ELEMENTS OF STATISTICAL THEORY

Primitive Notions of Probability

Modern probability theory is founded on a set of axioms which were first propounded in their definitive form by Kolmogorov (1936). The notions that underlie this axiomatic system can be attributed to two distinct sources. The first and most fruitful source is the deductive analysis of games of chance. We shall use this kind of analysis to outline some of the results in probability theory that we shall later derive from the axioms.

The analysis of games of chance is not a sufficient foundation for a theory of probability, for the reason that it is founded on the concept of equally probable elementary outcomes. An example of such outcomes is provided by the six faces of a die that have an equal chance of falling uppermost when the die is tossed. Another example concerns the equal probabilities of finding any of the 52 cards at the top of a well-shuffled deck.

Such equi-probable events are artificial contrivances for the purposes of gambling, and we should not expect to use them as a basis of a general theory of probability. Nevertheless, attempts have been made, notably by J.M. Keynes (1921) in his *Treatise on Probability,* to build upon such foundations.

The second source of ideas are the simple rules of accountancy that accompany the construction of empirical frequency tables which record and categorise events that can be subjected to statistical analysis. Prime examples are provided by statistics of mortality classified according to time, place and cause.

Attempts have been made to extrapolate from the notion of the relative frequencies of the occurrence of statistical events to an underlying abstract concept of the probability. A notable attempt to define probabilities as the limits of relative frequencies was made by Richard von Mises (1928) in his book *Probability Statistics and Truth.* However, the modern consensus is that such endeavours are flawed and that they are bound to be afflicted by circular reasoning. It seems that, in order to give a meaningful definition of the limit of a relative frequency, we must invoke the very notions of probability which it is our purpose to define.

The Natural Laws of Probability

In tossing a single die, we might seek to determine the probability of getting either a 4, which is the event A_4, or a 6, which is the event A_6. The question concerns the compound event $A_4 \cup A_6$. This is read as "either the event A_4 or the event A_6 or both A_4 and A_6".

Of course, we know that the two events cannot occur together: they are mutually exclusive, which is to say that $A_4 \cap A_6 = \emptyset$ is the null set or empty set. The constituent events A_4, A_6 each have a probability of 1/6 of occurring; for, unless we know that the die is loaded, we can only assume that the six faces of the die have equal probabilities of falling uppermost. Therefore, the answer to the question is given by

(1) $$P(A_4 \cup A_6) = P(A_4) + P(A_6) = \frac{2}{6}.$$

The principle at issue is the following:

(2) *The probability of the union of a set of mutually exclusive events is the sum of their separate probabilities.*

In this example of the tossing of a die, the six possible outcomes $A_1, \ldots, A_6$ are exhaustive as well as mutually exclusive. The set of all possible outcomes is the so-called sample space $\Omega = A_1 \cup \cdots \cup A_6$. Applying the result from (1) to this set gives

(3) $$P(\Omega) = P(A_1 \cup \cdots \cup A_6) = P(A_1) + \cdots + P(A_6) = 1.$$

Here the definition that $P(\Omega) = 1$ has been inveigled. It has the following meaning:

(4) *The probability of a certainty is unity.*

Now consider tossing a red die and a blue die together. For the red die, there are six possible outcomes which form the set $\{A_1, A_2, \ldots, A_6\}$. Each outcome gives rise to a distinct value of the random variable $x \in \{1, 2, \ldots, 6\}$, the value being obtained by counting the number of spots on the uppermost face. For the blue die, the set of outcomes is $\{B_1, B_2, \ldots, B_6\}$ and the corresponding random variable is $y \in \{1, 2, \ldots, 6\}$. A further set of events can be defined which correspond to the sum of the scores on the blue die and the red die. This is $\{C_k; k = 2, 3, \ldots, 12\}$ corresponding to the values of $x + y = z \in \{2, 3, \ldots, 12\}$.

Table 1. The outcomes from tossing a red die and a blue die, highlighting the outcomes for which the joint store is 5.

	1	*2*	*3*	*4*	*5*	*6*
1	2	3	4	(**5**)	7	7
2	3	4	(**5**)	6	7	8
3	4	(**5**)	6	7	8	9
4	(**5**)	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

In Table 1 above, the italic figures in the horizontal margin represent the scores x on the red die whilst the italic figures in the vertical margin represent the scores y on the blue die. The sums $x + y = z$ of the scores are the figures within the body of the table. The question to be answered is what is the probability that the joint score from tossing the red and the blue die together will be $z = 5$? The event that gives this score can be denoted by

$$C_5 = \Big\{(A_1 \cap B_4) \cup (A_2 \cap B_3) \cup (A_3 \cap B_2) \cup (A_4 \cap B_1)\Big\}. \tag{5}$$

But this is a union of a set of mutually exclusive events that are represented in the table by the cells labelled (**5**). By applying the probability law of (2), it is found that

$$P(C_5) = P(A_1 \cap B_4) + P(A_2 \cap B_3) + P(A_3 \cap B_2) + P(A_4 \cap B_1). \tag{6}$$

To find the value of this expression, it is necessary to evaluate the probability of the various constituent events $A_i \cap B_j$ for which $i + j = 5$. But A_i and B_j are statistically independent events such that the outcome of one does not affect the outcome of the other. Therefore

$$P(A_i \cap B_j) = P(A_i) \times P(B_j) = \frac{1}{36} \quad \text{for all} \quad i, j; \tag{7}$$

and it follows that

$$P(C_5) = P(A_1)P(B_4) + P(A_2)P(B_3) + P(A_3)P(B_2) + P(A_4)P(B_1) = \frac{4}{36}. \tag{8}$$

The principle that has been invoked in solving this problem, which has provided the probabilities of the events $A_i \cap B_j$, is the following:

(9) *The probability of the joint occurrence of two statistically independent events is the product of their individual or marginal probabilities. Thus, if A and B are statistically independent, then $P(A \cap B) = P(A) \times P(B)$.*

Now let us define a fourth set of events $\{D_k; k = 2, 3, \ldots, 12\}$ corresponding to the values of $x + y = z \geq k \in \{2, 3, \ldots, 12\}$. For example, there is the event

$$D_8 = C_8 \cup C_9 \cup \cdots \cup C_{12}, \tag{10}$$

whereby the joined score equals of exceeds eight, which has the probability

$$P(D_8) = P(C_8) + P(C_9) + \cdots + P(C_{12}) = \frac{15}{36}. \tag{11}$$

Table 2. The outcomes from tossing a red die and a blue die, highlighting the outcomes for which the score on the red die is 4, by the numbers enclosed by brackets [,] as well as the outcomes for which the joint score exceeds 7, by the numbers enclosed by parentheses (,).

	1	*2*	*3*	*4*	*5*	*6*
1	2	3	4	[**5**]	7	7
2	3	4	5	[**6**]	7	(**8**)
3	4	5	6	[**7**]	(**8**)	(**9**)
4	**5**	**6**	$\underline{7}$	[(**8**)]	(**9**)	(**10**)
5	6	7	(**8**)	[(**9**)]	(**10**)	(**11**)
6	7	(**8**)	(**9**)	[(**10**)]	(**11**)	(**12**)

In Table 2, the event D_8 corresponds to the set of cells in the lower triangle that bear numbers in boldface surrounded by parentheses. The question to be asked is what is the value of the probability $P(D_8|A_4)$ that the event D_8 will occur when the event A_4 is already know to have occurred? Equally, we are seeking the probability that $x + y \geq 8$ given that $x = 4$.

The question concerns the event $D_8 \cap A_4$; and therefore one can begin by noting that $P(D_8 \cap A_4) = 3/36$. But it is not intended to consider this event within the entire sample space $\Omega = \{A_i \cap B_j; i, j = 1, \ldots, 6\}$, which is the set of all possible outcomes—the event is to be considered only within the narrower context of A_4, as a sub event or constituent event of the latter.

Since the occurrence of A_4 is now a certainty, its probability, according to (4), has a value of unity; and the probabilities of its constituent events must sum to unity, given that they are mutually exclusive and exhaustive. This can be achieved by re-scaling the probabilities in question, and the appropriate scaling factor is $1/P(A_4)$. Thus the conditional probability that is sought is given by

(12) $$P(D_8|A_4) = \frac{P(D_8 \cap A_4)}{P(A_4)} = \frac{3/36}{1/6} = \frac{1}{2}.$$

By this form of reasoning, we can arrive at the following law of probability:

(13) *The conditional probability of the occurrence of the event A given that the event B has occurred is the probability of their joint occurrence divided by the probability of B. Thus*

$$P(A|B) = P(A \cap B)/P(B).$$

To elicit the final law of probability, we shall consider the probability of the event $A_4 \cup D_8$, which is the probability of getting $x = 4$ for the score on the red die or of getting $x + y \geq 8$ for the joint score, or of getting both of these outcomes at the same time. Since $A_4 \cap D_8 \neq \emptyset$, the law of probability under (2), concerning mutually exclusive outcomes, cannot be invoked directly. It would lead to the double counting of those events that are indicated in Table 2 by the cells bearing numbers that are surrounded both by brackets and by parentheses. Thus $P(A_4 \cup D_8) \neq P(A_4) + P(D_8)$. The avoidance of double counting leads to the formula

$$\begin{aligned} P(A_4 \cup D_8) &= P(A_4) + P(D_8) - P(A_4 \cap D_8) \\ &= \frac{6}{36} + \frac{15}{36} - \frac{3}{36} = \frac{1}{2}. \end{aligned} \tag{14}$$

By this form of reasoning, we can arrive at the following law of probability:

(15) *The probability that either of events A and B will occur, or that both of them will occur, is equal to the sum of their separate probabilities less the probability or their joint occurrence. Thus*

$$P(A \cup B) = P(A) + P(B) - P(A \cup B).$$

The Calculus of Events

The experiments of the previous section have finite numbers of elementary outcomes or sample points. There are six outcomes from tossing a single die and there are 36 outcomes from tossing a pair of dice, which give rise to the individual cells of Tables 1 and 2. These outcomes have also been described as events, for the reason that they have each been recognised in their own right.

An event may comprise several outcomes which are not recognised individually. An example is provided by the event that the two dice have different numbers of spots on the uppermost faces. One can recognise the difference without counting the spots. The same is true of the complementary event that the dice have the same number of spots on their uppermost faces.

In order to extend the model of probability beyond the case of equiprobable outcomes, and to provide a axiomatic basis, there are two steps that must be accomplished. The first is to establish the rules by which statistical events are constructed from the elementary outcomes of an experiment, and the second is to state the rules by which probabilities are assigned to such events.

It transpires that the first step is by far the more difficult, albeit that the difficulties arise only when there are allowed to be an infinite number of outcomes within the sample space Ω. To cope with these difficulties, a burden of mathematical formalities must be carried which can be negelected when dealing with experiments that have finite numbers of outcomes. Such formalities are usually overlooked in elementary accounts of probability theory. We shall being by refining our account of finite experiments.

The set of individual outcomes of a finite experiment, which together constitute the sample space, may be denoted by $\{\omega_1, \ldots, \omega_n\} = \Omega$. The events are subsets of the sample space. There is no presumption that all subsets of Ω constitute events. However, the set $\mathcal{F}$ of all of the events that are recognised must constitute a field which is subject to the following conditions:

$$\begin{aligned} &\text{(a)} \quad \text{the sample space} \quad \Omega \quad \text{is included} \quad \mathcal{F}, \\ &\text{(b)} \quad \text{if} \quad A, B \in \mathcal{F} \quad \text{then} \quad A \cup B \in \mathcal{F}, \\ &\text{(c)} \quad \text{if} \quad A \in \mathcal{F} \quad \text{then} \quad A^c \in \mathcal{F}. \end{aligned} \tag{16}$$

The complement of the sample space Ω is the empty set $\emptyset$; and, according to (c), this must also belong to $\mathcal{F}$. There is no mention here of the condition that if $A, B \in \mathcal{F}$ then $A \cap B \in \mathcal{F}$. In fact, this is an implication of the conditions (b) and (c). For, if $A, B \in \mathcal{F}$, then (c) implies that $A^c, B^c \in \mathcal{F}$ whence $A^c \cup B^c \in \mathcal{F}$ and $(A^c \cup B^c)^c = A \cap B \in \mathcal{F}$, where the equality follows from De Morgan's rule. It follows from the properties of a field that

$$\text{if} \quad A_1, \ldots, A_n \in \mathcal{F} \quad \text{then} \quad \bigcup_{i=1}^{n} A_i \in \mathcal{F}. \tag{17}$$

This indicates that the field $\mathcal{F}$ is closed under finite unions and hence, according to the preceding argument, it is closed under finite intersections.

Example. The set of all subsets of Ω, described as the power set, clearly constitutes a field. Even when n is a small number, the number $N = \sum_{i=0}^{n} {}^nC_i = 2^n$ of the elements of the power set can be large.

The number N is derived by considering the numbers of ways of selecting i elements from n when i ranges from 1 to n. These numbers are the coefficients of the binomial expansion $(a+b)^n = \sum_{i=1}^{n} {}^nC_i a^n b^{n-1}$. Setting $a = b = 1$ gives $N = 2^n$. In the case of a single toss of a die, there is $N = 2^6 = 64$. When a pair of dice are thrown together there is $N = 2^{36} = 68,719,476,736$.

Since the power set comprises all conceivable events, it is, in one sense, the fundamental event set. However, it may be far too large and unwieldy for the purpose at hand. Consider the example where the event of interest, denoted A, is that the uppermost faces of two dice that are thrown together have equal numbers of spots. Then the relevant event set is $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$, which has only four members. It can be confirmed, with reference to Tables 1 and 2, that $P(A) = 1/6$ and $P(A^c) = 5/6$.

Now let us consider the case where the sample space Ω comprises an infinite number of outcomes. A leading example concerns the experiment of tossing a coin repeatedly until the first head turns up. One would expect this to occur within a small number of trials; but there is no reason to exclude the possibility the one might have to continue tossing the coin indefinitely. The set of all possible outcome is therefore the infinite set $\Omega = \{\omega_1, \omega_2, \ldots\}$ wherein ω_i denotes the outcome that the first $i - 1$ tosses are tails and the ith toss is a head.

Within this context, we might seek to assign a probability to the event $A = \{\omega_2 \cup \omega_2 \cup \cdots\}$ that the first head occurs after an even number of tosses. This event represents the union of a countable infinity of the elementary outcomes within Ω. In order to construct a model which accommodates such events, it is necessary to extend the concept of a field.

The necessary extension, which is known as a σ-field, satisfies following condition in addition to the conditions of (16):

$$\text{(18)} \qquad \text{if} \quad A_1, A_2, \ldots \in \mathcal{F} \quad \text{then} \quad \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

The conditions defining a σ-field also imply that the field is closed in respect of countably infinite intersections of its members. The objective of calculating the probability of events that are represented by countably infinite intersections or unions is often rendered practicable through the availability of analytic expressions for power series.

Example. As in the case of a finite experiment, one might be tempted to regard the power set of $\Omega = \{\omega_1, \omega_2, \ldots\}$ as the fundamental event set. However, it transpires that the power set has uncountably many elements; and this is too large to allow probabilities to be assigned to all of its members.

The power set contains all possible sets that are selections of the elements of Ω. Any such selection A corresponds to a decimal binary number $0.d_1 d_2 \cdots = \sum_{i=1}^{\infty} d_i 2^{-i}$ in the unit interval $[0, 1)$ on the real line. The correspondence can be established by the following rule. If $\omega_i \in A$, then the ith binary digit of this number takes the value $d_i = 1$. Otherwise it takes the value $d_i = 0$. Conversely, any binary digit in $[0, 1)$ corresponds to a selection of subsets from Ω. The rule it to include ω_i in the set A if $d_i = 1$ and to exclude it if $d_1 = 0$. Since the unit interval contains an uncountable infinity of elements, so too does the power set.