PART ONE

BASIC PROBABILITY

CAMBRIDGE

Cambridge University Press
978-1-107-60358-5 - Probability and Statistics by Example: I: Basic Probability and Statistics: Second Edition
Yuri Suhov and Mark Kelbert
Excerpt
More information

1

Discrete outcomes

1.1 A uniform distribution

Lest men suspect your tale untrue,
Keep probability in view.
J. Gay (1685–1732), English poet.

In this section we use the simplest (and historically the earliest) probabilistic model where there are a finite number m of possibilities (often called *outcomes*) and each of them has the same probability $1/m$. A collection A of k outcomes with $k \leq m$ is called an *event* and its probability $\mathbb{P}(A)$ is calculated as k/m:

$$\mathbb{P}(A) = \frac{\text{the number of outcomes in } A}{\text{the total number of outcomes}}. \tag{1.1.1}$$

An empty collection has probability zero and the whole collection one. This scheme looks deceptively simple: in reality, calculating the number of outcomes in a given event (or indeed, the total number of outcomes) may be tricky.

Worked Example 1.1.1 You and I play a coin-tossing game: if the coin falls heads I score one, if tails you score one. In the beginning, the score is zero. (i) What is the probability that after $2n$ throws our scores are equal? (ii) What is the probability that after $2n+1$ throws my score is three more than yours?

Solution The outcomes in (i) are all sequences

$$HHH\dots H, THH\dots H, \dots, TTT\dots T$$

formed by $2n$ subsequent letters H or T (or, 0 and 1). The total number of outcomes is $m = 2^{2n}$, each carries probability $1/2^{2n}$. We are looking for outcomes where the number of Hs equals that of Ts. The number k of such outcomes is $(2n)!/n!n!$ (the number of ways to choose positions for n Hs among $2n$ places available in the sequence). The probability in question is $\dfrac{(2n)!}{n!n!} \times \dfrac{1}{2^{2n}}$.

In (ii), the outcomes are the sequences of length $2n+1$, 2^{2n+1} in total. The probability equals

$$\frac{(2n+1)!}{(n+2)!(n-1)!} \times \frac{1}{2^{2n+1}}. \qquad \square$$

Worked Example 1.1.2 A tennis tournament is organised for 2^n players on a knock-out basis, with n rounds, the last round being the final. Two players are chosen at random. Calculate the probability that they meet: (i) in the first or second round, (ii) in the final or semi-final, and (iii) the probability they do not meet.

Solution The sentence 'Two players are chosen at random' is crucial. For instance, one may think that the choice has been made after the tournament when all results are known. Then there are 2^{n-1} pairs of players meeting in the first round, 2^{n-2} in the second round, two in the semi-final, one in the final and $2^{n-1} + 2^{n-2} + \cdots + 2 + 1 = 2^n - 1$ in all rounds.

The total number of player pairs is $\begin{pmatrix} 2^n \\ 2 \end{pmatrix} = 2^{n-1}(2^n - 1)$. Hence the answers:

$$\text{(i)} \quad \frac{2^{n-1} + 2^{n-2}}{2^{n-1}(2^n - 1)} = \frac{3}{2(2^n - 1)}, \qquad \text{(ii)} \quad \frac{3}{2^{n-1}(2^n - 1)},$$

and

$$\text{(iii)} \quad \frac{2^{n-1}(2^n - 1) - (2^n - 1)}{2^{n-1}(2^n - 1)} = 1 - \frac{1}{2^{n-1}}. \qquad \square$$

Worked Example 1.1.3 There are n people gathered in a room.

(i) What is the probability that two (at least) have the same birthday? Calculate the probability for $n = 22$ and 23.

(ii) What is the probability that at least one has the same birthday as you? What value of n makes it close to $1/2$?

Solution The total number of outcomes is 365^n. In (i), the number of outcomes not in the event is $365 \times 364 \times \cdots \times (365 - n + 1)$. So, the probability that all birthdays are distinct is $\big(365 \times 364 \times \cdots \times (365 - n + 1)\big)/365^n$ and that two or more people have the same birthday

$$1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$

For $n = 22$:

$$1 - \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{344}{365} = 0.4927,$$

and for $n = 23$:

$$1 - \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{343}{365} = 0.5243.$$

In (ii), the number of outcomes not in the event is 364^n and the probability in question $1 - (364/365)^n$. We want it to be near $1/2$, so

$$\left(\frac{364}{365}\right)^n \approx \frac{1}{2}, \quad \text{i.e. } n \approx -\frac{1}{\log_2(364/365)} \approx 252.61. \qquad \square$$

Worked Example 1.1.4 Mary tosses $n+1$ coins and John tosses n coins. What is the probability that Mary gets more heads than John?

Solution We must assume that all coins are unbiased (as it was not specified otherwise). Mary has 2^{n+1} outcomes (all possible sequences of heads and tails) and John 2^n; jointly 2^{2n+1} outcomes that are equally likely. Let $H_{\rm M}$ and $T_{\rm M}$ be the number of Mary's heads and tails and $H_{\rm J}$ and $T_{\rm J}$ John's, then $H_{\rm M} + T_{\rm M} = n + 1$ and $H_{\rm J} + T_{\rm J} = n$. The events $\{H_{\rm M} > H_{\rm J}\}$ and $\{T_{\rm M} > T_{\rm J}\}$ have the same number of outcomes, thus $\mathbb{P}(H_{\rm M} > H_{\rm J}) = \mathbb{P}(T_{\rm M} > T_{\rm J})$.

On the other hand, $H_{\rm M} > H_{\rm J}$ if and only if $n - H_{\rm M} < n - H_{\rm J}$, i.e. $T_{\rm M} - 1 < T_{\rm J}$ or $T_{\rm M} \leq T_{\rm J}$. So event $H_{\rm M} > H_{\rm J}$ is the same as $T_{\rm M} \leq T_{\rm J}$, and $\mathbb{P}(T_{\rm M} \leq T_{\rm J}) = \mathbb{P}(H_{\rm M} > H_{\rm J})$.

But for any (joint) outcome, either $T_{\rm M} > T_{\rm J}$ or $T_{\rm M} \leq T_{\rm J}$, i.e. the number of outcomes in $\{T_{\rm M} > T_{\rm J}\}$ equals 2^{2n+1} minus that in $\{T_{\rm M} \leq T_{\rm J}\}$. Therefore, $\mathbb{P}(T_{\rm M} > T_{\rm J}) = 1 - \mathbb{P}(T_{\rm M} \leq T_{\rm J})$. To summarise:

$$\mathbb{P}(H_{\rm M} > H_{\rm J}) = \mathbb{P}(T_{\rm M} > T_{\rm J}) = 1 - \mathbb{P}(T_{\rm M} \leq T_{\rm J}) = 1 - \mathbb{P}(H_{\rm M} > H_{\rm J}),$$

whence $\mathbb{P}(H_{\rm M} > H_{\rm J}) = 1/2$. $\square$

Solution Suppose that the final toss belongs to Mary. Let x be the probability that Mary's number of heads equals John's number of heads just before the final toss. By the symmetry, the probability that Mary's number of heads exceeds that of John just before the final toss is $(1-x)/2$. This implies that the Mary's number of heads exceeds that of John by the end of the game equals $(1-x)/2 + x/2 = 1/2$. $\square$

Solution By the end of the game Mary has either more heads or more tails than John because she has more tosses. These two cases exclude each other. Hence, the probability of each case is 1/2 by the symmetry argument. $\square$

Worked Example 1.1.5 You throw $6n$ six-sided dice at random. Show that the probability that each number appears exactly n times is

$$\frac{(6n)!}{(n!)^6}\left(\frac{1}{6}\right)^{6n}.$$

Solution There are 6^{6n} outcomes in total (six for each die), each has probability $1/6^{6n}$. We want n dice to show one dot, n two, and so forth. The number of such outcomes is counted by fixing first which dice show one: $(6n)!/[n!(5n)!]$. Given n dice showing one, we fix which remaining dice show two: $(5n)!/[n!(4n)!]$, etc. The total number of desired outcomes is the product that equals $(6n)!(n!)^6$. This gives the answer. $\square$

In many problems, it is crucial to be able to spot recursive equations relating the cardinality of various events. For example, for the number f_n of ways of tossing a coin n times so that successive tails never appear: $f_n = f_{n-1} + f_{n-2}$, $n \geq 3$ (a Fibonacci equation).

Worked Example 1.1.6 (i) Determine the number g_n of ways of tossing a coin n times so that the combination HT never appears. (ii) Show that $f_n = f_{n-1} + f_{n-2} + f_{n-3}$, $n \geq 3$, is the equation for the number of ways of tossing a coin n times so that three successive heads never appear.

Solution (i) $g_n = 1 + n$; 1 for the sequence $HH\dots H$, n for the sequences $T\dots TH\dots H$ (which includes $T\dots T$).

(ii) The outcomes are 2^n sequences $(y_1, \dots, y_n)$ of H and T. Let A_n be the event {no three successive heads appeared after n tosses}, then f_n is the cardinality $\#A_n$. Split: $A_n = B_n^{(1)} \cup B_n^{(2)} \cup B_n^{(3)}$, where $B_n^{(1)}$ is the event {no three successive heads appeared after n tosses, and the last toss was a tail}, $B_n^{(2)} =$ {no three successive heads appeared after n tosses, and the last two tosses were TH} and $B_n^{(3)} =$ {no three successive heads appeared after n tosses, and the last three tosses were THH}.

Clearly, $B_n^{(i)} \cap B_n^{(j)} = \emptyset$, $1 \leq i \neq j \leq 3$, and so $f_n = \#B_n^{(1)} + \#B_n^{(2)} + \#B_n^{(3)}$.

Now drop the last digit y_n: $(y_1, \dots y_n) \in B_n^{(1)}$ if and only if $y_n = T$, $(y_1, \dots y_{n-1}) \in A_{n-1}$, i.e. $\# B_n^{(1)} = f_{n-1}$. Also, $(y_1, \dots y_n) \in B_n^{(2)}$ if and only if $y_{n-1} = T$, $y_n = H$, and $(y_1, \dots y_{n-2}) \in A_{n-2}$. This allows us to drop the two last digits, yielding $\# B_n^{(2)} = f_{n-2}$. Similarly, $\# B_n^{(3)} = f_{n-3}$. The equation then follows. $\square$

Worked Example 1.1.7 In a Cambridge cinema n people sit at random in the first row. The row has $N \geq n$ seats. Find the probability of the following events:

(a) that no two people sit next to each other;
(b) that each person has exactly one neighbour; and
(c) that, for every pair of distinct seats symmetric relative to the middle of the row, at least one seat from the pair is vacant.

Now assume that n people sit at random in the two first rows of the same cinema, with $2N \geq n$. Find the probability of the following events:

(d) that at least in one row no two people sit next to each other;
(e) that in the first row no two people sit next to each other and in the second row each person has exactly one neighbour; and
(f) that, for every pair of distinct seats in the second row, symmetric relative to the middle of the row, at least one seat from the pair is vacant.

In parts (d)–(f) *you may find it helpful to use indicator functions specifying limits of summation.*

Solution We assume that $n \geq 1$. In parts (a)–(c), the total number of outcomes equals $\binom{N}{n}$, and all of them have the same probability. (All people are indistinguishable.) Then:

(a) The answer is $\binom{N-n+1}{n} \Big/ \binom{N}{n}$ if $N \geq 2n - 1$ and 0 if $N < 2n - 1$. In fact, to place n people in N seats so that no two of them sit next to each other, we scan the row from left to right (say) and affiliate, with each of n seats taken, an empty seat positioned to the right. Place an extra empty seat to the right of the person in the position to the right end of the row. This leaves $N - n + 1$ virtual positions where we should place n objects. The objects are empty seats to the right of occupied ones.

(b) First, assume $n = 2l$ is even. Then we have $n/2 = l$ pairs of neighbouring occupied seats, and with each of $n/2$ of them we again affiliate an empty seat to the right. Thus the answer is $\binom{N-n+1}{n/2} \Big/ \binom{N}{n}$ if $N \geq 3n/2 - 1$ and 0 if $N < 3n/2 - 1$.

If n is odd, the probability in question equals 0.

Cambridge University Press
978-1-107-60358-5 - Probability and Statistics by Example: I: Basic Probability and Statistics: Second Edition
Yuri Suhov and Mark Kelbert
Excerpt
More information

(c) First, assume that N is even. Then, if we require that all people sit in the left-hand side of the row, the number of outcomes is $\binom{N/2}{n}$. Now, for each person we have 2 symmetric choices. Hence, the answer is $2^n\binom{N/2}{n}\Big/\binom{N}{n}$ if $N \geq 2n$ and 0 if $N < 2n$.

When N is odd, there is a seat in the middle: it can be taken or vacant. Thus, the answer in this case is:

$$\begin{array}{ll} \left[2^n\binom{(N-1)/2}{n} + 2^{n-1}\binom{(N-1)/2}{n-1}\right]\Big/\binom{N}{n} & \text{if } N \geq 2n+1, \\ 2^{n-1}\binom{(N-1)/2}{n-1}\Big/\binom{N}{n} & \text{if } N = 2n-1, \\ 0 & \text{if } N < 2n-1. \end{array}$$

In parts (d)–(f), we have in total

$$\binom{2N}{n} = \sum_{0\leq k\leq n} \mathbf{1}(n-N \leq k \leq N)\binom{N}{k}\binom{N}{n-k}$$

outcomes, again of equal probability.

(d) The answer is therefore

$$\sum_{0\leq k\leq n}\Bigg[2\binom{N-k+1}{k}\binom{N}{n-k}\mathbf{1}(n-N \leq k \leq (N+1)/2)$$
$$-\binom{N-k+1}{k}\binom{N-n+k+1}{n-k}\mathbf{1}(n-(N+1)/2 \leq k \leq (N+1)/2)\Bigg]\Bigg/\binom{2N}{n}.$$

(e) The answer is

$$\sum_{0\leq l\leq n/2}\binom{N-2l+1}{l}\binom{N-n+2l+1}{n-2l}$$
$$\times\,\mathbf{1}(n/2-(N+1)/4 \leq l \leq (N+1)/3)\Bigg/\binom{2N}{n}.$$

(f) For N even, the answer is

$$\sum_{0\leq k\leq n} 2^k\binom{N/2}{k}\binom{N}{n-k}\mathbf{1}(n-N \leq k \leq N/2)\Bigg/\binom{2N}{n}.$$

For N odd, the answer is

$$\sum_{0\leq k\leq n}\Bigg[\mathbf{1}(k \leq (N-1)/2)2^k\binom{(N-1)/2}{k}$$
$$+\,\mathbf{1}(k \leq (N+1)/2)2^{k-1}\binom{(N-1)/2}{k-1}\Bigg]$$
$$\times\binom{N}{n-k}\mathbf{1}(n-N \leq k)\Bigg/\binom{2N}{n}. \qquad \square$$

1.2 Conditional Probabilities. Bayes' Theorem. Independent trials

Probability theory is nothing but common sense reduced to calculation.
P.-S. Laplace (1749–1827), French mathematician.

Clockwork Omega
(From the series *'Movies that never made it to the Big Screen'.*)

From now on we adopt a more general setting: our outcomes do not necessarily have equal probabilities $p_1, \dots, p_m$, with $p_i > 0$ and $p_1 + \cdots + p_m = 1$.

As before, an *event* A is a collection of outcomes (possibly empty); the *probability* $\mathbb{P}(A)$ of event A is now given by

$$\mathbb{P}(A) = \sum_{\text{outcome } i \in A} p_i = \sum_{\text{outcome } i} p_i I(i \in A). \tag{1.2.1}$$

($\mathbb{P}(A) = 0$ for $A = \emptyset$.) Here and below, I stands for the *indicator function*, viz.:

$$I(i \in A) = \begin{cases} 1, & \text{if } i \in A, \\ 0, & \text{otherwise.} \end{cases}$$

The probability of the total set of outcomes is 1. The total set of outcomes is also called the whole, or full, event and is often denoted by Ω, so $\mathbb{P}(\Omega) = 1$. An outcome is often denoted by ω, and if $p(\omega)$ is its probability, then

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) = \sum_{\omega \in \Omega} p(\omega) I(\omega \in A). \tag{1.2.2}$$

As follows from this definition, the probability of the union

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) \tag{1.2.3}$$

for any pair of disjoint events A_1, A_2 (with $A_1 \cap A_2 = \emptyset$). More generally,

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) \tag{1.2.4}$$

for any collection of pair-wise disjoint events (with $A_j \cap A_{j'} = \emptyset$ for all $j \neq j'$). Consequently, (i) the probability $\mathbb{P}(A^{\mathrm{c}})$ of the complement $A^{\mathrm{c}} = \Omega \setminus A$ is $1 - \mathbb{P}(A)$, (ii) if $B \subseteq A$, then $\mathbb{P}(B) \leq \mathbb{P}(A)$ and $\mathbb{P}(A) - \mathbb{P}(B) = \mathbb{P}(A \setminus B)$ and (iii) for a general pair of events A, B: $\mathbb{P}(A \setminus B) = \mathbb{P}\big(A \setminus (A \cap B)\big) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$.

Cambridge University Press
978-1-107-60358-5 - Probability and Statistics by Example: I: Basic Probability and Statistics: Second Edition
Yuri Suhov and Mark Kelbert
Excerpt
More information

Furthermore, for a general (not necessarily disjoint) union:

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} \mathbb{P}(A_i);$$

a more detailed analysis of the probability $\mathbb{P}(\bigcup A_i)$ is provided by the exclusion–inclusion formula (1.3.1); as follows.

Given two events A and B with $\mathbb{P}(B) > 0$, the *conditional probability* $\mathbb{P}(A|B)$ of A given B is defined as the ratio

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{1.2.5}$$

At this stage, the conditional probabilities are important for us because of two formulas. One is the formula of complete probability: if $B_1, \ldots, B_n$ are pair-wise disjoint events partitioning the whole event Ω, i.e. have $B_i \cap B_j = \emptyset$ for $1 \leq i < j \leq n$ and $B_1 \bigcup B_2 \bigcup \cdots \bigcup B_n = \Omega$, and in addition $\mathbb{P}(B_i) > 0$ for $1 \leq i \leq n$, then

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \cdots + \mathbb{P}(A|B_n)\mathbb{P}(B_n). \tag{1.2.6}$$

The proof is straightforward:

$$\mathbb{P}(A) = \sum_{1\leq i\leq n} \mathbb{P}(A \cap B_i) = \sum_{1\leq i\leq n} \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)}\mathbb{P}(B_i) = \sum_{1\leq i\leq n} \mathbb{P}(A|B_i)\mathbb{P}(B_i). \quad \square$$

The point is that often it is conditional probabilities that are given, and we are required to find unconditional ones; also, the formula of complete probability is useful to clarify the nature of (unconditional) probability $\mathbb{P}(A)$. Despite its simple character, this formula is an extremely powerful tool in literally all areas dealing with probabilities. In particular, a large portion of the theory of Markov chains is based on its skilful application.

Representing $\mathbb{P}(A)$ in the form of the right-hand side (RHS) of (1.2.6) is called conditioning (on the collection of events $B_1, \ldots, B_n$).

Another formula is the *Bayes formula* (or *Bayes' Theorem*) named after T. Bayes (1702–1761), an English mathematician and cleric. It states that *under the same assumptions as above, if in addition* $\mathbb{P}(A) > 0$, *then the conditional probability* $\mathbb{P}(B_i|A)$ *can be expressed in terms of probabilities* $\mathbb{P}(B_1), \ldots, \mathbb{P}(B_n)$ *and conditional probabilities* $\mathbb{P}(A|B_1), \ldots, \mathbb{P}(A|B_n)$ *as*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum\limits_{1\leq j\leq n} \mathbb{P}(A|B_j)\mathbb{P}(B_j)}. \tag{1.2.7}$$