

Quantitative Methods – I

A.Y. 2021-22

Practice 7

Lorenzo Cavallo

For any clarification/meeting: cavallo@istat.it

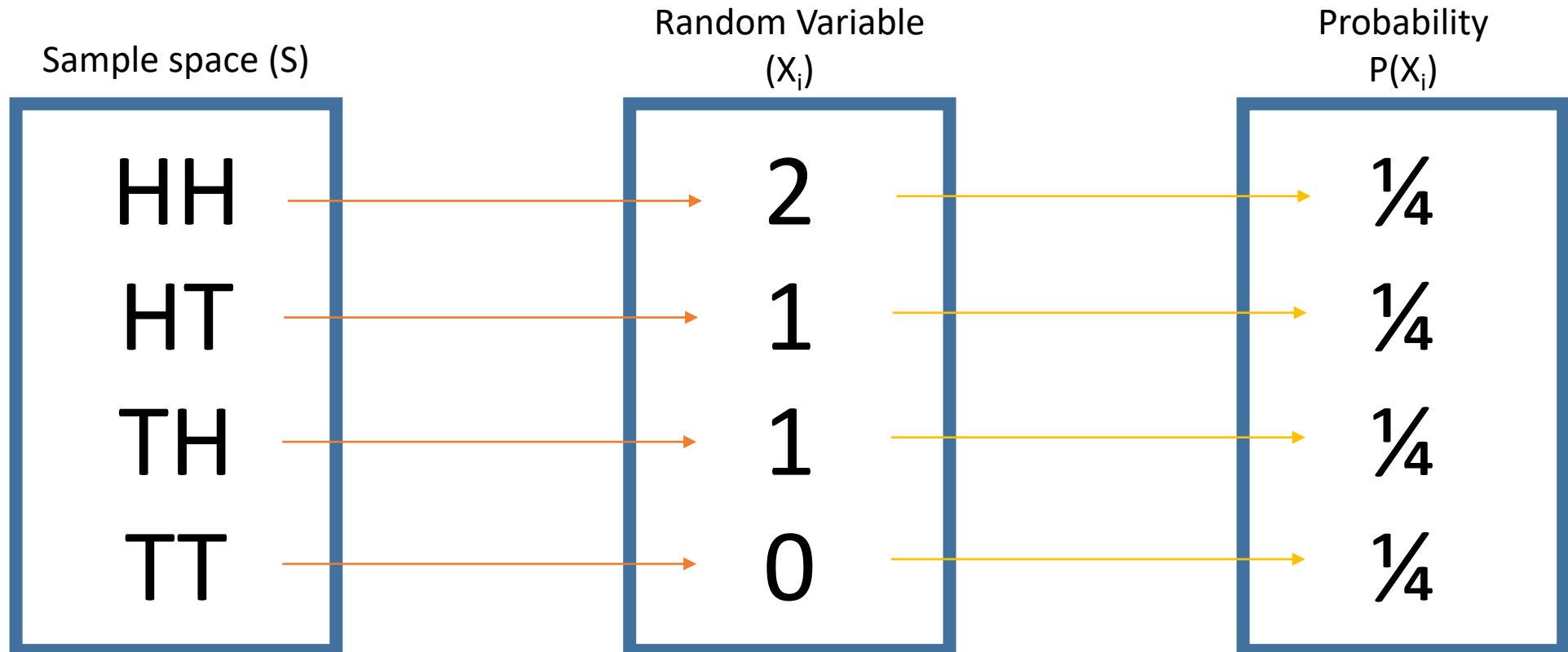
THEME #1

Random Variables & Probability Distribution

Random Variable

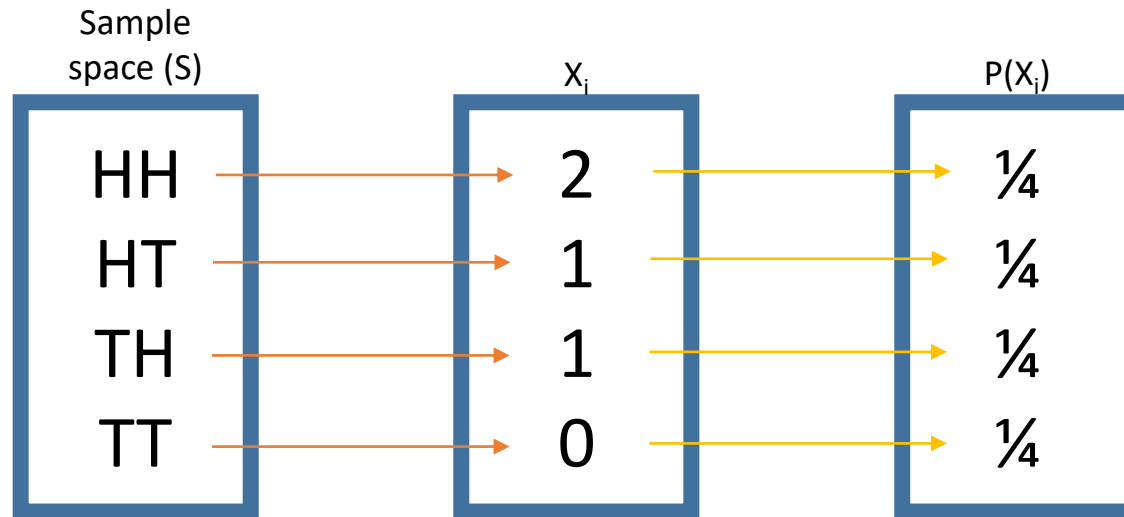
A variable whose value is determined by the outcome of an experiment.

Ex. Number of Head from the toss of two coins



Probability distributions

Lists all the possible values that the random variable can assume and their corresponding probabilities.

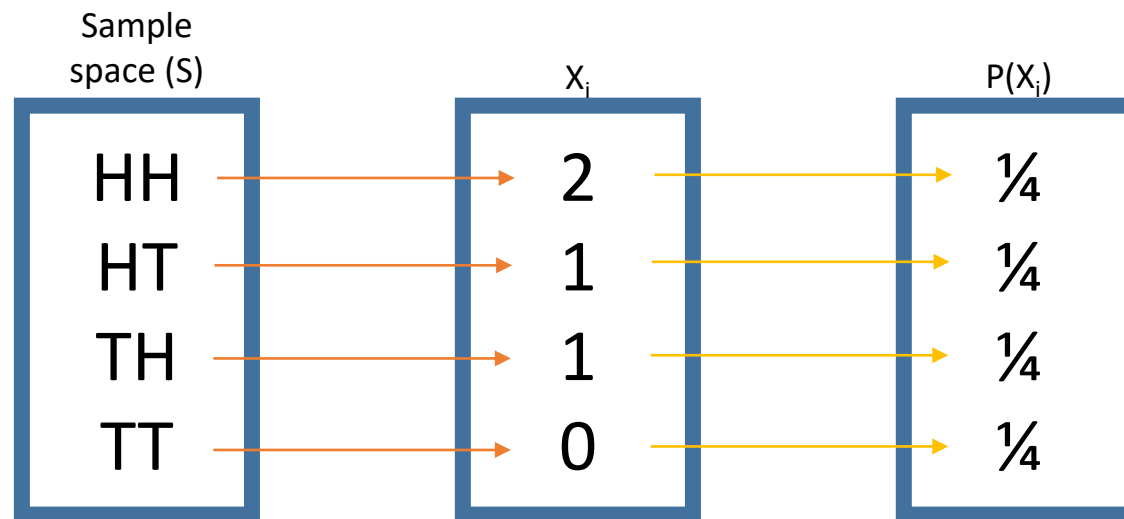


X_i	$P(X_i)$
0	$\frac{1}{4}=0.25$
1	$\frac{2}{4}=0.5$
2	$\frac{1}{4}=0.25$
Tot.	1.0

Probability distributions

2 conditions:

1. $0 \leq P(X_i) \leq 1$
2. $\sum P(X_i) = 1$



X _i	P(X _i)
0	1/4=0.25
1	2/4=0.5
2	1/4=0.25
Tot.	1.0

Exercise 1

Each of the following tables lists certain values of x and their probabilities. Determine whether or not each table represents a valid probability distribution.

a)

x_i	$P(x_i)$
0	0.18
1	0.01
2	0.29
3	0.37

b)

x_i	$P(x_i)$
2	0.35
3	0.24
4	0.18
5	0.23

c)

x_i	$P(x_i)$
7	0.65
8	0.50
9	-0.15

Solution

a) Each probability listed in this table is in the range 0 to 1, it satisfies the first condition of a probability distribution.

But, the sum of all probabilities is not equal to 1.0 because $\Sigma P(x_i) = 0.18 + 0.01 + 0.29 + 0.37 = 0.85$ and the second condition is not satisfied.

This table **does not represent** a valid probability distribution.

Exercise 1

Each of the following tables lists certain values of x and their probabilities. Determine whether or not each table represents a valid probability distribution.

a)

x_i	$P(x_i)$
0	0.18
1	0.01
2	0.29
3	0.37

b)

x_i	$P(x_i)$
2	0.35
3	0.24
4	0.18
5	0.23

c)

x_i	$P(x_i)$
7	0.65
8	0.50
9	-0.15

Solution

b) Each probability listed in this table is in the range 0 to 1.

$$\text{Also, } \Sigma P(x_i) = 0.35 + 0.24 + 0.18 + 0.23 = 1.0$$

Consequently, this table **represents** a valid probability distribution.

Exercise 1

Each of the following tables lists certain values of x and their probabilities. Determine whether or not each table represents a valid probability distribution.

a)

x_i	$P(x_i)$
0	0.18
1	0.01
2	0.29
3	0.37

b)

x_i	$P(x_i)$
2	0.35
3	0.24
4	0.18
5	0.23

c)

x_i	$P(x_i)$
7	0.65
8	0.50
9	-0.15

Solution

c) The sum of all probabilities listed in this table is equal to 1.0, but one of the probabilities is negative.

This violates the first condition of a probability distribution.

Therefore, this table **does not represent** a valid probability distribution.

Exercise 2

We toss 3 coins. Let X be random variable that counts the number of heads. Obtain the probability distribution of X .

Also calculate the probability of having :

1. two heads;
2. no head;
3. more than one head;
4. at least one head;
5. less than three heads;
6. at most one head.

Solution

The 8 possible outcomes, w , of the experiment and the probabilities with which they occur, $P(w)$, are:

w	$P(w)$
HHH	1/8
HHT	1/8
HTH	1/8
HTT	1/8
THH	1/8
THT	1/8
TTH	1/8
TTT	1/8

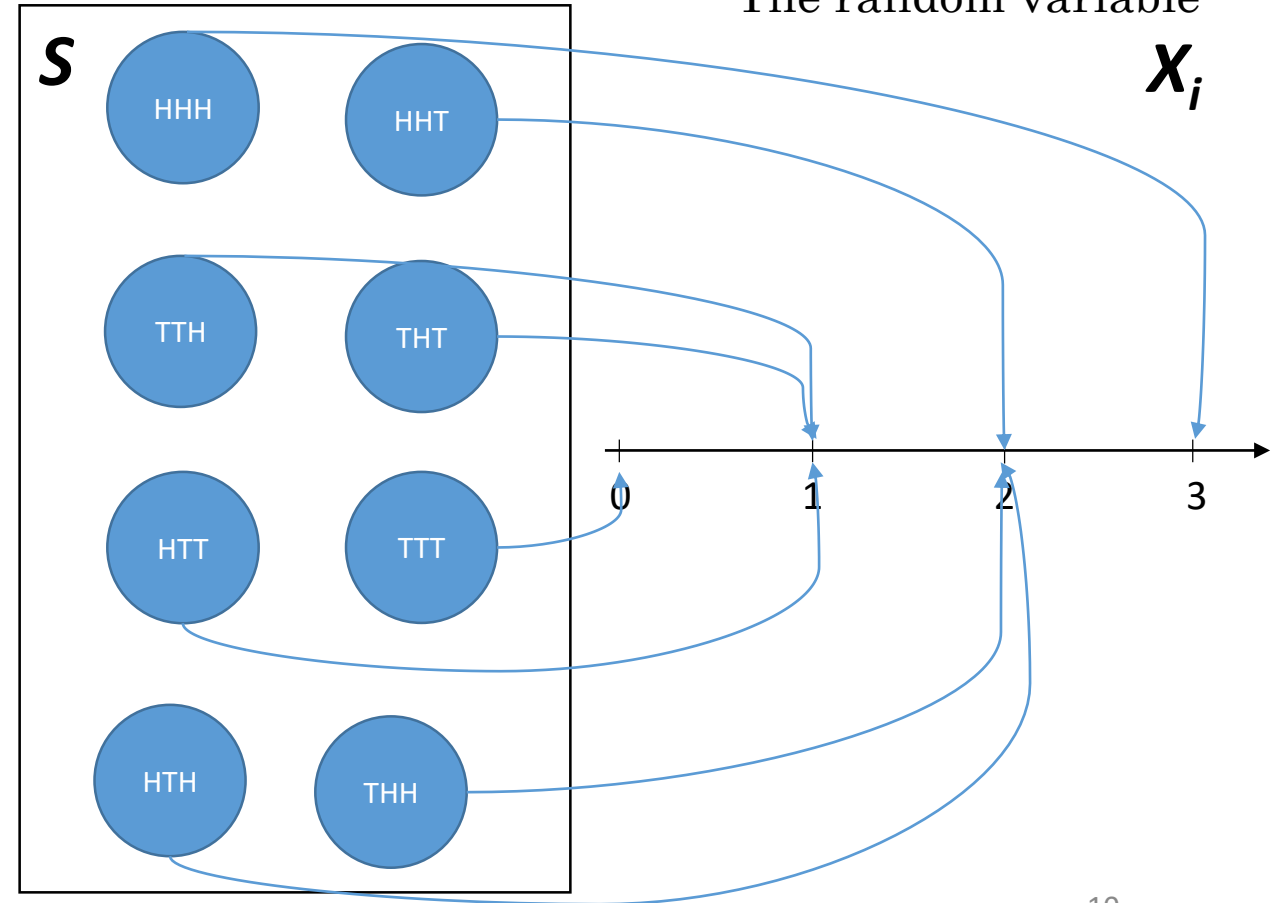
Exercise 2

We toss 3 coins. Let X be random variable that counts the number of heads. Obtain the probability distribution of X .

Solution

X is the random variable that counts the number of heads.

w	$P(w)$	X_i
HHH	$1/8$	3
HHT	$1/8$	2
HTH	$1/8$	2
THH	$1/8$	2
HTT	$1/8$	1
THT	$1/8$	1
TTH	$1/8$	1
TTT	$1/8$	0



Exercise 2

We toss 3 coins. Let X be random variable that counts the number of heads. Obtain the probability distribution of X .

Also calculate the probability of having :

1. two heads;
2. no head;
3. more than one head;
4. at least one head;
5. less than three heads;
6. at most one head.

Solution

w	P (w)	X_i
HHH	1/8	3
HHT	1/8	2
HTH	1/8	2
THH	1/8	2
HTT	1/8	1
THT	1/8	1
TTH	1/8	1
TTT	1/8	0

So, the probability distribution of X is:

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

Exercise 2

We toss 3 coins. Let X be random variable that counts the number of heads. Obtain the probability distribution of X .

Also calculate the probability of having :

1. two heads;
2. no head;
3. more than one head;
4. at least one head;
5. less than three heads;
6. at most one head.

Solution

From this probability distribution we can calculate the probability from 1. to 6.:

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

1. $P(X = 2) = 3/8$

2. $P(X = 0) = 1/8$

3. $P(X > 1) = P(X = 2) + P(X = 3) = 3/8 + 1/8 = 4/8 = 1/2$

4. $P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) = 1/8 + 3/8 + 3/8 = 7/8$

5. $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 7/8$

or $P(X < 3) = 1 - P(X = 3)$ [2nd condition of prob. distr.] $= 1 - 1/8 = 7/8$

6. $P(X \leq 1) = P(X = 0) + P(X = 1) = 1/8 + 3/8 = 4/8 = 1/2$

Exercise 3

The following table lists the probability distribution of the number of breakdowns per week for a machine based on past data.

Breakdowns per week	0	1	2	3
Probability	0.15	0.20	0.35	0.30

Find the probability that the number of breakdowns for this machine during a given week is:

- a) exactly 2
- b) less than 2
- c) more than 1
- d) at most 1

Solution

- a) $P(X=2) = 0.35 = 35\%$
- b) $P(X < 2) = P(X=0) + P(X=1) = 0.15 + 0.20 = 0.35 = 35\%$
- c) $P(X > 1) = P(X=2) + P(X=3) = 0.35 + 0.30 = 0.65 = 65\%$
- d) $P(X \leq 1) = P(X=0) + P(X=1) = 0.15 + 0.2 = 0.35 = 35\%$

THEME #2

Expected value and Variance of random variables

Expected Value

$E(X)$ is the mean of the probability distribution

$$E(X) = \mu = \sum x_i \cdot P(x_i)$$

Variance and Standard Deviation

They measure the spread of the probability distribution:

$$\text{Var}(X) = \sigma^2 = E(X - E(X))^2 = E(X^2) - E(X)^2$$

$$\text{For discrete, } V(X) = \sigma^2 = \sum x_i^2 \cdot P(x_i) - \mu^2$$

Standard Deviation

$$SD(X) = \sigma = \sqrt{\sigma^2}$$

Exercise 4

Let X be the random variable indicating the number of textbooks required by a randomly chosen statistical course.

The r.v. X can assume only the values between 0 and 5.

The table below provides the probability distribution of the random variable:

X_i	0	1	2	3	4	5
$P(X = x_i)$?	0.45	0.24	0.12	0.09	0.05

- What is the of the missing probability?
- What is the probability that a course will require 2 or more textbooks?
- What is the probability that a course require 2 or 3 textbooks?
- Find the expected value of the random variable X .
- Find the variance and the standard deviation of the random variable X .

Exercise 4

Let X be the random variable indicating the number of textbooks required by a randomly chosen statistical course.

The r.v. X can assume only the values between 0 and 5.

The table below provides the probability distribution of the random variable:

X_i	0	1	2	3	4	5
$P(X = x_i)$?	0.45	0.24	0.12	0.09	0.05

Solution

a. What is the of the missing probability?

The probability distributions $P(X = x)$ have 2 characteristics:

1. For each value of X , $0 \leq P(Xi) \leq 1$
2. $\sum P(Xi) = 1$

From the second characteristic we have:

$$P(X = 0) + \sum P(X = x_i) = 1 \text{ then } P(X = 0) = 1 - \sum P(X = x_i) = 1 - 0.95 = 0.05$$

The probability of $P(X=0)$ is 0.05

Exercise 4

Let X be the random variable indicating the number of textbooks required by a randomly chosen statistical course.

The r.v. X can assume only the values between 0 and 5.

The table below provides the probability distribution of the random variable:

X_i	0	1	2	3	4	5
$P(X = x_i)$	0.05	0.45	0.24	0.12	0.09	0.05

Solution

b. What is the probability that a course will require 2 or more textbooks?

$P(X \geq 2)$ is a sum of probability:

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = 0.24 + 0.12 + 0.09 + 0.05 = 0.5 = 50\%$$

or the probability of the complementary event: $1 - P(X < 2) = 1 - P(X=0) - P(X=1) = 1 - 0.05 - 0.45 = 0.5$

c. What is the probability that a course require 2 or 3 textbooks?

The two event are disjoint so:

$$P(X=2 \text{ or } 3) = P(X=2 \cup X=3) = P(X=2) + P(X=3) = 0.24 + 0.12 = 0.36 = 36\%$$

Exercise 4

Let X be the random variable indicating the number of textbooks required by a randomly chosen statistical course.

The r.v. X can assume only the values between 0 and 5.

The table below provides the probability distribution of the random variable:

X_i	0	1	2	3	4	5
$P(X = x_i)$	0.05	0.45	0.24	0.12	0.09	0.05

Solution

d. Find the expected value of the random variable X .

For a discrete random variable the expected value $E(X)$ is: $\mu = E(X) = \sum x_i \cdot P(X = x_i)$

X_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
0	0.05	0.00
1	0.45	0.45
2	0.24	0.48
3	0.12	0.36
4	0.09	0.36
5	0.05	0.25
	1.00	1.90

So the expected value of X is:

$$E(X) = \mu = 1.9$$

Exercise 4

Let X be the random variable indicating the number of textbooks required by a randomly chosen statistical course.

The r.v. X can assume only the values between 0 and 5.

The table below provides the probability distribution of the random variable:

X_i	0	1	2	3	4	5
$P(X = x_i)$	0.05	0.45	0.24	0.12	0.09	0.05

Solution

e. Find the variance and the standard deviation of the random variable X

For a discrete random variable the variance is:

$$\sigma^2 = \text{Var}(X) = \sum (x_i - \mu)^2 \cdot P(X = x_i)$$

or

$$\sigma^2 = E(X^2) - \mu^2 = E(X^2) - (E(X))^2 = \sum x_i^2 \cdot P(X = x_i) - \mu^2$$

Use the second notation...

Exercise 4

Let X be the random variable indicating the number of textbooks required by a randomly chosen statistical course.

The r.v. X can assume only the values between 0 and 5.

The table below provides the probability distribution of the random variable:

X_i	0	1	2	3	4	5
$P(X = x_i)$	0.05	0.45	0.24	0.12	0.09	0.05

Solution

e. (cont.)

X_i	$P(X = x_i)$	$x_i^2 \cdot P(X = x_i)$
0	0.05	$0^2 \cdot 0.05$
1	0.45	$1^2 \cdot 0.45$
2	0.24	$2^2 \cdot 0.24$
3	0.12	$3^2 \cdot 0.12$
4	0.09	$4^2 \cdot 0.09$
5	0.05	$5^2 \cdot 0.05$
	1.00	5.18

$$\sigma^2 = E(X^2) - \mu^2 = \sum x_i^2 \cdot P(X = x_i) - \mu^2$$

$$\sigma^2 = 5.18 - 1.90^2 = 5.18 - 3.61 = 1.57$$

σ is the square root of σ^2

$$\sigma = \sqrt{1.57} = 1.25$$

THEME #3

**Bernoulli, Binomial and Poisson
probability distribution**

Bernoulli Probability Distribution

Only 2 values (1=“success”, 0 = “failure”). 1 with probability p , 0 with probability $(1 - p)$

Probability function:

$$P(x_i) = p^{x_i} \times (1 - p)^{(1-x_i)}$$

$$E(X) = p \quad \text{and} \quad V(X) = p \times (1 - p)$$

ex. Accept or decline an investment, vote yes or no on a ballot, etc.

Binomial Random Variable

Represents the number of successes in n Bernoulli experiments : a) independent b) with equal p

Probability function:

$$P(x) = \binom{n}{x} p^x \times (1 - p)^{(n-x)}$$

$$E(X) = n \times p \quad \text{and} \quad V(X) = n \times p \times (1 - p)$$

ex. Number of heads obtained tossing a coin 10 times.

Exercise 4

A club membership is renewed with probability 0.65. Let X be the random variable that represents the decision of each member to renew the subscription.

What is the distribution of X ? Obtain $E(X)$ and $V(X)$.

Solution

The distribution is a Bernoulli probability distribution because the possible event are only 2: “renew the membership” (event 0) or “not renew the membership” (event 1).

The probability distribution is:

$$P(X=x)=p^x \cdot (1-p)^{(1-x)} = 0.65^x \cdot (1-0.65)^{(1-x)}$$

The expected value of X for a Bernoulli distribution is:

$$E(X) = p$$

$$\text{So, } E(X) = p = 0.65 = 65\%$$

The variance of X for a Bernoulli distribution is:

$$\text{Var}(X)= p \cdot (1-p)$$

$$\text{So, } \text{Var}(X)= p \cdot (1-p) = 0.65 \cdot (1-0.65) = 0.65 \cdot 0.35 = 0.2275 = 22.75\%$$

Exercise 5

Based on his experience a travel agent believes that only 60% of the tickets booked then are bought.

If 8 customers have gone independently to the travel agency for a reservation, determine the probability that:

- a. five buy a ticket;
- b. everyone buys a ticket;
- c. at least 2 buy a ticket.

Solution

The X is a Binomial because the customers can only BUY (event 0) or NOT BUY (event 1) a ticket.

The r.v. X is a Binomial distribution $X \sim \text{Bin}(n, p) = \text{Bin}(8; 0.6)$

where $n = 8$ (the customers) and $p = 0.6$ (the probability of buying a ticket)

The probability function is:

$$P(X = x) = \binom{n}{x} p^x \cdot (1 - p)^{(n - x)}$$

- a. The probability that $x = 5$ customers buy a ticket is:

$$\begin{aligned} P(X = 5) &= \binom{n}{x} p^x \cdot (1 - p)^{(n - x)} = \frac{n!}{(n-x)! x!} p^x \cdot (1 - p)^{(n - x)} = \binom{8}{5} 0.60^5 \cdot (1 - 0.60)^{(8 - 5)} = \\ &= 8!/(5! \cdot 3!) \cdot 0.078 \cdot 0.064 = 56 \cdot 0.078 \cdot 0.064 = 0.28 = 28\% \end{aligned}$$

Exercise 5

Based on his experience a travel agent believes that only 60% of the tickets booked then are bought.

If 8 customers have gone independently to the travel agency for a reservation, determine the probability that:

- a. five buy a ticket;
- b. everyone buys a ticket;
- c. at least 2 buy a ticket.

Solution

b. The probability that all customers buy a ticket is:

$$\begin{aligned} P(X = 8) &= \binom{n}{x} p^x \cdot (1 - p)^{(n - x)} = \binom{8}{8} 0.60^8 \cdot (1 - 0.60)^{(8 - 8)} = \\ &= 8!/(8! \cdot 0!) \cdot 0.0168 \cdot 1 = 1 \cdot 0.0168 = 1.68\% \end{aligned}$$

c. The probability that at least two customers buy a ticket, $P(X \geq 2)$, is:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = \\ &= 1 - P(X=1) - P(X=0) = \\ &= 1 - \binom{8}{1} 0.60^1 \cdot (1 - 0.60)^{(8 - 1)} - \binom{8}{0} 0.60^0 \cdot (1 - 0.60)^{(8 - 0)} = \\ &= 1 - 8!/(7! \cdot 1!) \cdot 0.6 \cdot 0.4^7 - 8!/(0!8!) \cdot 0.6^0 \cdot 0.4^8 = \\ &= 1 - 8 \cdot 0.6 \cdot 0.0016 - 1 \cdot 1 \cdot 0.0006 = 1 - 0.0077 - 0.0006 = 0.9917 = 99.17\% \end{aligned}$$

From the complementary statement of probability

Poisson Probability Distribution

Represents the number of occurrences in a given interval. Occurrences have to be random and independent

Probability function: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

with λ as average number of occurrences in a given interval and x number of occurrences (in the same interval)

$$E(X) = V(X) = \lambda$$

ex. Number of telemarketing phone calls received in a day

Exercise 6

On average a household receives 3 telemarketing phone calls per month. Find the probability that a randomly selected household receives:

- exactly 2 calls during a given month;
- exactly 1 call during a given week (assume 4 weeks each month);
- Find the expected value

Solution

To calculate the points a. b. and c. we have to use the Poisson probability distribution

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where λ is the number of telemarketing calls received per month ($\lambda = 3$)

- We have to calculate $P(X = 2)$.

Using the Poisson distribution, $P(X = 2) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{3^2 e^{-3}}{2!} = 9 \cdot 0.049 / 2 = 0.2205 = 22.5\%$

- If the calls per month are 3, the calls per week are:

$$\lambda = 3 \text{ calls per month} / 4 \text{ weeks} = 0.75 \text{ calls per week}$$

So, $P(X = 1) = 0.75^1 \cdot e^{(-0.75)} / 1! = 0.75 \cdot 0.4724 = 0.3543 = 35.43\%$

- For the Poisson distribution: $E(X) = V(X) = \lambda = 3$

Hypergeometric Distribution

The hypergeometric distribution is very similar to the binomial distribution.

Represents the number of successes in n not independent Bernoulli experiments

Values (integers): $0, 1, 2, \dots, n$

Probability function:

$$P(X = x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

with:

N = population size,

n = sample size,

r = nr successes in population,

x = nr successes in sample

Exercise 8

A firm has 12 employees, 7 females and 5 males. The company is planning to send 3 of them to a conference. Find the probability that all 3 of them are female.

Solution

Single employee is a Bernoulli experiment (1 if female, 0 otherwise), but in this case the event is not independent (each employee is different)

The hypergeometric distribution is:

$$P(X = x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

$N=12, n=3, r=7, x=3$

$$P(X = x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{7}{3} \binom{12-7}{3-3}}{\binom{12}{3}} = \frac{7}{44} = 0.1591$$

Recap on Discrete Random Variables

For Binomial and Bernoulli
sometimes you can find
q instead of (1-p)

Bernoulli Probability Distribution

$$P(x_i) = p^{x_i} (1 - p)^{(1-x_i)} \quad E(X) = p \quad \text{and} \quad \text{Var}(X) = p \cdot (1 - p)$$

Binomial Probability Distribution

$$P(x) = \binom{n}{x} p^x (1 - p)^{(n-x)}$$

with n as total number of occurrences and x number of selected occurrences

$$E(X) = n \cdot p \quad \text{and} \quad \text{Var}(X) = n \cdot p \cdot (1 - p)$$

Poisson Probability Distribution

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

with λ as average number of occurrences in a given interval and x number of occurrences

$$E(X) = \text{Var}(X) = \lambda$$

Exercise 9

For each point, indicate first what probability distribution is used to solve the problem.

- a. 60% of people who purchase sports cars are men. If 10 sports car owners are randomly selected, find the probability that exactly 7 are men.
- b. The evaporation of water for a pool in California is (on average) 2 centimeters per month (30 days). Find the probability that the pool has lost 5 centimeters of water in August.
- c. 1% of the product sold by Apple have some bugs. Find the probability that for 100 products sold at most 1 have bugs.
- d. A company receives on average 3 returns of defective product per month. Find the probability that a randomly selected product sold in October could be defective.

Exercise 9

Solution

- a. 60% of people who purchase sports cars are men. If 10 sports car owners are randomly selected, find the probability that exactly 7 are men.

- a. We have to use the Binomial Distribution because the possible outcome are only 2: men who **PURCHASE** sports car (event 1) and men who **NOT PURCHASE** sports car (event 0).

Identify 'n' and 'x' from the problem.

n is the number of randomly selected items, that in this case is sports car owners, and is 10, and x, the number of men you are asked to “find the probability” for, that is 7

Find “p” the probability of success and “1-p” or “q” the probability of failure.

We are given p = 60%, or 0.6. therefore, the probability of failure is $1 - 0.6 = 0.4$ (40%)

$$P(X=7) = \frac{n!}{(n-x)! x!} \times p^x \times (1-p)^{n-x} = \frac{10!}{3! 7!} \times 0.6^7 \times 0.4^3 = 120 \times 0.0279936 \times 0.064 = 0.215 = 21.5\%$$

- b. The evaporation of water for a pool in California is (on average) 2 centimeters per month (30 days).

Find the probability that the pool has lost 5 centimeters of water in August.

- b. We have to use the Poisson Distribution because we have number of occurrence (cm of evaporation) in a given interval of time (a month).

Identify 'λ' and 'x' from the problem.

λ=2 (evaporation per month) and x=5 (evaporation in August)

$$P(X = 5) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{2^5 \times e^{-2}}{5!} = 0.3609 = 3.6\%$$

Exercise 9

Solution

- c. 1% of the product sold by Apple have some bugs. Find the probability that for 100 products sold at most 1 have bugs.
- c. We have to use the Binomial Distribution because the possible outcome are only 2: the product **has a bug** (event 1) and the product **has NOT a bug** (event 0). Identify 'n' and 'x' from the problem.

$$n=100 \quad x=1$$

The probability are: $p = 0.01$ and $(1-p)=0.99$

$$P(X=1) = \frac{n!}{(n-x)! x!} \times p^x \times (1-p)^{n-x} = \frac{100!}{99! 1!} \times 0.01^1 \times 0.99^{99} = 100 \times 0.01 \times 0.3697 = 0.3697 = 36.97\%$$

- d. A company receives on average 3 returns of defective product per month. Find the probability that a randomly selected product sold in October could be defective.
- d. We have to use the Poisson Distribution because we have number of occurrence (return of products) in a given interval of time (a month). Identify 'λ' and 'x' from the problem.

$\lambda=3$ (return products per month) and $x=1$ (defective product sold in October)

$$P(X = 1) = \frac{\lambda^x e^{-\lambda}}{x!} = 3^1 \times e^{-3} / 1! = 3 \times 0,049 / 1 = 0,147 = 14,7\%$$

Exercise 10

On average a household receives 3 telemarketing phone calls per month. Find the probability that a randomly selected household receives:

- a. exactly 2 calls during a given month;
- b. exactly 1 call during a given week (assume 4 weeks each month);
- c. Find the expected value

Solution

To calculate the points a. b. and c. we have to use the Poisson probability distribution

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where λ is the number of telemarketing calls received per month ($\lambda = 3$)

- a. We have to calculate $P(X = 2)$.

Using the Poisson distribution, $P(X = 2) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{3^2 e^{-3}}{2!} = 9 \cdot 0.049 / 2 = 0.2205 = 22.5\%$

- b. If the calls per month are 3, the calls per week are:

$$\lambda = 3 \text{ calls per month} / 4 \text{ weeks} = 0.75 \text{ calls per week}$$

So, $P(X = 1) = 0.75^1 \cdot e^{(-0.75)} / 1! = 0.75 \cdot 0.4724 = 0.3543 = 35.43\%$

- c. For the Poisson distribution: $E(X) = \text{Var}(X) = \lambda = 3$