UNIVERSITÉ DU LUXEMBOURG

LUXEMBOURG INSTITUTE OF HEALTH
RESEARCH DEDICATED TO LIFE

Bioinformatics group, Proteome and Genome Research Unit

BIOSTATISTICS

Lecture 3

Continuous Probability Distributions

dr. Petr Nazarov

petr.nazarov@lih.lu

6-03-2020

OUTLINE

- **Continuous probability distribution**
  - a continuous probability distribution
  - uniform probability distribution
  - normal probability distribution
  - exponential probability distribution

RANDOM VARIABLES

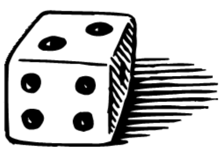
Random variable
A numerical description of the outcome of an experiment.

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.

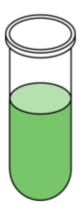
Roll a die



Number of calls to a reception per hour

Time between calls to a reception

Volume of a sample in a tube



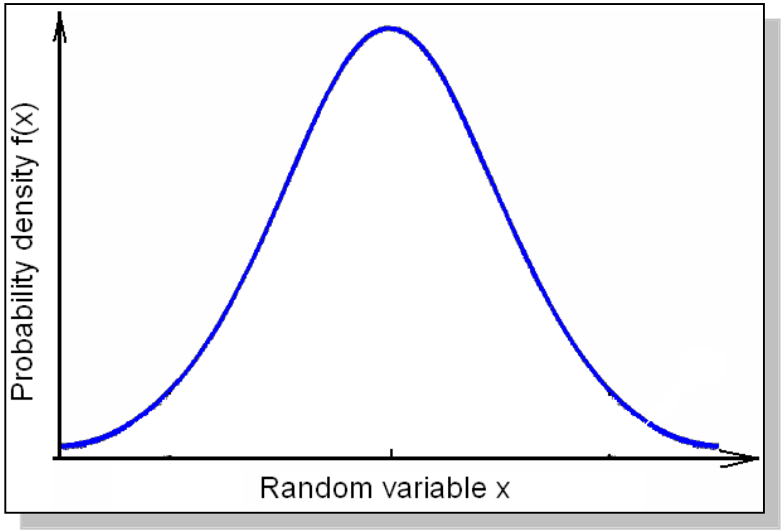
Weight, height, blood pressure, etc

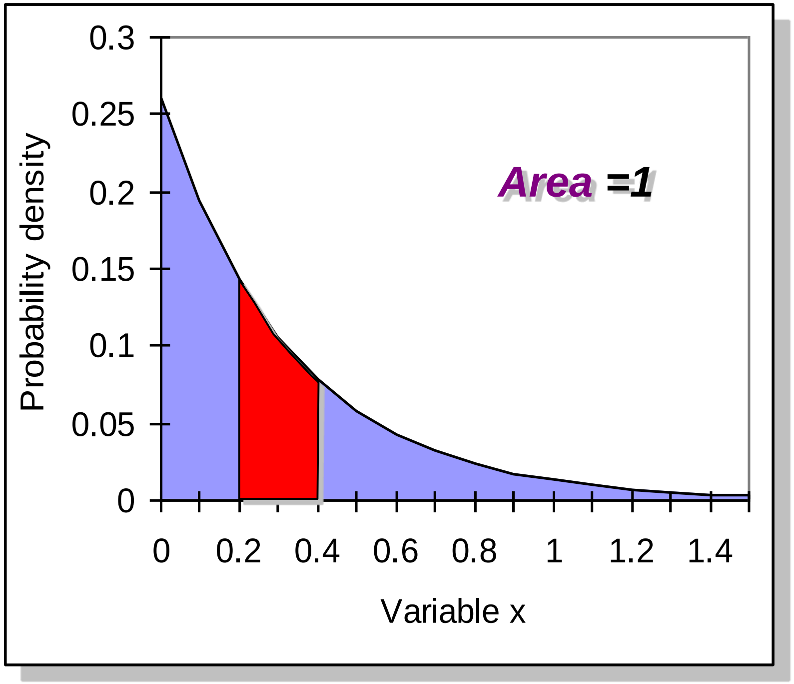
CONTINUOUS PROBABILITY DISTR.

Probability Density

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.



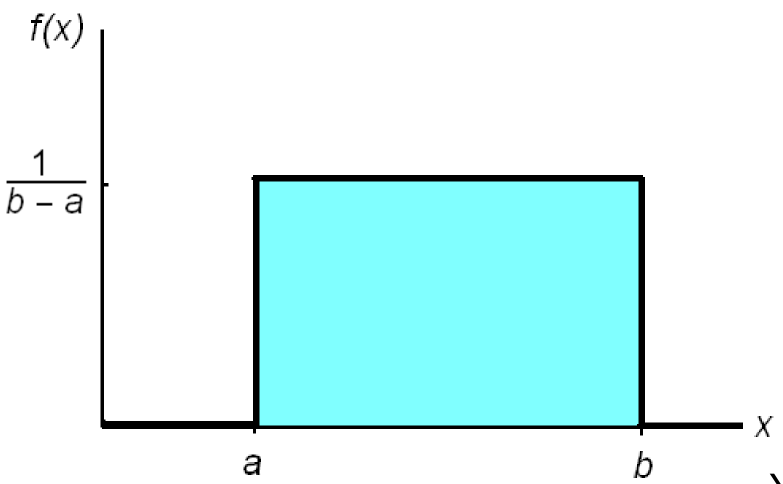


$$\int_x f(x) = 1$$

UNIFORM DISTRIBUTION

Uniform Probability Distribution

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.

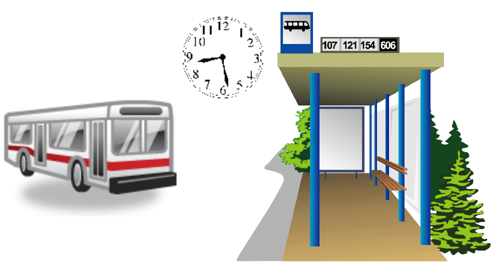


$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } a \le x \le b \\ 0, & \text{elsewhere} \end{cases}$$

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

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

You can generate a uniform random number b/w 0 and 1 using Excel function `=RAND()`



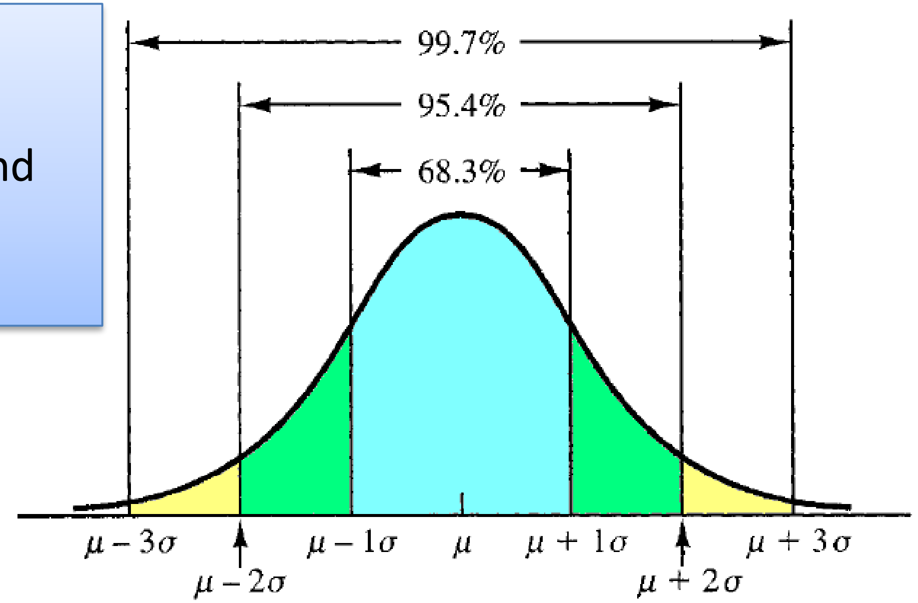
Example
Bus 3 goes every 10 minutes. You are coming to Ketten bus station, having no idea about precise timetable. What is the distribution for the time, you may wait there?

NORMAL DISTRIBUTION

Normal Probability Density Function

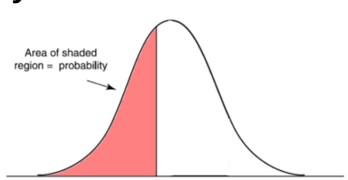
Normal (Gaussian) probability distribution
A continuous probability distribution. Its probability density function is bell shaped and determined by its mean μ and standard deviation σ.

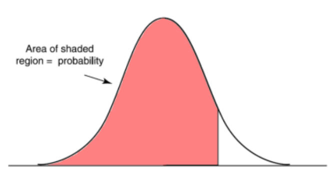
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



In Excel 2010 use the function:

- `= NORM.DIST(x,m,s,false)` for probability density function *(almost never used!!!)*
- `= NORM.DIST(x,m,s,true)` for cumulative probability function of normal distribution (area left to x, or probability for a value to be less then x)





`Excel 2003: NORMDIST`

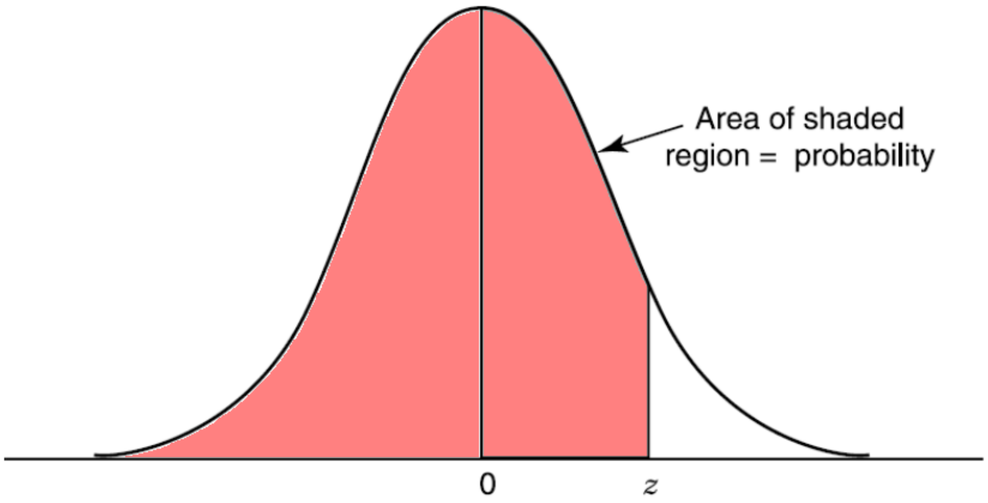
NORMAL DISTRIBUTION

Standard Normal Probability Distribution

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

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

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



In Excel 2010 use the function:

- `= NORM.S.DIST(z)`

`Excel 2003: NORMSDIST`

NORMAL DISTRIBUTION

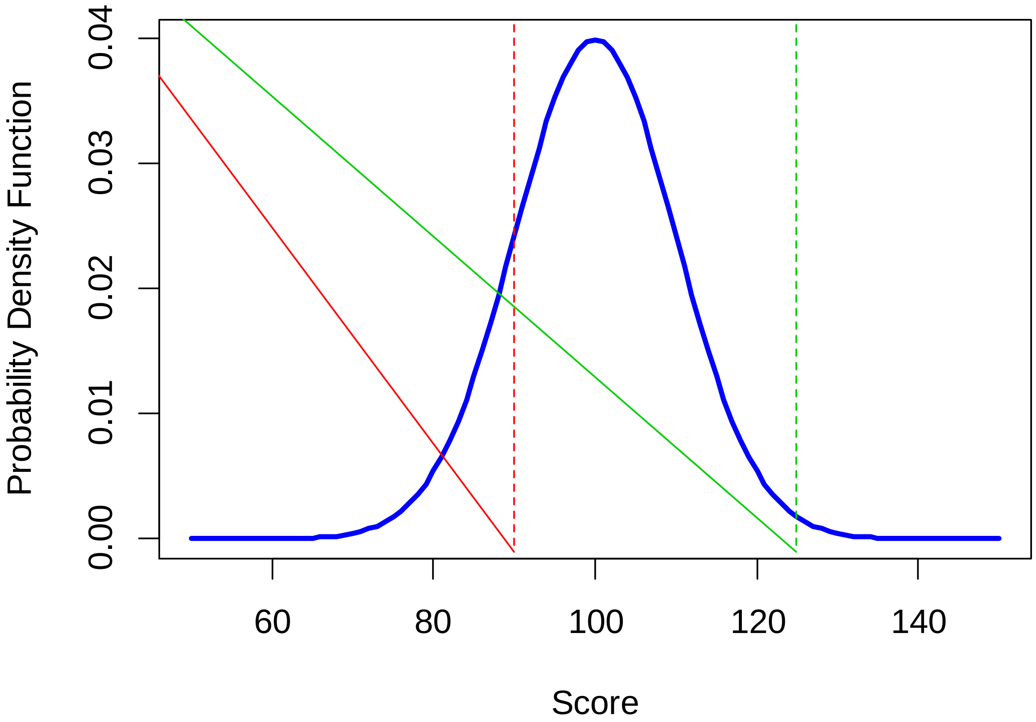
Example: Aptitude Test

Example

Suppose that the score on an aptitude test are normally distributed with a mean of 100 and a standard deviation of 10. (Some original IQ tests were purported to have these parameters.)

What is the probability that a randomly selected score is below 90?

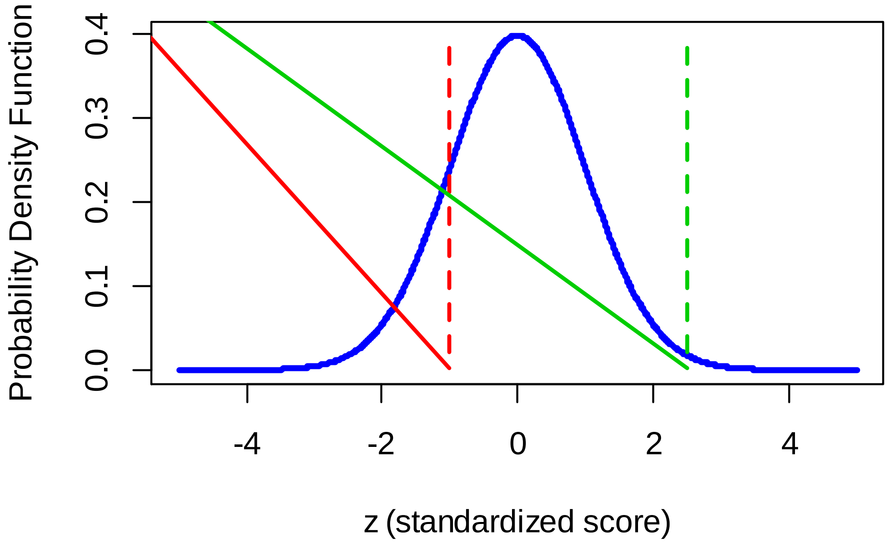
What is the probability that a randomly selected score is above 125?



Glover & Mitchel. An introduction to biostatistics

NORMAL DISTRIBUTION

Example: Aptitude Test



1. Let’s transfer from Normal distribution to Standard Normal, then z, corresponding to 90 will be

$$z_{x=90} = \frac{90-100}{10} = -1$$

$$z_{x=125} = \frac{125-100}{10} = 2.5$$

2. Calculate the area under the curve before z = -1 :

P(x<90) = P(z< -1) = NORM.S.DIST(-1;TRUE) = **0.159**

P(x>125) = P(z>2.5) = 1- P(z<2.5) =1- NORM.S.DIST(2.5,TRUE) = **0.006**

Alternatively in Excel

```
= NORM.DIST(90,100,10,true)
= 1-NORM.DIST(125,100,10,true)
```

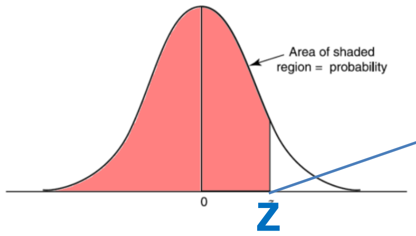
NORMAL DISTRIBUTION

Example: Inverted situation

Example

Suppose that the score on an aptitude test are normally distributed with a mean of 100 and a standard deviation of 10.

Find the score cutting top 5% respondent?



Assume that we know red area (probability *p*). Then limiting z can obtained using:

- **= NORM.S.INV(p)**
- **= NORM.INV(p,m,s)**

Two equal ways to solve:

(A) Classical (via z-score),

(A) = NORM.S.INV(1-0.05) = **1.64** (z-score)

= 1.64 * 10 + 100 = **116** (transform z into test score)

Excel 2003: NORMSINV

(B) Computational (via normal distrib.)

(B) = NORM.INV(1-0.05, 100, 10) = **116**

Excel 2003: NORMINV

Glover & Mitchel. An introduction to biostatistics

NORMAL DISTRIBUTION

Example of a Theoretical Question

A random variable *x* is distributed normally. Its probability density function is described by the following equation. Define the mean and variance of *x*.

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

$$f(x) = \frac{1}{\sqrt{6\pi}} e^{-\frac{(x-5)^2}{6}}$$

EXPONENTIAL DISTRIBUTION

Exponential Probability Distribution

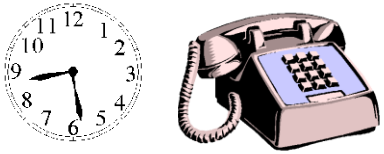
Example

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?

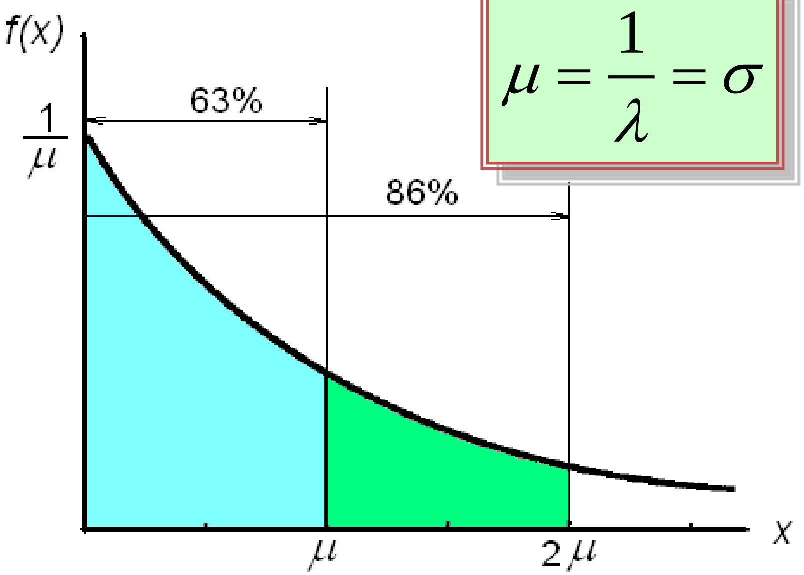
Exponential probability distribution

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

Time between calls to a reception



Generally: distance/time between events in a Poisson process with mean rate λ



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

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

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

Cumulative probability function

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

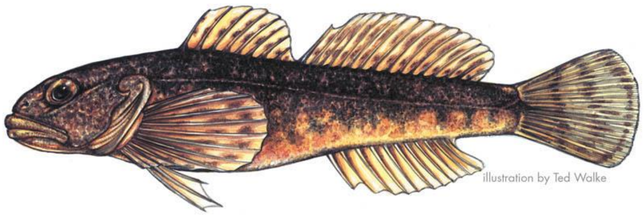
EXPONENTIAL DISTRIBUTION

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 fishes per trolling**. Each trolling take **~30 minutes**.

Find the probability of catching no fish in the next hour

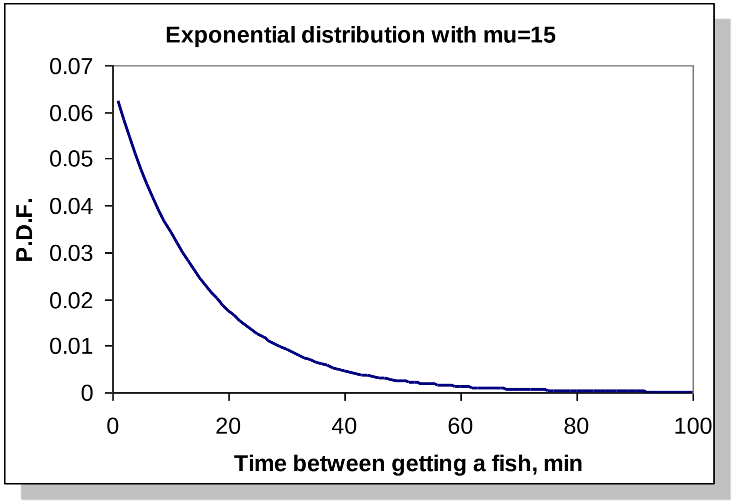


In Excel use the function:

- **= EXPON.DIST(x,1/mu,true)**

  Excel 2003: EXPONDIST

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



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

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

= 1-EXPON.DIST(60;1/15;TRUE) = 0.02

Take Home Message

Solving Tasks Ask These Questions:

Is distribution discrete or continuous?

Discrete

Binomial (n, p)	Sample of ***n*** objects with the same probability of event ***p***
Hyper-geometric (N, r, n)	Group of ***N*** objects with known number of events ***r*** among it. Sample of ***n*** objects is selected from this group.
Poisson (μ)	Counting events in space, time or along other measure

Continuous

See task. Usually it is mentioned.

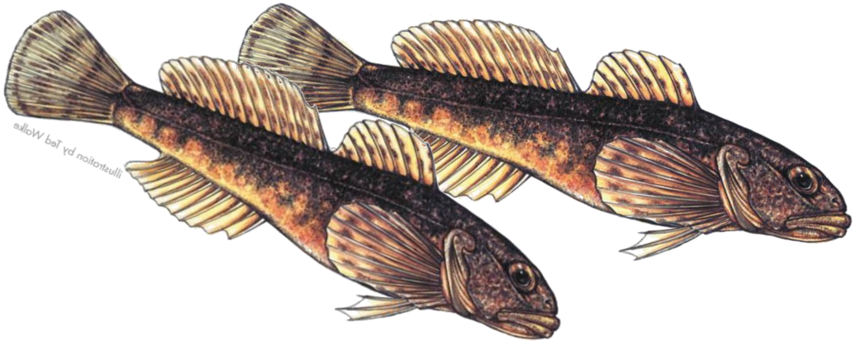
Otherwise – bell shaped are often approximated by normal, right-skewed - exponential.

Normal (μ, σ)

Exponential (μ) or (λ)

$\lambda = 1/\mu$

Thank you for your attention



to be continued…