARORSZÁG KORMÁNYA
Európai Unió
Európai Szociális Alap
BEFEKTETÉS A JÖVŐBE

Experiment: Experiment (trial) is the observation of a phenomenon.

Phenomenon: can be artificial (caused by people) or in the nature

- deterministic: only one possible outcome
- stochastic: more than one possible outcomes

Examples: lottery, insurance, guarantee, rainfall, infected files, water level, exit pools, number of faults, number of infected people, ...

- Elementary event: a possible outcome of the observation (ω)
- Sample space: set of the possible outcomes (Ω)

- Event: a subset of Ω (A)

  we say that the event A occurs during an experiment, if the possible outcome observed during the experiment is an element of the set A
- Sure event: Ω (always occurs), impossible event: $\emptyset$ empty set (never occurs)

1. Trial: Flip a coin.

   Elementary events: Head, Tail

   $\omega_1 = H, \quad \omega_2 = T,$

   $\Omega = \{H, T\}$

   events: $\{\emptyset\}, \Omega, \{H\}, \{T\}$

2. Trial: Roll a dice

   Elementary events: 1 point, 2 points,....., 6 points on the top
   $\omega_1 = 1, \cdots, \omega_6 = 6$
   $\Omega = \{1, 2, 3, 4, 5, 6\}$
   $A = \{2, 4, 6\}$ (even number)
   $B = \{1, 3, 5\}$ (odd number)
   $C = \{2, 3, 5\}$ (prime)
   $D = \{3, 6\}$ (can be divided by 3)

3. trial: Lottery (90 numbers, 5 of them are thrown)

- Elementary events: gain, no gain

$$\omega_1 = Y \quad \omega_2 = N \qquad \Omega = \{Y, N\}$$

- Elementary events: the number of drawn numbers we crossed

$$\omega_1 = 0 \qquad \omega_2 = 1 \quad \omega_3 = 2 \quad \omega_4 = 3 \quad \omega_5 = 4 \quad \omega_6 = 5$$

$$\Omega = \{0, 1, 2, 3, 4, 5\}$$

- Elementary events: the set of drawn numbers

  $\omega_1 = \{1, 2, 3, 4, 5\}$, $\omega_2 = \{21, 23, 45, 67, 89\} \dots$

  $\Omega = \{\{i_1, i_2, i_3, i_4, i_5\} \mid 1 \leq i_j \leq 90, \text{ integers, } i_1 < i_2 < i_3 < i_4 < i_5\}$ (set of sets)

- Elementary events: the set of drawn numbers in the order of the draws = set of sequences

  $\omega_1 = (1, 2, 3, 4, 5)$, $\omega_2 = (5, 4, 3, 2, 1)$, $\omega_3 = (15, 24, 3, 42, 1) \dots$

  $\Omega = \{(i_1, i_2, i_3, i_4, i_5) \mid 1 \leq i_j \leq 90, j = 1, 2, 3, 4, 5 \text{ different integers}\}$

4. Trial: flip a coin twice

   Elementary events:
   $\omega_1 = (H, H)$, $\omega_2 = (H, T)$, $\omega_3 = (T, H)$, $\omega_4 = (T, T)$

   $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$ (set of sequences)
   Events:

   - There is at least one head :$A = \{(H, H), (H, T), (T, H)\}$
   - There are more heads than tails: $B = \{(H, H)\}$
   - There is one more head than tails $C = \emptyset$

5. Trial: measure the level of Danube in Budapest.

   a possible outcome: $\omega =$726 cm.
   $\Omega = \mathbb{R}_0^+$ (infinite set!)

   - Danger of flood: $A = \{\omega : \omega \geq 750\,\mathrm{cm}\} \subset \Omega$
   - Flood: $B = \{\omega : \omega \geq 950\,\mathrm{cm}\} \subset \Omega$
   - Danger of flood but not flood: $C = \{\omega : 750 \leq \omega < 950\} \subset \Omega$

6. Trial: measure a quantity and determine the difference (with sign) between the exact value and the measured value.

a possible outcome: $\omega = 1.5$

$\Omega = \mathbb{R}$ (infinite set!)

- the measurement is exact: $A = \{0\}$,
- under-measured value: $B = \{\omega : \omega < 0\}$,
- the measured value is at least 5 more than the exact one: $C = \{\omega : \omega \geq 5\}$

Union of events: A and B are events

$$A \cup B \textit{ occurs if at least one of them occurs}.$$

$A \cup B$ contains all elementary events which are in A or B.

$$A \cup B \text{ occurs if } A \textbf{ OR } B \text{ occurs.}$$

not excluding **or!**

Intersection of events: A and B are events

$$A \cap B \text{ occurs if both A and B occur.}$$

If the observed possible outcome is in $A \cap B$, then it is in A and also in B .

$$A \cap B \text{ occurs if } A \textbf{ AND } B \text{ occur}$$

A and$B$ are mutually exclusive if $A \cap B = \emptyset$.

<u>Difference of the events:</u> A and$B$ events

$A \setminus B$ occurs if A occurs but B does not

$A \setminus B$ contains the elementary events which are in A but (=and) not in B

$$
\begin{aligned}
A &= (A \setminus B) \cup (A \cap B) \\
\emptyset &= (A \setminus B) \cap (A \cap B)
\end{aligned}
$$

Complement of an event: A event,

$$\overline{A} \text{ occurs if } A \text{ does not occur}.$$

$$\overline{A} = \Omega \setminus A \qquad A \setminus B = A \cap \overline{B}$$

De Morgan rules

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

Definition

a set of certain subsets of Ω $\mathcal{A}$ is called σ algebra if

$\Omega \in \mathcal{A}$

if $A \in \mathcal{A}$, then $\overline{A} \in \mathcal{A}$

if $A_i \in \mathcal{A}$, $i = 1, 2, 3, \ldots.$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Theorem

If $\mathcal{A}$ is σ algebra, then

- $\emptyset \in \mathcal{A}$
- *if* $A \in \mathcal{A}$ *and* $B \in \mathcal{A}$ *then*
  $A \cup B \in \mathcal{A} \qquad (A \cup B = A \cup B \cup \emptyset \cup \cdots)$
- *if* $A \in \mathcal{A}$ *and* $B \in \mathcal{A}$ *then* $A \cap B \in \mathcal{A} \qquad \left(A \cap B = \overline{\overline{A} \cup \overline{B}}\right)$
- *if* $A_i \in \mathcal{A}$, $i = 1, 2, 3, ...,$ *then* $\bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{A} \qquad \left(\bigcap\limits_{i=1}^{\infty} A_i = \overline{\bigcup\limits_{i=1}^{\infty} \overline{A_i}}\right)$
- *if* $A \in \mathcal{A}$ *and* $B \in \mathcal{A}$ *then* $A \setminus B \in \mathcal{A} \qquad (A \setminus B = A \cap \overline{B})$

σ algebra: closed concerning operations, events: elements of $\mathcal{A}$, they have probability

Concept of probability
Relative frequency

experiment, A is an event

the experiment is repeated n times

the frequency of A in case of n trials: $k_A(n)$

relative frequency of A : $\dfrac{k_A(n)}{n}$

one can see a kind of stability

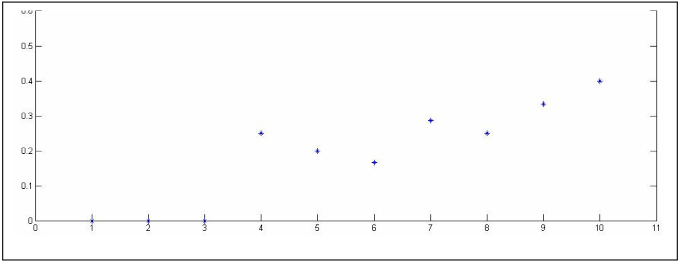
if $n \to \infty$, then $\dfrac{k_A(n)}{n} \to ?$ (as if it had a limit).

Trial: flip a coin

results of 10 repetitions: T,T,T,H,T,T,H,T,H,H

event A: the result of a flip is H.

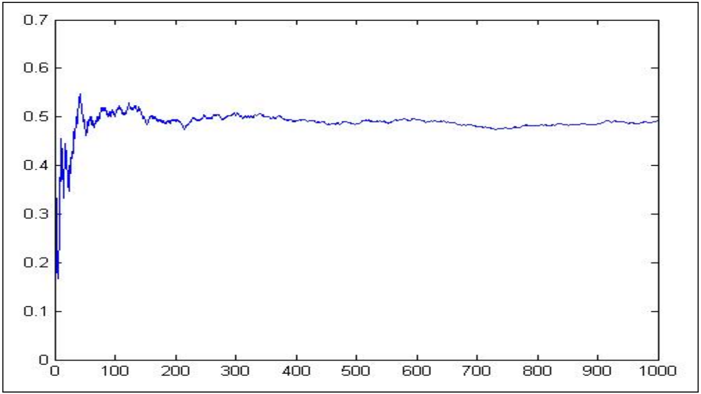
trial	T	T	T	H	T	T	H	T	H	H
n	1	2	3	4	5	6	7	8	9	10
$k_A(n)$	0	0	0	1	1	1	2	2	3	4
$\frac{k_A(n)}{n}$	0	0	0	0.25	0.2	0.17	0.27	0.25	0.33	0.4



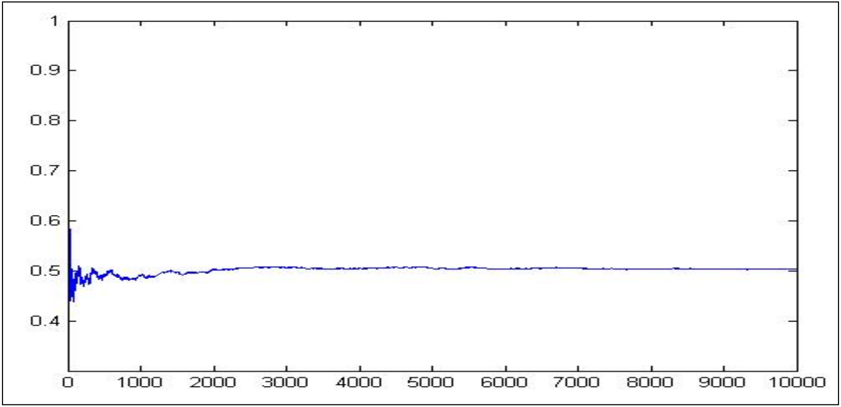
Relative frequencies in case of 10 trials

Concept of probability

Relative frequency - example iii)



Relative frequencies in case of 1000 trials



Relative frequencies in case of 10000 trials

- $0 \leq k_A(n) \leq n \Rightarrow 0 \leq \dfrac{k_A(n)}{n} \leq 1$
- $k_\Omega(n) = n \Rightarrow \dfrac{k_\Omega(n)}{n} = 1$
- if A, B are events for which $A \cap B = \emptyset$, then

$$k_{A\cup B}(n) = k_A(n) + k_B(n),$$

therefore

$$\frac{k_{A\cup B}(n)}{n} = \frac{k_A(n)}{n} + \frac{k_B(n)}{n}.$$

Definition

Probability is a function mapping from σ algebra of events to $\mathbb{R}$

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

with the following properties (axioms):
I) $0 \leq P(A)$ for all $A \in \mathcal{A}$.
II) $P(\Omega) = 1$.
III) If A_i i=1,2,....$\in \mathcal{A}$ for which $A_i \cap A_j = \emptyset, i \neq j$,then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

NOTE: $P(A)$ is a real number!

1. $P(\emptyset) = 0$

2. If $A_i \in \mathcal{A}$, $i = 1, 2, \ldots.n$ and $A_i \cap A_j = \emptyset, i \neq j$, then
$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \ .$$

3. If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

4. $P(\overline{A}) = 1 - P(A)$.

5. If $B \subset A$, then $P(A \backslash B) = P(A) - P(B)$.
6. If $B \subset A$, then $P(B) \leq P(A)$.
7. $P(A) \leq 1$ for any event.

8. $P(A \setminus B) = P(A) - P(A \cap B)$

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

10. $P(A \cup B) \leq P(A) + P(B)$

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

12. $$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j\in\{1,2,\cdots,n\}} P(A_i \cap A_j) + \sum_{i<j<k\in\{1,2,\cdots,n\}} P(A_i \cap A_j \cap A_k) \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

- Ω non empty finite set, $|\Omega| = n$
- $\mathcal{A} =$ the set of all subsets of Ω
- $$P(A) := \frac{|A|}{|\Omega|} = \frac{\text{number of "good" outcomes}}{\text{number of all outcomes}}$$

Axioms I., II., III. are satisfied.

- $0 \leq P(A)$,
- $P(\Omega) = \dfrac{|\Omega|}{|\Omega|} = 1$,
- if $A_1 \cap A_2 = \emptyset$, then

$$P(A_1 \cup A_2) = \frac{|A_1 \cup A_2|}{|\Omega|} = \frac{|A_1| + |A_2|}{|\Omega|} = P(A_1) + P(A_2).$$

NOTE: $P(\{\omega\}) = \dfrac{1}{n} \Rightarrow$ all possible outcomes are equally likely.
Axioms are satisfied $\Rightarrow$ so are the consequences.

- $\Omega \subset \mathbb{R}^n$ with finite measure, $\mu(\Omega) \neq 0; \mu(\Omega) \in \mathbb{R}_0^+$.
  μ denotes the measure (length, area, volume, ...)

- $\mathcal{A}$ is the set of measurable subsets of Ω.

- $$P(A) := \frac{\mu(A)}{\mu(\Omega)}$$

Axioms I. II. and III. are satisfied

- $0 \leq \dfrac{\mu(A)}{\mu(\Omega)}$,
- $P(\Omega) = \dfrac{\mu(\Omega)}{\mu(\Omega)} = 1$,
- *if* $A_i \cap A_j = \emptyset$, *then*

$$P(\bigcup_{i=1}^{\infty} A_i) = \frac{\mu\left(\bigcup\limits_{i=1}^{\infty} A_i\right)}{\mu(\Omega)} = \frac{\sum\limits_{i=1}^{\infty} \mu(A_i)}{\mu(\Omega)} = \sum_{i=1}^{\infty} P(A_i)\Bigg)$$

NOTE: $P(A)$ is proportional to the measure of A.

Ω is a sample space,
A and B are events

- We know that B occurs (information!)
- Given that B occurs compute the probability that A occurs
- If B occurs, then the elementary event is in B.
- If A occurs, then the elementary event is in A.
- The observed outcome must be in $A \cap B$, while the sample space is restricted to B.

Definition

Let $P(B) > 0$, the conditional probability of A given B is

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

Note: In the nominator: the probability of the intersection.
In the denominator: the probability of the condition.

I. $0 \leq P(A|B)$, *as* $P(A|B) := \dfrac{P(A \cap B)}{P(B)}, 0 \leq P(A \cap B), 0 < P(B)$.

II. $P(\Omega|B) = \dfrac{P(\Omega \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$.

III. $$P(\bigcup_{i=1}^{\infty} A_i|B) = \frac{P\left(\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup\limits_{i=1}^{\infty}(A_i \cap B)\right)}{P(B)} =$$

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

Axioms I.,II., III. are satisfied $\Rightarrow$so are the consequences

NOTE:The condition is fixed.

- $P(\overline{A}|B) = 1 - P(A|B)$
- $P(A \setminus C \mid B) = P(A|B) - P(A \cap C \mid B)$
- $P(A \cup C \mid B) = P(A|B) + P(C|B) - P(A \cap C \mid B)$
- Let $P(A \cap B) > 0$. Then

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

Theorem of total probability

Definition

$B_1, B_2, ..., B_n$ are called partition of Ω if $\bigcup\limits_{i=1}^{n} B_i = \Omega$, and $B_i \cap B_j = \emptyset$, $i \neq j$.

Theorem

Let $B_1, B_2, ..., B_n$ *be a partition of* Ω *and* $P(B_i) > 0$, $i = 1, 2, 3, ..., n$. *Then for any event* A

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i).$$

There are three shifts in a factory. 45% of all products are manufactured by the morning shift, 35% of all products are manufactured by the afternoon shift, 20% are manufactured by the evening shift. A product manufactured by the morning shift is substandard with probability 0.04, a product manufactured by the afternoon shift is substandard with probability 0.06, and a product manufactured by the evening shift is substandard with probability 0.08. Choose a product from the entire set of products. Compute the probability that the chosen product is substandard.

Let
S = the product is substandard.
B_1 = product is manufactured by the morning shift
B_2 = product is manufactured by the afternoon shift
B_3 = product is manufactured by the evening
B_1, B_2, B_3 is a partition.

Theorem of total probability

Example

$$P(B_1) = 0.45, P(B_2) = 0.35, P(B_3) = 0.2,$$

$$P(S|B_1) = 0.04,\ P(S|B_2) = 0.06,\ P(S|B_3) = 0.08,$$

$$P(S) = P(S|B_1) \cdot P(B_1) + P(S|B_2) \cdot P(B_2) + P(S|B_3) \cdot P(B_3) =$$
$$= 0.04 \cdot 0.45 + 0.06 \cdot 0.35 + 0.08 \cdot 0.2 = 0.055$$

Bayes theorem

Further question: Given that the chosen product is substandard, compute the probability that it is manufactured by the morning/afternoon/evening shift.

$P(B_1|S) =?$, $P(B_2|S) =?$, $P(B_3|S) =?$

Bayes Theorem

Theorem

Suppose that $B_1, B_2, ..., B_n$ is a partition with $P(B_i) > 0$, $i = 1, 2, 3, ..., n$. Then for any event A with $P(A) > 0$

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A|B_j) \cdot P(B_j)}.$$

Recall: $P(A|B_i) \cdot P(B_i) = P(A \cap B_i)$.

$$P(B_1|S) = \frac{P(S|B_1)P(B_1)}{P(S)} = \frac{0.04 \cdot 0.45}{0.055} = 0.327,$$

$$P(B_2|S) = \frac{P(S|B_2)P(B_2)}{P(S)} = \frac{0.06 \cdot 0.35}{0.055} = 0.382,$$

$$P(B_3|S) = \frac{P(S|B_3)P(B_3)}{P(S)} = \frac{0.08 \cdot 0.2}{0.055} = 0.291.$$

Comparison of the conditional and total (unconditional) probabilities

If we know that an event occurs, then the conditional probability of another event

- may increase as compared to the unconditional probability
- may decrease as compared to the unconditional probability
- it may happen that it does not change as compared to the unconditional probability

Comparison of the conditional and total probabilities

Example i)

Roll twice a fair dice.

- Let $A =$ the sum of the results is 8,
  $A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$
  $B =$ the difference of the results is 2
  $B = \{(1,3),(2,4),(3,5),(4,6),(6,4),(5,3),(4,2),(3,1)\}$
  Then
  $$P(B|A) = \frac{2}{5} > P(B) = \frac{8}{36} = 0.222.$$

Comparison of the conditional and total probability

Example ii)

- Let $C =$ the difference of the rolls is at most 1,
  $C = \left\{ \begin{array}{c} (1,1),(1,2),(2,1),(2,2),(2,3),(3,2),(3,3),(3,4), \\ (4,3),(4,4),(4,5),(5,4),(5,5),(5,6),(6,5),(6,6) \end{array} \right\}$
  $D =$ the sum of the results equals 7
  $D = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

  Then
  $$P(C|D) = \frac{2}{6} = 0.333 < P(C) = \frac{16}{36} = 0.444$$

- Let $E =$ the difference of the rolls is at least 4
  $E = \{(5,1),(1,5),(2,6),(6,2),(1,6),(6,1)\}$
  $F =$ the first roll equals 5
  $F = \{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\}$
  Then
  $$P(F|E) = \frac{1}{6} = \frac{6}{36} = P(F)$$

Definition

A and B are called independent is

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

Theorem

If $P(A) > 0, P(B) > 0$, moreover A and B are independent, then

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

Theorem

If $P(A) > 0$ and $P(B) > 0$, furthermore $A \cap B = \emptyset$, then A and B can not be independent!

$P(A \cap B) = P(\emptyset) = 0$, $P(A) \cdot P(B) \neq 0, P(A \cap B) \neq P(A) \cdot P(B) \Rightarrow A$ and B are not independent!

IMPORTANT NOTE: Independent$\neq$mutually exclusive

Theorem

If A and B are independent, then so are A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$.

Proof

$$\begin{aligned} P(A \cap \overline{B}) &= P(A \setminus B) = P(A) - P(A \cap B) = \\ &= P(A) - P(A) \cdot P(B) = P(A) \cdot (1 - P(B)) = P(A) \cdot P(\overline{B}) \end{aligned}$$

$$\begin{aligned} P(\overline{A} \cap \overline{B}) &= P(\overline{A \cup B}) = 1 - P(A \cup B) = \\ &= 1 - (P(A) + P(B) - P(A \cap B)) = P(\overline{A}) \cdot P(\overline{B}) \end{aligned}$$

■

Definition

Events A, B, and C are called independent if

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

$$P(A \cap C) = P(A) \cdot P(C)$$

$$P(B \cap C) = P(B) \cdot P(C)$$

(pairwise independent)

moreover

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

Choose one number of the set $\{1, 2, 3, 4\}$.

Let $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$.

Then $P(A) = \frac{1}{2}$,

$P(B) = \frac{1}{2}$,

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

$$\left.\begin{aligned} P(A \cap B) &= P(\{1\}) = \frac{1}{4} = P(A) \cdot P(B) \\ P(A \cap C) &= P(\{1\}) = \frac{1}{4} = P(A) \cdot P(C) \\ P(B \cap C) &= P(\{1\}) = \frac{1}{4} = P(B) \cdot P(C) \end{aligned}\right\} \Rightarrow \textit{pairwise} \text{ independent}$$

but

$$P(A \cap B \cap C) = P(\{1\}) = \frac{1}{4} \neq \frac{1}{8} = P(A) \cdot P(B) \cdot P(C).$$

Definition

Events A_i, $i \in I$ are called independent if for any $n \geq 1$ integer, and $\{k_1, k_2, ..., k_n\} \subset I$ indices

$$P(A_{k_1} \cap A_{k_2} \cap ... \cap A_{k_n}) = P(A_{k_1}) \cdot P(A_{k_2}) \cdot .... \cdot P(A_{k_n})$$

Experiments are called independent if the events connected to them are independent.

Random variables

Let
Ω sample space
$\mathcal{A}$ σ algebra of events
P probability
Consider the functions

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

$\xi(\omega)$ depends on the result of the random experiment, ($\omega \in \Omega$),
therefore ξ is a random quantity (random variable)

Flip a fair coin;
if the result is head, then I gain 5HUF;
if the result is tail, I pay 3 HUF.
Let ξ be the money I gain.
$\Omega = \{H, T\}, \mathcal{A} = 2^{\Omega}$; P classical probability
$\xi(H) = 5, \quad \xi(T) = -3,$

therefore $\quad \xi : \Omega \to \mathbb{R} \quad \xi(\omega) = \begin{cases} 5 & \text{if } \omega = H \\ -3 & \text{if } \omega = T \end{cases}$

ξ can be 5 and -3, moreover
$P(\xi = 5) = P(\{H\}) = 0.5, \; P(\xi = -3) = P(\{T\}) = 0.5.$
The probability of gain $P(\xi > 0) = P(\xi = 5) = 0.5$

Roll a fair dice, let the gain be the square of the number of points.

ξ is the square of the number of points
$\Omega = \{1, 2, 3, 4, 5, 6\}$,
$\mathcal{A} = 2^{\Omega}$ (all subsets of Ω),
P classical probability
$\xi(1) = 1,\ \xi(2) = 4,\ \xi(3) = 9,\ \xi(4) = 16,\ \xi(5) = 25,\ \xi(6) = 36.$
Therefore $\qquad \xi : \Omega \to \mathbb{R}, \quad \xi(i) = i^2.$

possible values of ξ are 1, 4, 9, 16, 25 and 36 and

$$P(\xi = 1) = P(\{1\}) = \tfrac{1}{6} \quad P(\xi = 4) = P(\{2\}) = \tfrac{1}{6}, \quad \cdots$$
$$\cdots \quad P(\xi = 36) = P(\{6\}) = \tfrac{1}{6}.$$

The gain is at most 10 HUF:

$$P(\xi \le 10) = P(\xi = 1) + P(\xi = 4) + P(\xi = 9) = \tfrac{1}{2}.$$

Roll twice a fair dice, let ξ be the difference between the points
$\Omega = \{(1,1),(1,2),.....,(6,6)\}$, $\mathcal{A} = 2^{\Omega}$, P classical probability.
$\xi((1,1)) = 0$, $\xi((2,2)) = 0$, $\xi((2,1)) = 1$, $\xi((6,1)) = 5, ...$

$$\xi : \Omega \to \mathbb{R}, \quad \xi\left((i_1, i_2)\right) = |i_1 - i_2|.$$

The possible values of ξ are 0, 1, 2, 3, 4 and 5.
$P(\xi = 0) = P(\{(1,1),(2,2),....(6,6)\}) = \frac{6}{36}$,
$P(\xi = 1) = P(\{(1,2),(2,1),(3,2),(2,3),\cdots,(5,6),(6,5)\}) = \frac{10}{36}$,
$P(\xi = 2) = P(\{(1,3),(3,1),(4,2),(2,4),\cdots,(4,6),(6,4)\}) = \frac{8}{36}$,
$P(\xi = 3) = P(\{(1,4),(4,1),(5,2),(2,5),(3,6),(6,3)\}) = \frac{6}{36}$,
$P(\xi = 4) = P(\{(1,5),(5,1),(2,6),(6,2)\}) = \frac{4}{36}$,
$P(\xi = 5) = P(\{(6,1),(1,6)\}) = \frac{2}{36}$.

$$P(1 \leq \xi < 3) = P(\xi = 1) + P(\xi = 2) = \frac{1}{2}.$$

Choose two numbers with replacement out of $\{1, 5, 12\}$, ξ is the sum of the numbers.
$\Omega = \{(1,1), (1,5), \cdots, (12,12)\}$,
$\mathcal{A} = 2^{\Omega}$,
P classical probability
$\xi((1,1)) = 2,\ \xi((1,5)) = 6,\ \xi((5,5)) = 10, \ldots$
$\xi : \Omega \to \mathbb{R}, \quad \xi((i_1, i_2)) = i_1 + i_2.$

The possible values of ξ are 2, 6, 10, 13, 17 and 24, moreover
$P(\xi = 2) = P(\{(1,1)\}) = \frac{1}{9}$,
$P(\xi = 6) = P(\{(1,5),(5,1)\}) = \frac{2}{9}$,
$P(\xi = 10) = P(\{(5,5)\}) = \frac{1}{9}$,
$P(\xi = 13) = P(\{(1,12),(12,1)\}) = \frac{2}{9}$,
$P(\xi = 17) = P(\{(5,12),(12,5)\}) = \frac{2}{9}$,
$P(\xi = 24) = P(\{(12,12)\}) = \frac{1}{9}$.
The sum is at least 15: $P(\xi \geq 15) = P(\xi = 17) + P(\xi = 24) = \frac{1}{3}$.

Choose two numbers without replacement out of $\{1, 5, 12\}$,let ξ be the minimum of the numbers.
$\Omega = \{\{1,5\}, \{1,12\}, \{5,12\}\}$,
$\mathcal{A} = 2^{\Omega}$, P classical probability.
$\xi(\{1,5\}) = 1,\ \xi(\{1,12\}) = 1,\ \xi(\{5,12\}) = 5$
$\xi : \Omega \to \mathbb{R}, \quad \xi(\{i_1, i_2\}) = \min\{i_1, i_2\}$.

The possible values of ξ are 1 and 5

$P(\xi = 1) = P(\{\{1, 5\}, \{1, 12\}\}) = \frac{2}{3}$,

$P(\xi = 5) = P(\{\{5, 12\}\}) = \frac{1}{3}$,

The minimum is prime

$$P\left(\xi \in \{2, 3, 5, 7, \cdots\}\right) = P(\xi = 5) = \frac{1}{3}.$$

- The values ξ can be listed: $x_1,\ x_2,\ x_3,\ \cdots$
- The probabilities
$$p_1 = P(\xi = x_1),\ p_2 = P(\xi = x_2),\ p_3 = P(\xi = x_3),\ \cdots$$
can be computed.

The following inequalities hold

$$\begin{aligned} 0 &\leq p_1,\ p_2,\ p_3,\ \cdots \\ 1 &= p_1 + p_2 + p_3 + \cdots \end{aligned} \tag{1}$$

moreover

$$P(\xi \in A) = \sum_{x_i \in A} p_i \qquad A \subset \mathbb{R} \tag{2}$$

especially

$$P(\xi < x) = \sum_{x_i < x} p_i \qquad x \in \mathbb{R} \tag{3}$$

The random variable ξ is called discrete, if the set of its possible values is finite or countable infinite.
If ξ is discrete we give its distribution if we list its possible values $(x_1, x_2, ...)$ and the probabilities $p_i = P(\xi = x_i) \quad i = 1, 2, \cdots .$

$$\text{Notation:} \quad \begin{pmatrix} x_1, & x_2, & x_3, & . & . \\ p_1, & p_2, & p_3, & . & . \end{pmatrix}$$

Examples i)-v) are discrete r.v-s.
The probabilities satisfy (1) and (2) we do not need the mapping $\omega \mapsto \xi(\omega)$.

Shoot on a circle with radius R by geometric probability. Let ξ be the distance between the centre of the circle and the point.

- $\Omega = \left\{ (u, v) \in \mathbb{R}^2 \mid u^2 + v^2 \leq R^2 \right\}$ the points of the circle
- $\mathcal{A}$ $-$ the set of the subsets of the circle which has area
- P $-$ geometric probability
- $\xi : \Omega \to \mathbb{R} \quad \xi\left((u, v)\right) = \sqrt{u^2 + v^2}$
  the possible values of ξ are the numbers of the interval $[0, R]$

Random variables

Example vi)

- $P(\xi = 0) = \frac{T(centre)}{T(\Omega)} = \frac{0}{R^2\pi} = 0$
- $P(\xi = R) = \frac{T(\text{ circle line with radius } R)}{T(\Omega)} = \frac{0}{R^2\pi} = 0$
- $P(\xi = R/2) = \frac{T(\text{circle line with radius } R/2)}{T(\Omega)} = \frac{0}{R^2\pi} = 0$
- $P(\xi = x) = \frac{T(\text{circle line with radius x })}{T(\Omega)} = \frac{0}{R^2\pi} = 0$
- All possible values have probability zero!!!!

Instead of $P(\xi = x)$ consider the probabilities $P(\xi < x)$!

- $P(\xi < 0) = P(\emptyset) = 0$
- $P(\xi < R) = \dfrac{T(\textit{open circle }\text{)with radius } R)}{T(\Omega)} = \dfrac{R^2\pi}{R^2\pi} = 1$
- $P(\xi < R/2) = \dfrac{T(\textit{open circle with radius } \mathrm{R}/2)}{T(\Omega)} = \dfrac{(R/2)^2\pi}{R^2\pi} = 0.25$
- $P(\xi < x) = \dfrac{T(\textit{open circle with radius } x\ )}{T(\Omega)} = \dfrac{x^2\pi}{R^2\pi} = \dfrac{x^2}{R^2}$, if $0 < x < R$
- $P(\xi < x) = P(\emptyset) = 0 \quad \text{if } x < 0$
- $P(\xi < x) = P(\Omega) = 1 \quad \text{if } R < x.$

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

Definition

The function $\xi : \Omega \to \mathbb{R}$ is called random variable if for any $x \in \mathbb{R}$

$$\{\omega : \xi(\omega) < x\} \in \mathcal{A}$$

Remark

The previous examples satisfy this.

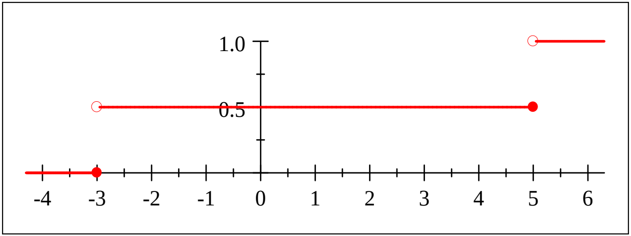
Definition

The cumulative distribution function of ξ is $F : \mathbb{R} \to \mathbb{R}$ for which

$$F(x) := P(\xi < x) = P\left(\{\omega : \xi(\omega) < x\}\right) \quad x \in \mathbb{R}$$

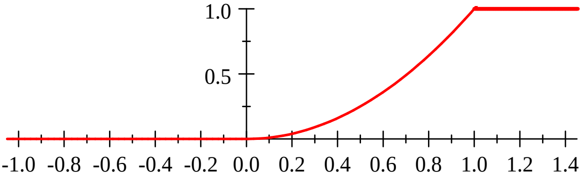
$$F(x) = P(\xi < x) = \begin{cases} 0 & \text{if } x \le -3 \\ \frac{1}{2} & \text{if } -3 < x \le 5 \\ 1 & \text{if } 5 < x \end{cases}$$

Graph:



$$F(x) = P(\xi < x) = \begin{cases} 0 & \text{if } x \le 0 \\ \dfrac{x^2}{R^2} & \text{if } 0 < x \le R \\ 1 & \text{if } R < x \end{cases}$$

Graph, if $R = 1$:



ξ r.v., F c.d.f., $F(x) := P(\xi < x) \quad x \in \mathbb{R}$

A) $0 \leq F(x) \leq 1$ (it is a probability)

B) $F(x)$ is monotone increasing

$$\Big(\text{if } x < y, \text{ then } \{\omega : \xi(\omega) < x\} \subset \{\omega : \xi(\omega) < y\}\Big)$$

C) $\lim\limits_{x \to \infty} F(x) = 1, \quad \lim\limits_{x \to -\infty} F(x) = 0$

D) F is continuous from left hand side, that is $\lim\limits_{x \to a-} F(x) = F(a)$.

Remark

B) and C) imply A)

Conditions B), C) and D) are sufficient:

Theorem

If the function $F : \mathbb{R} \to \mathbb{R}$ satisfies B), C), D) then there exists a random variable ξ with c.d.f. F.

ξ random variable with c.d.f. F, $F(x) := P(\xi < x)$

Theorem

$$P\left(\xi \in (-\infty, a)\right) = P(\xi < a) = F(a)$$

$$P\left(\xi \in [a, \infty)\right) = P(a \le \xi) = P(\overline{\xi < a}) = 1 - F(a)$$

$$P(a \le \xi < b) = F(b) - F(a)$$

ξ random variable with c.d.f. F, $F(x) := P(\xi < x)$

Theorem

$$P(\xi = a) = \lim_{x \to a+} F(x) - F(a)$$

Corollary

- *If F is a continuous function in a then* $P(\xi = a) = 0$.
- *if F is a continuous function in every* $x \in \mathbb{R}$, *then* ξ *takes every value with probability 0.*

ξ is a random variable with c.d.f. F, $F(x) := P(\xi < x)$

Definition

The r.v. ξ is called continuous r.v. if there exists such a function $f : \mathbb{R} \to \mathbb{R}$, which is continuous except from at most finite points, for which

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

f is called probability density function (p.d.f).

Remark

- *If ξ is continuos random variable, then $F(x)$ is continuous in every $x \in \mathbb{R}$.*
- *If f is continuous in x, then F is differentiable in x, moreover $F'(x) = f(x)$.*
- *If the value of f is changed at a (some) point, then F does not change.*

ξ continuous r.v., f is its p.d.f.

I.) $f(x) \geq 0$ (except from some points) (derivative of an increasing function)

II.) $\int\limits_{-\infty}^{\infty} f(t)dt = 1$

$$\left(\text{as } \int\limits_{-\infty}^{\infty} f(t)dt = \lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x) = 1 - 0\right)$$

Theorem

If $f : \mathbb{R} \to \mathbb{R}$ satisfies the above properties I.) and II.) then there exists a continuous r.v. ξ which has p.d.f. f.

$$P(a \leq \xi < a + \Delta a) = F(a + \Delta a) - F(a)$$

$$\frac{P(a \leq \xi < a + \Delta a)}{\Delta a} = \frac{F(a + \Delta a) - F(a)}{\Delta a}$$

$$\lim_{\Delta a \to 0+} \frac{F(a + \Delta a) - F(a)}{\Delta a} = F'(a) = f(a)$$

therefore

$$P(a \leq \xi < a + \Delta a) \approx f(a) \cdot \Delta a$$

Theorem

$$P(a \leq \xi < b) = \int_a^b f(t)dt$$

$$P(a \leq \xi < b) = F(b) - F(a) = \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt = \int_a^b f(t)dt.$$

Remark

F is continuous in every point, therefore $P(\xi = x) = 0$ *for any* $x \in \mathbb{R}$

Corollary

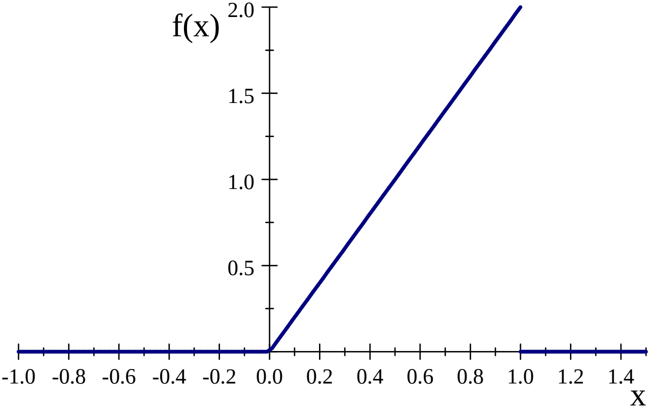
$$P(a \leq \xi < b) = P(a \leq \xi \leq b) = P(a < \xi < b) = P(a < \xi \leq b) = \\ = F(b) - F(a).$$

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \dfrac{x^2}{R^2} & \text{if } 0 < x \leq R \\ 1 & \text{if } R < x \end{cases}$$

$$f(x) = F'(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{2x}{R^2} & \text{if } 0 < x < R \\ 0 & if\ R < x \end{cases}$$

if $x = 0$, then F is differentiable
if $x = R$, then F is not differentiable at x.

Let $R = 1$



Definition

The r.v.-s ξ and η are called identically distributed if they have the same c.d.f.-s.

Theorem

If ξ and η are discrete, moreover identically distributed, then they have same possible values and the same probabilities belonging to them.

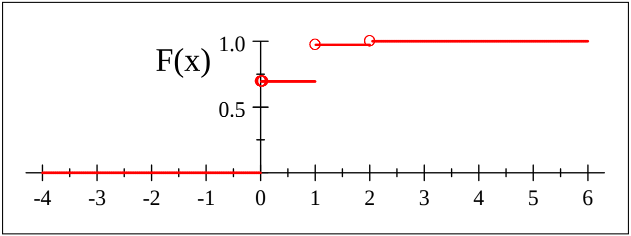
For example: $\xi : \begin{pmatrix} -1 & 2 & 11 \\ 0.4 & 0.25 & 0.35 \end{pmatrix}$, $\eta : \begin{pmatrix} -1 & 2 & 11 \\ 0.4 & 0.25 & 0.35 \end{pmatrix}$

Theorem

If ξ and η are continuous r.v.-s and moreover they are identically distributed, then the p.d.f. of ξ equals the p.d.f. of η.

Roll twice a fair dice. Let ξ be the number of "6".
Then
$\Omega = \{(1,1), ..., (6,6)\}, \xi((1,1)) = 0; \xi((5,6)) = 1; ..., \xi((6,6)) = 2;$

$$\xi : \begin{pmatrix} 0 & 1 & 2 \\ \frac{25}{36} & \frac{11}{36} & \frac{1}{36} \end{pmatrix}$$

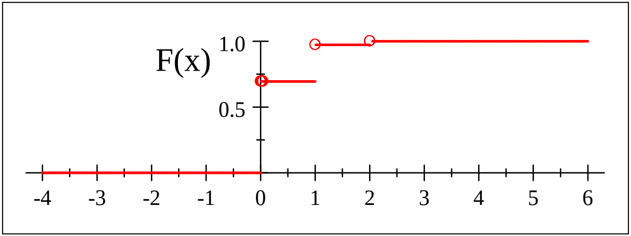


Random variables

Identically distributed r.v.-s - example ii)

Choose 1 number from the set .Let the gain be 2 HUF, if the number equals 1, let the gain be 2, if the number is at least 2 and at most 12, the gain is 1, in any other case we do not gain. Let the gain be denoted by η.
$\Omega = \{1, ..., 36\}, \eta(1) = 2; \eta(2) = 1; \eta(12) = 1; \eta(13) = 0; ...; \eta(36) = 0;$

$$\eta : \begin{pmatrix} 0 & 1 & 2 \\ \frac{25}{36} & \frac{11}{36} & \frac{1}{36} \end{pmatrix}$$



The c.d.f.-s of ξ and η and are the same functions, therefore ξ and η are

$(\Omega, \mathcal{A}, P)$, ξ, η r.v.-s

Definition

The r.v.-s ξ, η are called independent if for any values of $x, y \in \mathbb{R}$
$P((\xi < x) \cap (\eta < y)) = P(\xi < x) \cdot P(\eta < y)$.

Remark

Independence of ξ and η means the independence of the events $\{\xi < x\}$ and $\{\eta < y\}$ $x, y \in \mathbb{R}$.

$(\Omega, \mathcal{A}, P)$, ξ, η r.v.-s

Theorem

Let ξ and η be discrete r.v.-s with possible values x_i i=1,.... and y_j

j=1,..., respectively.
ξ and η are independent if and only if

$$P((\xi = x_i) \cap (\eta = y_j)) = P(\xi = x_i) \cdot P(\eta = y_j)$$

for all $i = 1, ...$ and $j = 1, ...$

$(\Omega, \mathcal{A}, P)$, ξ, η r.v.-s

Theorem

Let ξ and η continuous r.v.-s with p.d.f. $f(x)$ and $g(y)$,respectively. ξ and η are independent, if and only if

$$\frac{\partial^2 P(\xi < x \cap \eta < y)}{\partial x \partial y} = f(x) \cdot g(y)$$

for every $x \in \mathbb{R}, y \in \mathbb{R}$.

1. expectation (expected value)
2. variance/dispersion
3. mode
4. median
5. covariance/correlation coefficient

Definition

Let ξ *discrete r.v.* $\xi : \begin{pmatrix} x_1 & x_2 & x_3 & . & . \\ p_1 & p_2 & p_3 & . & . \end{pmatrix}$.
The expectation of ξ *is*

$$E(\xi) = \sum_{i=1}^{\infty} x_i \cdot p_i$$

supposing $\sum_{i=1}^{\infty} |x_i| \cdot p_i < \infty$.

Remark

If the number of possible values is finite, then the sum is finite as well.

Remark

$\sum_{i=1}^{\infty} |x_i| \cdot p_i < \infty$ *implies* $\sum_{i=1}^{\infty} x_i \cdot p_i = E(\xi) < \infty$.

Definition

Let ξ be continuous random variable with p.d.f. f.
The expectation of ξ is

$$E(\xi) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

supposing $\displaystyle\int_{-\infty}^{\infty} |x| \cdot f(x)dx < \infty$.

Remark

$\displaystyle\int_{-\infty}^{\infty} |x| \cdot f(x)\, dx < \infty$ *implies* $E(\xi) < \infty$.

Let $\xi,\ \eta$ r.v.-s, $a,\ b,\ c \in \mathbb{R}$

1. If ξ and η are identically distributed, then $E(\xi) = E(\eta)$.
2. If $0 \leq \xi$, then $0 \leq E(\xi)$
3. $E(\xi + \eta) = E(\xi) + E(\eta)$
4. $E(c \cdot \xi) = c \cdot E(\xi)$.

Consequences:

5. If $\xi = c$, then $E(\xi) = c$
6. $E(a \cdot \xi + b) = a \cdot E(\xi) + b$
7. If $a \leq \xi \leq b$, then $a \leq E(\xi) \leq b$
8. If $\xi \leq \eta$, then $E(\xi) \leq E(\eta)$

9. If $\xi_1, \xi_2, ..., \xi_n$ are identically distributed, $E(\xi_1) = \ldots E(\xi_n) = m$, then

$$E\left(\sum_{i=1}^{n} \xi_i\right) = n \cdot m,$$

10. Furthermore

$$E\left(\frac{\sum\limits_{i=1}^{n} \xi_i}{n}\right) = m.$$

11. If ξ and η are independent, then

$$E(\xi \cdot \eta) = E(\xi) \cdot E(\eta).$$

12. If ξ is discrete, $\xi : \begin{pmatrix} x_1 & x_2 & x_3 & \cdots \\ p_1 & p_2 & p_3 & \cdots \end{pmatrix}$,
    $g : \mathbb{R} \to \mathbb{R}$ function, $g(\xi)$ exists, then

$$E\left(g(\xi)\right) = \sum_{i=1}^{\infty} g(x_i) \cdot p_i$$

    supposing $\sum_{i=1}^{\infty} |g(x_i)| \cdot p_i < \infty$.

13. Especially, $g(x) = x^2$,

$$E(\xi^2) = \sum_{i=1}^{\infty} x_i^2 \cdot p_i, \ \textit{supposing} \ \sum_{i=1}^{\infty} x_i^2 \cdot p_i < \infty.$$

14. If ξ is continuous random variable with p.d.f. f, $g:\mathbb{R}\to\mathbb{R}$; $g(\xi)$ exists,

$$E\left(g(\xi)\right)=\int\limits_{-\infty}^{\infty} g(x)\cdot f(x)dx$$

supposing $\int\limits_{-\infty}^{\infty} |g(x)|\cdot f(x)dx<\infty$.

15. Especially, $g(x)=x^2$

$$E(\xi^2)=\int\limits_{-\infty}^{\infty} x^2\cdot f(x)dx, \text{supposing } \int\limits_{-\infty}^{\infty} x^2\cdot f(x)dx<\infty.$$

Definition

The variance of ξ *is* $D^2(\xi) = E\left((\xi - E(\xi))^2\right)$, *if the expectations exist.*

Definition

The dispersion of ξ $D(\xi) = \sqrt{D^2(\xi)} = \sqrt{E\left((\xi - E(\xi))^2\right)}$*,if the expectations exist.*

Remark

$(\xi - E(\xi))^2$ *is nonnegative, therefore* $E(\xi - E(\xi))^2 \geq 0$, *the square root can be taken.*

1. If two random variables are identically distributed, then their variances are equal.
2. $D^2(\xi) = E(\xi^2) - (E(\xi))^2, \quad D(\xi) = \sqrt{E(\xi^2) - (E(\xi))^2}$
3. If $\mathrm{P}(\xi = c) = 1$, then $D^2(\xi) = 0 = D(\xi)$.
4. If $D^2(\xi) = 0 = D(\xi)$, then $P(\xi = c) = 1$.
5. $D^2(a \cdot \xi + b) = a^2 \cdot D^2(\xi), \;\; D(a \cdot \xi + b) = |a| \cdot D(\xi)$

6. If ξ and η are independent, then

$$D^2(\xi+\eta) = D^2(\xi) + D^2(\eta).$$

$$D(\xi+\eta) \neq D(\xi) + D(\eta)!$$

7. If $\xi_1, \xi_2, \ldots, \xi_n$ are independent, identically distributed random variables, $D^2(\xi_1) = \ldots = D^2(\xi_n) = \sigma^2$ then

$$D^2\left(\sum_{i=1}^{n} \xi_i\right) = n \cdot D^2(\xi_1) = n \cdot \sigma^2,$$

$$D\left(\sum_{i=1}^{n} \xi_i\right) = \sqrt{n} \cdot D(\xi_1) = \sqrt{n} \cdot \sigma$$

8. If $\xi_1, \xi_2, \ldots, \xi_n$ are independent, identically distributed random variables, $D(\xi_1) = \ldots = D(\xi_n) = \sigma$, then

$$D^2\left(\frac{\sum_{i=1}^{n} \xi_i}{n}\right) = \frac{\sigma^2}{n},$$

$$D\left(\frac{\sum_{i=1}^{n} \xi_i}{n}\right) = \frac{\sigma}{\sqrt{n}}$$

Definition

Let ξ be discrete r.v.. The mode of ξ is its most likely possible value.

$\xi : \begin{pmatrix} x_1 & x_2 & x_3 & \cdots \\ p_1 & p_2 & p_3 & \cdots \end{pmatrix}$, *the mode of ξ is x_i, if $p_i \geq p_j$, $j = 1, 2, \ldots$.*

Remark

The mode of ξ may not be unique (multimode distribution).

Remark

Mode is one of the possible values.

Definition

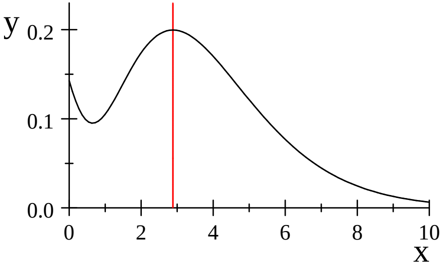
Let ξ be continuous r.v. with p.d.f. f. The mode of ξ is the argument of the local maximum value of f.

Remark

The mode of ξ may not be unique (multimode distribution).

$$f(x) = \begin{cases} \frac{1}{7}\left(e^{-x} + x^3 \cdot e^{-x}\right) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

mode: 0, 2.879



Definition

Let ξ be a r.v.. The the median of ξ is x, if

$$P(\xi \leq x) \geq 0.5,$$

and

$$P(\xi \geq x) \geq 0.5.$$

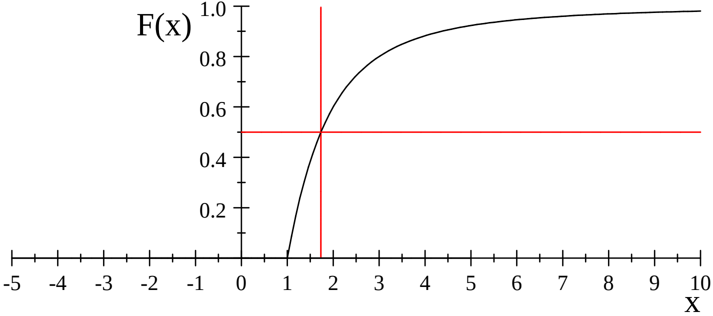
Statement

If ξ is continuous r.v., then median of ξ is x if $F(x) = 0.5$.

Proof $P(\xi \leq x) = P(\xi < x) = F(x) \geq 0.5$,
$P(\xi \geq x) = 1 - F(x) \geq 0.5$, $0.5 \geq F(x) \Rightarrow F(x) = 0.5$. ∎

$$F(x) = \begin{cases} \frac{x^2-1}{x^2+1} & \text{if } 1 \leq x \\ 0 & \text{otherwise} \end{cases}$$

median: 1.73



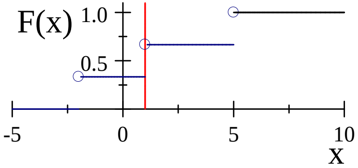
$$F(x) = \frac{x^2-1}{x^2+1} = 0.5,\ x^2 = 3,\ \ x = \sqrt{3} = 1.73$$

Statement

If ξ is discrete and $F(x) \neq 0.5$, then the median of ξ is the value x where F jumps the level 0.5.

Proof $P(\xi \geq x) \geq 0.5 \Rightarrow F(x) \leq 0.5$,
$P(\xi \leq x) = F(x) + P(\xi = x) = \lim_{u \to x+} F(u) \geq 0.5$ ∎

$$\xi : \begin{pmatrix} -2 & 1 & 5 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} ; F(x) = \begin{cases} 0 & \text{if } x \leq -2 \\ 1/3 & \text{if } -2 < x \leq 1 \\ 2/3 & \text{if } 1 < x \leq 5 \\ 1 & \text{if } 5 < x \end{cases}$$



median: 1

Statement

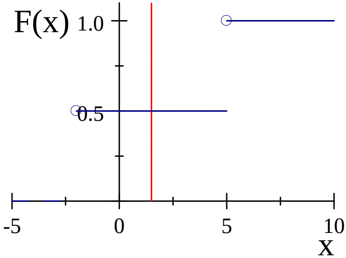
If the c.d.f. of c $F(x)$ takes the value 0.5 for $a < x \leq b$, then the median of ξ is any point in $(a, b]$

Usually $\dfrac{a+b}{2}$ is used.

Proof $P(\xi \leq x) \geq P(\xi < x) = F(x) = 0.5; P(\xi \geq x) = 1 - F(x) = 0.5.$

■

$$\xi : \begin{pmatrix} -2 & 5 \\ 1/2 & 1/2 \end{pmatrix};\ F(x) = \begin{cases} 0 & \text{ha } x \le -2 \\ 1/2 & \text{ha } -2 < x \le 5 \\ 1 & \text{ha } 5 < x \end{cases}$$



median: $\frac{-2+5}{2} = 1.5$

Let ξ , η be r.v.-s, $E(\xi), E(\eta)$ exist

Definition

The covariance of ξ and η is

$$cov(\xi, \eta) = E\left((\xi - E(\xi)) \cdot (\eta - E(\eta))\right)$$

if the expectation exists.

1. $cov(\xi, \eta) = cov(\eta, \xi)$
2. $cov(\xi, \xi) = E\left((\xi - E(\xi)) \cdot (\xi - E(\xi))\right) = D^2(\xi)$
3. $cov(\xi, \eta) = cov(\xi - E(\xi), \eta - E(\eta))$
4. $cov(c\xi + a, d\eta + b) = cd cov(\xi, \eta)$
5. $cov(c\xi + a, c\xi + a) = c^2 D^2(\xi)$
6. If $\xi = c$, *then* $cov(\xi, \eta) = 0$.

7. $cov(\xi, \eta) = E(\xi \cdot \eta) - E(\xi) \cdot E(\eta))$,
   as $cov(\xi, \eta) = E\left((\xi - E(\xi)) \cdot (\eta - E(\eta))\right)$
   $= E\left(\xi \cdot \eta - \xi \cdot E(\eta) - E(\xi) \cdot \eta) + E(\xi)E(\eta)\right)$.
8. If $\mathrm{E}(\xi) = 0$, then $cov(\xi, \eta) = E(\xi \cdot \eta)$
9. $\mathrm{D}^2(\xi + \eta) = D^2(\xi) + D^2(\eta) + 2cov(\xi, \eta)$

Statement

If $E(\xi^2)$ *and* $E(\eta^2)$ *exist, then* $cov(\xi, \eta)$ *exists and* $|cov(\xi, \eta)| \leq D(\xi) \cdot D(\eta)$.
Equality holds if and only if $\xi = a\eta + b$ *or* $\eta = c\xi + d$.

Statement

If ξ and η are independent, then cov$(\xi, \eta) = 0$.

Statement

If cov (ξ, η) =0 $\nRightarrow$ ξ and η are independent.

Example

An example when $cov\,(\xi, \eta)$ =0, but ξ *and* η are not independent:

Let ξ be continuous r.v. with p.d.f. $f(x) = \begin{cases} 0.5, if \;\; -1 \le x \le 1 \\ 0, otherwise \end{cases}$.

Let $\eta = \xi^2$.Then ξ and η are not independent, but $cov(\xi, \eta) = 0$.

$E(\xi \cdot \eta) = E(\xi^3) = \int_{-1}^{1} x^3 \frac{1}{2} dx = 0.$
$E(\xi) = 0, E(\xi) \cdot E(\eta) = 0 \Longrightarrow$
$cov(\xi, \eta) = E(\xi \cdot \eta) - E(\xi) \cdot E(\eta) = 0.$

NOTE: The zero value of the covariance does not imply independence.

Definition

ξ, η are. r.v.-s, $D(\xi), D(\eta)$ are finite, $D(\xi) \neq 0, D(\eta) \neq 0$.
The correlation coefficient of ξ and η is $r(\xi, \eta) = \frac{cov(\xi,\eta)}{D(\xi) \cdot D(\eta)}$.
If $D(\xi) = 0$, then $P(\xi = E(\xi)) = 1$, and $cov(\xi, \eta) = 0$, let $r(\xi, \eta) = 0$, by definition.
If $D(\eta) = 0$, then let $r(\xi, \eta) = 0$, by definition.

1. $|r(\xi,\eta)| \leq 1$, Equality holds if and only if there is a linear relation between the two r.v.-s.
2. If ξ, η are independent, then $r(\xi,\eta) = 0$.
3. $|r(\xi,\eta)| = |r(a\xi + b, c\eta + d)|$.
4. $r(\frac{\xi - E(\xi)}{D(\xi)}, \frac{\eta - E(\eta)}{D(\eta)}) = cov(\frac{\xi - E(\xi)}{D(\xi)}, \frac{\eta - E(\eta)}{D(\eta)}) = r(\xi,\eta)$.

Frequently used discrete distributions

1. Characteristically distributed random variables.
2. Uniformly distributed discrete random variables
3. Binomially distributed random variables.
4. Poisson distributed random variables.
5. Hypergeometrically distributed random variables.

Definition

R.v. ξ is called characteristically distributed random variable with parameter $0 < p < 1$ if

$$\xi : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

Statement

$$E(\xi) = p,\ D(\xi) = \sqrt{p \cdot (1-p)}.$$

Example: A is an event, $P(A) = p$, $0 < p < 1$

$$\xi = \mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases} \qquad \left(\mathbf{1}_A = \begin{cases} 1 & \text{if} \quad A \quad \text{occurs} \\ 0 & \text{if} \quad A \quad \text{does not occur} \end{cases} \right).$$

Frequently used discrete distributions

Uniformly distributed r.v.

Definition

R.v. ξ is called uniformly distributed discrete random variable, if

$$\xi : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

Statement

$$E(\xi) = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}, \quad D(\xi) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2}.$$

Example: Roll a fair dice. Let ξ be the square of the result.

$\xi : \begin{pmatrix} 1 & 4 & 9 & 16 & 25 & 36 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$. All possible values are equally likely.

Definition

R.v. ξ is called binomially distributed random variable with parameters $2 \leq n$, and $0 < p < 1$ if

$$\xi : \begin{pmatrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{pmatrix}$$

with

$$p_k = P(\xi = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Statement

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

Application: Repeat an experiment $n \geq 2$ times independently, let A be an event with $0 < P(A) < 1$. Let ξ be the number of the occurrences of A during the experiments. Then ξ is binomially distributed r.v. with parameters n and $p = P(A)$ $(2 \leq n,\ 0 < P(A) < 1)$.

Example

There are N objects, S of them are of first quality, $N - S$ of them are of second quality. Sample n with replacement. Let ξ be the number of elements of first quality in the sample. Then ξ is binomially distributed r.v. with parameters n and $p = \dfrac{S}{N}$.

Statement

The sum of n independent characteristically distributed r.v. .with parameter p is binomially distributed r.v. with parameters n, *and* $p.(2 \leq n,\ 0 < p < 1)$.

Proof $\xi = \sum_{i=1}^{n} \mathbf{1}_i, \quad \mathbf{1}_1, \cdots, \mathbf{1}_n$ independent

the values in the sum are 0 or 1.

the possible values of $\sum_{i=1}^{n} \mathbf{1}_i$ are $0, 1, 2, \ldots, n$ and

$$P\left(\sum_{i=1}^{n} \mathbf{1}_i = k\right) = \binom{n}{k} \cdot P\left(\mathbf{1}_1 = 1\right) \cdot \cdots \cdot P\left(\mathbf{1}_k = 1\right) \cdot$$
$$P\left(\mathbf{1}_{k+1} = 0\right) \cdots P\left(\mathbf{1}_n = 0\right) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}. \quad \blacksquare$$

Repeat an experience n times, let A be a fixed event and let ξ be the number of occurrences of A during the experiments. Then

$$\xi = \sum_{i=1}^{n} \mathbf{1}_i,$$

with

$$\mathbf{1}_i = \begin{cases} 1 & \text{if } A \text{ occurs during the } i^{\text{th}} \text{ experiment} \\ 0 & \text{if } A \text{ does not occur during the } i^{\text{th}} \text{ experiment} \end{cases}$$

and $\mathbf{1}_i \quad i = 1, 2, \ldots, n$ are independent.

Statement

If ξ is binomially distributed r.v. with parameters n and p, then

$$E(\xi) = np, \qquad D(\xi) = \sqrt{n \cdot p \cdot (1-p)}\ .$$

Proof $E(\xi) = E\left(\sum_{i=1}^{n} \mathbf{1}_i\right) = \sum_{i=1}^{n} E(\mathbf{1}_i) = \sum_{i=1}^{n} p = n \cdot p$

$$D^2(\xi) = D^2\left(\sum_{i=1}^{n} \mathbf{1}_i\right) = \sum_{i=1}^{n} D^2(\mathbf{1}_i) = \sum_{i=1}^{n} p \cdot (1-p) = n \cdot p \cdot (1-p)$$

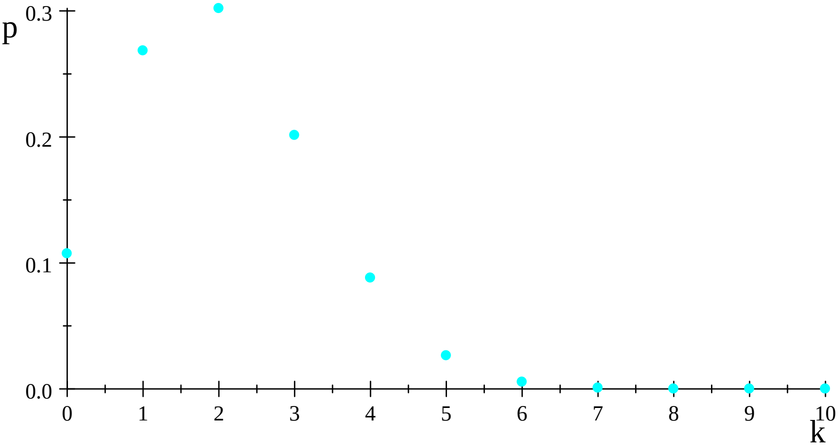
$D(\xi) = \sqrt{n \cdot p \cdot (1-p)}$. ■

Statement

If ξ is binomially distributed r.v. with parameters n and p, then the mode of ξ is

- $[(n+1)\cdot p]$, *if* $(n+1)\cdot p$ *is not integer*
- $(n+1)\cdot p$ *and* $(n+1)\cdot p-1$, *if* $(n+1)\cdot p$ *is integer.*

Let $n = 10$, $p = 0.2$, $(n+1) \cdot p = 2.2$ is not integer $\Rightarrow$ mode$= 2$



Statement

If ξ_n are binomially distributed r.v.-s with parameters n and p_n; k is fixed and $\lim\limits_{n\to\infty} np_n = \lambda > 0$.
Then

$$P(\xi_n = k) = \binom{n}{k} p_n^k \cdot (1 - p_n)^{n-k} \xrightarrow{n\to\infty} \frac{\lambda^k}{k!} e^{-\lambda}.$$

Definition

R.v. ξ is called Poisson distributed r.v. with parameter $\lambda > 0$, if

$$\xi : \begin{pmatrix} 0 & 1 & \dots \\ p_0 & p_1 & \dots \end{pmatrix} \text{ with } p_k = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad k = 0, 1, 2, \cdots .$$

Statement

$$0 \le p_k, \text{ and } \sum_{k=0}^{\infty} p_k = 1.$$

Proof $\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$ ■

$$p_k = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad k = 0, 1, 2, \cdots .$$

Statement

If ξ is Poisson distributed r.v. with $\lambda > 0$ then

$$E(\xi) = \lambda, \qquad D(\xi) = \sqrt{\lambda}.$$

Statement

If ξ is Poisson distributed with $\lambda > 0$, then the mode of ξ is

- $[\lambda]$, *if λ is not integer,*
- λ *and* $\lambda - 1$, *if λ is integer.*

Frequently used discrete distributions

Poisson distributed r.v. ii)

Statement

If ξ_1 is Poisson distributed with parameter $\lambda_1 > 0$ and ξ_2 is Poisson distributed r.v. with $\lambda_2 > 0$, moreover ξ_1 and ξ_2 are independent, then their sum $\xi_1 + \xi_2$ is also Poisson distributed with parameter $\lambda_1 + \lambda_2$.

Definition

The r.v. ξ is called hypergeometrically distributed random variable with parameters N, S, n, $(2 \le S < N,\ 1 \le n \le S,\ n \le N - S\ )$ if

$$\xi : \begin{pmatrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{pmatrix}$$

with

$$p_k = P(\xi = k) = \frac{\binom{S}{k} \cdot \binom{N-S}{n-k}}{\binom{N}{n}} \quad k = 0, 1, 2, \cdots, n$$

Example

Sampling without replacement:

N elements, S is of first quality, $N-S$ is of second quality.
Choose n out of them without replacement. $(1 \leq n \leq S, n \leq N-S)$
Let ξ be the number of elements of first quality in the sample.
Then the possible values of ξ are $0, 1, \cdots, n$, and

$$P(\xi = k) = \frac{\binom{S}{k} \cdot \binom{N-S}{n-k}}{\binom{N}{n}} \quad k = 0, 1, 2, \cdots, n.$$

Statement

If the r.v. ξ is hypergeometrically distributed with N, S, n then

$$E(\xi) = n \cdot \frac{S}{N}$$

$$D(\xi) = \sqrt{n \cdot \frac{S}{N} \cdot (1 - \frac{S}{N}) \cdot \left(1 - \frac{n-1}{N-1}\right)}$$

Statement

If the r.v.-s ξ_N are hypergeometrically distributed r.v.-s with N, S_N, $\lim\limits_{N\to\infty} \dfrac{S_N}{N} = p$, moreover n , k is fixed and then

$$P(\xi_N = k) = \frac{\binom{S_N}{k} \cdot \binom{N - S_N}{n-k}}{\binom{N}{n}} \xrightarrow{N\to\infty} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Frequently used continuous distributions

1. uniform
2. exponential
3. Weibull
4. normal (Gauss)
5. Chi-square
6. Erlang

Definition

R.v. ξ is called uniformly distributed on the interval $[a, b]$ ($a<b$), if its p.d.f. is

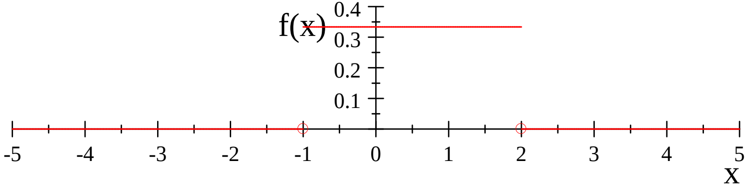
$$f(x) = \begin{cases} c & \text{if} \quad a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Remark

$\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow c = \dfrac{1}{b-a} > 0.$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if} \quad a \le x \le b \\ 0 & \textit{otherwise} \end{cases}$$

$a = -1, \; b = 2$



Statement

If ξ is uniformly distributed r.v. on [a,b], then

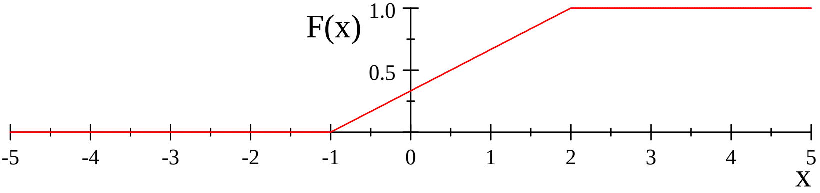
$$F(x) = \begin{cases} 0 & \text{if} \quad x \le a \\ \dfrac{x-a}{b-a} & \text{if} \quad a < x \le b \\ 1 & \text{if} \quad b < x \end{cases}$$

Proof

$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{a}^{x} \frac{1}{b-a} dt = \frac{1}{b-a} \left[t\right]_a^x = \dfrac{x-a}{b-a}, \quad \text{if} \quad a \le x \le b.$ ■

$$F(x) = \begin{cases} 0 & \text{if} \quad x \le a \\ \dfrac{x-a}{b-a} & \text{if} \quad a < x \le b \\ 1 & \text{if} \quad b < x \end{cases}$$

$a = -1, \; b = 2$



Statement

If ξ is continuous uniformly distributed r.v. on $[a, b]$ and $a \leq c < d \leq b$, then

$$P(c < \xi < d) = \frac{d-c}{b-a}.$$

Proof $P(c < \xi < d) = F(d) - F(c) = \dfrac{d-a}{b-a} - \dfrac{c-a}{b-a} = \dfrac{d-c}{b-a}.$

(geometric probability) ∎

Statement

If ξ is continuos uniformly distributed r.v. on $[a, b]$, then

$$\begin{aligned} E(\xi) &= \frac{a+b}{2}, \\ D(\xi) &= \frac{b-a}{\sqrt{12}} \end{aligned}$$

Definition

ξ is exponentially distributed r.v., with parameter $\lambda > 0$ if its p.d.f. is

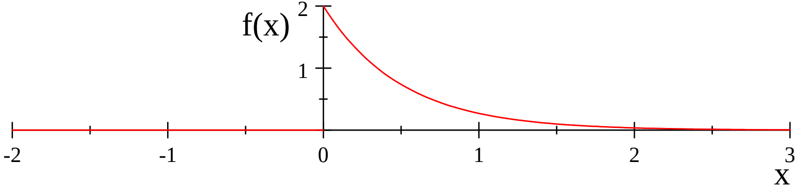
$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{if} \quad 0 \leq x \\ 0 & \text{otherwise} \end{cases}$$

Remark

$$0 \leq f(x),\ \int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \lambda \cdot e^{-\lambda x} dx = \lambda \left[\frac{e^{-\lambda x}}{-\lambda} \right]_0^{\infty} = 0 - (-1) = 1.$$

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{if} \quad 0 \leq x \\ 0 & \textit{otherwise} \end{cases}$$

$\lambda = 2$



Statement

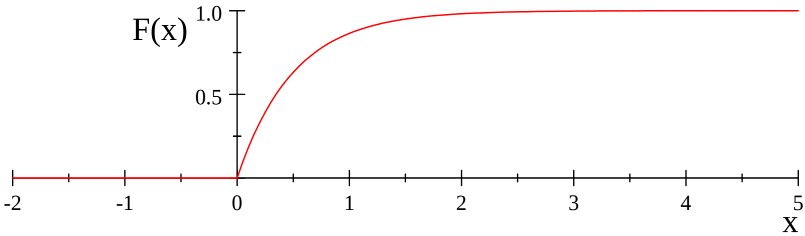
If ξ is exponentially distributed r.v. with $\lambda > 0$ then

$$F(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ 1 - e^{-\lambda x} & \text{if} \quad 0 < x \end{cases}$$

Proof $F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x \lambda \cdot e^{-\lambda t} dt = \left[\frac{e^{-\lambda t}}{-1} \right]_0^x =$
$-e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}$ if $0 < x$. ∎

$$F(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ 1 - e^{-\lambda x} & \text{if} \quad 0 < x \end{cases}$$

$\lambda = 2$



Statement

If ξ exponentially distributed with a $\lambda > 0$, then

$$E(\xi) = \frac{1}{\lambda}, \qquad D(\xi) = \frac{1}{\lambda}.$$

Proof $E(\xi) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x} dx = \left[x \cdot \frac{\lambda e^{-\lambda x}}{-\lambda} \right]_0^{\infty}$

$- \int_0^{\infty} 1 \cdot \frac{\lambda e^{-\lambda x}}{-\lambda} dx = 0 + \int_0^{\infty} 1 \cdot e^{-\lambda x} dx = \left[\frac{e^{-\lambda x}}{-\lambda} \right]_0^{\infty} = 0 - (-\frac{1}{\lambda}) = \frac{1}{\lambda}.$

■

Statement

If ξ is exponentially distributed r.v. with $\lambda > 0$, then for any $0 \leq x,\ 0 \leq y$

$$P(\xi > x + y \mid \xi > x) = P(\xi > y).$$

Proof $P(\xi > y) = 1 - F(y) = e^{-\lambda y}$.

$$P(\xi > x + y \mid \xi > x) = \frac{P\left(\xi > x + y \cap \xi > x\right)}{P\left(\xi > x\right)} = \frac{P(\xi > x + y)}{P\left(\xi > x\right)} =$$

$$\frac{1 - F(x + y)}{1 - F(x)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y}. \quad \blacksquare$$

Statement

If ξ has only nonnegative values, F is differentiable, $F(x) \neq 1$, $\lim\limits_{x \to 0+} F'(x) = \lambda > 0$, and

$$P(\xi > x + y \mid \xi > x) = P(\xi > y) \quad 0 \leq x,\ y$$

then ξ is exponentially distributed r.v. with parameter λ.

Statement

Let $\xi_1, \xi_2, ..., \xi_n, \ldots$ be independent, exponentially distributed r.v. with λ, $0 \leq T$ fixed. Let

$$\eta_T = \begin{cases} 0 & \text{if} \quad T < \xi_1 \\ 1 & \text{if} \quad \xi_1 \leq T \quad \text{and} \quad T < \xi_1 + \xi_2 \\ \vdots & \vdots \\ k & \text{if} \quad \sum_{i=1}^{k} \xi_i \leq T \quad \text{and} \quad T < \sum_{i=1}^{k+1} \xi_i \\ \vdots & \vdots \end{cases} .$$

Then η_T is Poisson distributed r.v. with parameter $\lambda \cdot T$.

Frequently used continuous distributions

Exponentially distributed r.v. - Relation between exponential and Poisson distribution ii)

|————————————————————————T

|________|_____________|_________|_____|_________

$0 < - - - > < - - - - - - - - - > < - - - - - - - - - - >$

$\ldots\ldots\ldots\xi_1\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\xi_2\ldots\ldots\ldots\ldots\ldots\xi_1+\xi_2\ldots\ldots\ldots\ldots\xi_3\ldots\ldots\ldots\xi_1+\xi_2+\xi_3$

$\xi_1 =$ lifetime of the first part $=$ the time point of the first change

$\xi_2 =$ lifetime of the second part, $\xi_1 + \xi_2 =$ the time point of the second change

$\xi_3 =$ lifetime of the third part, $\xi_1 + \xi_2 + \xi_3 =$ the time point of the third change

the sum of the first and second r.v. is less than T, but the sum of 3 r.v.-s is more than T

the number of changes to T is the number of r.v.-s whose sum is less than T but increasing the number of the r.v.-s by 1, their sum exceeds the value of T.

Definition

The distribution of the r.v. ξ is called Weibull with parameters $\lambda > 0$ and $k > 0$, if its c.d.f. is

$$F(x) = \begin{cases} 1 - e^{-(\lambda \cdot x)^k} & \text{if} \quad 0 \leq x \\ 0 & \text{otherwise} \end{cases}$$

Remark

F is c.d.f. as it is continuous, monotone increasing and its limit at -∞ equals 0, at ∞ equals 1.

Statement

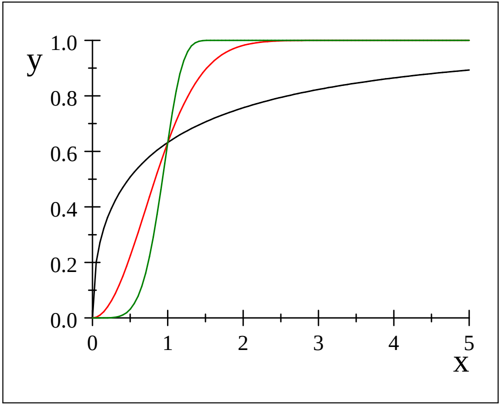
If the distribution of ξ is Weibull distribution, then its p.d.f is

$$f(x) = \begin{cases} \lambda \cdot k \cdot (\lambda \cdot x)^{k-1} e^{-(\lambda \cdot x)^k} & \textit{if} \quad 0 \leq x \\ 0 & \textit{otherwise} \end{cases}$$

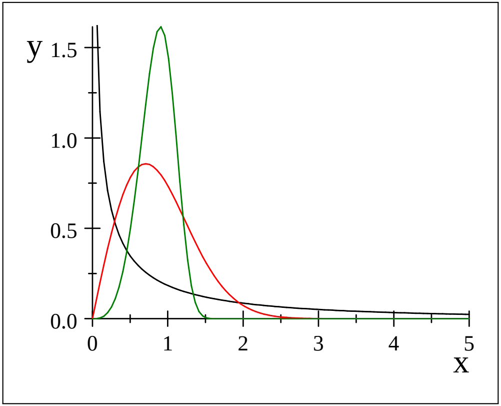
and $E(\xi) = \frac{1}{\lambda}\Gamma(1 + \frac{1}{k})$.

$\Gamma(s) =: \int_0^\infty t^s - 1 \cdot e^{-t} ds; \Gamma(s) = (s-1) \cdot \Gamma(s-1)$; generalization of s!
Note: Weibull distribution is a generalization of exponential distribution (k=1).

c.d.f.-s in case of $\lambda = 1$, $k = 0.5 - black$; $k = 2 - red$; $k = 5$ - green



p.d.f.-s in case of $\lambda = 1$, $k = 0.5 - black$; $k = 2 - red$; $k = 5$ - green



Definition

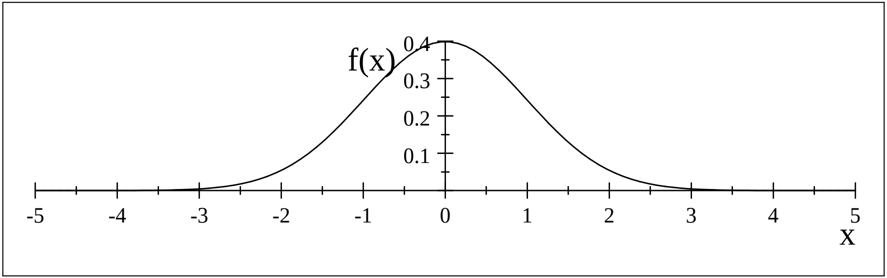
The r.v. ξ is called standard normally distributed if its p.d.f. is
$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ *for any $x \in \mathbb{R}$.*
Notation: $\xi \sim N(0,1)$

Remark

$0 < \varphi(x)$, $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi} \Rightarrow \int_{-\infty}^{\infty} \varphi(x) dx = 1.$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\varphi(0) = \frac{1}{\sqrt{2\pi}} = 0.398\,94$$



Statement

If $\xi \sim N(0,1)$, *then its c.d.f. is* $\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$.

The values are contained in tables.

x	$\Phi(x)$
0	0.5
1	0.8413
2	0.9772
4	1

Statement

$\Phi(-x) = 1 - \Phi(x)$.

Statement

If $\xi \sim N(0,1)$, *then*

$$E(\xi) = 0, \qquad D(\xi) = 1$$

Proof $E(\xi) = \int_{-\infty}^{\infty} x \cdot \varphi(x)dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx =$
$\frac{1}{\sqrt{2\pi}} \left[-e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} = \frac{1}{\sqrt{2\pi}}(0-0) = 0.$
$E(\xi^2) = \int_{-\infty}^{\infty} x^2 \cdot \varphi(x)dx = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx =$
$\frac{1}{\sqrt{2\pi}} \left[-x \cdot e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 + 1 = 1.$
$D^2(\xi) = E(\xi^2) - (E(\xi))^2 = 1 \Rightarrow D(\xi) = 1.$ ∎

Statement

If $\xi \sim N(0,1)$, *then* $\eta = -\xi \sim N(0,1)$.

Proof $F_\eta(x) = P(\eta < x) = P(-\xi < x) = P(0 < \xi + x) = P(-x < \xi) = 1 - \Phi(-x) = 1 - (1 - \Phi(x)) = \Phi(x)$.

$f_\eta(x) = (F_\eta(x))' = (\Phi(x))' = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. ■

Definition

Let $\xi \sim N(0,1)$, $m \in \mathbb{R}$, $0 < \sigma$. $\eta = \sigma \cdot \xi + m$ is called normally (Gauss) distributed r.v. with parameters m and σ. Notation: $\eta \sim N(m, \sigma)$.

Remark

In case of $m = 0$, $\sigma = 1$ η is standard normally distributed.

Remark

Let $\xi \sim N(0,1)$, $m \in \mathbb{R}$ $\quad a < 0 \quad \eta = a \cdot \xi + m$ is normally distributed. as $\eta = -a \cdot (-\xi) + m$, $0 < -a$ and $-\xi \sim N(0,1)$.

Statement

If $\eta \sim N(m, \sigma)$, then

$$F_\eta(x) = \Phi(\frac{x-m}{\sigma}),$$

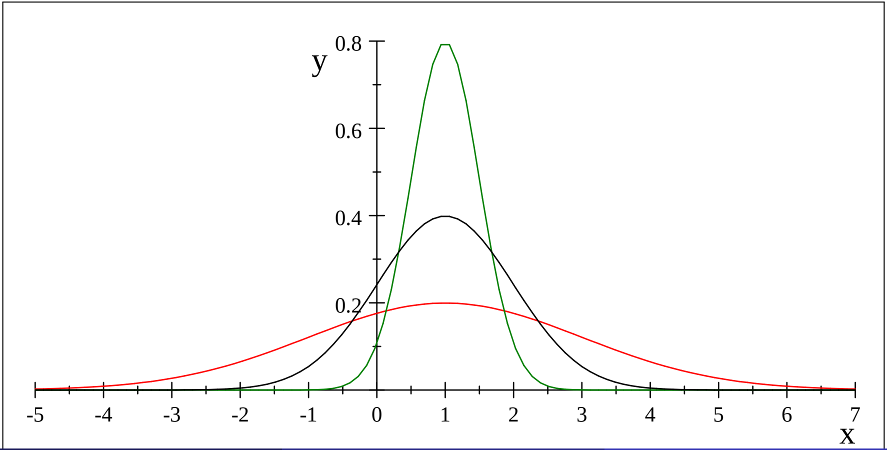
and

$$f_\eta(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Proof $F_\eta(x) = P(\eta < x) = P(\sigma\cdot\xi + m < x) = P(\sigma\cdot\xi < x - m) = P(\xi < \frac{x-m}{\sigma}) = \Phi\left(\frac{x-m}{\sigma}\right)$.

$$f_\eta(x) = \left(F_\eta(x)\right)' = \Phi'\left(\frac{x-m}{\sigma}\right)\cdot\frac{1}{\sigma} = \varphi(\frac{x-m}{\sigma})\cdot\frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{\left(\frac{x-m}{\sigma}\right)^2}{2}}.$$

m=1, $\sigma = 0.5$: green m=1, $\sigma = 2$: red, m=1, $\sigma = 1$ black



Statement

If $\eta \sim N(m, \sigma)$, then

$$E(\eta) = m, \qquad D(\eta) = \sigma.$$

Proof $E(\eta) = E(\sigma \cdot \xi + m) = \sigma \cdot E(\xi) + m = \sigma \cdot 0 + m = m.$
$D(\eta) = D(\sigma \cdot \xi + m) = \sigma \cdot D(\xi) = \sigma \cdot 1 = \sigma.$ ■

Statement

($k \cdot \sigma$ rule)
If $\eta \sim N(m, \sigma)$, then $P(m - k \cdot \sigma < \eta < m + k \cdot \sigma) = 2\Phi(k) - 1$.

Proof $P(m - k \cdot \sigma < \eta < m + k \cdot \sigma) = F_\eta(m + k \cdot \sigma) - F_\eta(m - k \cdot \sigma) = \Phi(\frac{m + k\sigma - m}{\sigma}) - \Phi(\frac{m - k\sigma - m}{\sigma}) = \Phi(k) - \Phi(-k) = 2 \cdot \Phi(k) - 1$. ∎

Corollary

$k = 1 : P(m - \sigma < \eta < m + \sigma) = 2\Phi(1) - 1 = 0.6826$
$k = 2 : P(m - 2\sigma < \eta < m + 2\sigma) = 2\Phi(2) - 1 = 0.9544$
$k = 3 : P(m - 3\sigma < \eta < m + 3\sigma) = 2\Phi(3) - 1 = 0.9974$

Statement

If $\eta \sim N(m, \sigma)$, then $\theta = a \cdot \eta + b \sim N(a \cdot m + b, |a| \cdot \sigma)$, if $a \neq 0$.

Proof

$\theta = a \cdot \eta + k = a(\sigma\xi + m) + b = a\sigma\xi + am + b = (-a\sigma)(-\xi) + am + b$
where $\xi \sim N(0, 1)$. ■

Statement

If $\eta_1 \sim N(m_1, \sigma_1)$, $\eta_2 \sim N(m_2, \sigma_2)$, independent, then

$$\eta_1 + \eta_2 \sim N\left(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right).$$

Statement

If $\eta_i \sim N(m, \sigma)$ *i=1,2,...,n are independent, then*

$$\sum_{i=1}^{n} \eta_i \sim N(n \cdot m, \sqrt{n} \cdot \sigma).$$

Statement

If $\eta_i \sim N(m, \sigma)$ *i=1,2,...,n are independent, then*

$$\frac{1}{n} \sum_{i=1}^{n} \eta_i \sim N\left(m, \frac{\sigma}{\sqrt{n}}\right) .$$

Definition

Let $\xi \sim N(0,1)$. Then $\eta = \xi^2$ is called Chi-square distributed r.v. with degree of freedom 1.

Statement

If η is Chi-square distributed r.v. with degree of freedom 1, then

$$F_\eta(x) = \begin{cases} 2\Phi(\sqrt{x}) - 1 & \text{if } 0 \le x \\ 0 & \text{if } x < 0 \end{cases},$$

Proof $F_\eta(x) = P(\eta < x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2 \cdot \Phi(\sqrt{x}) - 1.$ ■

Statement

If η is Chi-square distributed r.v. with degree of freedom 1, then

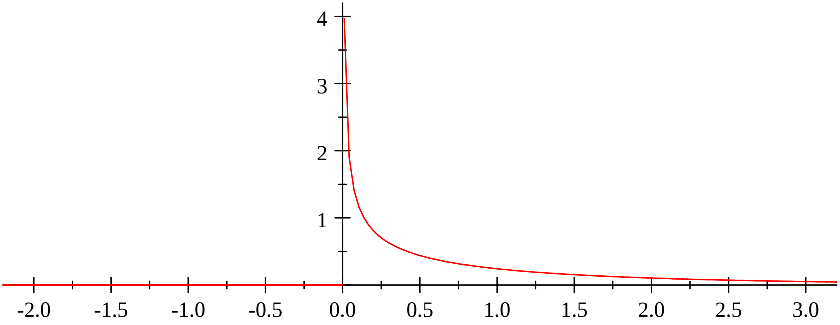
$$f_\eta(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\cdot\sqrt{x}} e^{-\frac{x}{2}} & \text{if } 0 < x \\ 0 & \text{if } x \leq 0 \end{cases} \quad \text{and } E(\eta) = 1, D(\eta) = \sqrt{2}.$$

Proof $f_\eta(x) = \left(F_\eta(x)\right)' = \left(2\Phi(\sqrt{x}) - 1\right)' = 2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(\sqrt{x})^2}{2}} \left(\sqrt{x}\right)' =$

$$2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(\sqrt{x})^2}{2}} \cdot \frac{1}{2\sqrt{x}}.$$

$E(\eta) = E(\xi^2) = D^2(\xi) = 1.$ ∎

p.d.f:



Statement

Let ξ and η be independent continuous random variables with p.d.f.-s f_ξ and f_η, respectively. Then $\xi + \eta$ is also continuous random variable and its p.d.f. is

$$f_{\xi+\eta}(z) = \int_{-\infty}^{\infty} f_\xi(z-x) f_\eta(x)dx = \int_{-\infty}^{\infty} f_\xi(x) f_\eta(z-x)dx$$

Definition

Let $\xi_1 \sim N(0,1)$, $\xi_2 \sim N(0,1)$ be independent r.v.-s. Then $\eta = \xi_1^2 + \xi_2^2$ is called Chi-square distributed r.v. with degree of freedom 2.

Statement

The p.d.f. of $\eta = \xi_1^2 + \xi_2^2$ is

$$f_\eta(x) = \begin{cases} 0.5e^{-\frac{x}{2}} & \text{if } 0 < x \\ 0 & \text{if } x \le 0 \end{cases},$$

and

$$E(\eta) = 2, D(\eta) = 2.$$

Proof apply convolution formula,

$$f_{\xi_1}(z-x)=\begin{cases}\frac{1}{\sqrt{2\pi}\cdot\sqrt{z-x}}e^{-\frac{z-x}{2}} & if \quad 0<z-x\\ 0 & if \quad z-x\leq 0\end{cases}, f_{\xi_2}(x)=$$

$$\begin{cases}\frac{1}{\sqrt{2\pi}\cdot\sqrt{x}}e^{-\frac{x}{2}} & if \quad 0<x\\ 0 & if \;\; x\leq 0\end{cases}$$

$$f_{\xi_1+\xi_2}(z)=\int_{-\infty}^{\infty}f_{\xi_1}(z-x)\cdot f_{\xi_2}(x)dx=$$

$$\int_0^z\frac{1}{\sqrt{2\pi}\cdot\sqrt{z-x}}e^{-\frac{z-x}{2}}\cdot\frac{1}{\sqrt{2\pi}\cdot\sqrt{x}}e^{-\frac{x}{2}}dx=\frac{1}{2\pi}\exp\left(-\frac{z}{2}\right)\int_0^z\frac{1}{\sqrt{(z-x)x}}dx$$

$$\int_0^z\frac{1}{\sqrt{(z-x)x}}dx=\int_0^z\frac{1}{\sqrt{zx-x^2}}dx, 1-\frac{x}{\frac{z}{2}}=a, \int_0^z\frac{1}{\sqrt{zx-x^2}}dx=\int_{-1}^{1}\frac{1}{\sqrt{1-a^2}}da=$$

$$\left[\arcsin a\right]_{-1}^{1}=\pi,$$

$$f_{\xi_1+\xi_2}(z)=\frac{1}{2}\exp\left(-\frac{z}{2}\right), 0<z \quad \blacksquare$$

Definition

Let $\xi_i \sim N(0,1), i = 1, 2, \ldots, n,$ *be independent r.v.-s. Then* $\eta = \sum_{i=1}^{n} \xi_i^2$ *is called Chi-square distributed r.v. with degree of freedom n.*

Statement

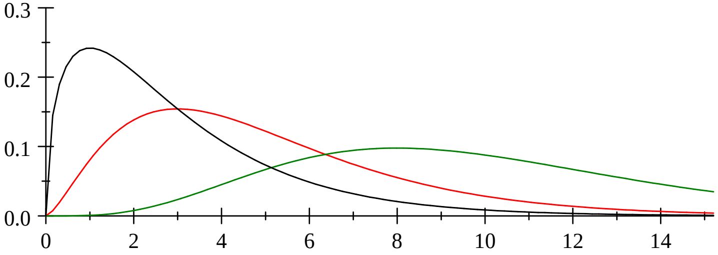
If η *is Chi-square distributed r.v. with degree of freedom n, then its p.d.f. is*

$$f_\eta(x) = \begin{cases} c \cdot x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{if } 0 < x \\ 0, & \text{if } x \leq 0 \end{cases}$$

and

$$E(\eta) = n \qquad D(\xi) = \sqrt{2n}$$

p.d.f. $f_\eta(x)$;
$n = 3$ - black, $n = 5 - red$, $n = 10$ - green



Definition

Let ξ_i i=1,2,...,n independent exponentially distributed r.v.-s with parameters $0<\lambda$. Then $\eta = \sum_{i=1}^{n} \xi_i$ is called Erlang distributed r.v. with degree of freedom n.

Statement

Let ξ_1 and ξ_2 independent exponentially distributed r.v.-s with parameter 0<λ. Then the p.d.f. of $\eta = \xi_1 + \xi_2$

$$f_\eta(z) = \begin{cases} \lambda^2 z \exp(-\lambda z), & 0 < z \\ 0 & \textit{otherwise} \end{cases}$$

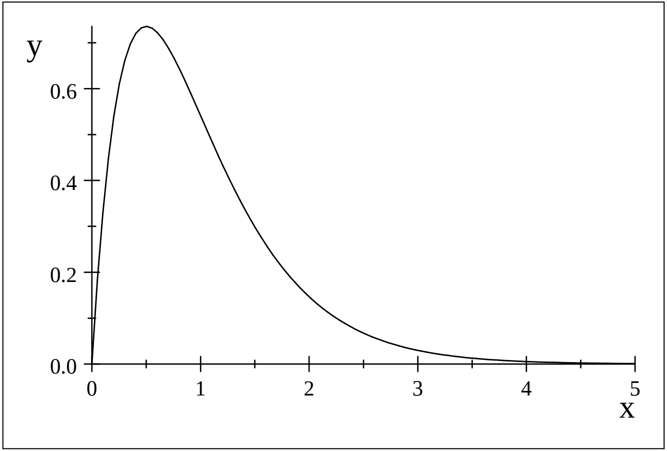
Proof $f_{\xi_1}(z-x) = \begin{cases} 0, \text{ \textit{if} } z < x \\ \lambda \exp(-\lambda(z-x)), \text{ \textit{if} } 0 \le z - x \end{cases},$

$f_{\xi_2}(x) = \begin{cases} 0, \text{\textit{if} } x < 0 \\ \lambda \exp(-\lambda x), \text{ \textit{if} } 0 \le x \end{cases},$

$f_{\xi_1}(x) \cdot f_{\xi_2}(x) = \begin{cases} 0, \text{\textit{if} } x < 0 \text{ or } z < x \\ \lambda^2 \exp(-\lambda z), \text{ if } 0 \le \mathrm{x} \le z \end{cases}$

$f_{\xi_1+\xi_2}(z) = \int_{-\infty}^{\infty} f_{\xi_1}(z-x) \cdot f_{\xi_2}(x) dx = \int_0^z \lambda^2 \exp(-\lambda z) dx =$
$= \lambda^2 z \exp(-\lambda z), 0 < z$ ■

$\lambda = 2; f_\eta(z) = 2^2 z \exp(-2z);$

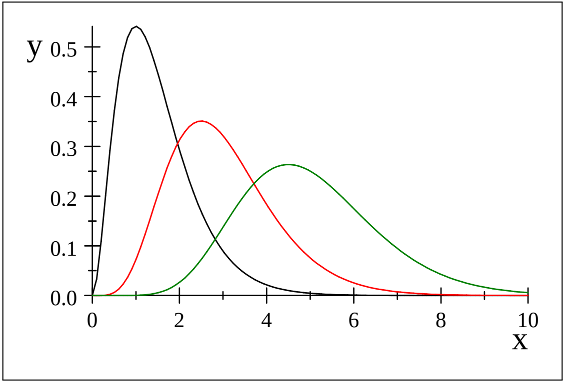


Statement

If ξ_i are independent exponentially distributed r.v.-s with parameter $0<\lambda$, then the p.d.f. of $\sum_{i=1}^{n} \xi_i$ is

$$f_n(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} \exp(-\lambda x), \ 0 < x \\ 0, \ \textit{otherwise} \end{cases}$$

$\lambda = 2; n = 3-$ *black*, $n = 6 -$ *red*, $n = 10 -$ *green*



Statement (Markov inequality)

Let $0 \le \xi$, $E(\xi)$ *be finite and* $0 < \lambda$, *then*

$$P(\xi \geq \lambda) \leq \frac{E(\xi)}{\lambda}.$$

Proof

$$\xi \geq \lambda \cdot \mathbf{1}_{\xi \geq \lambda} = \begin{cases} \lambda & \text{if } \xi \geq \lambda \\ 0 & \textit{if } \xi < \lambda \end{cases}.$$

$$E(\xi) \geq E(\lambda \cdot \mathbf{1}_{\xi \geq \lambda}) = \lambda \cdot E(\mathbf{1}_{\xi \geq \lambda}) = \lambda \cdot P(\xi \geq \lambda),$$

$$\frac{E(\xi)}{\lambda} \geq P(\xi \geq \lambda).$$

■

Statement (Csebisev inequality)

If $D(\eta)$ is finite, $0 < \varepsilon$, then

$$P(|\eta - E(\eta)| \geq \varepsilon) \leq \frac{D^2(\eta)}{\varepsilon^2}.$$

Proof As $0 \leq (\eta - E(\eta))^2$, and $E\left((\eta - E(\eta))^2\right) = D^2(\eta)$ is finite, $\lambda = \varepsilon^2$, apply Markov inequality:

$$\frac{D^2(\eta)}{\varepsilon^2} = \frac{E\left((\eta - E(\eta))^2\right)}{\varepsilon^2} \geq P\left((\eta - E(\eta))^2 \geq \varepsilon^2\right) = P(|\eta - E(\eta)| \geq \varepsilon)$$

■

Statement (Another form of Chebisev inequality)

If $D(\eta)$ *is finite,* $0 < \varepsilon$*, then*

$$P(|\eta - E(\eta)| < \varepsilon) \geq 1 - \frac{D^2(\eta)}{\varepsilon^2}.$$

Proof

$$P(|\eta - E(\eta)| < \varepsilon) = P(\overline{|\eta - E(\eta)| \geq \varepsilon}) =$$

$$1 - P(|\eta - E(\eta)| \geq \varepsilon) \geq 1 - \frac{D^2(\eta)}{\varepsilon^2}.$$

■

Statement

If $D(\eta)$ is finite, $D(\eta) \neq 0$, $0 < k$, then

$$P\left(|\eta - E(\eta)| \geq k \cdot D(\eta)\right) \leq \frac{1}{k^2}$$

$$P\left(|\eta - E(\eta)| < k \cdot D(\eta)\right) \geq 1 - \frac{1}{k^2}$$

Proof Let $\varepsilon = k \cdot D(\eta)$ Then $\dfrac{D^2(\eta)}{\varepsilon^2} = \dfrac{D^2(\eta)}{(k \cdot D(\eta))^2} = \dfrac{1}{k^2}$. ■

Remark

We get bounds for the probabilities, not the exact values. We do not need the distribution.

$$P\left(|\eta - E(\eta)| < kD(\eta)\right) \geq 1 - \frac{1}{k^2}$$

Remark

Large difference-small probability, small difference-large probability.

Especially:

$$\begin{array}{lll} k=1 & 1-\frac{1}{k^2}=0 & P\left(|\eta - E(\eta)| < D(\eta)\right) \geq 0 \\ k=2 & 1-\frac{1}{k^2}=0.75 & P\left(|\eta - E(\eta)| < 2D(\eta)\right) \geq 0.75 \\ k=3 & 1-\frac{1}{k^2}=0.89 & P\left(|\eta - E(\eta)| < 3D(\eta)\right) \geq 0.89 \end{array}$$

Statement

If $\xi_i \quad i = 1, 2, \ldots$ *are independent identically distributed r.v.-s,* $E(\xi_i) = m, \; D(\xi_i) = \sigma$ *are finite, then for any* $0 < \varepsilon$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - m\right| \geq \varepsilon\right) \to 0, \; \textit{if } n \to \infty.$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - m\right| < \varepsilon\right) \to 1, \; \textit{if } n \to \infty$$

Proof Apply the Chebysev's inequality for $\eta = \frac{1}{n}\sum_{i=1}^{n}\xi_i$ ■

Statement (Bernoulli theorem)

Repeat an experiment n times independently, let $k_A(n)$ be the frequency of the event A, $0 < p = P(A) < 1$. Then for any $0 < \varepsilon$

$$P\left(\left|\frac{k_A(n)}{n} - p\right| \geq \varepsilon\right) \to 0,\ \textit{if } n \to \infty$$

and

$$P\left(\left|\frac{k_A(n)}{n} - p\right| < \varepsilon\right) \to 1,\ \textit{if } n \to \infty.$$

Proof $k_A(n)$ is binomially distributed with parameters n, and $p = P(A)$.
$k_A(n)$ is the sum of n independent characteristically distributed r.v.-s.
$E(\mathbf{1}_{A_i}) = P(A) = p$, $D(\mathbf{1}_{A_i}) = \sqrt{p(1-p)}$
Apply the law of large numbers. ■

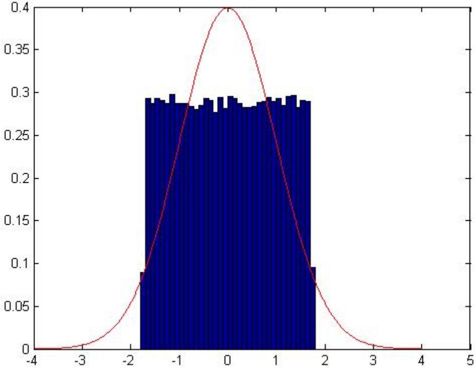
Remark

The relative frequency is close to the probability if the number of experiments is large.

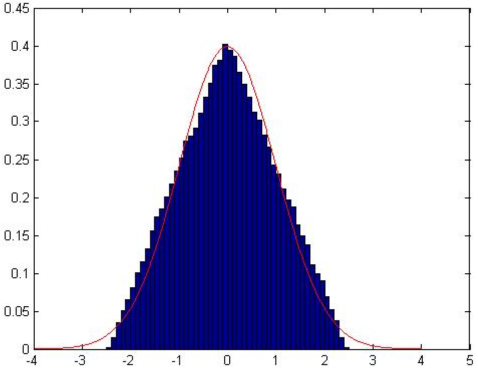
Application: in case of

- pools,
- computer simulations,
- sampling.

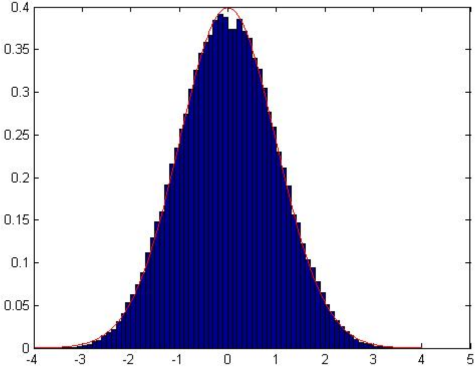
Random numbers generated by uniform distribution (transformation: expectation 0 and dispersion 1)



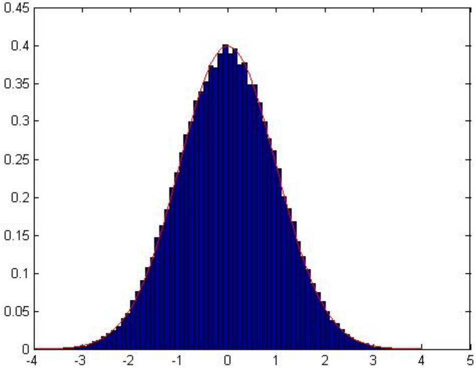
Random numbers (after standardization) - sums of two uniformly distributed random variables



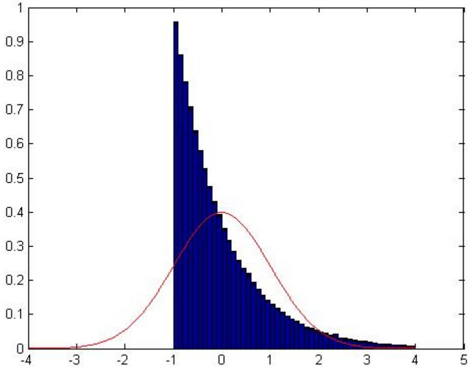
Random numbers (after transformation) - sums of five uniformly distributed random variables



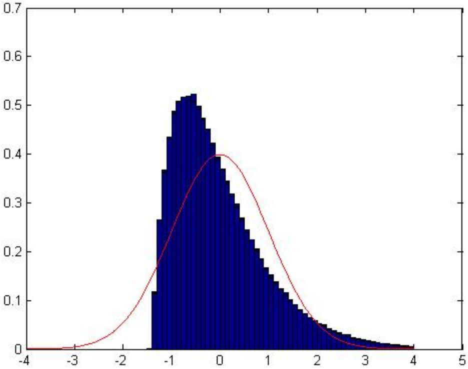
Random numbers (after transformation) - sums of 100 uniformly distributed random variables



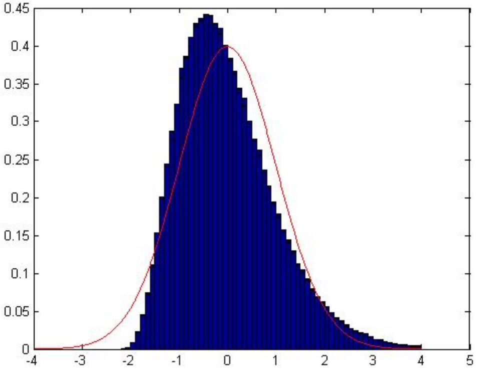
Random numbers generated by exponential distribution and transformed (0 expectation and 1 dispersion)



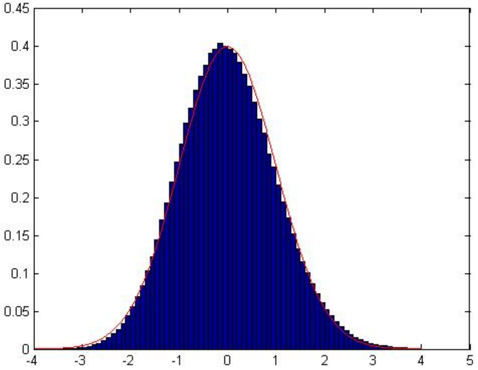
Random numbers (after transformation) - sums of 2 exponentially distributed random variables



Random numbers (after transformation) - sums of 5 exponentially distributed random variables



Random numbers (after transformation)- sums of 100 exponentially distributed random variables



Let ξ_i $i = 1, 2, 3, \ldots$be independent identically distributed r.v.-s.
$E(\xi_i) = m$, $D(\xi_i) = \sigma$.
Take the sum $\sum_{i=1}^{n} \xi_i$.
By the properties of the expectation

$$E\left(\frac{\sum_{i=1}^{n} \xi_i - nm}{\sqrt{n} \cdot \sigma}\right) = \frac{E\left(\sum_{i=1}^{n} \xi_i - nm\right)}{\sqrt{n} \cdot \sigma} = \frac{E\left(\sum_{i=1}^{n} \xi_i\right) - nm}{\sqrt{n} \cdot \sigma} = 0.$$

By the properties of the dispersion

$$D\left(\frac{\sum_{i=1}^{n} \xi_i - nm}{\sqrt{n} \cdot \sigma}\right) = \frac{D\left(\sum_{i=1}^{n} \xi_i - nm\right)}{\sqrt{n} \cdot \sigma} = \frac{D\left(\sum_{i=1}^{n} \xi_i\right)}{\sqrt{n} \cdot \sigma} = 1.$$

$\dfrac{\sum_{i=1}^{n} \xi_i - nm}{\sqrt{n} \cdot \sigma}$ is called standardized sum.

Theorem

Let $\xi_i \quad i = 1, 2, 3, \ldots$ *independent identically distributed r.v.-s,*
$E(\xi_i) = m$, $D(\xi_i) = \sigma$.
Then

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} \xi_i - nm}{\sqrt{n}\cdot\sigma} < x\right) = \Phi(x) \quad \text{for any } x \in \mathbb{R}.$$

Remark

There is no requirement for the distribution of ξ_i*, it can be arbitrary. On the left hand side one can see a sequence of c.d.f.-s. The limit is the standard normal c.d.f..*

Let $\xi_i \quad i = 1, 2, 3, \ldots$ be independent identically distributed r.v.-s,
$E(\xi_i) = m$, $D(\xi_i) = \sigma$.
Then

$$F_{sum}(y) = \approx \Phi\left(\frac{y - nm}{\sqrt{n} \cdot \sigma}\right).$$

$$F_{sum}(y) = P\left(\sum_{i=1}^{n} \xi_i < y\right) = P\left(\sum_{i=1}^{n} \xi_i - nm < y - nm\right) =$$
$$P\left(\frac{\sum_{i=1}^{n} \xi_i - nm}{\sqrt{n} \cdot \sigma} < \frac{y - nm}{\sqrt{n} \cdot \sigma}\right) \approx \Phi\left(\frac{y - nm}{\sqrt{n} \cdot \sigma}\right).$$

$$F_{sum}(y) \approx \Phi\left(\frac{y - nm}{\sqrt{n}\cdot\sigma}\right)$$

- On the left hand side there is a c.d.f..
- On the right hand side there is a **normal** c.d.f..
- The parameters of the normal c.d.f. are: nm and $\sqrt{n}\cdot\sigma$.
- The approximation is good for $100 \leq n$ in case of (almost) any distribution.
- If the distribution of ξ_i $i = 1, 2, 3, \ldots$ is about symmetrical, then it can be applied in case of $30 \leq n$.

Let $\xi_i \quad i = 1, 2, 3, \ldots$ be independent identically distributed r.v.-s,
$E(\xi_i) = m$, $D(\xi_i) = \sigma$.
Let $\eta = \frac{\sum_{i=1}^{n} \xi_i}{n}$ (average).
Then

$$F_\eta(y) \approx \Phi\left(\frac{y-m}{\frac{\sigma}{\sqrt{n}}}\right).$$

$$F_\eta(y) = P(\eta < y) = P(\eta - m < y - m) = P\left(\frac{\eta - m}{\frac{\sigma}{\sqrt{n}}} < \frac{y-m}{\frac{\sigma}{\sqrt{n}}}\right) \approx$$

$$\Phi\left(\frac{y-m}{\frac{\sigma}{\sqrt{n}}}\right).$$

$$F_\eta(y) \approx \Phi\left(\frac{y-m}{\frac{\sigma}{\sqrt{n}}}\right)$$

- On the left hand side there is a c.d.f..
- On the right hand side there is a **normal** c.d.f..
- The parameters of the normal c.d.f. are m and $\dfrac{\sigma}{\sqrt{n}}$.
- These are just the expectation and dispersion of the average.

Theorem

Let A be an event with $P(A) = p$, 0<p<1. Let $k_A(n)$ be the frequency of the event A if we have n independent experiments. Then

$$P\left(a \leq k_A(n) < b\right) \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)$$

Proof $k_A(n)$ is the sum of n independent characteristically distributed random variables.
The c.d.f. of the sum is approximately normal distribution function with parameters $m = E\left(k_A(n)\right) = np, \sigma = D\left(k_A(n)\right) = \sqrt{np(1-p)}$. ■

If $k_A(n)$ is the frequency of event A having n independent experiments, $P(A) = p$, 0<p<1, then

$$P\left(k_A(n) = i\right) = P\left(i \le k_A(n) < i+1\right) \approx$$

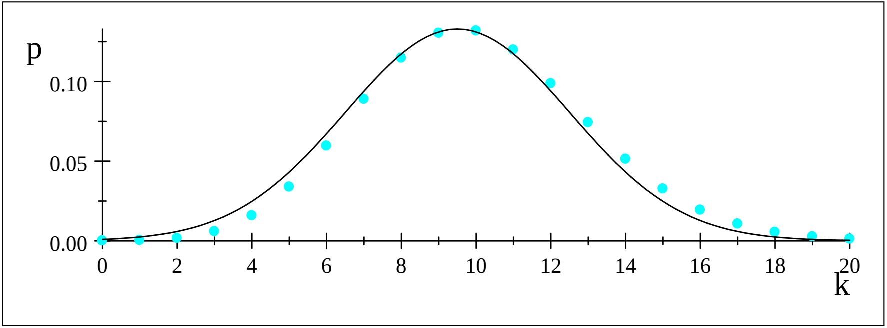
$$\Phi\left(\frac{i+1-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{i-np}{\sqrt{np(1-p)}}\right) =$$

$$\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(z-np)^2}{2\cdot np(1-p)}} \cdot \frac{1}{\sqrt{np(1-p)}} \approx$$

$$\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(i+0.5-np)^2}{2np(1-p)}} \cdot \frac{1}{\sqrt{np(1-p)}}, z \in [i, i+1]$$

The probability can be approximated by the values of normal p.d.f. with parameters np and $\sqrt{np(1-p)}$.

$n = 100, p = 0.1$



The exact probabilities and the approximate p.d.f.

Statement

Let η be Poisson distributed r.v. with 10$\leq \lambda$. Then

$$F_\eta(x) = P(\eta < x) \approx \Phi(\frac{x-\lambda}{\sqrt{\lambda}})$$

Statement

Let η be Poisson distributed with 30$\leq \lambda$ parameter. Then

$$P(\eta = k) = \frac{\lambda^k}{k!}e^{-\lambda} \approx \Phi(\frac{k+1-\lambda}{\sqrt{\lambda}}) - \Phi(\frac{k-\lambda}{\sqrt{\lambda}})$$

Theorem

If $k_A(n)$ is the frequency of the event A having n independent experiments, $P(A) = p$, $0<p<1$, then

$$\lim_{n\to\infty} P\left(\frac{\frac{1}{n}\cdot k_A(n) - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} < x\right) = \Phi(x).$$

If $k_A(n)$ is the frequency of the event A having n independent experiments, $P(A) = p$, $0 < p < 1$,
$\eta_n = \frac{k_A(n)}{n}$ is the relative frequency of A, then

$$F_{\eta_n}(y) \approx \Phi\left(\frac{y-p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}}\right).$$

Note: The parameters of the approximate normal distribution are just the expectation of the relative frequency and the dispersion of the relative frequency.

Example

Roll a fair dice. The gain is the square of the results, if the result is odd, and the let the gain be 0 if the result is even number. Compute the probability that the sum of the gains is at least 550 and less than 600 after 100 games.At least how much is the gain with probability 0.99?

Solution

Let ξ_1 be the gain in the first game, ξ_2 be the gain in the second game,...,ξ_{100} be the gain in the 100th game.

$$\xi_i : \begin{pmatrix} 0 & 1 & 9 & 25 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}, i = 1, 2, 3, ..., 100.$$

Solution

For every i=1,2,...,100, ξ_i is random quantity, ξ_i are independent identically distributed r.v.-s.

The question concerns the sum $\sum_{i=1}^{100} \xi_i$.

The question can be answered by the help of the c.d.f. of $\eta = \sum_{i=1}^{100} \xi_i$.

We don't know the c.d.f. $F_\eta(x)$*, but it can be approximated by normal c.d.f. according to the central limit theorem (CLTH).*

Solution

In order to be able to use CLTH we have to know the parameters of the approximate normal distribution.
The expectation of the normal distribution is the expectation of the sum.
The dispersion of the normal distribution is the dispersion of the sum.

$$E(\xi_i) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} = 5.8333 \Longrightarrow$$

$$E(\sum_{i=1}^{100} \xi_i) = 100 \cdot 5.8333 = 583.33$$

$$E(\xi_i^2) = 0 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{6} + 9^2 \cdot \frac{1}{6} + 25^2 \cdot \frac{1}{6} = 117.83$$

$$D^2(\xi_i) = 117.83 - 5.8333^2 = 83.803 \Longrightarrow$$

$$D(\xi_i) = \sqrt{83.803} = 9.1544 \Longrightarrow$$

$$D(\sum_{i=1}^{100} \xi_i) = \sqrt{100} \cdot 9.1544 = 91.544$$

Example

a) Compute the probability that the sum of the gains is at least 550 and less than 600 after 100 games .

Solution

$$P(550\leq \sum_{i=1}^{100} \xi_i < 600) = F_\eta(600) - F_\eta(550) \approx$$
$$\Phi\left(\frac{600-583.33}{91.544}\right) - \Phi\left(\frac{550-583.33}{91.544}\right) = 0.214\,35.$$

b) At least how much is the gain with probability 0.99?

Solution

x=? $P(\sum_{i=1}^{100} \xi_i \geq x) = 0.99, 1 - F_\eta(x) = 0.99, F_\eta(x) = 0.01,$
as $F_\eta(x) \approx \Phi(\frac{x-583.33}{91.544})$,
we solve the equation $\Phi(\frac{x-583.33}{91.544}) = 0.01$.
$\Phi(\frac{x-583.33}{91.544}) = 0.01 \Longrightarrow \frac{x-583.33}{91.544} = -2.32 \Rightarrow x = 370.95, x = 371.$

Example

In a factory the used material during a day is a r.v. with p.d.f.

$$f(x)=\begin{cases} 1-\frac{x}{2}, & \text{if } 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}.$$

If the daily needs are independent, compute the probability that at least 25 units of material is used during 30 days.

Solution

Let ξ_1 be the amount of material used during the first day, ξ_2 be the amount of material used during the second day,...., ξ_{30} be the amount of material used during the 30th day,.
For every i=1,2,...,30 ξ_i is random quantity, ξ_i are independent identically distributed r.v.-s.
The question concerns $\sum\limits_{i=1}^{30} \xi_i$ and it can be answered by the help of the c.d.f. of $\eta = \sum\limits_{i=1}^{30} \xi_i$
We do not know F_η, but it can be approximated by normal c.d.f. according to CLTH.

Solution

We have to know the expectation and the dispersion of the normal distribution. They are the expectation and the dispersion of the sum.

$E(\xi_i) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^2 x \cdot (1 - \frac{x}{2})dx = \frac{2}{3} \Longrightarrow E(\sum_{i=1}^{30} \xi_i) = 30 \cdot \frac{2}{3} = 20$

$E(\xi_i^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^2 x^2 \cdot (1 - \frac{x}{2})dx = 0.66667$

$D^2(\xi_i) = \frac{2}{9} \Longrightarrow D(\xi_i) = \sqrt{\frac{2}{9}} = 0.4714 \Longrightarrow$

$D(\sum_{i=1}^{30} \xi_i) = \sqrt{30} \cdot 0.4714 = 2.5820$

Example

Compute the probability that at least 25 units of material is used during 30 days.

Solution

a) The question is $P(\sum_{i=1}^{30} \xi_i \geq 25) = 1 - F_\eta(25) \approx 1 - \Phi(\frac{25-20}{2.5820}) = 0.264.$

Example

For how many days 50 units amount of material are enough with probability 0.99 ?

Solution

n=? $\Theta = \sum_{i=1}^{n} \xi_i$: $P(\sum_{i=1}^{n} \xi_i \leq 50) = 0.99$; $F_\Theta(50) = 0.99$;

Use normal c.d.f. instead of $F_\Theta(x)$, *the expectation is* $E(\sum_{i=1}^{n} \xi_i) = n \cdot \frac{2}{3}$,

the dispersion is $D(\sum_{i=1}^{n} \xi_i) = \sqrt{n} \cdot 0.4714$.

Now $F_\Theta(x) \approx \Phi(\frac{x-0.6667 \cdot n}{\sqrt{n} \cdot 0.4714})$,

$\Phi(\frac{50-0.6667 \cdot n}{\sqrt{n} \cdot 0.4714}) = 0.99$, $\frac{50-0.6667 \cdot n}{\sqrt{n} \cdot 0.4714} = 2.32$, $n = 62.072$.

For 62 days it is enough 50 units of material with probability 0.99.

Example

For a flight, 500 tickets are sold. If every passenger is at the board with probability 0.9 independently of each other, compute the probability that at least 460 people are at the board.

Solution

Let η_{500} the number of people at he board in case of 500 sold tickets. η_{500} is binomially distributed r.v. with n=500, p=0.9 parameters. The question is: $P(460 \leq \eta_{500})$.

Solution

We have 2 ways:
a) $P(460 \leq \eta_{500}) = \sum_{k=460}^{500} P(\eta_{500} = k) = \sum_{k=460}^{500} \binom{500}{k} \cdot 0.9^k \cdot (1-0.9)^{500-k} = 0.075$

difficulty: computing and summing 41 probabilities.

Solution

b) η_{500} *can be considered as a sum of 500 characteristically distributed r.v.-s. Therefore CLTH can be applied. Expectation of 1 member:* p, *dispersion:* $\sqrt{p(1-p)}$, *therefore the expectation of the sum is* np, *dispersion of the sum is* $\sqrt{np(1-p)}$. *Consequently*
$F_{\eta_{500}}(x) \approx \Phi(\frac{x-500\cdot 0.9}{\sqrt{500\cdot 0.9\cdot 0.1}})$.
Now $P(460 \leq \eta_{500}) = 1-F_{\eta_{500}}(460) \approx 1-\Phi(\frac{460-500\cdot 0.9}{\sqrt{500\cdot 0.9\cdot 0.1}}) = 0,06\,8$.
Note: $E(\eta_{500}) = 500 \cdot 0.9 = 450.0$; $D(\eta_{500}) = \sqrt{500 \cdot 0.9 \cdot 0.1} = 6.708\,2$.

Example

Which is value for which the number of people at the board is less than with probability 0.99 ?

Solution

x=? $P(\eta_{500} < x) = 0.99$
$F\eta_{500}(x) = 0.99$, $\Phi(\frac{x-500\cdot 0.9}{\sqrt{500\cdot 0.9\cdot 0.1}}) = 0.99$,

$\frac{x-500\cdot 0.9}{\sqrt{500\cdot 0.9\cdot 0.1}} = 2.33$,
$x = 465.63$; $x = 466$.

Example

How many tickets could be sold if we want to insure that the number of people at the board is less than 500 with probability 0.999?

Solution

Let n be the number of sold tickets, η_n is the number of people at the board,
η_n is binomially distributed r.v. with n and p=0.9;
n=? for which $P(\eta_n < 500) = 0.999$.
$F\eta_n(500) = 0.999, E(\eta_n) = n \cdot 0.9, D(\eta_n) = \sqrt{n \cdot 0.9 \cdot 0.1}$

$\Phi(\frac{500 - n \cdot 0.9}{\sqrt{n \cdot 0.9 \cdot 0.1}}) = 0.999$, $\frac{500 - n \cdot 0.9}{\sqrt{n \cdot 0.9 \cdot 0.1}} = 3.09$, $n = 531$.

Example

Compute the probability that, in case of 500 sold tickets, there are exactly 450 people at the board.

Solution

a) Exact value $\binom{500}{450} \cdot 0.9^{450} \cdot 0.1^{50} = 0.05\,937\,1.$
b) Approximate value applying c.d.f. by the help of CLTH:
$P(\eta_{500} = 450) = P(450 \leq \eta_{500} < 451) \approx$
$\Phi\left(\frac{451-500\cdot 0.9}{\sqrt{500\cdot 0.9\cdot 0.1}}\right) - \Phi\left(\frac{450-500\cdot 0.9}{\sqrt{500\cdot 0.9\cdot 0.1}}\right) = 0.05\,925\,1.$
Not too much difference.

Solution

c) approximate value applying p.d.f.

Normal p.d.f. with expectation 450, and dispersion 6.708 2

$$f_{\eta_{500}}(x) = \frac{1}{\sqrt{2\pi}\cdot 6.7082} exp(-\frac{(x-450)^2}{2\cdot 6.7082^2})$$

$$f_{\eta_{500}}(450) = \frac{1}{\sqrt{2\pi}\cdot 6.7082} \exp(-\frac{(450-450)^2}{2\cdot 6.7082^2}) = 0.059471$$

$$f_{\eta_{500}}(451) = \frac{1}{\sqrt{2\pi}\cdot 6.7082} \exp(-\frac{(451-450)^2}{2\cdot 6.7082^2}) = 0.58814$$

$$f_{\eta_{500}}(450.5) = \frac{1}{\sqrt{2\pi}\cdot 6.7082} \exp(-\frac{(450.5-450)^2}{2\cdot 6.7082^2}) = 0.059306$$

good coincidences.

Examples

The misprints of a book on a page are Poisson distributed r.v.-s with expectation 0.7. The numbers of misprints on the separate pages are independent r.v.-s. Compute the probability that in a book of 250 pages there are more than 200 misprints.

Solution

η_{250} is the number of misprints on 250 pages. η_{250} is Poisson distributed r.v. with expectation $250 \cdot 0.7 = 175$ *and dispersion* $\sqrt{250 \cdot 0.7} = 13.229$*.*
Therefore $F_{\eta_{250}}(x) \approx \Phi(\frac{x-175}{13.229})$.
$P(\eta_{250} > 200) = P(\eta_{250} \geq 201) = 1 - F_{\eta_{250}}(201) \approx 1 - \Phi(\frac{210-175}{13.229}) =$
$1 - 0.99592 = 0.00408$

Example

Which is the value for which the number of misprints is less with probability 0.9?

Solution

x=? $P(\eta_{250} < x) = 0.9$, $\Phi(\frac{x-175}{13.229}) = 0.9$, $\frac{x-175}{13.229} = 1.28$, $x = 191.93$, $x = 192$.

Example

Compute the probability that there are 170 misprints on 250 pages.

Solution

a) applying c.d.f.:

$P(\eta_{250} = 170) = P(170 \leq \eta_{250} < 171) \approx \Phi(\frac{171-175}{13.229}) - \Phi(\frac{170-175}{13.229}) = 0.0$
$2\,845\,5$

b) applying p.d.f.:

$\varphi(170.5) = \frac{1}{\sqrt{2\pi}\cdot 13.229} \exp(-\frac{(170.5-175)^2}{2\cdot 175} = 0.02\,846\,1.$

Exact value: $P(\eta_{250} = 170) = \frac{175^{170}}{170!} \cdot \exp(-175) = 0.02\,845\,5.$

Multidimensional random variables

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be given, consider the functions $\underline{\xi}$: $\Omega \rightarrow \mathbb{R}^n$.
$\underline{\xi}(\omega) = (\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega))-$ n dimensional point!

Definition

$\underline{\xi} : \Omega \rightarrow \mathbb{R}^n$ is called multidimensional random variable (m.d.r.v.) if for any $(x_1, ..., x_n) \in \mathbb{R}^n$
$\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, ..., \xi_n(\omega) < x_n\} \in \mathcal{A}$

Remark

As $\{\omega : \xi_1(\omega) < x_1 \cap \xi_2(\omega) < x_2 \cap .... \cap \xi_n(\omega) < x_n\}$ *is an event, it has probability!*
$\xi_i : \Omega \rightarrow \mathbb{R}$ *is a r.v. i=1,2,...,n*

Remark

As $\{\xi_1 < x_1\} = \cup_{m=1}^{\infty} \{\xi_1(\omega) < x_1, \xi_2(\omega) < m, \xi_n(\omega) < m\}$

There are 100 products, 20 are of I class, 30 are of II class, 40 are of III class, 10 are of IV class. Choose 6 items without replacement. Let $\underline{\xi}=(\xi_1,\xi_2)$, ξ_1 be the number of the products of I class in the sample, ξ_2 is the number of the products of II class in the sample.
The items are numbered, 1,2,...,20 are of I class, 21,...,50 are of II class,51,...90 are of III class, 91,...,100 are of IV class.

Possible outcomes: the subsets containing 6 elements from $\{1, 2, .., 100\}$.
events: subset of Ω
P: classical probability.

- $\underline{\xi}$: $\Omega \to \mathbb{R}^2$
- $\underline{\xi}(\{5, 21, 31, 41, 51, 100\}) = (1, 3)$
- $\underline{\xi}$ maps the set of the possible outcomes into a subset of $\mathbb{R}^2$.
- Any value of $\underline{\xi}$ is a two dimensional point.
- The value of $\underline{\xi}$ is a random value, as the possible outcome is random.

- The possible values of $\underline{\xi}$ are the pairs (0,0), (0,1),...(0,6),(1,0),(1,1),...,(5,0),(5,1),(6,0).
- The possible values can be listed.
- Probabilities belonging to the possible values are can be computed:

$$P(\underline{\xi} = (i,j)) = \frac{\binom{20}{i}\binom{30}{j}\binom{50}{6-(i+j)}}{\binom{100}{6}},\ 0 \leq i,j \leq 6, integers, i+j \leq 6.$$

- The margins: $\xi_1(\{5, 21, 31, 41, 51, 100\}) = 1, ...,$ $\xi_2(\{5, 21, 31, 41, 51, 100\}) = 3, ...$
- ξ_1: $\Omega \to \mathbb{R}$ r.v., ξ_2: $\Omega \to \mathbb{R}$ r.v.;
- The possible values of ξ_1 are 0,1,2,3,4,5,6 ; probabilities are: $P(\xi_1 = i) = \frac{\binom{20}{i}\binom{80}{6-i}}{\binom{100}{6}}$.
- The possible values of ξ_2 are 0,1,2,3,4,5,6; probabilities are: $P(\xi_2 = j) = \frac{\binom{30}{j}\binom{70}{6-j}}{\binom{100}{6}}$.

Example

Choose two numbers independently from the interval [0,1] by geometric probability.
Let $\underline{\xi}=(\xi_1, \xi_2)$, where ξ_1 is the sum of the numbers, ξ_2 is the difference of the numbers.

- $\Omega = [0,1]\mathrm{x}[0,1]$,
- events: the subsets of Ω which have area,
- P : area
- $\underline{\xi}$: $\Omega \to \mathbb{R}^2$,
- $\underline{\xi}((0.5, 0.5)) = (1, 0)$, $\underline{\xi}((0.9, 0.5)) = (1.4, 0.4)$,....

- The possible values of $\underline{\xi}$:
  $\left\{(y_1, y_2) \in \mathbb{R}^2 : 0 \leq y_1 \leq 2, 0 \leq y_2 \leq 1\right\} \subset [0,2]x[0,1]$
- It is not a finite or a countable infinite set
- $\xi_1((0.5, 0.5)) = 1; \xi_2((0.5, 0.5)) = 0 \Longrightarrow \underline{\xi}((0.5, 0.5)) = (1, 0)$
- P($\underline{\xi}$=(1.5,0.3)=P$((\{(0.9, 0.6), (0.6, 0.9)\}) =$
  $= \frac{T(\{(0.9,0.6),(0.6,0.9)\})}{T(\Omega)} = 0.$
- P($\underline{\xi}$=(y_1, y_2))=0, $\forall (y_1, y_2) \in \mathbb{R}^2$.

Definition

$\underline{\xi}$: $\Omega \to \mathbb{R}^n$ m.d.r.v. is discrete if it has finite or countable infinitely many possible values.

Definition

We give the distribution of a discrete m.d.r.v. if we list its possible values and the probabilities belonging to the possible values.

Definition

The distribution of ξ_i are called marginal distributions.

Remark

Example I) is discrete, Example II) is not.

Following the steps of one dimensional r.v.-s,

- instead of $P(\underline{\xi} = \underline{x})$=$\mathrm{P}\left\{\omega : \underline{\xi}(\omega) = \underline{x})\right\})$, $\underline{x} \in \mathbb{R}^n$ we investigate $\mathrm{P}(\underline{\xi} < \underline{x})$!
- $\underline{x}$ =$(x_1, x_2, ...x_n)$;
  $\left\{\underline{\xi} < \underline{x}\right\}$=$\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, ..., \xi_n(\omega) < x_n\}$;

Remark

As it is required that $\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, ..., \xi_n(\omega) < x_n\} \in \mathcal{A}$, *therefore it has probability.*

Definition

$\underline{\xi}$: $\Omega \to \mathbb{R}^n$ m.d.r.v., the joint c.d.f. of $\underline{\xi}$ is the function F:$\mathbb{R}^n \to \mathbb{R}$ for which

$F(x_1, x_2, ..., x_n) = P(\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, ..., \xi_n(\omega) < x_n\}) = P(\xi_1 < x_1, \xi_2 < x_2, ..., \xi_n < x_n)$

Multidimensional random variables

Properties of the joint cumulative distribution function (j.c.d.f.) i)

1. $0 \leq F(x_1, x_2, ..., x_n) \leq 1$
2. F is monotone increasing in all its variables
3. $\lim\limits_{x_i \to \infty, \forall i=1,2,...n} F(x_1, x_2, ..., x_n) = 1, \lim\limits_{\exists i: x_i \to -\infty} F(x_1, x_2, ..., x_n) = 0$
4. it is continuous from left hand side in all variables

Remark

These properties are not sufficient if $1 < n$.

Multidimensional random variables

Properties of the joint cumulative distribution function ii)

5. if n=2, then for $F(x_1, x_2) = P(\xi_1 < x_1, \xi_2 < x_2)$ the inequality

   $$0 \leq F(x_1 + \Delta x_1, x_2 + \Delta x_2) - F(x_1 + \Delta x_1, x_2) - F(x_1, x_2 + \Delta x_2) + F(x_1, x$$

   is satisfied for all $0 < \Delta x_1, 0 < \Delta x_2$.

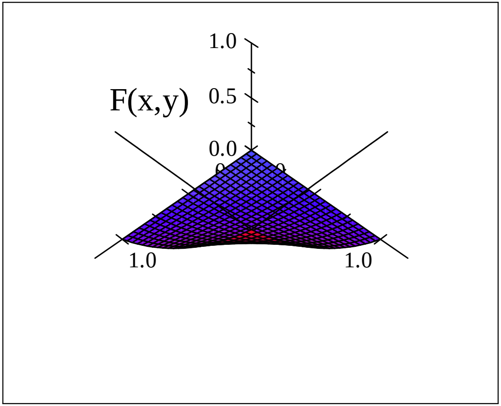
Remark

Properties 1., 2., 3., 4., 5 are sufficient in two dimensions!

Remark

If F is continuous in every point, then $P(\underline{\xi} = \underline{x}) = 0$ for all $\underline{x} \in \mathbb{R}^n$.

$$F(x,y)=\begin{cases} 0,\ \textit{if}\ \text{x<0 or y<0} \\ \frac{x^2y+y^2x}{2},\ \textit{if}\quad 0\le x\le 1, 0\le y\le 1 \\ \frac{y+y^2}{2},\ \textit{if}\ 1<x, 0\le y\le 1 \\ \frac{x+x^2}{2},\ \textit{if}\ 1<y, 0\le x\le 1 \\ 0,\ \textit{if}\ \text{1<x and 1<y} \end{cases}$$


1.0
0.5
0.0
F(x,y)
1.0
1.0

Definition

$\underline{\xi}$: $\Omega \to \mathbb{R}^n$ is called continuous m.d.r.v. if there exists such a function $f : \mathbb{R}^n \to \mathbb{R}$, continuous almost everywhere, for which

$$F(x_1, x_2, ..., x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} f(t_1, ..., t_n) dt_n ... dt_1.$$

The function f is called joint probability density function (j.p.d.f.)

Remark

If f is changed in "some" points, the integral does not change.

Remark

If f is continuous at $\underline{x}$, then F is differentiable at $\underline{x}$.

Remark

If F is differentiable at $\underline{x}$, then

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, x_2, ..., x_n)}{\partial x_1 ... \partial x_n}.$$

Theorem

Let $\underline{\xi}$: $\Omega \to \mathbb{R}^n$ *be continuous m.d.r.v. with p.d.f. f. Then*

$$0 \leq f(x_1, ...x_n), \tag{4}$$

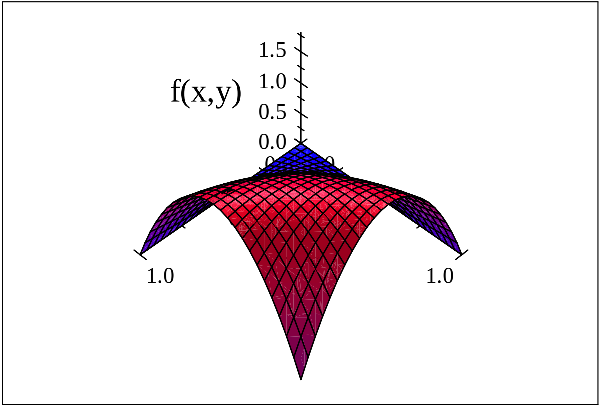
and

$$\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty} f(t_1, ..., t_n)dt_n...dt_1 = 1. \tag{5}$$

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ *is a continuous function except from "some" points and f satisfies properties 4 and 5, then there exists such a m.d.r.v.* $\underline{\xi}$ *whose p.d.f is f.*

$$f(x,y) = \begin{cases} \frac{1}{0.138\,89}(1-xy)\cdot xy, 0 < x < 1, 0 < y < 1 \\ 0, \textit{otherwise} \end{cases}$$



Theorem

$\underline{\xi}$: $\Omega \to \mathbb{R}^n$ is continuous m.d.r.v. with joint p.d.f. f. Then for $A \subset \mathbb{R}^n$

$$P(\underline{\xi} \in A) = \int_A f(t_1, ..., t_n) dt_n ... dt_1.$$

Remark

- *Where the j.p.d.f is large, $\underline{\xi}$ takes values frequently around these values.*
- *Where the j.p.d.f. is small, then $\underline{\xi}$ takes values rarely around these values.*
- *If $f(\underline{x}) = 0$, for every $\underline{x} \in A$, then $P(\underline{\xi} \in A) = 0$.*

Definition

$\underline{\xi}$ and $\underline{\eta}$ are m.d.r.v.-s. They are called identically distributed if their j.c.d.f.-s are equal.

Remark

If $\underline{\xi}$ and $\underline{\eta}$ are discrete, then they have the same possible values and with the same probabilities.

Remark

If $\underline{\xi}$ and $\underline{\eta}$ are continuous m.d.r.v., then their j.p.d.f.-s are the equal.

Example

Let $\underline{\xi}=(\xi_1,\xi_2)$ be a m.d.r.v. with

$$f(x_1,x_2)=\begin{cases} x_1+x_2, & \text{if } 0\le x_1\le 1 \text{ and } 0\le x_2\le 1 \\ 0, & \text{otherwise} \end{cases}.$$

Compute the probability of $\xi_2 < \frac{\xi_1}{2}$.

Solution

$$P(\xi_2 < \tfrac{\xi_1}{2}) = \int_A f(x_1,x_2)dx_2dx_1, \text{ with } A=\left\{(x_1,x_2)\in\mathbb{R}^2 : x_2 < \tfrac{x_1}{2}\right\}$$

$$\int_A f(x_1,x_2)dx_2dx_1 = \int_0^1\int_0^{\frac{x_1}{2}}(x_1+x_2)dx_2dx_1 = \int_0^1\left[x_1x_2+\frac{x_2^2}{2}\right]_0^{\frac{x_1}{2}}dx_1 =$$

$$\int_0^1\left(x_1\frac{x_1}{2}+\frac{x_1^2}{8}\right)dx_1 = \int_0^1\frac{5}{8}x_1^2dx_1 = \frac{5}{8}\left[\frac{x_1^3}{3}\right]_0^1 = 0.208\,33$$

Definition

$\underline{\xi} = (\xi_1, \xi_2, ..., \xi_n)$ *m.d.r.v. The distribution of* ξ_i *is called the ith marginal distribution.*

Definition

The c.d.f. of ξ_i is the ith marginal c.d.f., that is $\mathrm{F}_i(x_i) = P(\xi_i < x_i)$.

Theorem

$$F_i(x_i) = \lim_{\forall x_j \to \infty, j \neq i} F(x_1, x_2, ..., x_i, ... x_n)$$

Remark

If $\underline{\xi}$ *is discrete, then* ξ_i *is also discrete for any i=1,2,...,n.*
If $\underline{\xi}$ *is continuous m.d.r.v., then* ξ_i *is also continuous r.v. for any i=1,2,...,n.*

Theorem

If $\underline{\xi}$ is continuous m.d.r.v., then ξ_i is continuous r.v., i=1,2,...,n, moreover

$$F_i(x_i) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\mathbf{x}_i} .. \int_{-\infty}^{\infty} f(t_1, ..., t_i, ...t_n) dt_n ... dt_i ... dt_1,$$

$$f_i(x_i) = \int_{-\infty}^{\infty} .... \int_{-\infty}^{\infty} f(t_1, ..., x_i, ...t_n) dt_n ... \mathbf{dt}_{i+1} \mathbf{dt}_{i-1} ... dt_1$$

Definition

$\underline{\xi}$ and $\underline{\eta}$ are m.d.r.v.-s. $\underline{\xi}$ and $\underline{\eta}$ are called independent if for any values of 1<r and 0<m, for any indices $(k_1,k_2,...,k_r)$ and $(s_1,s_2,...,s_m)$, with notation $\theta = (\xi_{k_1},...,\xi_{k_r},\eta_{s_1},...,\eta_{s_m})$

$$F_\theta = F_{\xi_{k_1}} \cdot ... \cdot F_{\xi_{k_r}} \cdot F_{\eta_{s_1}} \cdot ... F_{\eta_{s_m}}$$

Remark

In case of independence, the j.c.d.f. is the product of the marginal c.d.f.-s.

Theorem

$\underline{\xi}$ is an n dimensional r.v., $\underline{\eta}$ is an m dimensional r.v.-s. $\underline{\xi}$ and $\underline{\eta}$ are independent if and only if

$$P\left(\left(\underline{\xi} \in A\right) \cap \left(\underline{\eta} \in B\right)\right) = P(\underline{\xi} \in A) \cdot P(\underline{\eta} \in B)$$

$A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$.

Theorem

If ξ and η are continuous r.v.-s, then $(\xi\ ,\eta)$ is continuous m.d.r.v. Let us denote its j.p.d.f. by $f(x_1, x_2)$. ξ and η are independent, if and only if

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

for $(x_1, x_2) \in \mathbb{R}^2$ except from "some" points.

Proof ξ and η are independent, if and only if $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$.
If F is differentiable at (x_1, x_2),then
$f(x_1, x_2) = \frac{\partial^2 F(x_1,x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 F_1(x_1) F_2(x_2)}{\partial x_1 \partial x_2} = \frac{\partial F_1(x_1)}{\partial x_1} \cdot \frac{\partial F_2(x_2)}{\partial x_2} = f_1(x_1) \cdot f_2(x_2)$ ■

Theorem

If $\underline{\xi}$ and $\underline{\eta}$ are independent, g and h are such functions for which $g(\underline{\xi})$ and $h(\underline{\eta})$ exist, then $g(\underline{\xi})$ and $h(\underline{\eta})$ are also independent.

Theorem

If $\underline{\xi}=c$, then $\underline{\xi}$ is independent from any m.d.r.v..

Theorem

Let ξ, η *be r.v.-s,the j.p.d.f. of* (ξ, η) *is* $f(x_1, x_2)$. *Then*

$$E(\xi \cdot \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 \cdot x_2 \cdot f(x_1, x_2) dx_2 dx_1,$$

supposing $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x_1 \cdot x_2| \cdot f(x_1, x_2) dx_2 dx_1 < \infty$.

Theorem

Let ξ, η *be discrete r.v.-s with possible values* (x_i, y_j), *i=1,2,...,j=1,2,....* *Then*

$$E(\xi \cdot \eta) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i y_j P(\xi = x_i, \eta = y_j),$$

supposing $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |x_i \cdot y_j| \cdot P(\xi = x_i, \eta = y_j) < \infty.$

Example

Let the j.p.d.f. of (ξ_1, ξ_2) be

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 \text{ if } 0 \le x_1 \le 1, 0 \le x_2 \le 1 \\ 0 \text{ otherwise} \end{cases}.$$

Compute $cov(\xi, \eta)$.

Solution

We need $E(\xi_1 \cdot \xi_2)$, $E(\xi_1)$ and $E(\xi_2)$.

$E(\xi_1 \cdot \xi_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 \cdot x_2 \cdot f(x_1, x_2) dx_2 dx_1 =$

$\int_0^1 \int_0^1 x_1 \cdot x_2 \cdot (x_1 + x_2)\, dx_2 dx_1 =$

$\int_0^1 \int_0^1 \left(x_1^2 \cdot x_2 + x_2^2 x_1\right) dx_2 dx_1 = \int_0^1 \left[\frac{x_1^2 \cdot x_2^2}{2} + \frac{x_2^3 x_1}{3}\right]_0^1 dx_1 = \int_0^1 \frac{x_1^2}{2} + \frac{x_1}{3} dx_1 =$

$\left[\frac{x_3^2}{6} + \frac{x_1^2}{6}\right]_0^1 = \frac{1}{3}.$

Solution

$E(\xi_1) = \int_{-\infty}^{\infty} x_1 f_1(x_1) dx_1,$

$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 = \int_0^1 f(x_1, x_2) dx_2 =$

$\begin{cases} \int_0^1 x_1 + x_2 dx_2 = x_1 + \frac{1}{2}, \textit{ if } 0 \leq x_1 \leq 1 \\ 0, \ \textit{otherwise} \end{cases}$

$E(\xi_1) = \int_{-\infty}^{\infty} x_1 \cdot f_1(x_1) dx_1 = \int_0^1 x_1(x_1 + 0.5) dx_1 = \left[\frac{x_1^3}{3} + \frac{x_1^2}{4}\right]_0^1 = \frac{7}{12}.$

Similarly, $E(\xi_2) = \frac{7}{12}.$

$cov(\xi, \eta) = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = -\frac{1}{144}.$

Definition

Let $\underline{\xi}$ be an n dimensional r.v., then $E(\underline{\xi}) = (E(\xi_1), E(\xi_2), ..., E(\xi_n))$.

Definition

Let $\underline{\xi}$ be an m dimensional r.v., $\underline{\eta}$ n dimensional r.v., then $cov(\underline{\xi}, \underline{\eta}) = \mathbb{C} \in \mathbb{R}^{nxm}$ is a matrix, which has the elements $c_{i,j} = cov(\xi_i, \eta_j)$.

Definition

(auto-covariance matrix) Let $\underline{\xi}$ be a m.v.r.v.. Then

$$cov(\underline{\xi}, \underline{\xi}) = \mathbb{C} = (c_{i,j})_{i=1,\ldots,n,j=1,\ldots,n}. \in \mathbb{R}^{nxn}$$

with the elements

$$c_{i,j} = cov(\xi_i, \xi_j).$$

- $c_{i,j} = c_{j,i}$
- $c_{i,i} = D^2(\xi_i)$
- positive definite
- *If* $\underline{\eta} = A\underline{\xi} + \underline{c}$, then

$$cov(\underline{\eta}, \underline{\eta}) = A \cdot cov(\underline{\xi}, \underline{\xi})A^T.(A \in \mathbb{R}^{mxn}, \underline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ . \\ \xi_n \end{pmatrix})$$

Definition

Let ξ_i i=1,2,...,k be independent r.v.-s with standard normal distributions.

Let $\underline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ . \\ . \\ \xi_k \end{pmatrix}$, $A \in \mathbb{R}^{nxk}$, $\underline{m} \in \mathbb{R}^n$.

Consider

$$\underline{\eta} = A\underline{\xi} + \underline{m}.$$

$\underline{\eta}$ *is called n dimensional r.v. with normal distribution.*

The above definition is the generalization of one dimensional normal distribution (recall $\eta = \sigma\xi + m$).

Theorem

If $\underline{\eta}$ is m.d.r.v. with normal distribution, then $B \cdot \underline{\eta} + \underline{b}$ is also m.d.r.v. with normal distribution.

Proof $B \cdot \underline{\eta} + \underline{b} = B \cdot \left(A\underline{\xi} + \underline{m}\right) + \underline{b} = BA \cdot \underline{\xi} + B\underline{m} + \underline{b}$ ■

Theorem

If $\underline{\eta}$ is m.d.r.v. with normal distribution, then every coordinate η_i is one dimensional r.v. with normal distribution.

Proof $B_1 = \begin{pmatrix} 1 & 0 & . & . & 0 \end{pmatrix}$, $B \cdot \underline{\eta} = \eta_1$,
$B_i = \begin{pmatrix} 0 & . & 1 & . & 0 \end{pmatrix}$, .$B_i \underline{\eta} = \eta_i$ ∎

Theorem

If $\underline{\eta}$ is a m.d.r.v. with normal distribution then the sum, the difference, the linear transformations are normally distributed r.v.

Proof $\eta_1 + \eta_2 = \begin{pmatrix} 1 & 1 & . & . & 0 \end{pmatrix} \underline{\eta}$,
$\eta_1 - \eta_2 = \begin{pmatrix} 1 & -1 & . & . & 0 \end{pmatrix} \underline{\eta}$. ■

Theorem

If η_1, η_2 are independent normally distributed r.v.-s, then $\eta_1 + \eta_2$ is also normally distributed.

Proof $\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$,
where $\xi_1 \tilde{} N(0,1), \xi_2 \tilde{} N(0,1)$, ξ_1, ξ_2 are independent.
Apply the above theorem. ■

Theorem

$$E(\underline{\eta}) = \underline{m}$$

$$cov(\underline{\eta}, \underline{\eta}) = A \cdot A^T$$

Proof $E(\underline{\eta}) = E(A\underline{\xi} + \underline{m}) = A \cdot E(\underline{\xi}) + \underline{m} =$
$A \cdot \begin{pmatrix} 0, & 0, & . & . & . & ,0 \end{pmatrix}^T + \underline{m} = \underline{m},$
$cov(\underline{\eta}, \underline{\eta}) = cov(A \cdot \underline{\xi}, A \cdot \underline{\xi}) = A \cdot cov(\underline{\xi}, \underline{\xi}) \cdot A^T =$
$= A \cdot I \cdot A^T = A \cdot A^T$. ■

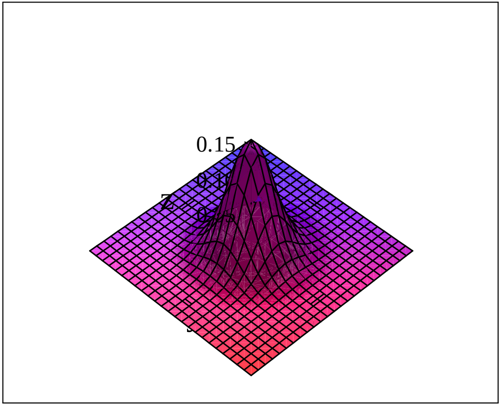
Theorem

If $\underline{\eta}=A\underline{\xi}+m$ *m.d.r.v. with normal distribution, moreover* $A \cdot A^T$ *is invertible, then the j.p.d.f of* $\underline{\eta}$ *is*

$$f_{\underline{\eta}}(x_1, x_2, ..., x_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(A \cdot A^T)}} \cdot$$

$$\cdot \exp(-\frac{1}{2}(x_1 - m_1, ..., x_n - m_n)\left(A \cdot A^T\right)^{-1} \begin{pmatrix} x_1 - m_1 \\ . \\ . \\ . \\ x_n - m_n \end{pmatrix}$$

$$\underline{m} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; cov(\underline{\eta}, \underline{\eta}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; f_{\underline{\eta}}(x_1, x_2) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

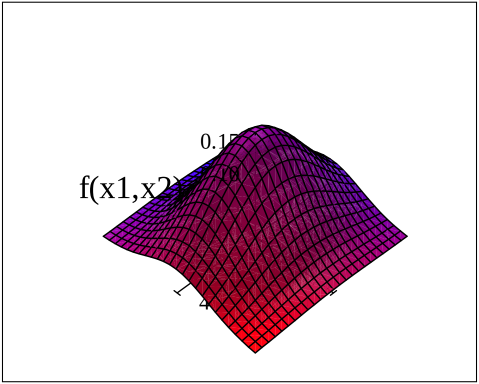


Multidimensional normal distribution

J.p.d.f. - example ii)

$$\underline{m} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$cov(\underline{\eta}, \underline{\eta}) = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}, f_{\underline{\eta}}(x_1, x_2) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{2 \cdot 3} \exp\left(-\frac{\frac{x_1^2}{9} + \frac{x_2^2}{4}}{2}\right)$$

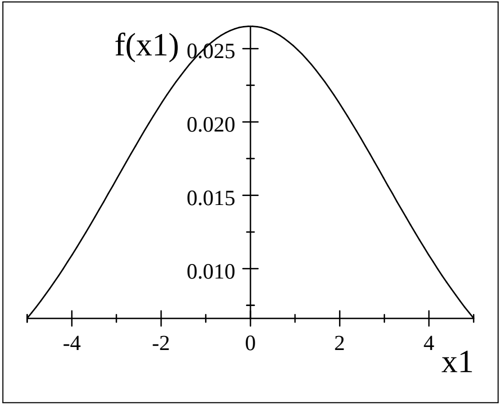


Multidimensional normal distribution

J.p.d.f. - example iii)

$$\underline{m} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ cov(\underline{\eta}, \underline{\eta}) = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix},\ x_2 = 0,$$

$$f_{\underline{\eta}}(x_1, x_2) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{2 \cdot 3} \exp\left(-\frac{\frac{x_1^2}{9} + 0}{2}\right)$$

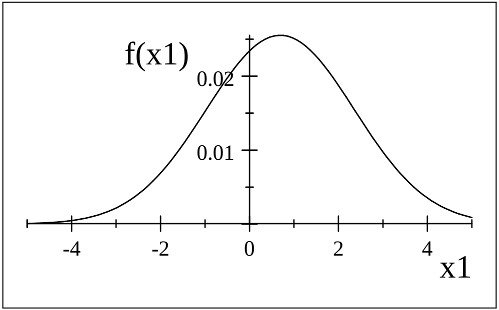


Multidimensional normal distribution

J.p.d.f. - example iv)

$$\underline{m}=\begin{pmatrix} 0 \\ 0 \end{pmatrix},\ cov(\underline{\eta},\underline{\eta}) = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix},$$

$$x_1 + x_2 = 1 : f_{\underline{\eta}}(x_1, x_2) = f_{\underline{\eta}}(x_1) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{2\cdot 3} \exp\left(-\frac{\frac{x_1^2}{9} + \frac{(1-x_1)^2}{4}}{2}\right)$$



Example

Let $\underline{\xi}=\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ be a two dimensional r.v. with normal distribution,
$\mathrm{E}(\xi_1) = 0$, $\mathrm{E}(\xi_2) = 10$, $\mathrm{D}(\xi_1) = 2$, $\mathrm{D}(\xi_2) = 3$, $\mathrm{r}(\xi_1, \xi_2) = 0.5$.
Which is more probable: $\underline{\xi}$ is around (-1,8) or (2,11)?

Solution

$$cov(\underline{\xi},\underline{\xi})=\begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix} = AA^T,$$

$$\left(AA^T\right)^{-1} = \frac{1}{\det AA^T}\begin{pmatrix} 9 & -3 \\ -3 & 4 \end{pmatrix} = \frac{1}{27}\begin{pmatrix} 9 & -3 \\ -3 & 4 \end{pmatrix},$$

$$(check: \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix} \cdot \frac{1}{27}\begin{pmatrix} 9 & -3 \\ -3 & 4 \end{pmatrix} = \begin{pmatrix} 1.0 & 0 \\ 0 & 1.0 \end{pmatrix})$$

Solution

$$f_{\underline{\xi}}(x_1, x_2) = \frac{1}{\left(\sqrt{2\pi}\right)^2\sqrt{27}} \cdot \exp(-\frac{1}{2}\left(x_1 - 0, x_2 - 10\right)\frac{1}{27} \cdot \left(\begin{array}{cc} 9 & -3 \\ -3 & 4 \end{array}\right)\left(\begin{array}{c} x_1 - 0 \\ x_2 - 10 \end{array}\right))$$

$$f_{\underline{\xi}}(-1, 8) = \frac{1}{\left(\sqrt{2\pi}\right)^2\sqrt{27}} \cdot \exp(-\frac{1}{2}\left(-1 - 0, 8 - 10\right) \cdot \frac{1}{27} \cdot \left(\begin{array}{cc} 9 & -3 \\ -3 & 4 \end{array}\right)\left(\begin{array}{c} -1 - 0 \\ 8 - 10 \end{array}\right))$$

Solution

the exponent:

$$-\frac{1}{2}\left(-1-0, 8-10\right)\frac{1}{27}\begin{pmatrix} 9 & -3 \\ -3 & 4 \end{pmatrix}\begin{pmatrix} -1-0 \\ 8-10 \end{pmatrix} = -0.240\,74$$

$$f_{\underline{\xi}}(-1, 8) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{27}} \cdot \exp(-0.24074) = 0.02\,407\,6.$$

Solution

$$f_{\underline{\xi}}(x_1, x_2) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{27}} \cdot \exp(-\frac{1}{2}(x_1 - 0, x_2 - 10) \cdot \frac{1}{27} \cdot \begin{pmatrix} 9 & -3 \\ -3 & 4 \end{pmatrix} \cdot \begin{pmatrix} x_1 - 0 \\ x_2 - 10 \end{pmatrix})$$

$$f_{\underline{\xi}}(2, 11) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{27}} \cdot \exp(-\frac{1}{2}(2 - 0, 11 - 10) \cdot \frac{1}{27} \cdot \begin{pmatrix} 9 & -3 \\ -3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 2 - 0 \\ 11 - 10 \end{pmatrix})$$

Solution

exponent:

$$-\frac{1}{2}\left(2-0, 11-10\right)\frac{1}{27}\begin{pmatrix} 9 & -3 \\ -3 & 4 \end{pmatrix}\begin{pmatrix} 2-0 \\ 11-10 \end{pmatrix} = -0.518\,52$$

$$f_{\underline{\xi}}(2,11) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{27}} \cdot \exp(-0.518\,52) = 0.01\,823\,7.$$

$f_{\underline{\xi}}(2,11) < f_{\underline{\xi}}(-1,8) \Longrightarrow$ *around (-1,8) are more probable the values.*

Theorem

If $\underline{\xi}=\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ *is a two dimensional r.v. with normal distribution, moreover* $cov(\xi_1, \xi_2) = 0$*, then* ξ_1 *and* ξ_2 *are independent.*

Example

Let $\underline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ be a two dimensional r.v. with normal distribution, its expectation is $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$, the dispersions are 3 and 2, respectively, and $\operatorname{cov}(\xi_1, \xi_2) = 4$. Compute the probability of $\xi_1 + \xi_2 < 5$.

Solution

$\xi_1 + \xi_2 \,\tilde{}\, N(0+2, ?)$
$D^2(\xi_1 + \xi_2) = cov(\xi_1 + \xi_2, \xi_1 + \xi_2) =$
$\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 9 & 4 \\ 4 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 21.0,$
therefore $\xi_1 + \xi_2 \,\tilde{}\, N(2, \sqrt{21})$.
$P(\xi_1 + \xi_2 < 5) = \Phi(\frac{5-2}{\sqrt{21}}) = \Phi(\ 0.654\,65) = 0.743\,65$

Example

Let $\underline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ be a two dimensional r.v. with normal distribution, its expectation is $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$,the dispersions are 3 and 2, respectively, and $\mathrm{cov}(\xi_1, \xi_2) = 4$, then compute the probability of $\xi_1 < \xi_2$.

Solution

$P(\xi_1 < \xi_2) = P(\xi_1 - \xi_2 < 0)$,
$\xi_1 - \xi_2 \tilde{\ } N(0-2, ?)$,
$D^2(\xi_1 - \xi_2) = cov(\xi_1 - \xi_2, \xi_1 - \xi_2) =$
$\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 9 & 4 \\ 4 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 5$,
$\xi_1 - \xi_2 \tilde{\ } N(-2, \sqrt{5})$,
$P(\xi_1 - \xi_2 < 0) = \Phi(\frac{0-(-2)}{\sqrt{5}}) = \Phi(0.894\,43) = 0.814\,45$.

Definition

Let A be an event, suppose $0 < P(A)$. The conditional c.d.f. of the m.d.r.v. $\underline{\xi}$ given A is the function

$$F_{\underline{\xi}}(\underline{x}|A) : \mathbb{R}^n \to \mathbb{R} : F_{\underline{\xi}}(\underline{x}|A) := P(\underline{\xi} < \underline{x}|A), \underline{x} \in \mathbb{R}^n.$$

Remark

As $0 < P(A)$*, therefore* $P(\underline{\xi} < \underline{x}|A) = \frac{P((\xi < x) \cap A)}{P(A)}$ *is well defined.*

Remark

As the conditional probability satisfies the axioms and all properties of the probability, the conditional c.d.f. satisfies the properties of the c.d.f.-s.

Definition

Let A be an event for which 0<P(A), and suppose that the conditional distribution of $\underline{\xi}$ is continuous. Then the derivative of the conditional c.d.f. of $\underline{\xi}$ is called conditional density function of $\underline{\xi}$ and it is denoted by $f_{\underline{\xi}}(\underline{x}|A)$.

Remark

The conditional density function satisfies the properties (4) and (5) of the joint probability density functions.

Example

Roll twice a fair dice. Let ξ be the sum of the results, let A be the event that the difference between the results equals 1. Determine the conditional distribution of ξ given A.

Conditional distribution

Example I ii)

Solution

If A holds, then the possible values of ξ are 3,5,7,9,11,

$$P(\xi = 3|A) = \frac{P(\xi=3 \cap A)}{P(A)} = \frac{P(\{(1,2),(2,1)\}}{P(A)} = \frac{\frac{2}{36}}{\frac{10}{36}} = 0.2;$$

$$P(\xi = 5|A) = \frac{P(\xi=5 \cap A)}{P(A)} = \frac{P(\{(3,2),(2,3)\}}{P(A)} = \frac{\frac{2}{36}}{\frac{10}{36}} = 0.2;$$

$$P(\xi = 7|A) = \frac{P(\xi=7 \cap A)}{P(A)} = \frac{P(\{(3,4),(4,3)\}}{P(A)} = \frac{\frac{2}{36}}{\frac{10}{36}} = 0.2;$$

$$P(\xi = 9|A) = \frac{P(\xi=9 \cap A)}{P(A)} = \frac{P(\{(4,5),(5,4)\}}{P(A)} = \frac{\frac{2}{36}}{\frac{10}{36}} = 0.2;$$

$$P(\xi = 11|A) = \frac{P(\xi=9 \cap A)}{P(A)} = \frac{P(\{(4,5),(5,4)\}}{P(A)} = \frac{\frac{2}{36}}{\frac{10}{36}} = 0.2.$$

Statement

Let ξ *and* η *be discrete r.v.-s* ξ*:* $\begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ p_1 & p_2 & . & . & . & p_n \end{pmatrix}$,
$\eta : \begin{pmatrix} y_1 & y_2 & . & . & . & y_m \\ q_1 & q_2 & . & . & . & q_m \end{pmatrix}$, $0 < q_j, P(\xi = x_i, \eta = y_j) = p_{i,j}$. *Then the possible vales of* ξ *given* $\eta = y_j$ *form a subset of* $\{x_1, x_2, ..., x_n\}$ *and the probabilities belonging to them are*

$$P(\xi = x_i | \eta = y_j) = \frac{P(\xi = x_i \cap \eta = y_j)}{P(\eta = y_j)} = \frac{p_{i,j}}{q_j}.$$

Remark

The conditional distribution can be expressed by the joint distribution and the marginal distributions.

Example

Let the lifetime of a spare part (ξ) be exponentially distributed r.v. with expectation 10 hours. Determine the conditional c.d.f. and the conditional p.d.f. of ξ given A=$\{5 < \xi\}$.

Solution

The conditional c.d.f. is

$$F_{\xi}(x|A) = P(\xi < x|A) = \frac{P((\xi<x)\cap A)}{P(A)} = \frac{P((\xi<x)\cap(5<\xi))}{P(5<\xi)} = \frac{P(5<\xi<x)}{P(5<\xi)} =$$

$$= \begin{cases} \frac{F(x)-F(5)}{1-F(5)}, & \text{if } 5<x \\ 0, & \text{if } x \leq 5 \end{cases} =$$

$$= \begin{cases} \frac{\exp(-0.5)-\exp(-0.1x)}{\exp(-0.5)}, & \text{if } 5<x \\ 0, & \text{if } x \leq 5 \end{cases} =$$

$$\begin{cases} 1-\exp(-0.1x+0.5), & \text{if } x<5 \\ 0, & \text{if } x \leq 5 \end{cases} .$$

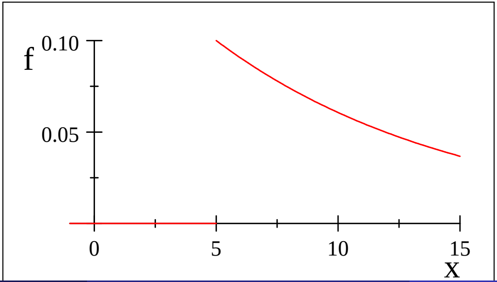
Conditional distribution

Example II iii)

Solution

The conditional p.d.f. is

$$f_\xi(x|A) = \left(F_\xi(x|A)\right)' = \begin{cases} 0.1 \cdot \exp(-0.1x + 0.5), & \text{if } 5<x \\ 0, & \text{if } x \leq 5 \end{cases}$$



Definition

Let ξ and η be continuous r.v.-s, the joint distribution is continuous, j.p.d.f. is f(x,y). The conditional p.d.f. of ξ given $\eta = y$ is defined as

$$f_\xi(x|\eta = y) = \begin{cases} \frac{f(x,y)}{f_\eta(y)}, \ \mathit{if} \ 0 < f_\eta(y) \\ 0, \ \mathit{if} \ f_\eta(y) = 0. \end{cases}.$$

Statement

If $0 < f_\eta(y)$, then $f_\xi(x|\eta = y)$ a is density function.

Proof $0 \le f_\xi(x|\eta = y)$, as the nominator and denominator are nonnegative.
If $0 < f_\eta(y)$, then $\int_{-\infty}^{\infty} f_\xi(x|\eta = y)dx = \int_{-\infty}^{\infty} \frac{f(x,y)}{f_\eta(y)}dx =$
$= \frac{1}{f_\eta(y)} \int_{-\infty}^{\infty} f(x, y)dx = \frac{f_\eta(y)}{f_\eta(y)} = 1.$ ∎

Theorem

Let (ξ,η) be a m.d.r.v. with two dimensional normal distribution, $m_1 = E(\xi)$, $m_2 = E(\eta)$, $\sigma_1 = D(\xi)$, $\sigma_2 = D(\eta)$, $r = r(\xi,\eta)$. Then the conditional distribution of ξ given $\eta = y$ is also normal distribution with expectation

$$m_1 + r \cdot \frac{\sigma_1}{\sigma_2}(y - m_2)$$

and dispersion

$$\sigma_1 \cdot \sqrt{1 - r^2}.$$

Example

The height and the weight of a man is a two dimensional normally distributed r.v. with expectations 175 cm and 75 kg, with dispersions 10 cm and 15kg, the correlation coefficient is 0.6. Given that the height is 180 cm, compute the probability that the weight is more than 100 kg.

Solution

Let ξ denote the weight, η the height.
The conditional distribution of ξ given $\eta = 180$ is normal distribution with expectation

$$m_1 + r\frac{\sigma_1}{\sigma_2}(y - m_2) = 75 + 0.6 \cdot \frac{15}{10}(180 - 175) = 79.5,$$

and with dispersion

$$\sigma = \sigma_1 \cdot \sqrt{1 - r^2} = 15 \cdot \sqrt{1 - 0.6^2} = 12.0.$$

$P(\xi > 100|\eta = 180) = 1 - \Phi\left(\frac{100-79.5}{12}\right) = 0.043\,79.$

Conditional distribution

Multidimensional normal distribution - example II i)

Example

Given that the height is 160 cm, compute the probability that the weight is more than 100 kg.

Solution

The conditional distribution of ξ given $\eta = 160$ is normal distribution with expectation

$$m_1 + r\frac{\sigma_1}{\sigma_2}(y - m_2) = 75 + 0.6 \cdot \frac{15}{10}(160 - 175) = 61.5,$$

and with dispersion

$$\sigma = \sigma_1 \cdot \sqrt{1 - r^2} = 15 \cdot \sqrt{1 - 0.6^2} = 12.0.$$

$P(\xi > 100 | \eta = 180) = 1 - \Phi\left(\frac{100-61.5}{12}\right) = 0.066753.$

Definition

Let ξ and η be discrete r.v.-s with ξ: $\begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ p_1 & p_2 & . & . & . & p_n \end{pmatrix}$,
$\eta : \begin{pmatrix} y_1 & y_2 & . & . & . & y_m \\ q_1 & q_2 & . & . & . & q_m \end{pmatrix}$, $0 < q_j$, $P(\xi = x_i, \eta = y_j) = p_j$. Then the conditional expectation of ξ given $\eta = y_j$ is defined as the expectation defined by the conditional distribution of ξ given $\eta = y_j$, namely

$$E(\xi|\eta = y_j) = \sum_i x_i \cdot P(\xi = x_i|\eta = y_j).$$

Remark

$E(\xi|\eta = y_j)$ is the function of the condition.

Definition

Let (ξ, η) be a two dimensional continuous random variable. Then the conditional expectation of ξ given $\eta = y$ is defined as the expectation computed by the conditional p.d.f. of ξ given $\eta = y$, namely

$$E(\xi|\eta = y) = \begin{cases} \int_{-\infty}^{\infty} x \cdot f_\xi(x|\eta = y)dx, \ \textit{if} \ \ f_\eta(y) > 0 \\ 0, \textit{if} \ \ \ f_\eta(y) = 0 \end{cases}$$

Remark

$E(\xi|\eta = y)$ is the function of the condition y.

Theorem

Let ξ and η be discrete r.v.-s, possible values of η are y_j j=1,2,..,m. Then

$$E(\xi) = \sum_{j=1}^{m} E(\xi|\eta = y_j) \cdot P(\eta = y_j).$$

Theorem

Let ξ and η be continuous r.v.-s. Then

$$E(\xi) = \int_{-\infty}^{\infty} E(\xi|\eta = y) \cdot f_\eta(y) dy.$$

Theorem

Let ξ and η be r.v.-s with finite dispersion, and let $H^*(y) = E(\xi|\eta = y)$. We want to find the function $H(y)$ for which $E((H(\eta))^2)$ is finite, moreover for which $E((\xi - H(\eta))^2)$ is minimal, then the $H(y) = H^*(y)$. That is

$$\min_{H:E(H(\eta)^2)<\infty} E((\xi - H(\eta))^2) = E((\xi - H^*(\eta))^2)$$

where $H^*(\eta) = E(\xi|\eta)$ denotes the random variable which takes value $E\acute{o}(\xi|\eta = y)$ if $\eta = y$.

Remark

If we want to approximate the r.v. ξ by a function of η then the "best" approximation is given by the function computed by the conditional expectation.

Statement

Let (ξ,η) be a two dimensional normally distributed r.v.. Then

$$H^*(\eta) = E(\xi|\eta) = a\eta + b$$

and

$$E(\eta|\xi) = E(\eta|\xi) = c\xi + d$$

Example

The income and the outcome of a factory (ξ, η) is a two dimensional r.v. with j.p.d.f.

$$f(x,y) = \begin{cases} x+y, \ \ \textit{if} \ 0 \le x \le 1, 0 \le y \le 1 \\ 0 \end{cases}.$$

Determine the conditional p.d.f. of the outcome (η) given the income equals 0.3.

Solution

$$\textit{Recall } f_\xi(x) = \int_{-\infty}^{\infty} f(x,y)dy = \begin{cases} x+0.5, \textit{ if } 0<x<1 \\ 0, \textit{ otherwise} \end{cases}$$

$$f_\eta(y|\xi = x) = \begin{cases} \frac{f(x,y)}{f_\xi(x)}, \textit{ if } 0<y<1 \\ 0, \textit{ otherwise} \end{cases} =$$

$$\begin{cases} \frac{f(x,y)}{x+0.5}, \textit{ if } 0<y<1 \\ 0, \textit{ otherwise} \end{cases} \quad (0 < x < 1)$$

$$f_\eta(y|\xi = 0.3) = \frac{f(0.3,y)}{f_\xi(0.3)} = \frac{y+0.3}{0.3+0.5}, \textit{ if } 0<y<1.$$

Example

Given that the income is 0.3, compute the probability that the outcome is between 0.5 and 0.8 ?

Solution

$$P(0.5<\eta<0.8|\xi = 0.3) = \int_{0.5}^{0.8} f_\eta(y|\xi = 0.3)\,dy \int_{0.5}^{0.8} \frac{y+0.3}{0.8}\,dy = 0.356\,25$$

Example

Given that the income is 0.3, compute the expectation of the outcome.

Solution

$$E(\eta|\xi = 0.3) = \int_{-\infty}^{\infty} y \cdot f_\eta(y|\xi = 0.3) = \int_0^1 y \cdot \frac{y+0.3}{0.8} dy = 0.604\,17.$$

Example

Given that the income equals x, 0<x<1, compute the expectation of the outcome.

Solution

$$E(\eta|\xi = x) = \int_0^1 y \cdot \frac{x+y}{x+0.5} dy =$$
$$\frac{x}{x+0.5} \cdot \left[\frac{y^2}{2}\right]_0^1 + \frac{1}{x+0.5} \cdot \left[\frac{y^3}{3}\right]_0^1 =$$
$$0.5 \cdot \frac{x}{x+0.5} + \frac{1}{3} \cdot \frac{1}{x+0.5}. \ (0 < x < 1).$$

Example

Check the property $E(E(\eta|\xi)) = E(\eta)$.

Solution

$$E(\eta) = \int\limits_{-\infty}^{\infty} y \cdot f_\eta(y)dy = \int\limits_0^1 y \cdot (y + 0.5)dy = 0.583\,33$$

$$E(E(\eta|\xi)) = \int\limits_0^1 E(\eta|\xi = x) \cdot f_\xi(x) =$$

$$\int\limits_0^1 (0.5 \cdot \tfrac{x}{x+0.5} + \tfrac{1}{3} \cdot \tfrac{1}{x+0.5}) \cdot (x + 0.5)dx =$$

$$\int\limits_0^1 (0.5 \cdot x + \tfrac{1}{3})dx = 0.583\,33.$$

Example

Which function of the income should be applied for approximating the outcome, if we want to minimize the expectation of the squared difference between the outcome and its approximate value?

Solution

As

$$E(\eta|\xi = x) = \frac{x}{x+0.5} \cdot \frac{1}{2} + \frac{1}{x+0.5} \cdot \frac{1}{3},$$

therefore we should apply

$$E(\eta|\xi) = \frac{\xi}{\xi+0.5} \cdot \frac{1}{2} + \frac{1}{\xi+0.5} \cdot \frac{1}{3}.$$

Example

Compute the expectation of the squared difference between the outcome and its approximation.

Solution

$$E((\eta - E(\eta|\xi))^2) =$$
$$E\left(\left(\eta - \left(\frac{\xi}{\xi+0.5} \cdot \frac{1}{2} + \frac{1}{\xi+0.5} \cdot \frac{1}{3}\right)\right)^2\right) =$$
$$E(\eta^2) - 2 \cdot E\left(\eta \cdot \left(\frac{\xi}{\xi+0.5} \cdot \frac{1}{2} + \frac{1}{\xi+0.5} \cdot \frac{1}{3}\right)\right) +$$
$$E\left(\left(\frac{\xi}{\xi+0.5} \cdot \frac{1}{2} + \frac{1}{\xi+0.5} \cdot \frac{1}{3}\right)^2\right)$$

Solution

$$E(\eta^2) = \int_0^1 y^2 (y+0.5)\, dy = 0.416\,67$$

$$E\left(\eta \cdot \left(\frac{\xi}{\xi+0.5} \cdot \frac{1}{2} + \frac{1}{\xi+0.5} \cdot \frac{1}{3}\right)\right) =$$

$$\int_0^1 \int_0^1 y \cdot \left(\frac{x}{x+0.5} \cdot \frac{1}{2} + \frac{1}{x+0.5} \cdot \frac{1}{3}\right) \cdot (x+y)\, dy dx = 0.340\,96.$$

Solution

$E\left(\left(\frac{\xi}{\xi+0.5}\cdot\frac{1}{2}+\frac{1}{\xi+0.5}\cdot\frac{1}{3}\right)^2\right)=$
$\int_0^1\left(\frac{x}{x+0.5}\cdot\frac{1}{2}+\frac{1}{x+0.5}\cdot\frac{1}{3}\right)^2(x+0.5)dx=0.340\,96.$
$E((\eta-E(\eta|\xi))^2)=0.416\,67-2\cdot 0.340\,96+0.340\,96=0.075\,71.$

Let ξ and η be r.v.-s. We would like to find the **linear** function of ξ which approximates η "as well as possible".
More exactly: determine the values a and b for which $E((\eta - (a\xi + b))^2)$ is minimal.
$\eta_1 = a\xi + b$; we are seaking the solution of $\underset{a,b}{min}(E((\eta - \eta_1)^2$.
η : dependent variable,
ξ : independent variable - explanatory variable
The minimizer will be denoted by $\widehat{\eta}$ (approximation of η).
Simple linear regression: one dependent variable is approximated by a linear function of one independent variable.

Theorem

If $D^2(\xi)$ and $D^2(\eta)$ are finite and $D^2(\xi) \neq 0 \neq D^2(\eta)$, then the solution of the above minimum problem is

$$a = \frac{cov(\xi, \eta)}{D^2(\xi)}, b = E(\eta) - aE(\xi)$$

therefore the best approximation is

$$\widehat{\eta} = \frac{cov(\xi, \eta)}{D^2(\xi)}(\xi - E(\xi)) + E(\eta)$$

The error of the approximation is $E((\eta - E(\widehat{\eta}))^2) = (1 - r^2)D^2(\eta)$.

Remark

$$\frac{\widehat{\eta} - E(\eta)}{D(\eta)} = r\frac{\xi - E(\xi)}{D(\xi)}$$

.

Statement

If there is a linear relationship between ξ and η,then $\widehat{\eta} = \eta$, and

$$E((\widehat{\eta} - \eta)^2) = D^2(\eta) \cdot (1 - 1^2) = 0.$$

Statement

If ξ and η are independent, then

$$\widehat{\eta} = E(\eta)$$

and $E((\widehat{\eta} - \eta)^2) = D^2(\eta) = D^2(\eta)(1 - 0^2)$.

Statement

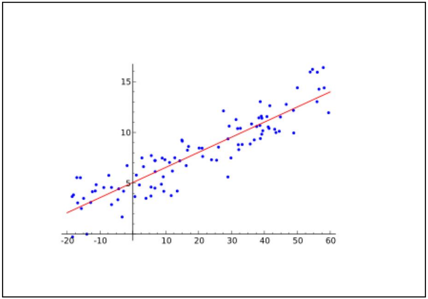
$E(\widehat{\eta}) = E(\eta)$ *and* $cov(\eta - \widehat{\eta}, \widehat{\eta}) = 0$.

Remark

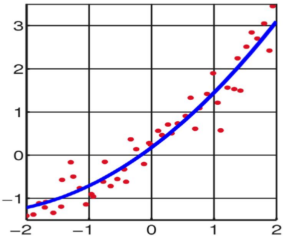
η is splitted into a component parallel to ξ and another one which is orthogonal to ξ.(scalar product: covariance)

- amount of the rainwater- yield in agriculture
- IQ of mother - IQ of the child
- support at the local election- support at the country's election
- height - mass
- grade in microeconomics - grade in mathematics
- grade in microeconomics+ grade in macroeconomics - grade in mathematics
- temperature - gas consumption
- income - outcome
- ...

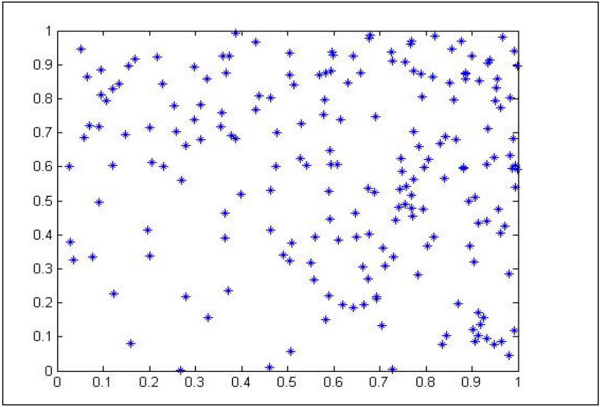
Which type of functions is worth using for approximation?



Which type of functions is worth using for approximation?



Which type of functions is worth using for approximation?



Statement

Having pairs of data (x_i, y_i) *with* $P(\xi = x_i, \eta = y_i) = \frac{1}{n}$, *then*

$E(\xi \cdot \eta) = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{n}, \qquad E(\xi) = \overline{x}, \qquad E(\eta) = \overline{y},$

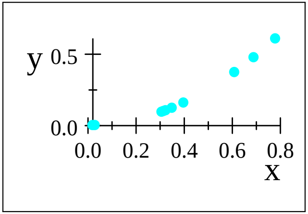
$cov(\xi,\eta) = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{n} - \overline{x} \cdot \overline{y},$

$D(\xi) = \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{n} - \overline{x}^2}, \; D(\eta) = \sqrt{\sum_{i=1}^{n} \frac{y_i^2}{n} - \overline{y}^2}.$

Now

$$\widehat{\eta} = \frac{\frac{\sum_{i=1}^{n} x_i \cdot y_i}{n} - \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} \frac{x_i^2}{n} - \overline{x}^2}(\xi - \overline{x}) + \overline{y}$$

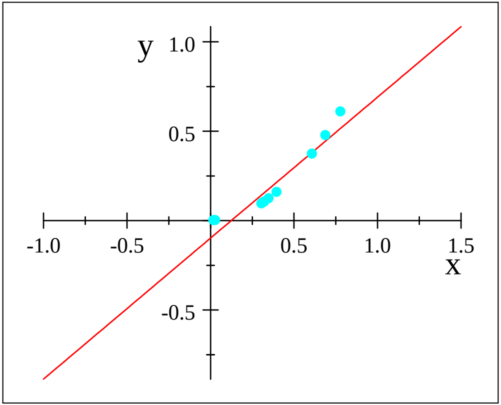
data: $(0.34855;\ 0.34855^2)$ $(0.30736;\ 0.30736^2)$
$(0.78392;\ 0.78392^2)$ $(0.32371;\ 0.32371^2)$
$(0.31560;\ 0.31560^2)$ $(3.0718\cdot\ 10^{-2};(3.0718\cdot\ 10^{-2})^2)$
$(0.69065;\ 0.69065^2)$ $(0.607\ 36;\ 0.60736^2)$
$(0.39860;\ 0.39860^2)$ $(2.1681\cdot 10^{-2};\ (2.1681\cdot 10^{-2})^2)$



$E(\xi) = \overline{x} = 0.382\ 81$, $E(\eta) = \overline{y} = 0.204\ 11$,
$D^2(\xi) = 6.\ 395\ 4 \cdot 10^{-2}$, $\mathrm{cov}(\xi, \eta) = 5.0443 \cdot 10^{-2}$
regression function:

$$\begin{aligned} y &= \frac{5.0443 \cdot 10^{-2}}{6.3954 \cdot 10^{-2}}(x - 0.38281) + 0.20411 \\ y &= 0.78874x - 9.7834 \cdot 10^{-2} \end{aligned}$$

Simple linear regression
Example - fitting for data iv)



Example

Let ξ be uniformly distributed r.v. on [0,1],let $\eta = \xi^2$.
Determine the linear regression of η by ξ.

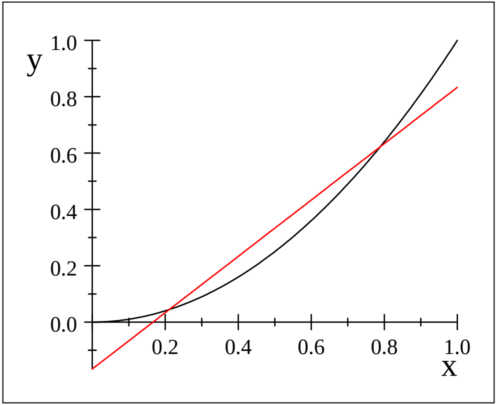
Solution

coefficients: $a = \frac{cov(\xi,\eta)}{D^2(\xi)}$, $b = E(\eta) - aE(\xi)$.
$E(\xi) = 0.5$, $D(\xi) = 1/\sqrt{12}$, $E(\eta) = \int_0^1 x^2 dx = 0.33333$,
$E(\xi \cdot \eta) = E(\xi^3) = \int_0^1 x^3 dx = 0.25$;
$cov(\xi,\eta) = 0.25 - \frac{1}{2} \cdot \frac{1}{3} = 8.3333 \times 10^{-2}$;
$D(\eta) = 0.29814$;

Simple linear regression

Example v)

$$\widehat{\eta} = \frac{8.3333 \times 10^{-2}}{\frac{1}{12}}(\xi - 0.5) + \frac{1}{3} = \xi - \frac{1}{6}$$



Example

Let the height and the weight of a man be a two-dimensional normally distributed r.v.. The height is ξ˜ N(177,10), the weight is η˜N(75, 12) and correlation coefficient is 0.7.
a) Determine the best linear approximation of the height by the help of weight.
b) Determine the best approximation of the height by the help of weight.

Solution

ξ ~N(177, 10) (height), η ~N(75, 12) (weight), $r(\xi, \eta) = 0.7$.
a) linear regression:

$$\frac{\widehat{\eta} - E(\eta)}{D(\eta)} = r(\xi, \eta) \cdot \frac{\xi - E(\xi)}{D(\xi)}$$

$$\frac{\widehat{\eta} - 75}{12} = 0.7 \cdot \left(\frac{\xi - 177}{10}\right)$$

$$\widehat{\eta} = 8.4\left(\frac{\xi - 177}{10}\right) + 75$$

Solution

b) In case of m.d. normal distribution the best approximation is linear, therefore the best approximation is

$$\widehat{\eta} = 8.4\left(\frac{\xi - 177}{10}\right) + 75.$$

Example

c) If the height of a man is 165 cm, determine the approximate value of the weight.

Solution

$\widehat{\eta} = 8.4 \cdot \frac{165-177}{10} + 75 = 64.92$

Example

d) Determine the best linear approximation of the weight by the help of height.

Solution

$$\frac{\widehat{\xi} - E(\xi)}{D(\xi)} = r(\xi, \eta) \cdot \frac{\eta - E(\eta)}{D(\eta)}$$

$$\frac{\widehat{\xi} - 177}{10} = 0.7 \cdot \frac{\eta - 75}{12}$$

$$\widehat{\xi} = 10 \cdot 0.7 \cdot \frac{\eta - 75}{12} + 177 = \frac{7}{12}\eta + 133.5$$

Example

If the weight of a person is 82kg how much is the approximate value of the height?

Solution

$\widehat{\zeta} = \frac{7}{12} \cdot 82 + 133.5 = 181.33$

Example

Let $\underline{\xi} = (\xi_1, \xi_2)$ be a m.d.r.v. with j.p.d.f.

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, 0 < x_1 < 1, 0 < x_2 < 1 \\ 0, \ \textit{otherwise} \end{cases}$$

Determine the linear regression of ξ_2 by the help of ξ_1

Solution

we need $E(\xi_1)$,$D(\xi_1)$,$E(\xi_2)$, $cov\ (\xi_1,\xi_2)$

$E(\xi_1) = \int_0^1 x_1(x_1 + 0.5)dx_1 = 0.583\,33 = E(\xi_2)$

$D(\xi_1) = \sqrt{\int_0^1 x_1^2(x_1 + 0.5)dx_1 - 0.583\,33^2} = 0.276\,39 = D(\xi_2)$

$E(\xi_1 \cdot \xi_2) = \int_0^1 \int_0^1 x_1 \cdot x_2 \cdot (x_1 + x_2)dx_1 dx_2 = 0.33333$

$cov(\xi_1, \xi_2) = 0.33333 - 0.58333^2 = -6.\,943\,9 \times 10^{-3}$

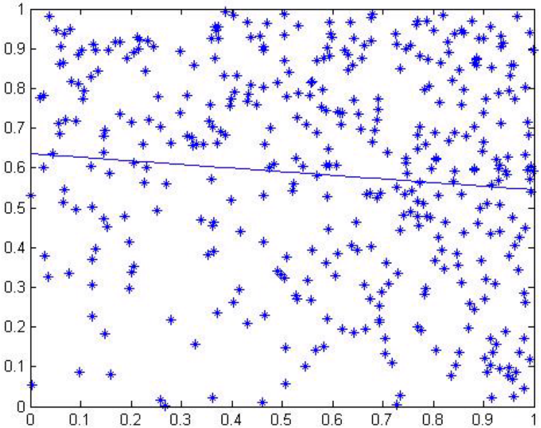
Solution

$$\begin{aligned}\widehat{\xi_2} &= \frac{-6.9439 \times 10^{-3}}{0.27639^2}(\xi_1 - 0.58333) + 0.58333 \\ &= -9.0899 \times 10^{-2} \cdot \xi_1 + 0.63635.\end{aligned}$$

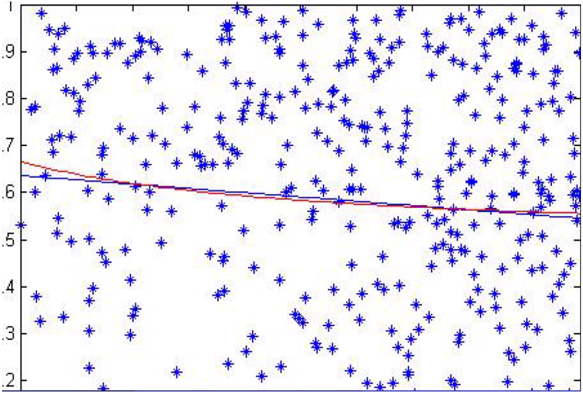
The function of the best linear approximation:

$$y = -9.0899 \times 10^{-2} \cdot x + 0.63635$$

The best linear approximation:



The best (nonlinear) approximation $E(\xi_2|\xi_1 = x_1) = \frac{x_1}{x_1+0.5} \cdot \frac{1}{2} + \frac{1}{x_1+0.5} \cdot \frac{1}{3}$



Let $\underline{\xi}$ be n dimensional r.v. (written in column vector)
Let η be a r.v.;
We suppose, that the dispersions are finite.
We want to find the linear function of $\underline{\xi}$ which approximates the best the r.v. η.

$$\underline{a} \in \mathbb{R}^n =?, b \in \mathbb{R} =?, \quad \textit{for which } E((\eta - (a^T \cdot \underline{\xi} + b))^2) \textit{ is minimal}.$$

- one dependent variable η,
- more than one independent (explanatory) variables (ξ_i, i=1,2,...,n).

Statement

The solution of the above problem

$$a^T = cov(\eta, \underline{\xi}) \cdot cov(\underline{\xi}, \underline{\xi})^{-1}$$

$$b = E(\eta) - a^T E(\underline{\xi}).$$

Remark

The expectation of the squared difference is

$$E((\eta - \widehat{\eta})^2) = D^2(\eta) - cov(\eta, \underline{\xi}) \cdot cov(\underline{\xi}, \underline{\xi})^{-1} \cdot cov(\eta, \underline{\xi})^T$$

Example

The lifetime of a machine is a r.v. with expectations 100 units, dispersion 10 units. The lifetime is approximated by the help of the time elapsed from the shopping and the time turned on. Determine the best linear approximation of the lifetime if we know the expectations, dispersions and covariances.

Example

η : a the lifetime of the machine
ξ_1 : the time elapsed from the shopping
ξ_2 : the time turned on

$$\underline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

$\mathrm{E}(\xi_1) = 80$, $\mathrm{E}(\xi_2) = 50$,
$\mathrm{D}(\xi_1) = 30$, $\mathrm{D}(\xi_2) = 20$, $\mathrm{r}(\xi_1,\xi_2) = 13$,
$\mathrm{cov}\,(\eta,\, \xi_1) = 6$, $\mathrm{cov}\,(\eta,\, \xi_2) = -50$

Multivariate linear regression

Example - computations i)

Solution

$cov\ (\eta, \underline{\xi}) = (6,\text{-}50)$

$$cov\ (\underline{\xi}, \underline{\xi})^{-1} = \begin{pmatrix} 900 & 200 \\ 200 & 400 \end{pmatrix}^{-1} = \begin{pmatrix} 0.00125 & 6.25 \cdot 10^{-4} \\ 6.25 \cdot 10^{-4} & 2.8125 \cdot 10^{-3} \end{pmatrix}$$

$a^T = cov\ (\eta, \underline{\xi}) \cdot cov\ (\underline{\xi}, \underline{\xi})^{-1}$

$$\begin{pmatrix} 6 & -50 \end{pmatrix} \cdot \begin{pmatrix} 0.00125 & 6.25 \cdot 10^{-4} \\ 6.25 \cdot 10^{-4} & 2.8125 \cdot 10^{-3} \end{pmatrix} = (0.03875,\text{-}0.14438)$$

Solution

$E(\eta)$–a^T $E(\underline{\xi})$ =100 -(0.03875 ,-0.14438)$\cdot \begin{pmatrix} 80 \\ 50 \end{pmatrix}$ =104.12.

$$\widehat{\eta} = (0.03875, -0.14438) \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + 104.12 =$$
$$0.03875\xi_1 - 0.14438\xi_2 + 104.12.$$

The slides were made by the support of the project

$$EFOP - 3.4.3 - 16 - 2016 - 00009.$$

The author expresses her thanks for the financial support.