Statistical Techniques. Lab 1

Plan for today

- Review of Probability Theory
- Random variables
- Density. Distributions
- Expected value
- Some tasks (pen & paper)
- Short introduction to Statistics
- Python Implementations
- Some tasks (Python)
- Summary and time for questions

Objectives

- To recall Probability and solve some tasks
- To have introduction to Python in Statistics
- To learn some libraries for statistics in Python
- To solve some tasks using Python

Review of Probability Theory

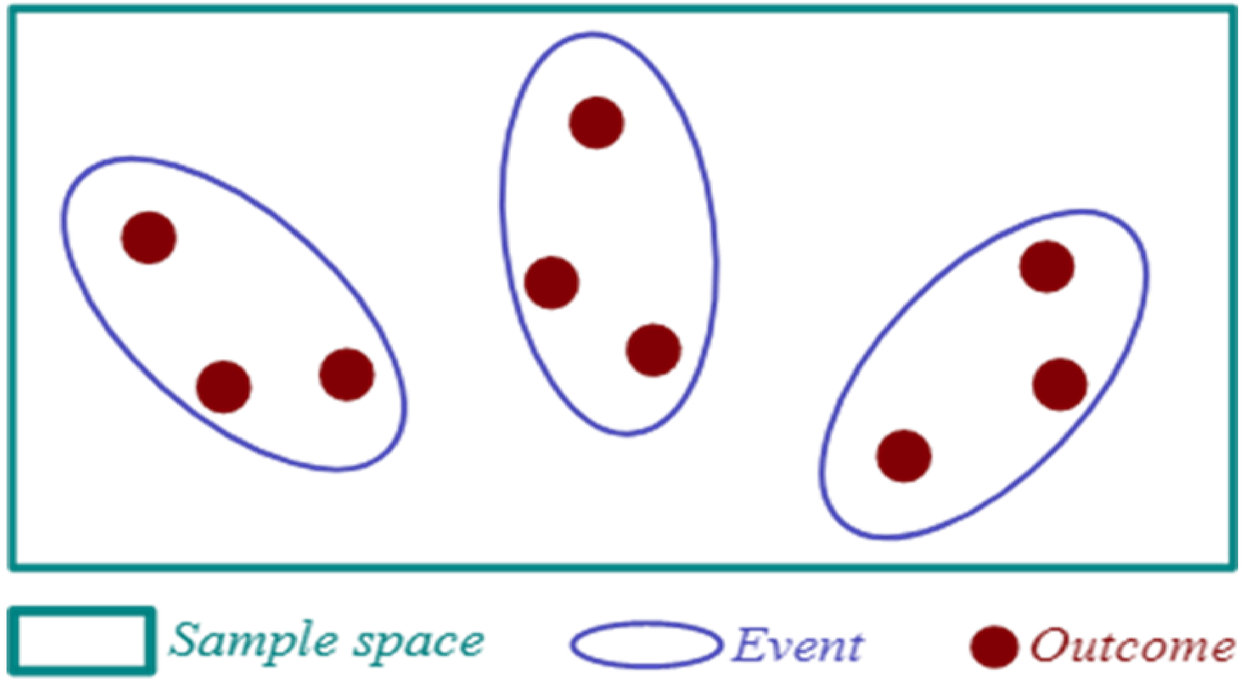
Be sure, that you know:

- Outcome
- Sample space
- Event

Don’t confuse these terms!

Sample space

Outcome: A result of a random experiment.
Sample Space: The set of all possible outcomes.
Event: A subset of the sample space.



Time to practice

- **Example 1.** Write down the sample space for each experiment below:
- • Tossing a coin: S =
- • Rolling a die: S =
- • Drawing a card from a deck: S =

- **Example 2**. Find the sample space for each experiment below:
- • Throw a coin twice: S =
- • Throw two dice: S =
- • Throw a coin repeatedly until a head first appears: S=

Simple probability

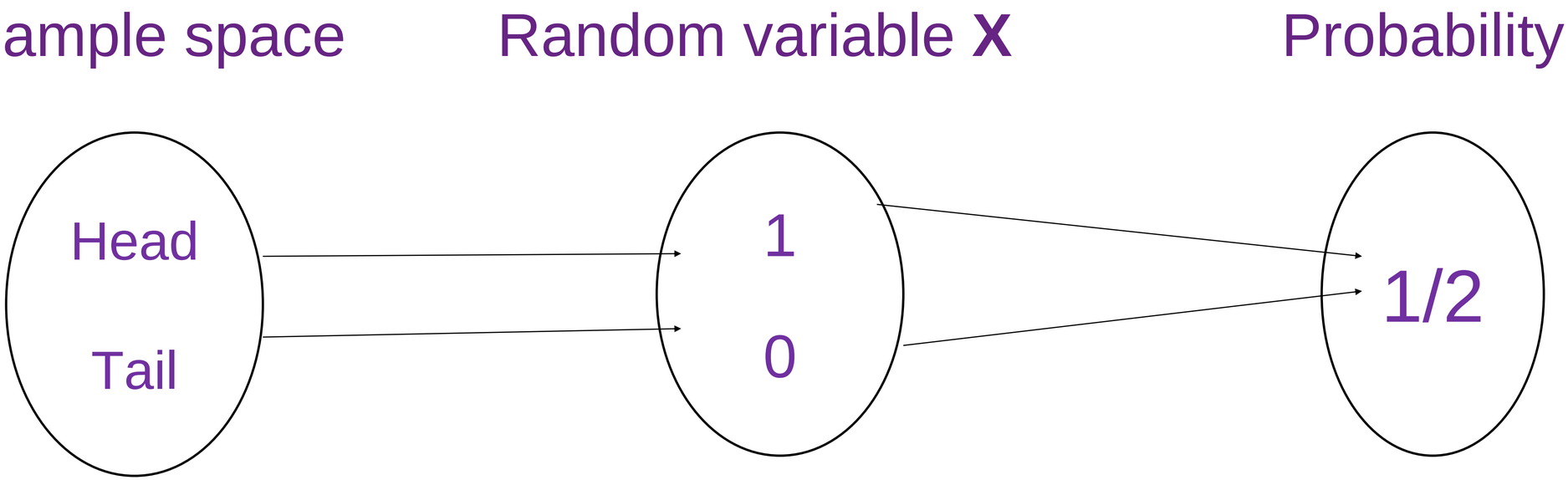
$$P(A) = \frac{n(A)}{n(S)} = \frac{no.of\ outcomes\ in\ A}{total\ no.of\ outcomes}$$

Example 3. On a six-sided die, each side has a number between 1 and 6. What is the probability of throwing a 3 or a 4?

a. 1 in 6
b. 1 in 3
c. 1 in 2
d. 1 in 4

Random variables

- X, is a numerical quantity whose value is determined be a random experiment



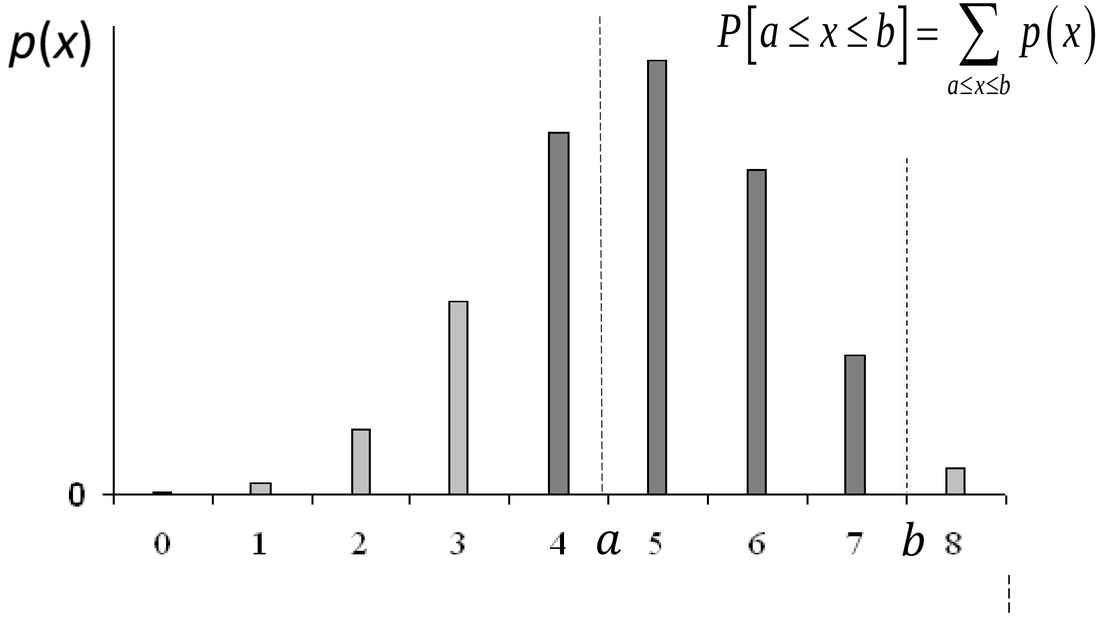
Two types of Random variables

- Continuous Random Variables (CDF (Cumulative Distribution Function), PDF (Probability Density Function))
- Discrete Random Variables (CDF (Cumulative Distribution Function), PMF (Probability mass function))

Probability mass function (pmf, $p(x)$)

- For a discrete random variable *X* the probability distribution is described by the **probability mass function *p*(*x*),** which has the following properties:

1. $0 \le p(x) \le 1$
2. $\sum_{x} p(x) = \sum_{i=1}^{\infty} p(x_i) = 1$
3. $P[a \le x \le b] = \sum_{a \le x \le b} p(x)$



Graph: Discrete Random Variable

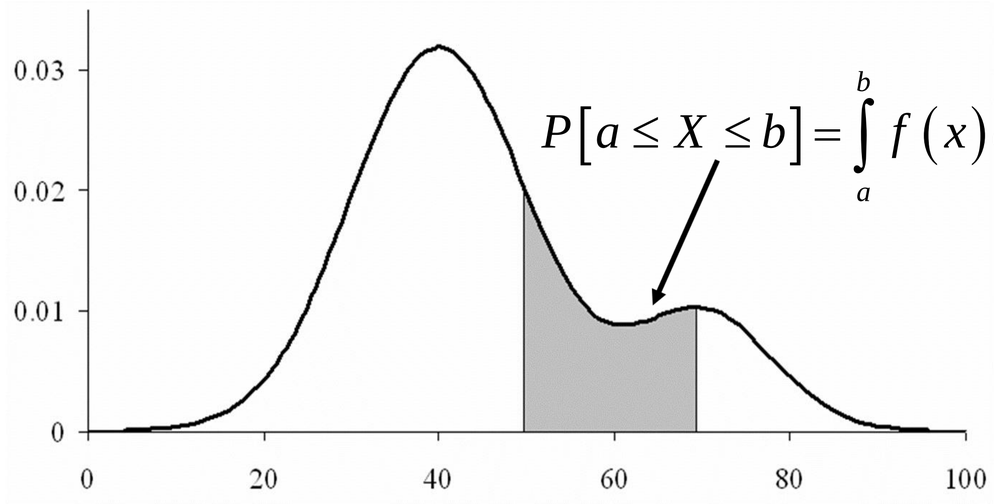
Probability Density Function (pdf, $f(x)$)

For a continuous random variable X the probability distribution is described by the **probability density function** $f(x)$, which has the following properties:

1. $f(x) \geq 0$

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

3. $P[a \leq X \leq b] = \int_{a}^{b} f(x)\,dx.$



The cumulative distribution function $F(x)$ (CDF)

- For Discrete Random Variables

$$F(x) = P[X \leq x] = \sum_{u \leq x} p(u)$$

- For Continuous Random Variables

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(u)\,du$$

Expected value

Let *X* denote a **discrete** random variable with probability mass function *p*(*x*) (probability density function *f*(*x*) if *X* is **continuous**) then the expected value of *X*, *E*(*X*) is defined to be:

$$E(X) = \sum_x xp(x) = \sum_i x_i p(x_i)$$

and if *X* is **continuous** with probability density function $f(x)$

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$

Time to practice

Example 4

The random variable X giving the number of passengers (excluding the driver) per car in rush hour traffic has the following probability function:

x_i	0	1	2	3	4
$P(X = x_i)$	0.7	p_2	0.1	0.05	0.05

1. Find p_2.
2. What is the probability that the number of passengers is at least 2?
3. Determine and sketch the distribution function $F(x)$ of the random variable X.
4. Find and interpret
   a. $F(3.9) - F(0.05)$
   b. Expected value $E(X)$?

Time to practice

Example 5.

Let the random variable X have the following distribution function:

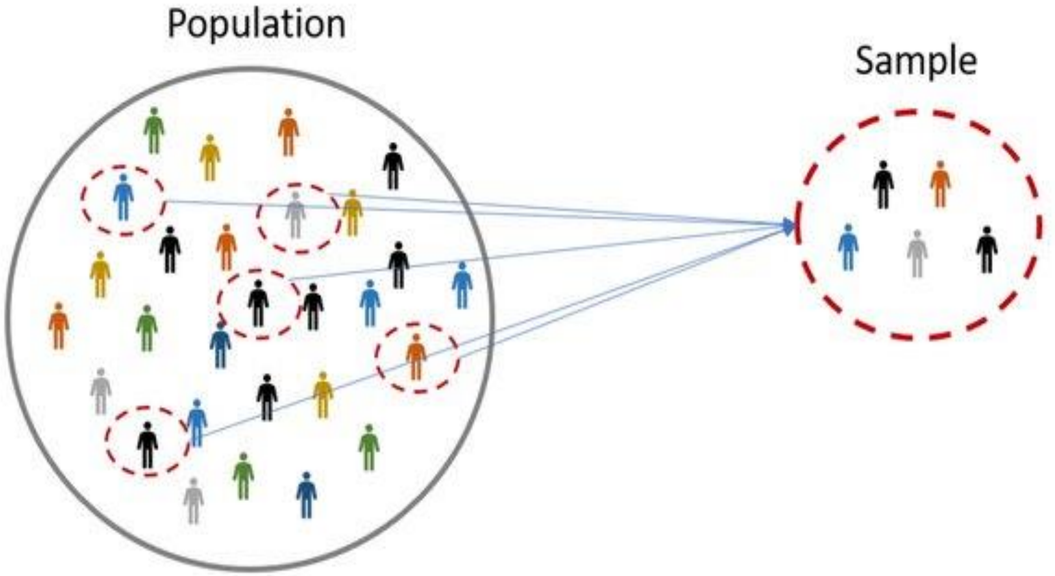
$$F(t) = \begin{cases} 0, & t < 0 \\ kt^2, & 0 \le t \le \frac{3}{4} \\ 1, & t > \frac{3}{4} \end{cases}$$

1. Find k, pdf?
2. What is the probability that that X lies in the interval $\left[0; \frac{1}{2}\right]$?
3. Find $E(t)$ (expected value)?

Introduction to Statistics

- A **population** is the entire collection of objects or outcomes about which information is sought.

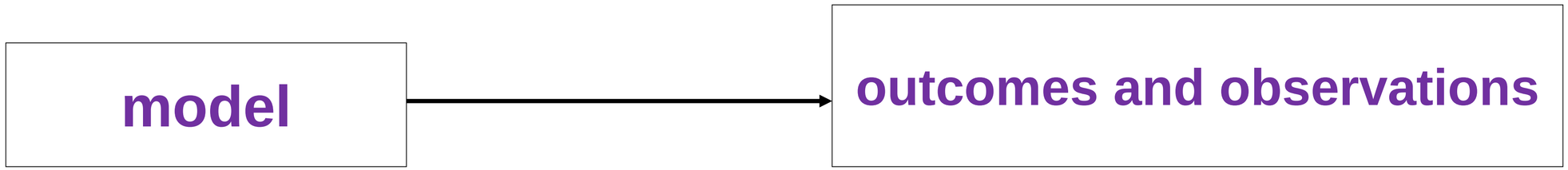
- A **sample** is a subset of a population, containing the objects or outcomes that are actually observed.



Introduction to Statistics

- In probability theory

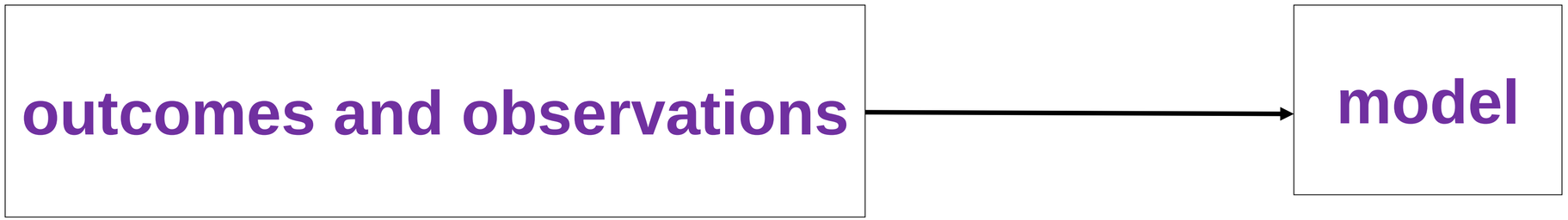
The model is known. We are interested in predicting the outcomes and observations of the phenomena



Introduction to Statistics

- In statistics

The model is unknown. The outcomes and observations of the phenomena have been observed.
We are interested in determining the model from the observations



Python Implementations

- Let’s have some practice.

Click here

Summary

Check yourself:

1.) What is the difference between pdf and cdf?

2.) What is a sample? Give your own example.

3.) What Python libraries do we use for Statistics?

Thank you