# PROBABILITY DISTRIBUTIONS

KAROLINA TURA-GAWRON, PHD

# AGENDA

1. Probability rules
2. Random variables
3. Distribution parameters
4. Special discrete probability distributions
5. Special continous probability distributions
6. Central Limit Theorem

# PROBABILITY – QUICK INTRODUCTION

For equally likely outcomes,
*Probability= Number of successful outcomes/Number of possible outomes*

**Example:**

Two fair coins are tossed. Show the possible outcomes. Find the probability that two heads are obtained (A).

Heads or Tails?

Possible outcomes: $\Omega$={(HH), (TT), (HT), (TH)}, n($\Omega$)=4

A={(HH)}, n(A)=1, P(A)=n(A)/n($\Omega$)=0.25

# THE PROBABILITY RULES – QUICK INTRODUCTION (1)

**Events A and B are said to be combined if they can occur at the same time.**
**Probability rule for combined events:**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

**Example:**

In a class of 20 children, 4 of the 9 boys and 3 of the 11 girls are in the athletic team. A person from the class is chosen to be in the 'egg and spoon' race on Sports Day. Find the probability that the person chosen is a female or in the athletics team.

A- female

B- in athletic team

P(A)=11/20, P(B)=7/20 P(A and B)=3/11 P(A or B)=11/20+7/20-3/11=0.63

# THE PROBABILITY RULES – QUICK INTRODUCTION (2)

Events A and B are said to be exclusive if they can not occur at the same time.
Probability rule for exclusive events:

$$P(A \cup B) = P(A) + P(B).$$

**Example:**

In race in which there are no dead heats, the probability that John wins is 0.3, the probability that Paul wins is 0.2 and the probability that Mark wins is 0.4. Find the probability that John or Mark wins.

A- John wins, P(A)=0.3

B- Mark wins, P(B)=0.4

P(A and B)= 0.3+0.4= 0.7

# THE PROBABILITY RULES – QUICK INTRODUCTION (3)

If A and B are two elements, not necessarily from the same experiment, than the conditional probability that A occurs, given that B has already occured, is written:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

If either of the events A and B can occur without being affected by the other, then the two events are independent:

$$P(A \cap B) = P(A) * P(B).$$

**Example:**

In a group of 60 students, 20 study History (A), 24 study French (B), 8 study both History and French (A and B). Are the events ‘a student studies History’ and ‘a student studies French’ independent?

P(A)=20/60, P(B)=24/60, P(A and B)=8/60

(20/60)*(24/60)=0.13=8/60=P(A and B) The events are independent

# 1. RANDOM VARIABLE (ONE DIMENSION) (1)

**Random variable (*X*)**- variable whose possible values are numerical outcomes of a random phenomenon. There are two types of random variables, **discrete** and **continuous.**

An example „Measuring sth”:

Measuring sth (height, temperature) → random event

Value as an outcome of measuring (1.73 cm, -12 degrees) → random variable

**Notification:**

Random variables: (X, Y, ...),

Values of random variables: (x, y, ...).

**Examples:**

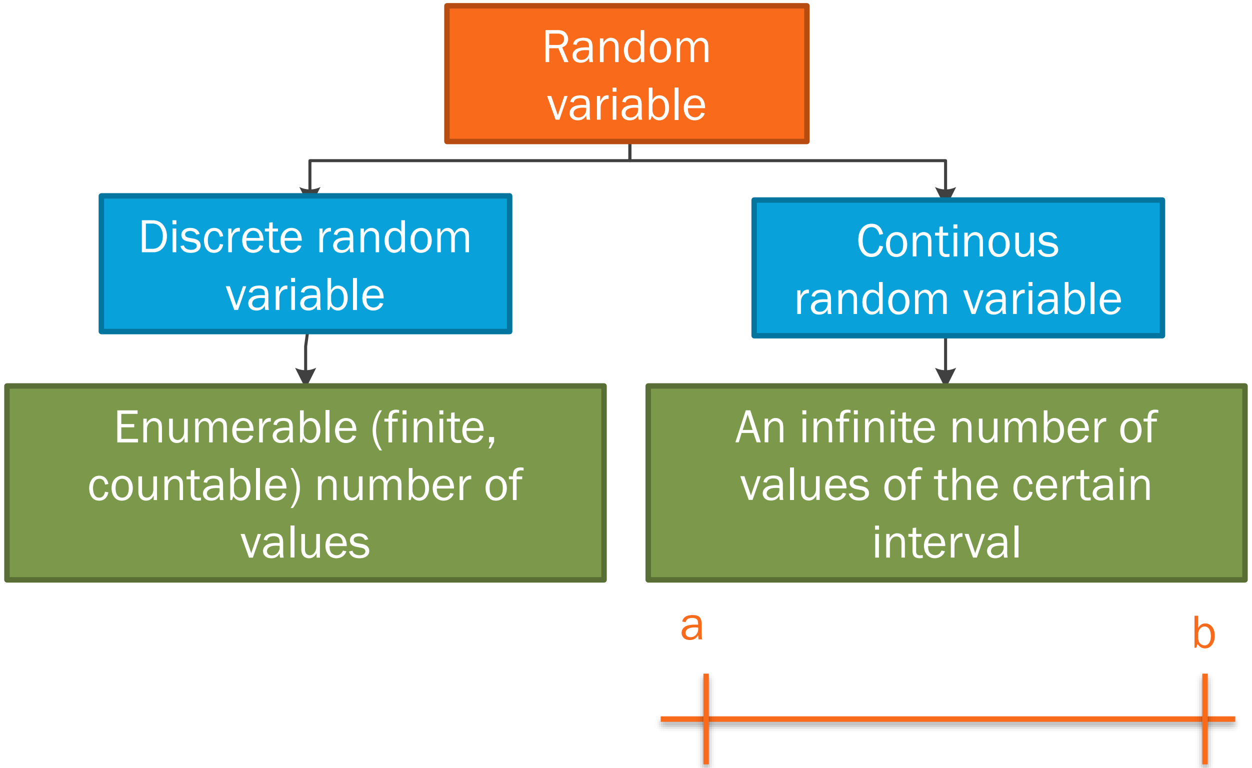
Unbiased dice is thrown. Number of ”six” in *n* throws.

Number of throws necessary to get the first „six”.

Orb’s number, on each is the placed the electron.

Human’s mood: .−1- mad;0- netral;1- happy

# 1. RANDOM VARIABLE (ONE DIMENSION) (2)



# 1.PROBABILITY DISTRIBUTION FUNCTION (P.D.F.)

**Probability distribution-** a mutually exclusive list of all the events that can result from a chance process and the corresponding probability of each event occuring (gives the probability of each possible value of the variable)

Discrete random variable

Probability distribution function

$$\sum p_i = 1$$

$$\sum_{all\ x} P(X = x) = 1$$

Continous random variable

Probability density function

$$f(x) \geq 0 \ \wedge \ \int_{-\infty}^{+\infty} f(x)dx = 1$$

$$P(a < X \leq b) = P(a \leq X \leq b) =$$

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

# 1. CUMULATIVE DISTRIBUTION FUNCTION (C.D.F)

Cumulative distribiution function–F(x): F(x)=P(X<x)

F(x) attributes:

1. $0 \le F(x) \le 1$
2. *F(x)* is no-declining function
3. *F(x)* is at least leftmostly continous
4. $\lim_{x \to -\infty} F(x) = 0 \quad \vee \quad \lim_{x \to +\infty} F(x) = 1$

# CUMULATIVE DISTRIBUTION FUNCTION (C.D.F)

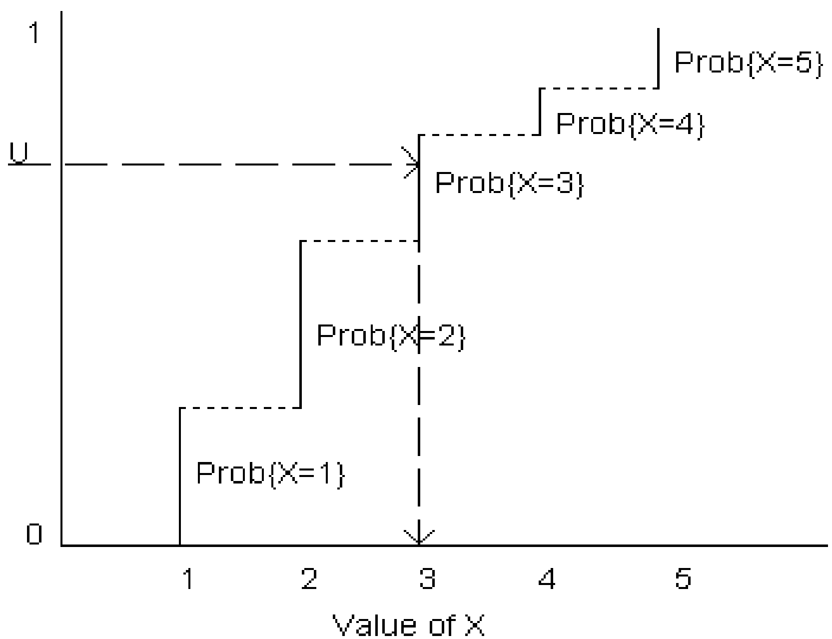
DISCRETE RANDOM VARIABLE

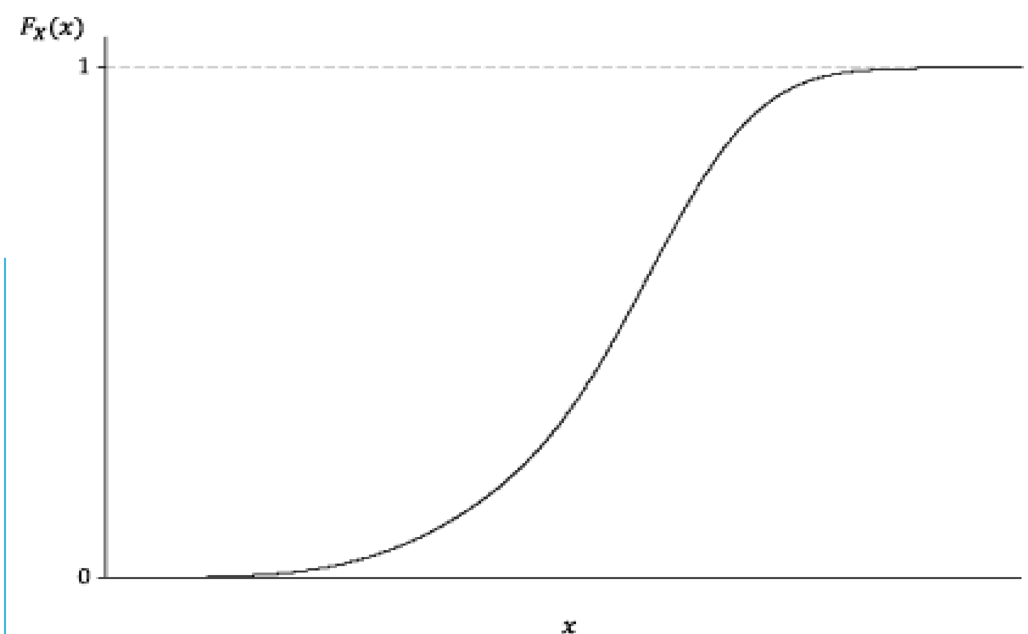
**Step-function**, leftmostly continous in their steps

CONTINOUS RANDOM VARIABLE

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

Probability density function



# FORMULAS

DISCRETE RANDOM VARIABLE

$$P(X \le x) = \lim_{k \to x^+} F(k) = F(x+)$$

$$P(X > x) = 1 - F(x+)$$

$$P(X < x) = F(x)$$

$$P(X = x) = F(x+) - F(x)$$

$$P(a \le X < b) = F(b) - F(a)$$

$$P(a < X \le b) = F(b) - F(a) + P(X = b) - P(X = a)$$

$$P(a \le X \le b) = F(b) - F(a) + P(X = b)$$

$$P(a < X < b) = F(b) - F(a) - P(X = a)$$

CONTINOUS RANDOM VARIABLE

$$P(a < X \le b) = P(a \le X \le b) =$$

$$= P(a \le X < b) = P(a < X < b)$$

# THE EXPECTED VALUE

DISCRETE RANDOM VARIABLE

$$EX = m_1 = \sum_{i=1} x_i p_i$$

CONTINOUS RANDOM VARIABLE

$$EX = m_1 = \int_{-\infty}^{+\infty} xf(x)dx$$

It shows the central point of the distribution, (point around which the velues of the random variable are grouped (centroid)

**Attributes of the expected value:**

1.X, Y- random variable, for which EX and EY exist $\longrightarrow$ E(X+Y)=E(X)+E(Y)

3. Ec=c, c- constant

4. X, Y are independent $\longrightarrow$ E(XY)=E(X)E(Y)

5. E(aX+b)= aE(X)+b

# 2. VARIANCE

DISCRETE RANDOM VARIABLE

$$D^2(X) = \sum_i (x_i - m)^2 p_i$$

CONTINOUS RANDOM VARIABLE

$$D^2(X) = \int_{-\infty}^{+\infty} (x - m)^2 f(x)dx$$

Measures the dispersion of values of random variable around the ecpectaed value

$D^2(X)=E(X-EX)^2=E(X^2)-(EX)^2$

**The attributes of variance:**

1. $D^2(X-c)=D^2(X)$, c-constant
2. $D^2(c)=0$, c-constant
3. $D^2(X+Y)=D^2(X)+D^2(Y)$
4. $D^2(cX)=c^2D^2(X)$

# 2. STANDARD DEVIATION

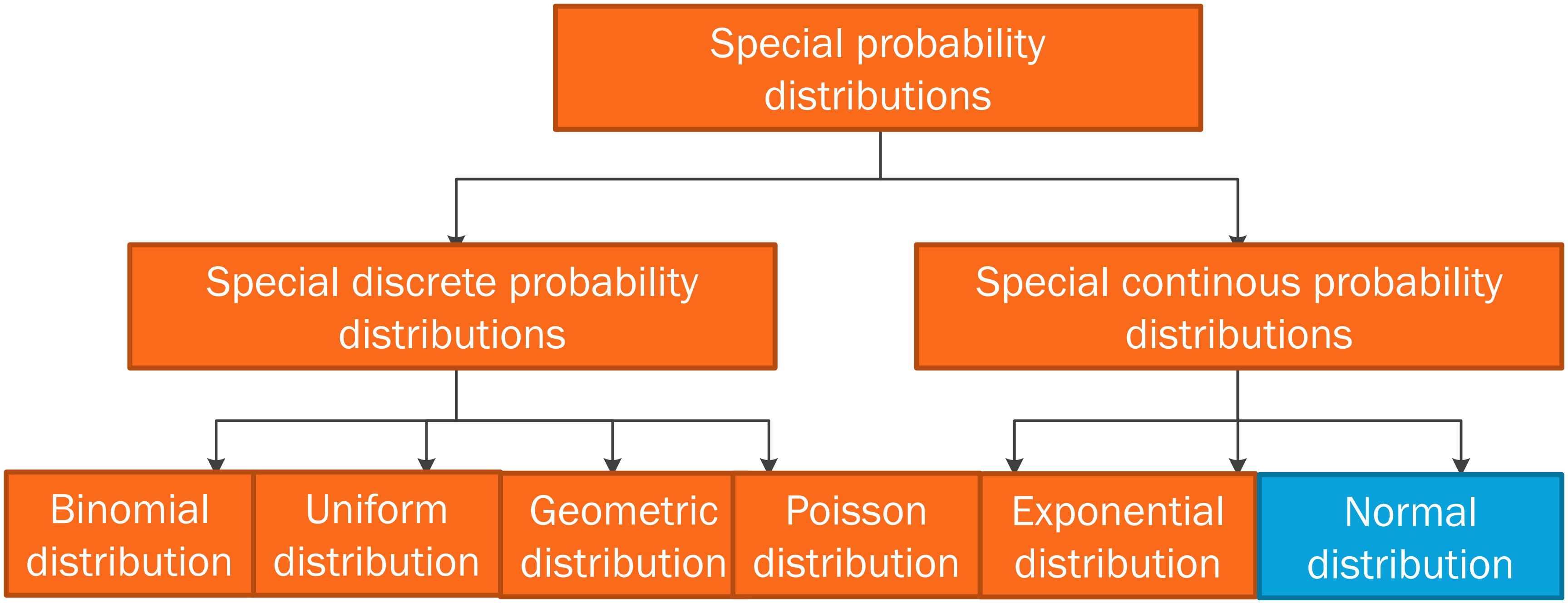
Root square of the variance

$$D(X) = \sqrt{D^2(X)}$$

**The attributes of the standard deviation:**

1. $D(aX+b)=|a|D(X)$
2. $D(X)>0$ lub $D(X)=0$

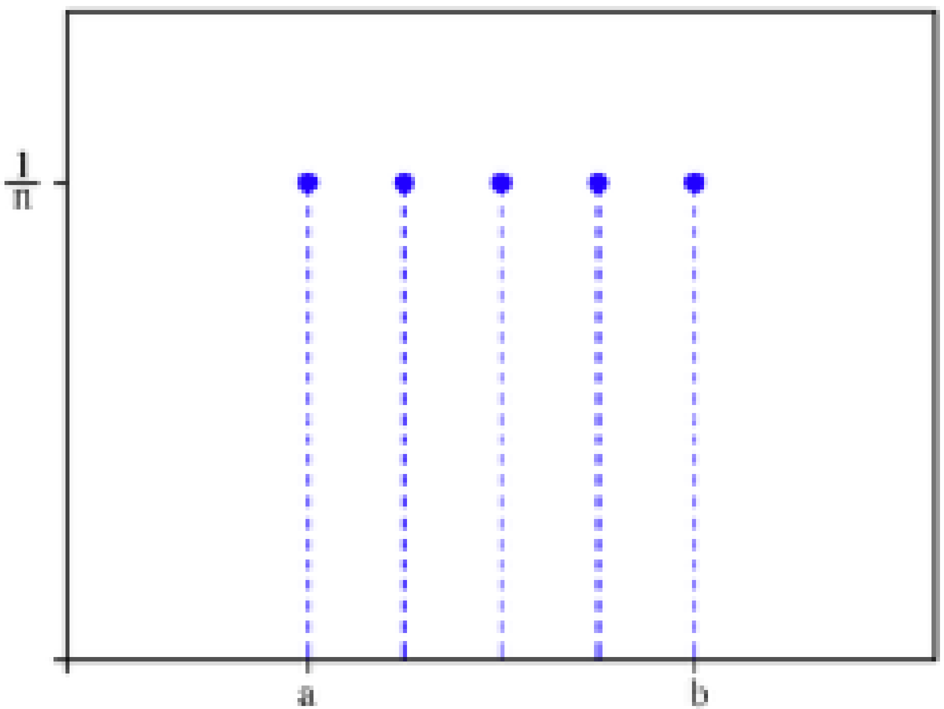
# SPECIAL PROBABILITY DISTRIBUTIONS



# SPECIAL DISCRETE PROBABILITY DISTRIBUTIONS

| | Geometric distribution | Uniform distribution | Binomial distribution | Discrete Poisson Distribution |
|---|---|---|---|---|
| **Probability distribution function** | $P(X=k)=q^{k-1}p$ | $P(X=x_i)=\frac{1}{n}$ | $P(X=k)=$ $=C_n^k p^k q^{n-k}$ $C_n^k=\frac{n!}{k!(n-k)!}$ | $np=\lambda$ $P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$ |
| **The expected value** | $\frac{1}{p}$ | $\frac{1}{n}\sum_{i=1}^{n}x_i$ | $np$ | $np=\lambda$ |
| **The standard deviation** | $\sqrt{\frac{q}{p^2}}$ | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-EX)^2}$ | $\sqrt{npq}$ | $\sqrt{np}=\sqrt{\lambda}$ |

# UNIFORM DISTRIBUTION

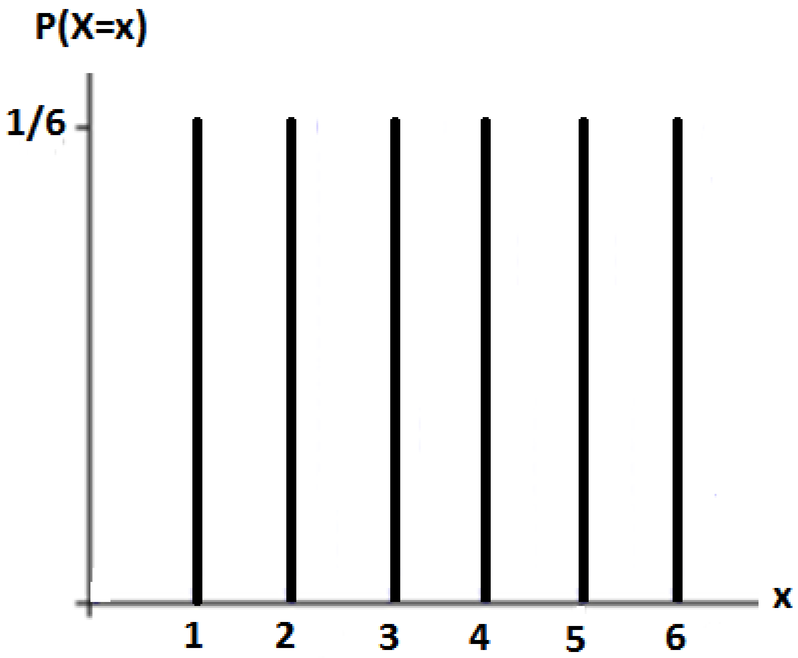



$$P(X = x_i) = \frac{1}{n}$$

- The discrete random variable X is defined over the set of n distinct values $x_1, x_2, x_3, \ldots, x_n$
- Each value is equally likely to occur

# EXAMPLE

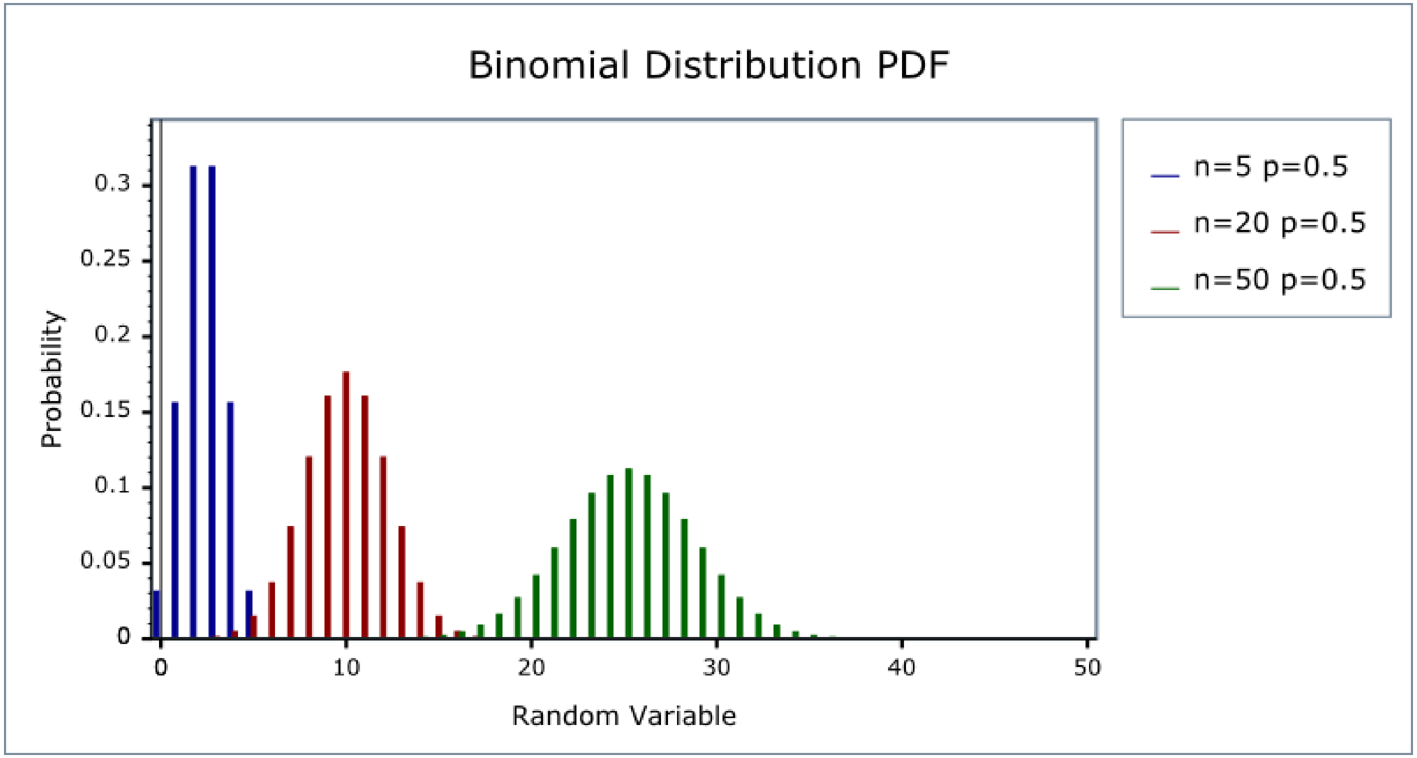
Throw an ordinary die. The probability distribution of X, the number on the die, is shown in the table and illustrated by the vertical line graph.




$$P(X = x) = \frac{1}{6} \quad for \quad x = 1, 2, 3, 4, 5, 6.$$

# BINOMIAL DISTRIBUTION

X~B(n,p)




- Finite number of trials *n*,
- Trials are independent,
- The outcome of each trial is deemed either a success (*p*) or a failure (*q*)
- The probability, *p*, of a succesful outcome is the same for each trial

# EXAMPLE

If we toss a die 5 times, what is the chance of obtaing 3 once (1,1,1).

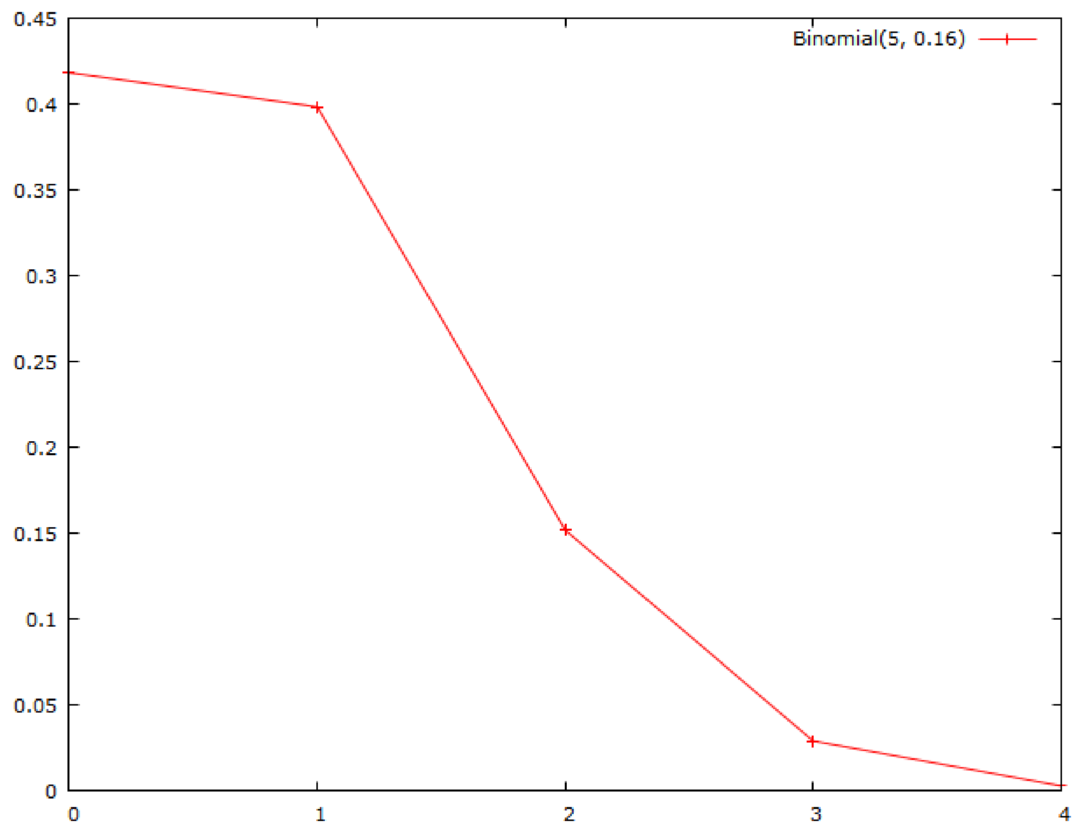
$$n = 5, k = 3, p = 1/6, q = 1 - p = 5/6$$

$$P(X = 3) = C_5^3 (1/6)^3 (5/6)^{5-3} = \frac{5!}{3!(5-3)!} * (1/6)^3 (5/6)^{5-3} = 0.032$$

$$P(X = k) =$$

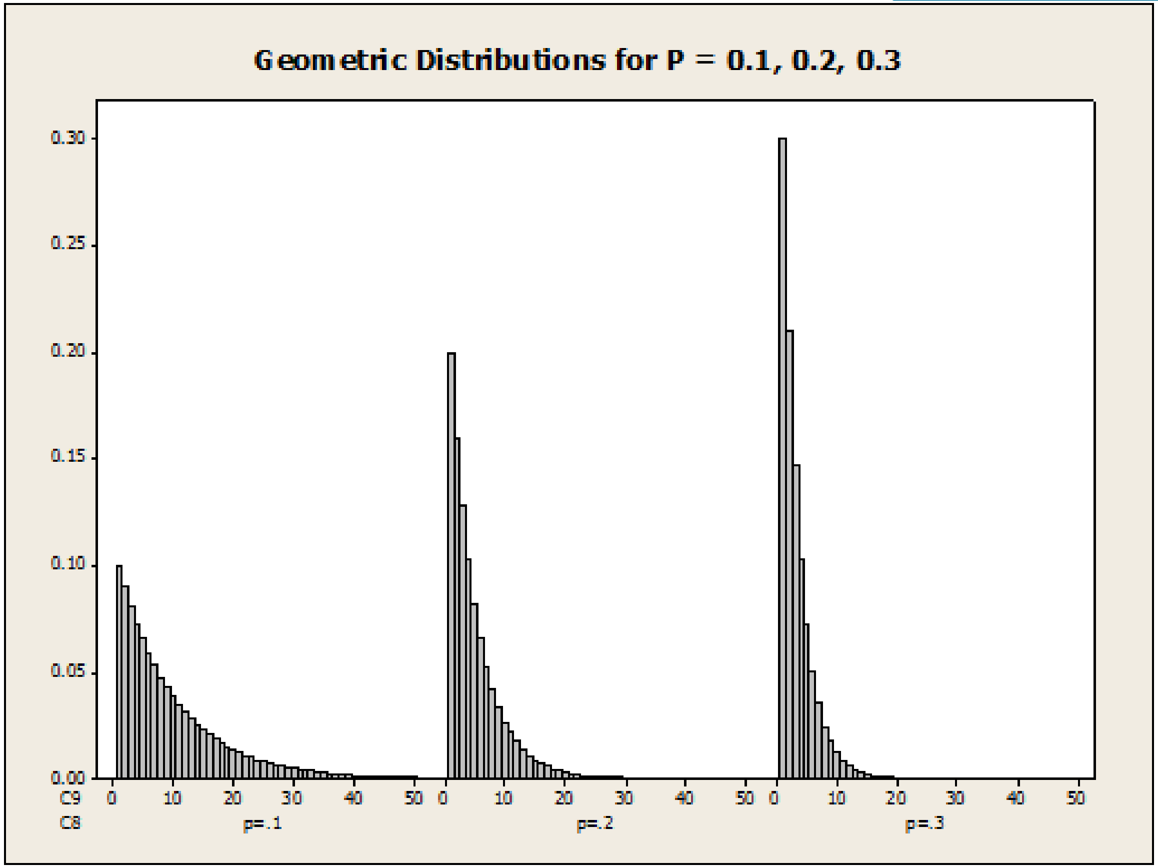
$$= C_n^k p^k q^{n-k}$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$



# THE GEOMETRIC DISTRIBUTION

X~Geo(p)




- Independent trials are carried out,
- The outcome of each trial is deemed either a success (*p*) or a failure (*q*)
- The probability, *p*, of a succesful outcome is the same for each trial
- The discrete random variable, X, is the number of trials needed to obtain the first succesfull outcome
- If above conditions are satisfied, X is said to follow a geometric distribution

# EXAMPLE

The sales of cereals „Lion” have increased rapidly after the promotion. The childern start to collect the plastic animals. Plastic models of anilmals are given away in packets of breakfast cereal. The probability that a packet contains a model of rabbit is 0.1. Consiider the probability distribution of X, the number of packets you open until you get a rabbit.

$$P(X = 1) = 0.1$$

$$P(X = 2) = 0.9 * 0.1 = 0.09$$

$$P(X = 3) = 0.9 * 0.9 * 0.1 = 0.081$$

$$P(X = 4) = 0.9 * 0.9 * 0.9 * 0.1 = 0.9^3 * 0.1$$

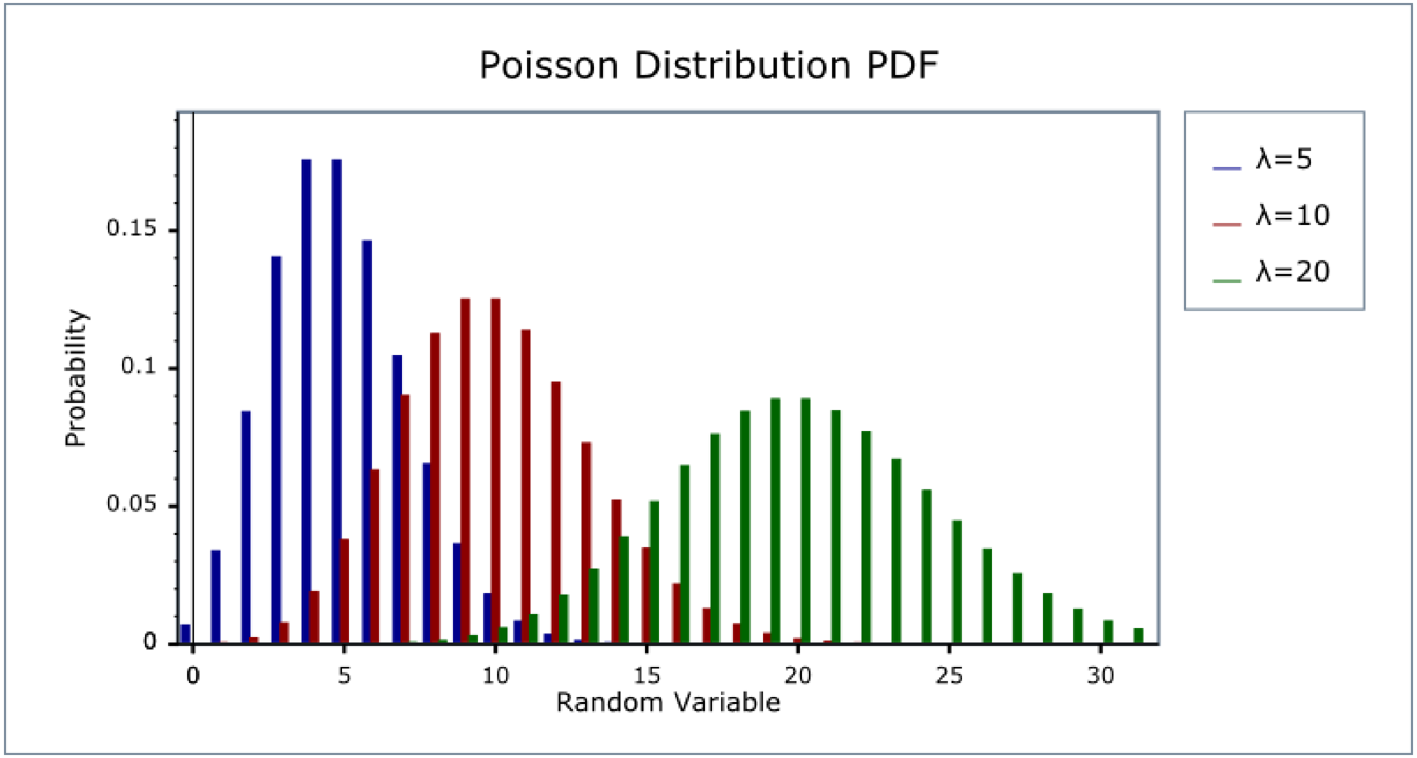
$$P(X = 5) = 0.9^4 * 0.1$$

$$P(X = 6) = 0.9^5 * 0.1$$

$$\ldots$$

# THE POISSON DISTRIBUTION

X~Po(λ)




- Events occur singly and at random in a given interval of time or space
- λ, the mean number of occurences in the given interval, is know and is finite.
- If above conditions are satisfied, X is said to follow a Poisson distribution

# EXAMPLES

- The number of emergency calls received by an ambulance control in an hour
- The number of vehicles approaching a motorway toll bridge in a five-minute interval
- The number of flaws in a metre length of material
- The number of white cupuscles on a slide

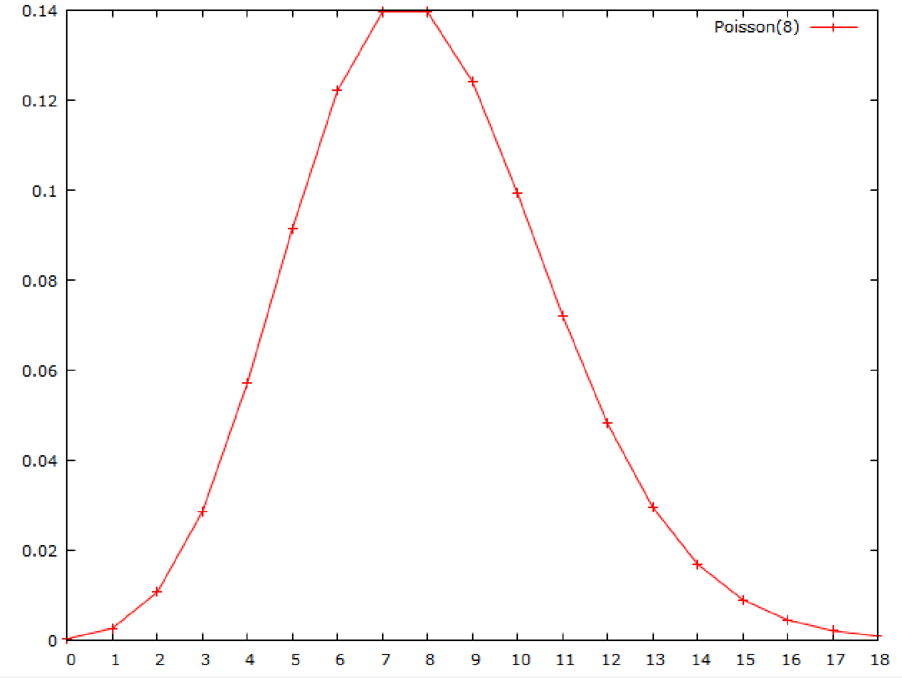
Each occurs randomly

# EXAMPLE

We are going to invest in school photocopiers. On average the school photocopier breaks down eight times during the school week (Monday to Friday). Assuming that the number of breakdowns can bt modelled by Poisson distribution, find the probability that it breaks down.

$$X \sim Po(8), \lambda = 8$$

$$P(X = 5) = \frac{8^5 e^{-8}}{5!} = 0.0916$$




$$np = \lambda$$

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

# CONTINOUS PROBABILITY DISTRIBUTIONS

# UNIFORM (RECTANGULAR) DISTRIBUTION

Continous random variable distributed uniformly in the range a<=x<=b.  X~R(a,b)

$$f(x) = \begin{cases} 0 & for \quad x < a \\ \dfrac{1}{b-a} & for \quad a \le x \le b \\ 0 & for \quad x > b \end{cases}$$

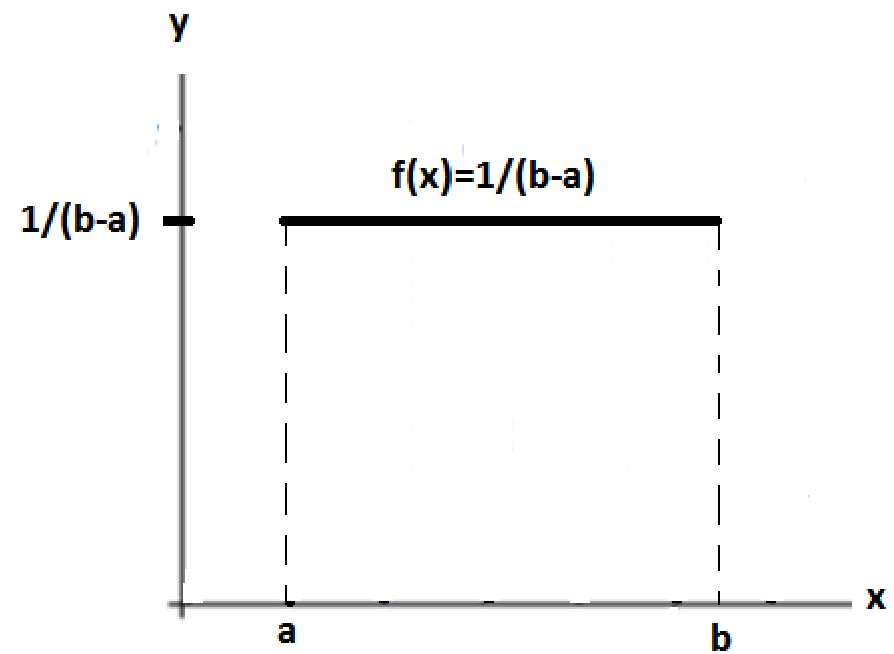
Probability density function

$$F(x) = \begin{cases} 0 & for \quad x \le a \\ \dfrac{x-a}{b-a} & for \quad a < x \le b \\ 1 & for \quad x > b \end{cases}$$

c.d.f.

Expected value: $\dfrac{a+b}{2}$

Standard deviation: $\dfrac{b-a}{2\sqrt{3}}$



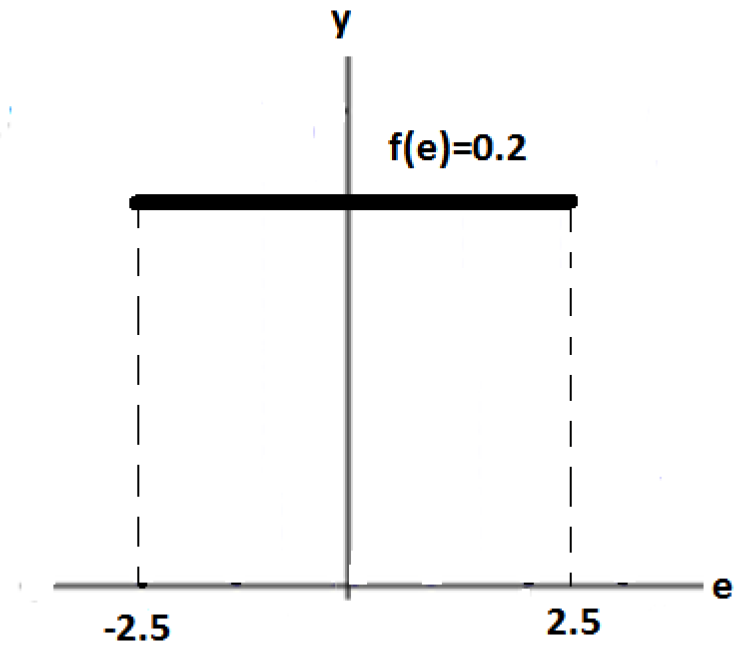
# EXAMPLE

The factory produces the metal rods. The manager of the factory decided to analyse the measurements' errors. The length of metal rods are measured to the nearest 5 mm. What is the distribution of the random variable E, the rounding error made when measuring? Give its probability density function f(e).The error is the difference between the true length and the recorded length after rounding to the nearest 5mm.

$$-2.5 \leq E < 2.5$$

$$E \sim R(-2.5, 2.5)$$

$$f(e) = \frac{1}{2.5-(-2.5)} = 0.2$$




$$f(x) = \begin{cases} 0 & for \quad x < a \\ \dfrac{1}{b-a} & for \quad a \leq x \leq b \\ 0 & for \quad x > b \end{cases}$$

# THE EXPONENTIAL DISTRIBUTION

A continous distribution closely related to the Poisson distribution. Recall that a Poisson random variable counts the number of occurances of an event during a given time interval. In contrast, an exponential random variable, T, can be used to measure the time that ellapses before the first occurance of an event where occurances of the event follow a Poisson distribution. Equivalently, an exponential random variable measure the time that elapses between the occurances of an event.

**Probability density function**

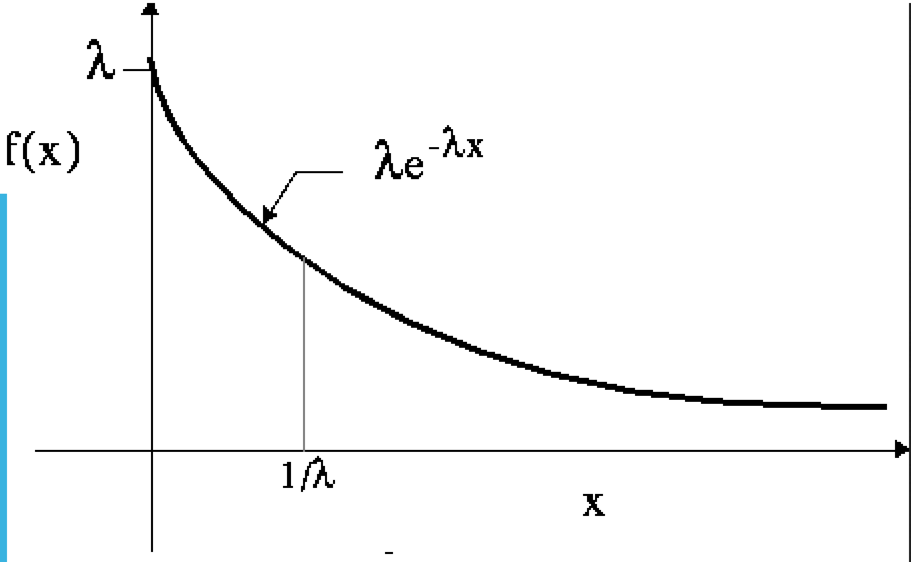
$$f(x) = \lambda e^{-\lambda x} \quad dla \quad x \geq 0$$

**Probability**

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

Expected value: $\frac{1}{\lambda}$

Standard deviation: $\frac{1}{\lambda}$



# NORMAL DISTRIBUTION

The random variable X with mean μ and variance $\sigma^2$ is normally distributed if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

$$\pi \approx\sim 3.14159\ldots$$

$$e \approx\sim 2.71828\ldots$$

**c.d.f.**

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv$$

Expected value: $\mu$
Standard deviation: $\sigma$
Skewness: 0
Kurtosis: 3

# NORMAL DISTRIBUTION

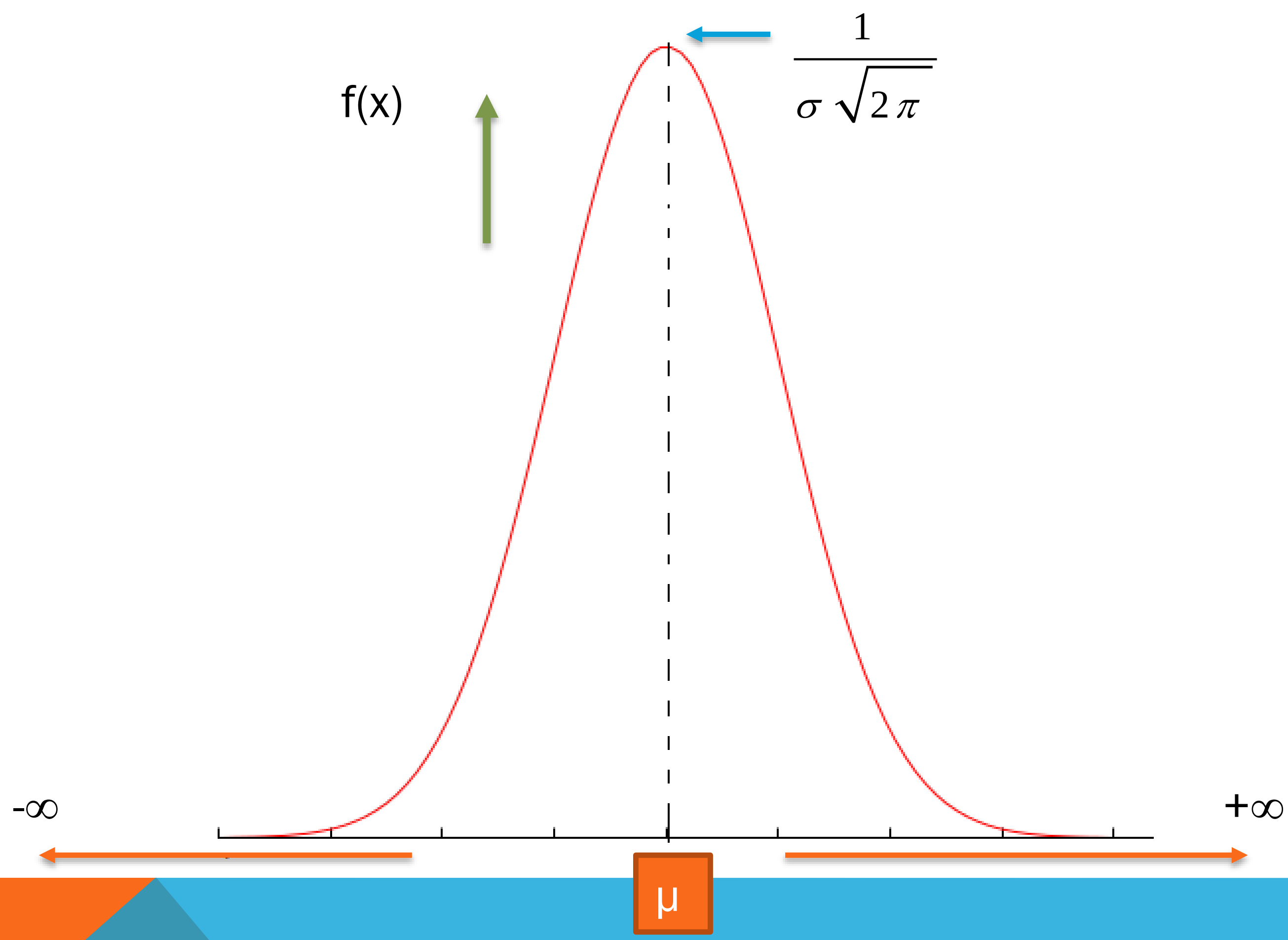
$X \sim N(\mu, \sigma^2)$

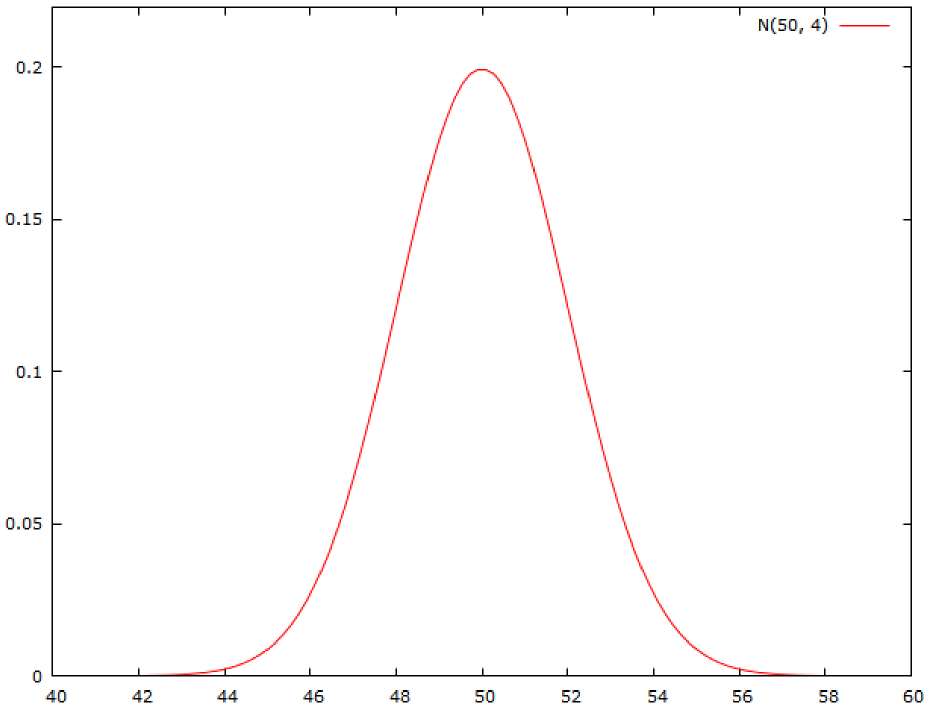
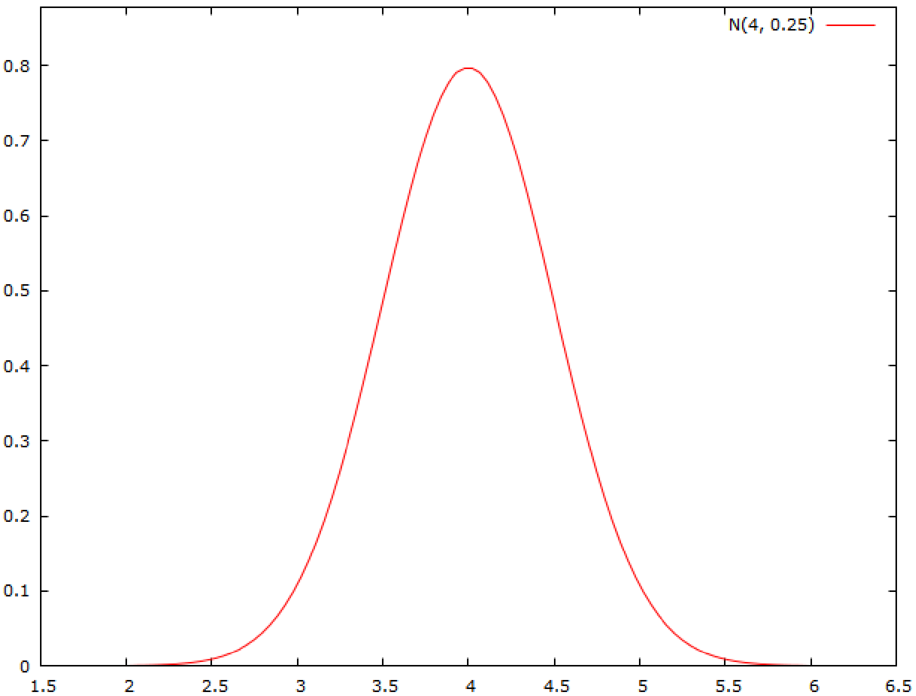
The normal distribution curve has the following features:

- It is bell shaped
- It is symmetrical about μ
- It extends from -∞ to ∞
- The maximum value of f(x) is

$$\frac{1}{\sigma\sqrt{2\pi}}$$

- The total area under the curve is 1


$\frac{1}{\sigma\sqrt{2\pi}}$
f(x)
-∞
+∞
μ


N(50, 4)
0.2
0.15
0.1
0.05
0
40
42
44
46
48
50
52
54
56
58
60

N(4, 0.25)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5

# THE STANDARD NORMAL VARIABLE, Z

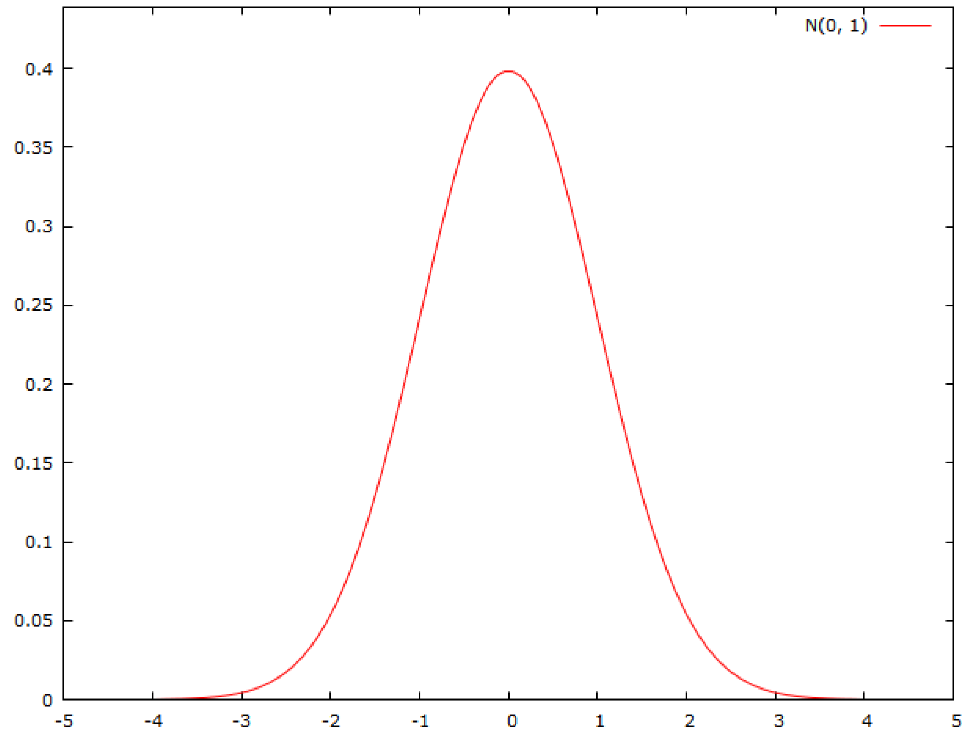
To standarise X, where X~N($\mu$,$\sigma^2$):

- substract the mean $\mu$,
- then divide by the standard deviation $\sigma$.

To obtain:

$$Z = \frac{X - \mu}{\sigma}$$

where Z~N(0,1).



# EXAMPLE

The company's profit has a normal distribution with the parameters: average 100 (thousands') Euro and standard deviation (thousands') Euro. Calculate the probability that the company will achieve a profit of over 120 (thousands') Euro.

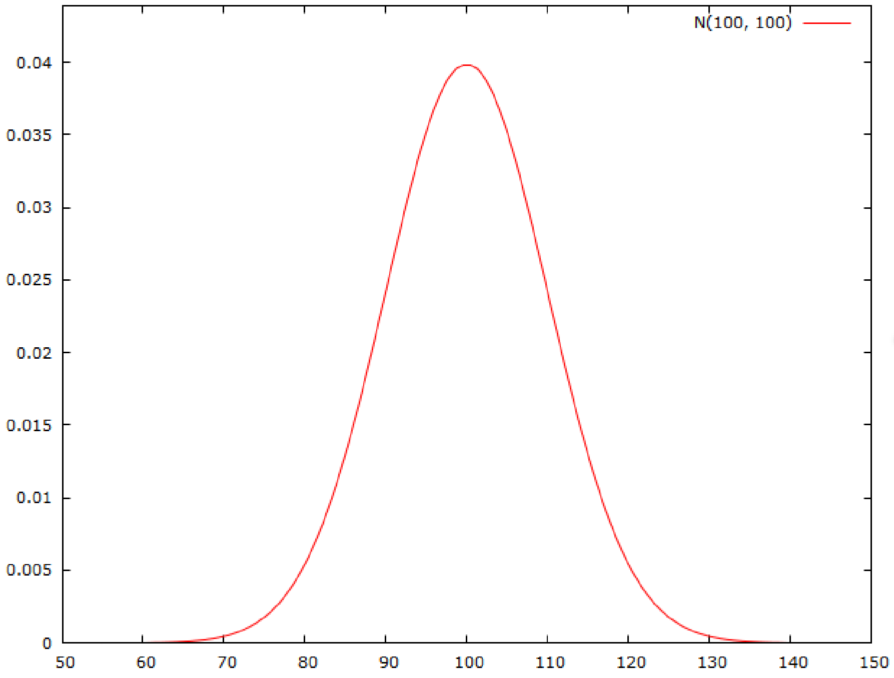
$$X = 120, \mu = 100, \sigma = 10,$$

$$X \sim N(100,100)$$

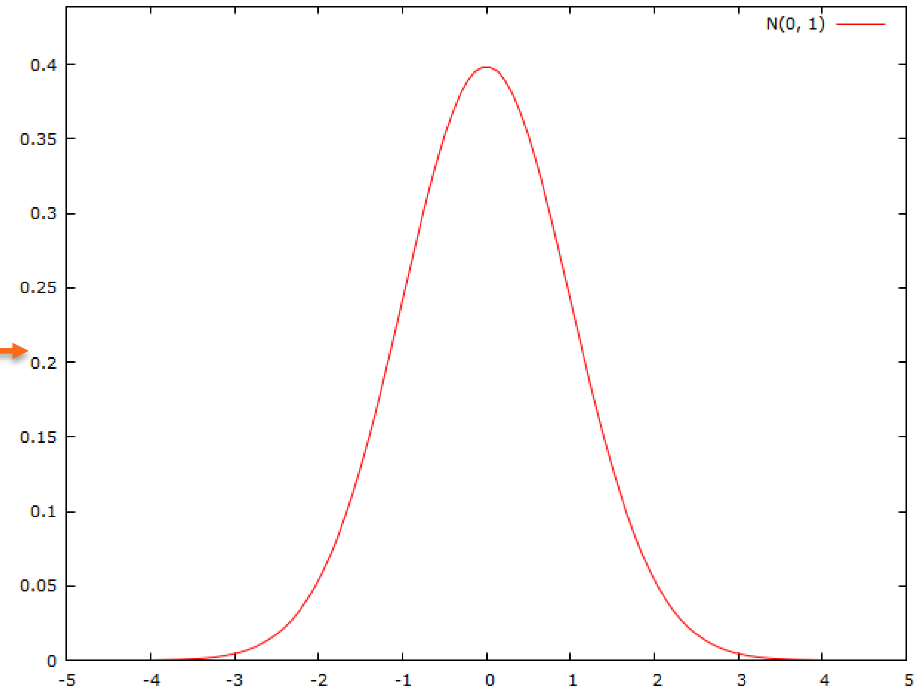
$$Z = \frac{X - \mu}{\sigma} = \frac{120 - 100}{10} = 2$$

$$Z \sim N(0,1)$$

# EXAMPLE


N(100, 100)
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
50
60
70
80
90
100
110
120
130
140
150

X


N(0, 1)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-5
-4
-3
-2
-1
0
1
2
3
4
5

Z

# CENTRAL LIMIT THEOREM

*The world is normal!*

# CENTRAL LIMIT THEOREM

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with finite mean $\mu$ and variance $\sigma^2$. If the sample size *n* is sufficiently large, then the sample mean $\bar{X}$ follows an approximate normal distribution:

- with mean $E(\bar{X}) = \mu_{\bar{X}} = \mu$
- and variance $Var(\bar{X}) = \sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}$

$$\bar{X} \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right) \quad as \quad n \to \infty$$

$or$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1) \quad as \quad n \to \infty$$

# "SUFFICIENTLY LARGE" – WHAT DOES IT MEAN?

- **If the distribution of the $X_i$ is symmetric, unimodal or continuous,**

**then a sample size *n* as small as 4 or 5 yields an adequate approximation.**

- **If the distribution of the $X_i$ is skewed,**

**then a sample size *n* of at least 25 or 30 yields an adequate approximation.**

- **If the distribution of the $X_i$ is extremely skewed,**

**then you may need an even larger *n* .**

# CENTRAL LIMIT THEOREM

The Central Limit Theorem states that the sampling distribution of the sampling means approaches a normal distribution as the sample size gets large

# TASK

**Suppose I toss a fair coin 100 times.** The mean and variance of the number of heads in a single coin toss are 1/2 and 1/4 respectively. Consequently, the total number of heads is approximately normally distributed with mean 50, variance 25, and standard deviation 5. From this information, and a table of values of $\Phi(x)$, I can estimate that I will get between 45 and 55 heads with probability about 0.7, and between 40 and 60 heads with probability about 0.95, without having to sum lots of binomial probabilities.

Thank you for your attention