Lecture 4.

Random variable and distribution of probability

The concept of random variable

Random variable is a function X, that attributes a real value x to a certain results of a random experiment.

$$\Omega = \{e_1, e_2, \ldots\}$$

$$X : \Omega \to R$$

$$X(e_i) = x_i \in R$$

Examples:

1) Coin toss: event 'head' takes a value of 1; event 'tails' - 0.
2) Products: event 'failure' - 0, well-performing – 1
3) Dice: '1' – 1, '2' – 2 etc.…
4) Interval [a, b]– a choice of a point of a coordinate 'x' is attributed a value, e.g. $\sin^2(3x+17)$ etc. ….

The concept of random variable

Random variable

Discrete

When the values of random variable X are isolated points on an number line

- **Toss of a coin**
- **Transmission errors**
- **Faulty elements on a production line**
- **A number of connections coming in 5 minutes**

Continuous

When the values of random variable cover all points of an interval

- **Electrical current, I**
- **Temperature, T**
- **Pressure, p**

Distribution of random variable

Distribution of random variable (probability distribution for discrete variables) is a set of pairs (x_i, p_i) where x_i is a value of random variable X and p_i is a probability, that a random variable X will take a value x_i

Example

Probability mass function for a single toss of coin.
Event corresponding to heads is attributed $x_1=1$; tails means $x_2=0$.

$$x_1 = 1 \quad p(X=1) = p(x_1) = \frac{1}{2}$$

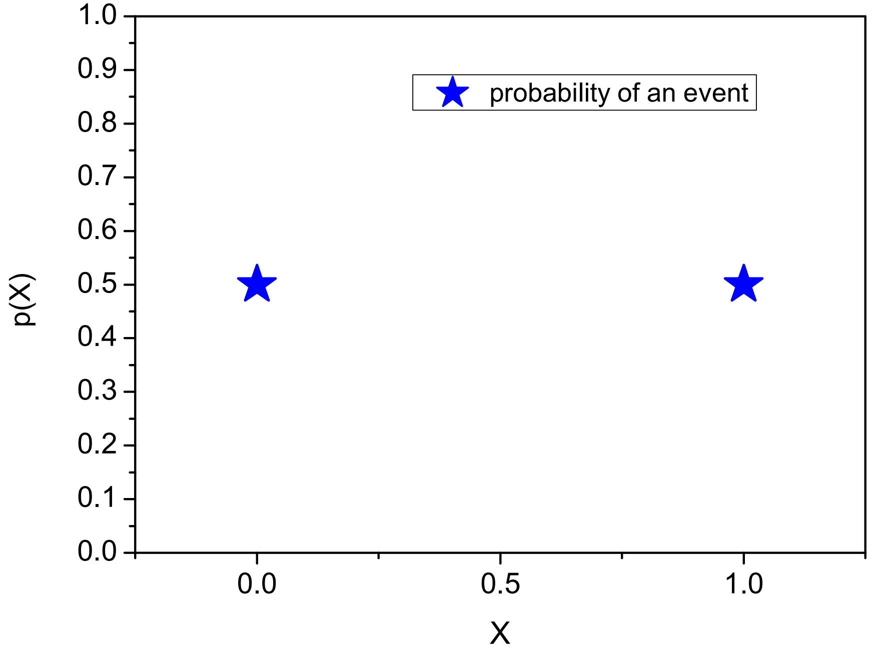
$$x_2 = 0 \quad p(X=0) = p(x_2) = \frac{1}{2}$$

Distribution of random variable

Example

Probability mass function for a single toss of coin is given by a set of the following pairs:

$$\{(1, \frac{1}{2}), (0, \frac{1}{2})\}$$



Random variable when discrete entails probability distribution also discrete.

Distribution of random variable

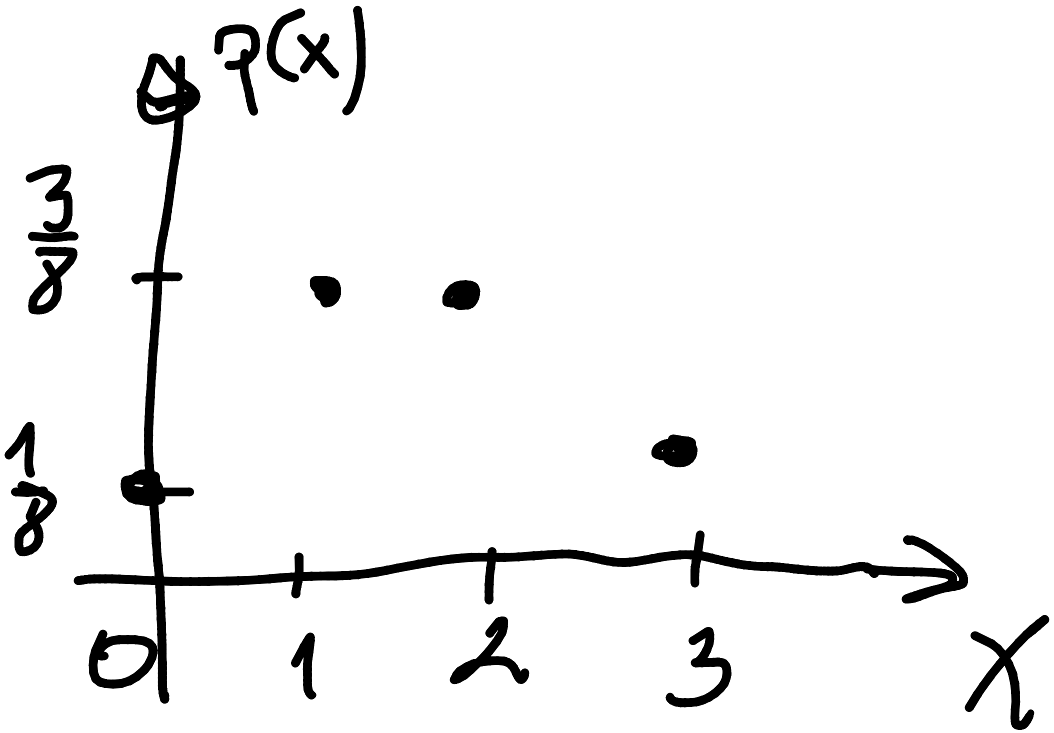
Example

A coin tossed three times. Random variable X takes values equal to the number of heads that can be obtained in this experiment. Please determine the probability distribution function of X.

TTT → 0
TTH → 1
THT → 1
HTT → 1
HHT → 2
HTH → 2
THH → 2
HHH → 3

X	0	1	2	3
$p(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$p(X) = \begin{cases} \frac{1}{8}; & X \in \{0,3\} \\ \frac{3}{8}; & x \in \{1,2\} \end{cases}$$


P(x)
3/8
1/8
0
1
2
3
X

Quantitative description of random variables

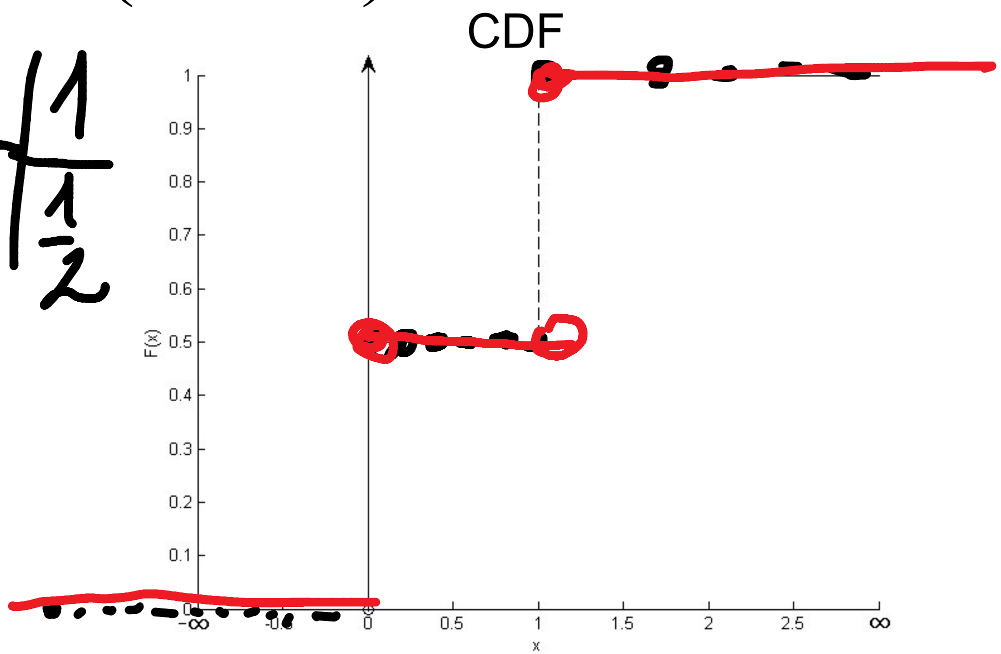
- Cumulative distribution function (CDF) F(x) is a probability of an event that the random variable X will assume a value smaller than or equal to x (at most x)

$$F(x) = P(X \le x)$$

Example

CDF of coin toss:

x	0	1
p(x)	$\frac{1}{2}$	$\frac{1}{2}$



$$F(x=0) = P(X \le 0) = \frac{1}{2}$$

$$F(x=1) = P(X \le 1) = 1$$

$$P(X) = \frac{1}{2} \; ; \; x \in \{0, 1\}$$

$$F(x) = \begin{cases} 0 \; ; & x \in (-\infty, 0) \\ \frac{1}{2} \; ; & x \in \langle 0, 1) \\ 1 \; ; & x \in \langle 1, +\infty) \end{cases}$$

Properties of CDF

1. $0 \le F(x) \le 1$

2. $F(-\infty) = 0$

3. $F(+\infty) = 1$

4. $x \le y \;\Rightarrow\; F(x) \le F(y)$

   non-decreasing function

5. F(x) has no unit

CDF of discrete variable

$$F(x) = P(X \le x) = \sum_{x_i \le x} p(x_i)$$

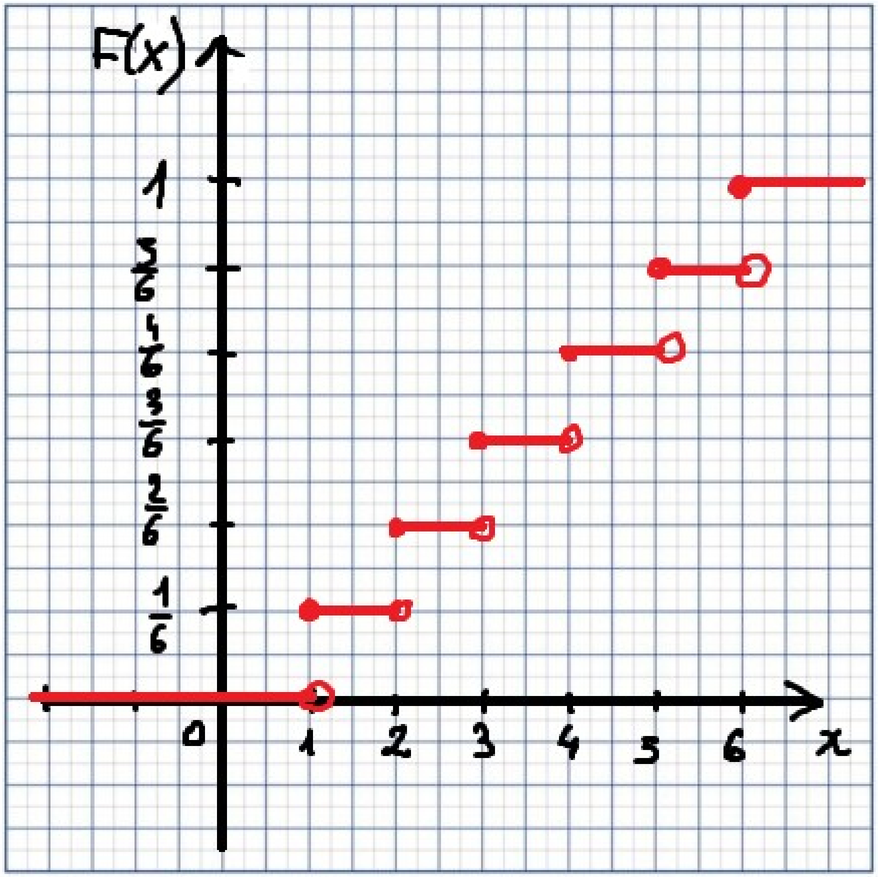
p* (x_i) – *probability mass function

Example

Throw a dice

X	1	2	3	4	5	6
p(X)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$F(x) = \begin{cases} 0; & x \in (-\infty, 1) \\ \frac{1}{6}; & x \in \langle 1, 2) \\ \frac{2}{6}; & x \in \langle 2, 3) \\ \vdots \end{cases}$$



CDF of discrete variable

$$F(x) = P(X \le x) = \sum_{x_i \le x} p(x_i)$$

$p(x_i)$ – probability mass function

Example

Determine CDF from the following probability mass function p(X):

X	-2	-1	$\frac{1}{2}$	3
p(X)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{4}$

$$F(x) = \begin{cases} 0 ; & x \in (-\infty, -2) \\ \frac{1}{8} ; & x \in \langle -2, -1) \\ \frac{1}{2} ; & x \in \langle -1, \frac{1}{2}) \\ \frac{3}{4} ; & x \in \langle \frac{1}{2}, 3) \\ 1 ; & x \in \langle 3, +\infty) \end{cases}$$

CDF of discrete variable

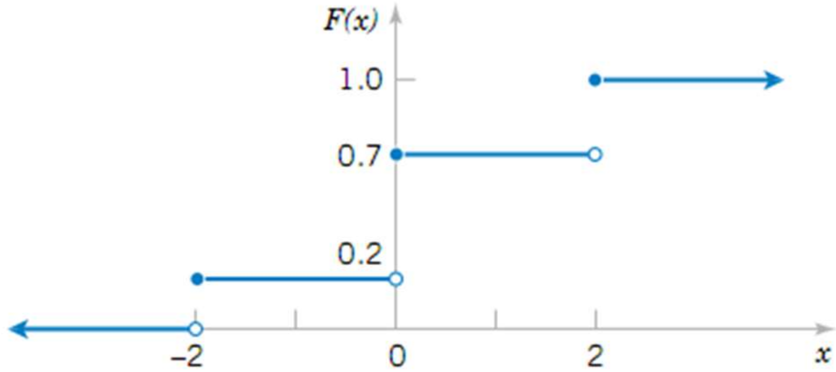
$$F(x) = P(X \le x) = \sum_{x_i \le x} p(x_i)$$

p (x_i) – probability mass function

Example

Determine probability mass function of X from the following cumulative distribution function F(x)

$$F(x) = \begin{cases} 0 & for \quad x < -2 \\ 0.2 & for \quad -2 \le x < 0 \\ 0.7 & for \quad 0 \le x < 2 \\ 1 & for \quad 2 \le x \end{cases}$$



From the plot, the only points to receive p(x)≠0 are -2, 0, 2.

$$p(-2) = 0.2 - 0 = 0.2 \qquad p(0) = 0.7 - 0.2 = 0.5 \qquad p(2) = 1.0 - 0.7 = 0.3$$

Numerical descriptors

Median ($x_{0.5}$) is the value separaitng lower and higher half of the sample / population.

$$\frac{22+22}{2} = 22$$

Example

For a discrete distribution : 19, 21, 21, 21, 22, 22, 23, 25, 26, 27
median is 22 (middle value or arithmetic average of two middle values)

Numerical descriptors

Mode represents the most frequently occurring value of random variable (x at which probability distribution attains a maximum)

Unimodal distribution has one mode (multimodal distributions – more than one mode)

Example: 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 - mode equals to 21 (which appears 3 times, i.e., the most frequently)

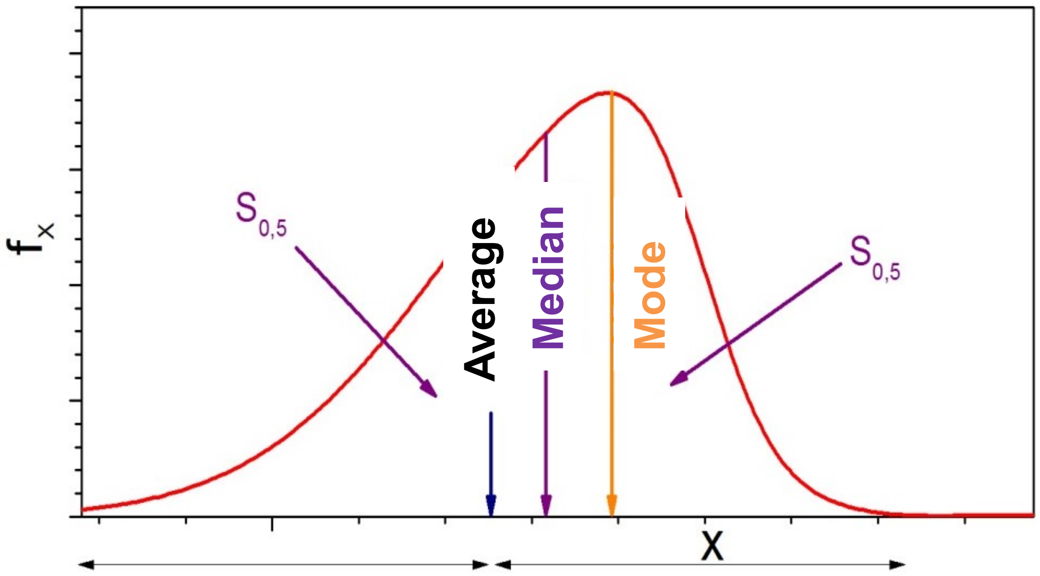
Average value

Arithmetic average:

x_i - belongs to a set of n – elements

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

19, 21, 21, 21, 22, 22, 23, 25, 26, 27 -
the arithmetic average is 22.7



Expected value

Symbols: m_1, E(X), μ, $\bar{x}$, $\hat{x}$

$$E(X) = \sum_i x_i p_i$$

The expected value is the probability-weighted average value of a random variable

Expected value

Symbols: m_1, E(X), μ, $\bar{x}$, $\hat{x}$

$$E(X) = \sum_i x_i p_i$$

The expected value is the probability-weighted average value of a random variable

A coin tossed three times. Let the random variable X take values equal to the number of heads that can be obtained in this experiment. Please find the expected value of X

x	0	1	2	3
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

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

$$= 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = 1.5$$

$$1.5 \cdot 100 = 150$$

Expected value

A grab bag contains 12 packages worth 80 cents apiece, 15 packages worth 40 cents apiece and 25 packages worth 30 cents apiece. Is it worthwhile to pay 50 cents for the privilege of picking one of the packages at random?

X	80	40	30
$P(X)$	$\frac{12}{52}$	$\frac{15}{52}$	$\frac{25}{52}$

$$12+15+25 = 52$$

$$E(X) = 80\cdot\frac{12}{52} + 40\cdot\frac{15}{52} + 30\cdot\frac{25}{52} \approx 44.4$$

Properties of E(X)

E(X) is a linear operator, i.e.:

1. $$E(\sum_i C_i X_i) = \sum_i C_i E(X_i)$$

In a consequence:

$E(C) = C$

$E(CX) = CE(X)$

$E(X_1+X_2)=E(X_1)+E(X_2)$

2. For independent variables $X_1, X_2, \dots X_n$

$$E(\prod_i X_i) = \prod_i E(X_i)$$

Variables are independent when:

$$f(X_1, X_2, \dots, X_n) = f_1(X_1) f_2(X_2) \cdot \dots \cdot f_n(X_n)$$

Variance

VARIANCE (dispersion) symbols: $\sigma^2(X)$, var(X), V(X), $D^2(X)$.
Standard deviation σ(x)

$$\sigma^2(X) \equiv \sum_i p_i (x_i - E(X))^2$$

Variance (or the standard deviation) is a measure of scatter of random variables around the expected value.

$$\sigma^2(X) = E(X^2) - E^2(X)$$

$$\sigma^2(X) \equiv \sum_i p_i (x_i - E(X))^2 =$$

$$= \sum p_i \cdot (x_i^2 - 2x_i E(X) + E^2(X)) =$$

$$= \sum [p_i x_i^2 - 2 p_i x_i E(X) + p_i E^2(X)] =$$

$$= \underbrace{\sum_i p_i x_i^2}_{E(X^2)} - 2E(X) \underbrace{\sum_i p_i x_i}_{E(X)} + E^2(X) \underbrace{\sum_i p_i}_{1} =$$

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

Variance

$$\sigma^2(X) = E(X^2) - E^2(X)$$

We roll a fair die. If we roll a 1, then we win $25, if you roll a 2, we win $5. If we roll a 3, we win nothing. If we roll a 4 or a 5, we loose $10, and if you roll a 6, we loose $15. Assuming that X is our profit in a single game, please calculate E(X) and D(X).

x	25	5	0	-10	-15
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

$$E(X) = 25\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + (-10)\cdot\frac{2}{6} + (-15)\cdot\frac{1}{6} =$$

$$= -\frac{5}{6} \approx -0.83\ \$$$

$$D^2(X) = E(X^2) - E^2(X) = 25^2\cdot\frac{1}{6} + 5^2\cdot\frac{1}{6} + 10^2\cdot\frac{2}{6} +$$

$$+ 15^2\cdot\frac{1}{6} - \left(-\frac{5}{6}\right)^2 = 178.47 \qquad D(X) = 13.4$$

Properties of $\sigma^2(X)$

Variance can be calculated using expected values only:

1. $$\sigma^2(X) = E(X^2) - E^2(X)$$

In a consequence we get:

$$\sigma^2(C) = 0$$

$$\sigma^2(CX) = C^2 \sigma^2(X)$$

$$\sigma^2(C_1X + C_2) = C_1^2 \sigma^2(X)$$

2. For independent variables $X_1, X_2, \dots X_n$

$$\sigma^2\left(\sum_i C_i X_i\right) = \sum_i C_i^2 \sigma^2(X)$$