Quantitative Methods – I (Statistics)

A. Y. 2023-24

Prof. Marco Stefanucci

Chapter 5
Discrete Random Variables

Discrete Random Variables: Outline

1. Random variables: definition, E(X), V(X)

2. Probability Distribution

3. Discrete random variables:
   a) Bernoulli
   b) Binomial
   c) Hypergeometric
   d) Poisson

Random Variable

A **random variable** is a numerical quantity that is generated by a random experiment.

We will denote random variables by capital letters, such as $\boldsymbol{X}$ or $\boldsymbol{Z}$, and the actual values that they can take by lowercase letters, such as $\boldsymbol{x}$ and $\boldsymbol{z}$.

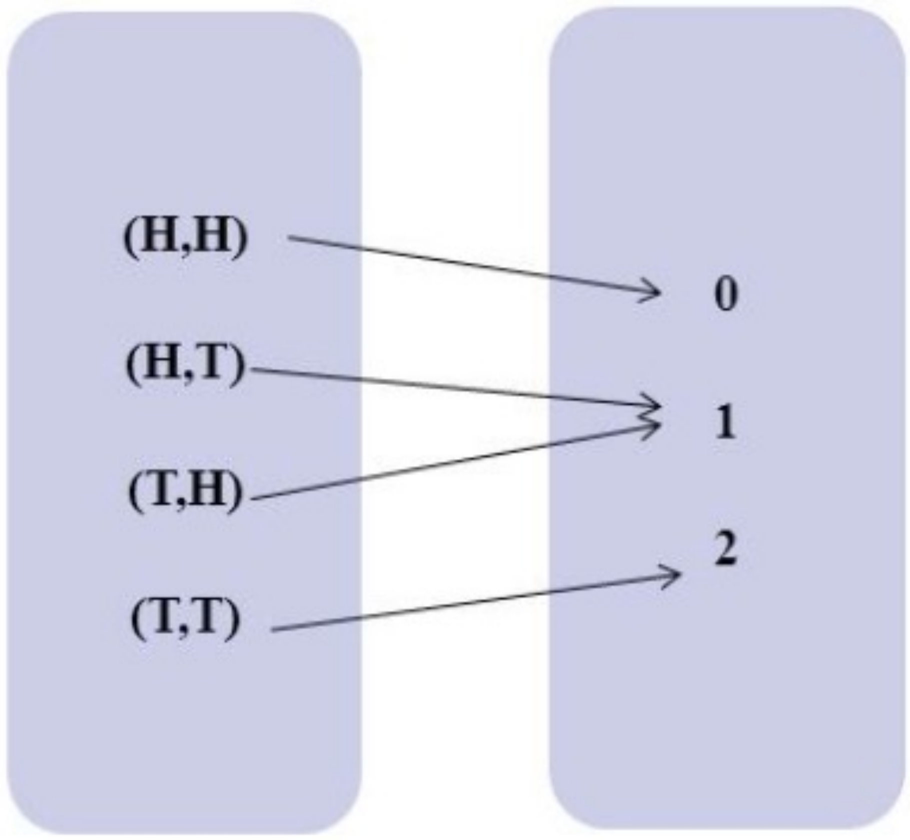
Random Variable

Experiment	Number X	Possible Values of X
Roll two fair dice	Sum of the number of dots on the top faces	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Flip a fair coin repeatedly	Number of tosses until the coin lands heads	1, 2, 3,4, ...
Measure the voltage at an electrical outlet	Voltage measured	$118 \le x \le 122$
Operate a light bulb until it burns out	Time until the bulb burns out	$0 \le x < \infty$

Random Variable: example

<u>A variable (denoted by capital letter, e.g. X) whose value is determined by the outcome of an experiment.</u>

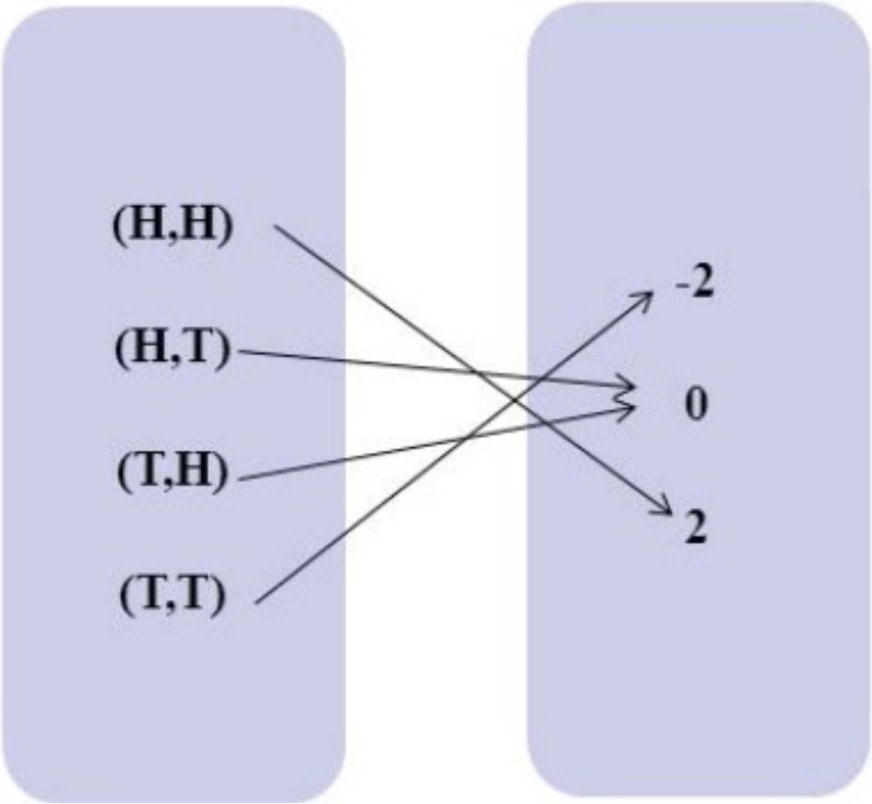
Example: consider the experiment of tossing a coin twice, and denote with X the variable counting the total number of Tails you get



Random Variable: example

<u>A variable (denoted by capital letter, e.g. X) whose value is determined by the outcome of an experiment.</u>

Example: consider the experiment of tossing a coin twice, and denote with X the variable giving the difference between the number of Heads minus the number of Tails



Random Variable: Discrete and Continuous

Discrete Random Variable

A random variable is called discrete if it has either a finite or a countable number of possible values.

Continuous Random Variable

A random variable is called continuous if its possible values contain a whole interval of numbers.

Random Variable: Discrete

A variable (denoted by capital letter, e.g. X) whose value is determined by the outcome of an experiment.

If only integer values are possible ➔ discrete random variable

If any value in given intervals ➔ continuous random variable

Examples DISCRETE: the number of…

- Heads obtained in 3 tosses of a coin
- Dots rolling a die
- Cars held by a household
- Customers who visit a bank during any given hour

Random Variable: Continuous

<u>A variable (denoted by capital letter, e.g. X) whose value is determined by the outcome of an experiment.</u>

If only integer values are possible ➔ discrete random variable

If any value in given intervals ➔ continuous random variable

Examples CONTINUOUS:

- length of a room, weight of a letter, height of a person
- time taken to commute from home to work, life of a battery
- Monetary amounts (technically discrete, but large number of unique values).

Probability distributions

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

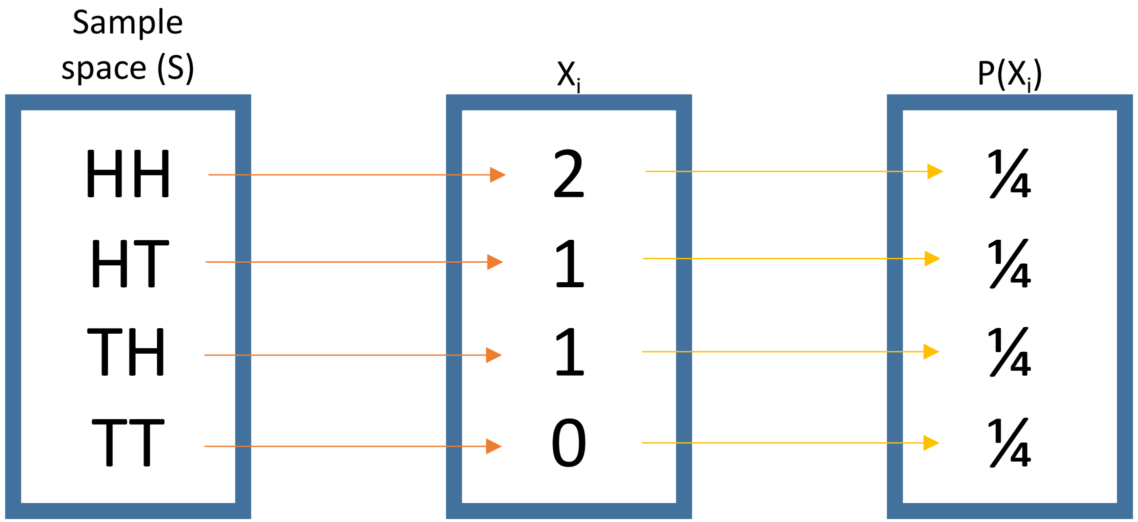
Ex. Number of Head from the toss of two coins

Sample space (S)	Random Variable (X_i)	Probability $P(X_i)$
HH	2	¼
HT	1	¼
TH	1	¼
TT	0	¼

Probability distributions

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

Ex. Number of Head from the toss of two coins



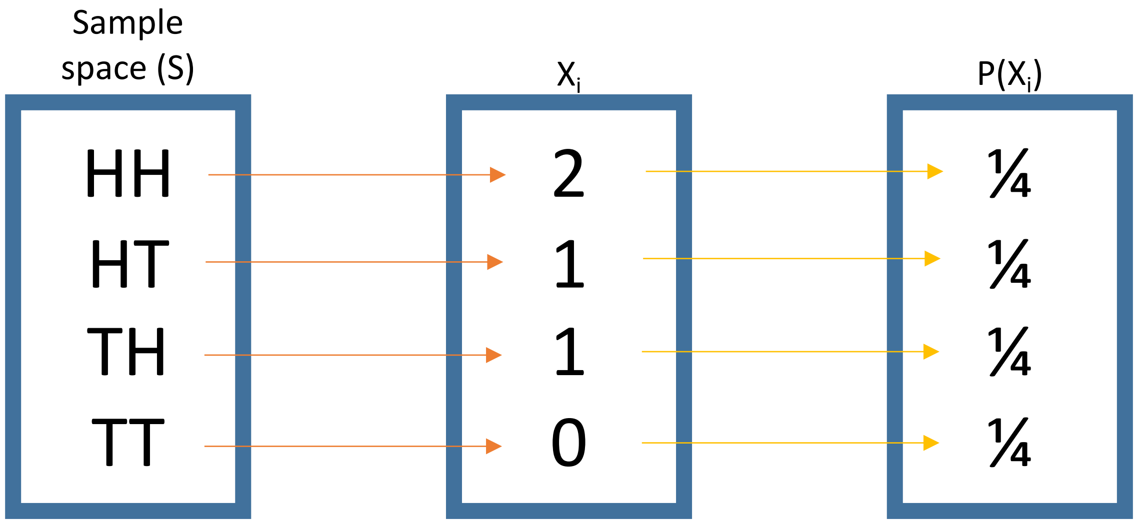
X_i	$P(X_i)$
0	¼ = 0.25
1	¼ + ¼ = 2/4 = 0.5
2	¼ = 0.25
Tot.	1.0

Probability distributions

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

2 conditions:

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



X_i	$P(X_i)$
0	¼ = 0.25
1	¼ + ¼ = 2/4 = 0.5
2	¼ = 0.25
Tot.	1.0

Probability distributions

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

2 characteristics:

1) For each value of X, $0 \leq P(X_i) \leq 1$
2) $\sum_{i=1}^{n} P(X_i) = 1$

Probability distributions

Only one of the following is a probability distribution. Which one?

(a)

x	$P(x)$
0	.08
1	.11
2	.39
3	.27

$\sum_{i=1}^{n} P(X_i) \neq 1$

(b)

x	$P(x)$
2	.25
3	.34
4	.28
5	.13

$0 \leq P(X_i) \leq 1$

$\sum_{i=1}^{n} P(X_i) = 1$

(c)

x	$P(x)$
7	.70
8	.50
9	−.20

$P(X_i) \leq 0$

Probability distributions: example

Ex.: 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 Sample Space of 8 possible outcomes of the experiment is:

$$\Omega=\{HHH, HHT, HTH , HTT, THH, THT, TTH, TTT\}$$

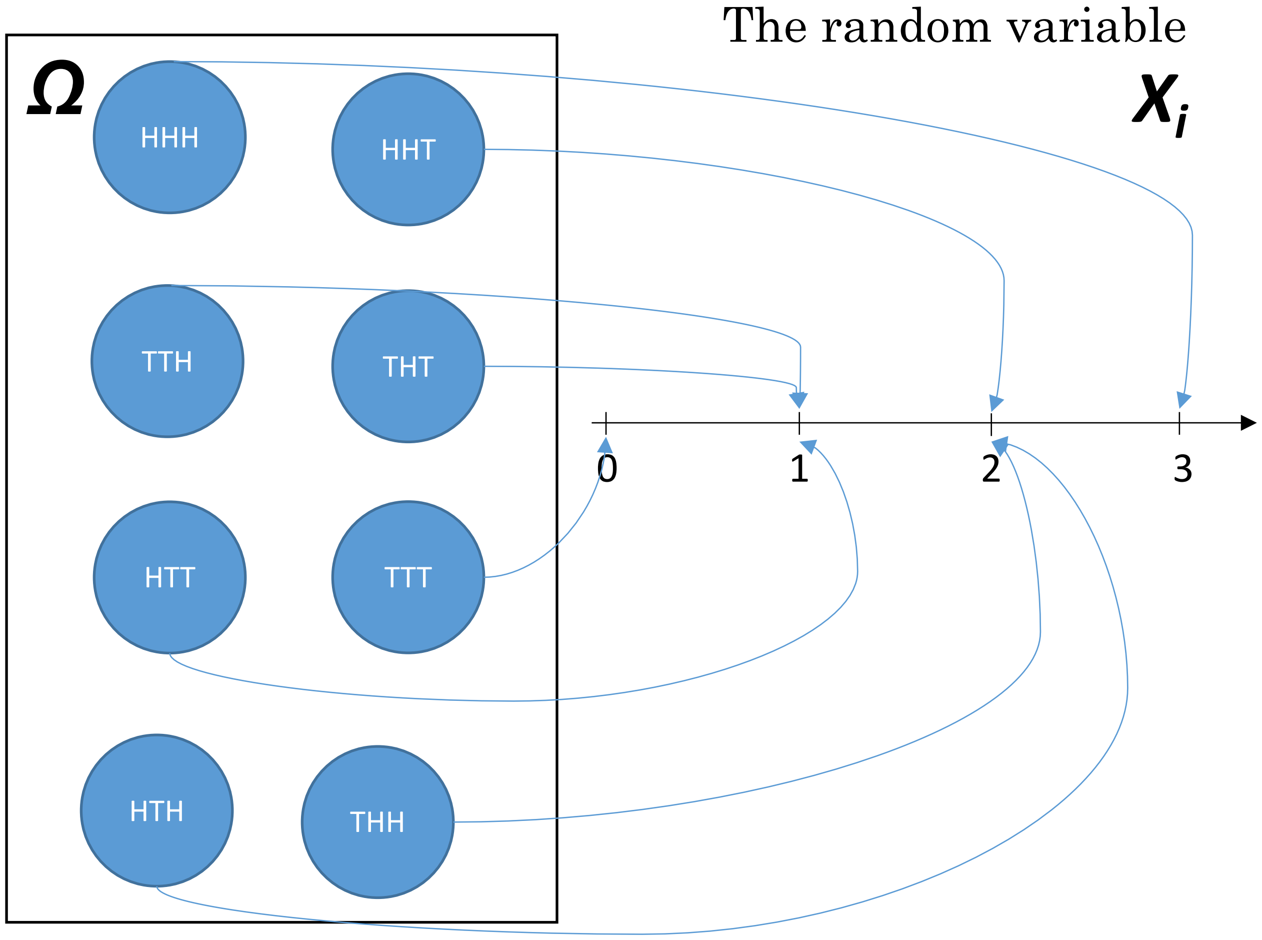
We call each outcome w_i

Probability distributions: example

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



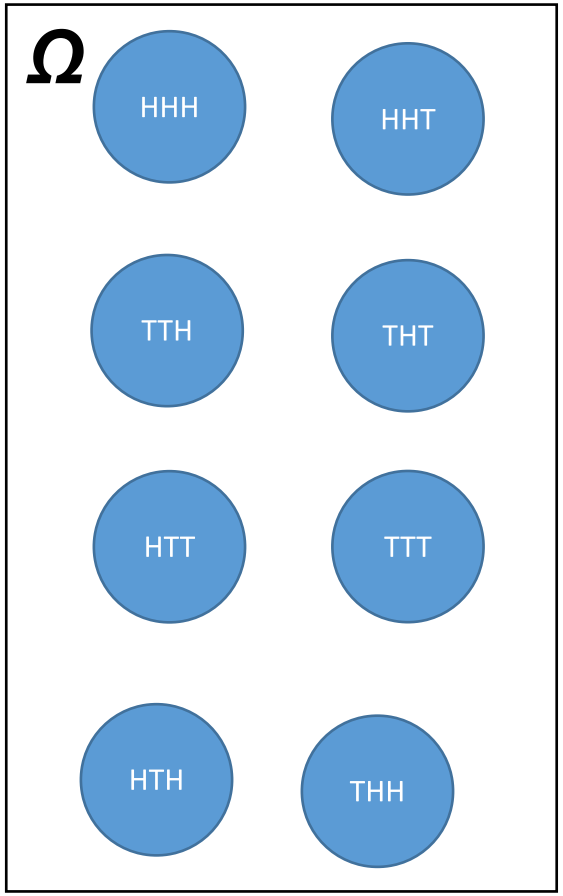
Probability distributions: example

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

X_i	$P(X = X_i)$ or $P(X_i)$
0	$\frac{1}{8}$
1	$\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$
2	$\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$
3	$\frac{1}{8}$



Probability distributions: example

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

Solution

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

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

Probability distributions: example

Ex: 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;

Solution

From this probability distribution we can calculate the probability from 1. to 3.:

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

Probability distributions: example

Ex: 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 :

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 4. to 6.:

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

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 $P(X < 3) = 1 - P(X = 3)$ [2^{nd} condition of prob. distr.] $= 1 - 1/8 = 7/8$
6. $P(X \leq 1) = P(X = 0) + P(X = 1) = 1/8 + 3/8 = 4/8 = 1/2$

Probability distributions: example

A pair of fair dice is rolled. Let X denote the sum of the number of dots on the top faces.

a. Construct the probability distribution of X.
b. Find $P(X \geq 9)$.
c. Find the probability that X takes an even value.

Probability distributions: example

Solution:
The sample space Ω of equally likely outcomes is

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

a. The possible values for X are the numbers 2 through 12.
$X = 2$ is the event $\{11\}$, so $P(2) = 1/36$.
$X = 3$ is the event $\{12, 21\}$, so $P(3) = 2/36$.
Continuing this way we obtain the table

x	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

This table is the probability distribution of X.

Probability distributions: example

Solution:

x	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b. The event $X \geq 9$ is the union of the mutually exclusive events $X = 9$, $X = 10$, $X = 11$, and $X = 12$. Thus

$$P\,(X \geq 9) = P\,(9) + P\,(10) + P\,(11) + P\,(12) = \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36}$$

Probability distributions: example

Solution:

x	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

c. Before we immediately jump to the conclusion that the probability that X takes an even value must be 0.5, note that X takes six different even values but only five different odd values.

$$P\,(X\ is\ even) = P\,(2) + P\,(4) + P\,(6) + P\,(8) + P\,(10) + P\,(12) =$$
$$= 1/36 + 3/36 + 5/36 + 5/36 + 3/36 + 1/36 = 18/36 = 0.5$$

Probability distributions: example

The probability that a randomly selected family holds 2 vehicles is…

$P(X = 2) = 0.425$

Number of Vehicles Owned x	Probability $P(x)$
0	.015
1	.235
2	.425
3	.245
4	.080
	$\Sigma P(x) = 1.000$

Probability distributions: example

The probability that a randomly selected family holds at most 1 vehicle is…

$P(X \leq 1) =$

$P(X = 0) + P(X = 1) =$

$0.015 + 0.235 = 0.25$

Number of Vehicles Owned x	Probability $P(x)$
0	.015
1	.235
2	.425
3	.245
4	.080
	$\Sigma P(x) = 1.000$

Probability distributions: example

The probability that a randomly selected family holds at least 2 vehicles is…

$P(X \geq 2) =$
$1 - P(X \leq 1) = 0.75$

Number of Vehicles Owned x	Probability $P(x)$
0	.015
1	.235
2	.425
3	.245
4	.080
	$\Sigma P(x) = 1.000$

Probability distributions: example

The probability that a randomly selected family holds 3 or more vehicles is…

$P(X \geq 3) =$
$P(X = 3) + P(X = 4) =$
$0.245 + 0.08 = 0.325$

Number of Vehicles Owned x	Probability $P(x)$
0	.015
1	.235
2	.425
3	.245
4	.080
	$\Sigma P(x) = 1.000$

Random Variable and Probability Distribution

Random Variable

It is a numerical quantity that is generated by a random experiment.

Probability Distribution

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

2 characteristics:

- $0 \leq P(X_i) \leq 1$
- $\sum_{i=1}^{n} P(X_i) = 1$

Cumulative Distribution Function (CDF)

Given a value x_0 the CDF gives the probability that X assumes values equal or lower than x_0

$$F(x_0) = P(X \leq x_0)$$

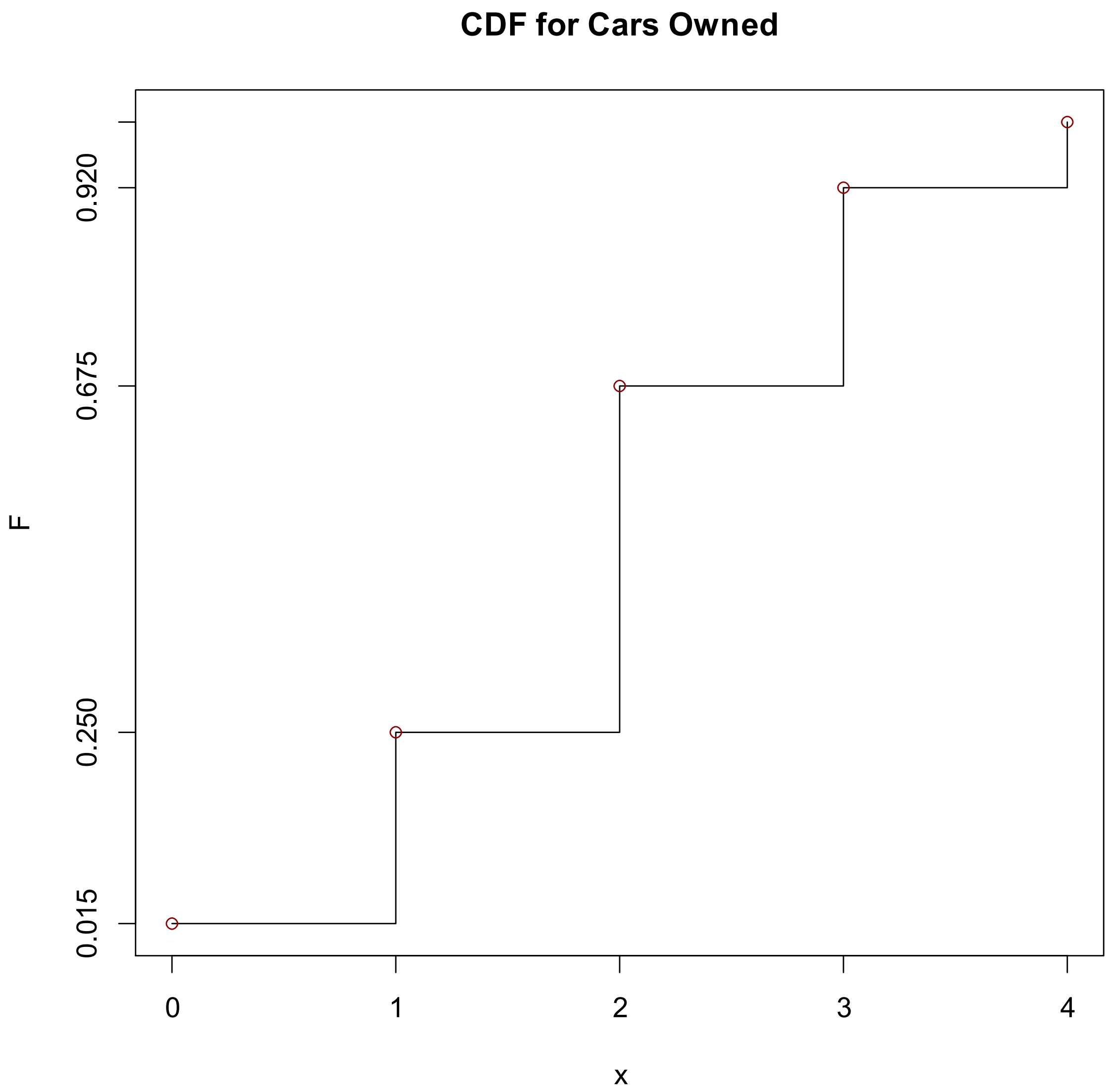
$\boldsymbol{x_i}$	$\boldsymbol{P(x_i)}$	$\mathbf{F}(\boldsymbol{x_i})$
0	0.015	0.015
1	0.235	$0.015 + 0.235 = 0.250$
2	0.425	$0.250 + 0.425 = 0.675$
3	0.245	$0.675 + 0.245 = 0.920$
4	0.080	$0.920 + 0.080 = \mathbf{1.000}$
Total	$\mathbf{1.000}$	

CDF: graphical representation

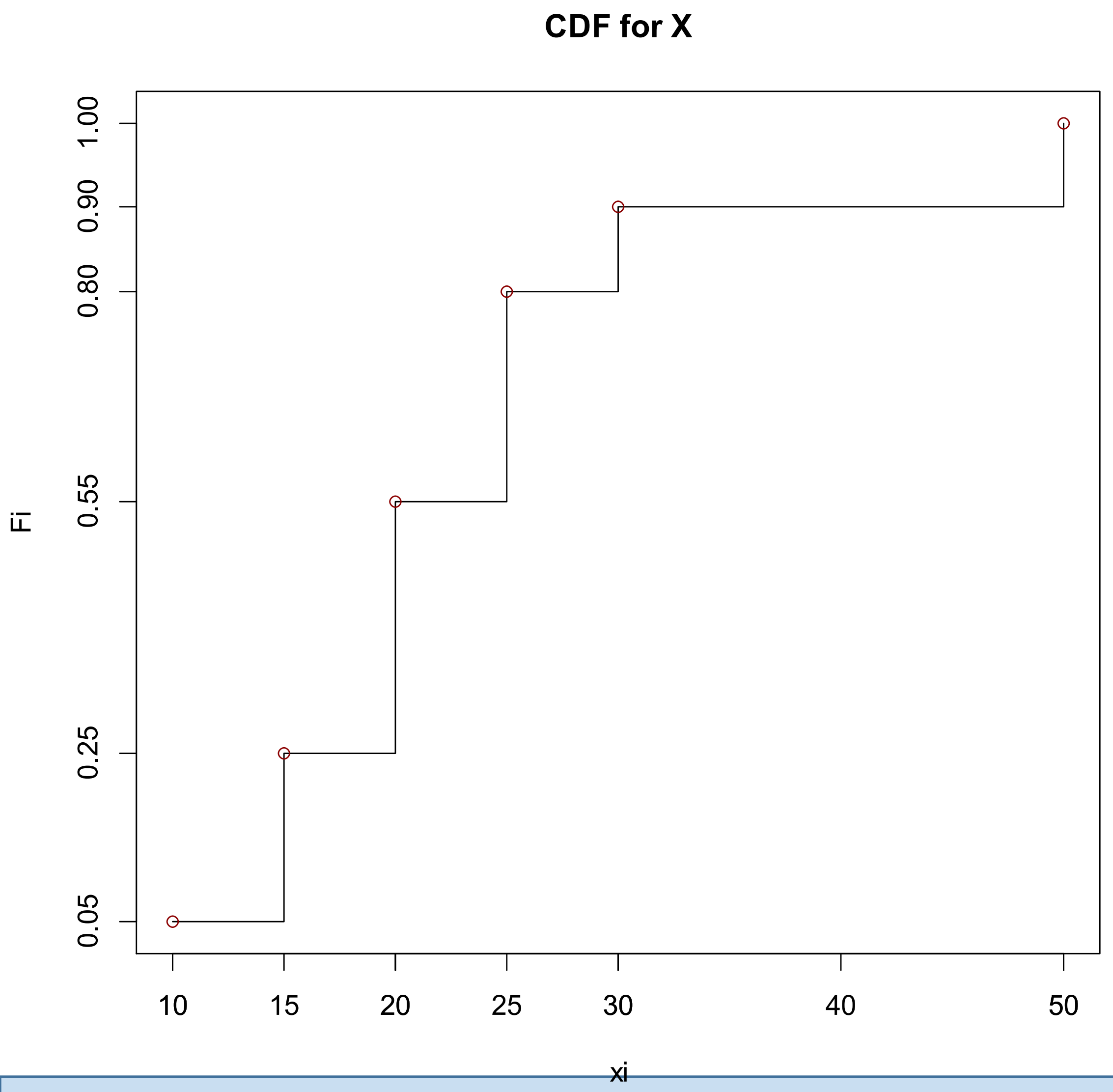
x_i	$P(x_i)$	$\mathrm{F}(x_i)$
0	0.015	0.015
1	0.235	0.250
2	0.425	0.675
3	0.245	0.920
4	0.080	**1.000**
Total	**1.000**	

A stepped line graph or step chart is a chart similar to a line graph, but with the line forming a series of steps between data points.

It is useful when you want to show the changes that occur at irregular intervals.



From CDF to probability distribution



➔

x_i	$F(x_i)$
10	0.05
15	0.25
20	0.55
25	0.80
30	0.90
50	1.00
Total	

Reverse path ➔

$P(x_i)$
0.05
0.20
0.30
0.25
0.10
0.10
1

Expected Value, Variance and Standard Dev.

Expected Value

E(X) is the mean of the probability distribution

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

Variance

They measure the spread of the probability distribution:

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

For discrete, $V(X) = \sigma^2 = \Sigma x_i^2 \cdot P(x_i) - \mu^2$

Standard Deviation

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

Expected Value

E(X) is the mean of the probability distribution

$$E(X) = \mu = \sum_{i=1}^{k} x_i P(x_i)$$

$\boldsymbol{x_i}$	$\boldsymbol{P(x_i)}$	$\boldsymbol{x_i P(X_i)}$
0	0.015	0.000
1	0.235	0.235
2	0.425	0.850
3	0.245	0.735
4	0.080	0.320
Total	$\boldsymbol{1.000}$	$\boldsymbol{2.140}$

$$\rightarrow \mu = E(X) = 2.14$$

Variance and Standard Deviation

They measure the spread of the probability distribution:

- **Variance, V(X)**

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

For discrete➔$\sigma^2 = \sum_{i=1}^{k}(x_i - E(X))^2 P(x_i)$

$$= \sum_{i=1}^{k} x_i^2 P(x_i) - E(X)^2$$

- **Standard Deviation, SD(X)**

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

σ roughly describes how far the variable falls on average from E(X)

Variance for discrete random variables: proof

$$\sigma^2 = \sum_{i=1}^{k} (x_i - E(X))^2 P(x_i) =$$

$$= \sum_{i=1}^{k} ({x_i}^2 + E(X)^2 - 2x_i E(X)) P(x_i) =$$

$$= \sum_{i=1}^{k} {x_i}^2 P(x_i) + E(X)^2 \sum_{i=1}^{k} P(x_i) - 2E(X) \sum_{i=1}^{k} x_i P(x_i) =$$

$$= \sum_{i=1}^{k} {x_i}^2 P(x_i) - E(X)^2$$

Variance and Standard Deviation: Example

x	$P(x)$	$(x_i - E(x))^2 P(x)$	$x^2 P(x)$
0	0.015	0.069	0.000
1	0.235	0.305	0.235
2	0.425	0.008	1.700
3	0.245	0.181	2.205
4	0.080	0.277	1.280
Total	**1.000**	**0.840**	**5.420**

$$V(X) = \sum_{i=1}^{k} (x_i - E(X))^2 P(x_i) = 0.84$$

$$V(X) = \sum_{i=1}^{k} x_i^2 P(x_i) - E(X)^2 = 5.42 - 2.14^2 = 0.84$$

$$SD(X) = \sqrt{0.84} = 0.917$$

Discrete Probability Distribution: Bernoulli

Bernoulli Probability Distribution

Only 2 values (1="success", 0 = "failure"). 1 with probability p, 0 with probability $(1 - p)$ (sometimes instead of $(1\text{-}p)$ you might find q)

Probability function:

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

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

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

Discrete Probability Distribution: Binomial

Binomial Random Variable

Represents the number of successes in n Bernoulli experiments:

a) independent
b) with equal p

Probability function:

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

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

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

Discrete Probability Distribution: Poisson

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 (or μ) 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

Discrete Probability Distribution: Hypergeometric

Hypergeometric Distribution

Represents the number of successes in n **not independent** Bernoulli experiments

Probability function:

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

with:
N = population size,
n = sample size (integer numbers),
r = nr. successes in population,
x = nr. successes in sample

Bernoulli Random Variable

Bernoulli trial: random experiment generating a sample space consisting of only two elementary outcomes, success (A) or failure ($\bar{A}$).

$$\Omega = (A; \bar{A})$$

A Bernoulli random variable X takes value 1 if the event success occurs and zero otherwise.

Let $p = P(A)$ denote the success probability and $(1 - p)$ or $q = P(\bar{A})$.

Bernoulli Probability Distribution

Only 2 values (1="success", 0 = "failure").

1 with probability p

0 with probability $(1-p)$ or q ➔

x	P(x)
0	1-p
1	p
Total	1

Probability function:

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

➔ $E(X) = p$

➔ $V(X) = p\times(1-p)$

Ex.: *a team will win a championship or not, a student will pass or fail an exam, a rolled dice will either show a 6 or any other number, etc.*

Bernoulli Probability Distribution: example

A gym membership is renewed with probability 0.7. 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).

x	P(x)
0	0.3
1	0.7
Total	1

1 (success)= renew

0 (failure) = not renewed

➔ $E(X) = 0.7$

➔ $V(X) = 0.7\times(0.3) = 0.21$

Binomial Trial

Binomial experiments are random experiments that consist of a fixed number of repeated trials, like tossing a coin 10 times, randomly choosing 10 people, rolling a die 5 times, etc.

These trials need to be independent in the sense that the outcome in one trial has no effect on the outcome in other trials.

In each of these repeated trials there is one outcome that is of interest to us (we call this outcome "success"), and each of the trials is identical in the sense that the probability that the trial will end in a "success" is the same in each of the trials.

So for example, if our experiment is tossing a coin 10 times, and we are interested in the outcome "heads" (our "success"), then this will be a binomial experiment, since the 10 trials are independent, and the probability of success is ½ in each of the 10 trials.

Binomial Trial

The requirements to be a binomial experiment are:

- a fixed number (n) of trials
- each trial must be independent of the others
- each trial has just two possible outcomes, called "success" (the outcome of interest) and "failure"
- there is a constant probability (p) of success for each trial, the complement of which is the probability (1 – p) of failure, sometimes denoted as q = (1 – p)

In binomial random experiments, the number of successes in n trials is random.

Binomial Random Variable

The random variable **X** that represents the number of successes in n **Binomial Trials** is called a **Binomial Random Variable**, and is determined by the values of n and p.

We say, "X is binomial with n = … and p = …"

"The Binomial Random Variable **X ~ B(n; p)** represents the number of successes in a sequence of n independent Bernoulli trials, all occurring with the same probability p of success."

X takes the values 0, 1, …, n.

Represents the number of successes in *n* Bernoulli experiments:

a) independent

b) with equal p

Random Experiments (Binomial or Not?)

Let's consider a few random experiments.

In each of them, we'll decide whether the random variable is binomial. If it is, we'll determine the values for n and p. If it isn't, we'll explain why not.

Ex. 1: A fair coin is flipped 20 times. X represents the number of heads.

X is binomial with $n = 20$ and $p = 0.5$

Ex. 2: You roll a fair die 50 times. X is the number of times you get a six.

X is binomial with $n = 50$ and $p = 1/6$

Random Experiments (Binomial or Not?)

Ex. 3: Roll a fair die repeatedly. X is the number of rolls it takes to get a six.

X is not binomial, because the number of trials is not fixed.

Ex. 4: Draw 3 cards at random, one after the other, without replacement, from a set of 4 cards consisting of one club, one diamond, one heart, and one spade. X is the number of diamonds selected.

X is not binomial, because the selections are not independent.

The probability (p) of success is not constant, because it is affected by previous selections.

Random Experiments (Binomial or Not?)

Ex. 5: Draw 3 cards at random, one after the other, with replacement, from a set of 4 cards consisting of one club, one diamond, one heart, and one spade. X is the number of diamonds selected.

Sampling with replacement ensures independence.

X is binomial with $n = 3$ and $p = ¼$

Ex. 6: Approximately 1 in every 20 children has a certain disease. Let X be the number of children with the disease out of a random sample of 100 children. Although the children are sampled without replacement, it is assumed that we are sampling from such a vast population that the selections are virtually independent.

X is binomial with $n = 100$ and $p = 1/20 = 0.05$

Random Experiments (Binomial or Not?)

Ex. 7: The probability of having blood type B is 0.1. Choose 4 people at random. X is the number with blood type B.

X is binomial with $n = 4$ and $p = 0.1$

Ex. 8: A student answers 10 quiz questions completely at random; the first five are true/false, the second five are multiple choice, with four options each. X represents the number of correct answers.

X is not binomial, because p changes from ½ to ¼

Binomial Probability Distribution

Represents the number of successes in n ***<u>independent</u>*** Bernoulli experiments with equal p (and q $= 1 - p$)

Values (integers): 0, 1, 2,…,n

Probability function:

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

With:

n = nr of trials, p = probability of success, x= number of successes

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

Binomial Variable: example

1. A six-sided die is rolled 12 times. What is the probability of getting a 4 five times?

Binomial Variable: example

1. A six-sided die is rolled 12 times. What is the probability of getting a 4 five times?

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

$n =$

$x =$

$p =$

$q =$

Binomial Variable: example

1. A six-sided die is rolled 12 times. What is the probability of getting a 4 five times?

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

$n = 12$

$x = 5$

$p = 1/6$

$q = 5/6$

Binomial Variable: example

1. A six-sided die is rolled 12 times. What is the probability of getting a 4 five times?

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

$n = 12$ $\quad p = 1/6$

$x = 5$ $\quad q = 5/6$

$$\binom{n}{r} = {}_nC_r = \frac{n!}{(n-r)!\,r!} = \binom{12}{5} = {}_{12}C_5 = \frac{12!}{(12-5)!\,5!}$$

$$= 792$$

Binomial Variable: example

1. A six-sided die is rolled 12 times. What is the probability of getting a 4 five times?

$$P(x) = \binom{n}{x} p^x q^{n-x} \qquad n = 12 \quad x = 5 \qquad p = 1/6 \quad q = 5/6$$

$$P(5) = 792\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^{12-5}$$

$$= 792\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^7 = 0.0284$$

Binomial Variable: example

2. A multiple choice test contains 20 questions with answer choices A, B, C, and D. Only one answer choice to each question represents a correct answer. Find the probability that a student will answer exactly 6 questions correct if he makes random guesses on all 20 questions.

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

$n = 20$ $\quad$ $p = 0.25$

$x = 6$ $\quad$ $q = 0.75$

Binomial Variable: example

2. A multiple choice test contains 20 questions with answer choices A, B, C, and D. Only one answer choice to each question represents a correct answer. Find the probability that a student will answer exactly 6 questions correct if he makes random guesses on all 20 questions.

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

$n = 20$ $\quad p = 0.25$

$x = 6$ $\quad q = 0.75$

$${}_{20}C_6 = \frac{20!}{(20-6)!\,6!} = \frac{20!}{14!\,6!} =$$

$$= \frac{\cancel{20} \times 19 \times \cancel{18} \times 17 \times \cancel{16} \times 15 \times \cancel{14!}}{\cancel{14!} \cdot \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times 1} = \boxed{38,760}$$

Binomial Variable: example

2. A multiple choice test contains 20 questions with answer choices A, B, C, and D. Only one answer choice to each question represents a correct answer. Find the probability that a student will answer exactly 6 questions correct if he makes random guesses on all 20 questions.

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

$n = 20$ $\quad$ $x = 6$ $\quad$ $p = 0.25$ $\quad$ $q = 0.75$

$$P(6) = 38760(0.25)^6(0.75)^{14} = 0.168609$$

$$= 16.86\%$$

Binomial Probability Distribution

The random variable **X** that represents the number of successes in n **Binomial Trials** is called a **Binomial Random Variable**, and is determined by the values of n and p.

Represents the number of successes in ***n*** Bernoulli experiments:

a) independent
b) with equal ***p***

Probability function:

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

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

Binomial Variable: example

5% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players.

1) What is the distribution of X?

X is a binomial variable, with $n = 3, p = 0.05$

2) What are all the possible values of X?

All the integers between 0 and n ➔ from 0 to 3

Binomial Variable: example

5% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players.

3) What is the probability that none of the DVD is defective?

$$P(X=0)=\binom{3}{0}0.05^{0}(0.95)^{3-0}=\frac{3\times2\times1}{1\times3\times2\times1}1(0.95)^{3}=0.8574$$

4) What is the probability that only one of the DVD is defective?

$$P(X=1)=\binom{3}{1}0.05^{1}(0.95)^{3-1}=0.1354$$

Binomial Variable: example

5% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

$$X = (0, 1, 2, 3)$$

$$P(X=0) = \binom{3}{0} 0.05^0 (0.95)^{3-0} = \frac{3 \times 2 \times 1}{1 \times 3 \times 2 \times 1} 1(0.95)^3 = 0.8574$$

$$P(X=1) = \binom{3}{1} 0.05^1 (0.95)^{3-1} = 0.1354$$

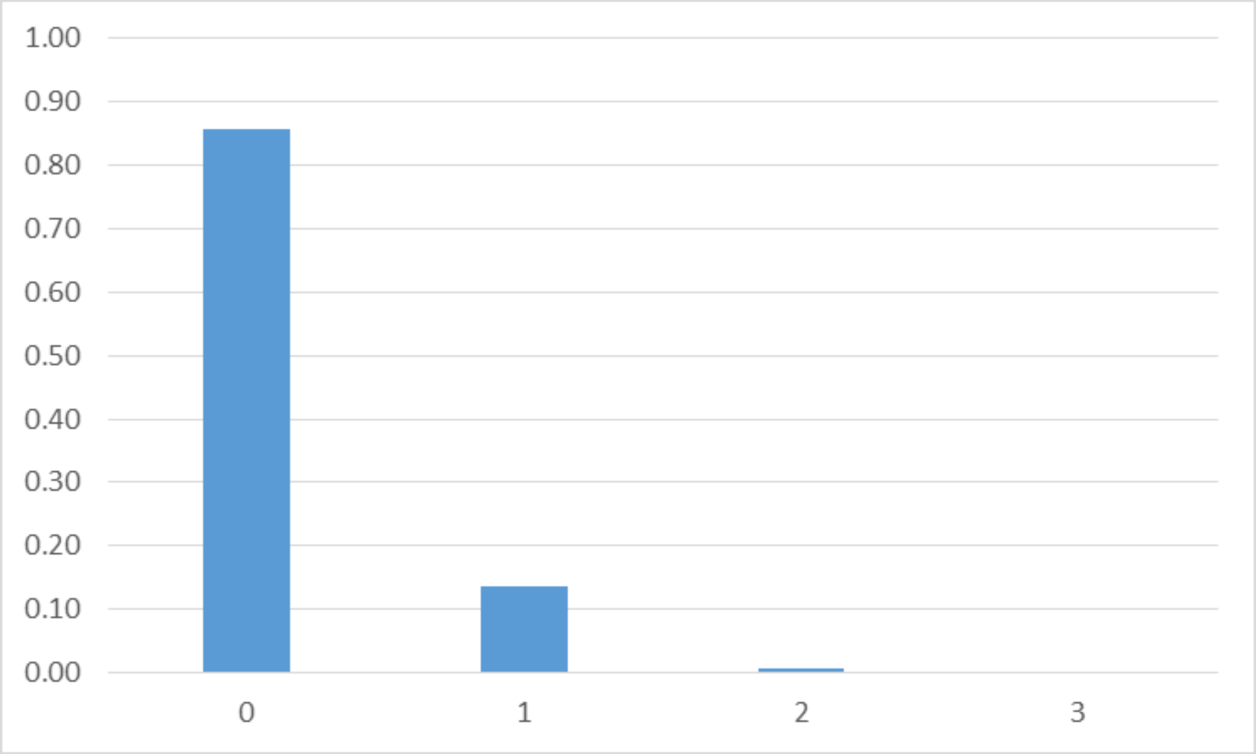
$$P(X=2) = \binom{3}{2} 0.05^2 (0.95)^{3-2} = 0.0071$$

$$P(X=3) = \binom{3}{3} 0.05^3 (0.95)^0 = 0.0001$$

Binomial Variable: example

5% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

x	P(x)
0	0.8574
1	0.1354
2	0.0071
3	0.0001
Total	1



Binomial Variable: example

5% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. Compute E(X) and V(X)

Binomial Variable: example

5% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. Compute E(X) and V(X)

x	P(x)	xP(x)	x^2P(x)
0	0.8574	0.0000	0.0000
1	0.1354	0.1354	0.1354
2	0.0071	0.0143	0.0285
3	0.0001	0.0004	0.0011
Total	1	0.15	0.165

➔ $E(X) = 0.15$

$E(X) = n \cdot p = 3 \cdot 0.05 = 0.15$

➔ $\mathrm{V(X)} = E(X^2) - E(X)^2 = 0.165 - 0.15^2 = 0.1425$

$V(X) = n \cdot p \cdot (1-p) = 3 \cdot 0.05 \cdot 0.95 = 0.1425$

Binomial Variable: example

10% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

Binomial Variable: example

10% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

$$X = (0, 1, 2, 3)$$

$$P(X = 0) = \binom{3}{0} 0.10^0 (0.90)^{3-0} = \frac{3\times2\times1}{1\times3\times2\times1} 1(0.90)^3 = 0.729$$

$$P(X = 1) = \binom{3}{1} 0.10^1 (0.90)^{3-1} = 0.243$$

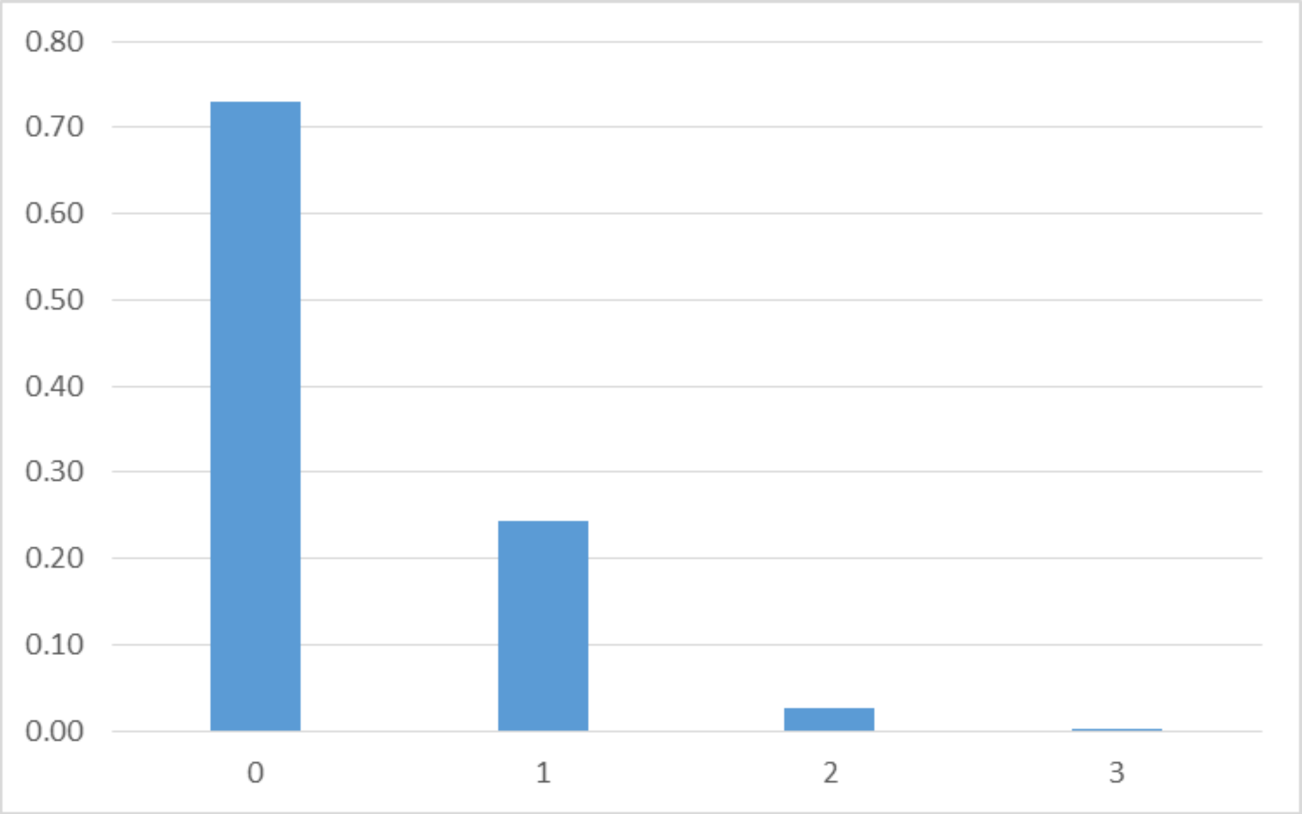
$$P(X = 2) = \binom{3}{2} 0.10^2 (0.90)^{3-2} = 0.027$$

$$P(X = 3) = \binom{3}{3} 0.10^3 (0.90)^0 = 0.001$$

Binomial Variable: example

10% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

x	P(x)
0	0.7290
1	0.2430
2	0.0270
3	0.0010
Total	1



Binomial Variable: example

30% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

Binomial Variable: example

30% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

$$X = (0, 1, 2, 3)$$

$$P(X = 0) = \binom{3}{0} 0.30^0 (0.70)^{3-0} = \frac{3\times2\times1}{1\times3\times2\times1} 1(0.70)^3 = 0.343$$

$$P(X = 1) = \binom{3}{1} 0.30^1 (0.70)^{3-1} = 0.441$$

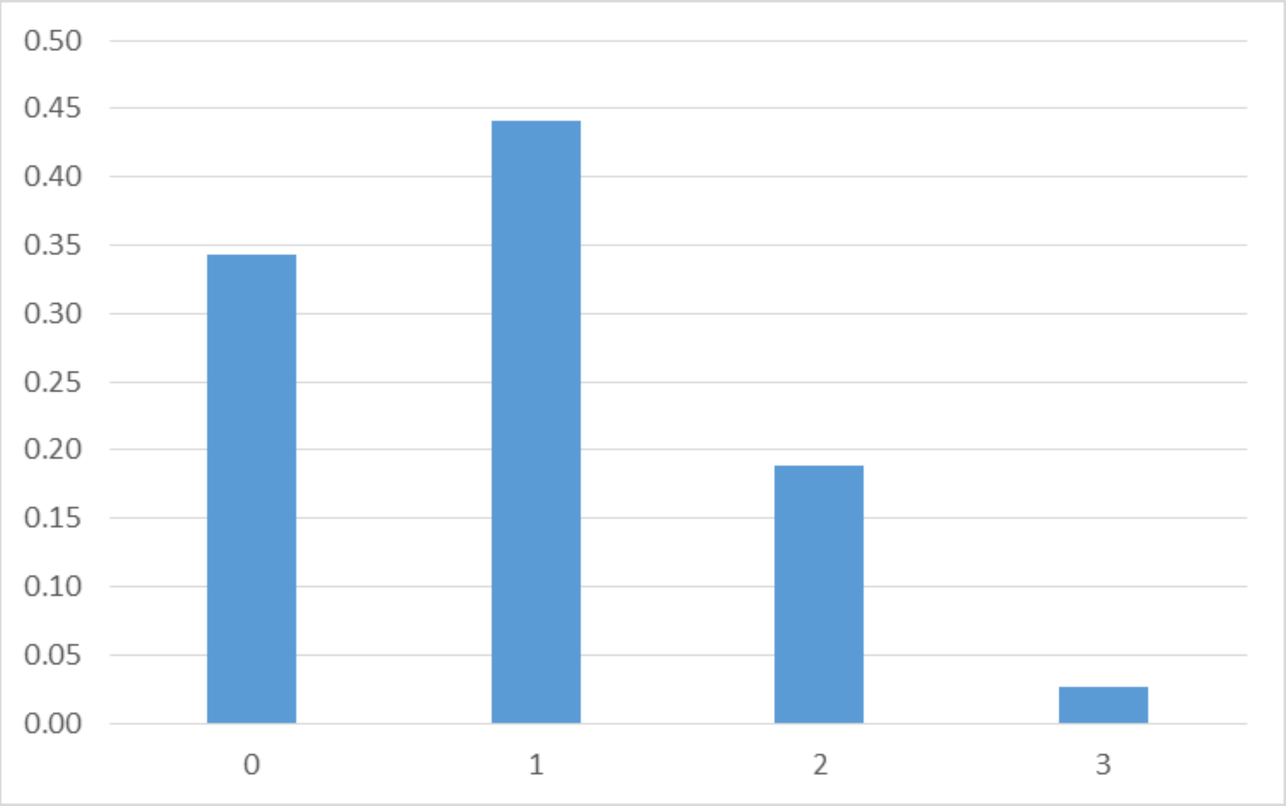
$$P(X = 2) = \binom{3}{2} 0.30^2 (0.70)^{3-2} = 0.189$$

$$P(X = 3) = \binom{3}{3} 0.30^3 (0.70)^0 = 0.027$$

Binomial Variable: example

30% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

x	P(x)
0	0.3430
1	0.4410
2	0.1890
3	0.0270
Total	1



Binomial Variable: example

50% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

Binomial Variable: example

50% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

$$X \ = \ (0, 1, 2, 3)$$

$$P(X = 0) = \binom{3}{0} 0.50^0 (0.50)^{3-0} = \frac{3\times2\times1}{1\times3\times2\times1} 1(0.70)^3 = 0.125$$

$$P(X = 1) = \binom{3}{1} 0.50^1 (0.50)^{3-1} = 0.375$$

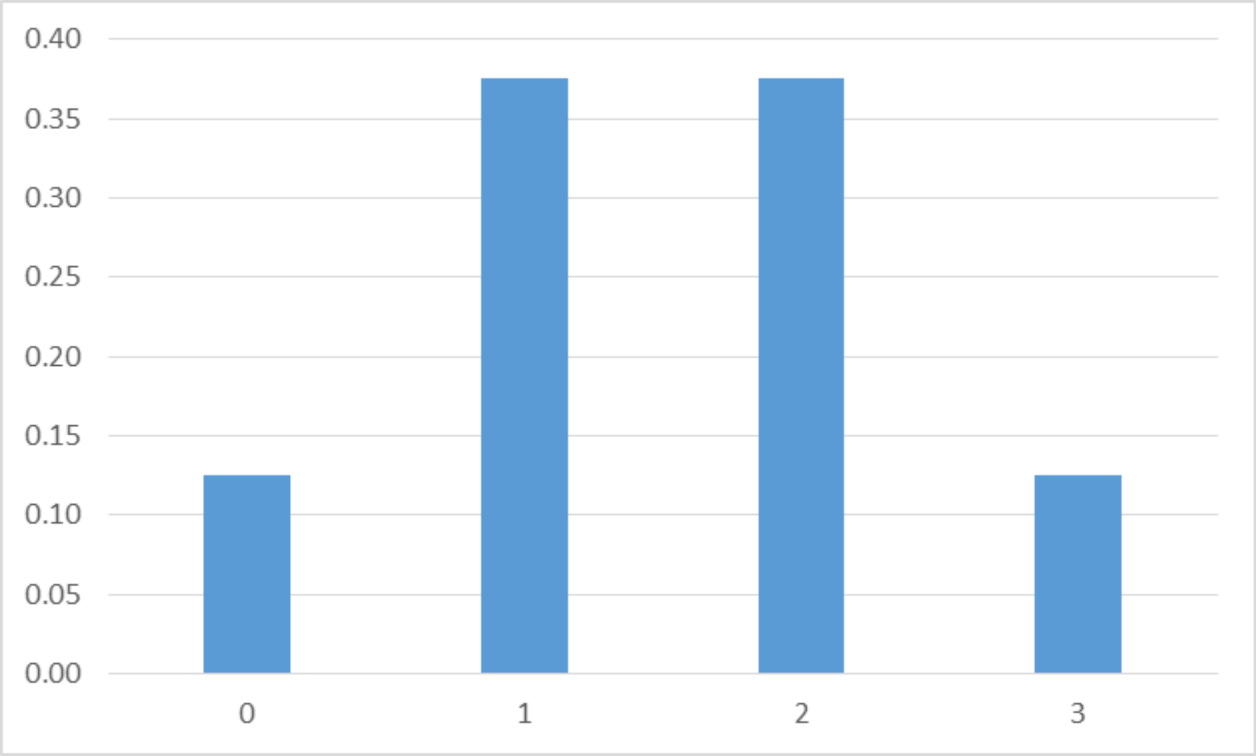
$$P(X = 2) = \binom{3}{2} 0.50^2 (0.50)^{3-2} = 0.375$$

$$P(X = 3) = \binom{3}{3} 0.50^3 (0.50)^0 = 0.125$$

Binomial Variable: example

50% percent of all DVD players manufactured by a large electronics company are defective. Let X be the number of defective DVD found in a random sample of 3 DVD players. What is the distribution of X?

x	P(x)
0	0.1250
1	0.3750
2	0.3750
3	0.1250
Total	1



Binomial Variable: example

2% of the packages mailed through Amazon do not arrive within the specified time. Suppose that 10 packages are mailed. Find the probability that:

(a) 1 will not arrive within the specified time

(b) at most 1 will not arrive within the specified time

Binomial Variable: example

2% of the packages mailed through Amazon do not arrive within the specified time. Suppose that 10 packages are mailed. Find the probability that:

(a) 1 will not arrive within the specified time

$$P(X = 1) = \binom{10}{1} 0.02^1 (0.98)^{10-1} = 0.1667$$

(b) at most 1 will not arrive within the specified time

$$P(X \le 1) = P(X = 0) + P(X = 1) =$$
$$= \binom{10}{0} 0.02^0 (0.98)^{10-0} + \binom{10}{1} 0.02^1 (0.98)^{10-1} =$$
$$= 0.8171 + 0.1667 = 0.9838$$

Binomial Variable: another example

Historical data show that 12% percent of all credit card holders of a US bank eventually become delinquent. A random sample of 3 credit card holders is extracted. Derive the probability distribution and the CDF:

Binomial Variable: another example

Historical data show that 12% percent of all credit card holders of a US bank eventually become delinquent. A random sample of 3 credit card holders is extracted. Derive the probability distribution and the CDF :

x	P(x)	F(x)
0	0.681	0.681
1	0.279	0.960
2	0.038	0.998
3	0.002	1.000
Total	1	

Binomial Variable: another example

Historical data show that 12% percent of all credit card holders of a US bank eventually become delinquent. A random sample of 3 credit card holders is extracted. Derive the E(X) and the V(X):

$E(X) = n{\times}p = 3 \times 0.12 = 0.36$

$V(X) = n{\times}p{\times}(1-p) = 3 \times 0.12 \times 0.88 = 0.317$

x	P(x)	xP(x)	x^2P(x)
0	0.681	0	0
1	0.279	0.279	0.279
2	0.038	0.076	0.152
3	0.002	0.005	0.016
Total	1	0.360	0.446

➔E(X)=0.36

➔V(X)=$E(X^2) - E(X)^2 = 0.446 - 0.36^2 = 0.317$

Hypergeometric Random Variable

Represents the number of successes in n ***not independent*** Bernoulli experiments

Values (integers): 0, 1,2,…,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=number of successes in population, x=nr of successes in sample

Hypergeometric Random Variable: example

Ex. 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.

- Single employee is a Bernoulli experiment (1 if female, 0=otherwise)
- They are not independent (extracted without replacement)
- In this example: $N = 12$, $n = 3$, r=7, x=3

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

Hypergeometric Random Variable: example

Ex. 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 at most 1 of them is a female:

$$P(X \leq 1) = P(X = 0) + P(X = 1) =$$

$$= \frac{\binom{7}{0}\binom{12-7}{3-0}}{\binom{12}{3}} + \frac{\binom{7}{1}\binom{12-7}{3-1}}{\binom{12}{3}} = \frac{1}{22} + \frac{7}{22} = 0.0455 + 0.3182 = 0.3637$$

Poisson Probability Distribution

Represents the number of occurrences ***<u>in a given interval</u>***. Occurrences have to be ***<u>random</u>*** and ***<u>independent</u>***

Example: number of telemarketing phone calls received ***<u>in a day</u>***

Values (integers): 0, 1,2,…,.. NO UPPER LIMIT!

Probability function:

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

With: μ = average number of occurrences in a given interval,

x = number of occurrences (<u>in the same interval</u>)

➔$E(X) = V(X) = \mu$

Poisson Probability Distribution

Poisson probability function:

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

It is possible to denote the average number of occurrences in a given interval (μ) also with λ, so

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

Poisson: example

1. A small business receives, on average, 12 customers per day. (a) What is the probability that the business will receive exactly 8 customers in one day?

Poisson: example

1. A small business receives, on average, 12 customers per day. (a) What is the probability that the business will receive exactly 8 customers in one day?

$\mu = 12 =$ $\qquad x = 8$

$$P(X = x) = \frac{\mu^x e^{-\mu}}{x!} \qquad e = 2.71828$$

Poisson: example

1. A small business receives, on average, 12 customers per day. (a) What is the probability that the business will receive exactly 8 customers in one day?

$\mu = 12 =$ $\qquad x = 8$

$$P(X = x) = \frac{\mu^x e^{-\mu}}{x!} \qquad e = 2.71828$$

$$P(X = 8) = \frac{12^8 e^{-12}}{8!} = 0.065523$$

$$\approx 6.55\%$$

Poisson: example

A washing machine breaks down an average of ***three times per year***. Find the probability that ***during one year*** this machine will have:

1) exactly 2 breakdowns

Poisson: example

A washing machine breaks down an average of ***<u>three times per year</u>***. Find the probability that ***<u>during one year</u>*** this machine will have:

1) exactly 2 breakdowns

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

$$P(X = 2) = \frac{3^2 e^{-3}}{2!} = 0.2240$$

Poisson: example

A washing machine breaks down an average of ***three times per year***.
Find the probability that ***during one year*** this machine will have:

2) At most 1 breakdown

Poisson: example

A washing machine breaks down an average of ***three times per year***.
Find the probability that ***during one year*** this machine will have:

2) At most 1 breakdown

$$P(X \leq x) = \sum_{i=0}^{x} \frac{\mu^i e^{-\mu}}{i!}$$

$$P(X \leq 1) = P(X = 0) + P(X = 1) =$$

$$= \frac{3^0 e^{-3}}{0!} + \frac{3^1 e^{-3}}{1!} = 0.0498 + 0.1494 = 0.1992$$

Poisson: interval of time has to be the same!

On average a household receives 10 telemarketing phone calls per month. Find the probability that a randomly selected household receives:

1) exactly 6 of such phone calls during a given month.

Poisson: interval of time has to be the same!

On average a household receives 10 telemarketing phone calls per month. Find the probability that a randomly selected household receives:

1) exactly 6 of such phone calls during a given month.

$$P(X = 6) = \frac{10^6 e^{-10}}{6!} = 0.063$$

Poisson: interval of time has to be the same!

On average a household receives 10 telemarketing phone calls per **month**. Find the probability that a randomly selected household receives:

2) exactly 3 of such phone calls during a given **week** (assume 4 weeks each month)

Poisson: interval of time has to be the same!

On average a household receives 10 telemarketing phone calls per **month**. Find the probability that a randomly selected household receives:

2) exactly 3 of such phone calls during a given **week** (assume 4 weeks each month)

$$\mu = \frac{10}{4} = 2.5$$

$$P(X = 3) = \frac{2.5^3 e^{-2.5}}{3!} = 0.214$$

Linear Combination of Random Variables

Let X and Y be 2 random variables, their linear combination is

$$aX + bY$$

- If $a = b = 1$ ➔ linear combination is the sum
- If $a = 1, b = -1$ ➔ linear combination is the difference

Then

$$E(aX + bY) = aE(X) + bE(Y)$$

$$V(aX + bY) = a^2 \cdot E(X) + b^2 \cdot E(Y) + 2a \cdot b \cdot Cov(XY)$$

Linear Combinations: Expected Value

$$\boxed{E(aX+b) = a \cdot E(X) + b}$$

This leads to the following:

1. For any constant a,
   $E(aX) = a \cdot E(X)$.

2. For any constant b,
   $E(X+b) = E(X) + b$.

Linear Combinations: Expected Value (proof)

$$E[aX + bY] = \sum_{i=1}^{k}\sum_{j=1}^{h}(ax_i + by_j)p_{ij} =$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{h} ax_i p_{ij} + \sum_{i=1}^{k}\sum_{j=1}^{h} by_j p_{ij} = a\sum_{i=1}^{k} x_i \sum_{j=1}^{h} p_{ij} + b\sum_{j=1}^{h} y_j \sum_{i=1}^{k} p_{ij} =$$

$$= a\sum_{i=1}^{k} x_i\, p_{i\cdot} + b\sum_{j=1}^{h} y_j\, p_{\cdot j} = a\mu_X + b\mu_Y$$

Linear Combinations: Variance (proof) 1/2

$$Var(aX+bY)=\sum_{i=1}^{k}\sum_{j=1}^{h}\left(ax_i+by_j-E[aX+bY]\right)^2 p_{ij}$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{h}\left(ax_i+by_j-a\mu_X-b\mu_Y\right)^2 p_{ij}$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{h}\left[a\left(x_i-\mu_X\right)+b\left(y_j-\mu_Y\right)\right]^2 p_{ij}$$

$$=\cdots$$

Linear Combinations: Variance (proof) 2/2

$Var(aX + bY) = \cdots$

$$= \sum_{i=1}^{k}\sum_{j=1}^{h}\Big[a^2(x_i-\mu_X)^2 + b^2(y_j-\mu_Y)^2$$

$$+ 2ab(x_i-\mu_X)(y_j-\mu_Y)\Big]p_{ij}$$

$$= a^2\sum_{i=1}^{k}\sum_{j=1}^{h}(x_i-\mu_X)^2 p_{ij} + b^2\sum_{i=1}^{k}\sum_{j=1}^{h}(y_j-\mu_Y)^2 p_{ij}$$

$$+ 2ab\sum_{i=1}^{k}\sum_{j=1}^{h}(x_i-\mu_X)(y_j-\mu_Y)p_{ij}$$

$$= a^2\sum_{i=1}^{k}(x_i-\mu_X)^2\sum_{j=1}^{h}p_{ij} + b^2\sum_{j=1}^{h}(y_j-\mu_Y)^2\sum_{i=1}^{k}p_{ij} + 2ab\sigma_{XY}$$

$$= a^2\sum_{i=1}^{k}(x_i-\mu_X)^2 p_{i\cdot} + b^2\sum_{j=1}^{h}(y_j-\mu_Y)^2 p_{\cdot j} + 2ab\sigma_{XY}$$

$$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}.$$