MSc European Economy and Business Law

Statistics Pre-course 2022

Topic 2: Elements of Probability

Francesca Centofanti

francesca.centofanti@students.uniroma2.eu

Outline

- Introduction
- Set Theory
- Probability Axioms
- Fundamentals of Probability
- Combinatory Calculus
- Introduction to Random Variables

Introduction: Definitions of Probability

1. **Randomness**: suggests unpredictability. *Example*: when a coin is tossed, the outcome is uncertain (the outcome could be either an observed head (H) or an observed tail (T).
2. **Random experiment**: experiment in which the outcome on each trial is uncertain and distinct. *Examples*: rolling a die, selecting items at random from a manufacturing process to examine for defects, selection of numbers by a lottery machine, etc.)
3. **Sample space**: list of possible outcomes Ω. *Example*: when we toss a coin, we have two possible outcomes summarized by $\{H, T\}$.
4. **Event**: subset of the sample space. *Example*: a fair regular six-sided die is rolled, then $\Omega = \{1, 2, 3, 4, 5, 6\}$, but we may be concerned only with even numbers (events) $A = \{2, 4, 6\}$.

Now we can define the *classical* **probability** of an event A (if the outcomes in a sample space are equally likely to occur):

$$P(A) = \frac{Number\ \ of\ \ simple\ \ events\ \ in\ \ A}{Total\ \ number\ \ of\ \ \ simple\ \ events\ \ in\ \ the\ \ sample\ \ space}$$

Example

- When a child is born, the gender of the child is either a boy (B) or a girl (G), summarized by $\Omega = \{B, G\}$. If we consider a two-child family, the possibilities can be summarized by $\Omega = \{BB, BG, GB, GG\}$.
- If a two-child family is selected at random, what is the probability of there being two boys in the family?
- The event of two boys occurs once, and there are four simple events in the sample space, thus

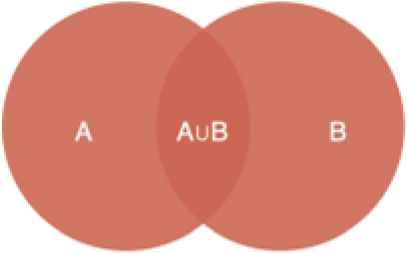
$$P(BB) = \frac{1}{4} = 0.25$$

Set Theory: Basic notions

Mathematically speaking *events* are **sets**

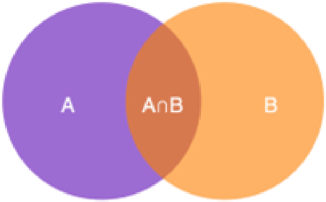
- **Definition of set**: finite or infinite collection of objects
  1. **Finite**: contains a finite number of objects
  2. **Infinite**: contains an infinite number of objects
- **Cardinality**: number of objects that belong to such set. Example: if $E = \{1, 2, 3\}$, the cardinality of E is written as $\#E = 3$.
- **Union of sets** ($A \cup B$): given two events A, B, everything that is in either A, B or in both
- **Intersection of sets** ($A \cap B$): given two events A, B, everything that is in both A and B
- **Complement sets** ($A^c or \ \overline{A}$): everything that is not in A

- **Union of sets** $(A \cup B)$: $A =$ "the die returns an even number", $B =$ "the die returns a 5" $\Rightarrow A \cup B = \{2, 4, 5, 6\}$

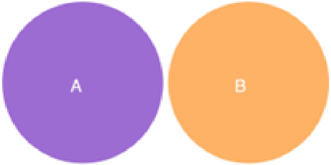


Examples

- **Intersection of sets** $(A \cap B)$: $A =$ "the die returns an even number", $B =$ "the die returns a number smaller than 5" $\Rightarrow A \cap B = \{2, 4\}$

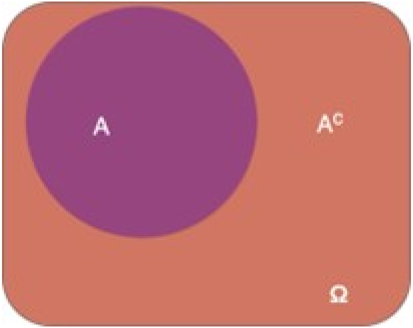


- If $A \cap B = \emptyset$, A and B are **disjoint**



Examples

- **Complement sets** (A^c *or* $\overline{A}$): $A =$ "the die returns an even number" $\Rightarrow A^c =$ "the die returns an odd number"



Probability Axioms

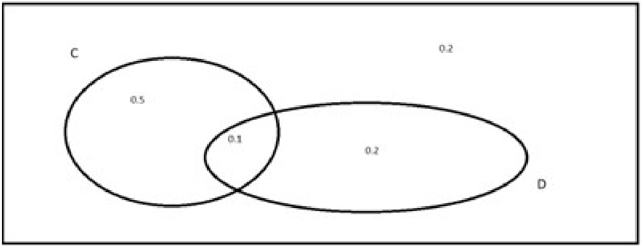
- The **probability** is a *set function* on which the following holds:
  1. $0 \leq P(A) \leq 1$
  2. $P(\Omega) = 1$
  3. $P(\emptyset) = 0$
- Consequently:
  1. $P(A^c) = 1 - P(A)$
  2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

REMEMBER! If A, B are *disjoint*, then $P(A \cup B) = P(A) + P(B)$

Example: Use Graphical Representation

In a sample of 100 college students, 60 said that they own a car, 30 said that they own a stereo, and 10 said that they own both a car and a stereo. Compute probabilities for these events and depict this information on a **Venn diagram**.

- Let C be the event that a student owns a car and let D be the event that a student owns a stereo. Thus:
  1. $P(C) = 0.6$
  2. $P(D) = 0.3$
  3. $P(C \cap D) = 0.1$



Fundamentals of Probability: Relative Frequencies

The *relative frequency probability* of an event's occurring is the proportion of times the event occurs over a given number of trials

- If A is the event in which we are interested, then the relative frequency probability of A's occurring is:

$$P(A) = \frac{frequency \;\; of \;\; occurence}{number \;\; of \;\; trials}$$

- *Example*: for the first 43 presidents of the United States, 26 were lawyers. What is the probability of randomly selecting from those 43 presidents a president who was a lawyer?

Let A represent the event of a president being a lawyer. Thus, since there are 43 presidents and 26 were lawyers, then:

$$P(A) = \frac{26}{42} = 0.605$$

Fundamentals of Probability: Conditional Probability

Probability is a measure of *uncertainty* on the result of a random experiment, so it is affected by any additional information on the outcome

- Let A and B be two events, if we know that B happened, we can *update* the probability of A as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example: let C be the event that a student owns a car, and let D be the event that a student owns a stereo. Thus $P(C) = 0.6$, $P(D) = 0.3$, and $P(C \cap D) = 0.1$. What is the probability of a student's having a stereo given the student has a car?

$$P(D|C) = \frac{P(D \cap C)}{P(C)} = \frac{0.1}{0.6} = 0.167$$

Fundamentals of Probability: Dependence/Independence

- If knowing an event B does not affect our probability evaluation of A, then we say that A and B are **independent**

$$P(A|B) = P(A)$$

- Combining this to the definition of **conditional probability**, we can derive the **factorization criterion** to assess if two events are independent:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$$

- If there is a relation between events (**dependence**), recalling the conditional probability formula, we have:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(A|B)P(B)$$

Combinatorics is a branch of mathematics which is about **counting**

- To compute probabilities we will often need a method for determining in how many ways a given phenomenon can happen
- **Fundamental principle of combinatorics**: given an experiment with n possible outcomes and and another experiment with m possible outcomes, then the *combination* of the two experiment has $m \times n$ possible outcomes

Example: tossing coin 1 has two possible outcomes and the same holds for coin 2. Thus, there are 2×2 possible outcomes.

Combinatory Calculus: Permutations

Given a set of n objects, a **permutation** is a given ordering of those objects. The number of possible permutations of n objects is equal to $n!$

- *Example*: There are $n = 9$ students attending the statistics precourse. Let us suppose that all the students have to seat on a chair and let us suppose that the chairs are disposed on a straight line. How many possible lines can be formed by changing the position of each student?

The first student can choose his position in , different ways, the second student has 8 possibilities, the third student has 7 possibilities and so on. Thus, there are

$$9 \times 8 \times 7 \times 6 \times ... \times 1 = 9!$$

ways to place 9 students in a line.

Combinatory Calculus: Dispositions

Given a set of n objects a **disposition** is a way of choosing k elements out of a set of n elements *without repetitions* and *taking* in account of their *order*. The number of possible of k disposition out of n objects is equal to

$$\frac{n!}{(n-k)!}$$

- *Example*: let us consider the example of the students enrolled in the statistics precourse. Let us now suppose that only $k = 6$ chairs are available for $n = 9$ students. How many ways can we dispose $n = 9$ students on $k = 6$ chairs?

The first student can choose a chair in 9 different ways, the second one can choose his chair in 8 different ways and so on. Thus there are

$$\frac{9 \times 8 \times 7 \times ... \times 3 \times 2 \times 1}{3 \times 2 \times 1} = 9 \times 8 \times 7 \times 6 \times 5 \times 4 = \frac{9!}{3!}$$

ways to dispose 9 students on six chairs.

Combinatory Calculus: Combinations

Given a set of n objects a **combination** is a way of choosing k elements out of a set of n elements *without repetitions* and *without taking* in account of their *order*. The number of possible k combinations out of n objects is equal to

$$\frac{n!}{(n-k)!k!}$$

- *Example*: Let's go back to the example of the 9 students enrolled in the statistics precourse. Let us suppose that we only have 6 chairs on which the students can sit. In how many different ways can we dispose $n = 9$ students on $k = 6$ chairs regardless of the order of the chairs?

We found that, if we keep into account the ordering, we can dispose the students in $\frac{n!}{(n-k)!}$ different ways. If we are not interested in their ordering, thus disposition $1, 2, 3, 4, 5, 6$ is equivalent to $6, 2, 3, 4, 5, 1$. So, we have to divide $\frac{n!}{(n-k)!}$ by the number of possible permutations of k objects which is equal to $k!$

Random Variables: Introduction

A *random variable* X is any **function** from the sample space to the real numbers

$$X : \Omega \rightarrow \mathbb{R}$$

- As the random variable is defined on the sample space, we can associate a *probability* to the values that the random variable assumes

Example: we introduced the sample space Ω and we know how to compute probabilities associated to subset of the sample space ($P =$ "Primary School", $S =$"Secondary School", $M =$"Master"). Now we can use the random variable (r.v.) X to randomly pick up a student and get back in output his/her years of schooling. Thus, there is a *probability law* associated to its values x (*probability distribution*)

Random Variables: Definitions

- X: *random variable* (random function)
- x: *realization* of the random variable (the number we get after we observe the result of the random experiment)
- χ: *support* of the random variable (all the possible values assumed by X)

Example: toss a three coins. X is the number of tails. $\chi = \{0, 1, 2, 3\}$
REMEMBER! A random variable is a *number*: we can do all sorts of operations with it!

Random Variables: Definitions

- **Discrete**: random variable which represents numbers found by counting (χ is countable). *Example*: number of marbles in a jar, number of students present, number of heads when tossing two coins, etc.
- **Continuous**: random variables which are found from measuring (X represents an infinite number of values on the number line). *Example*: height of a group of people, distance travelled while grocery shopping, student test scores, etc.
- **Distribution** (p_x): convenient way of summarizing single outcomes probabilities

Example: Ω represents the sets of all the possible outcomes after tossing two coins; X represents the number of tails after tossing 2 coins; $\chi = \{0, 1, 2, 3\}$

- **Discrete Random Variable**:
  1. *Probability mass distribution*: $p_x = P(X = x) = 0 \ \forall x \in \chi$
  2. *Cumulative distribution function*:
     $F_X(x) = P(X \le x) = \sum_{y \le x} P(X = y) = \sum_{y \le x} p_y$

Example: What is the probability of **exactly** 1 head?
$P(X = 1) = p_1 = 3/8$.
What is the probability of **at most** two heads?
$P(X \le 2) = F_X(2) = p_0 + p_1 + p_2 = 7/8$

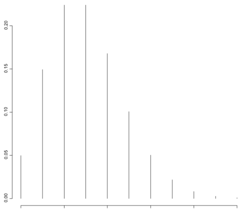
- **Continuous Random Variable**:
  1. *Cumulative distribution function*:
     $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)dt \ \forall x \in \chi$
  2. *Probability density distribution*: $f_X(x) = \frac{dF_X(x)}{dx} \ \forall x \in \chi$

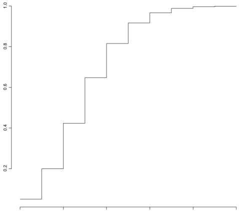
REMEMBER! If χ is not countable, it is not possible to put mass on any value $x \in \chi$ $(P(X = x) = 0 \ \forall x \in \chi)$

Discrete Random Variables: Properties

- **Probability mass distribution**:
  1. $p_x \geq 0$
  2. $p_x \leq 1$
  3. $\sum_x p_x = 1$

- **Cumulative distribution function**:
  1. $0 \leq F(x) \leq 1$
  2. F is non-decreasing
  3. F is right continuous





REMEMBER! A *discrete random variable* takes values in a **discrete set** (*p.m.f.* tells us exactly how much probability is assigned to each **point**). The *probability* that the random variable lies in a given interval is given by the **sum of masses** associated to the points composing the interval

Discrete Random Variables: Expected Value

The *expected value* for a random variable is the long-term mean or average value of the random variable

$$\mathbb{E}[X] = \sum_{x \in \chi} x \cdot p(x)$$

- The larger the number of observations, the closer the expected value will be to the average value of the observations generated by the random process

Example: you flip a fair coin. Every time you get a head, you lose 1\$. Every time you get a Tail you get 2\$. The random variable (gain) distribution is given by

$$P(x) = \begin{cases} 1/2 & if \quad x = -1 \\ 1/2 & if \quad x = 2 \end{cases}$$

and the expected value is given by

$$\mathbb{E}[X] = -1 \cdot 1/2 + 2 \cdot 1/2$$

Discrete Random Variables: Variance

The *variance* is the mean squared deviation of a random variable from its own mean

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

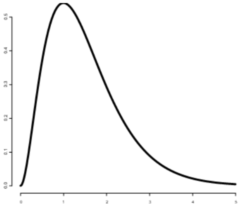
- The variance gives a measure of how spread out a random variable is

Example: X discrete r. v. with p. m. f. p_x has variance

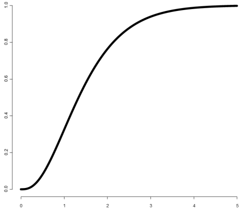
$$\mathbb{V}[X] = \sum_{x \in \chi} (x - \mathbb{E}[X])^2 p_x$$

Continuous Random Variables: Properties

- **Probability density function**:
  1. $f_X \geq 0$
  2. $f_X(x)$ needs not be ≤ 1
  3. $\int_{-\infty}^{\infty} f_X(x)dx = 1$



- **Cumulative distribution function**:
  1. $0 \leq F(x) \leq 1$
  2. F is non-decreasing
  3. F is right continuous



REMEMBER! A *continuous random variable* takes values over a **continuous range**. In this case the **mass** assigned to each point is given by the so called *density function*. To calculate the probability that the random variable lies in a given interval we have to calculate the **area** under

Continuous Random Variables: Expected Value

The *Mean* or *Expected Value* is the "average" of the elements in the support of X, weighted by the probability of each outcome

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

- **Properties** (valid also for *discrete* random variables):
  1. $\mathbb{E}[c] = c$ with c as a constant
  2. $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$
  3. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
  4. $\mathbb{E}[X - \mathbb{E}[X]] = 0$
  5. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Continuous Random Variables: Variance

The *variance* of a (continuous) random variable tells us of how much the variable oscillates around the mean

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- Example: X continuous r.v. with p.d.f. $f_X(x)$ has variance

$$\mathbb{V}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) dx$$

- **Properties** (valid also for *discrete* random variables)
  1. $\mathbb{V}[X]$ is always non-negative (it is 0 only when X is a constant)
  2. Standard deviation (square root of the variance) $sd(X) = \sqrt{\mathbb{V}[X]} \Rightarrow$ describes how far values of the random variable fall, on the average, from the expected value of the distribution
  3. $\mathbb{V}[aX + b] = a^2\mathbb{V}[X] \Rightarrow$ variance is insensitive to the location of the distribution but depends only on its scale

Additional Summary Statistics for Random Variables: Mode, Median, Covariance

- **Mode**: the value that is "more likely", i.e. the value that maximizes the density
- **Median**: the value that "splits in half" the distribution, i.e. m s.t. $P(X \leq m) = P(X > m) = 0.5$
- **Covariance**: measure of association between two random variables

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

REMEMBER! $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2Cov(X, Y)$

Random Variables: Independence

Two random variables X, Y are *independent* if

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leq x \cap Y \leq y) \\ &= P(X \leq x)P(Y \leq y) \\ &= F_X(x)F_Y(y) \qquad\qquad \forall x, y \in \mathbb{R} \end{aligned}$$

- If X and Y are **independent**, the value of one does not affect the value of the other:
  1. $p_{x_1,...,x_n} = p_{x_1} \times ... \times p_{x_n}$ for discrete random variables
  2. $f_{X_1,...,X_n}(x_1,...,x_n) = f_{X_1}(x_1) \times ... \times f_{X_n}(x_n)$
- Factorization Criterion: $F_{X,Y}(x,y) = F_X(x)F_Y(y)\ \forall x, y \in \mathbb{R}$

REMARK: if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$