2. Probabilities

2.1 Experiments, Events, Sample Space

Since in this book we are concerned with the analysis of data originating from experiments, we will have to state first what we mean by an experiment and its result. Just as in the laboratory, we define an experiment to be a strictly followed procedure, as a consequence of which a quantity or a set of quantities is obtained that constitutes the result. These quantities are continuous (temperature, length, current) or discrete (number of particles, birthday of a person, one of three possible colors). No matter how accurately all conditions of the procedure are maintained, the results of repetitions of an experiment will in general differ. This is caused either by the intrinsic statistical nature of the phenomenon under investigation or by the finite accuracy of the measurement. The possible results will therefore always be spread over a finite region for each quantity. All of these regions for all quantities that make up the result of an experiment constitute the *sample space* of that experiment. Since it is difficult and often impossible to determine exactly the accessible regions for the quantities measured in a particular experiment, the sample space actually used may be larger and may contain the true sample space as a subspace. We shall use this somewhat looser concept of a sample space.

Example 2.1: Sample space for continuous variables

In the manufacture of resistors it is important to maintain the values R (electrical resistance measured in ohms) and N (maximum heat dissipation measured in watts) at given values. The sample space for R and N is a plane spanned by axes labeled R and N. Since both quantities are always positive, the first quadrant of this plane is itself a sample space. ■

S. Brandt, *Data Analysis: Statistical and Computational Methods for Scientists and Engineers*, DOI 10.1007/978-3-319-03762-2__2,

Example 2.2: Sample space for discrete variables

In practice the exact values of R and N are unimportant as long as they are contained within a certain interval about the nominal value (e.g., $99\,\mathrm{k}\Omega < R < 101\,\mathrm{k}\Omega$, $0.49\,\mathrm{W} < N < 0.60\,\mathrm{W}$). If this is the case, we shall say that the resistor has the properties R_n, N_n. If the value falls below (above) the lower (upper) limit, then we shall substitute the index n by $-(+)$. The possible values of resistance and heat dissipation are therefore R_-, R_n, R_+, N_-, N_n, N_+. The sample space now consists of nine points:

$$\begin{array}{lll} R_- N_-, & R_- N_n, & R_- N_+, \\ R_n N_-, & R_n N_n, & R_n N_+, \\ R_+ N_-, & R_+ N_n, & R_+ N_+. \end{array} \quad \blacksquare$$

Often one or more particular subspaces of the sample space are of special interest. In Example 2.2, for instance, the point R_n, N_n represents the case where the resistors meet the production specifications. We can give such subspaces names, e.g., $A, B, \ldots$ and say that if the result of an experiment falls into one such subspace, then the *event* A (or $B, C, \ldots$) has occurred. If A has not occurred, we speak of the complementary event $\bar{A}$ (i.e., not A). The whole sample space corresponds to an event that will occur in every experiment, which we call E. In the rest of this chapter we shall define what we mean by the probability of the occurrence of an event and present rules for computations with probabilities.

2.2 The Concept of Probability

Let us consider the simplest experiment, namely, the tossing of a coin. Like the throwing of dice or certain problems with playing cards it is of no practical interest but is useful for didactic purposes. What is the probability that a "fair" coin shows "heads" when tossed once? Our intuition suggests that this probability is equal to $1/2$. It is based on the assumption that all points in sample space (there are only two points: "heads" and "tails") are equally probable and on the convention that we give the event E (here: "heads" or "tails") a probability of unity. This way of determining probabilities can be applied only to symmetric experiments and is therefore of little practical use. (It is, however, of great importance in statistical physics and quantum statistics, where the equal probabilities of all allowed states is an essential postulate of very successful theories.) If no such perfect symmetry exists—which will even be the case with normal "physical" coins—the following procedure seems reasonable. In a large number N of experiments the event A is observed to occur n times. We define

$$P(A) = \lim_{N\to\infty} \frac{n}{N} \tag{2.2.1}$$

as the probability of the occurrence of the event A. This somewhat loose *frequency definition* of probability is sufficient for practical purposes, although it is mathematically unsatisfactory. One of the difficulties with this definition is the need for an infinity of experiments, which are of course impossible to perform and even difficult to imagine. Although we shall in fact use the frequency definition in this book, we will indicate the basic concepts of an axiomatic theory of probability due to KOLMOGOROV [1]. The minimal set of axioms generally used is the following:

(a) To each event A there corresponds a non-negative number, its probability,

$$P(A) \geq 0 \quad . \tag{2.2.2}$$

(b) The event E has unit probability,

$$P(E) = 1 \quad . \tag{2.2.3}$$

(c) If A and B are *mutually exclusive* events, then the probability of A *or* B (written $A+B$) is

$$P(A+B) = P(A) + P(B) \quad . \tag{2.2.4}$$

From these axioms* one obtains immediately the following useful results. From (b) and (c):

$$P(\bar{A}+A) = P(A) + P(\bar{A}) = 1 \quad , \tag{2.2.5}$$

and furthermore with (a):

$$0 \leq P(A) \leq 1 \quad . \tag{2.2.6}$$

From (c) one can easily obtain the more general theorem for mutually exclusive events A, B, C, ...,

$$P(A+B+C+\cdots) = P(A) + P(B) + P(C) + \cdots \quad . \tag{2.2.7}$$

It should be noted that summing the probabilities of events combined with "or" here refers only to mutually exclusive events. If one must deal with events that are not of this type, then they must first be decomposed into mutually exclusive ones. In throwing a die, A may signify even, B odd, C less than 4 dots, D 4 or more dots. Suppose one is interested in the probability for the

*Sometimes the definition (2.3.1) is included as a fourth axiom.

event A *or* C, which are obviously not exclusive. One forms A *and* C (written AC) as well as AD, BC, and BD, which are mutually exclusive, and finds for A or C (sometimes written $A \dot{+} C$) the expression $AC + AD + BC$. Note that the axioms do not prescribe a method for assigning the value of a particular probability $P(A)$.

Finally it should be pointed out that the word probability is often used in common language in a sense that is different or even opposed to that considered by us. This is subjective probability, where the probability of an event is given by the measure of our belief in its occurrence. An example of this is: "The probability that the party A will win the next election is $1/3$." As another example consider the case of a certain track in nuclear emulsion which could have been left by a proton or pion. One often says: "The track was caused by a pion with probability $1/2$." But since the event had already taken place and only one of the two kinds of particle could have caused that particular track, the probability in question is either 0 or 1, but we do not know which.

2.3 Rules of Probability Calculus: Conditional Probability

Suppose the result of an experiment has the property A. We now ask for the probability that it also has the property B, i.e., the probability of B under the condition A. We define this *conditional probability* as

$$P(B|A) = \frac{P(A\,B)}{P(A)} \quad . \tag{2.3.1}$$

It follows that

$$P(A\,B) = P(A)\,P(B|A) \quad . \tag{2.3.2}$$

One can also use (2.3.2) directly for the definition, since here the requirement $P(A) \neq 0$ is not necessary. From Fig. 2.1 it can be seen that this definition is reasonable. Consider the event A to occur if a point is in the region labeled A, and correspondingly for the event (and region) B. For the overlap region both A and B occur, i.e., the event (AB) occurs. Let the area of the different regions be proportional to the probabilities of the corresponding events. Then the probability of B under the condition A is the ratio of the area AB to that of A. In particular this is equal to unity if A is contained in B and zero if the overlapping area vanishes.

Using conditional probability we can now formulate the *rule of total probability*. Consider an experiment that can lead to one of n possible mutually exclusive events,

$$E = A_1 + A_2 + \cdots + A_n \quad . \tag{2.3.3}$$

The probability for the occurrence of any event with the property B is

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i) \quad , \tag{2.3.4}$$

as can be seen easily from (2.3.2) and (2.2.7).

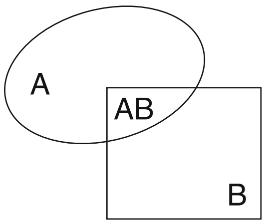

Fig. 2.1: Illustration of conditional probability.

We can now also define the *independence* of events. Two events A and B are said to be independent if the knowledge that A has occurred does not change the probability for B and vice versa, i.e., if

$$P(B|A) = P(B) \quad , \tag{2.3.5}$$

or, by use of (2.3.2),

$$P(A\,B) = P(A) P(B) \quad . \tag{2.3.6}$$

In general several decompositions of the type (2.3.3),

$$\begin{aligned} E &= A_1 + A_2 + \cdots + A_n \quad , \\ E &= B_1 + B_2 + \cdots + B_m \quad , \\ &\vdots \\ E &= Z_1 + Z_2 + \cdots + Z_\ell \quad , \end{aligned} \tag{2.3.7}$$

are said to be independent, if for all possible combinations $\alpha, \beta, \ldots, \omega$ the condition

$$P(A_\alpha B_\beta \cdots Z_\omega) = P(A_\alpha) P(B_\beta) \cdots P(Z_\omega) \tag{2.3.8}$$

is fulfilled.

2.4 Examples

2.4.1 Probability for *n* Dots in the Throwing of Two Dice

If n_1 and n_2 are the number of dots on the individual dice and if $n = n_1 + n_2$, then one has $P(n_i) = 1/6$; $i = 1, 2$; $n_i = 1, 2, \ldots, 6$. Because the two dice are independent of each other one has $P(n_1, n_2) = P(n_1) P(n_2) = 1/36$. By considering in how many different ways the sum $n = n_i + n_j$ can be formed one obtains

$$\begin{aligned}
P_2(2) &= P(1,1) = 1/36 \quad , \\
P_2(3) &= P(1,2) + P(2,1) = 2/36 \quad , \\
P_2(4) &= P(1,3) + P(2,2) + P(3,1) = 3/36 \quad , \\
P_2(5) &= P(1,4) + P(2,3) + P(3,2) + P(4,1) = 4/36 \quad , \\
P_2(6) &= P(1,5) + P(2,4) + P(3,3) + P(4,2) \\
&\quad + P(5,1) = 5/36 \quad , \\
P_2(7) &= P(1,6) + P(2,5) + P(3,4) + P(4,3) \\
&\quad + P(5,2) + P(6,1) = 6/36 \quad , \\
P_2(8) &= P_2(6) = 5/36 \quad , \\
P_2(9) &= P_2(5) = 4/36 \quad , \\
P_2(10) &= P_2(4) = 3/36 \quad , \\
P_2(11) &= P_2(3) = 2/36 \quad , \\
P_2(12) &= P_2(2) = 1/36 \quad .
\end{aligned}$$

Of course, the normalization condition $\sum_{k=2}^{12} P_2(k) = 1$ is fulfilled.

2.4.2 Lottery 6 Out of 49

A container holds 49 balls numbered 1 through 49. During the drawing 6 balls are taken out of the container consecutively and none are put back in. We compute the probabilities $P(1)$, $P(2)$, …, $P(6)$ that a player, who before the drawing has chosen six of the numbers 1, 2, …, 49, has predicted exactly 1, 2, …, or 6 of the drawn numbers.

First we compute $P(6)$. The probability to choose as the first number the one which will also be drawn first is obviously 1/49. If that step was successful, then the probability to choose as the second number the one which is also drawn second is 1/48. We conclude that the probability for choosing six numbers correctly in the order in which they are drawn is

$$\frac{1}{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = \frac{43!}{49!} \quad .$$

The order, however, is irrelevant. Since there are 6! possible ways to arrange six numbers in different orders we have

$$P(6) = \frac{6!43!}{49!} = \frac{1}{\binom{49}{6}} = \frac{1}{C_6^{49}} \quad .$$

That is exactly the inverse of the number of combinations C_6^{49} of 6 elements out of 49 (see Appendix B), since all of these combinations are equally probable but only one of them contains only the drawn numbers.

We may now argue that the container holds two kinds of balls, namely 6 balls in which the player is interested since they carry the numbers which he selected, and 43 balls whose numbers the player did not select. The result of the drawing is a sample from a set of 49 elements of which 6 are of one kind and 43 are of the other. The sample itself contains 6 elements which are drawn without putting elements back into the container. This method of sampling is described by the hypergeometric distribution (see Sect. 5.3). The probability for predicting correctly ℓ out of the 6 drawn numbers is

$$P(\ell) = \frac{\binom{6}{\ell}\binom{43}{6-\ell}}{\binom{49}{6}} \quad , \qquad \ell = 0, \ldots, 6 \quad .$$

2.4.3 Three-Door Game

In a TV game show a candidate is given the following problem. Three rooms are closed by three identical doors. One room contains a luxury car, the other two each contain a goat. The candidate is asked to guess behind which of the doors the car is. He chooses a door which we will call A. The door A, however, remains closed for the moment. Of course, behind at least one of the other doors there is a goat. The quiz master now opens one door which we will call B to reveal a goat. He now gives the candidate the chance to either stay with the original choice A or to choose remaining closed door C. Can the candidate increase his or her chances by choosing C instead of A?

The answer (astonishing for many) is yes. The probability to find the car behind the door A obviously is $P(A) = 1/3$. Then the probability that the car is behind one of the other doors is $P(\bar{A}) = 2/3$. The candidate exhausts this probability fully if he chooses the door C since through the opening of B it is shown to be a door without the car, so that $P(C) = P(\bar{A})$.

http://www.springer.com/978-3-319-03761-5

Data Analysis
Statistical and Computational Methods for Scientists
and Engineers
Brandt, S.
2014, XX, 523 p. 134 illus., Hardcover
ISBN: 978-3-319-03761-5