STATISTICS

Module 2 *of*
MATHEMATICS AND STATISTICS

Marina Santacroce

marina.santacroce@unibocconi.it
Department of Decision Sciences, Bocconi University

L3. Introduction to probability. Probability spaces and axioms.

Outline

Example

Consider the *experiment* consisting in **rolling a *fair* 6 sided die**.

Let us focus on $E = \{\text{rolling an even number}\}$

- The probability of E $\mathbb{P}(E) = \frac{1}{2}$.

$$\mathbb{P}(E) = \frac{\textit{number of outcomes in } E}{\textit{number of possible outcomes}} = \frac{3}{6}$$

- *number of* "outcomes in E"$= \{\text{rolling an even number}\} = \{2, 4, 6\} \Longrightarrow 3$
- *number of* "possible outcomes"$=\{1, 2, 3, 4, 5, 6\} \Longrightarrow 6$

Classical definition of probability

Figure: Pierre Simon Laplace

Def. 1 The **probability of any event** E equals the proportion of the outcomes in E, *if* all the outcomes of the experiment are *equally likely* to occur.

$$\mathbb{P}(E) = \frac{number \;\; of \;\; outcomes \;\; in \;\; E}{number \;\; of \;\; all \;\; possible \;\; outcomes}$$

$\Longrightarrow$ This definition is *operative* but has some *drawbacks*:

- all the outcomes of the experiment must be *equally likely* to occur; "equally likely" translates into *fair* die in our example
- *tautology*;
- the number of all possible outcomes must be finite.

Imagine to "*roll a die*" n times and let n(E) be the number of times the event E (the outcome is even) occurred.

As n gets large, one expects the *relative frequency* of the event E,

$$\frac{n(E)}{n} \text{ close to } \frac{1}{2} \quad \to \mathbb{P}(E)$$

> "*In a sequence of identical trials, the relative frequency of an event gets close to its probability and the approximation improves as the number of trials increases.*"

This belief translates to the empirical level one of the central result of probability theory known as the

Law of large numbers

and it suggests another definition given in terms of experimental results.

Frequentist definition (R. Von Mises)

Def. 2 The *probability of an event E* is given by the *limiting relative frequency* of times that E occurs, as the number of trials tends to infinite,

$$\mathbb{P}(E) = \lim_{n\to\infty} \frac{n(E)}{n}.$$

$\Longrightarrow$ Like the classical, the frequentist definition is *operative* but it has the following *drawbacks*:

- one supposes the experiment can be repeatedly performed *under exactly the same conditions* an unlimited number of times;
- it is somehow "assumed" the sequence $\{a_n = \frac{n(E)}{n}, n \geq 1\}$ has a constant limit.

Subjective probability (de Finetti/Savage)

In the picture **Bruno de Finetti**

Def. 3 The *subjective* probability of an event E is a measure of the individual's degree of belief that the event E occurs.

To be applied one should think to the probability $\mathbb{P}(E)$ as the *stake on the event E*, i.e. how much one is willing to bet on E gaining

- 1 if *E occurs*
- 0 if *E does not occur*

The probabilities of the events should be assigned in such a way that there should not be a system of bets with a certain positive gain.

Axiomatic definition (A.N. Kolmogorov (1933))

Which are the common features of the three definitions given above?

The probability of an event $\mathbb{P}(E)$ is

- $0 \leq \mathbb{P}(E) \leq 1$
- if E surely occurs then $\mathbb{P}(E) = 1$
- additivity: if E and F are two disjoint events then $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$.

(in the example $E = \{\text{rolling an even number}\}$ and $F = \{\text{rolling } 3\}$)

The **modern** *probability theory* is based on the **axiomatic definition** due to **A.N. Kolmogorov** (1933).

Sample space and events

Def. The **sample space S** of an experiment (whose outcome is uncertain) is the *set of all possible outcomes* of the experiment.

Examples:

1. *flipping twice a coin* $\Rightarrow$ $S = \{(H,H),(H,T),(T,H),(T,T)\}$
2. *the lifetime (in years) of your pet* $\Rightarrow$ $S = \{x : 0 \leq x < \infty\}$
3. *the order of finish in a horse race with 12 horses* $\Rightarrow$ $S = \mathscr{P}(\{1,2,\ldots,12\})$

Def. A **subset** E of the sample space S is an **event**.[1].

If the outcome of the experiment is in E, then we say that E has occurred.

Examples:

1. $E = \{(H,H),(H,T)\} = \{$ head appears on the first flip$\}$
2. $E = \{3 < x < 5.5\}$ is the event that your pet will live more than 3 years but won't live more than 5 years and 6 months.
3. $E = \{$outcomes in S starting with 7$\}$ is the event that the race is won by horse 7.

[1] Events form a *σ-algebra* of S: $\mathscr{E}$. A family of subsets of S is a *σ-algebra* $\Leftrightarrow$ (i) $S \in \mathscr{E}$; (ii) $E \in \mathscr{E} \Rightarrow E^c \in \mathscr{E}$; (iii) If $\{E_n, n \geq 1\} \in \mathscr{E} \Rightarrow \cup_n E_n \in \mathscr{E}$.

Events

- $\emptyset$ is the *null event* and S is the *certain event*;

Given two events E and F:

- $E \cup F$ is said the *union of E and F* which is the set of all outcomes either in E or F or in both E and F $\Rightarrow E \cup F$ occurs if either E or F occurs.
- $E \cap F$ is said the *intersection of E and F* which is the set of all outcomes that are both in E and $F \Rightarrow$ it occurs if both E and F occur.
- E^c is said the *complement set of E* which is the set of all the outcomes that are not in $E \Rightarrow$ it occurs if E does not occur. ($S^c = \emptyset$).
- $E - F = E \cap F^c$ is the set of the outcomes which are in E but not in $F \Rightarrow$ it occurs if $E \cap F^c$ occurs.

Moreover, given a sequence of events of S, $\{E_n, n \geq 1\}$, one defines

- $\bigcup_{n:1}^{\infty} E_n$ is the event which consists of all the outcomes in E_ns $\Rightarrow$ it occurs if at least one of the E_ns occurs.
- $\bigcap_{n:1}^{\infty} E_n$ is the event which consists of the outcomes which are in all of the events $E_n \Rightarrow$ it occurs if all the E_ns occur.

Def. If E and F are such that $E \cap F = \emptyset$, they are said **mutually exclusive**.

Algebra:

Commutative laws

$$E \cup F = F \cup E \qquad \textit{and} \qquad E \cap F = F \cap E$$

Associative laws

$$E \cup (F \cup G) = (E \cup F) \cup G$$
$$E \cap (F \cap G) = (E \cap F) \cap G$$

Distributive laws

$$E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$$

$$E \cap (F \cup G) = (E \cap F) \cup (E \cap G).$$

De Morgan's Laws

$$\left(\bigcup_{i:1}^{n} E_i\right)^c = \bigcap_{i:1}^{n} E_i^c \quad \textit{and} \quad \left(\bigcap_{i:1}^{n} E_i\right)^c = \bigcup_{i:1}^{n} E_i^c$$

Axioms of probability

Def. Given a sample space S and a family of events $\mathscr{E}$[2], a probability measure $\mathbb{P}$ is a function of the events which satisfies the 3 axioms:

A.1 $0 \leq \mathbb{P}(E) \leq 1, \ E \in \mathscr{E}$

A.2 $\mathbb{P}(S) = 1$

A.3 If $\{E_n, n \geq 1\}$ is a sequence of mutually exclusive events ($E_n \in \mathscr{E}$ and $E_n \cap E_m = \emptyset$ when $n \neq m$), then

$$\mathbb{P}(\bigcup_{n:1}^{\infty} E_n) = \sum_{n:1}^{\infty} \mathbb{P}(E_n) \quad (\textit{countable additivity})$$

The triplet $(S, \mathscr{E}, \mathbb{P})$ is a *probability space* or also *probability model*.

[2] the couple $(S, \mathscr{E})$ is called a measurable space

Properties of Probability

Direct consequences of the axioms include

P.1 $\mathbb{P}(\emptyset) = 0$

- If $E_1 = S$, $E_n = \emptyset$ when $n > 1$. From A.3 we have $\mathbb{P}(\overset{\infty}{\underset{n:1}{\cup}} E_n) = \mathbb{P}(S) + \sum_{n:2}^{\infty} \mathbb{P}(\emptyset)$. By A.1 and A.2 $\Rightarrow \mathbb{P}(\emptyset) = 0$.

P.2 For mutually exclusive events $E_1, E_2, \ldots E_n$, then

$$\mathbb{P}(\overset{n}{\underset{i:1}{\cup}} E_i) = \sum_{i:1}^{n} \mathbb{P}(E_i) \ \ (\textit{additivity})$$

- By A.2 to the sequence $E_1, \ldots E_n$, $E_i = \emptyset$ when $i \geq n + 1$. Recalling $\mathbb{P}(\emptyset) = 0 \Rightarrow$ P.2

Simple properties

P.3 $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$

- Note that $S = E \cup E^c$ and $E \cap E^c = \emptyset$.

$$\Rightarrow \; 1 = \mathbb{P}(S) = \mathbb{P}(E \cup E^c) = \mathbb{P}(E) + \mathbb{P}(E^c)$$

$\uparrow A.2$ $\qquad \uparrow P.2$ (*additivity*)

P.4 If $E \subset F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$

- Since $F = E \cup (F - E)$ and $E \cap (F - E) = \emptyset$.

$$\Rightarrow \; \mathbb{P}(F) = \mathbb{P}(E \cup (F - E)) = \mathbb{P}(E) + \mathbb{P}(F - E) \geq \mathbb{P}(E)$$

$\uparrow P.2$ (*additivity*) $\uparrow$ $A.1$ $(\mathbb{P}(F - E) \geq 0)$

P.5 $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$

- Since $F = (F \cap E) \cup (F \cap E^c)$ and $E \cup F = E \cup (F \cap E^c)$

(we write F and $E \cup F$ as union of mutually exclusive events).

$$\Rightarrow \mathbb{P}(E \cup F) = \mathbb{P}(E \cup (F \cap E^c)) = \mathbb{P}(E) + \mathbb{P}(F \cap E^c) \;\; (P.2 \text{ additivity})$$
$$\Rightarrow \mathbb{P}(F) = \mathbb{P}(F \cap E) + \mathbb{P}(F \cap E^c) \;\; (P.2 \text{ additivity})$$

Substituting we find P.5

Inclusion-Exclusion Identity

Generalizing Property P.5 to 3 events

$$\mathbb{P}(E \cup F \cup G) = \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}(E \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(E \cap F \cap G)$$

▷ Since $E \cup F \cup G = E \cup (F \cup G)$ *(associative law)*,
by P.5 we have

$$\mathbb{P}(E \cup F \cup G) = \mathbb{P}(E \cup (F \cup G)) = \mathbb{P}(E) + \mathbb{P}(F \cup G) - \mathbb{P}(E \cap (F \cup G))$$

by $E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$ *(distributive law)*

$$= \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(F \cap G) - \mathbb{P}((E \cap F) \cup (E \cap G))$$

Using P.5 we have $(*)$ and substituting the result follows.

$$(*) \quad \mathbb{P}((E \cap F) \cup (E \cap G)) = \mathbb{P}((E \cap F) + \mathbb{P}(E \cap G) - \mathbb{P}(E \cap F \cap G)$$

Inclusion-Exclusion Identity

Generalizing P.5 to an arbitrary finite number n of events we get the ***Inclusion-Exclusion Identity***

$$\mathbb{P}(\bigcup_{i:1}^{n} E_i) = \sum_{i:1}^{n} \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j) + \sum_{i<j<k} \mathbb{P}(E_i \cap E_j \cap E_k) - \cdots + (-1)^{n-1} \mathbb{P}(\bigcap_{i:1}^{n} E_i)$$

(*It can be proved by induction, i.e. it is known it holds true when $n = 2$, it is then supposed to be true for n events and shown to hold for $n + 1$*).

Example:

From a study the 10% of the students of LSE play an instrument in the leisure time, the 20% play some kind of sport, the 5% study a foreign language. Moreover, the 5% play an instrument and practice some sport, the 3% play an instrument and a foreign language and the 2% study a language and practice some sport; the 1% does the three things. Choosing a student at random, compute the probability that

1. he/she does at least one of these activities;
2. he/she does not do any of these activities;
3. he/she studies a language or plays an instrument but he/she does not practise any sport.