UNIVERSITÉ DU LUXEMBOURG

LUXEMBOURG INSTITUTE OF HEALTH
RESEARCH DEDICATED TO LIFE
Multiomics Data Science

BIOSTATISTICS

Lecture 2

Discrete Probability Distributions

Petr Nazarov

Email: petr.nazarov@lih.lu
Skype: pvn.public

26-02-2021

OUTLINE

Lecture 2

- **Random variables**
- **Discrete probability distributions**
  - discrete probability distribution
  - expected value and variance
  - discrete uniform probability distribution
  - binomial probability distribution
  - hypergeometric probability distribution
  - Poisson probability distribution

RANDOM VARIABLES

Random Variables

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

A random variable is always a numerical measure.

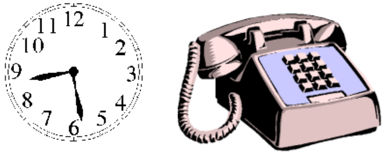
Discrete random variable
A random variable that may assume either a finite number of values or an infinite sequence of values.

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

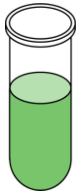
Roll a die

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

DISCRETE PROBABILITY DISTRIBUTIONS

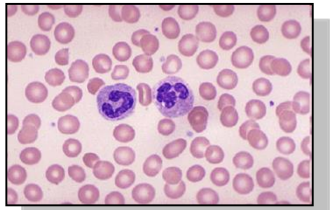
Probability Distribution and Probability Function

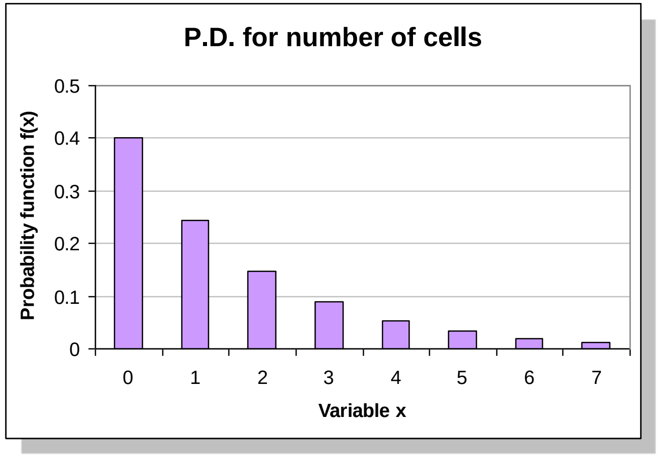
Probability distribution
A description of how the probabilities are distributed over the values of the random variable.

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

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

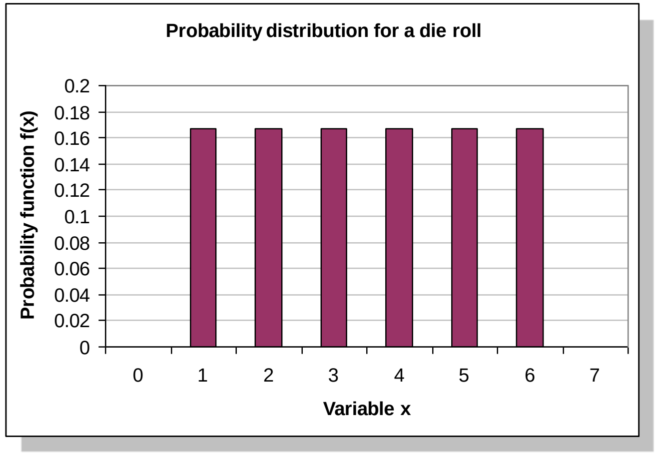
Number of cells under microscope
Random variable X:
x = 0
x = 1
x = 2
x = 3
…





Roll a die
Random variable X:

x = 1
x = 2
x = 3
x = 4
x = 5
x = 6



DISCRETE PROBABILITY DISTRIBUTIONS

Expected Value and Variance

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

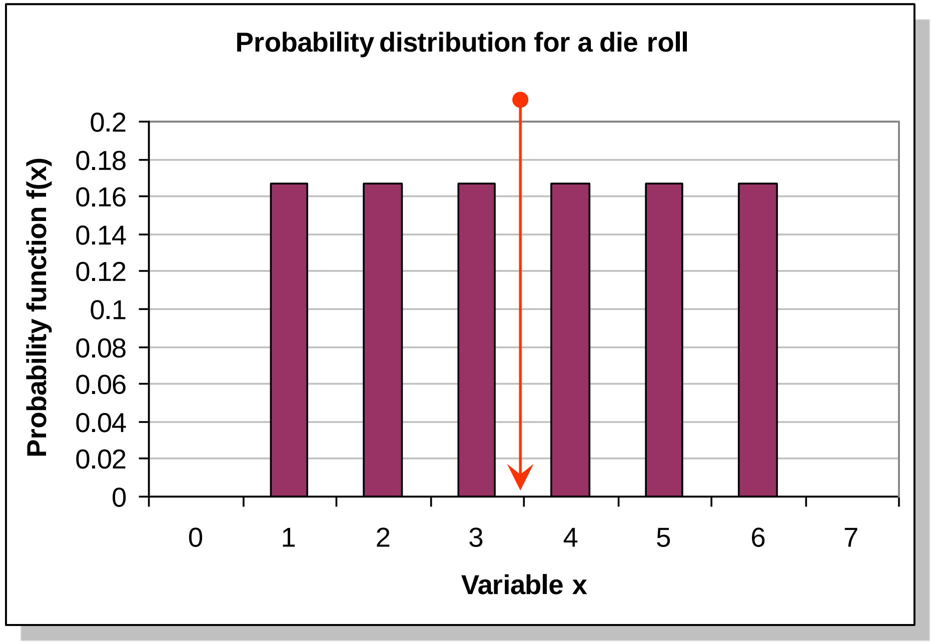
$$E(x) = \mu = \sum x f(x)$$

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

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

Roll a die
Random variable X:

x = 1
x = 2
x = 3
x = 4
x = 5
x = 6



DISCRETE UNIFORM DISTRIBUTION

Discrete Uniform Probability Function

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

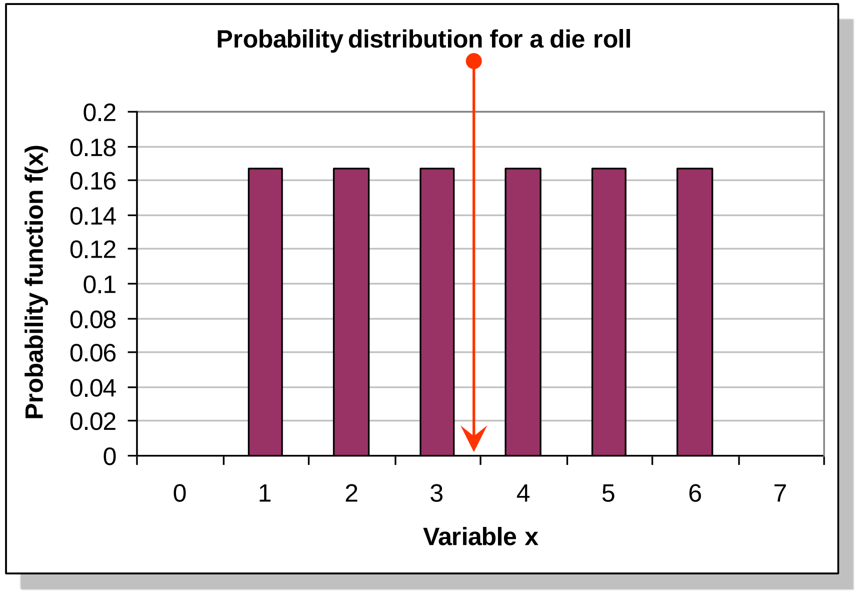
$$f(x) = \frac{1}{n}$$

n – number of values of x

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

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

$\mu = 3.5$
$\sigma^2 = 2.92$
$\sigma = 1.71$



BINOMIAL DISTRIBUTION

Binomial Experiment

Example

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

Binomial experiment

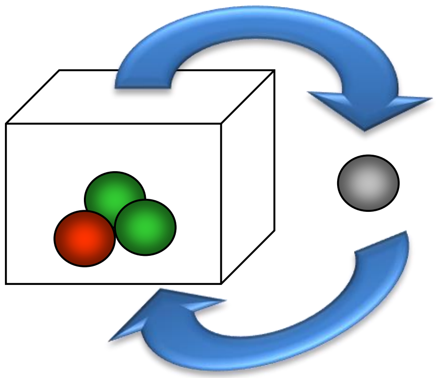
An experiment having the four properties:

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

BINOMIAL DISTRIBUTION

Binomial Probability Function

Binomial probability distribution

A probability distribution showing the probability of ***x*** successes in ***n*** trials of a binomial experiment, when the probability of success ***p*** does not change in trials.

$$C_x^n \equiv \binom{n}{x} \equiv \frac{n!}{x!(n-x)!}$$

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$$

$$0! = 1$$

Probability distribution for a binomial experiment

$$f(x,n,p) = C_x^n p^x (1-p)^{(n-x)}$$

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

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

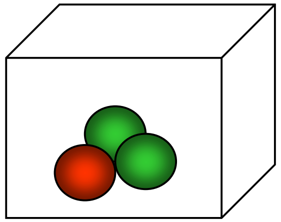
Excel:

```
BINOM.DIST(x,n,p,false)
```

R:

```
dbinom(x,n,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

BINOMIAL DISTRIBUTION

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?

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

p = 0.1
n = 5

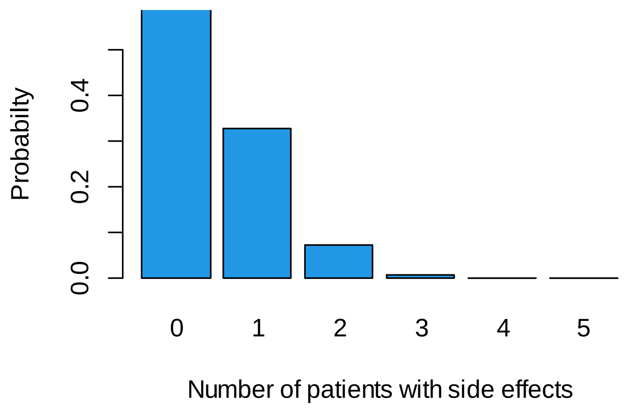
```
= BINOM.DIST(x,n,p,false)
```

1.

0	0.59049
1	0.32805
2	0.0729
3	0.0081
4	0.00045
5	0.00001

2. 0.91854

3. 0.5




```
p = 0.1
n = 5
x = 0:5
fx = dbinom(x, size=n, prob=p)
# Question 1 ->
barplot(fx, names.arg=x, col=4,
xlab="Number of patients with side effects",
ylab="Probabilty")
sum(dbinom(0:1,n,p)) # <- Question 2
n*p # <- Question 3
```

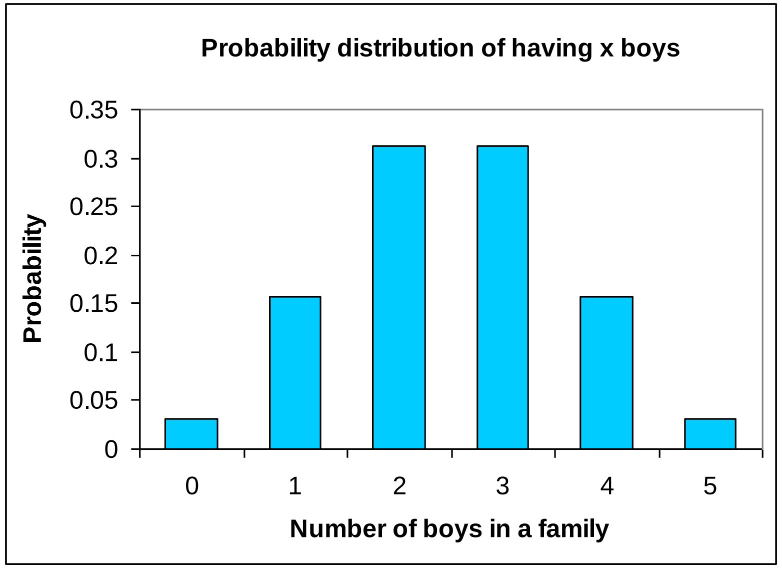
BINOMIAL DISTRIBUTION

Practical : Binomial Experiment

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 or 5) = P(0) + P(5)
= 0.03 + 0.03 = **0.06**

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

HYPERGEOMETRIC DISTRIBUTION

Hypergeometric Experiment

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.

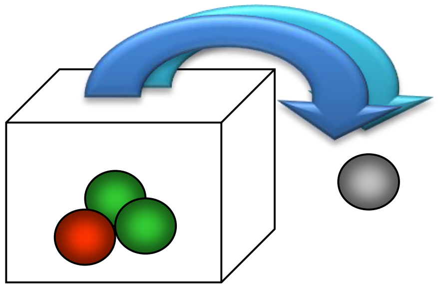
$$E(x) = \mu = n\left(\frac{r}{N}\right)$$

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

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

1. Assign x, n, r, N
2. Calculate p(x) using **R** with a slight modification:

```
dhyper(x,r,N-r,n)
```



n trials

1. Assign x, n, r, N
2. Calculate p(x) using **Excel**

```
= HYPGEOM.DIST (x,n,r,N,false)
```

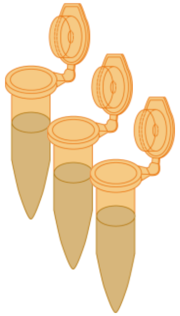
HYPERGEOMETRIC DISTRIBUTION

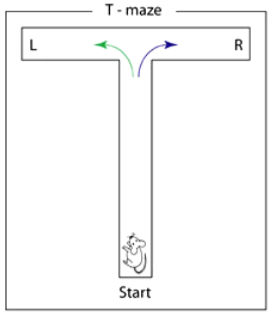
Example: Hypergeometric Distribution

Example

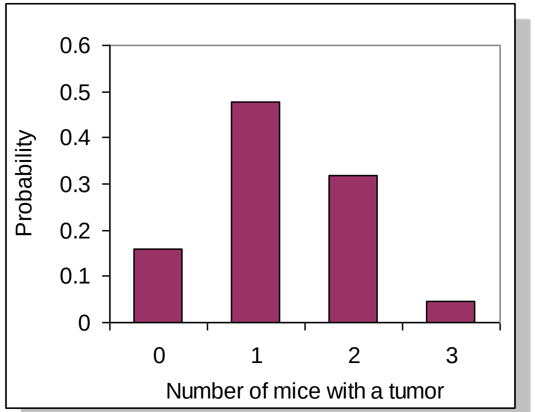
There are 12 mice, of which 5 have an early brain tumor. A researcher randomly selects 3 of 12.

1. What is the probability that none of these 3 has a tumor?
2. What is the probability that more than 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)=0.364

```
= HYPGEOM.DIST (x,n,r,N, false)
```

```
N = 12
r = 5
n = 3
x = 0:min(n,r)
fx = dhyper(x,r,N-r,n)
fx
barplot(fx, names.arg=x)

# Question 1
dhyper(0,r,N-r,n)

# Question 2
sum(dhyper(2:3,r,N-r,n))
```

POISSON DISTRIBUTION

Poisson Probability Function

Example

Number of calls to an Emergency Service is on average 3 per hour b/w 2 a.m. and 6 a.m. 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.

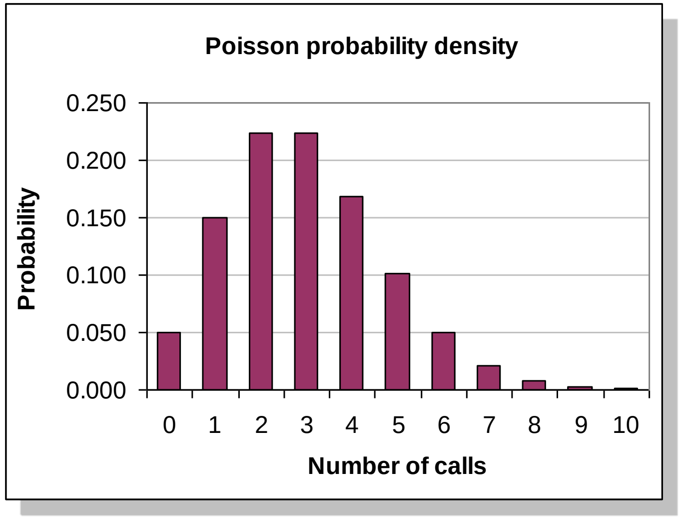
Poisson probability function
The function used to compute Poisson probabilities.

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

$$\mu = \sigma^2$$

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




```
= POISSON.DIST(x,mu,false)
```

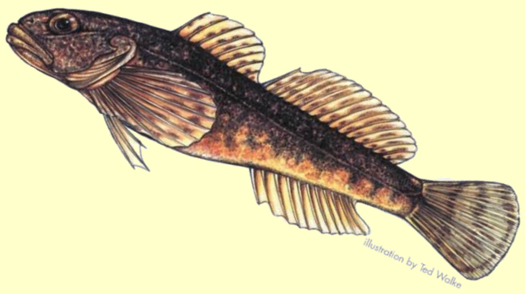
```
dpois(x,mu)
```

POISSON DISTRIBUTION

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 fishes per trolling.



1. Draw distribution

Find the probabilities of catching:

2. No fish;

3. Less than 4 fishes;

4. More than 1 fish.

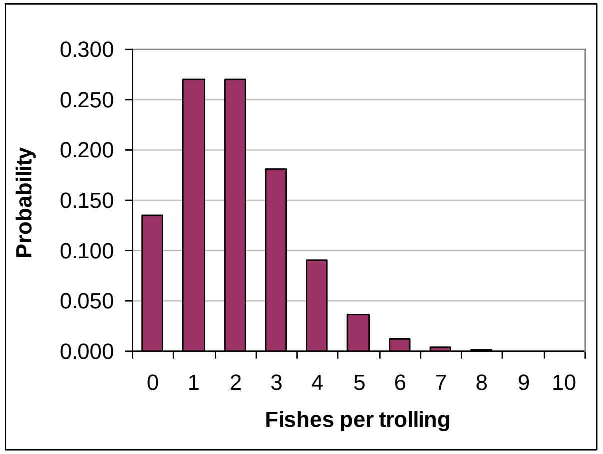
```
= POISSON.DIST(x,mu,false)
```

```
mu = 2
x = 0:10
fx = dpois(x,mu)
fx
# Question 1
barplot(fx, names.arg=x)

# Question 2
dpois(0,mu)

# Question 3
sum(dpois(0:3,mu))

# Question 4
1-sum(dpois(0:1,mu))
```



Q2.
P(0) = 0.135

Q3.
P(<4) = P(0)+P(1)+P(2)+P(3)=0.857

Q4.
P(>1) =1-P(0)-P(1)=0.594

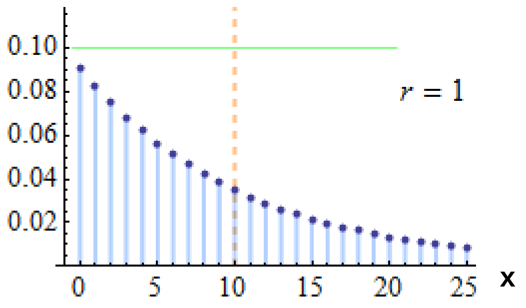
Glover , Mitchell, An Introduction to Biostatistics

NEGATIVE BINOMIAL DISTRIBUTION

Brief introduction

You toss a coin. How many heads will you observe before you got 3 tails?

How many blood samples you need to collect until you got 5 patients with a certain mutation?



Ref. Wikipedia

Negative binomial distribution *it can differ!*
A probability distribution showing the probability of ***x*** occurrences of successes until ***r****>0* failures are observed.

$$f(x,r,p) = C_x^{x+r-1} p^x (1-p)^r$$

$$E(x) = \frac{pr}{1-p} \qquad Var(x) = \frac{pr}{(1-p)^2}$$

Generally speaking:

- you have uncorrelated **and exclusive** events of 2 kind (A,B)
- Their probabilities do not change: P(A)=const, P(B)=const, P(A) + P(B) = 1
- You count x - number of A until certain number r of opposite B appear.

Here we denote *P(A)* = *p*, *P(B)* = *1-p*. People can use an inverted annotation, but then formulas will change!

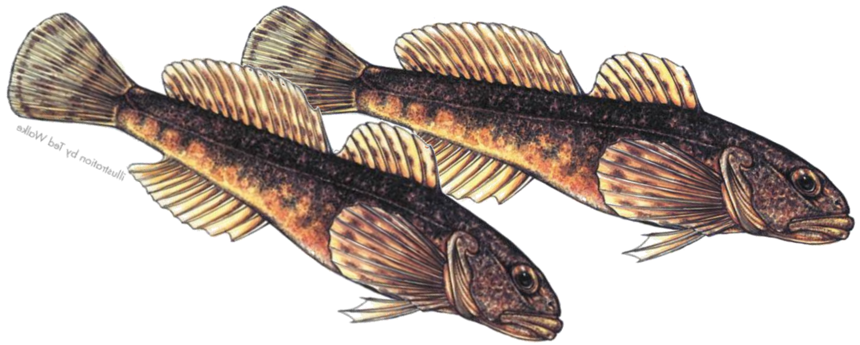
Why do we need NB ?
NB is similar with Poisson. But Poisson has mean = variance!
In the reality it is often not so. E.g. – **mutations in DNA**

`NEGBINOM.DIST (x,r,1-p,..)`

`dnbinom(...)`

QUESTIONS ?

Thank you for your attention



to be continued…

TWO GOATS AND A CAR







showman

player

1. P. guesses a door
2. S. looks inside and opens one of «goat» doors
3. P. should either keep his choice or change his mind.

There is a strategy to win!!!