Elements of Discrete Probability
1. From Frequencies to Probabilities

1 Introduction

The analytical sciences are concerned with the study of ”phenomena” in their most extensive connotation. The meaning of the word ”phenomenon” is ”something that appears”, i.e., it is observable, although the notion of observability is naturally extended from the physical experiments to the thought-experiments (namely, experiments o ccurring in a mental mo del of a physical reality).

If the phenomenon is a physical system (for example, a ro cket travelling in space), frequently we are able to mo del it so precisely that its behavior is extremely well described by the solution of a mathematical equation. In such case, in fact, the mechanism of the phenomenon is well understood and the ”causes” of the behavior (gravitational attractions, position and velocity of the rocket) are measurable with great accuracy. The cause-effect relationship is faithfully captured by a mathematical (differential) equation. We say that the behavior (effect) is fully determined by the given causes, i.e., it is *deterministic*.

However, such is not the norm. On a daily basis we experience phenomena and systems that hardly fall in the above category. Such phenomena may be either physical events (as the tossing of a coin) or pertain to complex social interactions (such as the behavior of a waiting line at a supermarket checkout counter). In all such cases, we recognize the action of a *large* number of causes and, most significantly, our inability to provide a description of them.

Consider, for example, the tossing of a coin. Even assuming that the experiment is carried out in the vacuum so that there is no interaction between coin and air , although there is no doubt that the motion of the coin is determined by the initial conditions impressed by the hand, we are in no position to specify the latter with adequate accuracy.

Or consider the behavior of the waiting-line at a check-out counter. A person joins a line at the completion of his/her shopping. In this case we have no way to accurately predict the time when this event may occur, for too many causes are at play (length of shopping list, distractions, etc.)

In conclusion, there are phenomena both with a large number of causes and with an unknown mechanism, so that their outcome is unpredictable (when compared with events of classical mechanics).

However, we detect some regularity in these phenomena that we cannot predict exactly. We can measure the (relative) frequencies of different outcomes, for example, the numbers of heads and tails when tossing a coin, or the length of the waiting line when measured at regular intervals. In most cases, the observed regularity (e.g., nearly equals number of heads and tails) is a suggestion to assume that the phenomenon is governed by a pro cess that is invariant in time (at least in the short run). This process is unknown and we do not even attempt to describe its mechanism. However, we assume that it produces its different outcomes in proportions approximating the observed relative frequencies. These proportions are called *probabilities* and are modelled on the basis of observed frequencies.

The above discussion is simply a motivation for the emergence of the notion of probability. We leave to different forums the difficult debate on the essential nature of probability (whether intrinsic in reality or a creation of the human mind).

2 Sample spaces

Suppose we have a box of s physically identical balls labeled 1, 2, . . . , s. Say $s = 4$. We will remove a ball at random from the box, note its number and then return the ball to the box. If we repeat this process 20 times, we may have a chart that looks something like this:

number:	1	2	3	4
count:	3	5	4	8

So in 20 repetitions of our experiment we have drawn the "1"-ball three times, the "2"-ball five times, and so on. From here we can calculate the fraction of the outcomes that resulted in a "1," "2," "3," or "4." Also known as the *relative frequency*, we simply divide our count by the total number of times we performed the experiment:

number:	1	2	3	4
relative frequency:	.15	.25	.2	.4

We can expect that after a sufficiently large number n of repetitions, the relative frequencies of all of the balls should approach some set numbers $p1,p_2, p_3, \ldots, p_s$ (which, according to our intuition, should all equal $1/s$). If we look at these p_i 's as relative frequencies, then we postulate that p_i is the likelihood that the ball labeled i will be drawn if we perform the experiment only once.

The previous example motivates the introduction of some fundamental notions. The most fundamental notion in probability theory is the "occurrence of an event" or simply "an event". The "simplest" events, called *elementary events* or *points* of the sample space, are the possible outcomes of an experiment, such as head, tail in tossing a coin or 1,2,3,4,5,6 in rolling a die.

The outcomes, or elementary events, form a set S , as evidenced by the notation used above.

Definition 1 *A* sample space *is a pair* $(S, \Pr)$*, where* S *is a set (called itself the sample space) of "elementary events " and* Pr *is a function* $\Pr : S \to [0, 1]$ *(called a* probability*), such that* $\sum_{a \in S} \Pr(a) = 1$*. (Note* $[0, 1]$ *is the set of real numbers* x*,* $0 \leq x \leq 1$ *.)*

Definition 2 *An event* E *is any subset of a sample space. The probability of an event* E *is the sum of the probabilities of its constituent elementary events, i.e.,*

$$\Pr(E) = \sum_{a \in E} \Pr(a)$$

In other words, any member of the power set $\mathcal{P}(S)$ is itself an event: For example, an even-valued outcome in rolling a die is the subset 2, 4, 6.

A useful way to picture a sample space is as a collection of individual points (in the plane), each representing an elementary event, with appended its corresponding probability. For example, in Figure 1, we have illustrated the sample space describing the rolling of two dice, where we have outlined as a subdomain the event "the sum is 5". Not shown is the assignment of probability to each point: if the dice are fair all points carry identical value, 1/36.

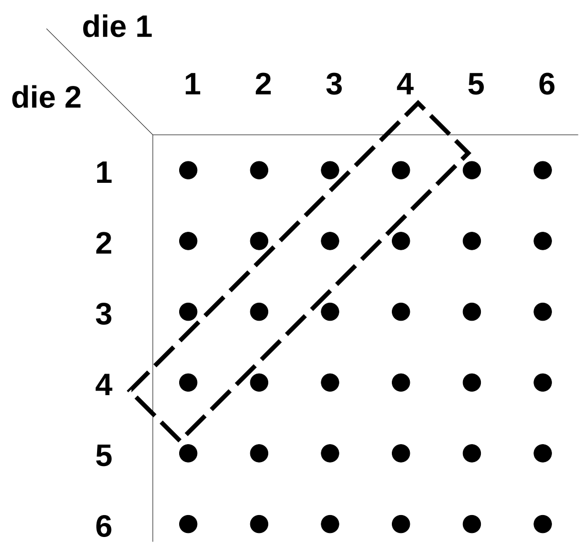

Figure 1: Illustration of the event "sum has value" 5 in the rolling of two dice.

The given set characterization immediately implies that any any set-theoretic expression whose variables are the elementary events is itself an event. The corresponding boolean algebra is refered to as the algebra of events.

In principle, the probability corresponding to any such set-theoretic expression is the sum of the probabilities of the elementary events contained in the defined set, and can be visualized as the area of the region defined in the appropriate Venn diagram.

For events A and B, two subsets of the sample space S , we shall focus on the probabilities $\Pr(A^c)$, $\Pr(A \cup B)$ and $\Pr(A \cap B)$. A^c is the event that A does not occur, i.e., Ac is the opposite event of A. Therefore,

$$\Pr(A^c) = \sum_{a \in S-A} = \sum_{a \in S} \Pr(a) - \sum_{a \in A} \Pr(a) = 1 - \Pr(A).$$

For example, what is the probability that a "randomly generated" binary string of 4 symbols do es not contain the symbol 0 [1]. There are 16 such strings (the sample space), and only one of them, 1111, satisfies the given condition. Thus the answer is 1 - 1 / 16 = 15/16 .

Next, consider the following experiment. A box contains a set of n balls, each of which is colored and carries an integer. We draw a ball from this box. B is the event that the ball we draw is odd (carries an odd number), and A is the event that the ball we draw is blue. What can we say about the event E that we draw an odd ball or a blue ball[2] . If A denotes the event "the ball is blue" and B the event "the ball is odd", then the subset of the sample space defining our event E is $A \cup B$. The probability $\Pr(A)$ that the ball is blue is given by the ratio b/n, where b is the number of blue balls; analogously, the probability $\Pr(B)$ that the ball is odd is given by o/n, where o is the number of odd balls. However, some blue ball may also have an odd number and should not be counted twice. Thus, after adding $\Pr(A)$ and $\Pr(B)$ we should subtract $\Pr(A \cap B)$, giving the probability that a ball be simultaneously blue and odd. We conclude that

$$\Pr(A \cup B) = \sum_{a \in A} \Pr(a) + \sum_{a \in} \Pr(a) - \sum_{a \in A \cap B} \Pr(a) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

which generalizes the principle of inclusion/exclusion (the latter simply counted the numbers of elements in each set; here we are allowing each point to have a different probability). As an immediate consequence, when the two events are "disjoint', that is, $A \cap B = \emptyset$, we have $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

The case just discussed can be characterized as a "disjoint occurrence"; what can we say about the probability of a "joint occurrence", that is, $\Pr(A \cap B)$? This case is more subtle, and will be discussed next.

[1] randomly generated here means that all possible 4-symbol strings appear with identical probability

[2] As usual, the "or" we refer to is an *inclusive* or meaning the phrase "a or b" really means "a or b or both"

3 Joint and conditional probability, Bayes' rule

We have again a box containing n balls of two colors (white and black), each also labelled with an integer. Of the n balls, w are white and e are even-numbered. Clearly, the probability that a ball drawn from the box is white is w/n, and that it is even-numbered is e/n. We also know that $e_w \leq w$ of the white balls, and $b_w \leq n - w$ of the black balls, are even-numbered. Consider now the following experiment. Suppose someone draws a ball from the box, reveals only its color and asks us to guess its numerical parity (that is, the guess corresponds to give a value to the probability of odd or even parity). For example, if we know that $e_w = 0$ (there is no even white ball) and we are told the ball is white, then we predict with certainty (probability 1) that the parity is odd. Although the case is extreme ($e_w = 0$), we see that the knowledge of the occurrence of an event (white ball) affects our knowledge of another event (i.e.,in the described case it alters $\Pr(odd)$ from $1 - e/n$ to 1). This motivates the introduction of the notion of *conditional probability*

Definition 3 *Let A and B be two events such that $\Pr(A) > 0$. The* conditional probability *of B given A, written $\Pr(B|A)$, is defined to be*

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

On the basis of this definition, we are now able to answer our box-and-ball question in general. Let A denote the event of drawing a white ball and let B denote the event of drawing an odd-labeled ball. In order to compute $\Pr(B|A)$, we must first compute the probabilities $\Pr(A \cap B)$ and $\Pr(A)$.

We know that there are a total of n balls, w of which are black, so $\Pr(A) = w/n$. The probability that a ball be simultaneously white and odd, i.e., the $\Pr(A \cap B)$, is the number $w - e_w$, divided by n. Thus

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{\frac{w-e_w}{n}}{\frac{w}{n}} = 1\frac{e_w}{w}$$

To gain further insight into the nature of "conditioning", consider now $\Pr(B)$, the probability that a ball is odd-numbered, whose value is clearly $1 - e/n$. We now note that

$$\Pr(A \cap B) = \Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$$

and dividing by $\Pr(B)$ we obtain

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

These identities are called **Bayes' rule** and can be used to infer a cause A from an observed effect B . The crucial point is that the observation of "effect" B, changes the *prior* probability $\Pr(A)$ into the *posterior* probability $\Pr(A|B)$.

Example 1. Suppose we have two fair coins. Coin 1 has heads on both sides and Coin 2 has heads on one side and tails on the other side. We select one of the two coins at random and, without looking at it, flip it. It comes up heads. What is the probability that we chose Coin 1?

Let H denote the event that the flip comes up heads (the observed effect) and let $C1$ denote the event that we chose coin 1 (a possible cause of H). To compute the desired $\Pr(B|H)$, we need to obtain $\Pr(H)$, $\Pr(H|C1)$, and $\Pr(C1)$. Event H is the union of two disjoint events: either we chose Coin 1 (event C1) and flipped heads (event H) or we chose Coin 2 (call this event C2) and flipped heads. So $H = (C1 \cap H) \cup (C2 \cup H)$, giving

$$\Pr(H) = \Pr(C1 \cap H) + \Pr(C2 \cap H)$$

. Using the definition of conditional probability, we have

$$\Pr(C1 \cap H) = \Pr(C1) \Pr(H|C1)$$

$$\Pr(C2 \cap H) = \Pr(C2) \Pr(H|C2)$$

By the way we choose the coin, $\Pr(A) = \Pr(B) = 1/2$. Equally simply we obtain $\Pr(H|C2) = 1/2$ (Coin 2 has heads and tails and is fair), but $\Pr(H|C1) = 1$ (Coin 1 has heads on both sides) so that we evaluate $\Pr(H)$ as

$$\Pr(H) = \Pr(C1) \Pr(H|C1) + \Pr(C2) \Pr(H|C2) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$$

We conclude therefore

$$\Pr(C1|H) = \frac{\Pr(H|C1) Pr(C1)}{\Pr(H)} = \frac{1 \times 1/2}{3/4} = \frac{2}{3}$$

The prior $\Pr(C1) = 1/2$ has been replaced by the posterior $\Pr(C1|H) = 2/3$.

Next, we shall analyze a more complex example, whose solution is not evident on the basis of intuition alone.

Example 2 In the not so remote past a game show called *Let's Make a Deal*, worked as follows. To win the grand prize, the Host would show the Contestant three doors, numbered 1,2 and 3. Behind one door there would be a prize; behind the others there was nothing. The Contestant was asked to choose the door he/she guessed concealed the prize. The Host would then open one of the *unselected* two doors to show that there was nothing behind that it. He would then ask the Contestant if he/she would like to switch from the originally chosen door to the one unopened door.

Let Pi be the event that the prize lies behind Door i, and let Dj be the event that the Host opens Door j $(i, j = 1, 2, 3)$.

Suppose the Contestant chooses Door 2. The Host opens Door 1. What is the probability that the prize lies behind Door 3 (given that the Host opened Door 1, an event observed by the Contestant)? From Bayes' rule applied to events we know that

$$\Pr(P3|D1) = \frac{\Pr(P3)\Pr(D1|P3)}{\Pr(D1)}$$

So, to answer our question, we must determine the terms $\Pr(P3), \Pr(D1|P3)$ and $\Pr(D1)$.

We first have

$$\Pr(D1) = \Pr(P1)\Pr(D1|P1) + \Pr(P2)\Pr(D1|P2) + \Pr(P3)\Pr(D1|P3)$$

Clearly, $\Pr(P1) = \Pr(P2) = \Pr(P3) = 1/3$, since it is equally likely that the prize lies behind any of the three doors. The crux is the analysis of the conditional probabilities (arrived at on the basis of the fact that the Host does not want to disclose the location of the prize):

- If the prize lies behind Door 1 (event $P1$) the Host certainly will not open Door 1 (event $D1$), so $\Pr(D1|P1) = 0$.
- If the prize lies behind Door 2 (event $P2$) the Host is equally likely to open Door 1 (event $D1$) or Door 3, so that $\Pr(D1|P2) = 1/2$.
- If the prize lies behind Door 3 (event $P3$) the Host will necessarily open Door 1 (event $D1$), since he cannot open either Door 2 chosen by the Contestant or Door 3 concealing the prize, so that $\Pr(D1|P3) = 1$.

We conclude

$$\Pr(D1) = \frac{1}{3} \cdot 0 + \frac{1}{3}\frac{1}{2} + \frac{1}{3} \cdot 1 = \frac{1}{2}$$

This argument also shows that

$$\Pr(P3)\Pr(D1|P3) = \frac{1}{3} \cdot 1 = \frac{1}{3}$$

so that we have the final answer:

$$\Pr(P3|D1) = \frac{1/3}{1/2} = \frac{2}{3}$$

The probability that the prize lies behind Door 3, given that the *Host opened Door 1*, is 2/3, i.e., the original prior 1/3 is to be replaced by the conditional 2/3.

We verify that

$$\Pr(P3|D2) = \frac{\Pr(P3)\Pr(D2|P3)}{\Pr(D2)} = \frac{1/3 \cdot 1/2}{1/2} = 1/3$$

4 Independence of events

The fact that the occurrence of an event A modifies our expectations about the occurrence of another event B, reveals some "connection" between A and B, as well as the absence of any modification would point to the absence of any connection. Such connection is technically called *(statistical) independence*.

Specifically:

Definition 4 *Two events are said to be independent if and only if*

$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

From this definition we see that if A and B are independent, then

$$\begin{aligned}
\Pr(B|A) &= \frac{\Pr(A \cap B)}{\Pr(A)} \\
&= \frac{\Pr(A)\Pr(B)}{\Pr(A)} \\
&= \Pr(B)
\end{aligned}$$

i.e., the probability of B (also called *marginal* probability) and its probability conditional on A coincide.

At this point, it is important to stress that the notions of "disjointness" and "independence" of events are completely distinct. Two events A and B are disjoint when $A \cap B = \emptyset$, that is, when the two events share no sample space point; A and B are independent when their probabilities satisfy the above definition. In fact, note that if two events are disjoint, they are certainly NOT independent, since the occurrence of one excludes the occurrence of the other. For example, consider the rolling of a die, where we have $\Pr(odd) = \Pr(even) = 1/2$. We also have $\Pr(odd \cap even) = 0 \neq \Pr(odd)\Pr(even)$ (knowing that the outcome is even, obviously excludes the alternative).

Contrapositively, if two events are independent, then they are not disjoint. However, we must avoid making the "inverse error": They may be not disjoint and yet not independent. Consider the following example: We have two coins, a fair one C1, and an unfair one C2, with probabilities 1/3 and 2/3 for "head (H)" and "tail (T)", respectively. The sample space has four points (see Figure 2) with $\Pr(HH) = 1/6, \Pr(HT) = 2/6, \Pr(TH) = 1/6$, and $\Pr(TT) = 2/6$. Define the following events:

1. $A1$: C1 comes up heads;
2. $A2$: The outcomes of C1 and C2 are different.

Clearly, $A1 = HH \cup HT$ and $\Pr(A1) = 1/6 + 2/6 = 1/2$; analogously, $A2 = HT \cup TH$ and $\Pr(A2) = 2/6 + 1/6 = 1/2$. However, $A1 \cap A2 = HT$, so that $\Pr(A1 \cap A2) = 1/3$. We conclude

$$\frac{1}{3} = \Pr(A1 \cap A2) \neq \Pr(A1)\Pr(A2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

as we wanted to illustrate.

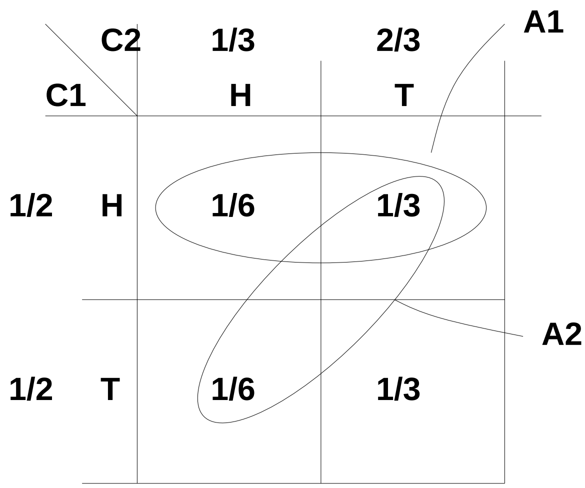

Figure 2: Events A1 and A2 are not disjoint and are not independent

So far we have considered "pairs" of events and have analyzed *pairwise independence*. We now consider sets of events.

Definition 5 *Events $A_1, A_2, \ldots, A_n$ are said to be* collectively *independent if for any subset $\{A_{j_1}, ..., A_{j_s}\}$ of their set*

$$\Pr(A_{j_1} \cap \ldots \cap A_{j_s}) = |Pr(A_{j_1}) \ldots \Pr(A_{j_s})$$

Finally, we note that pairwise independence and collective independence are themselves independent notions. Consider, the following modification of the preceding example: Coins C1 and C2 are both fair, so that each point of the sample space has probability 1/4. In addition to the previously defined events A1 and A2 consider (refer to Figure 3)

- A3: C2 comes up heads.

It is immediately recognized that $\Pr(A1) = \Pr(A2) = Pr(A3) = 1/2$; moreover, $\Pr(A1 \cap A2) = \Pr(HT) = 1/4$,$\Pr(A1 \cap A3) = \Pr(HH) = 1/4$, and $\Pr(A2 \cap A3) = \Pr(TH) = 1/4$, so that the three events are pairwise independent. But now

$$\Pr(A1 \cap A2 \cap A3) = \Pr(\emptyset) = 0 \neq \frac{1}{8}$$

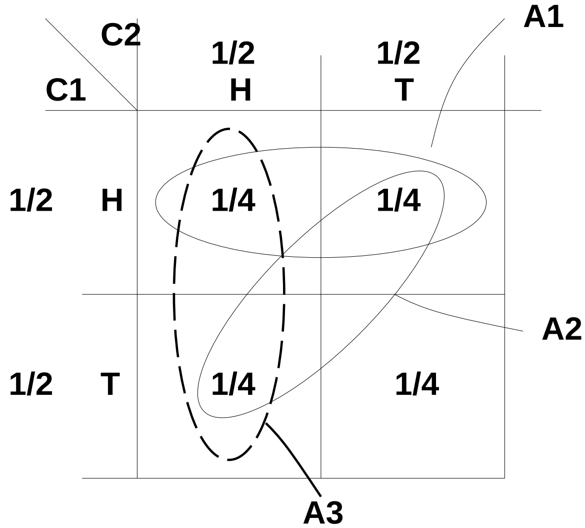

Figure 3: Events A1, A2, and A3 are pairwise but not collectively independent