ELEMENTARY PROBABILITY

Events and event sets. Consider tossing a die. There are six possible outcomes, which we shall denote by elements of the set $\{A_i; i = 1, 2, \ldots, 6\}$. A numerical value is assigned to each outcome or event:

$$\begin{aligned} A_1 &\longrightarrow 1, \\ A_2 &\longrightarrow 2, \\ &\vdots \\ A_6 &\longrightarrow 6. \end{aligned}$$

The variable $x = 1, 2, \ldots, 6$ that is associated with these outcomes is called a *random variable.*

The set of all possible outcomes is $S = \{A_1 \cup A_2 \cup \cdots \cup A_6\}$ is called the *sample space.* This is distinct from the elementary event set. For example, we can define the event $E \subset S$ to be a matter of getting an even number from tossing the die. We may denote this event by $E = \{A_2 \cup A_4 \cup A_6\}$, which is the event of getting a 2 or a 4 or a 6. These elementary events that constitute E are, of course, mutually exclusive.

The collectison of subsets of S that can be formed by taking unions of the events within S, and also by taking intersections of these unions, is described as a *sigma field,* or a σ-field, and there is also talk of a *sigma algebra.*

The random variable x is a described as a *scalar-valued set function* defined over the sample space S. There are numerous other random variables that we might care to define in respect of the sample space by identifying other subsets and assigning numerical values to the corresponding events. We shall do so presently.

Summary measures of a statistical experiment. Let us toss the die 30 times and let us record the value assumed by the random variable at each toss:

$$1, 2, 5, 3, \ldots, 4, 6, 2, 1.$$

To summarise this information, we may construct a *frequency table:*

x	f	r
1	8	8/30
2	7	7/30
3	5	5/30
4	5	5/30
5	3	3/30
6	2	2/30
	30	1

Here,

$$\begin{aligned} f_i &= \text{frequency}, \\ n &= \sum f_i = \text{sample size}, \\ r_i &= \frac{f_i}{n} = \text{relative frequency}. \end{aligned}$$

In this case, the order in which the numbers occur is of no interest; and, therefore, there is no loss of information in creating this summary.

We can describe the outcome of the experiment more economically by calculating various summary statistics.

First, there is the *mean* of the sample

$$\bar{x} = \frac{\sum x_i f_i}{n} = \sum x_i r_i.$$

The *variance* is a measure of the dispersion of the sample relative to the mean, and it is the average of the squared deviations. It is defined by

$$\begin{aligned} s^2 &= \sum \frac{(x_i - \bar{x})^2 f_i}{n} = \sum (x_i - \bar{x})^2 r_i \\ &= \sum (x_i^2 - x_i\bar{x} - \bar{x}x_i + \{\bar{x}\}^2) r_i \\ &= \sum x_i^2 r_i - \{\bar{x}\}^2, \end{aligned}$$

which follows since $\bar{x}$ is a constant that is not amenable to the averaging operation.

Probability and the limit of relative frequency. Let us consider an indefinite sequence of tosses. We should observe that, as $n \to \infty$, the relative frequencies will converge to values in the near neighbourhood of 1/6. We will recognise that the value $p_i = 1/6$ is the probability of the occurrence of a particular number x_i in the set $\{x_i; i = 1, \ldots, 6\} = \{1, 2, \ldots, 6\}$. We are tempted to define probability simply as the limit of relative frequency.

It has been recognised that this is an inappropriate way of defining probabilities. We cannot regard such a limit of relative frequency in the same light as an ordinary mathematical limit. We cannot say, as we can for a mathematical limit, that, when n exceeds a certain number, the relative frequency r_i is bound thereafter to reside within a limited neighbourhood of p_i. (Such a neighbourhood is denoted by the interval $(p + \epsilon_n, p - \epsilon_n)$, where ϵ_n is a number that gets smaller as n increases.)

The reason is that there is always a wayward chance that, by a run of aberrant outcomes, the value of r_i will break the bounds of the neighbourhood. All that we can say is that the probability of its doing so becomes vanishingly small as $n \to \infty$. But, in making this statement, we have invoked a notion of probability, which is the very thing that we have intended to define. Clearly, such a limiting process cannot serve to define probabilities.

We shall continue, for the present, to assume that we understand unquestionably that the probability of any number x_i arising from the toss of a fair die is $p_i = 1/6$ for all i. Indeed, this constitutes our definition of a fair die. We can replace all of the relative frequencies in the formulae above by the corresponding probabilities to derive a new set of measures that characterise the idealised notion of the long run tendencies of these statistics.

Population measures. We define the mathematical expectation or mean of the random variable associated with a toss of the die to be the number

$$E(x) = \mu = \sum x_i p_i = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5.$$

Likewise, we may define

$$\begin{aligned}V(x) = \sigma^2 &= \sum(\{x_i - E(x)\}^2 p_i \\ &= \sum(x_i^2 - x_i E(x) - E(x)x_i + \{E(x)\}^2)p_i \\ &= \sum x_i^2 p_i - \{E(x)\}^2,\end{aligned}$$

which we may call the true variance of the population variance. We may calculate that

$$V(x) = 13.125.$$

Random sampling. By calling this the population variance, we are invoking a concept of an ideal population in which the relative frequencies are precisely equal to the probabilities that we have defined. In that case, the sample experiment that we described at the beginning is a matter of selecting elements of the population at random, of recording their values and, thereafter, of returning them to the population to take their place amongst the other elements that have an equal chance of being selected. In effect, we are providing a definition of *random sampling.*

The union of mutually exclusive events. Having settled our understanding of probability, at least within the present context, let us return to the business of defining compound events.

We might ask ourselves, what is the probability of getting either a 2 or a 4 or a 6 in a single toss of the die? We can denote the event in question by $A_2 \cup A_4 \cup A_6$. The corresponding probability is

$$P(A_2 \cup A_4 \cup A_6) = P(A_2) + P(A_4) + P(A_6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

The result depends on the following principle:

> *The probability of the union of mutually exclusive events is the sum of their separate probabilities.*

Observe that $\cup$ (union) means *OR* in the inclusive sense. That is to say, $A_1 \cup A_2$ means A_1 *OR* A_2 *OR BOTH.* In this case, *BOTH* is not possible, since the events are mutually exclusive.

By extending the argument, we can see that

$$P(A_1 \cup A_2 \cup \cdots \cup A_6) = 1,$$

which is to say that the probability of a certainty is unity.

The intersection of events, or their joint or sequential occurrence. Let us toss a blue and a red die together or in sequence. For the red die, there are the outcomes $\{A_i; i = 1, 2, \ldots, 6\}$ and the corresponding numbers $\{x_i; i = 1, \ldots, 6\} = \{1, 2, \ldots, 6\}$. For the blue die, there are the outcomes $\{B_j; i = 1, 2, \ldots, 6\}$ and the corresponding numbers $\{y_j; i = 1, \ldots, 6\} = \{1, 2, \ldots, 6\}$. Also, we may define a further set of events $\{C_k; k = 2, 3, \ldots, 12\}$ and the corresponding numbers $\{z_{k=i+j} = x_i + y_j\} = \{2, 3, \ldots, 12\}$.

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

	1	*2*	*3*	*4*	*5*	*6*
1	2	3	4	(**5**)	6	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 of the event C_5, which is composed of the set of all outcomes for which the combined score is 5. The event is defined by

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

Observe that $\cap$ (intersection) means *AND*. That is to say $(A_1 \cap B_4)$ means that *BOTH* A_1 *AND* B_4 occur, either together or in sequence.

By applying the first rule of probability, we get

$$P(C_5) = P(A_1 \cap B_4) + P(A_2 \cap B_3) + P(A_3 \cap B_2) + P(A_4 \cap B_1)\},$$

which follows from the fact that, here, we have a union of mutually exclusive events. But A_i and B_j are independent events for all i, j, such that the outcome of one does not affect the outcome of the other. Therefore,

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

and it follows that

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

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:

> *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* $P(A \cap B) = P(A) \times P(B)$.

Conditional probabilities. 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

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

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

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

Table 2. The outcomes from tossing a red die and a blue die, indicating, by the numbers enclosed by brackets [,], the outcomes for which the score on the red die is 4, 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**]	6	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 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 $S = \{A_i \cap B_j; i, j = 1, \ldots, 6\}$, 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 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

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

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

> *The conditional probability of the occurrence of the event A given that the event B has occurred is given by the probability of their joint occurrence divided by the probability of B. Thus $P(A|B) = P(A \cap B)/P(B)$.*

The joint occurence of non-exclusive events. 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 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

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

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

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