Principles of Biostatistics and Informatics

4th Lecture: Frequently used distributions

27th September 2016

Dániel VERES

Population and Sample

Population

RANDOMNESS!

UNCERTAINTY!

Sample

Population
all the theoretically possible measurements

Sample
a subset of the population that could be examined

Questions in medical practice

Based on last year's data what is the probability that 4 flu vaccines will be sufficient if we are expecting a total number of 25 patients in our office?

How many births would be expected during the night if the yearly statistics shows 3000 deliveries?

How many student in our class will be able to do a hip replacement surgery based on their weight?

What is the probabiliy that a patient with 3.45 mmol/l K^+ level (it is out of the normal range) is healthy?

A flu/AIDS test is positive – what is the probability that I am truly ill?

Describing populations

How to feed it?

IGUANA

BISON

OSTRICH

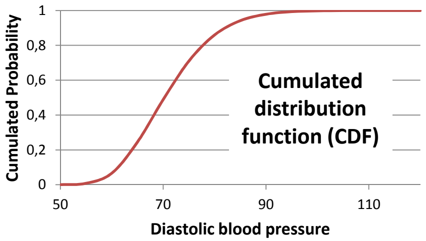
PANDA

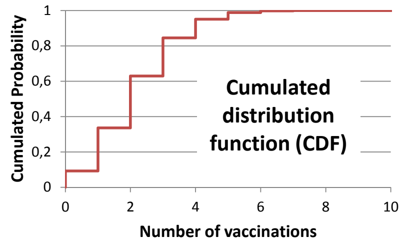
LADYBUG

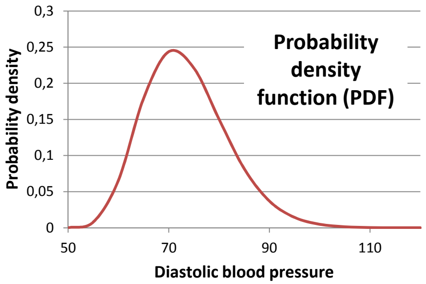
?

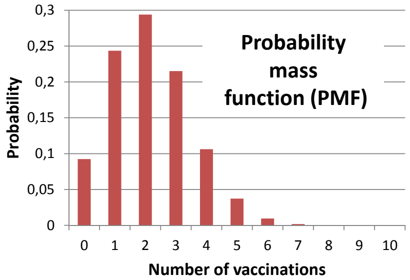
Theoretical Distributions

Continuous | **Discrete**

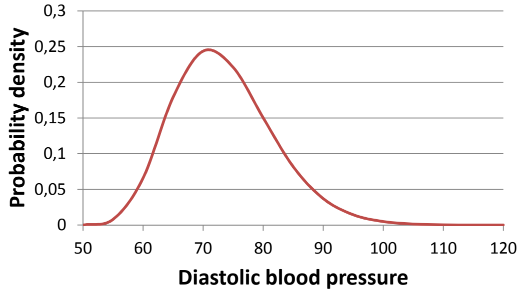








Theoretical Distributions



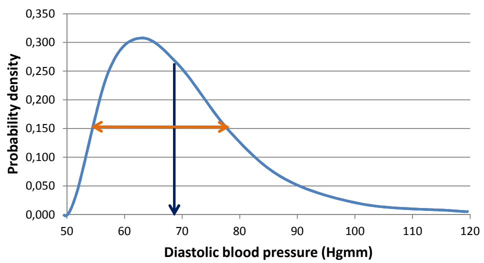
I know the probability for all value based on experiments. (very rare)

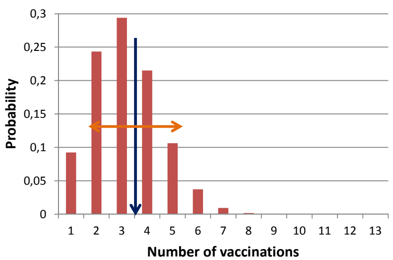
I can calculate (or estimate) the probability based on a few parameters using *special theoretical distributions.*

What are the parameters and which distribution should I use?

Parameters of Theoretical Distributions

Continuous | **Discrete**





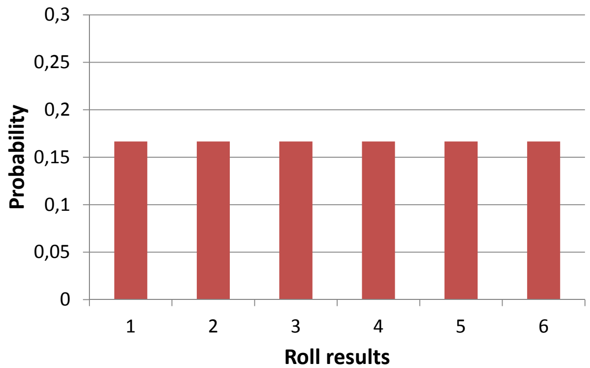
- ***Expected value(E) (location parameter)***

$$E(\xi)=\int_{-\infty}^{\infty} p_i \cdot x_i \qquad\qquad E(\xi)=\sum_{i=1}^{m} p_i \cdot x_i$$

- Theoretical variance (Var, D^2) (scale parameter)

$$Var(\xi)=E\left[\left(\xi-E(\xi)\right)^2\right]$$

Uniform Distribution



$$E(\xi)=\frac{1}{2}(a+b)$$

$$Var(\xi)=\frac{1}{12}(b-a)^2$$

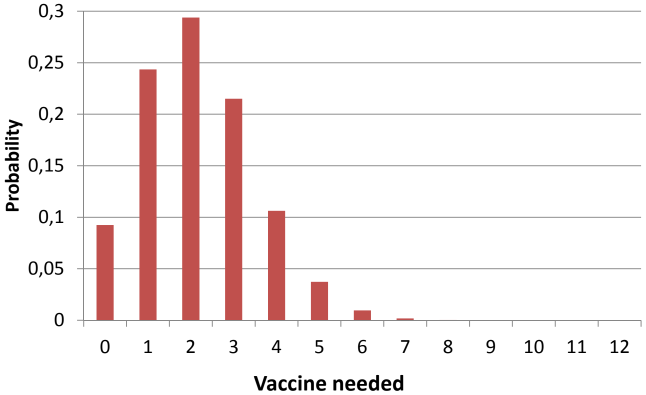
$$Var(\xi)=\frac{(b-a+1)^2-1}{12}$$

Distribution of a perfect die *(e.g.. probability of rolling 4)*

Ideal workload distribution throughout the day

Temperature distribution in an empty lecture hall

Binomial (Bernoulli) Distribution



$$E(\xi) = n \cdot p$$

$$Var(\xi) = n \cdot p \cdot (1-p)$$

$$P = \binom{n}{k} \cdot (p)^k \cdot (1-p)^{(n-k)}$$

Distribution of vaccine needed on a day

General: a phenomenon is repeated n times it occurs k times

If the probability of the occurrence is small it tends to Poisson distribution

If n is large and p is close to 0.5 it tends to Gaussian distribution

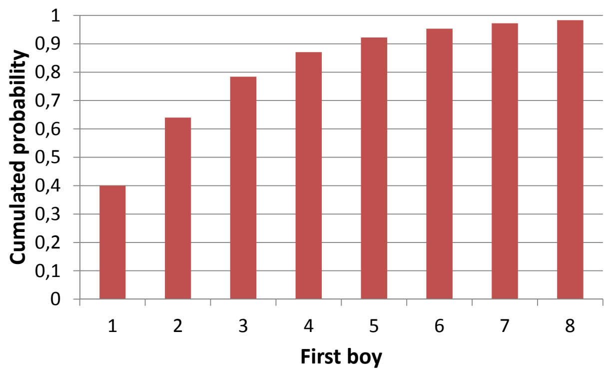
Probability Calculus Example

During last year's flu epidemic 402 out of the total 2989 patients who turned up at a doctor's office required vaccination. Based on last year's data what is the probability that 4 vaccines will be sufficient (exactly, i.e. no vaccines will be left), if we are expecting a total number of 25 patients?

$$P = \binom{n}{k} \cdot (p)^k \cdot (1-p)^{(n-k)} = \binom{25}{4} \cdot \left(\frac{402}{2989}\right)^4 \cdot \left(1 - \frac{402}{2989}\right)^{(25-4)} \approx 0{,}2$$

How to calculate (in excel)?

Geometric Distribution



$$E(\xi) = \frac{1}{p}$$

$$Var(\xi) = \sqrt{\frac{1-p}{p^2}}$$

$$P = p \cdot (1-p)^{(n-1)}$$

Independent sequence of Bernoulli trials

The first patient when we need a nurse's to help

The probability to get the first boy from a delivery during the night

To get the first kidney from a random sample that suits to the transplantation.

The Beginnings of probability calculus... Let's Play a Game

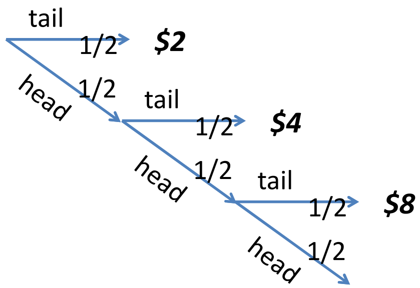
Coin tossing game:

- The pot starts at 2 dollars and it is *doubled* every time a head appears.
- The first time a tail appears, the game ends and Peter wins whatever is in the pot:
  - Peter wins \$2 if a *tail appears on the first toss*
  - Peter wins \$4 if a head appears on the first toss and a *tail on the second*
  - 8\$ if a head appears on the first two tosses and a *tail on the third*

Q: What would be a „fair" price for taking part in this game?

A: It would be the expected prize for one game.

The Beginnings...



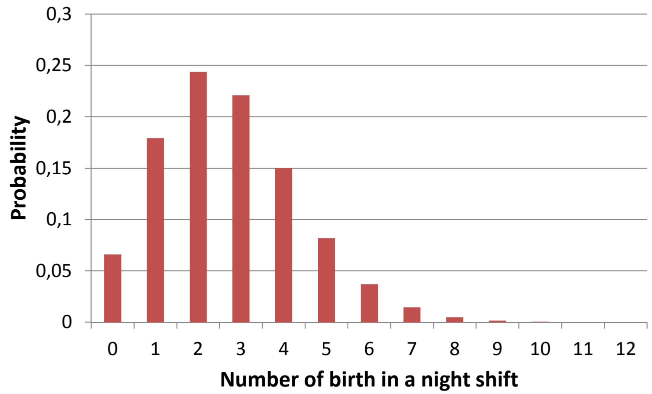
The theoretical „fair” price: **expected („mean”) infinite $ in one game!**

$$\frac{1}{2}\cdot 2+\frac{1}{2^2}\cdot 4+\ldots+\frac{1}{2^n}\cdot 2^n$$

In practice: we never win an infinite sum...

Buffon: Made 2048 tosses and won $9.82 on average (mean of the prizes). One million tossing: $10.94

Poisson Distribution



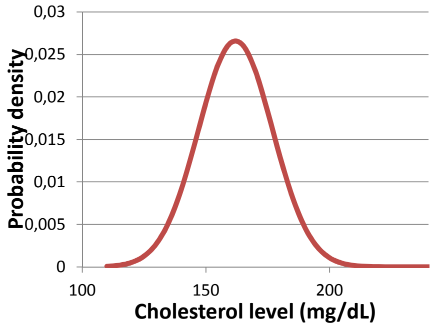
$$E(\xi)=\lambda$$

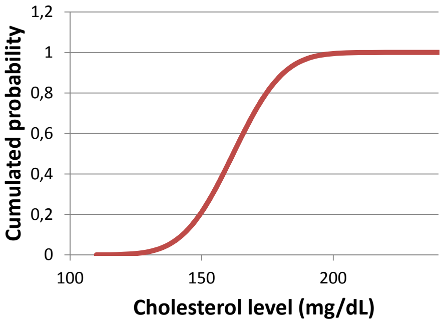
$$Var(\xi)=\lambda$$

$$P=\frac{\lambda^x}{x!}\cdot e^{-\lambda}$$

Number of births during night shift
Number of white blood cells in the field of view
Number of decayed atoms in a radioactive substance during a given time interval

Normal (Gaussian) Distribution I.





Cholesterol level, glucose level.....
Height, BMI...
Diastolic blood pressure of adults
......

$$E(\xi)=\mu$$

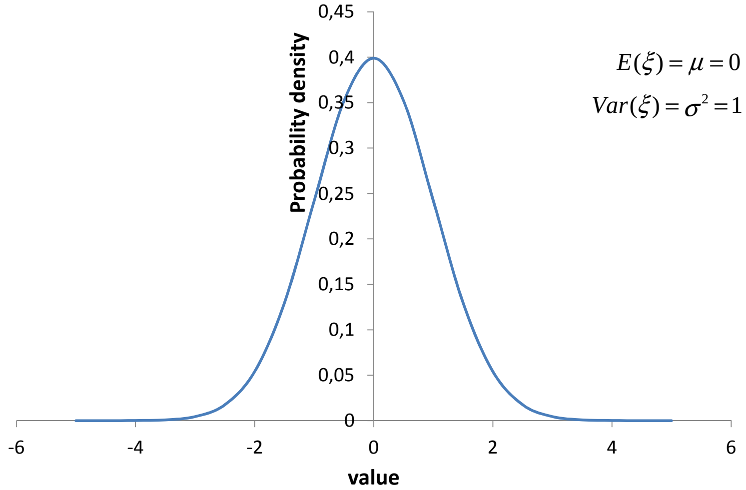
$$Var(\xi)=\sigma^2$$

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

Gaussian Distribution II.

Central limit theorem: for given conditions, adding a large number of independent variables yields a normally distributed variable.

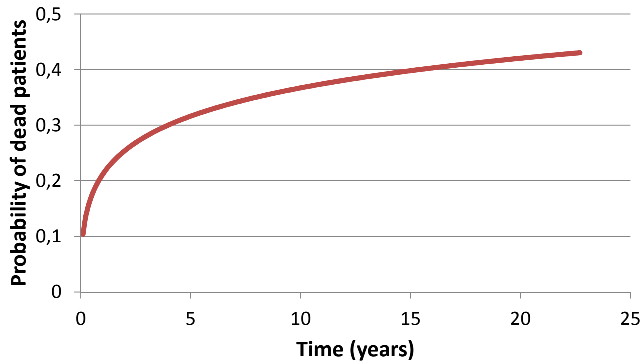
Standard normal distribution



$$E(\xi) = \mu = 0$$

$$Var(\xi) = \sigma^2 = 1$$

Lognormal (Galton) Distribution

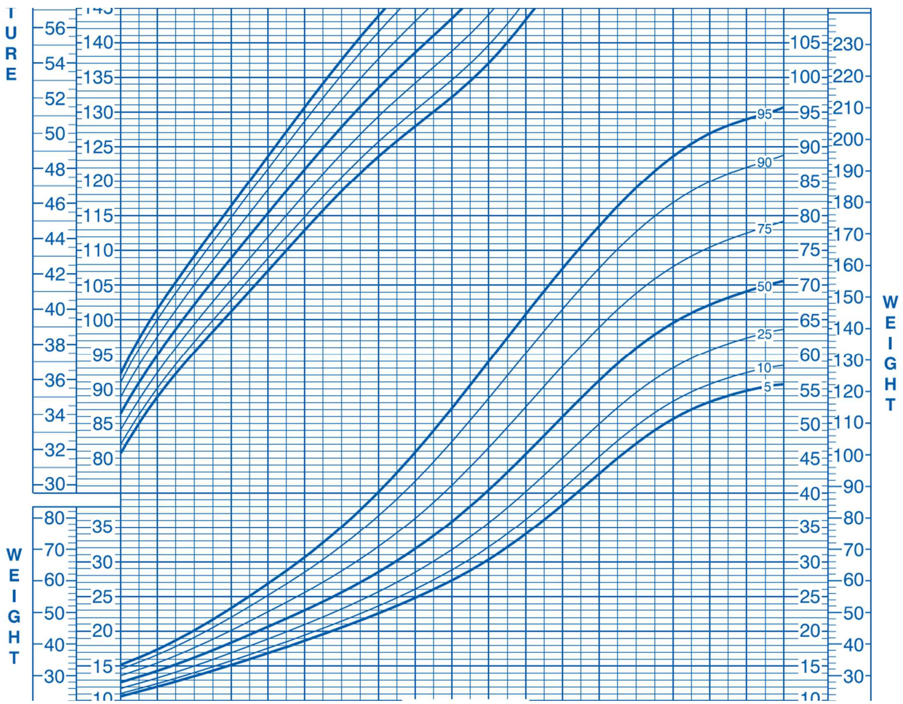


$$E(\xi) = e^{\mu+\sigma^2/2}$$

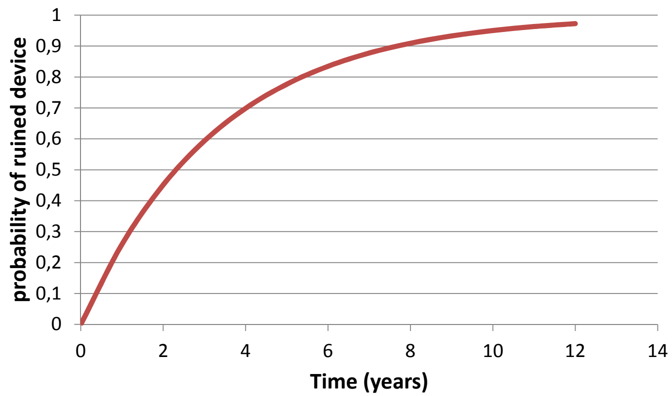
$$Var(\xi) = (e^{\sigma^2} - 1) \cdot e^{2\mu+\sigma^2}$$

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

Body mass in childhood
Survival time of a cancer



Exponential Distribution



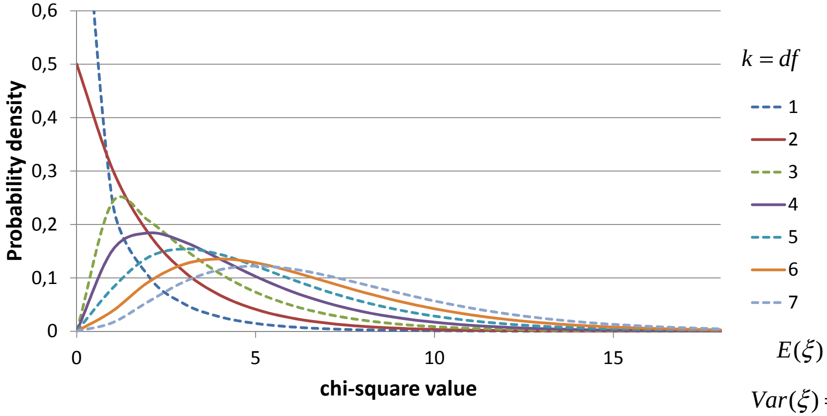
$$E(\xi) = \frac{1}{\lambda}$$

$$Var(\xi) = \frac{1}{\lambda^2}$$

$$P = \lambda \cdot e^{-\lambda x}$$

Anesthetic equipment operating time (before the first error).
Lifetime of the individual atoms in the course of radioactive decay.

Chi-square (χ^2) distributions

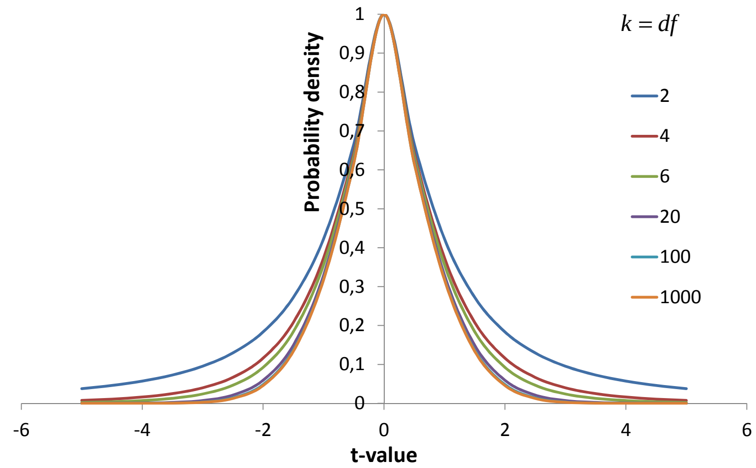


$E(\xi) = k$

$Var(\xi) = 2k$

Chi-square value: sum of the squares of independent, standard normal variables.

Student's t distributions



Transformations of distributions

- Addition of a constant

  $E(\eta) = E(\xi) + k \qquad Var(\eta) = Var(\xi)$
- Multiplication with a constant

  $E(\eta) = E(\xi) * k \qquad Var(\eta) = Var(\xi) * k^2$
- Standardization

  Addition then multiplication $\eta = (\xi - E(\xi)) * \frac{1}{\sqrt{Var(\xi)}} = \frac{(\xi - E(\xi))}{\sqrt{Var(\xi)}}$
- Addition of variables

  $E(\eta) = E(\xi) + E(\omega) \qquad Var(\eta) = Var(\xi) + Var(\omega) \leftarrow independent$

  Stable distribution: has the same distribution
- Multiplication of variables

  $E(\eta) = E(\xi) * E(\omega)$

Test Questions #1

- How you can calculate the expected value of a continuous distribution?
- How you can calculate the expected value of a discrete distribution?
- Which central tendency equal with the expected value in case of a population?
- Define the theoretical variance.
- What are the two indicators that define exactly a special distribution?
- How does the frequency distribution of a uniform distribution looks like?
- How does the frequency distribution of a Poisson distribution looks like?
- How does the frequency distribution of a Bernoulli distribution looks like?
- How does the frequency distribution of a Geometric distribution looks like?
- How does the frequency distribution of a Gaussian distribution looks like?
- How does the cumulative frequency distribution of a Gaussian distribution looks like?
- How does the frequency distribution of a exponential distribution looks like?
- How does the frequency distribution of a lognormal distribution looks like?
- Give two example for uniform distribution.
- Give two example for binomial distribution.
- Give two example for Poisson distribution.
- Give two example for normal distribution.
- Give two example for lognormal distribution.
- Give two example for geometric distribution.
- Give two example for exponential distribution.
- How you can calculate the expected value of a uniform distribution?
- How you can calculate the expected value of a binomial distribution?
- How you can calculate the expected value of a lognormal distribution?
- How you can calculate the expected value of a exponential distribution?
- How you can calculate the expected value of a Poisson distribution?
- How you can calculate the expected value of a Gaussian distribution?
- What is the central limit theorem?
- Why are the most of the medical variables normally distributed?

Test Questions #2

- Give a general description when we get a binomial distribution.
- Give a general description when we get a Poisson distribution.
- When we get lognormal distribution instead of normal distribution?
- How to convert lognormal distribution to normal distribution?