Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information

# 1

# Introductory remarks

This chapter contains a short outline of the history of probability and a brief account of the debate on the meaning of probability. The two issues are interwoven. Note that the material covered here already informally uses concepts belonging to the common cultural background and that will be further discussed below. After reading and studying this chapter you should be able to:

- gain a first idea on some basic aspects of the history of probability;
- understand the main interpretations of probability (classical, frequentist, subjectivist);
- compute probabilities of events based on the fundamental counting principle and combinatorial formulae;
- relate the interpretations to the history of human thought (especially if you already know something about philosophy);
- discuss some of the early applications of probability to economics.

## 1.1 Early accounts and the birth of mathematical probability

We know that, in the seventeenth century, probability theory began with the analysis of games of chance (a.k.a. gambling). However, dice were already in use in ancient civilizations. Just to limit ourselves to the Mediterranean area, due to the somewhat Eurocentric culture of these authors, dice are found in archaeological sites in Egypt. According to Svetonius, a Roman historian, in the first century, Emperor Claudius wrote a book on gambling, but unfortunately nothing of it remains today.

It is 'however' true that chance has been a part of the life of our ancestors. Always, and this is true also today, individuals and societies have been faced with unpredictable events and it is not surprising that this unpredictability has been the subject of many discussions and speculations, especially when compared with better predictable events such as the astronomical ones.

It is perhaps harder to understand why there had been no mathematical formalizations of probability theory until the seventeenth century. There is a remote



© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



possibility that, in ancient times, other treatises like the one of Claudius were available and also were lost, but, if this were the case, mathematical probability was perhaps a fringe subject.

A natural place for the development of probabilistic thought could have been physics. Real measurements are never exactly in agreement with theory, and the measured values often fluctuate. Outcomes of physical experiments are a natural candidate for the development of a theory of random variables. However, the early developments of physics did not take chance into account and the tradition of physics remains far from probability. Even today, formal education in physics virtually neglects probability theory.

Ian Hacking has discussed this problem in his book *The Emergence of Probability*. In his first chapter, he writes:

> A philosophical history must not only record what happened around 1660,[1] but must also speculate on how such a fundamental concept as probability could emerge so suddenly.
>
> Probability has two aspects. It is connected with the degree of belief warranted by evidence, and it is connected with the tendency, displayed by some chance devices, to produce stable relative frequencies. Neither of these aspects was self-consciously and deliberately apprehended by any substantial body of thinkers before the times of Pascal.

The official birth of probability theory was triggered by a question asked of Blaise Pascal by his friend Chevalier de Méré on a dice gambling problem. In the seventeenth century in Europe, there were people rich enough to travel the continent and waste their money gambling. de Méré was one of them. In the summer of 1654, Pascal wrote to Pierre de Fermat in order to solve de Méré's problem, and out of their correspondence mathematical probability theory was born.

Soon after this correspondence, the young Christian Huygens wrote a first exhaustive book on mathematical probability, based on the ideas of Pascal and Fermat, *De ratiociniis in ludo aleae* [1], published in 1657 and then re-published with comments in the book by Jakob Bernoulli, *Ars conjectandi*, which appeared in 1713, seven years after the death of Bernoulli [2]. It is not well known that Huygens' treatise is based on the concept of expectation for a random situation rather than on probability; expectation was defined as the fair price for a game of chance reproducing the same random situation. The economic flavour of such a point of view anticipates the writings by Ramsey and de Finetti in the twentieth century. However, for Huygens, the underlying probability is based on *equally possible cases*, and the fair price of the equivalent game is just a trick for calculations. Jakob Bernoulli's contribution to probability theory is important, not only for the celebrated theorem of *Pars IV*, relating the probability of an event to the

[1] This is when the correspondence between Pascal and Fermat took place.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



relative frequency of successes observed iterating independent trials many times. Indeed, Bernoulli observes that if we are unable to set up the set of all *equally possible cases* (which is impossible in the social sciences, as for causes of mortality, number of diseases, etc.), we can estimate unknown probabilities by means of observed frequencies. The 'rough' inversion of Bernoulli's theorem, meaning that the probability of an event is identified with its relative frequency in a long series of trials, is at the basis of the ordinary frequency interpretation of probability. On the other side, for what concerns foundational problems, Bernoulli is important also for the *principle of non-sufficient reason* also known as *principle of indifference*. According to this principle, if there are $g > 1$ mutually exclusive and exhaustive possibilities, indistinguishable except for their labels or names, to each possibility one has to attribute a probability $1/g$. This principle is at the basis of the logical definition of probability.

A probabilistic inversion of Bernoulli's theorem was offered by Thomas Bayes in 1763, reprinted in Biometrika [3]. Bayes became very popular in the twentieth century due to the *neo-Bayesian* movement in statistics.

It was, indeed, in the eighteenth century, namely in 1733, that Daniel Bernoulli published the first paper where probability theory was applied to economics, *Specimen theoriae novae de mensura sortis*, translated into English in Econometrica [4].

## 1.2 Laplace and the classical definition of probability

In 1812, Pierre Simon de Laplace published his celebrated book *Théorie analytique des probabilités* containing developments and advancements, and representing a summary of what had been achieved up to the beginning of the nineteenth century (see further reading in Chapter 3). Laplace's book eventually contains a clear definition of probability. One can argue that Laplace's definition was the one currently used in the eighteenth century, and for almost all the nineteenth century. It is now called the *classical definition*.

In order to illustrate the classical definition of probability, consider a *dichotomous variable*, a variable only assuming two values in an experiment. This is the case when tossing a coin. Now, if you toss the coin you have two possible outcomes: $H$ (for head) and $T$ (for tails). The probability $\mathbb{P}(H)$ of getting $H$ is given by the number of favourable outcomes, 1 here, divided by the total number of possible outcomes, all considered *equipossible*, 2 here, so that:

$$\mathbb{P}(H) = \frac{\text{\# of favourable outcomes}}{\text{\# of possible outcomes}} = \frac{1}{2}, \tag{1.1}$$

where # means *the number*. The classical definition of probability is still a good guideline for the solution of probability problems and for getting correct results

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



in many cases. The task of finding the probability of an event is reduced to a combinatorial problem. One must enumerate and count all the favourable cases as well as all the possible cases and assume that the latter cases have the same probability, based on the principle of indifference.

### 1.2.1 The classical definition at work

In order to use the classical definition, one should be able to list favourable outcomes as well as the total number of possible outcomes of an experiment. In general, some calculations are necessary to apply this definition. Suppose you want to know the probability of getting exactly two heads in three tosses of a coin. There are 8 possible cases: $(TTT, TTH, THT, HTT, HHT, HTH, THH, HHH)$ of which 3 exactly contain 2 heads. Then, based on the classical definition, the required probability is $3/8$. If you consider 10 tosses of a coin, there are already 1024 possible cases and listing them all has already become boring. The *fundamental counting principle* comes to the rescue.

According to the commonsensical principle, for a finite sequence of decisions, their total number is the product of the number of choices for each decision. The next examples show how this principle works in practice.

**Example** In the case discussed above there are 3 decisions in a sequence (choosing $H$ or $T$ three times) and there are 2 choices for every decision ($H$ or $T$). Thus, the total number of decisions is $2^3 = 8$.

Based on the fundamental counting principle, one gets the number of dispositions, permutations, combinations and combinations without repetition for $n$ objects.

**Example** (Dispositions with repetition) Suppose you want to choose an object $k$ times out of $n$ objects. The total number of possible choices is $n$ each time and, based on the fundamental counting principle, one finds that there are $n^k$ possible choices.

**Example** (Permutations) Now you want to pick an object out of $n$, remove it from the list of objects and go on until all the objects are selected. For the first decision you have $n$ choices, for the second decision $n-1$ and so on until the $n$th decision where you just have 1 to take. As a consequence of the fundamental counting principle, the total number of possible decisions is $n!$.

**Example** (Sequences without repetition) This time you are interested in selecting $k$ objects out of $n$ with $k \leq n$, and you are also interested in the order of the selected items. The first time you have $n$ choices, the second time $n-1$ and so on, until the

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



$k$th time where you have $n-k+1$ choices left. Then, the total number of possible decisions is $n(n-1)\cdots(n-k+1) = n!/(n-k)!$.

**Example** (Combinations) You have a list of $n$ objects and you want to select $k$ objects out of them with $k \leq n$, but you do not care about their order. There are $k!$ ordered lists containing the same elements. The number of ordered lists is $n!/(n-k)!$ (see above). Therefore, this time, the total number of possible decisions (possible ways of selecting $k$ objects out of $n$ irrespective of their order) is $n!/(k!(n-k)!)$. This is a very useful formula and there is a special symbol for the so-called *binomial coefficient*:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{1.2}$$

Indeed, these coefficients appear in the expansion of the $n$th power of a binomial:

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}, \tag{1.3}$$

where

$$\binom{n}{0} = \binom{n}{n} = 1, \tag{1.4}$$

as a consequence of the definition $0! = 1$.

**Example** (Combinations with repetition) Suppose you are interested in finding the number of ways of allocating $n$ objects into $g$ boxes, irrespective of the names of the objects. Let the objects be represented by crosses, $\times$, and the boxes by vertical bars. For instance, the string of symbols $|\times\times|\times||$ denotes two objects in the first box, one object in the second box and no object in the third one. Now, the total number of symbols is $n+g+1$ of which 2 are always fixed, as the first and the last symbols must be a $|$. Of the remaining $n+g+1-2 = n+g-1$ symbols, $n$ can be arbitrarily chosen to be crosses. The number of possible choices is then given by the binomial factor $\binom{n+g-1}{n}$.

**Example** (Tossing coins revisited) Let us consider once again the problem presented at the beginning of this subsection. This was: what is the probability of finding exactly two heads out of three tosses of a coin? Now, the problem can be generalized: what is the probability of finding exactly $n$ heads out of $N$ tosses of a coin ($n \leq N$)? The total number of possible outcomes is $2^N$ as there are 2 choices the

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



first time, two the second and so on until two choices for the $N$th toss. The number of favourable outcomes is given by the number of ways of selecting $n$ places out of $N$ and putting a head there and a tail elsewhere. Therefore

$$\mathbb{P}(\text{exactly } n \text{ heads}) = \binom{N}{n} \frac{1}{2^N}. \tag{1.5}$$

### *1.2.2 Circularity of the classical definition*

Even if very useful for practical purposes, the classical definition suffers from circularity. In order to justify this statement, let us re-write the classical definition: *the probability of an event is given by the number of favourable outcomes divided by the total number of possible outcomes considered equipossible*. Now consider a particular outcome. In this case, there is only 1 favourable case, and if $r$ denotes the total number of outcomes, one has

$$\mathbb{P}(\text{outcome}) = \frac{1}{r}. \tag{1.6}$$

This equation is the same for any outcome and this means that all the outcomes have the same probability. Therefore, in the classical definition, there seems to be no way of considering *elementary* outcomes with different probabilities and the equiprobability of all the outcomes is a consequence of the definition. Essentially, equipossibility and equiprobability do coincide. A difficulty with equiprobability arises in cases in which the outcomes have different probabilities. What about an unbalanced coin or an unfair die?

If the hidden assumption *all possible outcomes being equiprobable* is made explicit, then one immediately sees the circularity as probability is used to define itself, in other words *the probability of an event is given by the number of favourable outcomes divided by the total number of possible outcomes assumed equiprobable*. In summary, if the equiprobability of outcomes is not mentioned, it becomes an immediate consequence of the definition and it becomes impossible to deal with non-equiprobable outcomes. If, on the contrary, the equiprobability is included in the definition as an assumption, then the definition becomes circular. In other words, the equiprobability is nothing else than *a hypothesis* which holds in all the cases where it holds.

A possible way out from circularity was suggested by J. Bernoulli and adopted by Laplace himself; it is the so-called indifference principle mentioned above. According to this principle, if one has no reason to assign different probabilities to a set of exhaustive and mutually exclusive events (called outcomes or possibilities so far), then these events must be considered as equiprobable. For instance, in the case of

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



the coin, in the absence of further indication, one has the following set of equations

$$\mathbb{P}(H) = \mathbb{P}(T), \tag{1.7}$$

and

$$\mathbb{P}(H) + \mathbb{P}(T) = 1, \tag{1.8}$$

yielding $\mathbb{P}(H) = \mathbb{P}(T) = 1/2$, where the outcomes $H$ and $T$ are exhaustive (one of them must occur) and mutually exclusive (if one obtains $H$, one cannot get $T$ at the same time).

The principle of indifference may seem a beautiful solution, but it leads to several problems and paradoxes identified by J. M. Keynes, by J. von Kries in *Die Prinzipien der Wahrscheinlichkeitsrechnung* published in 1886 [5] and by Bertrand in his *Calcul des probabilités* of 1907 [6]. Every economist knows the *General Theory* [7], but few are aware of *A Treatise on Probability*, a book published by Keynes in 1921 and including one of the first attempts to present a set of axioms for probability theory [8]. Let us now consider some of the paradoxes connected with the principle of indifference. Suppose that one does not know anything about a book. Therefore the probability of the statement *this book has a red cover* is the same as the probability of its negation *this book does not have a red cover*. Again here one has a set of exhaustive and mutually exclusive events, whose probability is 1/2 according to the principle of indifference. However, as nothing is known about the book, the same considerations can be repeated for the statements *this book has a green cover*, *this book has a blue cover*, etc. Thus, each of these events turns out to have probability 1/2, a paradoxical result. This paradox can be avoided if one further knows that the set of possible cover colours is finite and made up of, say, $r$ elements. Then the probability of *this book has a red cover* becomes $1/r$ and the probability of *this book does not have a red cover* becomes $1 - 1/r$. Bertrand's paradoxes are subtler and they make use of the properties of real numbers. Already with integers, if the set of events is countable, the indifference principle leads to a distribution where every event has zero probability as $\lim_{r\to\infty} 1/r = 0$ but where the sum of these zero probabilities is 1, a puzzling result which can be dealt with using measure theory. The situation becomes worse if the set of events is infinite and non-countable. Following Bertrand, let us consider a circle and an equilateral triangle inscribed in the circle. What is the probability that a randomly selected chord is longer than the triangle side? Three possible answers are:

1. One of the extreme points of the chord can indifferently lie in any point of the circle. Let us then assume that it coincides with a vertex of the triangle, say vertex $A$. Now the chord direction can be selected by chance, that is we assume that the angle $\theta$ of the chord with the tangent to the circle in $A$ is uniformly distributed. Now the chord is longer than the side of the triangle only if its $\pi/3 < \theta < 2\pi/3$.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



The triangle defines three circle arcs of equal length, this means that the required probability is 1/3.

2. Fixing a direction, and considering all chords parallel to that direction, all chords whose distance from the centre is $< r/2$ are longer than the triangle side; then considering the distance from the centre of a chord parallel to the fixed direction uniformly distributed gives a probability equal to 1/2.
3. Random selection of a chord is equivalent to random selection of its central point. In order for the chord to be longer than the triangle side, the distance of its central point from the centre of the circle must be smaller than one-half of the circle radius. Then the area to which this point must belong is 1/4 of the circle area and the corresponding probability turns out to be 1/4 instead of 1/3.

An important vindication for the indifference principle has been given by E.T. Jaynes, who not only reverted Bertrand's so-called ill-posed problem into *The Well Posed Problem* [9], but also extended it to a more general Maximum Entropy Principle, one of major contemporary attempts to save Laplace's tradition against the frequentist point of view described below.

The history of thought has seen many ways for avoiding the substantial difficulties of the classical definition or for circumventing its circularity. One of these attempts is the frequentist approach, where probabilities are roughly identified with observed frequencies of outcomes in repeated independent experiments. Another solution is the subjectivist approach, particularly interesting for economists as probabilities are defined in terms of *rational bets*.

## 1.3 Frequentism

The principle of indifference introduces a logic or subjective element in the evaluation of probabilities. If, in the absence of any reason, one can assume equiprobable events, then if there are specific reasons one can make another assumption. Then probability assignments depend on one's state of knowledge of the investigated system. Empiricists opposed similar views and tried to focus on the outcomes of real experiments and to define probabilities in terms of frequencies. Roughly speaking, this viewpoint can be explained as follows, using the example of coin tossing. According to frequentists the probability of $H$ can be approximated by repeatedly tossing a coin, by recording the sequence of outcomes $HHTHTTHHTTTH\cdots$, counting the number of $H$s and dividing for the total number of trials

$$\mathbb{P}(H) \sim \frac{\#\text{ of } H}{\#\text{ of trials}}. \tag{1.9}$$

The ratio on the right-hand side of the equation is the *empirical relative frequency* of the outcome $H$, a useful quantity in descriptive statistics. Now, this ratio is hardly

© in this web service Cambridge University Press www.cambridge.org

Cambridge

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



exactly equal to 1/2 and the frequentist idea is to extrapolate the sequence of trials to infinity and to define the probability as

$$\mathbb{P}(H) = \lim_{\#\text{ of trials}\to\infty} \frac{\#\text{ of } H}{\#\text{ of trials}}. \tag{1.10}$$

This is the preferred definition of probability in several textbooks introducing probability and statistics to natural scientists and in particular to physicists. Probability becomes a sort of measurable quantity that does not depend on one's state of knowledge, it becomes *objective* or, at least, based on empirical evidence. Kolmogorov, the founder of axiomatic probability theory, was himself a supporter of frequentism and his works on probability theory have been very influential in the twentieth century.

The naïve version of frequentism presented above cannot be a solution to the problems discussed before. Indeed, the limit appearing in the definition of probability is not the usual limit defined in calculus for the convergence of a sequence. There is no analytic formula for the number of heads out of $N$ trials and nobody can toss a coin for an infinite number of times. Having said that, one can notice that similar difficulties are present when one wants to define real numbers as limits of Cauchy sequences of rational numbers. Following this idea, a solution to the objection presented above has been proposed by Richard von Mises; starting from 1919, he tried to develop a rigorous frequentist theory of probability based on the notion of *collective*. A collective is an infinite sequence of outcomes where each attribute (e.g. $H$ in the case of coin-tossing dichotomous variables) has a limiting relative frequency not depending on place selection. In other words, a collective is an infinite sequence whose frequencies have not only a precise limit in the sense of (1.10), but, in addition, the same limit must hold for *any* subsequence chosen in advance. The rationale for this second and most important condition is as follows: whatever strategy one chooses, or game system one uses, the probability of an outcome is the same. von Mises' first memoir was *Grundlagen der Wahrscheinlichkeitsrechnung*, which appeared in the fifth volume of *Mathematische Zeitschrift* [10]. Subsequently, he published a book, *Wahrscheinlichkeit, Statistik und Wahrheit. Einführung in die neue Wahrscheinlichkeitslehre und ihre Anwendungen* (Probability, statistics and truth. Introduction to the new probability theory and its applications) [11]. There are several difficulties in von Mises' theory, undoubtedly the deepest attempt to define randomness as the natural basis for a frequentist and objective view of probability.

The main objection to frequentism is that most events are not repeatable, and in this case it is impossible, even in principle, to apply a frequentist definition of probability based on frequencies simply because these frequencies cannot be measured at all. von Mises explicitly excluded these cases from his theory. In other words, given an event that is not repeatable such as *tomorrow it will rain*, it is a nonsense to ask for its probability. Notice that most of economics would fall

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51559-7 - Finitary Probabilistic Methods in Econophysics
Ubaldo Garibaldi and Enrico Scalas
Excerpt
More information



outside the realm of repeatability. If one were to fully accept this point of view, most applications of probability and statistics to economics (including most of econometrics) would become meaningless. Incidentally, the success of frequentism could explain why there are so few probabilistic models in theoretical economics. The late Kolmogorov explored a method to avoid problems due to infinite sequences of trials by developing a *finitary* frequentist theory of probability connected with his theory of information and computational complexity.

Note that the use of observed frequencies to calculate probabilities can be traced back to Huygens, who used mortality data in order to calculate survival tables, as well as bets on survival. The success of frequentism is related to the social success among statisticians and natural scientists of the methods developed by R.A. Fisher who was a strong supporter of the frequentist viewpoint. Namely, Fisher tried to systematically exclude any non-frequency notion from statistics. These methods deeply influenced the birth of econometrics, the only branch of economics (except for mathematical finance) making extensive use of probability theory. The frequentism adopted by a large majority of orthodox[2] statisticians amounts to refusing applications of probability outside the realm of repeatable events. The main effect of this attitude is not using a priori or initial probabilities within Bayes' theorem (*la probabilité de causes par l'evenements* in the Laplacean formulation). This attitude is just setting limits to the realm of probability, and has nothing to do with the definition of the concept. A sequence of $n$ trials such as $HHT \ldots H$ should be more correctly written as $H_1 H_2 T_3 \ldots H_n$, and almost everybody would agree that $\forall i, \mathbb{P}(H_i) = (\#\text{ of } H)/n$, if the only available knowledge is given by the number of occurrences of $H$ and by the number of trials $n$. However, usually, one is mainly interested in $\mathbb{P}(H_{n+1})$: the probability of an event not contained in the available evidence. There is no logical constraint in assuming that at the next trial the coin will behave as before, but after Hume's criticism of the Principle of Uniformity of Nature, this is a practical issue. Therefore, assuming (1.10) turns out to be nothing else than a *working hypothesis*.

## 1.4 Subjectivism, neo-Bayesianism and logicism

Frequentism wants to eliminate the subjective element present in the indifference principle. On the contrary, subjectivism accepts this element and amplifies it by defining probability as the degree of belief that each individual assigns to an event. This event need not refer to the future and it is not necessary that the

[2] The adjective orthodox is used by the neo-Bayesian E.T. Jaynes when referring to statisticians following the tradition of R.A. Fisher, J. Neyman and E. Pearson.

© in this web service Cambridge University Press www.cambridge.org