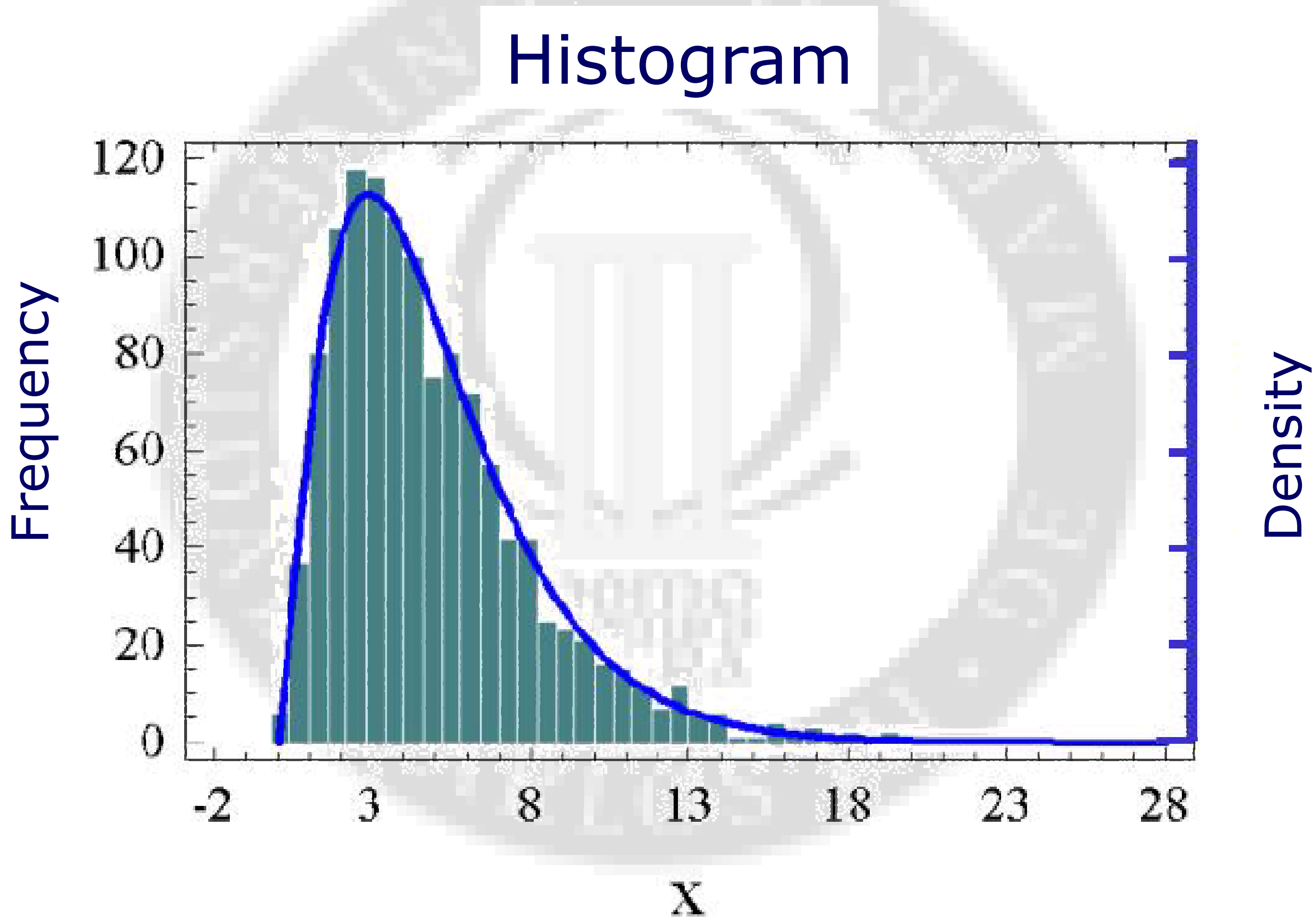
IV. Introduction to Random Variables



Chapter IV: Introduction to Random Variables

1. **Introduction**
2. **Univariate discrete random variables**
3. **Univariate continuous random variables**
4. **Characteristics measures of a random variables**

1. Introduction

What is a random variable?

Experiment: It is the process through which we obtain data given some **experimental conditions**. Getting new data, under the same conditions, means to repeat the experiment

Random experiment: It is an experiment where it is possible to obtain different outcomes even if the experiment is repeated under the same experimental conditions

Example: measuring the time a machine takes to complete a specific task, to observe the value obtained by rolling a dice...

Random variable: it is the outcome of a random experiment whose result is numeric. It is the numerical output variable of a random experiment

Example: Experiment - to observe the outcomes of rolling a die

Random variable – the numerical value that is obtained by rolling the dice (discrete variable)

Example: Experiment – to measure the time spent by a computer to access a network

Random variable – the numerical value of the time (continuous variable)

1. Introduction

Random variable: it is the numerical outcome of a random experiment

How is a random variable defined?

A Defining the sample space

B Having a set function that assigns probabilities to the different events of interest (*subsets of the sample space*)

C Having a function that maps the elements of the sample space to the real (or complex) numbers (the *random variable*)

Example: Experiment - to observe the outcomes of a coin tossing

Random variable – numerical value that is obtained by tossing a coin (discrete variable) 0 if the result is head, 1 if it is tail

- Sample space: {head, tail}
- Probability model: P(head)=P(tail)=0.5
- Random Variable : head -> 0, tail -> 1

1. Introduction

NOTATION:

<u>Radom variable</u>: We use the last capital letters of the alphabet to refer to a r..v.

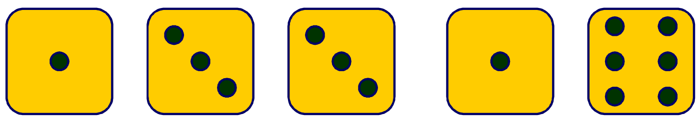
Example: X=outcome of a coin tossing

Y=outcome of rolling a die

<u>Actual values</u>: the actual observed values are called **<u>realizations</u>** of the random variable.

Example: Y=outcome of rolling a die

If we throw a die 5 times we will have 5 **realizations** of the random variables Y, for example 1,3,3,1,6.



realizations

Chapter IV: Introduction to Random Variables

1. Introduction
2. **Univariate discrete random variables**
3. Univariate continuous random variables
4. Characteristics measures of a random variables

2. Univariate discrete random variables

Let X be a discrete random variable

Let $x_1, x_2, \ldots, x_K$ be the possible values assumed by the variable

Example

X= value obtained by rolling a die

$$\begin{cases} x_1=1 & X_3=3 & x_5=5 \\ x_2=2 & X_4=4 & x_6=6 \end{cases}$$

X= state acceptable/defective of a manufactured item

$$\begin{cases} x_1=0 \text{ (acceptable)} \\ x_2=1 \text{ (defective)} \end{cases}$$

X= number of customers who arrive to a service point in a unit of time

$$\Big\{ \; x_1=0,\ x_2=1,\ \ldots \text{ ideally infinite}$$

A probability model is a way to assign probabilities to the events. It consists in defining one function

the probability function or the distribution function

Probability function

It is a function, $p(x)$, that assigns probabilities to each different value assumed by a discrete random variable X: $p(x_1)$, $p(x_2)$,..., $p(x_K)$

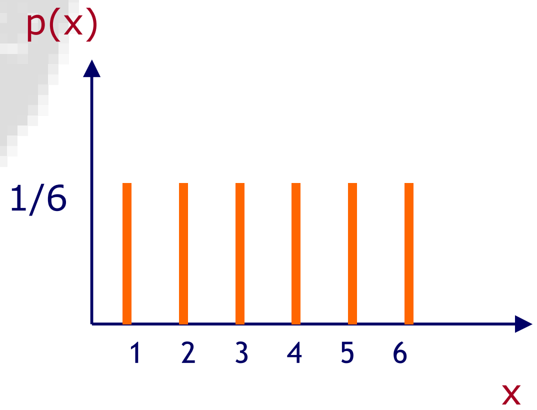
These probabilities may be obtained by indefinitely repeating a given random experiment and by calculating the relative frequencies

Example

X= outcomes of rolling a die

we have that $x=1,\dots,6$ and since of each values is equiprobable

We have a constant probability function: $p(x)=1/6$ for $x=1,\dots,6$.



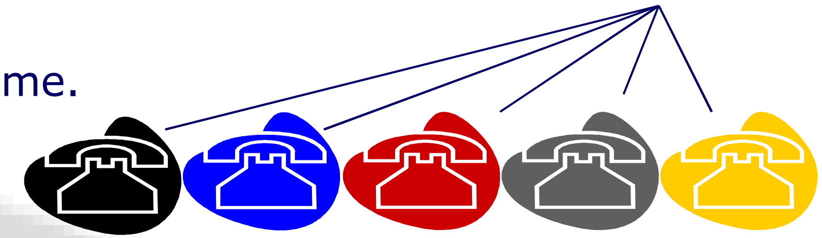
This kind of random variables, with K different and equiprobable values, are called **DISCRETE UNIFORM**

Example

A PBX (telephone exchange) has 5 lines.

Let X=number of busy lines in a unit of time.

Sample Space: X={0,1,2,3,4,5}



We assume that it receives on average 2 calls by unit of time and that each call ends in one unit of time. Then it is possible to show (see next chapter) that the probability function is:

P(X=0)=0.14

P(X=1)=0.27

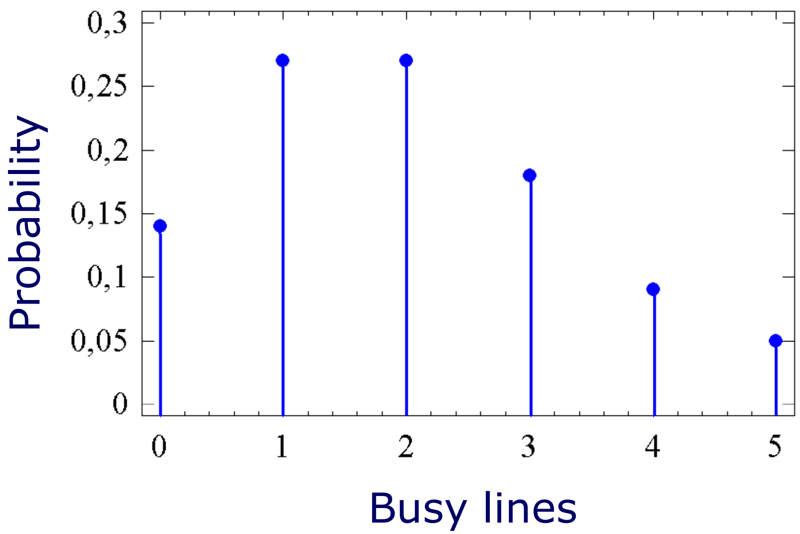
P(X=2)=0.27

P(X=3)=0.18

P(X=4)=0.09

P(X=5)=0.05

Probability function



Probability of event A: are there more than 2 busy lines?

P(A)=P[(X=3) U (X=4) U (X=5)]= P(X=3) + P(X=4) + P(X=5)=0.32

What is the probability to have an available line?

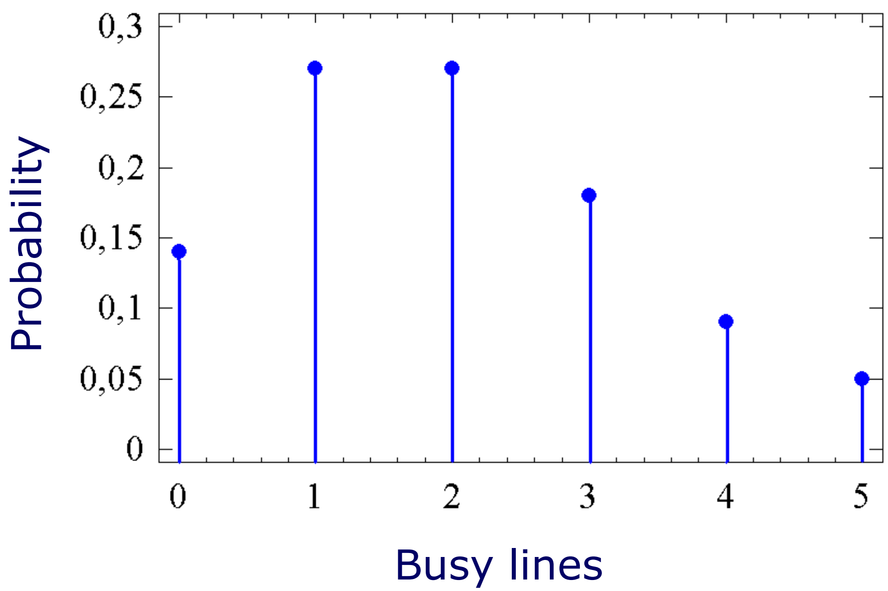
Distribution function

- It is a function F(x) defined in the whole real line
- At each point x, it is equal to the cumulative distribution up to that point, that is

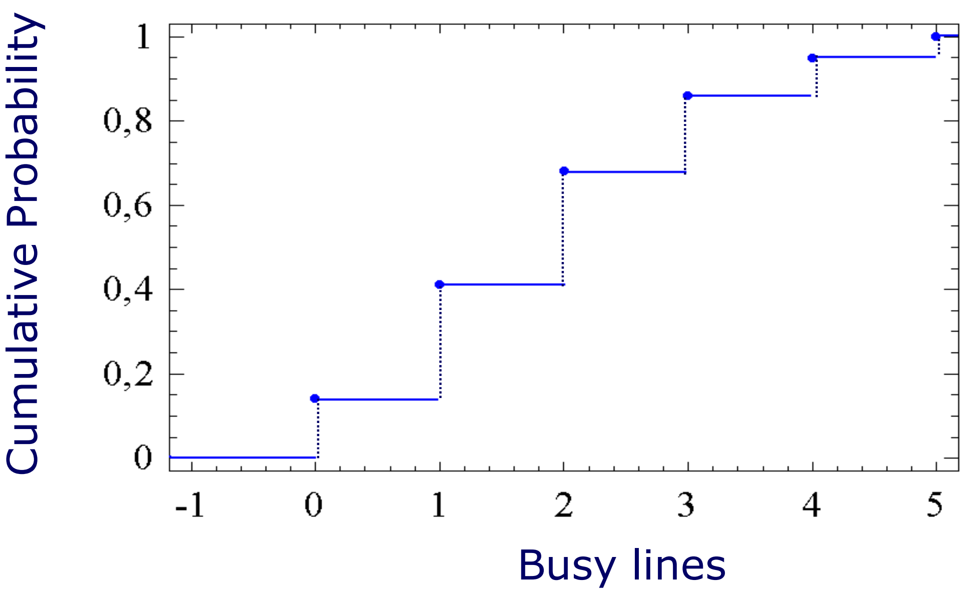
$$F(x)=P(X\leq x)$$

- $F(-\infty)=0$, $F(+\infty)=1$

Probability function



Distribution function

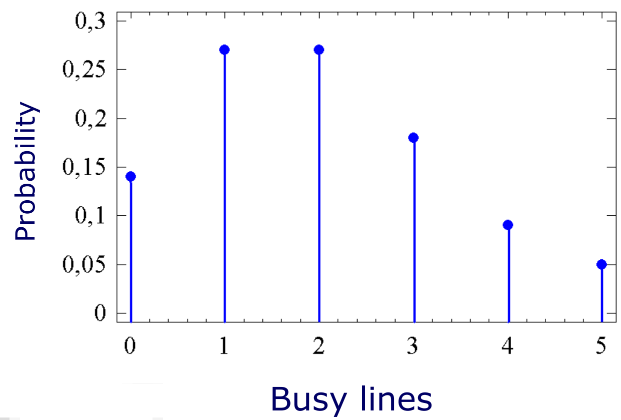


X= number of busy lines in a unit of time

Probability function

$$p(x) = \begin{cases} \mathbf{0.14} & \text{if } x=0 \\ \mathbf{0.27} & \text{if } x=1 \\ \mathbf{0.27} & \text{if } x=2 \\ \mathbf{0.18} & \text{if } x=3 \\ \mathbf{0.09} & \text{if } x=4 \\ \mathbf{0.05} & \text{if } x=5 \end{cases}$$

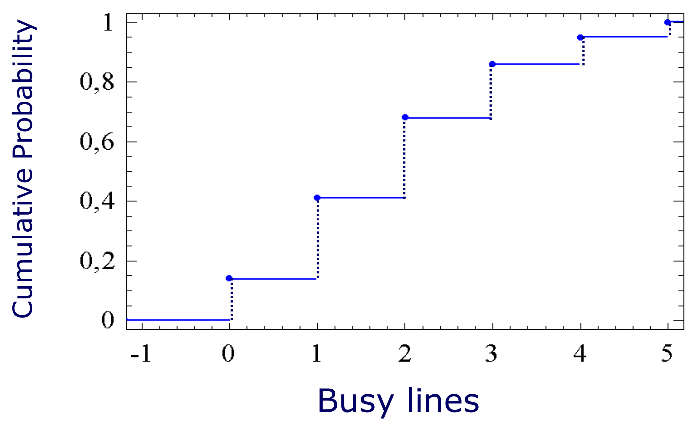
Probability function



Distribution function

$$F(x)=P(X\le x) = \begin{cases} 0 & \text{if } x<0, \text{then } P(X \le 0)=0 \\ 0.14 & \text{if } 0\le x<1, \text{then } P(X \le x)=P(X<0)+P(X=0) \\ 0.41 & \text{if } 1\le x<2, \text{then } P(X \le x)=P(X<1)+P(X=1) \\ 0.68 & \text{if } 2\le x<3, \text{then } P(X \le x)=P(X<2)+P(X=2) \\ 0.86 & \text{if } 3\le x<4, \text{then } P(X \le x)=P(X<3)+P(X=3) \\ 0.95 & \text{if } 4\le x<5, \text{then } P(X \le x)=P(X<4)+P(X=4) \\ 1.00 & \text{if } x\ge 5 \end{cases}$$

Distribution function



Chapter IV: Introduction to Random Variables

1. Introduction
2. Univariate discrete random variables
3. Univariate continuous random variables
4. Characteristics measures of a random variables

3. Univariate continuous random variable

Continuous random variables may assume uncountable infinite values

$P(X=x)=0$ always and therefore we are only interested in the probabilities of intervals $P(X>a)$, $P(a<X \leq b)$...

Examples

- X=values are randomly chosen in the interval [3,4], ← continuous uniform

$P(X=x)=1/\infty=0$

$P(X<3.5)=0.5$

- X= length of a piece
- X= time it takes to run a task
- X= arrival times between two successive customers

We can compute the probabilities of the events by using:

the density function or the distribution function

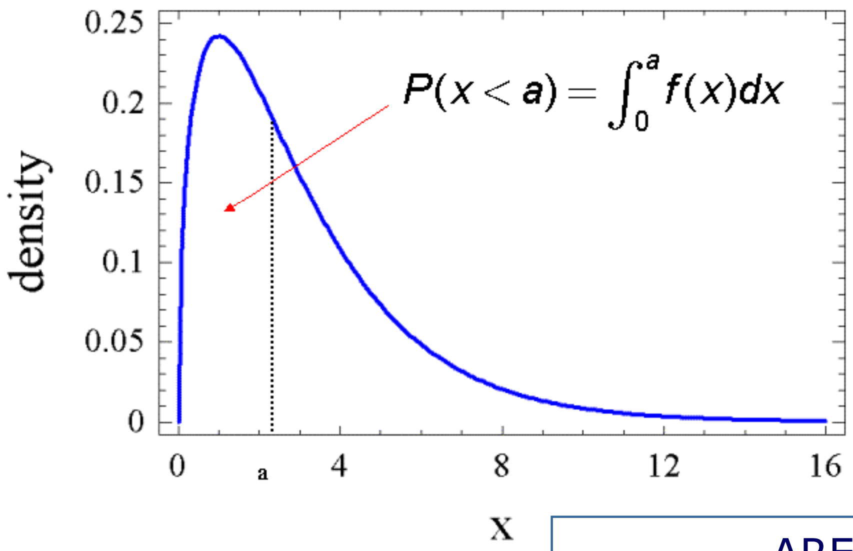
Density function

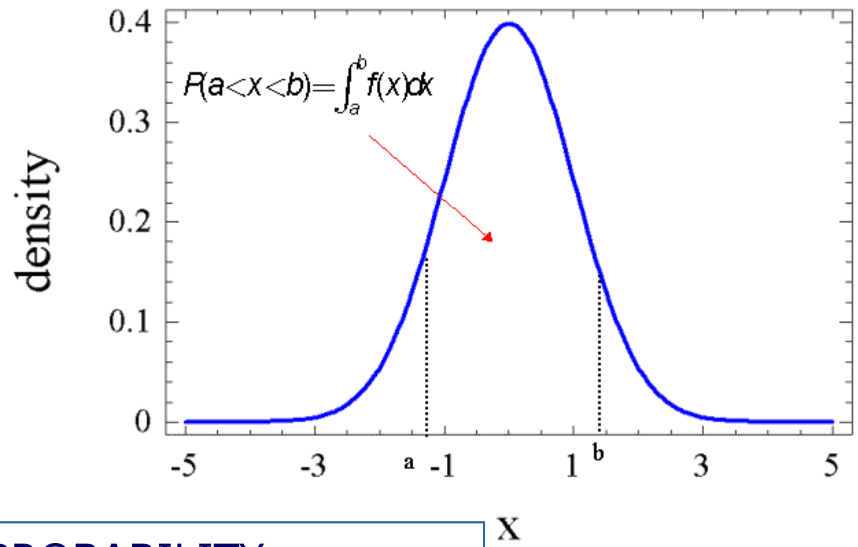
f(x): **Probability density function** of the random variable X at the point x

The density function f(x) is a mathematical function such as

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$$

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



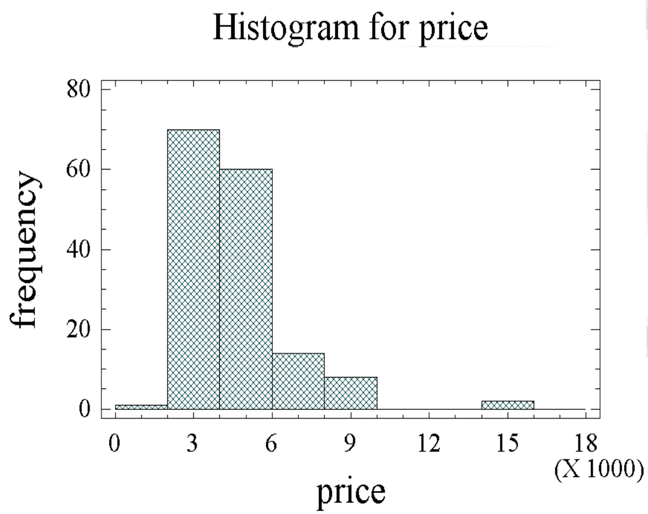


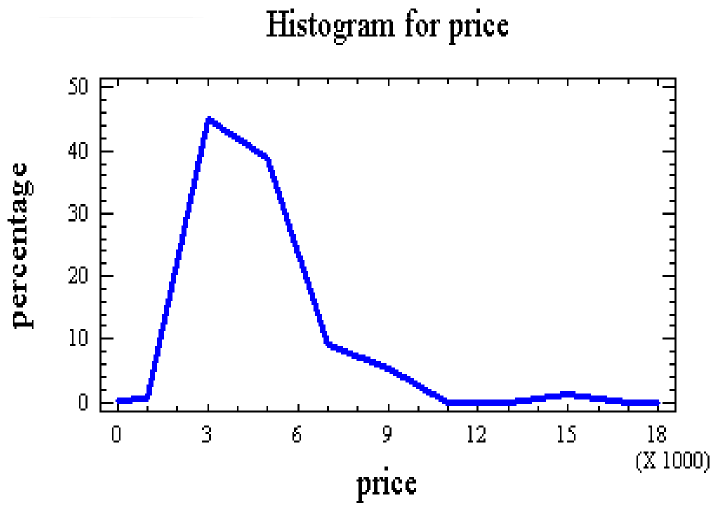
AREA=PROBABILITY

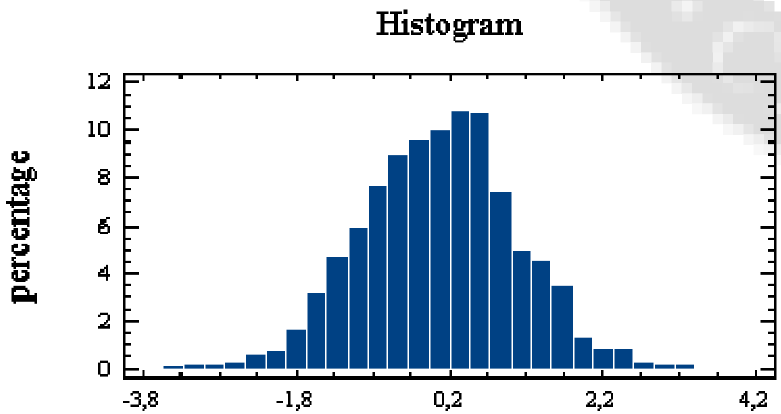
It can be computed as the rescaled limit of the polygon of relative frequencies when the size of the data set converges to infinity

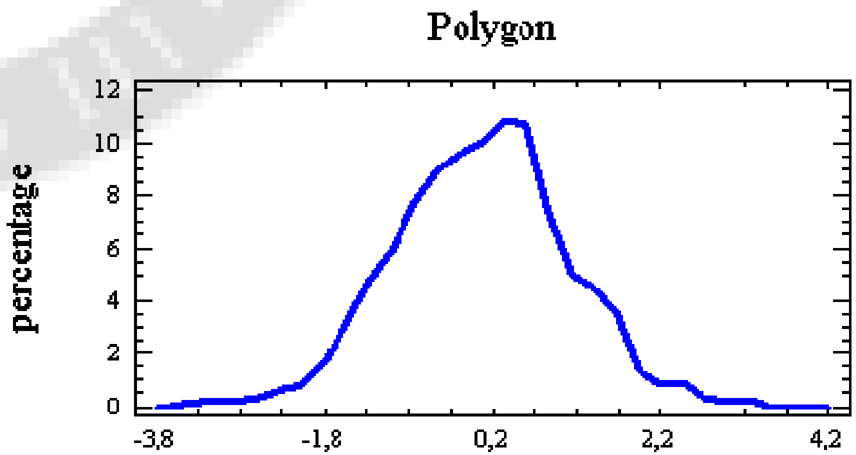
When we have more data:

- We need to use more classes to collect them
- and the polygon becomes smoother

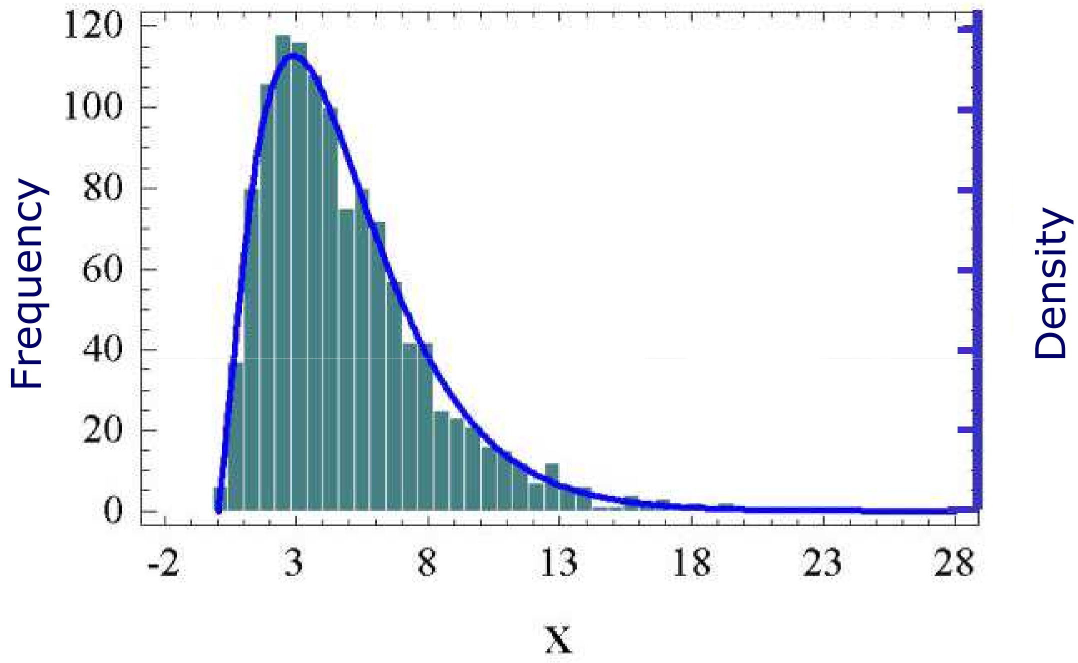








Histogram



Differences:

- Histogram or polygon: Describe only the n observed data (SAMPLE)

  It gives the frequency (absolute o relative) of each interval

- Density function: Describe the whole POPULATION

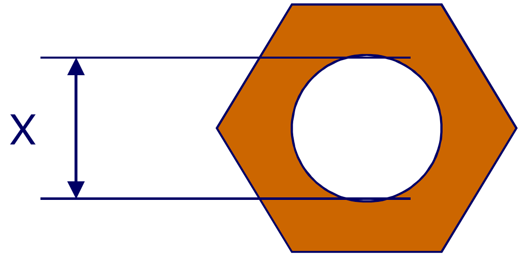
  It gives the probability density

  or probability by unit of measure in each point

(they can be measured on the same picture by using two different unit scales)

Example

X=length of manufactured item



$$f(x) = \begin{cases} k(x-1)(3-x) & \text{if } x \in [1,3] \\ 0 & \text{otherwise} \end{cases}$$

What value should take k?

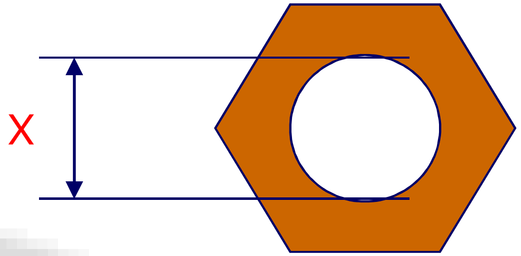
$$\boxed{\int_{-\infty}^{\infty} f(x)dx = 1}$$

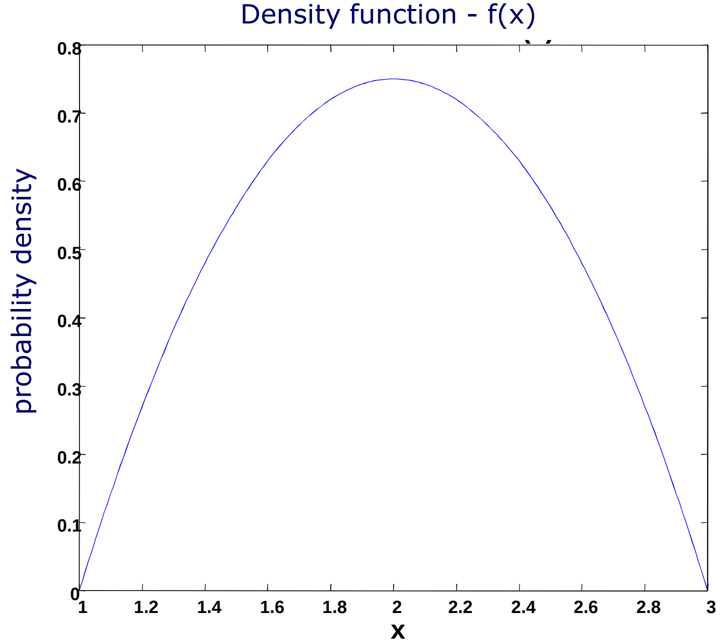
$$\int_{-\infty}^{\infty} k(x-1)(3-x)dx = \int_{-\infty}^{1} \cancelto{0}{f(x)dx} + \int_{1}^{3} f(x)dx + \int_{3}^{\infty} \cancelto{0}{f(x)dx}$$

$$= \int_{1}^{3} k(x-1)(3-x)dx = 1 \Rightarrow k = \frac{3}{4}$$

Example

X=length of manufactured item



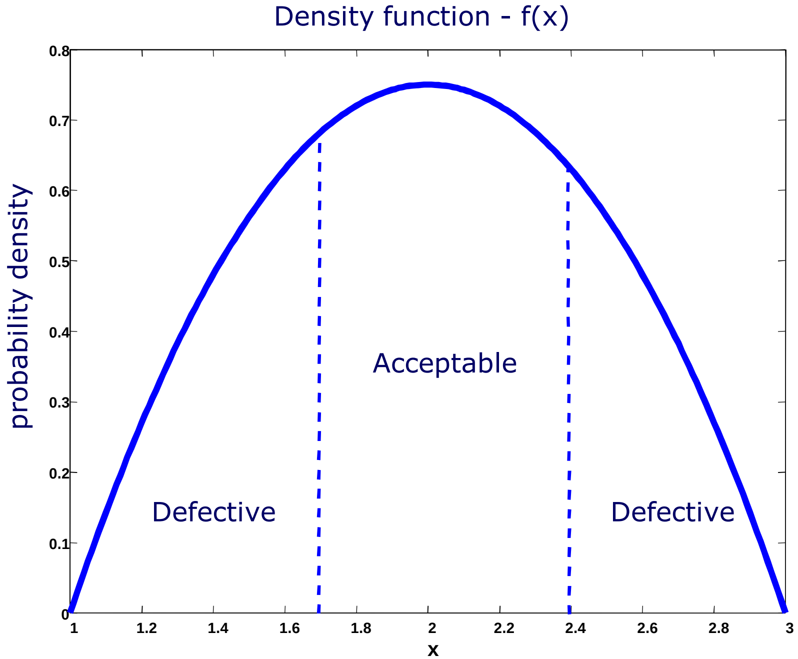


$$f(x) = \begin{cases} \frac{3}{4}(x-1)(3-x) & \text{if } x \in [1,3] \\ 0 & \text{otherwise} \end{cases}$$

We say that the item is acceptable when

X belongs to the interval [1.7;2.4]

What percentage of defective pieces is make?



$$P(\text{Acceptable}) = P(1.7 < X < 2.4)$$

$$= \int_{1.7}^{2.4} \frac{3}{4}(x-1)(3-x)dx = 0.502$$

We can compute the probabilities of the events by using:

the density function or the distribution function

Distribution function

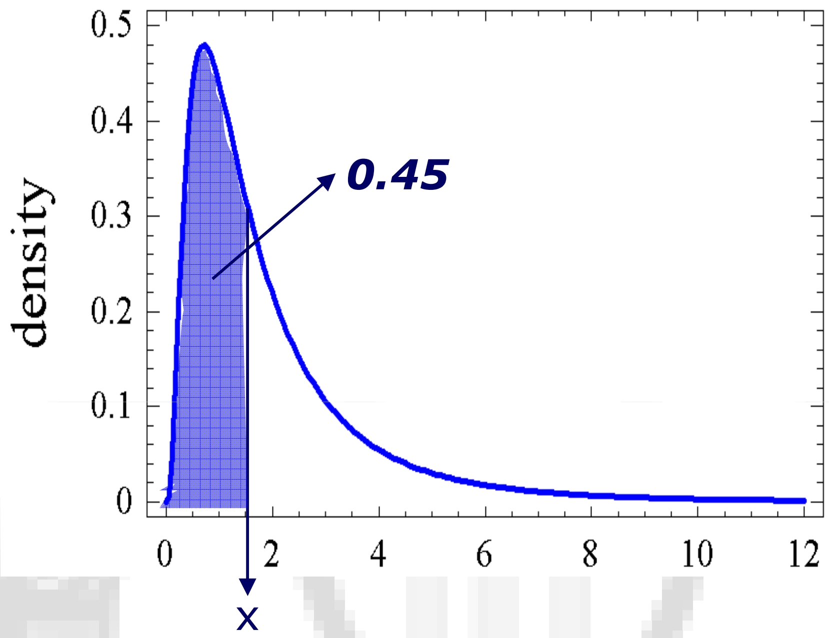
It has the same definition as in the discrete case

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \longrightarrow f(x) = \frac{dF(x)}{dx}$$

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = \int_{-\infty}^{x_2} f(x)\,dx - \int_{-\infty}^{x_1} f(x)\,dx$$

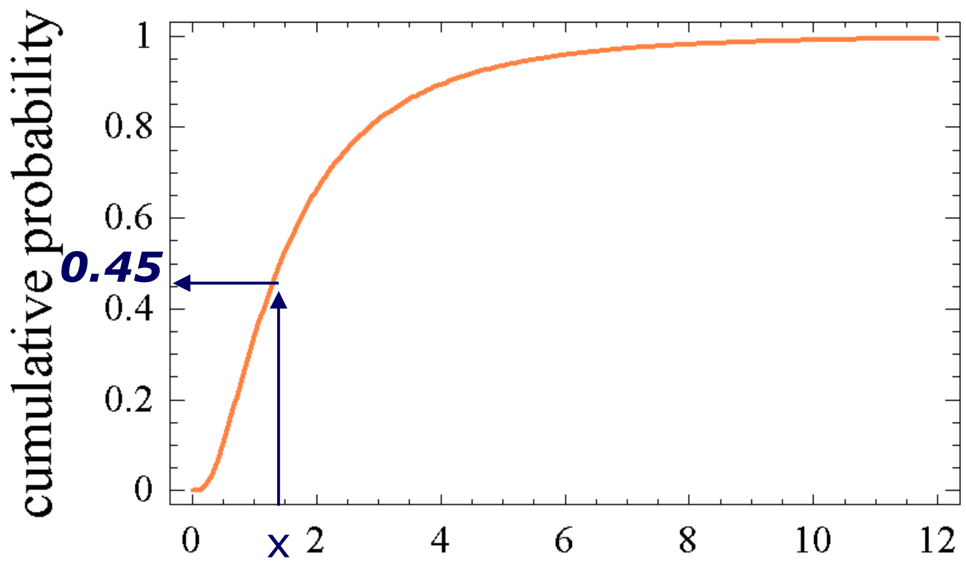
$$= F(x_2) - F(x_1).$$

f(x)



$$\int_0^x f(v)dv = 0.45$$

F(x)=P(X≤x)



Chapter IV: Introduction to Random Variables

1. Introduction
2. Univariate discrete random variables
3. Univariate continuous random variables
4. Characteristics measures of a random variables

4. Characteristics measures of the random variables

We are interested in measures that could summarize some important characteristics of the random variable

4.1 Measures of position

1. Mean
2. Median
3. Mode

4.2 Measures of dispersion

1. Variance
2. Quartiles

4.3 Covariance and correlation

4.4 Effects of the linear transformations

1. Mean, mathematical expectation, E(X), μ

It is the average of the data of an infinite dataset that represents the population

How can we compute it?

Discrete random variables:

From Chapter I: To calculate the mean of a dataset whose values are grouped into classes ...

If there are J different values which are repeated:

x_1, is repeated n_1 times

x_2, is repeated n_2 times

...

x_J, is repeated n_J times

$$\bar{x} = \sum_{j=1}^{J} x_j f_r(x_j).$$

where $fr(x_j)$ is the relative frequency of the value x_j

1. Mean, mathematical expectation, E(X), μ

It is the average of the data of an infinite dataset that represents the population

How can we compute it?

Discrete random variables:

Let X be a discrete random variable and let $x_1, x_2, \ldots, x_K$ be the different values that it can take

$$\mu = E(X) = \sum_{i=1}^{K} x_i p(x_i)$$

Example

A PBX has 5 lines.
Let X=number of busy lines in a unit of time.
Sample Space: X={0,1,2,3,4,5}

$$p(x) = \begin{cases} \mathbf{0.14} & \text{si } x=0 \\ \mathbf{0.27} & \text{si } x=1 \\ \mathbf{0.27} & \text{si } x=2 \\ \mathbf{0.18} & \text{si } x=3 \\ \mathbf{0.09} & \text{si } x=4 \\ \mathbf{0.05} & \text{si } x=5 \end{cases}$$

$$E(X) = 0 \times 0.14 + 1 \times 0.27 + 2 \times 0.27 + 3 \times 0.18 + 4 \times 0.09 + 5 \times 0.05 = \underline{\underline{1.96}}$$

1. Mean, mathematical expectation, E(X), μ

Discrete random variables:

$$\mu = E(X) = \sum_{i=1}^{K} x_i p(x_i)$$

Continuous random variables:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

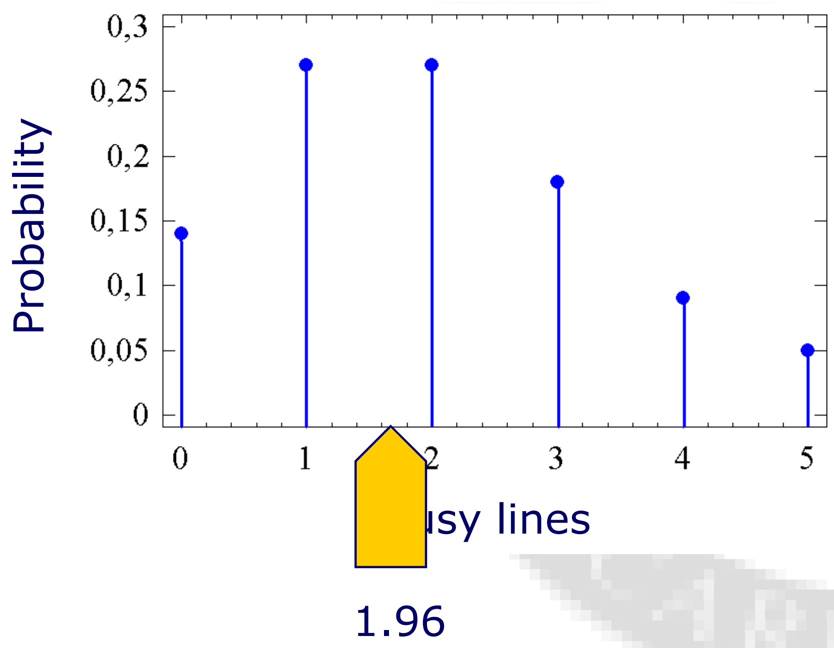
Example: Let X be a continuous r.v. defined in (0,1) and with density $f(x) = 12x^2(1-x)$

$$E(X) = \int_0^1 x f(x)\,dx = \int_0^1 12x^3(1-x)\,dx = 12\left(\frac{x^4}{4} - \frac{x^5}{5}\right)\Big]_0^1 = \frac{12}{20} = \frac{3}{5}$$

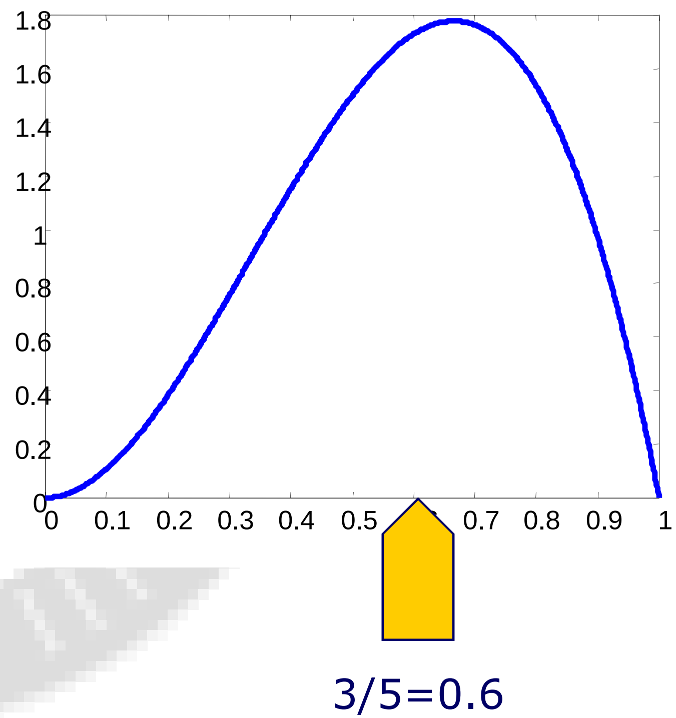
1. Mean, mathematical expectation, E(X), μ

It represents the position of the gravity centre for the entire population

Probability function



$$f(x) = 12x^2(1 - x)$$



1. Mean, mathematical expectation, E(X), μ

Some properties (similar to the sample mean):

$$E(g(X)) = \begin{cases} \sum_{i=1}^{K} g(x_i)p(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

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

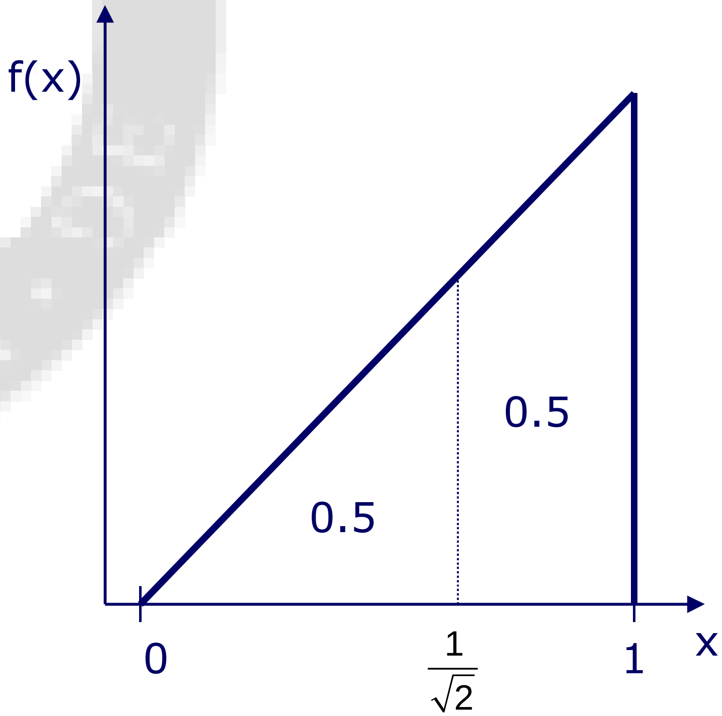
2. Median

The median of X is the value x_m that splits the range of the random values into two parts each one having probability 50%

$$F(x_m) = P(X \leq x_m) = 0.50$$

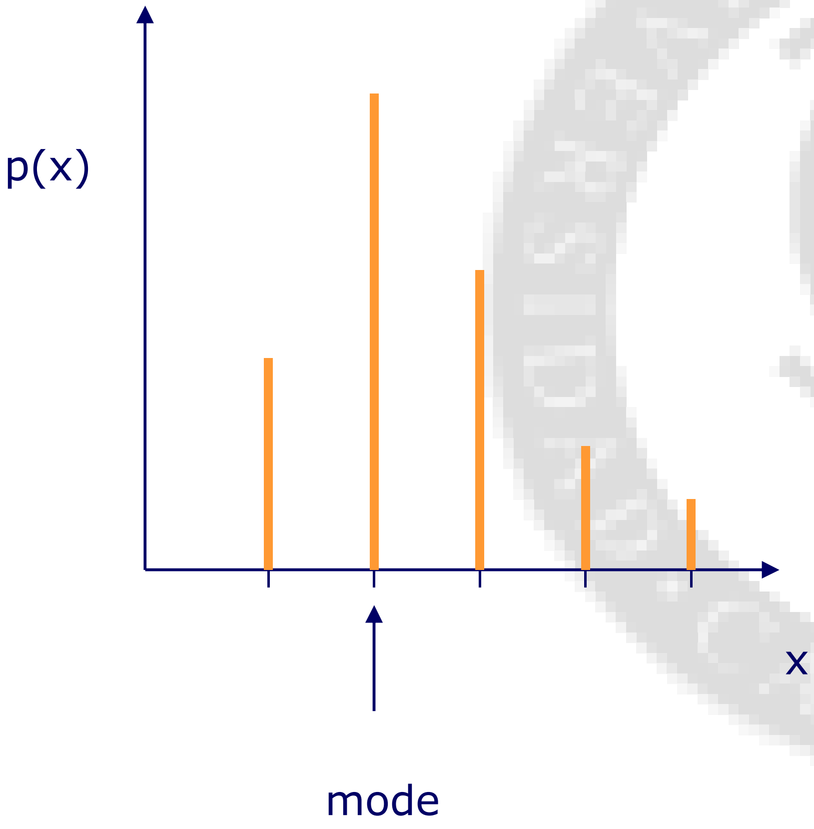
Example: Let X be a continuous r.v. in (0,1) with density function f(x)=2x

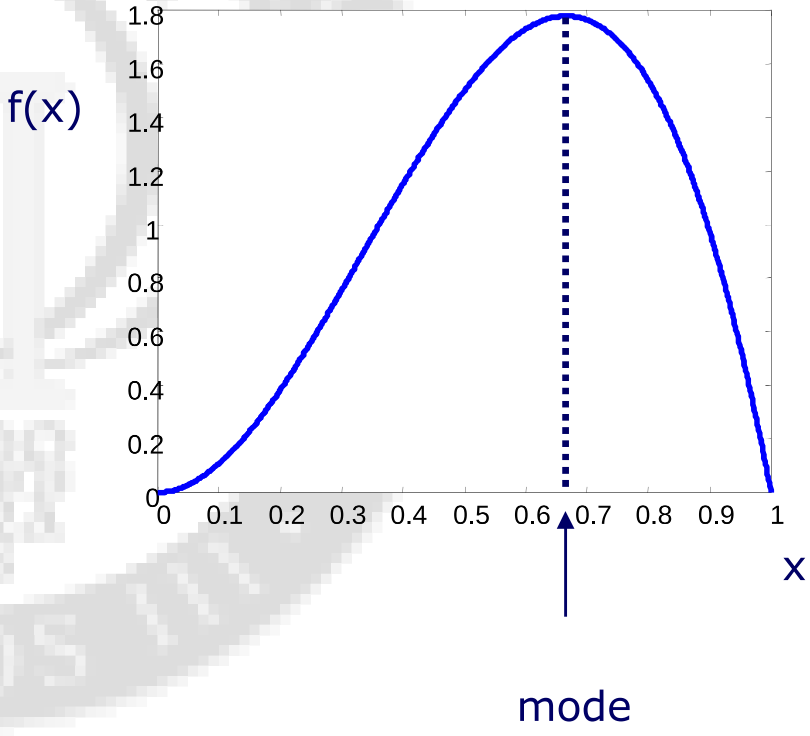
$$\int_0^{x_m} 2x\,dx = 0.5 \Rightarrow \left.\frac{2x^2}{2}\right]_0^{x_m} = x_m^2 = 0.5 \Rightarrow x_m = \frac{1}{\sqrt{2}}$$



3. Mode

The mode of X is the value x that maximizes p(x) or f(x)





1. Variance, var(X), σ^2

Variance of a dataset:

Sample mean of the squared deviations from the sample mean

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n},$$

Variance of a random variable:

(variance of the whole population)

Expected value of the squared deviations from the expectation

Random variable → X

Its mean or expectation → $E(X) = \mu$

$$\mu = E(X) = \sum_{i=1}^{K} x_i p(x_i)$$

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

Deviation from the mean → $X - \mu$

Squaring... → $(X - \mu)^2$

Its mean, or expectation, is the variance → $\sigma^2 = E[(X-\mu)^2]$

Standard deviation $= \sigma$; Coefficient of variation $= \sigma/|\mu|$

How to calculate the variance from p(x) or f(x)?

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

E(X)= μ

$$\mu = E(X) = \sum_{i=1}^{K} x_i p(x_i)$$

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$E(g(X)) = \begin{cases} \sum_{i=1}^{K} g(x_i) p(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

$$g(X) = (X-\mu)^2$$

$$\mathrm{var}(X) \equiv \sigma^2 = \begin{cases} \sum_{i=1}^{K} (x_i - \mu)^2 p(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Example: X= number of busy lines. X={0,1,2,3,4,5}.

$$p(x) = \begin{cases} \mathbf{0.14} & \text{if } x=0 \\ \mathbf{0.27} & \text{if } x=1 \\ \mathbf{0.27} & \text{if } x=2 \\ \mathbf{0.18} & \text{if } x=3 \\ \mathbf{0.09} & \text{if } x=4 \\ \mathbf{0.05} & \text{if } x=5 \end{cases}$$

$$E(X) = 0 \times 0.14 + 1 \times 0.27 + 2 \times 0.27 + 3 \times 0.18 + 4 \times 0.09 + 5 \times 0.05 = 1.96$$

$$\sigma^2 = \sum_{i=1}^{K} (x_i - \mu)^2 p(x_i)$$

$$Var(X) = (0 - 1.96)^2 \times 0.14 + (1 - 1.96)^2 \times 0.27 + (2 - 1.96)^2 \times 0.27$$
$$+ (3 - 1.96)^2 \times 0.18 + (4 - 1.96)^2 \times 0.09 + (5 - 1.96)^2 \times 0.05 = 1.82$$

Example: Let X be a continuous r.v. define in (0,1) with $f(x) = 12x^2(1-x)$

$$E(X) = \int_0^1 x f(x)dx = \int_0^1 12x^3(1-x)dx = 12\left(\frac{x^4}{4} - \frac{x^5}{5}\right)\Big]_0^1 = \frac{12}{20} = \frac{3}{5}$$

$$\text{var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_0^1 \left(x - \frac{3}{5}\right)^2 12x^2(1-x)dx = \frac{1}{25}$$

Other way to calculate the variance:

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

E(X) is a constant

Its expected value is itself

$$E\left[(X-E[X])^2\right] = E\left[X^2 - 2X\,E[X] + E[X]^2\right]$$
$$= E\left[X^2\right] - E[2X\,E[X]] + E\left[E[X]^2\right]$$
$$= E\left[X^2\right] - 2E[X]E[X] + E[X]^2$$
$$= E\left[X^2\right] - E\left[X\right]^2$$

Example: Let X be a continuous r.v. define in (0,1) with $f(x) = 12x^2(1-x)$

$$E(X) = \int_0^1 x f(x)dx = \int_0^1 12x^3(1-x)dx = 12\left(\frac{x^4}{4} - \frac{x^5}{5}\right)\Big]_0^1 = \frac{12}{20} = \frac{3}{5}$$

$$E[X^2] = \int_0^1 x^2 f(x)dx = \int_0^1 12x^4(1-x)dx = 12\left(\frac{x^5}{5} - \frac{x^6}{6}\right)\Big]_0^1 = \frac{12}{30} = \frac{2}{5}$$

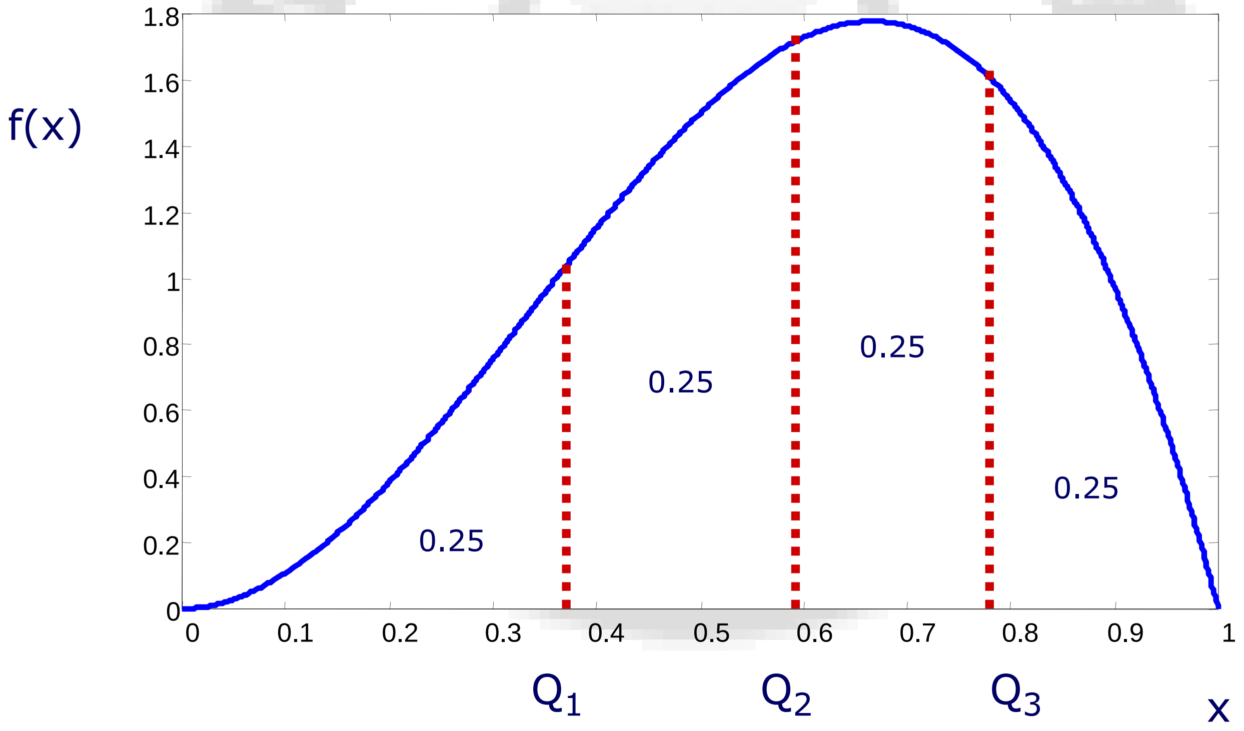
$$\mathrm{var}[X] = E[X^2] - (E[X])^2 = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25}$$

2. Quartiles Q_1, Q_2, Q_3

They are the values that separate the population into three groups of probability 0.25

They can be calculated in the same way as the median (Q_2). That is:

$$F(Q_1)=0.25; F(Q_3)=0.75$$



Working with two joint random variables (X,Y)

Discrete case

We need the bivariate probability mass function

$$\mathrm{p}(x,y)=\mathrm{P}(X=x,Y=y)$$

Continuous case

We need the bivariate density function

$$\mathrm{f}(x,y)=\lim_{\Delta x,\Delta y\to 0}\frac{\mathrm{P}(X\in[x,x+\Delta x),Y\in[y,y+\Delta y))}{\Delta x\,\Delta y}$$

Or the bivariate distribution function

$$\mathrm{F}(x,y)=\mathrm{P}(X\le x,Y\le y)$$

The formula to compute the expectation of a function h(x,y) is

$$\mathrm{E}[\mathrm{h}(X,Y)]=\sum_{x,y=-\infty}^{\infty}\mathrm{h}(x,y)\mathrm{p}(x,y)$$

$$\mathrm{E}[\mathrm{h}(X,Y)]=\iint_{(x,y)\in R^2}\mathrm{h}(x,y)\mathrm{f}(x,y)\,dx\,dy$$

In general given two functions h(x) and g(y) we have that

$$\mathrm{E}[\mathrm{h}(X)\mathrm{g}(Y)]\neq\mathrm{E}[\mathrm{h}(X)]\,\mathrm{E}[\mathrm{g}(Y)]$$

The equality holds in the case X and Y are independent. Indeed when X $\perp$ Y we have that

$$\mathrm{p}(x,y)\overset{\perp}{=}\mathrm{P}(X=x)\mathrm{P}(Y=y)$$
$$=\mathrm{p}_X(x)\mathrm{p}_Y(y)$$

$$\mathrm{f}(x,y)\overset{\perp}{=}\mathrm{f}_X(x)\mathrm{f}_Y(y)$$

and

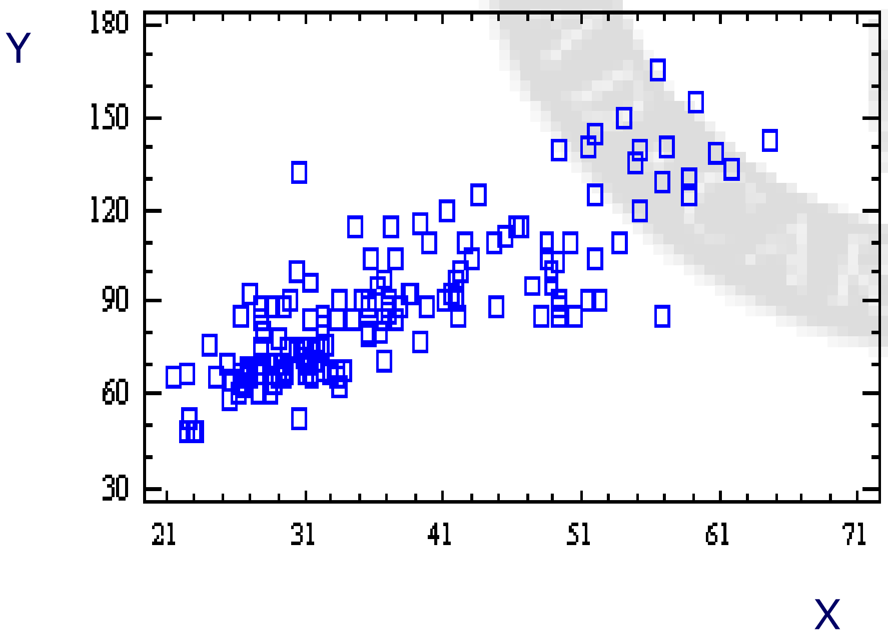
$$\mathrm{E}[\mathrm{h}(X)\mathrm{g}(Y)]\overset{\perp}{=}\mathrm{E}[\mathrm{h}(X)]\,\mathrm{E}[\mathrm{g}(Y)]$$

4.3 Covariance and correlation

Their definitions are similar to the once related to a sample dataset (Chapter II)

COVARIANCE, cov(X,Y): it is the numerical value that measures the linear relation degree bettween two variables belonging to the same population

- cov(X,Y)>0: positive linear relation between X and Y
- cov(X,Y)=0: the two variables have null linear relation
- cov(X,Y)<0: negative linear relation between X and Y



These bivariate data represent the a sample dataset belonging to a population whose variables have positive covariance

NOTE: the covariance of a sample dataset is different from the covariance of two variables of the same population. The former is computable by using the covariance formula from the descriptive statistics chapter. The second is theoretically computable if we know the probability models, but in reality we do not know it because in principle we cannot measure completely the whole population.

Although we can assume that the sample covariance may well approximate the population covariance if the size of the dataset is enough big. This is the basic concept of doing inference.

4.3 Covariance and correlation

Their definitions are similar to the once related to a sample dataset (Chapter II)

Covariance of a sample of bivariate data (Chapter II)

Sample mean of x

Sample mean of y

$$\mathrm{cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n}$$

Sample mean of the crossing products

If we consider the whole population we have to substitute the sample means by the random expectatins

$$\begin{aligned}\mathrm{cov}(X,Y) &= E\{[X-E(X)][Y-E(Y)]\}\\ &= E(XY)-E(X)E(Y).\end{aligned}$$

We can compute it from $f_X(x)$ and $f_Y(y)$

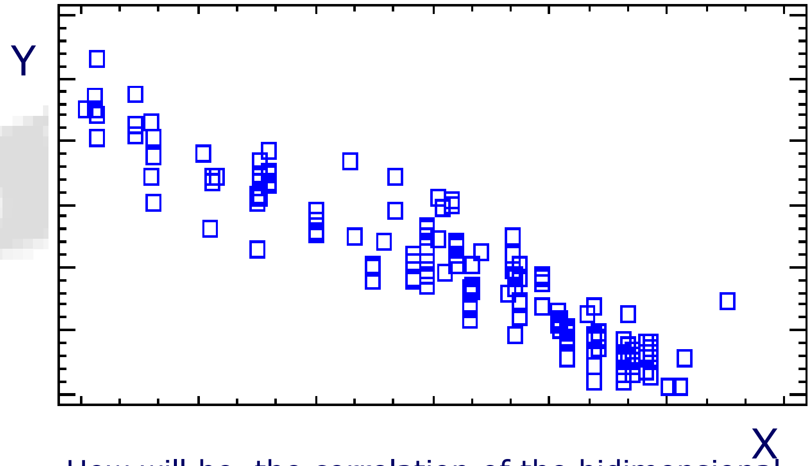
Difficult to compute. It is not enough to know $f_X(x)$ and $f_Y(y)$

Correlation, corr(X,Y), ρ : it is an adimensional numeric value that measures the linear relation degree between two variables belonging to the same population.

$$\rho = \operatorname{corr}(x,y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}$$

- Its value belongs to the interval [-1,1]
- corr(X,Y)>0: positive linear relation between X and Y
- corr(X,Y)=0: the two variables have null linear relation
- corr(X,Y)<0: negativelinear relation between X and Y

Same interpretation as for the case of the covariance



How will be the correlation of the bidimensional population from which these data where taken?

4.3 Covariance and correlation

A usual way to show information about covariance and correlation is by using matrices.

Covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \sigma_y^2 \end{bmatrix}$$

Correlation matrix

$$R = \begin{bmatrix} 1 & \mathrm{corr}(x,y) \\ \mathrm{corr}(y,x) & 1 \end{bmatrix}$$

The expectation is a sum and therefore it is a linear operator

$$E(a+bX) = a + bE(X)$$

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

$$\begin{aligned}\mathrm{Var}(a+bX) &= E\left\{[a+bX-E(a+bX)]^2\right\} \\ &= E\left[(a+bX-a-bE(X))^2\right] \\ &= E\left[b^2(X-E(X))^2\right] \\ &= b^2\mathrm{Var}(X),\end{aligned}$$

Similar to the sample mean and the sample variance

If X and Y are independent...

$$\mathrm{var}(aX + bY) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y)$$

$$\mathrm{var}(aX - bY) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y)$$

Why?

These results are theoretically interensting and useful for next chapters