

1

Historical Notes

1.1 Introduction

Probability is the measure of how likely a future event is. It is not by “chance” that most of the examples related to the understanding of probability are connected to objects like dices, cards, or coins. Historical events show that most primitive attempts on probability theory have the roots in gambling [1]. Given the implications that gambling had over time, particularly the social consequences of it, great efforts have been made to avoid or understand uncertainty. Historians have looked to Aristotle and beyond when searching for the origins of the probabilistic concepts [1]. The very first ideas for this fundamental principle may derive directly from Aristotle’s Ethics, where the concept of “justice” took new forms over time [1]. Later, the medieval poem *De Vetula* (“On the Old Woman”) appeared around the year 1250. This poem is a long thirteenth-century elegiac comedy that contains first written references on gambling [2]. The non-poetic content of *De Vetula* makes references to the connection between the number of combinations and the expected frequency of a given total [2]. Gerolamo Cardano (1501–1575) has made the first written references in defining odds as the ratio of favorable to unfavorable outcomes [1]. In his *Liber de Ludo Aleae* (“Book on Games of Chance”) published in 1564 or later, Cardano was among the first to approach probabilities in games of chance [1]. A few decades later, uncertain events related to gambling resulted in the well-known mathematical theory of probability formulated by Pierre de Fermat and Blaise Pascal (1654) [3]. Just 3 years later in 1657, Christian Huygens (1629–1695) published a dedicated book on probability theory related to problems associated with games of chance, entitled *De Ratiociniis in Ludo Aleae* (“On Reasoning in Games of Chance”) [4, 5]. A milestone contribution of Jakob Bernoulli (1654–1705) in probability theory was published post-mortem in 1713, under the title *Ars conjectandi* (“The Art of Conjecturing”) [6, 7]. Bernoulli was concerned with predicting the probability of unknown events [7]. In his work Bernoulli describes what today is known as the weak law of large numbers [7]. This law shows that the average of the results obtained from an



increasing number of trials should converge toward the expected value [7]. For instance, consider the flipping of an unbiased coin. As the number of flips goes to infinity, the proportion of heads will approach $1/2$. Let us consider another example: without our knowledge, x black balls and y white balls are placed in a jar (Figure 1.1a). To determine the proportions of black balls and white balls from the jar by experiment, a series of random draws must be performed. Whenever a black ball or a white ball is drawn, the observation is noted. The expected value of white versus black observations will converge toward the real ratio from the jar as the number of extractions increases. Therefore, Bernoulli proved that after many random draws (trials), the observations of white balls versus black balls will converge toward the real ratio of black balls versus white balls from the jar. Almost 100 years after Bernoulli, Pierre de Laplace (1749–1827) severs the thinking monopoly that gambling had on the probability theory [1–7]. In 1812, Laplace publishes the *Théorie Analytique des Probabilités* (“Analytical Theory of Probability”) in which it introduces probability to general sciences [8, 9].

1.2 On the Wings of Dependent Variables

Another milestone in probability theory was made almost 100 years later by Andrei Markov (1856–1922) [10, 11]. In the Bernoulli model, the outcome of previous events does not change the outcome of current or future events. Today it is obvious that the events are not independent in many cases; however, in the past, this was not that obvious (in the mathematical sense). The term of *dependent events*, or *dependent variables*, refers to those situations when the probability of an event is conditioned by the events that took place in the past. A colleague of Markov, namely Pavel Nekrasov, assumed that independence is a condition for the law of large numbers [12]. Following a dispute with Pavel Nekrasov, Markov considered that the law of large numbers

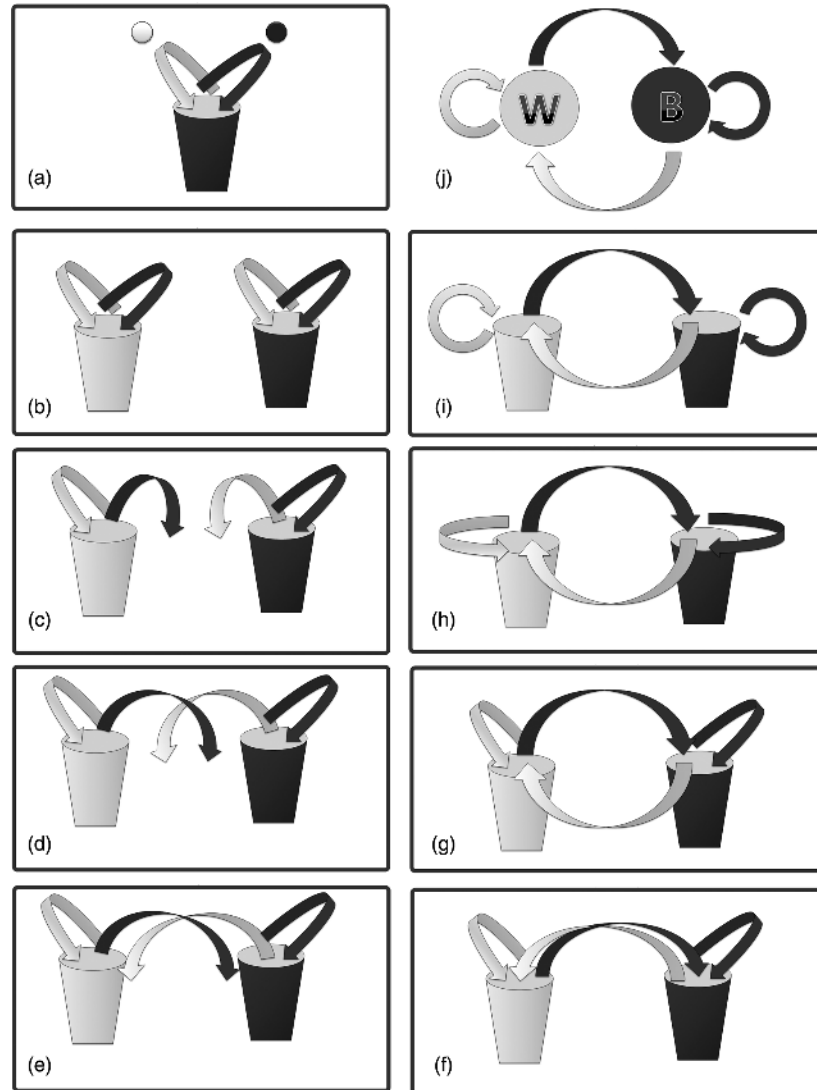


Figure 1.1 From Bernoulli model to the Markov model. (a) The Bernoulli model with a single jar filled with different proportions of black and white balls. The curved arrows show from which jar the extraction was done and to which jar the next draw will be made. Black curved arrows show the extraction and observation of black balls whereas white curved arrows show the extraction and observation of white balls. (b) Two Bernoulli jars. A white jar and a black jar, each filled with different proportions of black and white balls. Here, draws are still independent from one another. (c–f) Shows how dependence occurs between the two jars if the color of the curved arrows is “attracted” to the color of the jars. (g–j) Shows how the two jars system is transformed into a Markov diagram by changing the angle of viewing of the jars, from the side view to the top view.

4 | *Markov Chains*

can be also valid in the case of dependent variables. In 1906, Markov extends Bernoulli's results to dependent variables and began developing his reasoning about chains of linked probabilities (Figures 1.1a–j) [10]. Markov's connection with Bernoulli it is indirectly but deeply rooted in the history of the Academy of Sciences in St. Petersburg. Prior to Markov's time, the Academy included none other than the great Leonhard Euler (1707–1783) and the sons of Jakob Bernoulli, namely Nicholas Bernoulli (1687–1759) and Daniel Bernoulli (1700–1782) [12]. In 1907, Markov proved that the independence of random variables was not a required condition for the validity of the weak law of large numbers and the central limit theorem [10, 11]. For his demonstration, he envisioned a virtual machine (Figures 1.1f and 1.1j).

Let us consider two jars which represent the two states of a machine (Figure 1.1b). One is painted in black (state 1) and the other is painted in white (state 0). Both contain certain proportions of both white and black balls. First, an extraction of a ball from one of the jars is made; let us choose the black jar (state 1). If the black ball is drawn, then the next draw is made again from the black jar (state 1). If the white ball is drawn, then the next draw is made from the white jar (state 0). Let us consider that a white ball has been pulled. Therefore, the next draw is made from the white jar. If a white ball is drawn from the white jar, then the next draw is made again from the white jar (state 1). If a black ball is drawn from the white jar then the next draw is made from the black jar (state 0). Thus, these events may continue indefinitely. What can be immediately noticed is that the current extraction is dependent on the previous extraction. As long as both states of the machine are reachable (each jar contains both white and black balls), the number of visits to each jar will converge, as in the Bernoulli model, to a specific ratio. By this simple example, Markov showed that the law of large numbers applies in the case of dependent variables. But what does “both states are reachable” mean? If the black jar has only black balls inside, then all drawings are made from the black jar; therefore, the white jar is unreachable. However, what if draws are first started from the white jar? Eventually, after a number of draws, a black ball is drawn from the white jar. Once the black ball is drawn from the white jar, the next draw will be made from the black jar. Since the black jar contains in this case only black balls, from this point forward, all future draws will be made only from the black jar. Therefore, the white jar will be unreachable. In order to make the white jar reachable, the black jar must contain at least one white ball from the total number of balls (n). That one white ball it will provide a very small chance ($1/n > 0$), of switching the extraction of balls from the black jar to the white jar. Taking these observations into account, the probability of extracting a white ball from the black jar will be:

$$P[\text{white}] = \frac{1}{n}$$

Whereas the probability of extracting a black ball from the black jar will be:

$$P[\text{black}] = \frac{(n-1)}{n}$$

Also notice that:

$$P[\text{white}] + P[\text{black}] = \left(\frac{1}{n}\right) + \frac{(n-1)}{n} = 1$$

In 1913, by pencil and paper, Markov applied his method for a linguistic analysis of the first 20,000 letters from one of Pushkin's poems [12]. Thus, he showed that the letter probabilities in Pushkin's poem are not independent. This linguistic analysis sparked the interest of many scientists at that time and quickly brought a worldwide revolution in science and technology [12]. Many great minds preoccupied by uncertainty made their contribution over time to the probability theory. Nevertheless, what had begun as an analysis of gambling rooted in decadence is now the main weapon used for the progress of mankind.

1.3 From Bernoulli to Markov

Simple exemplifications are crucial for understanding the Markov process. A stochastic process is visually represented by state diagrams. Circles inside a diagram represent states while arrows indicate the probability of moving from one state to another (Figure 1.1j). In our days, state diagrams are usually taken for granted without any possible roots underlying these types of visual representations. As stated above, A. Markov was influenced by Jacob Bernoulli [10]. Therefore, let us start from Bernoulli's jar (Figure 1.1a). Consider an opaque jar filled with white and black balls whose proportions are unknown. Of course, a simple counting of the balls is not allowed. What is allowed is a repeated extraction of balls from the jar and the notation on paper of how many balls are white and how many balls are black. The question is: can a similar jar be filled with the same proportion of white and black balls? The expected value of white versus black observations will converge toward the real ratio from the jar, as the number of extractions increases. Also notice that in Bernoulli model, the outcome of previous events does not change the outcome of current or future events. Thus, all draws are independent of each other. Let us consider that in this Bernoulli model, the aim is to find out what is the probability of a specific sequence of draws, for instance black ball, white ball, black ball, black ball: $P[\text{BWBB}]$. If these four independent events (BWBB) have probabilities $P[B]$ and $P[W]$, the joint probability of all four events is the product:

$$P[\text{BWBB}] = P[B] \times P[W] \times P[B] \times P[B]$$

In other words, the probabilities are not linked. But what does that mean? How can probabilities be linked? Consider two jars which represent the two states

6 | Markov Chains

of a machine (Figure 1.1b). One jar is painted in black (state B) and the other jar is painted in white (state W). Both jars contain white balls and black balls in unknown proportions. Thus, the percentage of white versus black balls in each jar is the aim of this example. What can be done? Consider that two individuals are involved in an experiment. These two individuals are named Alice and Bob. Presumably, a first draw is made from one of the jars by Alice. A draw rule is also imposed, namely the color of the current ball indicates the color of the jar from which the next draw will be made (Figure 1.1f). By following this rule, suppose 80 draws are made by Alice behind a screen. Therefore, the interplay between jars (states) is not observable by Bob. However, Alice shows Bob the ball each time a black ball is drawn from the black jar, or a white ball is drawn from the white jar. The color of the ball is written on paper by Bob each time Alice shows him a ball. Thus, at the end of the 80th draw, Bob notes that Alice showed him the black ball 20 times and the white ball 30 times. Thus, Alice asks Bob to tell her the proportion of balls in each jar based exclusively on these observations. Taking into account the draw rule, Bob immediately realizes that each time Alice showed him the black ball she returned to the black jar, and each time she showed him the white ball she returned to the white jar. Initially, Bob imagines the jars as independent systems (Figures 1.1b and 1.1c). When regarded as an independent Bernoulli experiment with a single jar, the proportion of balls can be easily determined. Thus, the probability to draw a black ball ($P_{\text{white jar}}[B]$) or a white ball ($P_{\text{white jar}}[W]$) from the white jar equals 1:

$$P_{\text{white jar}}[B] + P_{\text{white jar}}[W] = 1$$

If the probability of extracting a white ball ($P_{\text{white jar}}[W]$) from the white jar is known, then the probability of extracting a black ball from the white jar can also be found, namely (Figure 1.2b):

$$P_{\text{white jar}}[B] = 1 - P_{\text{white jar}}[W]$$

Similarly, the same rules apply to the black jar. The probability of extracting a black or a white ball from the black jar is also equal to 1 (Figure 1.2c). Taking into consideration the above properties, Bob wonders how he can approximate the proportion of white versus black balls in each jar. In order to find out the proportion of balls from the two jars, he must deduce the unknown probabilities from the two jars, namely $P_{\text{black jar}}[W]$ and $P_{\text{white jar}}[B]$ (Table 1.1). First, it starts from what he knows and divides the number of observations by the total number of draws to obtain $P_{\text{black jar}}[B]$ and $P_{\text{white jar}}[W]$ (Table 1.1). Secondly, it uses the expression $1 - P_{\text{white jar}}[W]$ to find $P_{\text{white jar}}[B]$ and the expression $1 - P_{\text{black jar}}[B]$ to find $P_{\text{black jar}}[W]$ (Table 1.1). In the final step, Bob transforms the probabilities into percentages by multiplying the probability values with 100. Bob then shows his findings to Alice in the form of proportions: the white jar contains ~62.5% black balls and 37.5% white balls whereas the black jar contains ~75% white balls and 25% black balls. The value of $P_{\text{black jar}}[B]$ represents

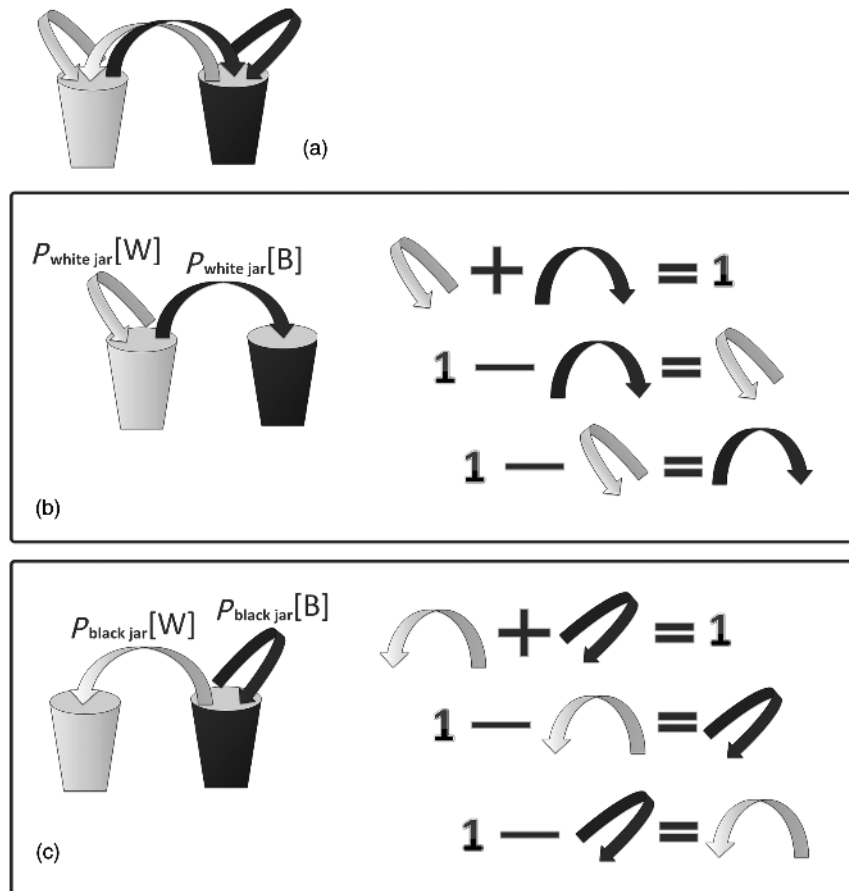


Figure 1.2 The two jars seen independently when disconnected from the Markov diagram. (a) The Markov model represented by two jars. (b) The white jar observed independently from the black jar. (c) The black jar observed independently from the white jar. The black arrows signify the transitions to the black jar, and the white arrows signify the transitions to the white jar.

the black balls in the black jar and the value of $P_{\text{white jar}}[W]$ represents the white balls in the white jar. Therefore, in this case, $P_{\text{white jar}}[B]$ and $P_{\text{black jar}}[W]$ counterparts represent the interplay between jars (states). Consequently, what was initially hidden behind the screen by Alice is now clear to Bob. Approximately 62.5% of the time the extractions were shifted from the white to the black jar and ~75% of the time the extractions were shifted from the black to the white jar. In other words, the number of trips from the white jar to the black jar represents the proportion of black balls inside the white jar (Figure 1.2b). In contrast,

8 | *Markov Chains***Table 1.1** Finding the unknown probabilities of the two jars.

In the black jar	In the white jar
$P_{\text{black jar}}[B] = 20/80 = 0.25$	$P_{\text{white jar}}[W] = 30/80 = 0.375$
$P_{\text{black jar}}[W] = 1 - P_{\text{black jar}}[B]$	$P_{\text{white jar}}[B] = 1 - P_{\text{white jar}}[W]$
$P_{\text{black jar}}[W] = 1 - 0.25$	$P_{\text{white jar}}[B] = 1 - 0.375$
$P_{\text{black jar}}[W] = 0.75$	$P_{\text{white jar}}[B] = 0.625$

the number of trips from the black jar to the white jar represents the proportion of white balls inside the black jar (Figure 1.2c). When the Markov rule applies, the two jars have linked probabilities. Thus, at least one parameter of each jar (state) must be known in order to rebuild the unknown transition probabilities from the system (Figures 1.2b and 1.2c).