CENTRE DE RECHERCHE PUBLIC SANTÉ

Microarray Center

APPLIED STATISTICS

Lecture 3

Probability Distributions

Petr Nazarov
petr.nazarov@crp-sante.lu

11-11-2009

- **Random variables**
- **Discrete probability distributions**
  - discrete probability distribution
  - expected value and variance
  - binomial probability distribution
  - Poisson probability distribution
  - hypergeometric probability distribution
- **Continuous probability distribution**
  - a continuous probability distribution
  - uniform probability distribution
  - normal probability distribution
  - exponential probability distribution

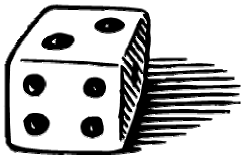
Random variable
A numerical description of the outcome of an experiment.

A random variable is always a numerical measure.

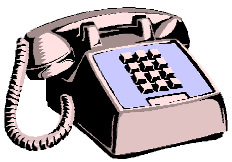
Discrete random variable
A random variable that may assume either a finite number of values or an infinite sequence of values.

Continuous random variable
A random variable that may assume any numerical value in an interval or collection of intervals.

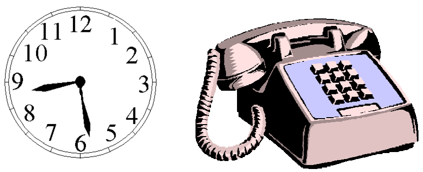
Roll a die



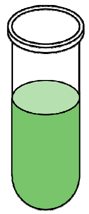
Number of calls to a reception per hour



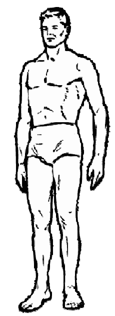
Time between calls to a reception



Volume of a sample in a tube



Weight, height, blood pressure, etc



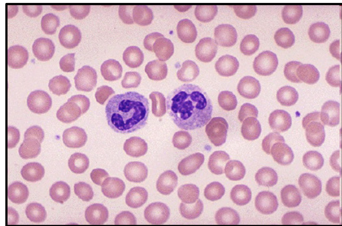
Discrete Probability Distribution

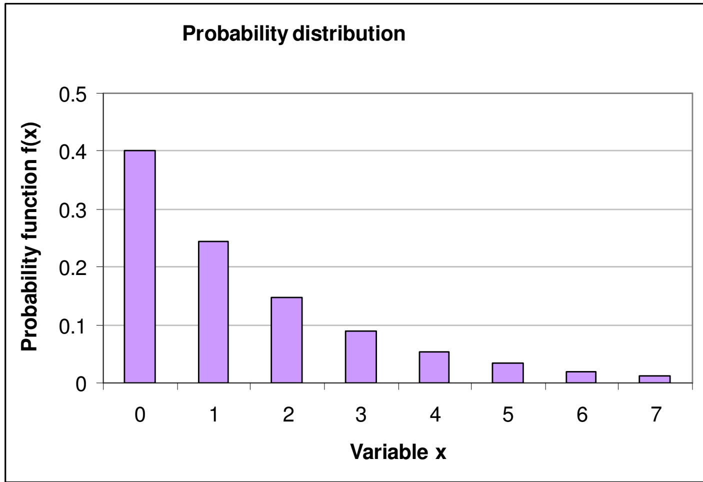
Probability distribution
A description of how the probabilities are distributed over the values of the random variable.

Probability function
A function, denoted by *f(x),* that provides the probability that *x* assumes a particular value for a discrete random variable.

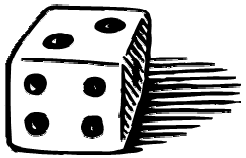
$$f(x) \geq 0$$
$$\sum f(x) = 1$$

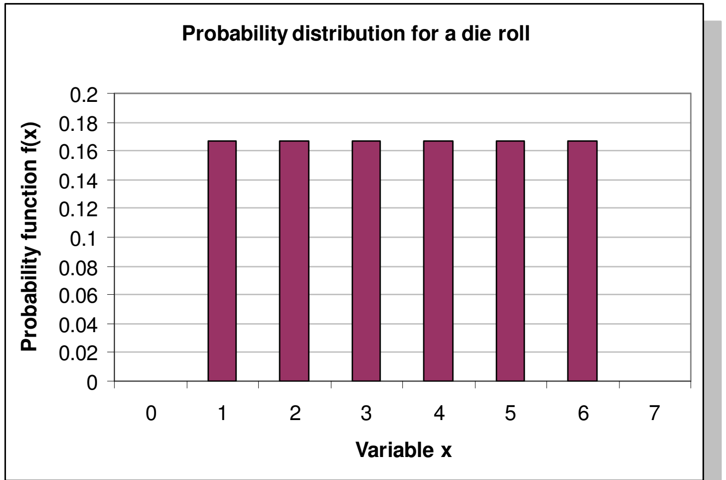
Number of cells under microscope
Random variable X:
x = 0
x = 1
x = 2
x = 3
…





Roll a die
Random variable X:
x = 1
x = 2
x = 3
x = 4
x = 5
x = 6





DISCRETE PROBABILITY DISTRIBUTIONS

Discrete Uniform Probability Function

Discrete uniform probability distribution
A probability distribution for which each possible value of the random variable has the same probability.

$$f(x) = \frac{1}{n}$$

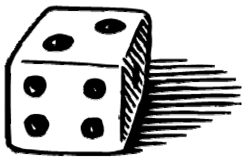
n – number of values of *x*

Expected value
A measure of the central location of a random variable, mean.

$$E(x) = \mu = \sum x f(x)$$

Variance
A measure of the variability, or dispersion, of a random variable.

$$\sigma^2 = \sum (x-\mu)^2 f(x)$$



x	$f(x)$
1	0.1667
2	0.1667
3	0.1667
4	0.1667
5	0.1667
6	0.1667

$\mu = \sum(x_i / n) = \sum(x_i) / n$

$\mu = 3.5$
$\sigma^2 = 2.92$
$\sigma^2 = 2.92$

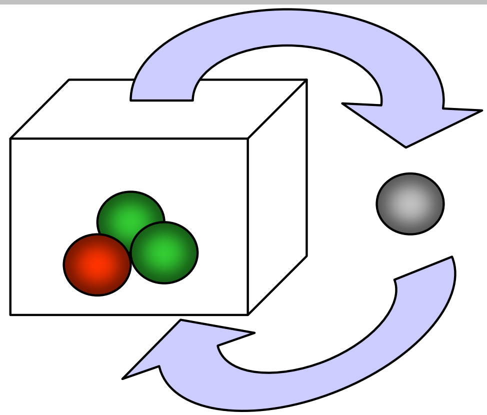
Example

Assuming that the probability of a side effect for a patient is 0.1. What is in a group of 3 patients none, 1, 2, or all 3 will get side effects?

Binomial experiment

An experiment having the four properties:

1. The experiment consists of a sequence of ***n* identical trials**.
2. **Two outcomes** are possible on each trial, one called success and the other failure.
3. The probability of a success ***p*** does not change from trial to trial. Consequently, the probability of failure, **1–*p***, does not change from trial to trial.
4. The trials are independent.



n trials

DISCRETE PROBABILITY DISTRIBUTIONS

Binomial Experiment

Binomial probability distribution
A probability distribution showing the probability of ***x*** successes in ***n*** trials of a binomial experiment.

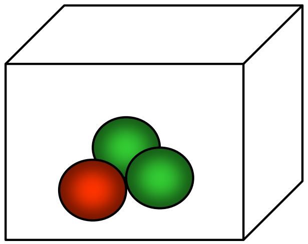
Probability distribution for a binomial experiment

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

$$E(x) = \mu = np$$

$$Var(x) = \sigma^2 = np(1-p)$$

Probability of red *p*(red)=1/3, 3 trials are given. Random variable = number of "red" cases



$$f(2) = \frac{3!}{2!(3-2)!}\left(\frac{1}{3}\right)^2\left(1-\frac{1}{3}\right)^{(3-2)}$$

f(0) = 8/27 = 0.296
f(1) = 4/9 = 0.444
f(2) = 2/9 = 0.222
f(3) = 1/27 = 0.037

Test: $\sum f(x) = 1$

Example: Binomial Experiment

Example

Assuming that the probability of a side effect for a patient is 0.1.

1. What is the probability to get none, 1, 2, etc. side effects in a group of 5 patients?
2. What is the probability that not more then 1 get a side effect
3. What is the expected number of side effects in the group?

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

p = 0.1

n = 5

In MS Excel use the function:

- **= BINOMDIST(x,n,p,false)**

Q1.

x	f(x)
0	0.59049
1	0.32805
2	0.0729
3	0.0081
4	0.00045
5	0.00001

Q2.

P(0 or 1) = P(0) + P(1)

= 0.59 + 0.33 = **0.92**

Q3.

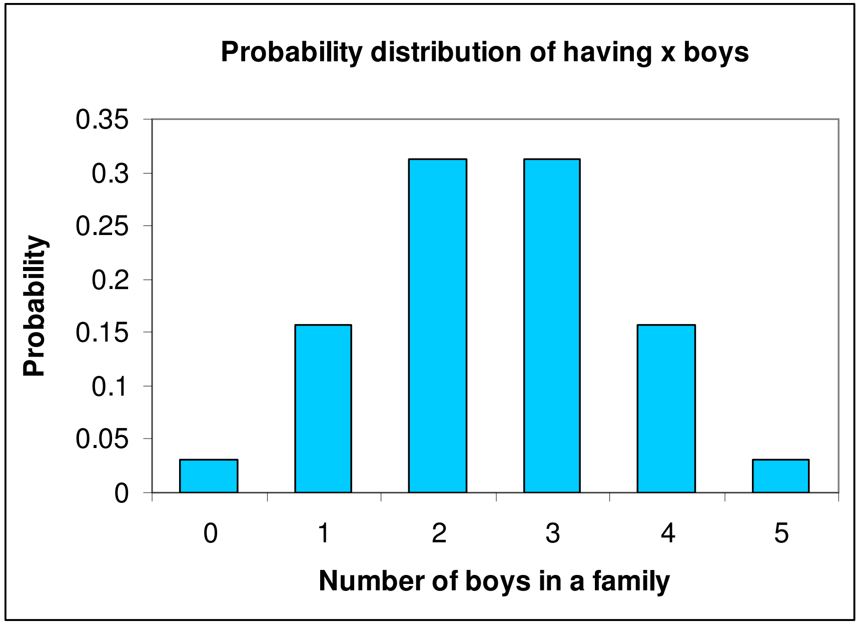
$\mu = 5 \times 0.1 = 0.5$

Practical

Assume the probability of getting a boy or a girl are equal.

1. Calculate the distribution of boys/girl in a family with **5 children**.
2. Plot the probability distribution
3. Calculate the probability of having all 5 children of only one sex

x	f(x)
0	0.03125
1	0.15625
2	0.3125
3	0.3125
4	0.15625
5	0.03125

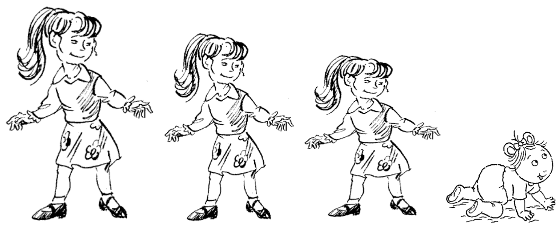


Q3.

$P(0 \text{ or } 5) = P(0) + P(5)$

$= 0.03 + 0.03 =$ **0.06**

Assume a family has 4 girls already. What is the probability that the 5th will be a girl?



?

Example

Number of calls to an Emergency Service is on average 3 per hour b/w 2.00 and 6.00 of working days. What are the probabilities to have 0, 5, 10 calls in the next hour?

Poisson probability distribution
A probability distribution showing the probability of *x* occurrences of an event over a specified interval of time or space.

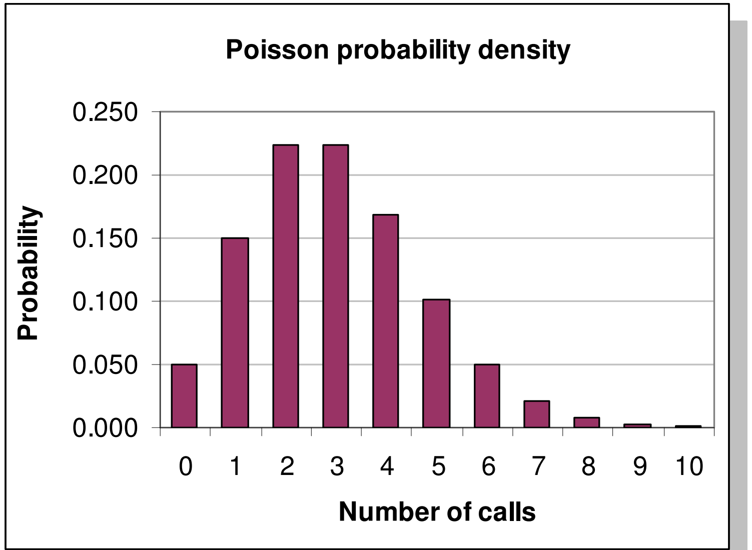
Poisson probability function
The function used to compute Poisson probabilities.

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

$$\mu = \sigma^2$$

where μ – expected value (mean)

x	f(x)
0	0.050
1	0.149
2	0.224
3	0.224
4	0.168
5	0.101
6	0.050
7	0.022
8	0.008
9	0.003
10	0.001



In Excel use the function:

- `= POISSON(x,mu,false)`

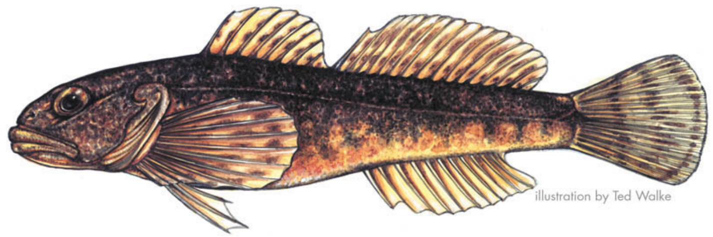
Example: Poisson Distribution for Fish Counting

Example

An ichthyologist studying the *spoonhead sculpin* catches specimens in a large bag seine that she trolls through the lake. She knows from many years experience that on averages she will catch 2 fish per trolling.

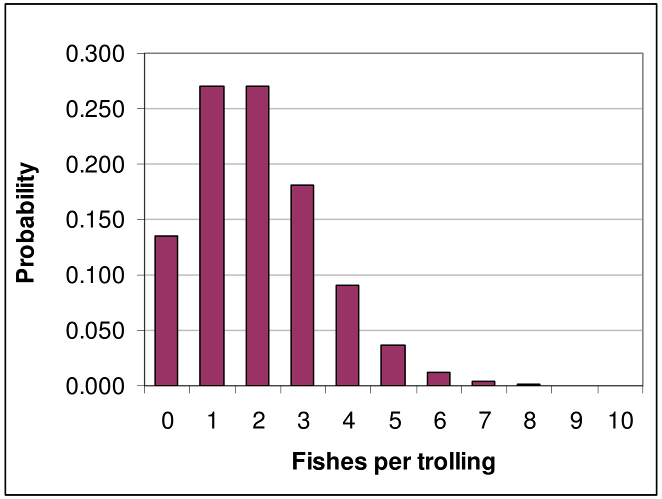
Find the probabilities of catching:

1. No fish;
2. Fewer than 4 fish;
3. More then 1 fish.



In Excel use the function:

- `= POISSON(x,mu,false)`



Q1.
P(0) = 0.135

Q2.
P(<4) = P(0)+P(1)+P(2)+P(3)=0.857

Q3.
P(>1) =1-P(0)-P(1)=0.857

Glover , Mitchell, An Introduction to Biostatistics

Example

There are 12 mice, of which 5 have an early brain tumor. A researcher randomly selects 3 of 12. What is the probability that none of these 3 has a tumor? What is the probability that more then 1 have a tumor?

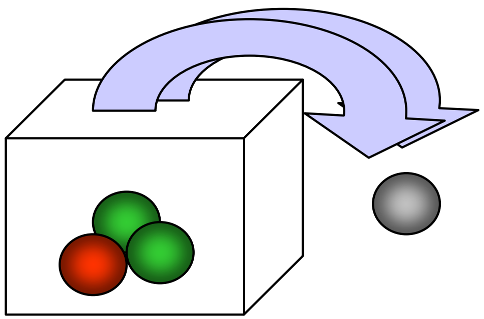
Hypergeometric experiment

A probability distribution showing the probability of *x* successes in *n* trials from a population *N* with *r* successes and *N-r* failures.

$$f(x) = \frac{C_x^r C_{n-x}^{N-r}}{C_n^N}, \quad for\ 0 \le x \le r$$

$$E(x) = \mu = n\left(\frac{r}{N}\right)$$

$$Var(x) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$



n trials

In Excel use the function:

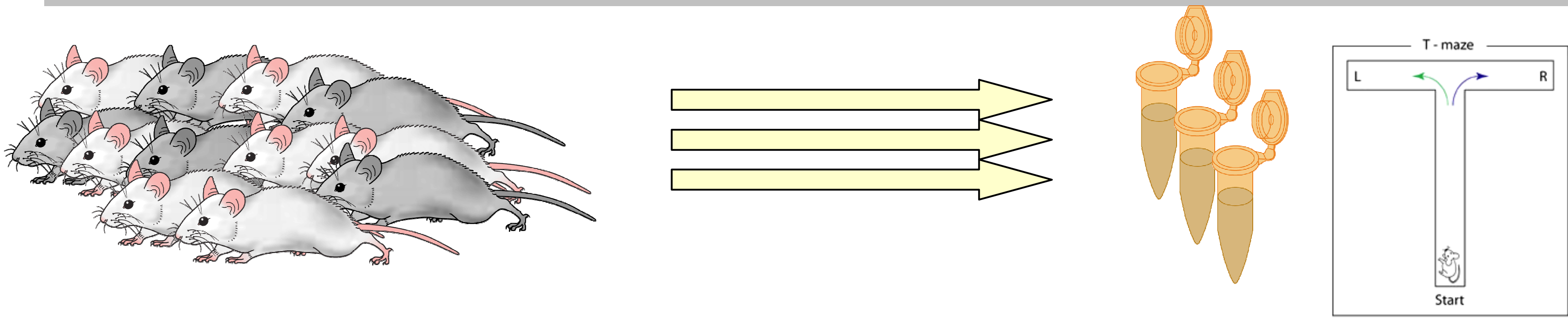
- `=HYPGEOMDIST(x,n,r,N)`

Example: Hypergeometric Distribution for Mice

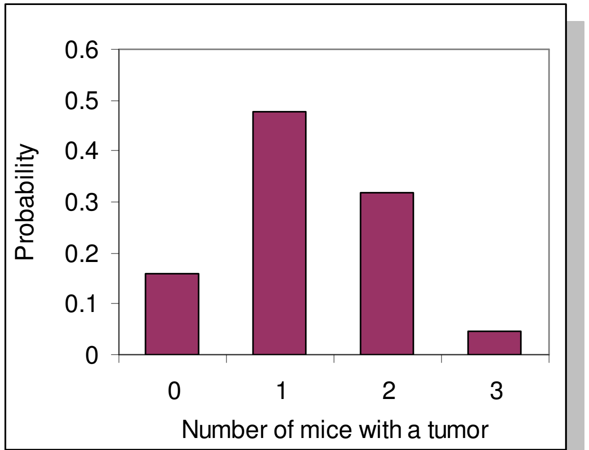
Example

There are 12 mice, of which 5 have an early brain tumor. A researcher randomly selects 3 of 12.

1. What is the probability that none of these 3 has a tumor?
2. What is the probability that more then 1 have a tumor?



x	f(x)
0	0.159
1	0.477
2	0.318
3	0.045



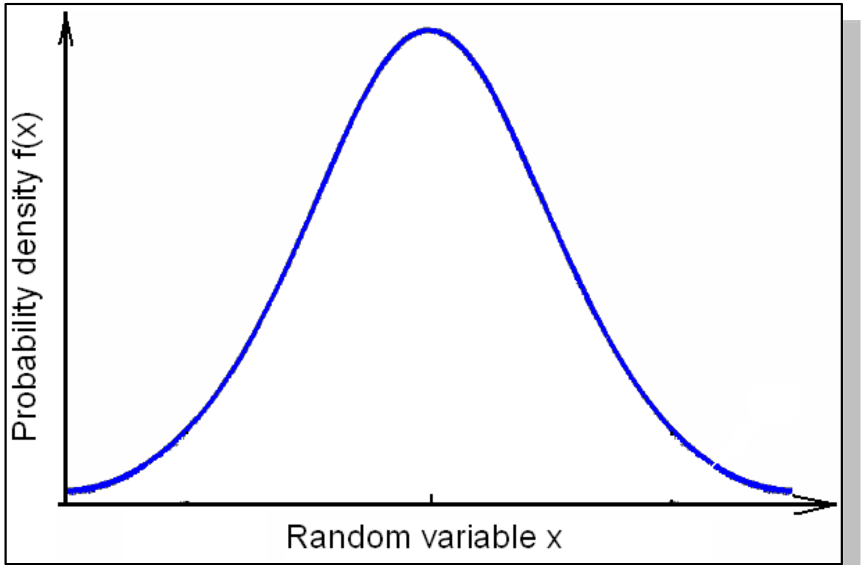
Q1.
P(0) =0.159

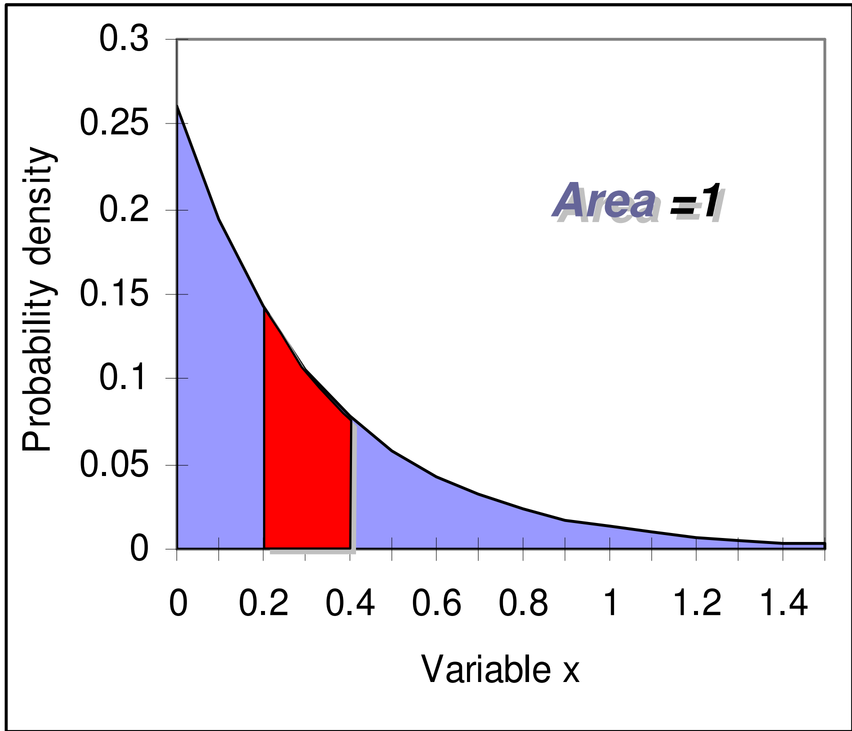
Q2.
P(>1) =P(2)+P(3)=0.464

In Excel use the function:

- `=HYPGEOMDIST(x,n,r,N)`

Probability density function
A function used to compute probabilities for a continuous random variable. The area under the graph of a probability density function over an interval represents probability.

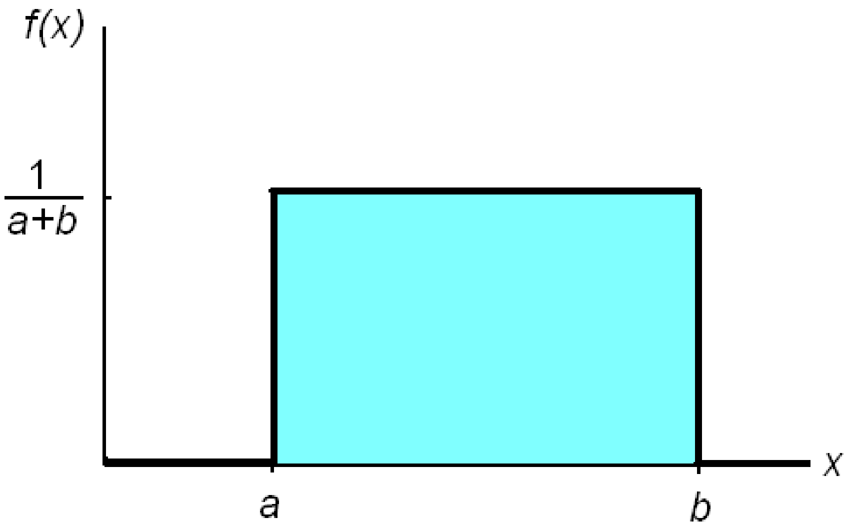




$$\int_x f(x)dx = 1$$

Uniform probability distribution

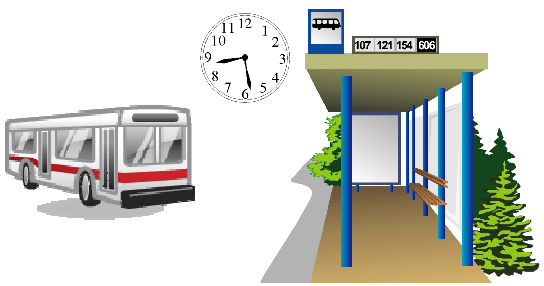
A continuous probability distribution for which the probability that the random variable will assume a value in any interval is the same for each interval of equal length.



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

$$E(x)=\mu=\frac{a+b}{2}$$

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

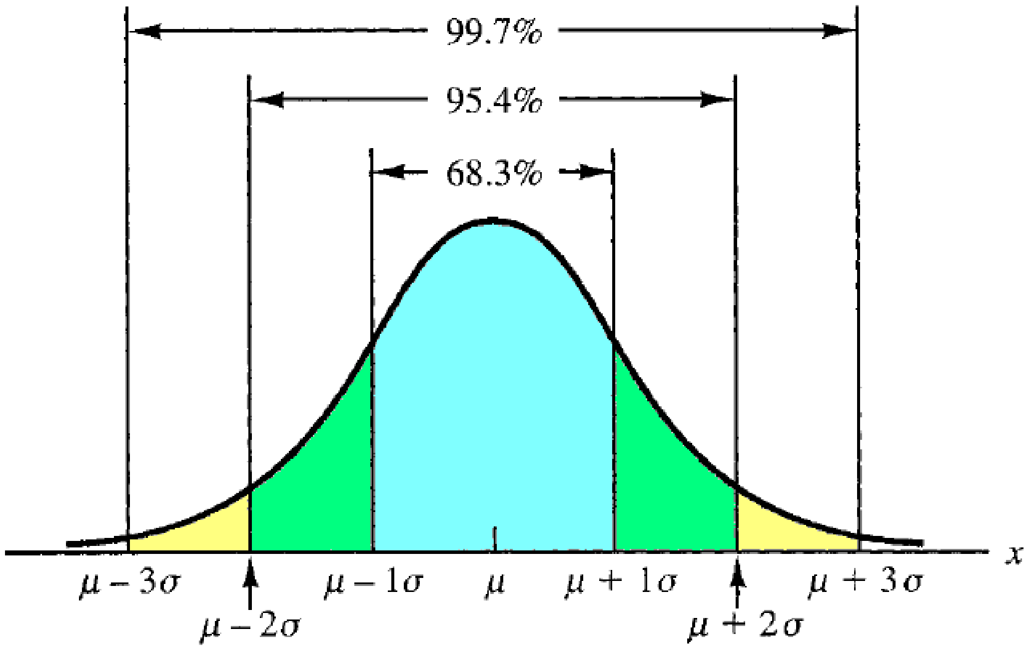


Example

The bus 22 goes every 7 minutes. You are coming to CHL bus station, having no idea about precise timetable. What is the distribution for the time, you may wait there?

Normal probability distribution
A continuous probability distribution. Its probability density function is bell shaped and determined by its mean μ and standard deviation σ.

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



In Excel use the function:

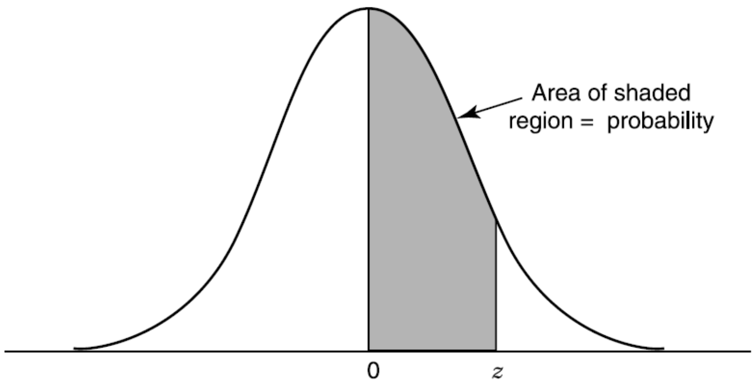
- **= NORMDIST(x,m,s,false)** for probability density function
- **= NORMDIST(x,m,s,true)** for cumulative probability function of normal distribution (area from left to x)

Standard Normal Probability Distribution

Standard normal probability distribution
A normal distribution with a mean of zero and a standard deviation of one.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$z = \frac{x - \mu}{\sigma}$$



z	.00	.01	.02	.03	.04	.05
.0	.0000	.0040	.0080	.0120	.0160	.0199
.1	.0398	.0438	.0478	.0517	.0557	.0596
.2	.0793	.0832	.0871	.0910	.0948	.0987
.3	.1179	.1217	.1255	.1293	.1331	.1368
.4	.1554	.1591	.1628	.1664	.1700	.1736
.5	.1915	.1950	.1985	.2019	.2054	.2088
.6	.2257	.2291	.2324	.2357	.2389	.2422
.7	.2580	.2612	.2642	.2673	.2704	.2734
.8	.2881	.2910	.2939	.2967	.2995	.3023
.9	.3159	.3186	.3212	.3238	.3264	.3289
1.0	.3413	.3438	.3461	.3485	.3508	.3531

In Excel use the function:

- `= NORMSDIST(z)-0.5`

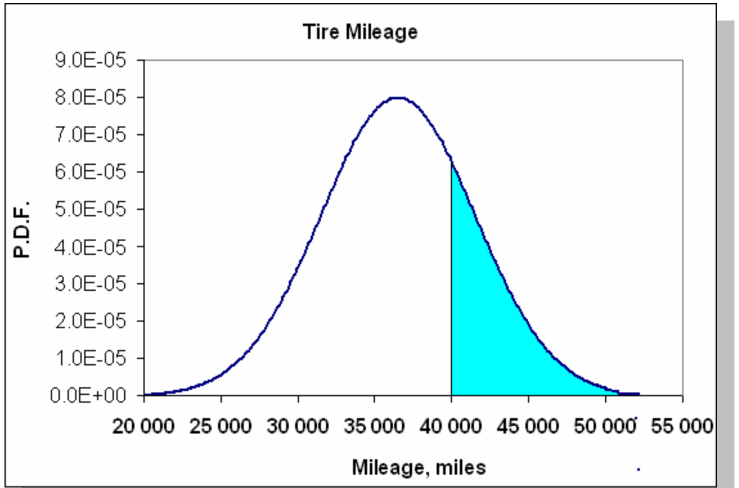
Example: Gear Tire Company

Example

Suppose the Grear Tire Company just developed a new steel-belted radial tire that will be sold through a chain of discount stores. Because the tire is a new product, Grear's managers believe that the mileage guarantee offered with the tire will be an important factor in the acceptance of the product. Before finalizing the tire mileage guarantee policy, Grear's managers want probability information about the number of miles the tires will last.

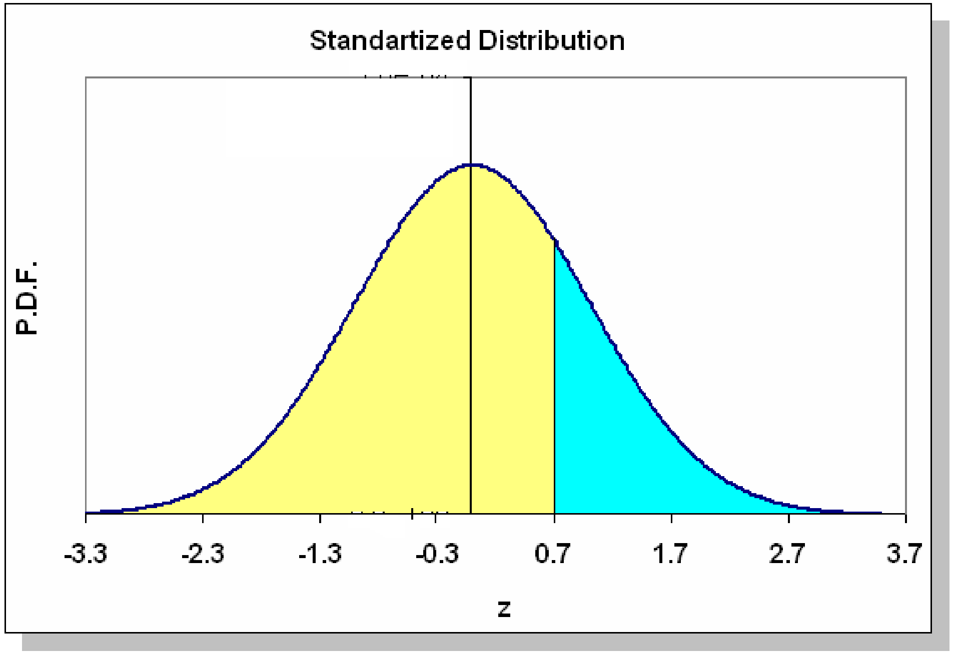
From actual road tests with the tires, Grear's engineering group estimates the mean tire mileage is **μ = 36 500** miles with a standard deviation of **σ = 5 000**. In addition, data collected indicate a normal distribution is a reasonable assumption.

What percentage of the tires can be expected to last more than 40 000 miles? In other words, what is the probability that a tire mileage will exceed 40 000?



Anderson et al Statistics for Business and Economics

Example: Gear Tire Company



1. Let's transfer from Normal distribution to Standard Normal, then z, corresponding to 40000 will be

$$z = \frac{40000 - 36500}{5000} = 0.7$$

2. Calculate the "blue" area P(z >0.7) using the table:

P(z>0.7) = 1 – P(z<0.7) = 1 – 0.5 – **P(0<z<0.7)** = 1 – 0.5 – 0.258 = **0.242**

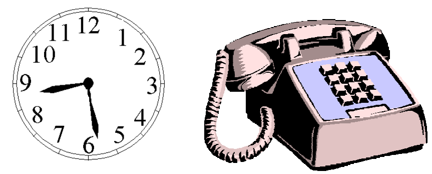
Alternatively in Excel

```
=1-NORMDIST(40000,36500,5000,true)
```

Example

Number of calls to an Emergency Service is on average 3 per hour b/w 2.00 and 6.00 of working days. What are the distribution of the time between the calls?

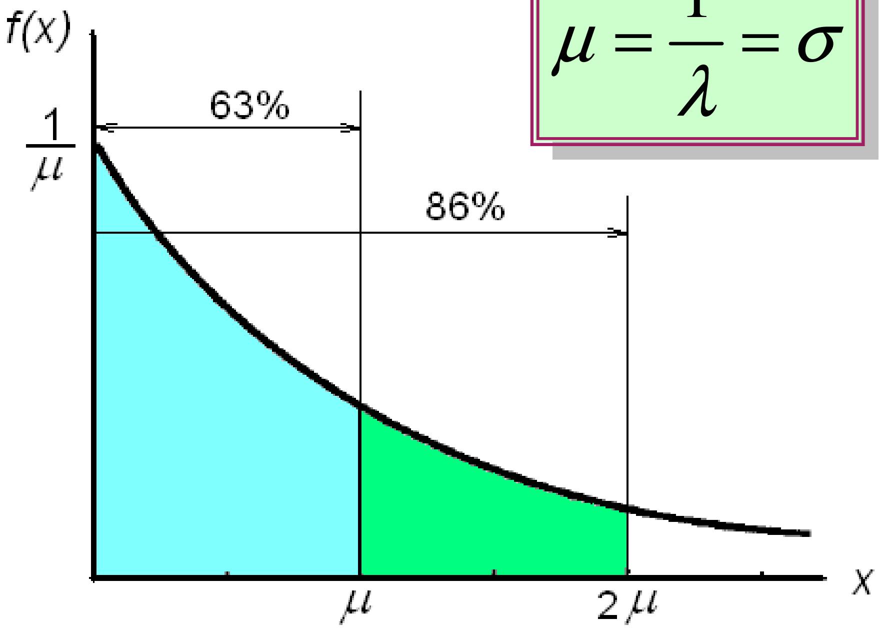
Time between calls to a reception



Exponential probability distribution

A continuous probability distribution that is useful in computing probabilities for the time between independent random events.

$$\mu = \frac{1}{\lambda} = \sigma$$



$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \qquad for\ x \geq 0, \mu > 0$$

$$f(x) = \lambda e^{-\lambda x}$$

Cumulative probability function

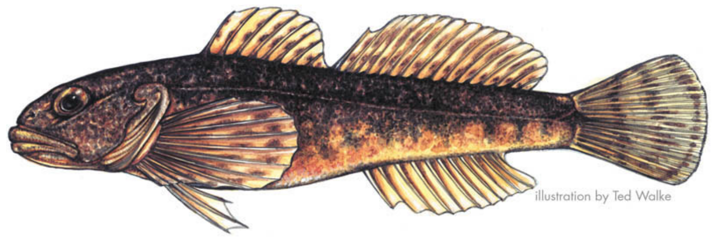
$$P(x \leq x_0) = F(x_0) = 1 - e^{-\frac{x_0}{\mu}}$$

Example: Exponential Distribution for Fish Counting

Example

An ichthyologist studying the *spoonhead sculpin* catches specimens in a large bag seine that she trolls through the lake. She knows from many years experience that on averages she will catch **2 fish per trolling**. Each trolling take ~**30 minutes**.

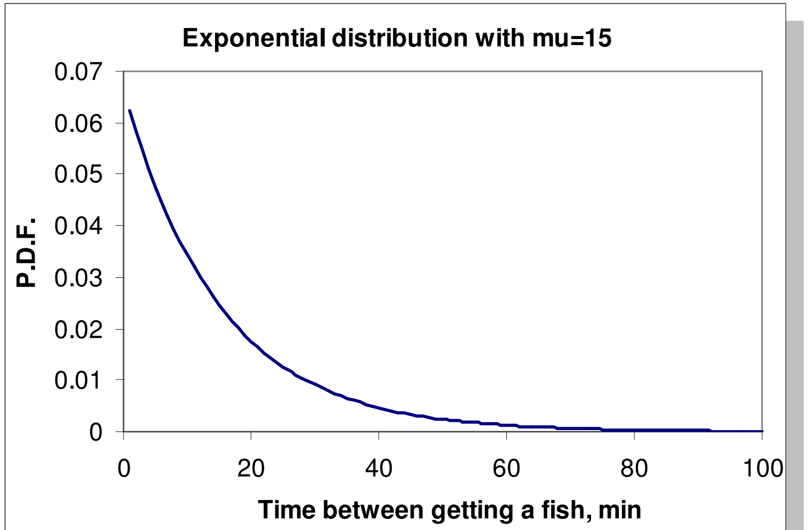
Find the probability of catching no fish in the next hour



In Excel use the function:

- `= EXPONDIST(x,1/mu,false)`

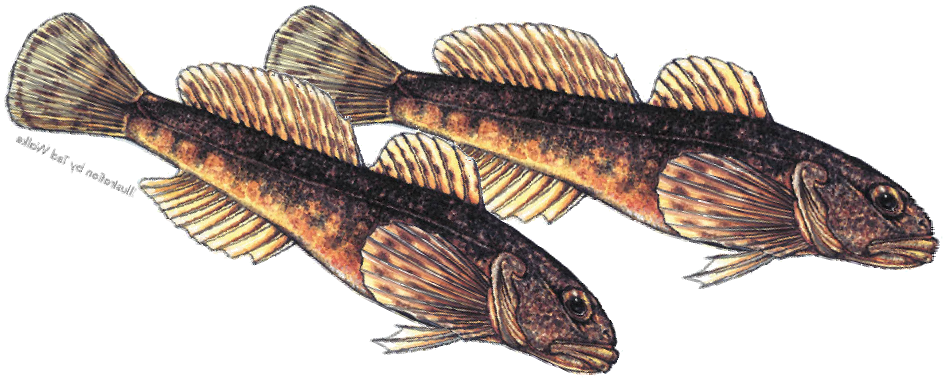
1. Let's calculate μ for this situation: $\mu = 30 / 2 = 15$ minutes



2. Use either a cumulative probability function or Excel to calculate:

$$P(x \geq 60) = 1 - P(x \leq 60) = 1 - F(60) = e^{-\frac{60}{15}} \approx 0.02$$

Thank you for your attention



to be continued...