UD 4

PROBABILITY DISTRIBUTIONS: discrete distrib.

Types of RANDOM VARIABLES

- Nature
  - QUALITATIVE
  - CUANTITATIVE
- Number of characteristics
  - ONE-DIMENSIONAL
  - K-DIMENSIONAL
- Set of values
  - DISCRETE
  - CONTINUOUS

Types (according to the set of values):

- **Discrete:** The sample space is formed by a finite number of values (or numerable infinite).

These values are usually the result of “counting” something:

X = number of defective pieces in a sample (finite)

X = number of throwings of a dice until obtaining for the first time 6 (numerable infinite)

X = number of readings in a diskette until it has an error (numerable infinite)

It can also be a code:

X= gender of a person E={0 (woman), 1 (man)}

X= supplier E={1 (supplier A), 2 (supplier B), 3 (supplier C)}

- **Continuous:** They are the result of measuring or evaluating a magnitude that takes values in a continuous scale. They can take any of the infinite values possible from a finite interval (e.g. uniform variable (a, b) or infinite (e.g. Normal variable)

X= weight of a person

X= electric power consumption of a machine

X= duration of a light bulb

…

In general, values of random variables can be obtained:

discrete random variables: by counting

continuous random variables: by measuring

Probability to find a person with a height of 170 cm?

DISCRETE DISTRIBUTIONS

Example: No. of vehicles stopped in a cycle of traffic lights

Probability Function (mass function): $P_X(x) = P(X=x) \quad \forall x \in E$

Nº v./c.	x	Nº data	$P_X(x)$	$P(X \leq x)$
2	0	1	0.1	0.1
5	1	2	0.2	0.3
0	2	3	0.3	0.6
1	3	2	0.2	0.8
3	4	1	0.1	0.9
1	5	1	0.1	1
2	$\geq$ 6	0	0	1
3		$\Sigma = 10$		
2				
4				

Probability Function

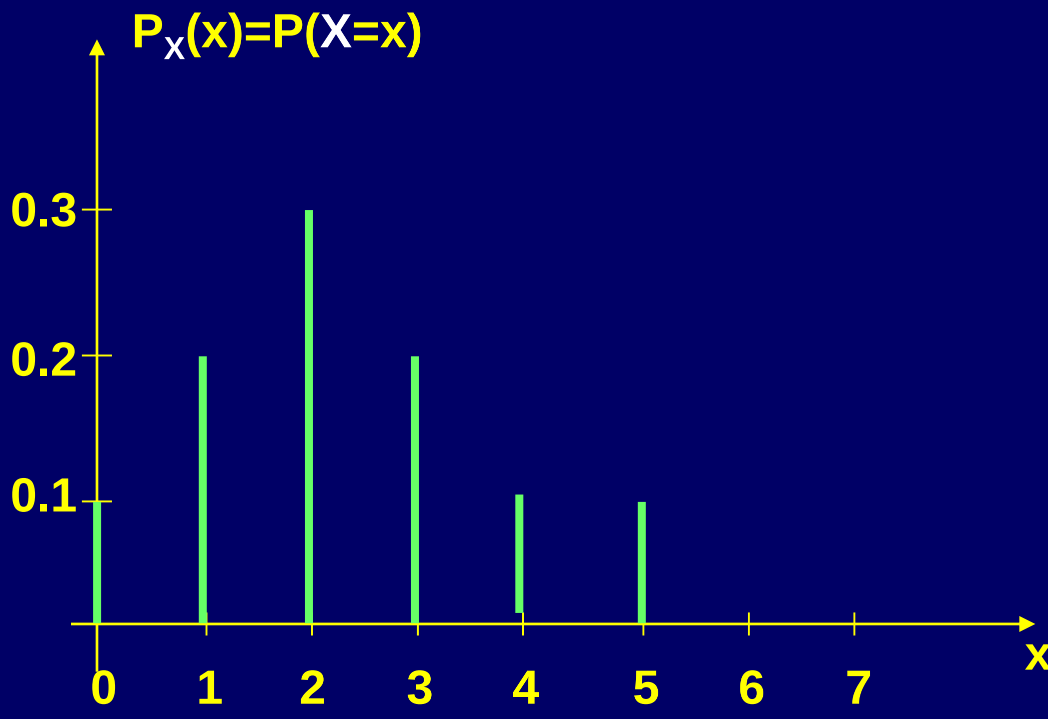
Distribution Function

DISCRETE DISTRIBUTIONS

Example: No. of vehicles stopped in a cycle of traffic lights

Probability Function: $P_X(x)=P(X=x) \qquad \forall x \in E$

x	$P_X(x)$
0	0.1
1	0.2
2	0.3
3	0.2
4	0.1
5	0.1
≥ 6	0



- **Characterization of <u>discrete variables:</u>**

PROBABILITY FUNCTION

X= sum of points obtained by throwing two dices

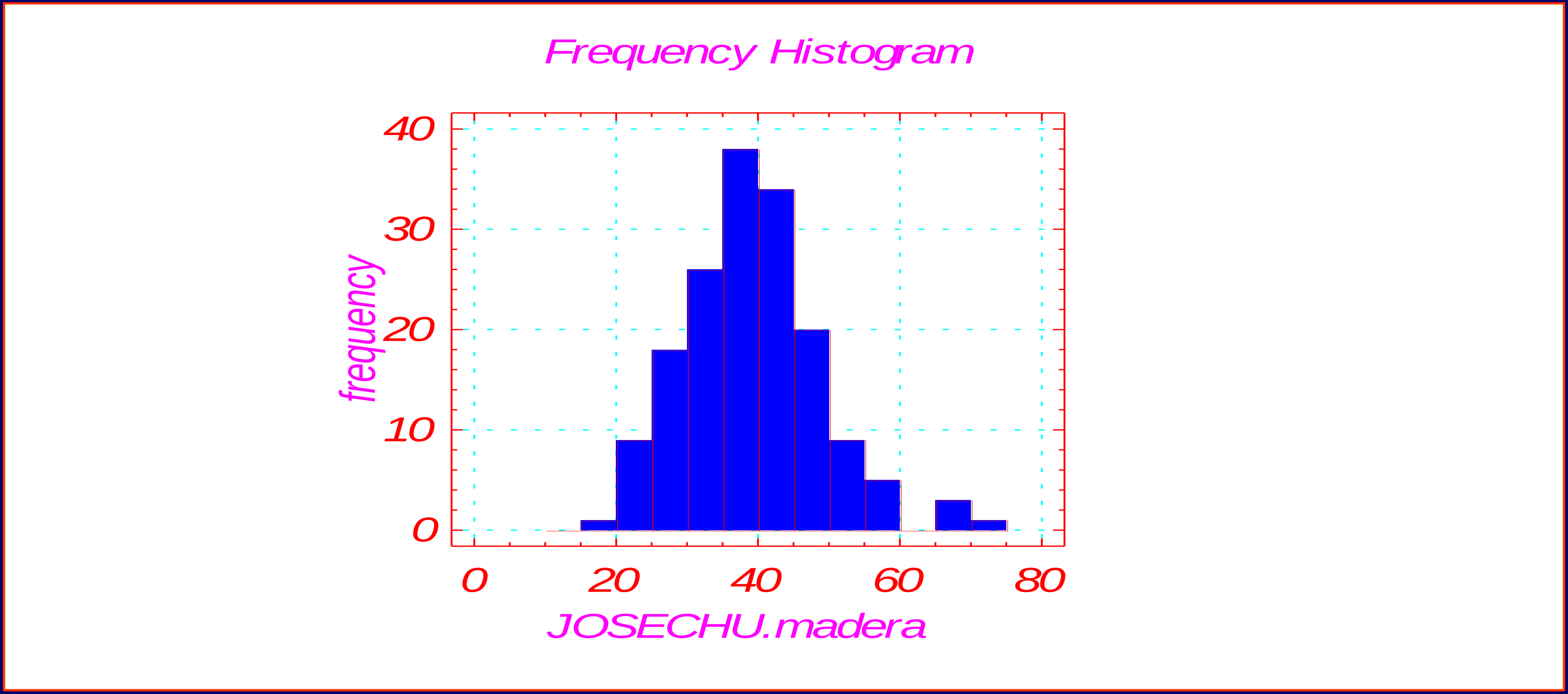
		D	A	D	O		1
		1	**2**	**3**	**4**	**5**	**6**
D	**1**	2	3	4	5	6	7
A	**2**	3	4	5	6	7	8
D	**3**	4	5	6	7	8	9
O	**4**	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

probabilities

x_i	$P(x_i)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

HISTOGRAMS

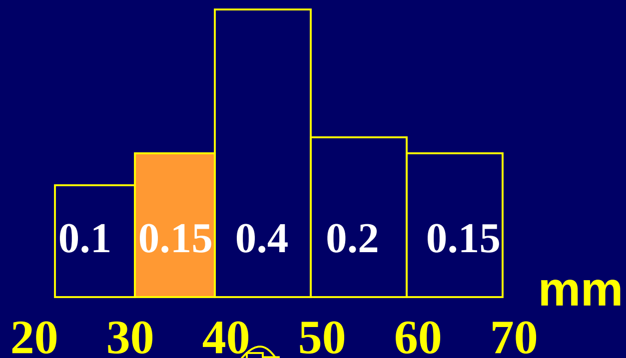
It is a graphical representation of one set of data (minimum 40-50 data) (frequency diagram)



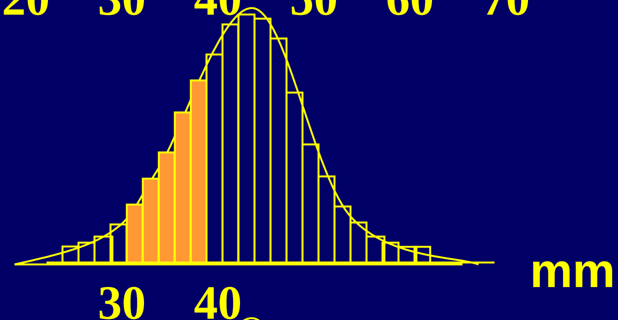
Continuous random variables: Density function

Example: length (mm) of a piece manufactured

Sample=100

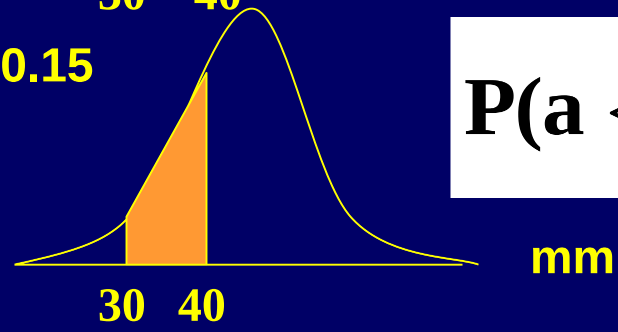


Dark area = 0.15



Probability (30<X<40) = 0.15

Dark area = 0.15



$$\mathbf{P(a < X \leq b) = \int_a^b f(x)dx}$$

- **Characterization of a <u>Continuous Variable</u>:**

DENSITY FUNCTION

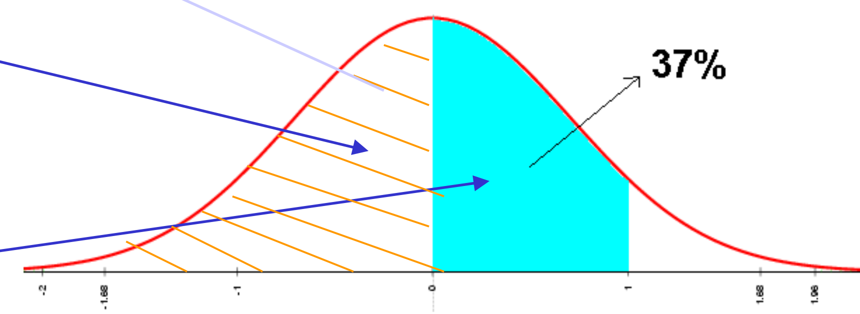
- **Continuous variables** (*the probability of a certain value is zero*, but not necessarily the probability within an interval*)*

Are characterized by means of the density function f(x) that allows calculation of probabilities in any interval

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

$$\forall x \in \Re \quad P(X \le x) = \int_{-\infty}^{x} f(x)dx$$

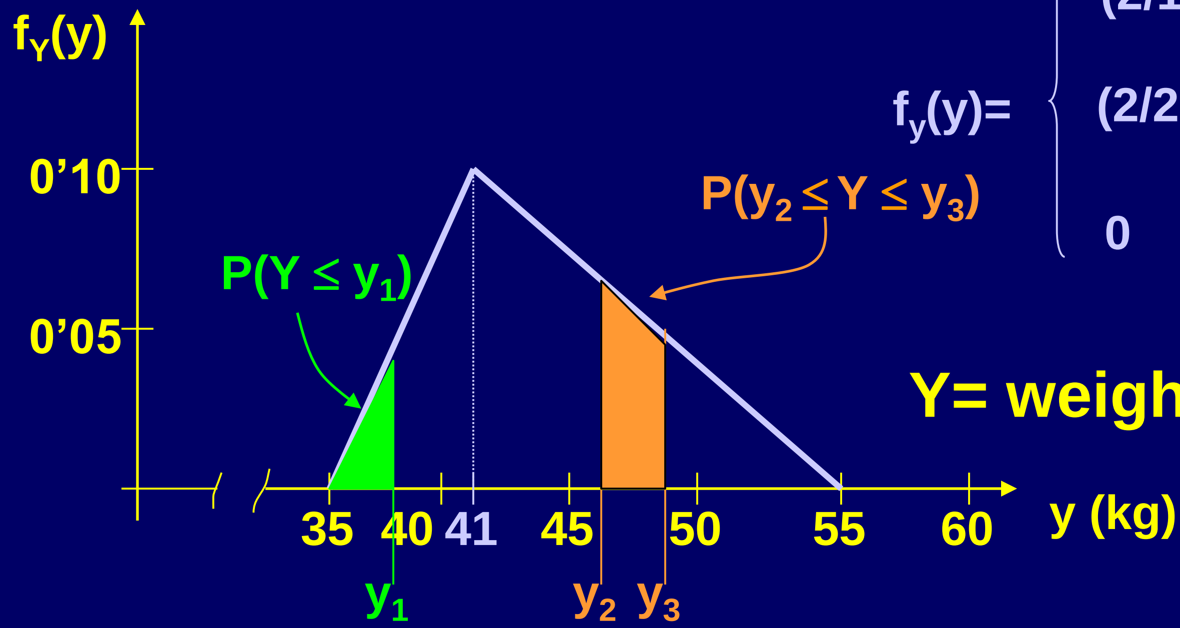
$$\forall [a,b] \in \Re \quad P(X \in [a,b]) = \int_{a}^{b} f(x)dx$$



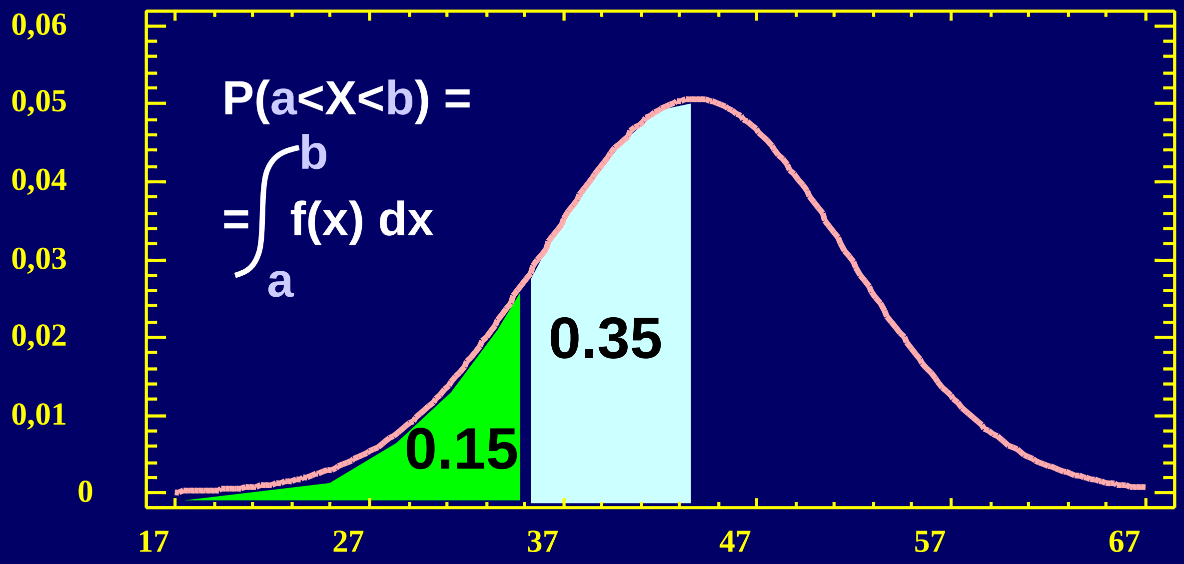
Probability in an interval: area under the curve of the density function and between the limits of the interval.

- P(X=x)=0 , but around that value there is a **density of probability** f(x), and the probability of a differential interval dx around x is: P(dx)=f(x)·dx

$$f_y(y) = \begin{cases} (2/120)(y-35) & 35 \le y \le 41 \\ (2/280)(55-y) & 41 \le y \le 55 \\ 0 & y \notin (35;55) \end{cases}$$



Y= weight of packages



DENSITY FUNCTION

Weight (kg)

MATHEMATICAL EXPECTATION. MOMENTS:

Expectation to earn money by betting 20 € in the primitive lottery?

X	RANDOM VARIABLE
h(X)	FUNCTION OF X
E(h(X))	MATHEMATICAL EXPECTATION (IDEALIZATION OF ARITMETIC MEAN OR AVERAGE).

IF X IS DISCRETE:

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

IF X IS CONTINUOUS:

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

MOMENTS WITH RESPECT TO THE ORIGIN OF ORDER u OF THE RANDOM VARIABLE X : $E(X^u)$

MEAN OF THE DISTRIBUTION: Average (mean) = m_X = E(X)

PROPERTY OF THE MEAN: $E(a_0+a_1X_1+ \dots +a_nX_n)=a_0+a_1E(X_1)+ \dots +a_nE(X_n)$

Parameters of position

- **Average, mean value or mathematical expectation**

a) Discrete variables

$$\mathrm{E(x) = m = \sum x_i P(x_i)} \quad \forall \mathrm{x_i} \text{ of sample space}$$

If X is the number of points obtained by throwing a dice

E(x)=1· 1/6+ 2 ·1/6+ 3 ·1/6+4 ·1/6+ 5 ·1/6+ 6 ·1/6 =21/6

b) Continuous variables

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

Example: $\quad If\ \ f(x) = \frac{1}{2}x \quad \forall x \in [0,\ 2] \quad and \quad 0\ in\ the\ rest$

$$E(x) = \int_{-\infty}^{+\infty} x\, f(x)dx = \int_{-\infty}^{0} x.0.dx + \int_{0}^{2} x.(\frac{1}{2}x)dx + \int_{2}^{+\infty} x.0.dx = \int_{0}^{2} x.(\frac{1}{2}x)dx = \frac{8}{6}$$

What "expectation" is there to earn money by betting 5 € to a number of roulette?

- Example:

Calculate the mean amount won per play, in a game that consists of paying in each play so many euros as points obtained if the throwing of a dice is even, and receiving so many euros as points obtained if the throwing is odd. In average, how much would we win in this game?

Solution:

X= No. points by throwing the dice E=(x_1=1, x_2=2, x_3=3, x_4=4, x_5=5, x_6=6)

$g(x_i)=(+x_i)$ if i=odd, $g(x_i)=(-x_i)$ if i i=even

$$E(g(x)) = 1\frac{1}{6} + (-2)\frac{1}{6} + 3\frac{1}{6} + (-4)\frac{1}{6} + 5\frac{1}{6} + (-6)\frac{1}{6} = -0{,}50 \text{ euros}$$

In average, 0.50 euros are lost per play

Central moments of order u of the distribution: $E(X-m)^u$

VARIANCE OF THE DISTRIBUTION: Variance = $\sigma_X^2 = E(X-m)^2 > 0$

PROPERTIES OF THE VARIANCE (for any kind of distribution):

1.- $\sigma^2(aX) = a^2\,\sigma^2(X)$ $\qquad$ $\sigma^2(b+aX) = a^2\,\sigma^2(X)$

2.- IF X, Y ARE 2 COMPONENTS OF A TWO-DIMENSIONAL VAR.:

$$\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y) \pm 2\,\mathrm{cov}(X,Y)$$

$$\mathrm{cov}(X,Y) = \sigma_{X,Y}^2 = E[(X-m_X)(Y-m_Y)]$$

$$\sigma^2(aX \pm bY) = a^2\,\sigma^2(X) + b^2\,\sigma^2(Y) \pm 2\,a\,b\,\mathrm{cov}(X,Y)$$

3.- IF X, Y ARE INDEPENDENT :

$$\mathrm{cov}(x,y) = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1}$$

$$\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y)$$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

QUESTION: If X is the weight of one potato and $\sigma^2(X)$ is the variance, what is the variance of a pack with 5 potatoes?

Dispersion parameters

- **Variance:** Measure of dispersion of the variable around the average

$$\sigma^2 = D^2(x) = E(x-m)^2 \quad = E\left(x^2 + m^2 - 2\cdot x\cdot m\right) = E\left(x^2\right) + m^2$$
$$-2\cdot m\cdot m = E\left(x^2\right) - \left[E(x)\right]^2$$

Calculation:

$$\sigma^2 = D^2(x) = \sum (x_i - m)^2 P(x_i) \quad \textit{in discrete variables}$$

$$\sigma^2 = D^2(x) = \int_x (x-m)^2 f(x)\,dx \quad \text{in continuous variables}$$

Standard deviation:

$\sigma = +\sqrt{\sigma^2}$ $\qquad$ $D(x) = +\sqrt{D^2(x)}$

SKEWNESS COEFFICIENT OF THE DISTRIBUTION:
CENTRAL MOMENT OF ORDER 3.

$$CA = \frac{E(X-m)^3}{\sigma^3}$$

KURTOSIS COEFFICIENT OF THE DISTRIBUTION:
CENTRAL MOMENT OF ORDER 4

$$CC = \frac{E(X-m)^4}{\sigma^4}$$

DISCRETE DISTRIBUTIONS

BINOMIAL DISTRIBUTION

- A: EVENT OF PROBABILITY P ASSOCIATED TO A CERTAIN RANDOM EXPERIMENT.

A={obtain a "4"}   P=P(A)= 1/6

- WE CARRY OUT n INDEPENDENT REPETITIONS OF THE EXPERIMENT.

n=5

- RANDOM VARIABLE X: NUMBER OF TIMES THAT EVENT A OCCURS (0, 1, 2 ,3, … , n).

Number of times that we obtain 4 after throwing 5 dices

X FOLLOWS A BINOMIAL DISTRIBUTION THAT DEPENDS ON TWO PARAMETERS: n AND P. $X \sim B(n,P)$

$X \sim B(5, 1/6)$

PROBABILITY FUNCTION OF X IS:

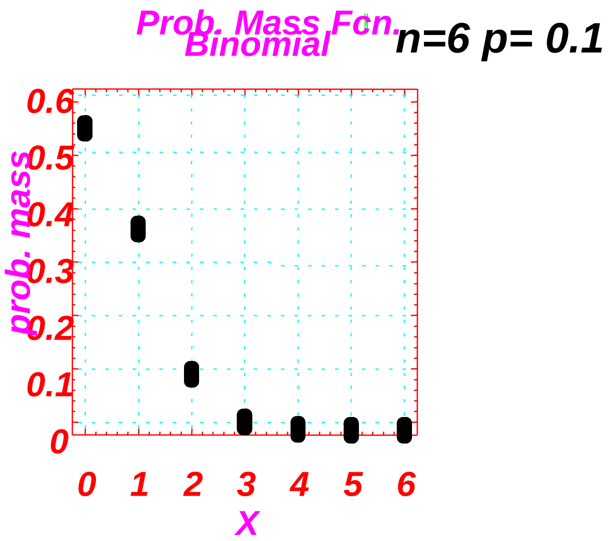
$$P(X=x) = \binom{n}{x} P^X (1-P)^{n-x}$$

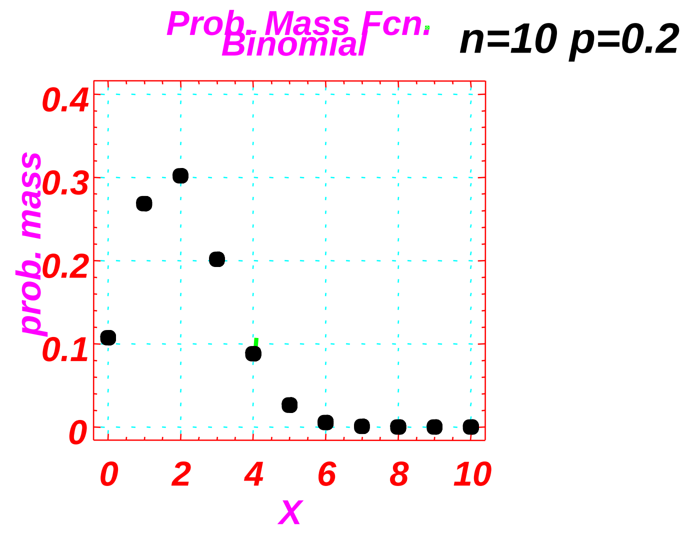
$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

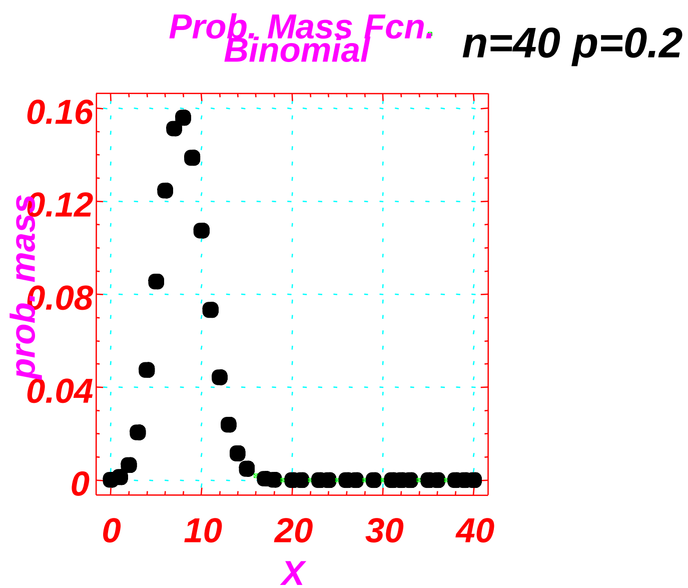
$$\mathbf{P(X=3)=\binom{5}{3}\cdot 0.17^3\cdot(1-0.17)^{5-3}=0.0322}$$

$$E[X] = np$$

$$\sigma^2[X] = np(1-p)$$


Prob. Mass Fcn.
Binomial
n=6 p= 0.1
prob. mass
0 0.1 0.2 0.3 0.4 0.5 0.6
0 1 2 3 4 5 6
X


Prob. Mass Fcn.
Binomial
n=10 p=0.2
prob. mass
0 0.1 0.2 0.3 0.4
0 2 4 6 8 10
X


Prob. Mass Fcn.
Binomial
n=40 p=0.2
prob. mass
0 0.04 0.08 0.12 0.16
0 10 20 30 40
X

PROBLEM:

Probability to obtain at least one 5 in six consecutive throwings of a dice.

PROBLEM:

One hunter has a probability of 40% to hit the target. What is the probability to hit the target with 3 shots (or less)?

POISSON DISTRIBUTION

The Poisson distribution is the limit of a Binomial distribution when:

- $n \to \infty$, $p \to 0$
- and the product $n \cdot p$ tends to a constant λ

The probability function of a Poisson distribution of parameter λ is:

$$\mathrm{Prob}(\mathrm{Poisson}(\lambda) = x) = \lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}$$

AVERAGE OF THE DISTRIBUTION: $E(X) = \lim_{np\to\lambda} np = \lambda$

VARIANCE : $\sigma^2(X) = \lim_{\substack{np\to\lambda \\ p\to 0}} np(1-p) = \lambda$

Poisson distributions are useful when we measure the number of occurrences of an event in a certain time or space.

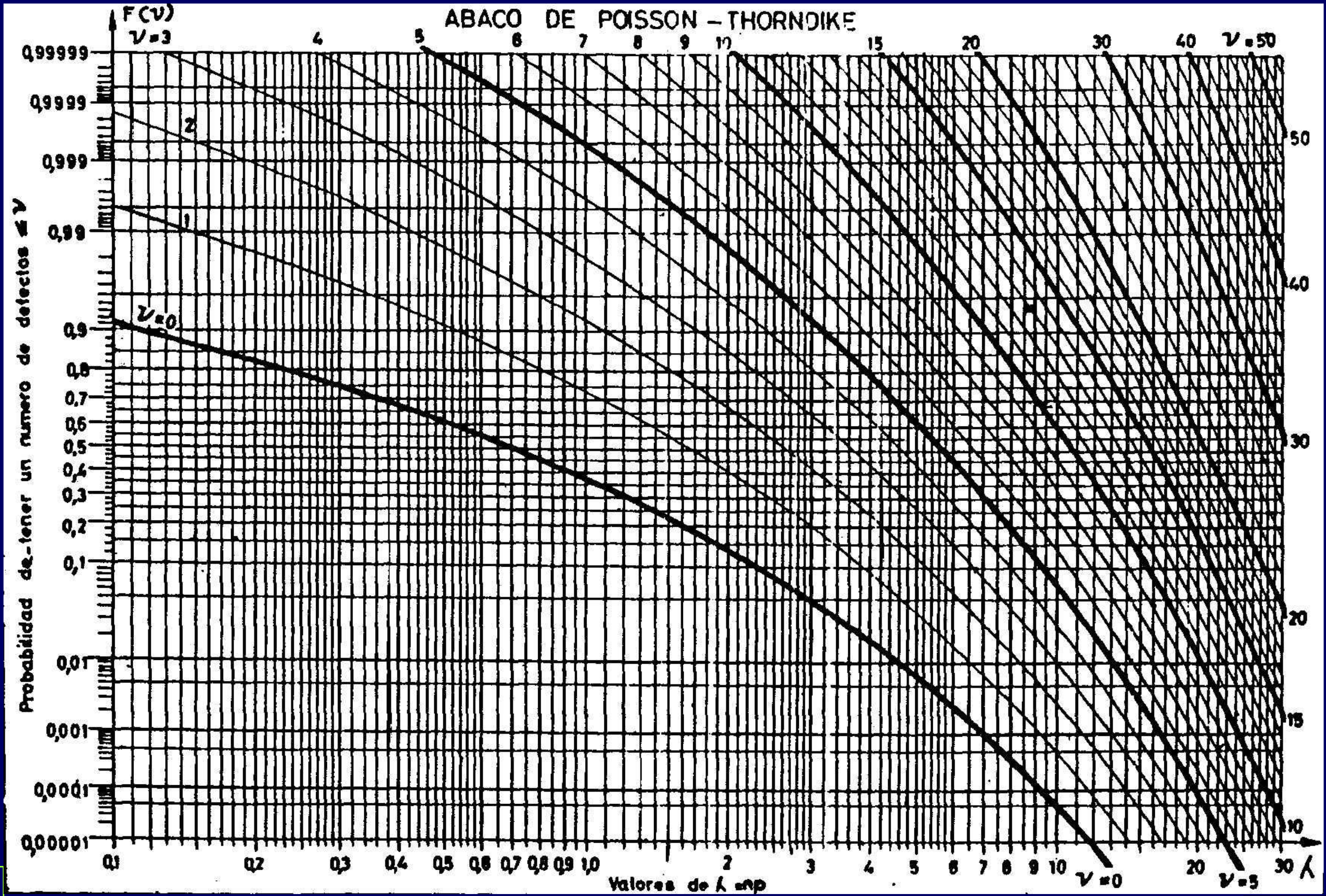
$X_i \sim Ps(\lambda_i)$; $Y = \Sigma X_i$ (independent) $\longrightarrow$ $Y \sim Ps(\lambda = \Sigma \lambda_i)$

EXAMPLES:

- Number of accidents per weekend
- No. of cars passing per min. through a bridge
- Number of errors in a printed page of book
- Number of errors in a program code

ABAC OF POISSON

CALCULATES P(X ≤ c) BEING X ~ POISSON (λ)



PROBLEMS

It is known that there are in average 3 mortal accidents per day in Spanish roads (excluding weekends).

a) What is the probability of occurring no mortal accidents next Tuesday?

b) What is the probability of occurring more than 7 mortal accidents next Tuesday?

Assuming that accidents in the different days of the week are independent and that the average number of mortal accidents in the weekend (Saturday and Sunday) is 15,

c) what is the probability of occurring less than 25 mortal accidents next week?

d) what is this probability, knowing that 8 mortal accidents have already occurred during the first two days?

UD 4 - part 2

CONTINUOUS DISTRIBUTIONS:

uniform, exponential

PROBLEM:

One train leaves from a station every 20 minutes. One traveler arrives at the station to take the train.

Variable X: waiting time (time between the traveler arrives and the train departure).

a) What is the minimum, maximum of X?

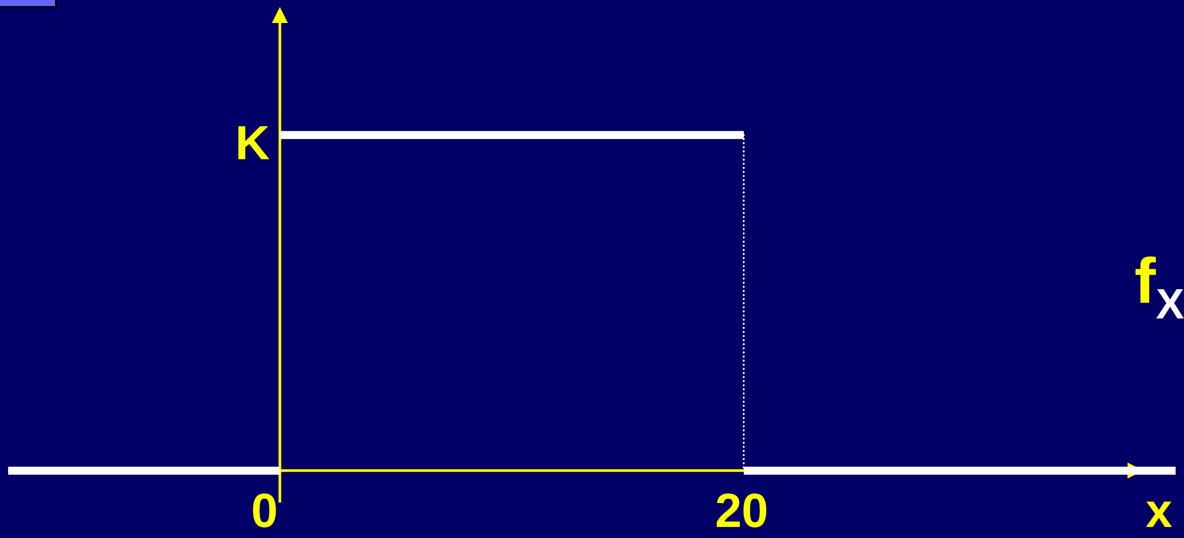
b) Draw a histogram for the values of X.

c) Calculate the density function of X.

d) What is the average and median of X?

e) $P(X<7)$

f) $P(X=7)$



$$f_X(x)=\begin{cases} 0 & x \le 0 \\ 1/20 & x \in (0,20) \\ 0 & x \ge 20 \end{cases}$$

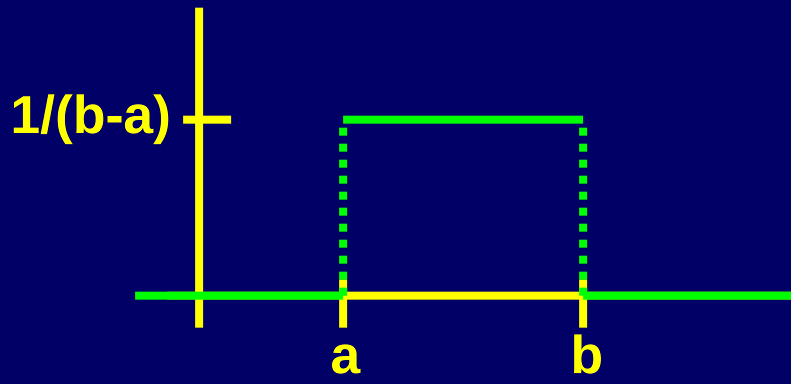
PROBLEM:

I have a file called "exercises.doc" in my laptop, but I cannot find it. I carry out a search and write down the <u>time</u> taken by the computer to find the file. If this operation is repeated many times, what would be the distribution of this variable?

UNIFORM DISTRIBUTION

One random variable is UNIFORM U(a,b) if the density function is constant.

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



AVERAGE :

$$\mu = \frac{a+b}{2}$$

VARIANCE:

$$\sigma^2 = \frac{(b-a)^2}{12}$$

PROBLEM:

Every time that there is an accident in a factory, they record X: time passed since the last accident. The average of X is 15 days.

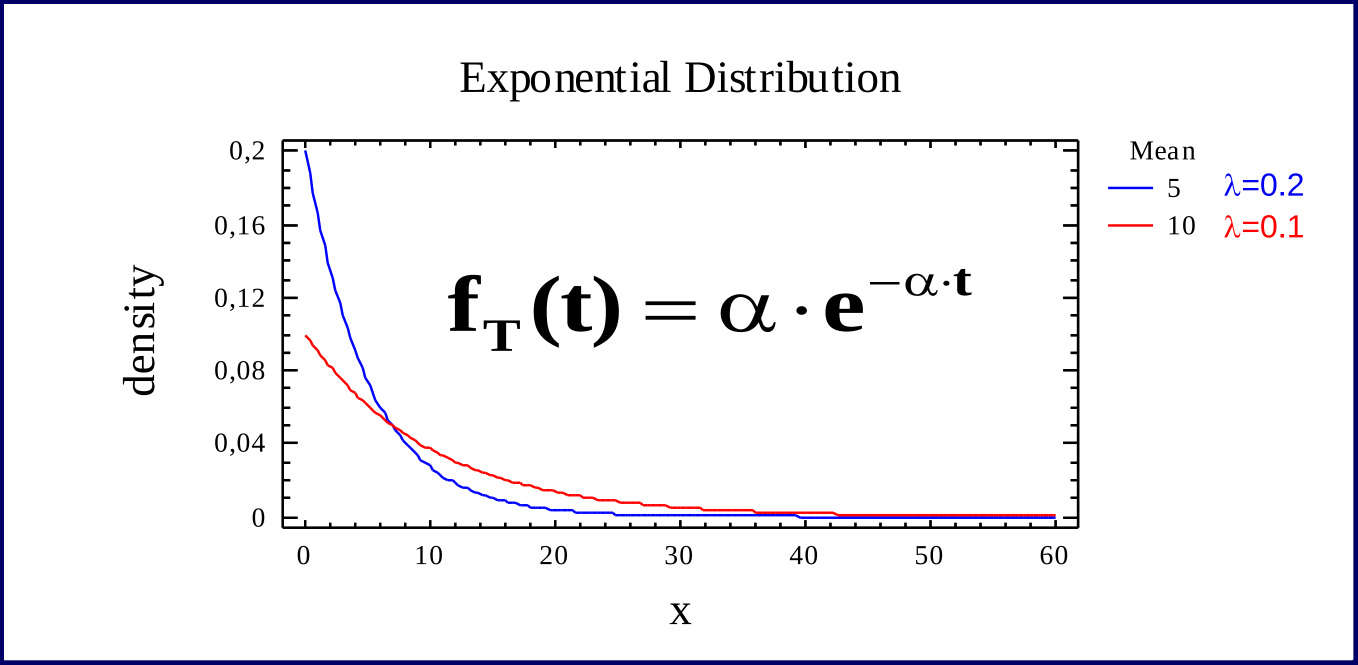
a) What is the minimum, maximum of X?
b) Draw a histogram for the values of X.
c) How would be the density function?
d) P(X=15)
e) P(X<15)

EXPONENTIAL DISTRIBUTION

Time until a Poisson event (accident, failure...)
- **X: time of operation of a light bulb until breakdown**
- **X: time of operation of a computer until failure**

Time between two Poisson events:
- **X: time between two consecutive failures of a printer**
- **X: time between two accidents in a factory**

In a shop: time of waiting until being attended



T follows an exponential distribution of parameter α

DENSITY FUNCTION:

$$f_T(t) = \alpha \cdot e^{-\alpha \cdot t}$$

$$\int_0^{\infty} \alpha \cdot e^{-\alpha t} dt = 1$$

$$P(T > t) = e^{-\alpha \cdot t} \qquad t > 0$$

$$P(T \leq t) = 1 - e^{-\alpha \cdot t}$$

AVERAGE:

$$E(T) = \int_0^{\infty} t \cdot f(t) dt = 1/\alpha$$

VARIANCE:

$$\sigma^2(T) = \int_0^{\infty} (t - m)^2 \cdot f(t) dt = \frac{1}{\alpha^2}$$

Observed values of time follow an exponential distribution if their average and standard deviation are similar.

EXERCISE: Calculate the median of the distribution:

$$\mathbf{X} \approx \mathbf{exp}(\alpha = \mathbf{3})$$

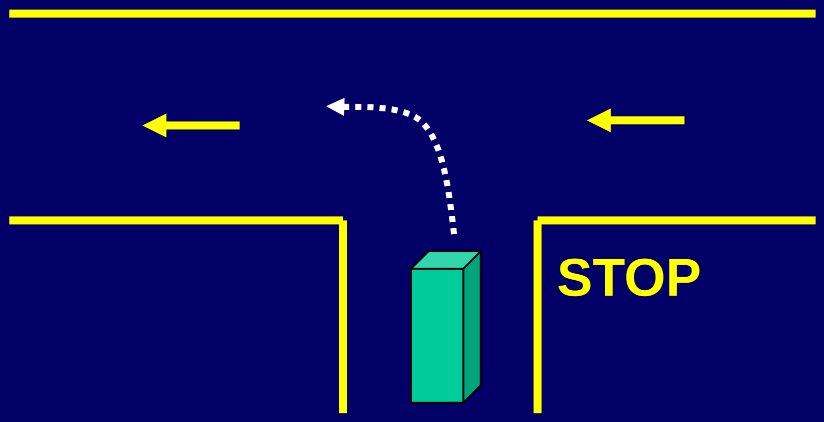
"The reliability of this TV is 80% at 2000 hours", means...

P (time until failure > 2000) = 0.8   True

P (time until failure < 2000) = 0.8   False

PROBLEM:

One car arrives at a crossroad and stops. At that time, the average number of vehicles passing per minute through the main road is 10. The driver needs 10 seconds without cars passing through to make the incorporation into the main road.



CALCULATE

a) Probability of passing more than one car in 3 min.

var. X: <u>number</u> of cars passing through in 3 min.

b) What is the distribution of the following variable?

var. T: <u>time</u> since the car stops at the crossroad until the first vehicle passes through.

c) If we know that 30 seconds ago passed the last vehicle, what is the probability of not having to wait when arriving to the crossroad?

$$P\left(T > \frac{40}{60} \middle/ T > \frac{30}{60} \right) = \ldots$$

If T : exp(α) →

The probability of a TV to work > 30 years <u>knowing</u> that has not failed in 29 years =
= probability to work > 1 year

$P(T>30 / T>29) = P(T>1)$

If failures of a machine are more likely as it ages (which often occurs), time until failure will <u>not</u> follow an exponential distribution.

If time until failures follows an exponential distribution it implies that failures are due to <u>accidental</u> causes and <u>not to aging</u>.

LACK OF MEMORY PROPETY:

THE FUTURE PERFORMANCE IS INDEPENDENT OF THE PAST PERFORMANCE:

$$P(T > t_0 + t)/T > t_0) = \frac{P(T > t_0 + t)}{P(T > t_0)} = \frac{e^{-\alpha(t_0+t)}}{e^{-\alpha t_0}} = e^{-\alpha t}$$

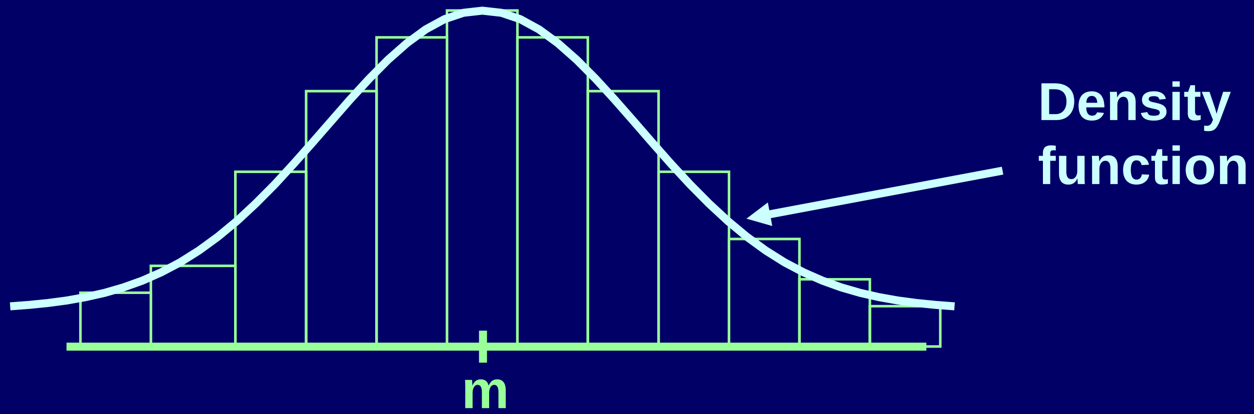
Therefore: $P(T > t_0 + t / T > t_0) = P(T > t)$

UD 4 - part 3

CONTINUOUS DISTRIBUTIONS: NORMAL distrib.

NORMAL DISTRIBUTION

Describes properly many continuous variables found in practice: most variables with a <u>symmetric</u> distribution



Typical histogram of a Normal distribution

$$X \sim N(m, \sigma)$$

$$f_X(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (-\infty < x < +\infty)$$

$E(X) = m$ $\qquad$ $\sigma(X) = \sigma$

PROPERTY OF ANY DISTRIBUTION:

If X and Y are two components of a two-dimensional random variable:

$$\sigma^2(a \cdot X \pm b \cdot Y) = a^2 \cdot \sigma^2(X) + b^2 \cdot \sigma^2(Y) \pm 2\,a\,b\,\mathrm{cov}(X,Y)$$

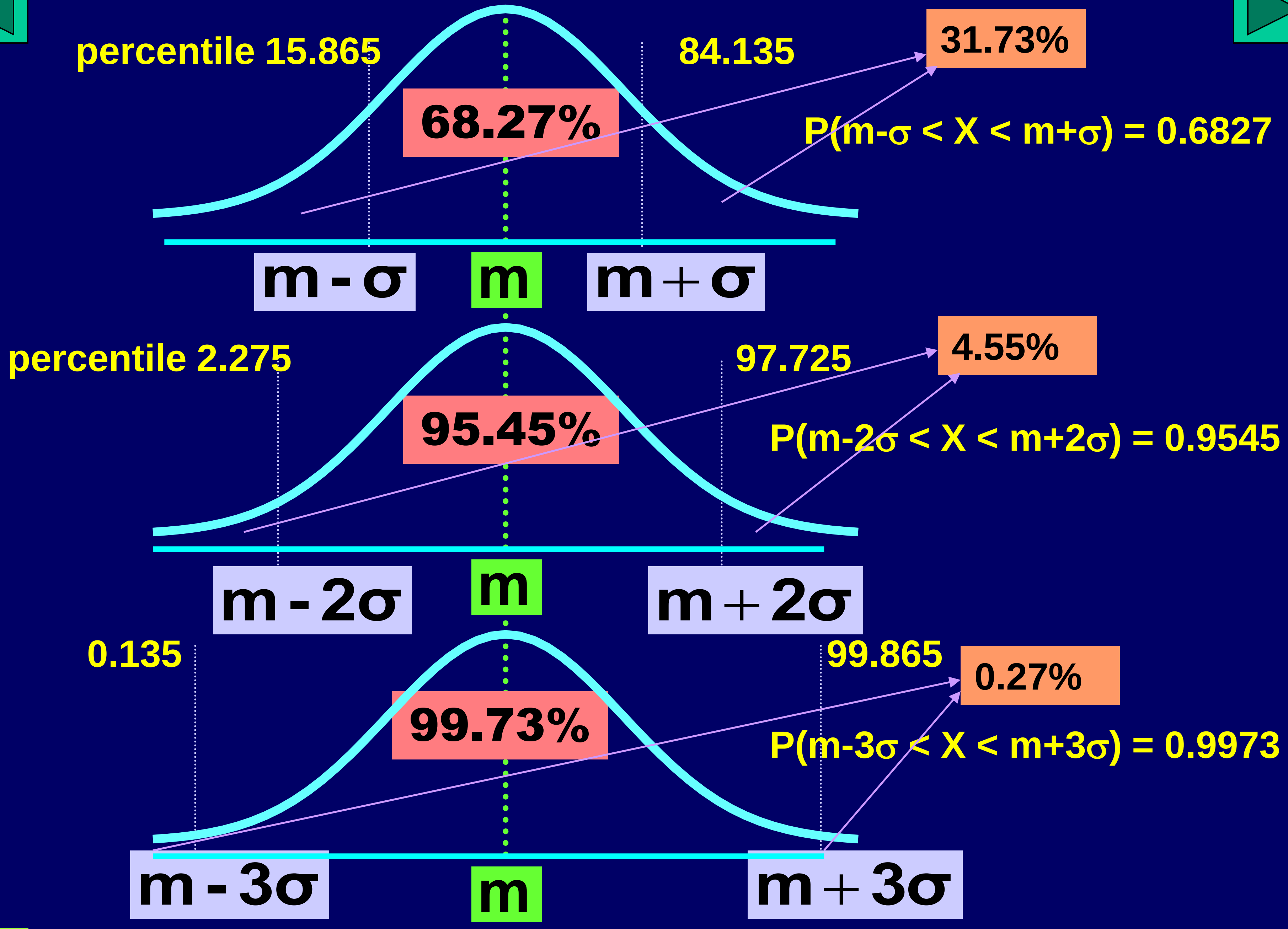
PROPERTIES OF THE NORMAL DISTRIBUTION:

1) If $X \sim N(m_x , \sigma_x)$

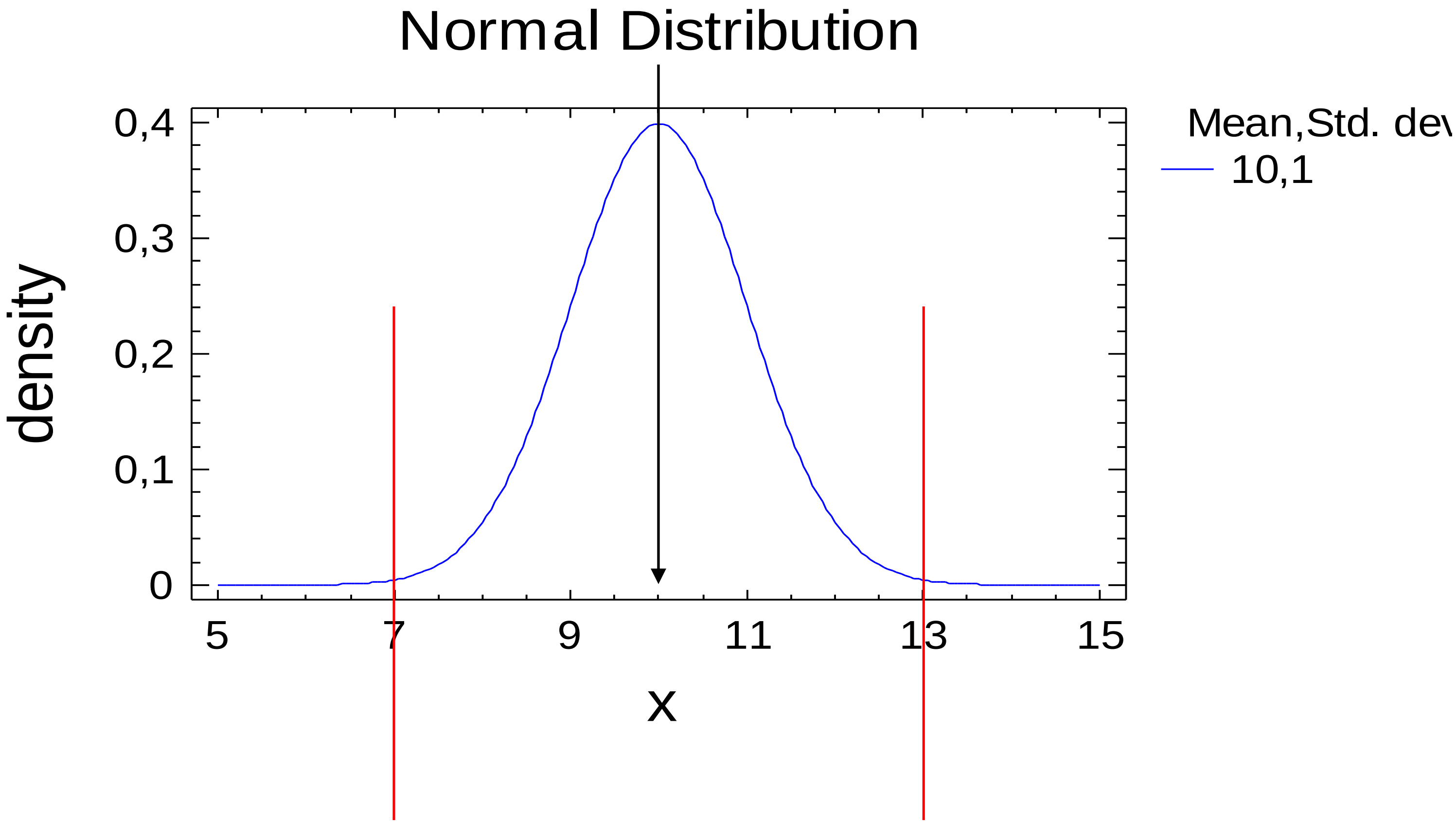
$$Y = a + bX \sim N\left(m_y = a + b\,m_x ,\ \sigma_y^2 = b^2 \sigma_x^2\right)$$

2) If $X \sim N(m_x , \sigma_x)$ and $Y \sim N(m_y , \sigma_y)$ independent

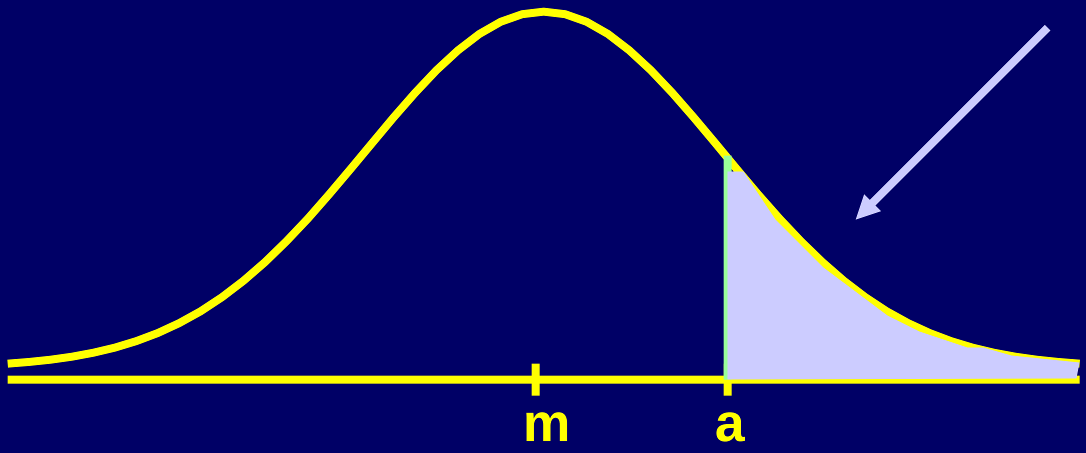
$$Z = X \pm Y \sim N\left(m_z = m_x \pm m_y ,\ \sigma_z^2 = \sigma_x^2 + \sigma_y^2\right)$$


percentile 15.865
84.135
31.73%
68.27%
P(m-σ < X < m+σ) = 0.6827
m - σ
m
m + σ
percentile 2.275
97.725
4.55%
95.45%
P(m-2σ < X < m+2σ) = 0.9545
m - 2σ
m
m + 2σ
0.135
99.865
0.27%
99.73%
P(m-3σ < X < m+3σ) = 0.9973
m - 3σ
m
m + 3σ

DENSITY FUNCTION of N (m,σ)= N(10,1)



How to calculate $P(X \sim N(m_x, \sigma_x) > a)$?



$$P(X > a) = \int_a^{+\infty} \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Cannot be solved analytically $\longrightarrow$ use table

$P(X \geqslant a) \equiv P(X > a)$ because $P(X = a) = 0$

If $X - m_x \longrightarrow E(X - m_x) = E(X) - m_x = 0$

If $\dfrac{X - m_x}{\sigma_x} \longrightarrow E\left(\dfrac{X - m_x}{\sigma_x}\right) = 0 \quad \sigma^2\left(\dfrac{X - m_x}{\sigma_x}\right) = \dfrac{\sigma_x^2}{\sigma_x^2} = 1$

$Z = \dfrac{X - m_x}{\sigma_x} \sim N(m_z = 0, \sigma_z = 1) \longrightarrow$ **STANDARD NORMAL DISTRIBUTION**

$$P(X > a) = P(X - m_x > a - m_x) = P\left(\frac{X - m_x}{\sigma_x} > \frac{a - m_x}{\sigma_x}\right) = P\left(Z > \frac{a - m_x}{\sigma_x}\right) =$$

$$P\left(Z > \frac{a - m_x}{\sigma_x}\right) = \int_{\frac{a - m_x}{\sigma_x}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Cannot be solved analytically

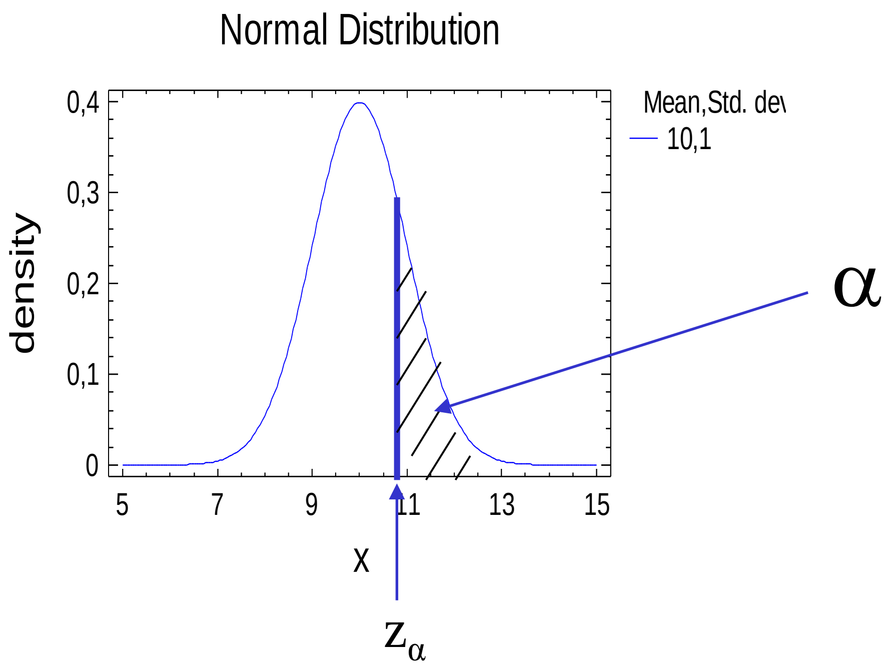
$\longrightarrow$ **Solved by numeric methods** $\longrightarrow$ **Table N(0,1)**

STANDARD NORMAL DISTRIBUTION

- The variable N(0,1) is sometimes called Z

Critical value z_α :

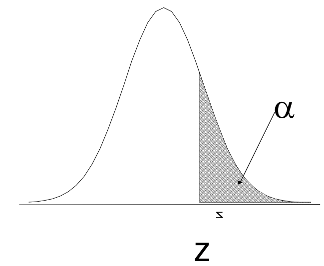
$P[N(0, 1) > z_\alpha] = \alpha$

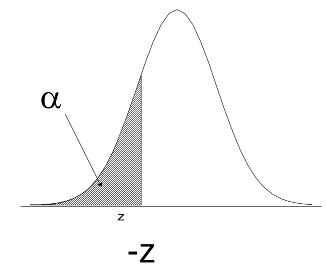


$$P(X \leq a) = P(N(0,1) \leq \frac{a-\mu}{\sigma}) = \phi(\frac{a-\mu}{\sigma})$$

STANDARD NORMAL DISTRIBUTION

z	0	1	2	3	4	5	6	7	8	9
3	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1057	0,1038	0,1020	0,1003	0,0985
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2207	0,2177	0,2148
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0.0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641

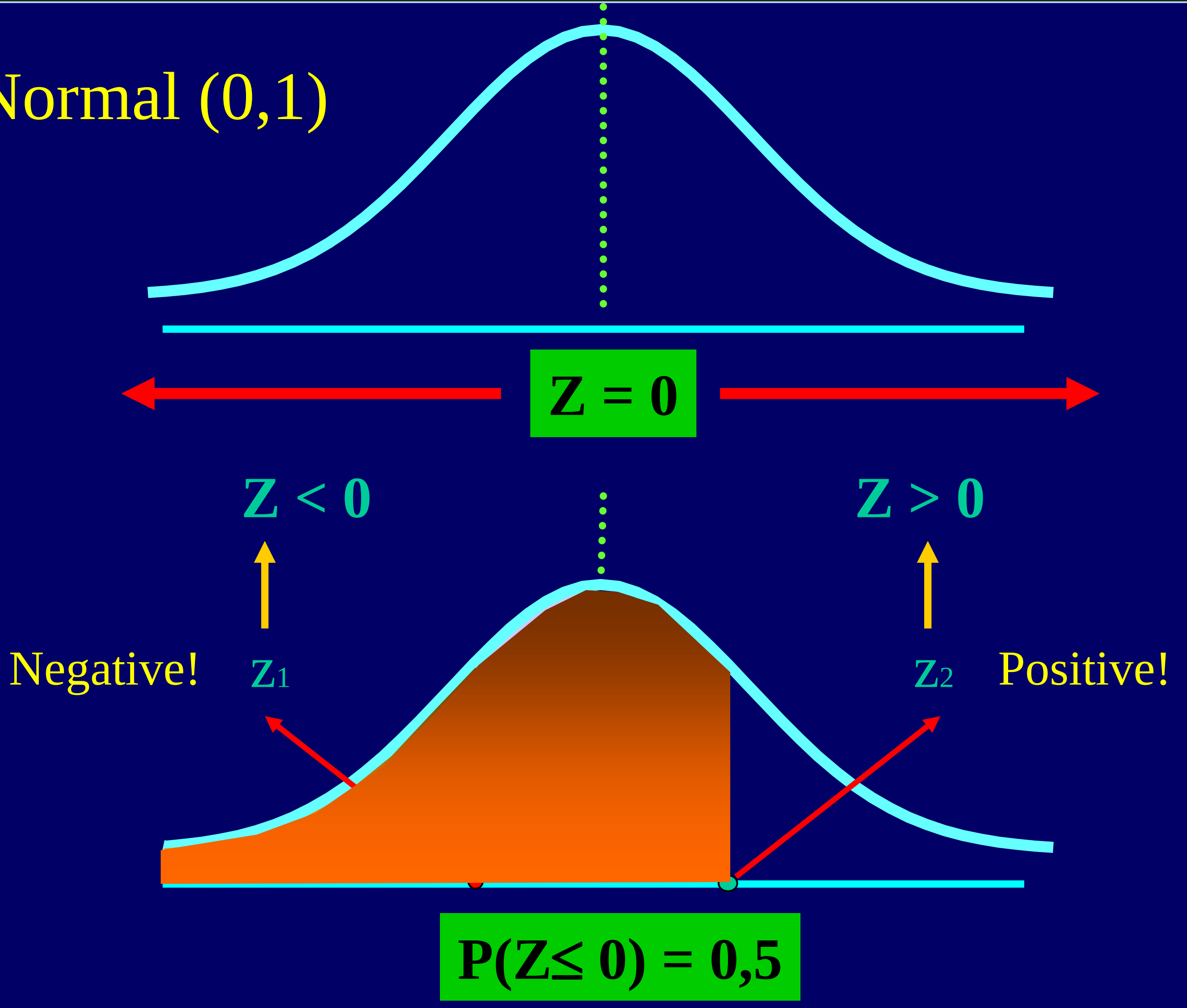




$$P(Z \leq z) = \int_{-\infty}^{z} \frac{\mathbf{1}}{\sqrt{\mathbf{2}\pi}} \cdot \mathbf{e}^{-\frac{t^2}{2}}\, dt$$

USE OF THE N(0, 1) TABLE

Z = Normal (0,1)

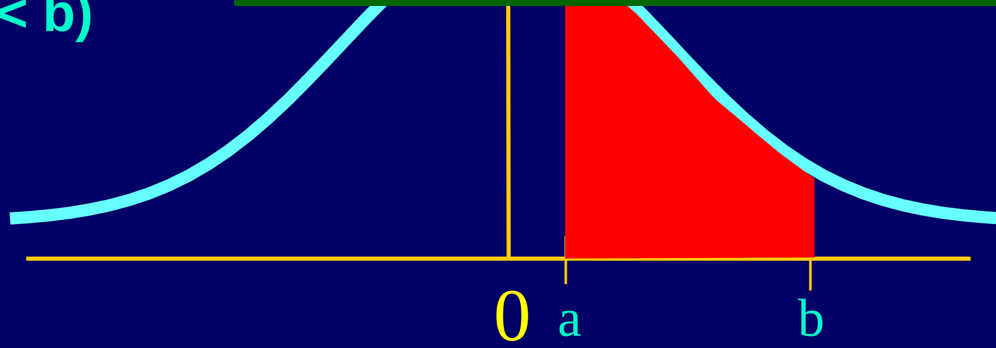


$P(Z \leq z_1) = 0,3$ $P(Z \leq z_2) = 0,7$

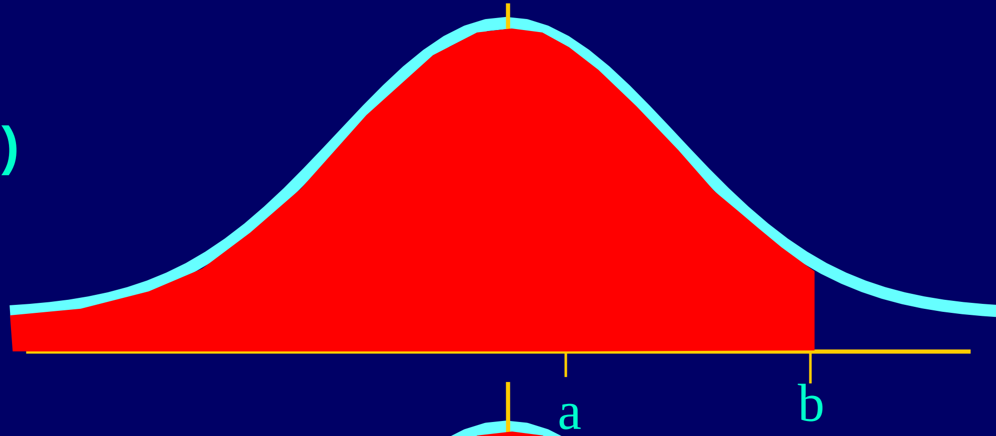
Normal (0,1)

P (a < N(0,1) < b) = P (N(0,1) < b) - P (N(0,1) < a)

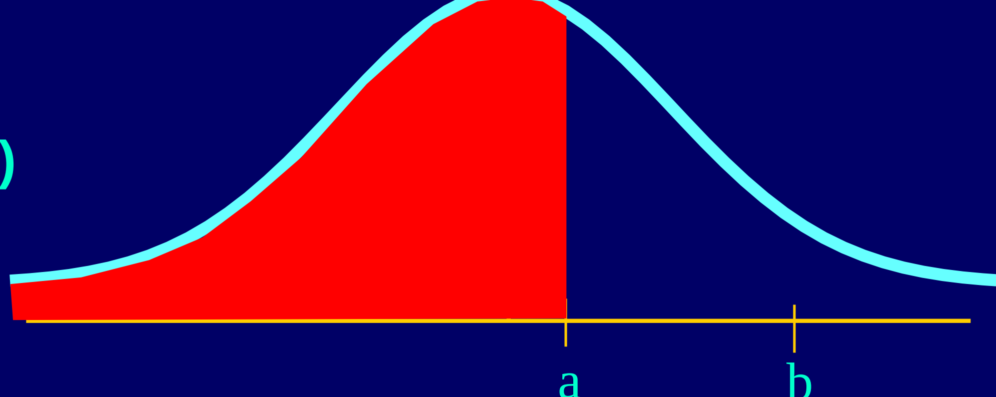
P (a < N(0,1) < b)



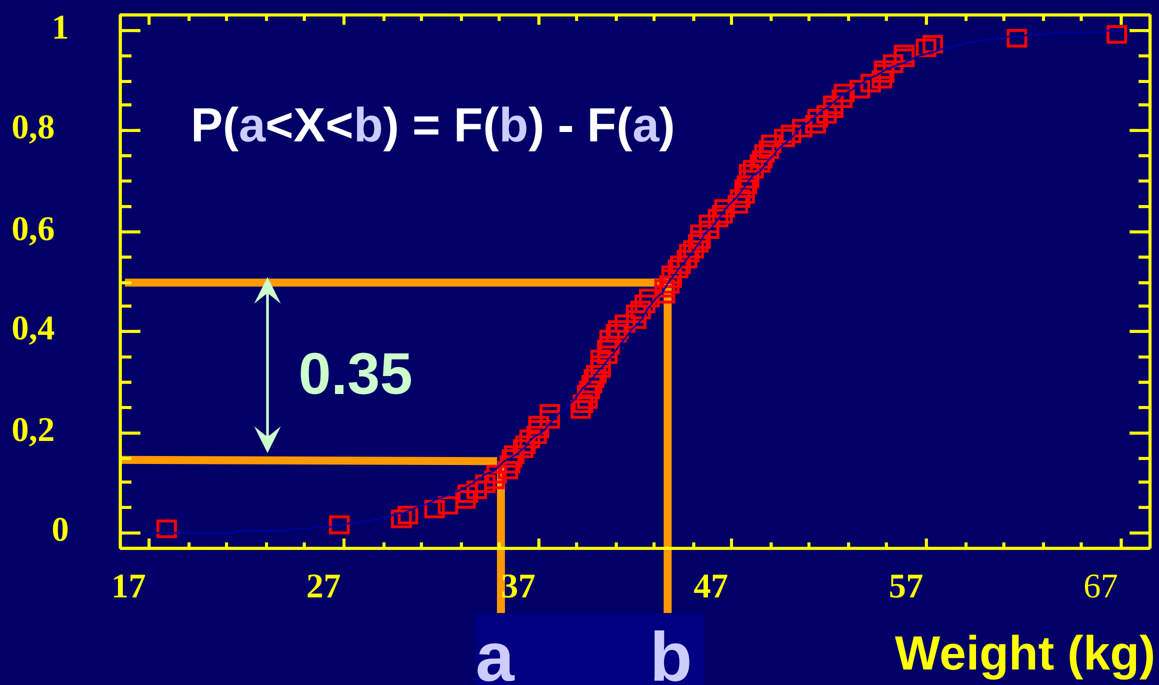
P (N(0,1) < b)



P (N(0,1) < a)

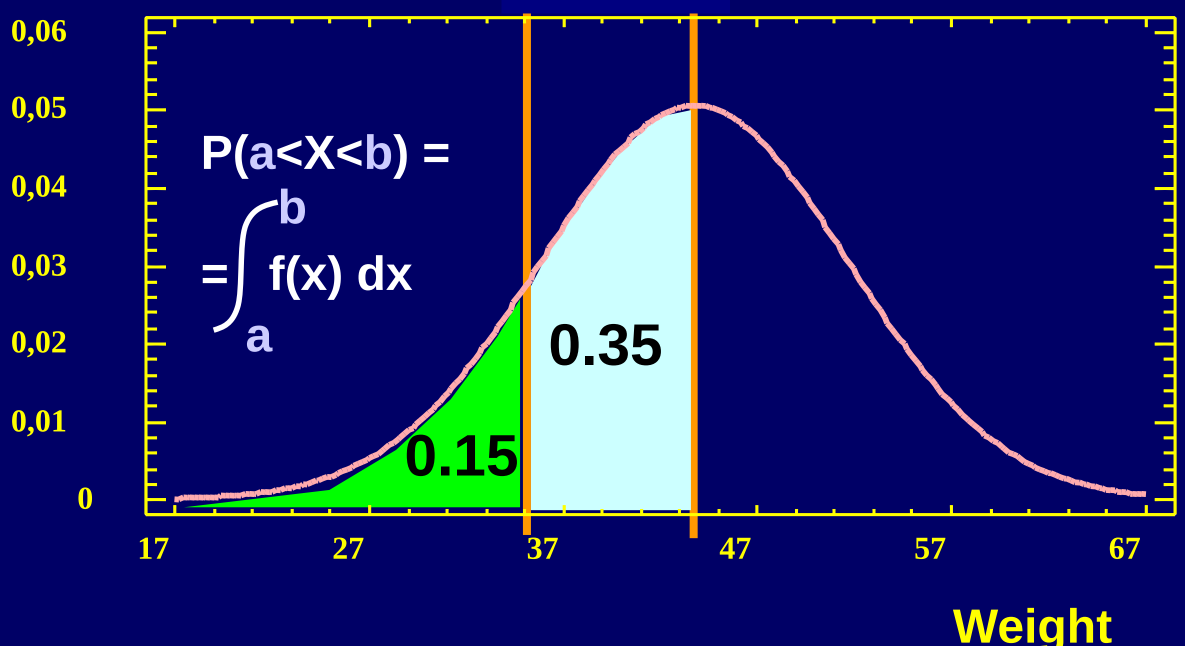


$F_R(x)=P(R\le x)$



DISTRIBUTION FUNCTION

F(x)

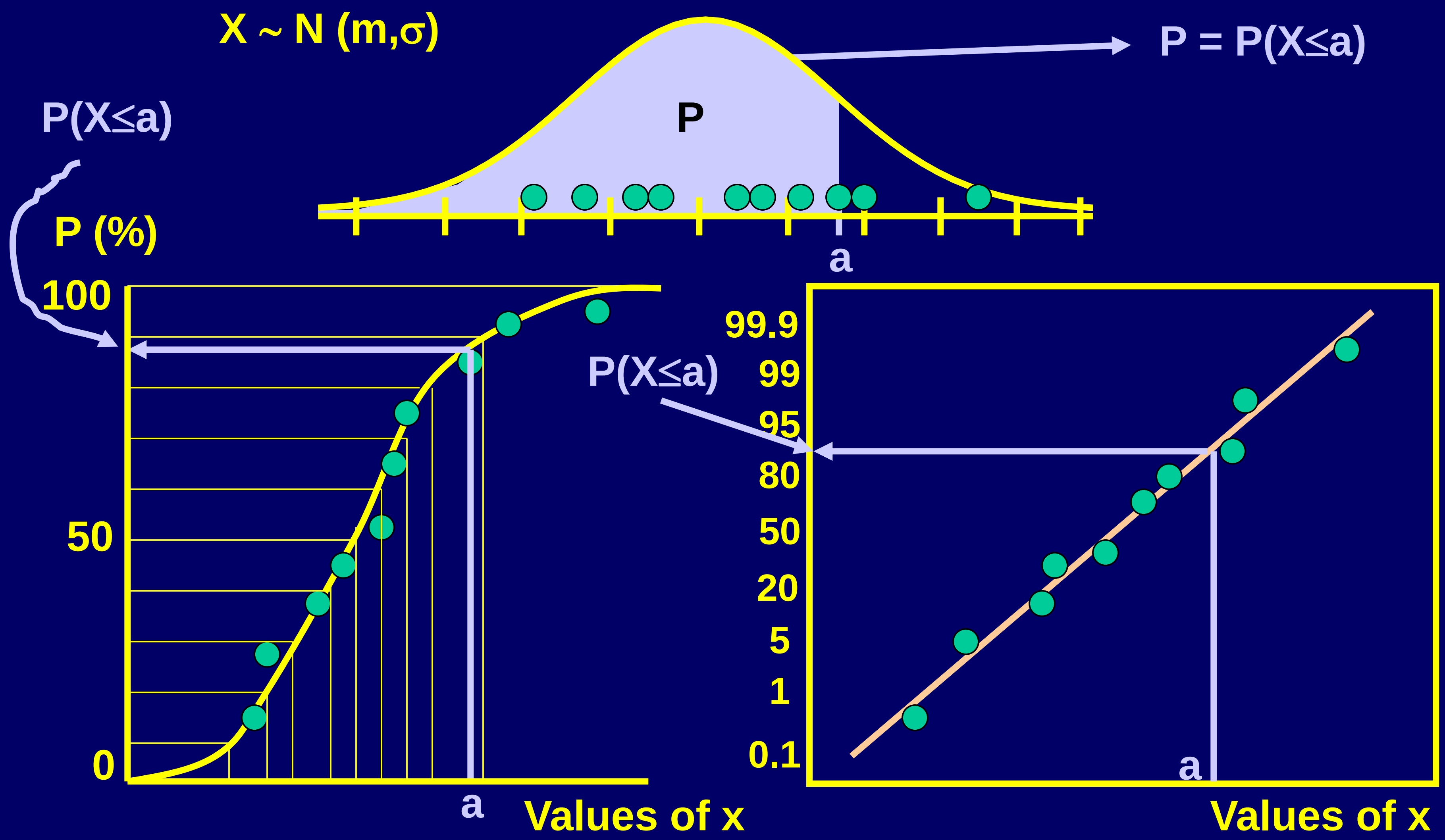


DENSITY FUNCTION

f(x)

$f(x)=dF(x)/dx$

NORMAL PROBABILITY PLOT:



Plot with ordinary scale

Normal probability plot

CONSTRUCTION OF A NORMAL PROBABILITY PLOT:

Resistance of 10 metallic pieces:

14.1 13.1 12.5 15.6 14.6 15 17 13.5 16.3 11.8 (Nw)

1) Sort the data from the lowest to the highest (increasing order)

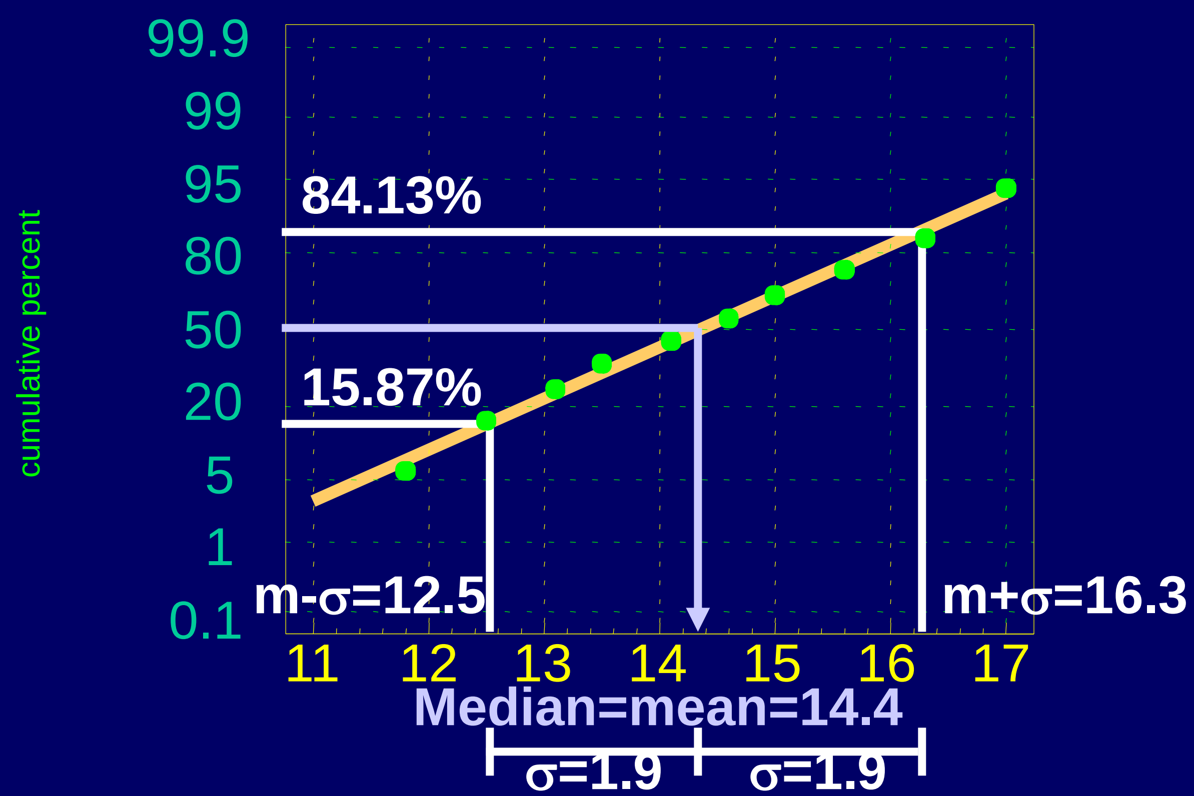
i (order)	1	2	3	4	5	6	7	8	9	10
	11.8	12.5	13.1	13.5	14.1	14.6	15	15.6	16.3	17
P(X≤x) (%)	5%	15%	25%	35%	45%	55%	65%	75%	85%	95%

2) Obtain the cumulated probabilities:

$$P(X \leq x)(\%) = \frac{i - \frac{1}{2}}{n} \cdot 100$$

3) Plot the values of X and P(X≤x) in the Normal Probability Plot

Normal Probability Plot

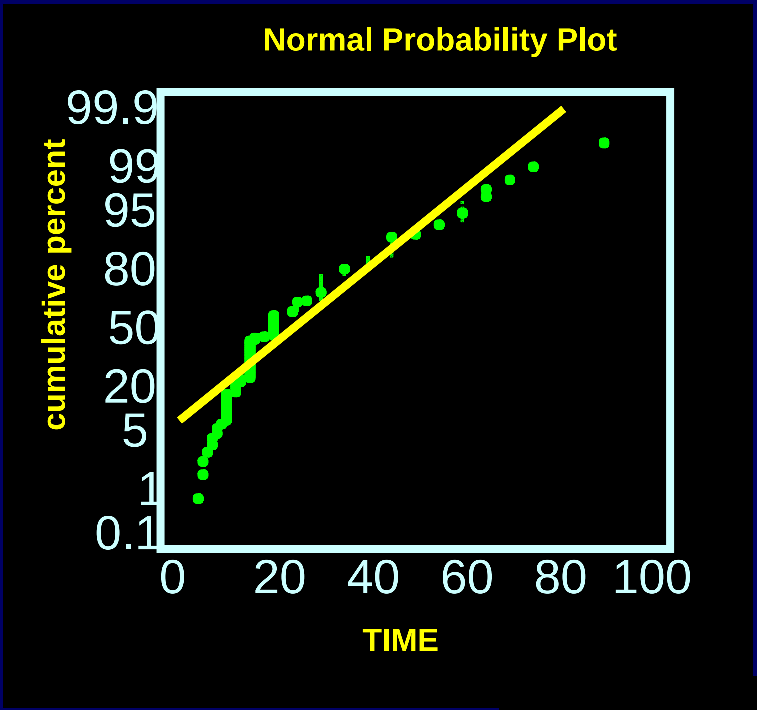


11.8	**12.5**	**13.1**	**13.5**	**14.1**	**14.6**	**15**	**15.6**	**16.3**	**17**
5%	**15%**	**25%**	**35%**	**45%**	**55%**	**65%**	**75%**	**85%**	**95%**

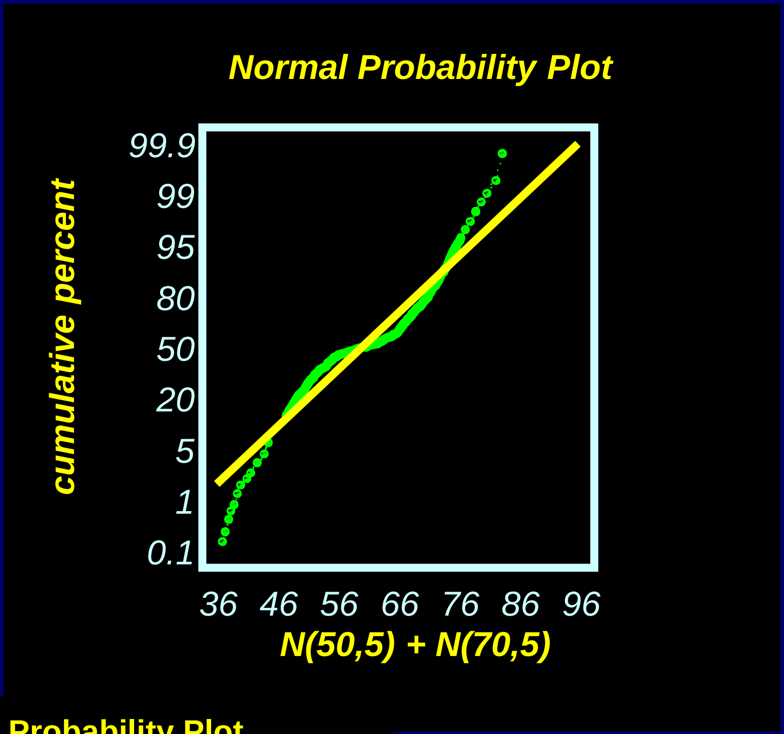
Data follow approx. a straight line $\longrightarrow$ **X: resistance ~ N(m_x,σ_x)**

How to estimate m_x? **How to estimate σ_x?**

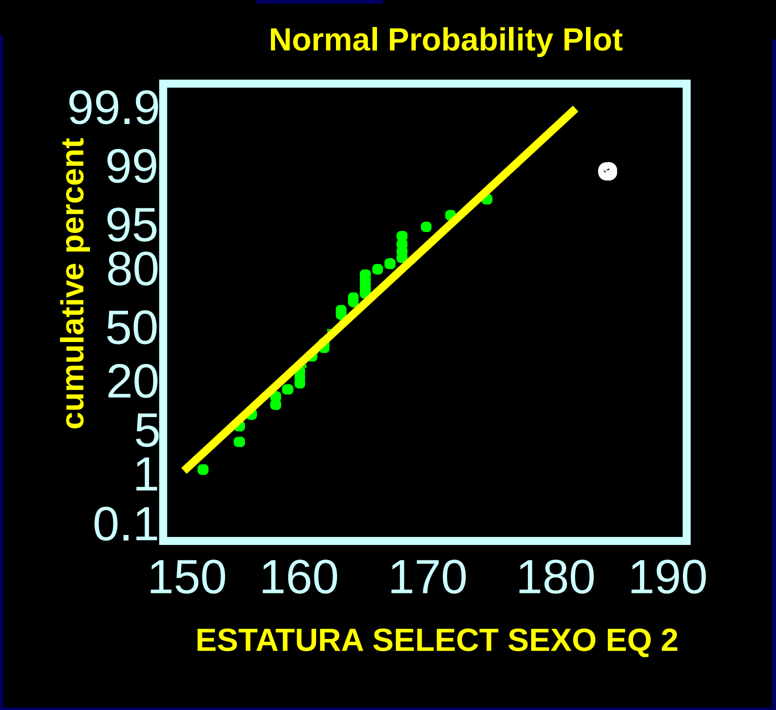
Therefore, X ~ N(m_x=14.4 , σ_x=1.9)



Positive skew



Mixture of 2 populations



Abnormal datum

NORMAL APPROXIMATIONS

CENTRAL LIMIT THEOREM

Under very general conditions, the addition of independent random variables, whatever type of distribution, tends to a normal distribution as the number of addends increases.

This theorem justifies why most distributions of continuous variables found in real problems are Normal.

$$Y = X_1 + X_2 + \ldots + X_N \xrightarrow[N \to \infty]{} Y \approx \text{Normal}$$

PROBLEM:

The resistance of an elevator in a factory is 10,000 kg. This elevator is used to carry packages with weight fluctuating uniformly between 40 and 60 kg. Calculate the maximum number of packages that can be loaded so that the probability to exceed the maximum load is lower than 0.1%.

Solution: $n_{max}=195$

APROXIMATION OF A BINOMIAL VARIABLE BY MEANS OF A NORMAL DISTRIBUTION

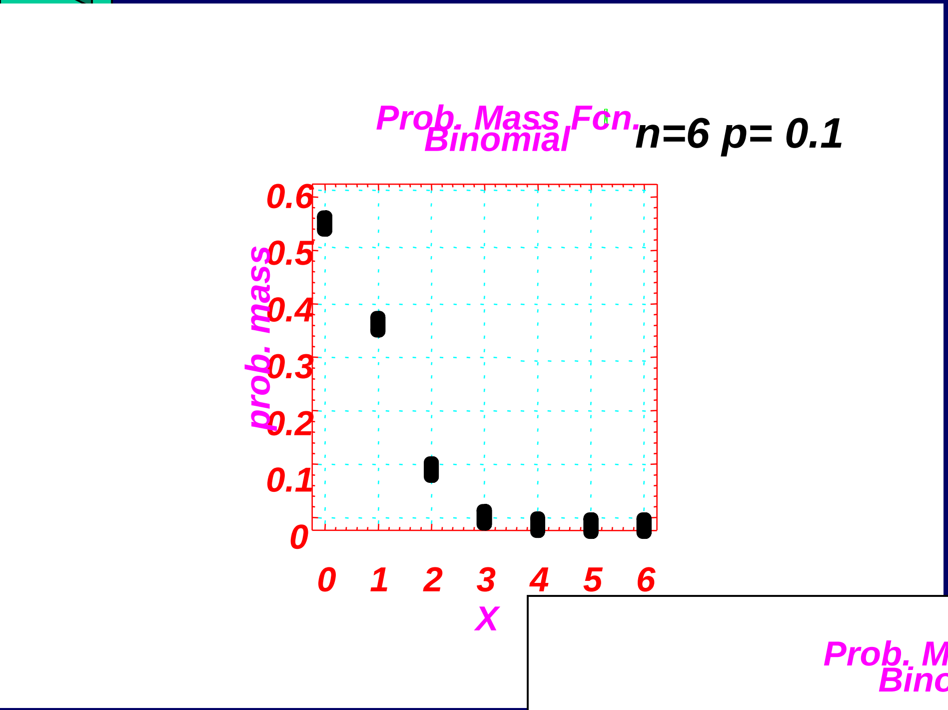
If $X \sim \text{Binomial}(n,p)$ => $E(X) = n \cdot p$; $\sigma^2(X) = n \cdot p \cdot (1-p)$

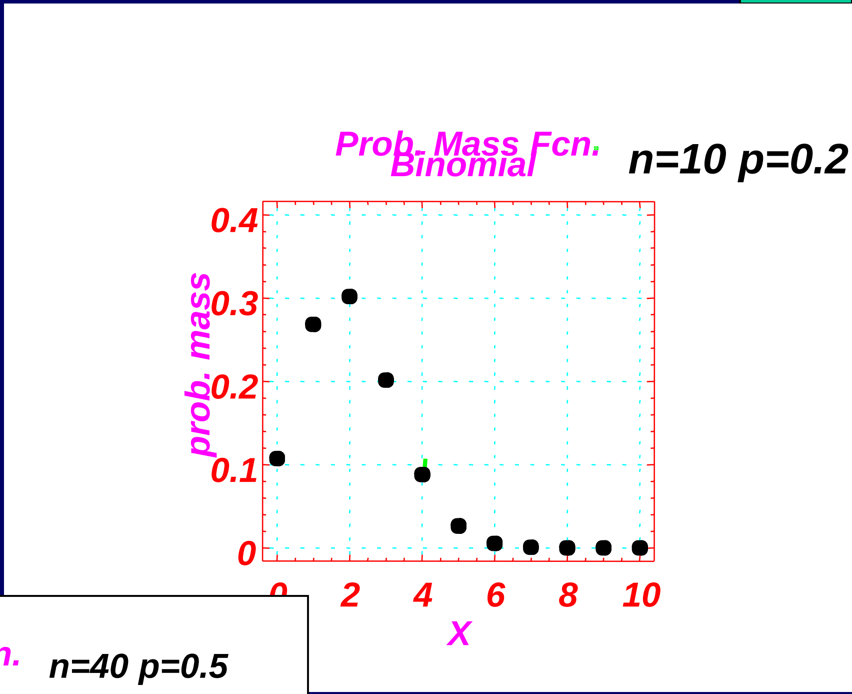
$X = Y_1 + Y_2 + \ldots + Y_n$; $Y_i \sim \text{Bernouilli}(p)$ independient

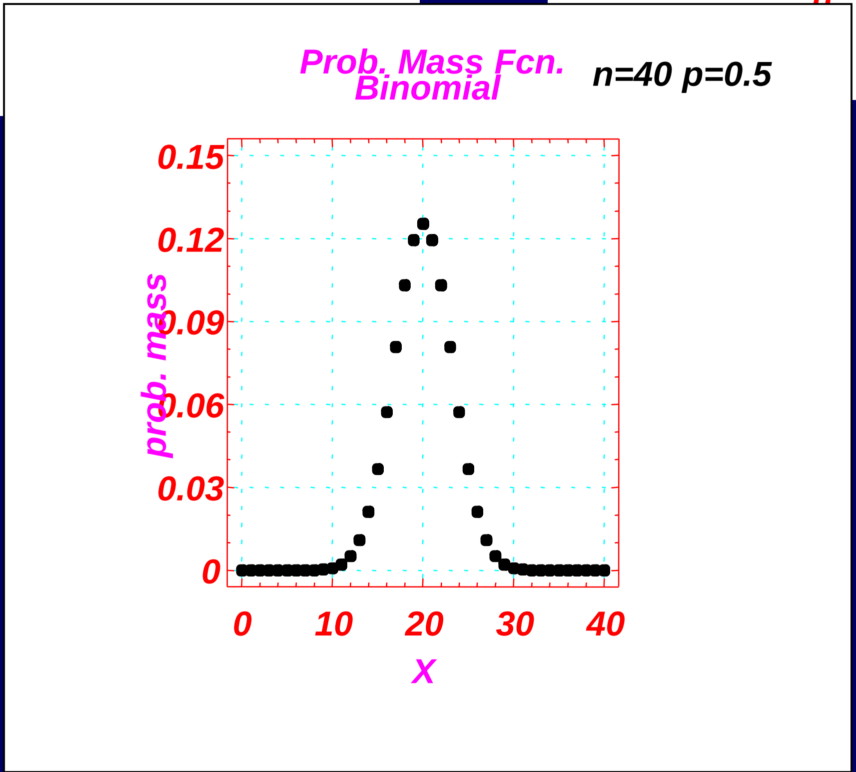
If n is high: $X \sim N[\, m = n \cdot p \, , \sigma^2 = n \cdot p \cdot (1-p) \,]$

The approximation is good if: $\sigma_X^2 = np(1-p) \geq 9$

$X \sim \text{Binomial}(n,p)$ --> if $n>50$ and $p<0.1$ --> $X \sim Ps(n \cdot p)$

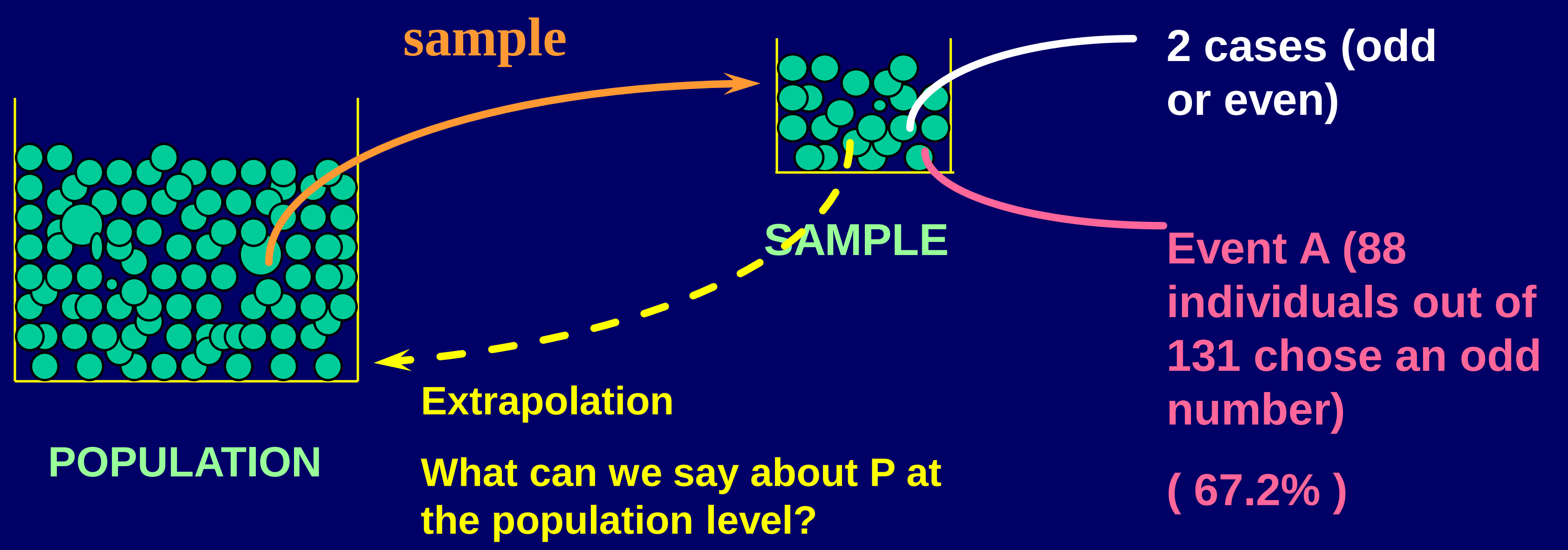

Prob. Mass Fcn.
Binomial
n=6 p= 0.1
prob. mass
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6
X


Prob. Mass Fcn.
Binomial
n=10 p=0.2
prob. mass
0.4
0.3
0.2
0.1
0
2 4 6 8 10
X


Prob. Mass Fcn.
Binomial
n=40 p=0.5
prob. mass
0.15
0.12
0.09
0.06
0.03
0
0 10 20 30 40
X

PROBLEM:

Assuming that the probability to say an odd digit (1, 3, 5, 7, 9) or an even digit (0, 2, 4, 6, 8) is the same, calculate approximately the probability that, among 131 digits randomly chosen by students, 88 or more are odd.



APROXIMATION OF A POISSON VARIABLE BY MEANS OF A NORMAL DISTRIBUTION

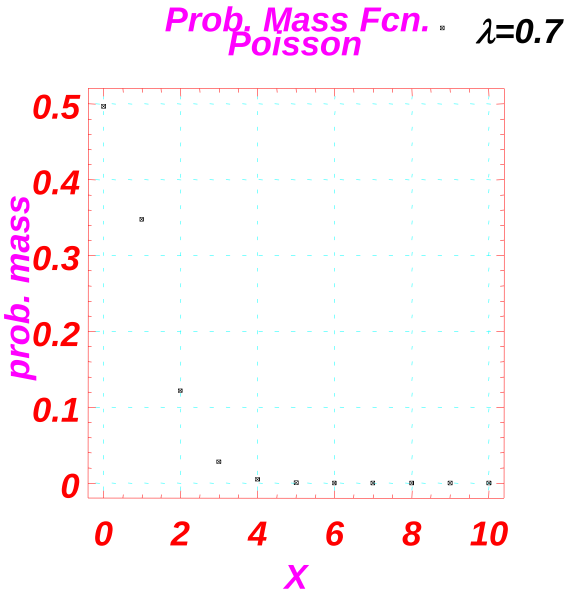
If $X \sim$ Poisson (λ) → $X \sim N\,[\,m = \lambda\,,\ \sigma^2 = \lambda\,]$

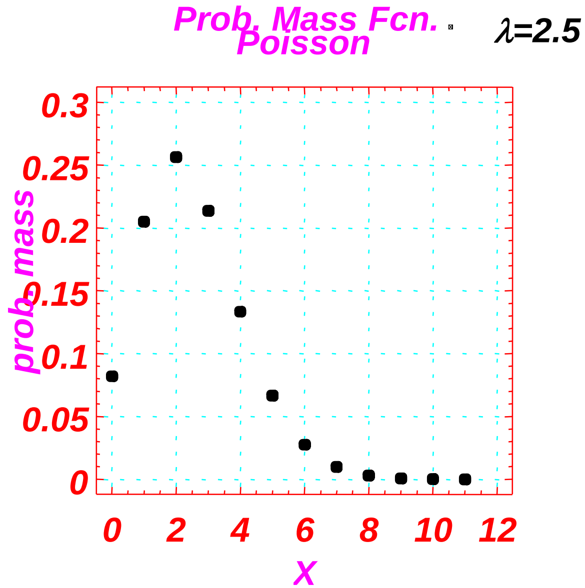
Follows approximately a N(0,1) distribution ← $\dfrac{X-\lambda}{\sqrt{\lambda}} \approx N(0,1)$

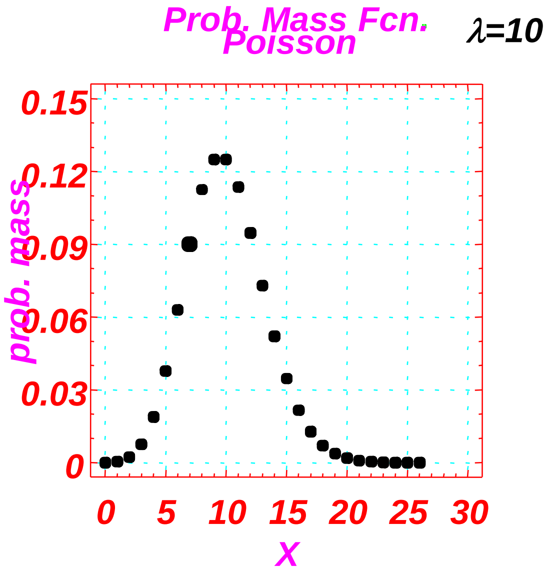
The approximation is good if: $\sigma_X^2 = \lambda \geq 9$

PROBLEM:

One consulting service receives in average 2 calls per minute. Calculate approximately the probability to receive more than 150 calls in one hour.


Prob. Mass Fcn.
Poisson
λ=0.7
prob. mass
X


Prob. Mass Fcn.
Poisson
λ=2.5
prob. mass
X


Prob. Mass Fcn.
Poisson
λ=10
prob. mass
X

CORRECTION OF CONTINUITY

$P[\ Ps(\lambda=36) < 3\ ] = P\ [\ N\ (36, 6) < 2.5\ ]$

$P[\ Ps(\lambda=36) \leq 3\ ] = P\ [\ N\ (36, 6) < 3.5\ ]$

$P[\ Ps(\lambda=36) = 3\ ] = P\ [N\ (36, 6) < 3.5] - P\ [N\ (36, 6) < 2.5]$