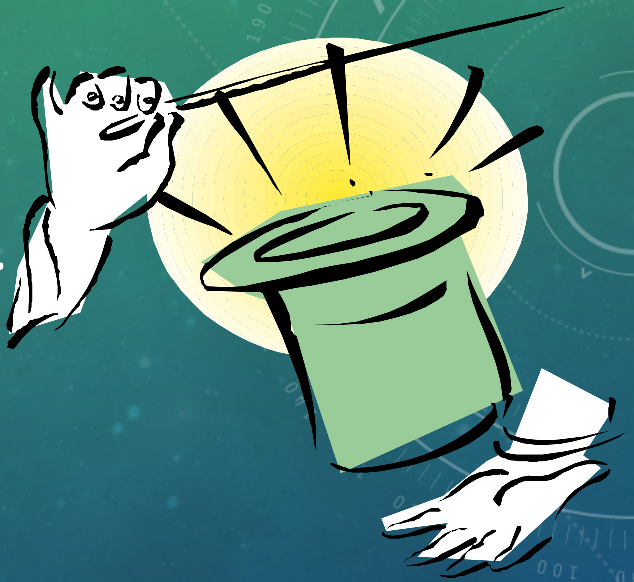
WEEK 4

PROBABILITY DISTRIBUTIONS:

DISCRETE DISTRIBUTIONS

RANDOM VARIABLE

- A variable which can take different values with given probabilities.



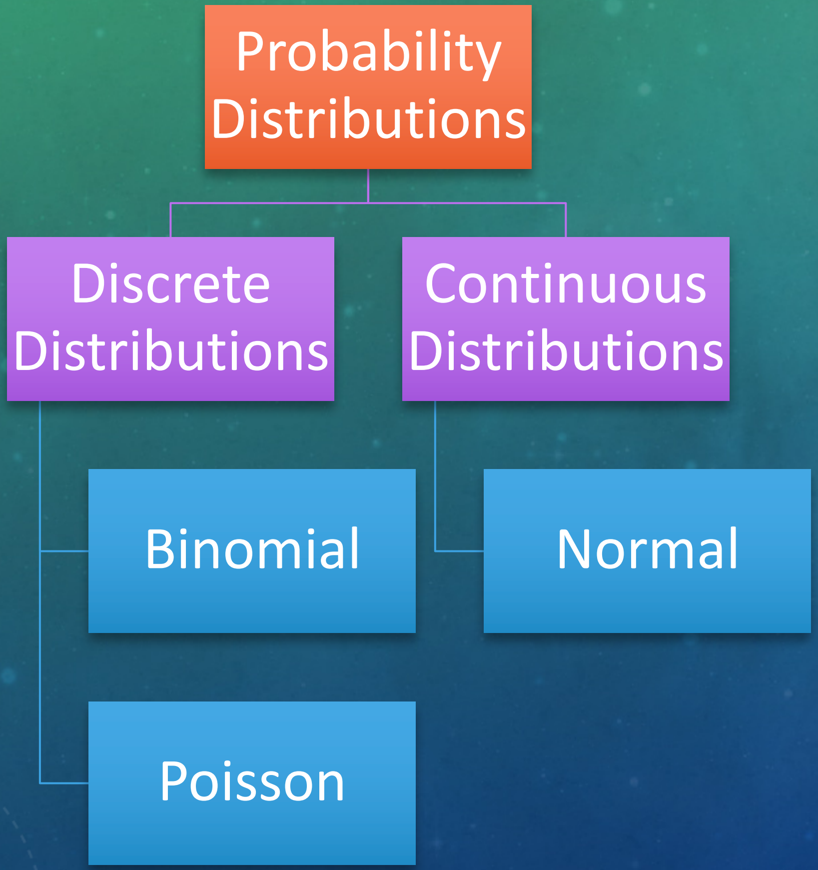
Random variable

- **Discrete** → it can take no more than countable number of values
  e.g. the number of eggs laid in a month, litter size, etc.
- **Continuous** → it can take any value in an interval
  e.g. the yearly milk yield for a farm
  or calf weight at the age of eight months

A discrete random variable with <u>only two possible values</u> is called a “**binary variable**” : e.g. diseased or healthy.

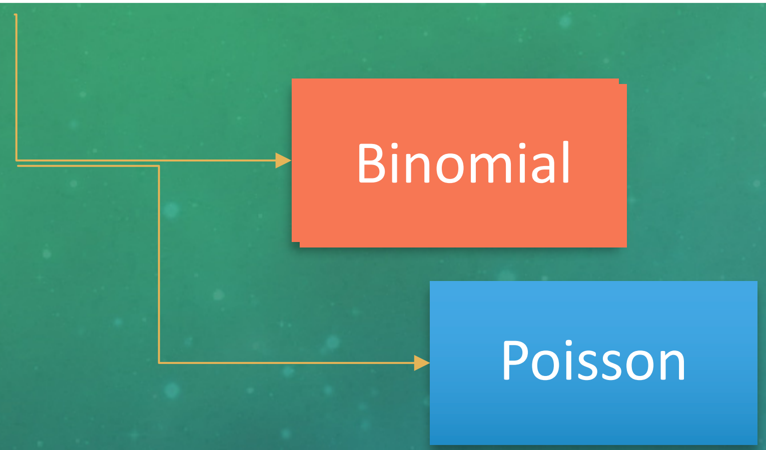
PROBABILITY DISTRIBUTION:

It comprises all the values that the "random variable" can take, with their associated probabilities.



The underlying foundation of most inferential statistical analysis is the concept of a probability distribution

DISCRETE PROBABILITY DISTRIBUTIONS



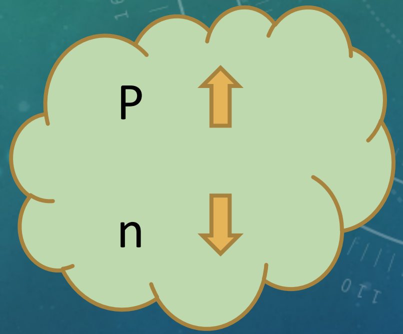
- For each possible value of a discrete random variable y we assign the probability P(y).
  - The probability distribution P(y) must satisfy the following two assumptions:

$$1)\ 0 \leq P(y) \leq 1$$

$$2)\ \sum_{(\text{all } y)} P(y) = 1$$

BINOMIAL DISTRIBUTION

- Relevant in the situation in which we are investigating a binary response. (E.g. animal shows a clinical sign of a particular disease or not)



- Success (event occurs): 1
- Failure (event does not occur): 0

} Binary response

Now assume that such single trial is repeated n times... → Binomial Distribution

➢ A binomial variable y is the number of successes in those n trials. It is the sum of n binary variables. The binomial probability distribution describes the distribution of different values of the variable y {0, 1, 2, ..., n} in a total of n trials.

CHARACTERISTICS OF A BINOMIAL DISTRIBUTION:

i. The experiment consists of n equivalent trials, independent of each other

ii. There are only two possible outcomes of a single trial, denoted with Y (yes) and N (no) or equivalently 1 and 0

iii. The probability of obtaining Y is the same from trial to trial, denoted with p. The probability of N is denoted with q, so $p + q = 1$

iv. The random variable y is the number of successes (Y) in the total of n trials.

BINOMIAL DISTRIBUTION

The probability distribution of a random variable y is determined by the parameter p and the number of trials n:

$$P(r) = \binom{n}{r} p^r q^{n-r} \Rightarrow P(r) = \left(\frac{n!}{(n-r)!\,r!}\right) p^r q^{n-r}$$

n= total number of events
r= number of success (event occurs)
n-r= number of failures (event does not occur)
p= probability of success in a single trial
q= 1-p = probability of failure in a single trial

The shape of the distribution depends on the parameter p. The binomial distribution is symmetric only when p = 0.5, and asymmetric in all other cases.

AN EXAMPLE

- Let's say that we took blood samples from six cattle randomly selected from the population. We know that the prevalence of bovine tuberculosis is 25%. What is the probability of <u>none</u> is positive for tuberculosis?

$$P(r) = \left(\frac{n!}{(n-r)!\,r!}\right) p^r q^{n-r}$$

n= total number of events
r= number of success (event occurs)
n-r= number of failures (event does not occur)
p= probability of success
q= probability of failure

SOLUTION:

$$P(0) = \left(\frac{6!}{(6-0)!\,0!}\right) 0.25^0 0.75^{6-0}$$

$$P(0) = \frac{729}{4096} = 0.1779 = 17{,}79\%$$

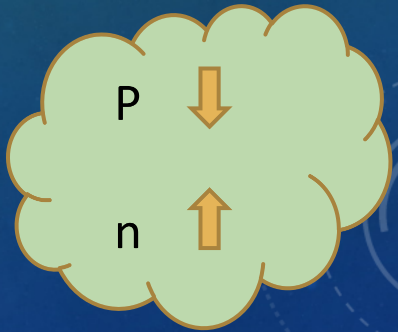
POISSON DISTRIBUTION

- The random variable of a Poisson distribution represents the *count* of the number of events occurring randomly and independently in time or space at a constant rate

- It is used to calculate the probability in health sciences, especially in the cases that are rarely seen.

e.g. the number of parasitic eggs per unit volume

$$P(r) = \frac{\bar{X}^r}{r!} e^{-\bar{X}}$$



$\bar{X}$ = average number of successes in a given time
r = number of success (event occurs)
e = base of the natural logarithm (e = 2.71828)

EXAMPLE

- Suppose that in a population of mice 3% have cancer. In a sample of 100 mice, what is the probability that more than one mouse has cancer?

P(y > 1) = 1 – P(y =0) – P(y = 1)

$\bar{X} = \lambda = 100\ (0.03) = 3$ (expectation, the mean is 3% of 100)

$$P(r) = \frac{\bar{X}^r}{r!} e^{-\bar{X}}$$

$$P(y = 0) = \frac{\bar{3}^0}{0!} 2.71828^{-\bar{3}} = 0.0498$$

$$P(y = 1) = \frac{\bar{3}^1}{1!} 2.71828^{-\bar{3}} = 0.149$$

$\bar{X}$ = average number of successes in a given time
r= number of success (event occurs)
e = base of the natural logarithm (e = 2.71828)

P(y > 1) = 1 – 0.0498 – 0.149 = 0.8012

The probability that in the sample of 100 mice more than one mouse has cancer is 0.8012