01 - Basics

- Probability gives a **numerical** description of how likely, or plausible, an event is to happen, or a proposition is true
    - it is a mathematical foundation for **statistics** and it is the doctrine of chances
    - *what are the chanches we get a 20?*
- to answer all questions we need a **model** that describes our dice and coins; in general the model is known but data are not
- Statistical Inference consists of methods to elaborate a collection of data to derive properties and parameters of an **underlying probability distribution** which isassumed to describe our data
    - *I flip a coint 10 times and get 10 heads, is the coin fair ?*
- The adopted language
    - a Proposition: must have an unambiguous meaning; must be of a simple logical type, i.e. TRUE or FALSE
    - A·B: logical product or conjunction --> both A and B are TRUE
    - A + B: logical sum or disjunction --> at least one of the propositions A, B are TRUE
    - A = B: the propostion on the left side has the same truth value as that on the right side
    - notA: denial of a propostion --> if A is true, A is false, and vice versa
- Definitions
    - Random experiment : an experiment with an outcome not completely predictable. When we repeat the experiment we may get a different result (e.g. coin tossing)
    - Outcome : the result of a single trial of the experiment
    - Sample space : the set of all possible outcomes of one single trial of the experiment Ω. The sample space containing everything we are considering in the analysis of the experiment is called Universe
    - Event : any set of possible outcomes of the experiment
        - independent events : two events are independent if the occurrence of the first does not affect the occurrence or non-occurrence of the second B For independent events, $P(A|B, I) = P(A|I)$, therefore $P(AB|I) = P(A|I)P(B|I)$. is not a property of the events themselves
        - if A and B are mutually exclusive events, (i.e. they have no outcome in common), $P(A + B|I) = P(A|I) + P(B|I)$. Mutually exclusive events contain no elements in common, and this is a property of the events
- Axiomatic Definition of Probability
- probabilities are real numbers between 0 and 1
- the higher the probability, the more likely it is to occur
- a probability equals to 1 means the event is certain to occur

- a probability of 0 means that the event cannot possibly occur

The following axioms are satisfied

1. $P(A|I) \geq 0$ for any event E
2. 2. $P(U|I) = 1$, is the probability of the universe (it means that some outcome occurs every time the experiment is performed)
3. $P(AB|I) = P(A|B, I) \cdot P(B|I) = P(B|A, I) \cdot P(A|I)$ --> (PRODUCT RULE)

Other properties, derived from the Axioms

1. $P(\emptyset) = 0$
2. $P(A|I) = 1 - P(A|I)$ --> (NORMALIZATION)
3. $P(A + B|I) = P(A|I) + P(B|I) - P(AB|I)$ --> (SUM RULE)
4. $P(A|I) = P(A, B|I) + P(A, B|I)$ --> (MARGINALIZATION)

- Bayes Theorem

  - 1 --> $P(B|A, I) = P(A|B, I)P(B|I) / (P(A|B, I)P(B|I) + P(A|B, I)P(B|I))$
  - let's imagine that we have a set of more than two events that partition the universe: $U = B_1 \cup B_2 \cup . . . \cup B_n$, and every distinct pair of the events are disjoint $B_j \cap B_i = \emptyset$, with j , i
  - an observable event A will be partitioned into parts A = unione($A \cap B_j$)
  - $P(B_j|A, I)=P(A|B_j, I)P(B_j|I)$ / SUM_j $P(A|B_j, I)P(B_j|I)$
  - the events A and B_j , (j = 1, . . . , n), are not treated symmetrically :
  - A is an observable event
  - B_j are considered not observable. We never know which one of them occurred
  - The marginal probabilities $P(B_j)$ are assumed known before we start and are called our prior probabilities
  - we often write Bayes' in its proportional form as $P(B_j|A, I) \propto P(A|B_j, I) \cdot P(B_j|I)$ --> (post, likel, prior)