1.3. Elementary Probability

Note. In this section we introduce some ideas of probability, where we consider a finite sample space of equally likely outcomes. We give examples which require us to count the outcomes in certain subsets of the sample space. We prove some elementary theorems and introduce conditional probability.

Note. We do not address the meaning of probability here. Instead, we refer to my online notes for Foundations of Probability and Statistics-Calculus Based (STAT 2050) on the Introduction to Chapter 2. Probability for some philosophical and historical introduction. In Section 2.1. Basic Ideas, a more general approach is taken to elementary probability, with the statement of three axioms of probability. In Section 2.3. Conditional Probability and Independence of Foundations of Probability and Statistics-Calculus Based, much of the material of the current section is covered. For a more advanced approach to probability, see my online notes for Intermediate Probability Theory (this would be a senior/graduate cross-listed class, though it is not currently a formal ETSU class). For information on a graduate-level approach to probability theory, see my notes for Measure Theory Based Probability (also not a formal ETSU class).

Note. We informally deal with probability by considering a collection D of equally likely "events" (or "outcomes"). With N as the number of these events that are "noteworthy," we associate the probability of N/D with the collection of noteworthy events.

Example 1.3.1. Suppose two dice are rolled, one red and the other green (this allows us distinguish between the two dice in the following discussion). Figure 1.3.1 gives the $6^2 = 36$ possible outcomes of this experiment (if the dice are fair, then each outcome is equally likely). Let $P(n)$ denote the probability of rolling (a sum of) n with two fair dice.

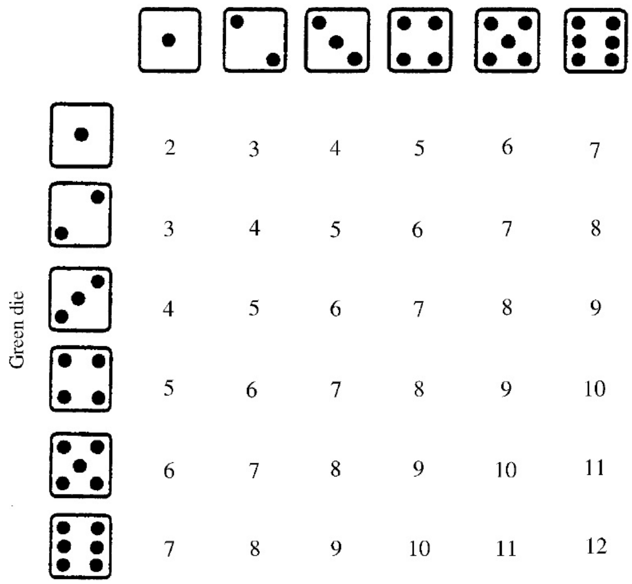

Figure 1.3.1. The 36 outcomes of rolling two dice.

We can read off the probabilities of all possible sums (that is, all noteworthy events) from Figure 1.3.1 as:

$$P(2) = P(12) = 1/36,\ \ P(3) = P(11) = 2/36,\ \ P(4) = P(10) = 3/36,$$

$$P(5) = P(9) = 4/36,\ \ P(6) = P(8) = 5/36,\ \ P(7) = 6/36.$$

Notice that the sum over all possible noteworthy events is 1, as expected.

Example 1.3.3. Suppose a (fair) coin is tossed 10 times. We consider the probability that half of the tosses result in heads (and the other half in tails); this is our "noteworthy event." Notice that there are a total of $D = 2^{10} = 1024$ possible outcomes of this experiment. Now an outcome of this experiment is equivalent to a word of length $n = 10$ on $k = 2$ different letters, say H and T. By Note 1.1.A, the number of such words containing 5 H's and 5 T's is $N = \binom{10}{5,5} = \frac{10!}{5!5!} = 252$. The probability of the noteworthy event is then $N/D = 252/1024 \approx 0.2461$. Similarly, if a coin is tossed n times, the probability that this results in r heads (and $n - r$ tails) is $\binom{n}{r, n-r}/2^n = \frac{n}{r!(n-r)!}/2^n = C(n,r)/2^n$. If we were interested in the probability that in n tosses, at most r heads result, then we would have $[C(n,0) + C(n,1) + \cdots + C(n,r)]/2^n$.

Definition 1.3.4. A nonempty set E of equally likely outcomes is a *sample space*. The number of elements in E is denoted $o(E)$. For any subset A of E, the probability of A is $P(A) = o(A)/o(E)$. If B is a subset of E, then $P(A \text{ or } B) = P(A \cup B)$, and $P(A \text{ and } B) = P(A \cap B)$.

Note. In this class, we are considering only equally likely outcomes, and this could certainly be generalized. Also, the number of elements in sample space E, $o(E)$, is often denoted $|E|$ or $\#E$. Next, we give a relationship between the two probabilities just defined.

Theorem 1.3.5. Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Corollary 1.3.6. Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then $P(A \text{ or } B) \leq P(A) + P(B)$, with equality if and only if A and B are disjoint.

Note/Definition. The *complement* of set A of noteworthy outcomes is the set $A^c = \{x \in E \mid x \notin A\}$. Since $A \cup A^2 = E$ and $A \cap A^c = \varnothing$, then $o(A) + o(A^c) = o(E)$. Therefore

$$\frac{o(A)}{o(E)} + \frac{o(A^c)}{o(E)} = \frac{o(A) + o(A^c)}{o(E)} = \frac{o(E)}{o(E)} = 1 = P(A) + P(A^c).$$

In other words, $p(A^c) = 1 - P(A)$.

Example 1.3.7. Suppose two fair dice are rolled, say a red one and a green one (as in Figure 1.3.1). What is the probability of rolling a 3 on the red die (which we denote $R3$) OR a 4 on the green die (denoted $G4$)? Well, first $P(R3) = P(G2) = 1/6$. When both die are rolled, by Figure 1.3.1, we have that there are $6^2 = 26$ possible equally likely outcomes, only one of which is $R3$ AND $G2$, so that $P(R3 \text{ and } G2) = 1/36$. Then by Theorem 1.3.5,

$$P(R3 \text{ or } G2) = P(R3) + P(G2) - P(R2 \text{ and } G2) = 1/6 + 1/6 - 1/36 = 11/36.$$

Example 1.3.9. Suppose a single (fair) die is rolled twice. What is the probability of getting a 3 OR a 2? Well, first to get *both* a 3 AND a 2, we could either get the 3 first and the 2 second, or the 2 first and the 3 second. Each of these events (from, say, Figure 1.3.1) had probability 1/36 so that $P(3 \text{ and } 2) = 2/36$. Since the probability of getting a 2 on a two rolls is 11/36 (by Example 1.3.7; and similar for getting a 3), then by Theorem 1.3.5 the probability (here, everything is based on two rolls) is

$$P(3 \text{ or } 2) = P(2) + P(3) - P(2 \text{ and } 3) = 11/36 + 11/36 - 2/36 = 5/9.$$

Definition 1.3.10. Let E be a fixed but arbitrary sample space. If A and B are subsets of E, the *conditional probability*,

$$P|A) = \begin{cases} P(B) & \text{if } A = \varnothing, \\ o(A \cap B)/o(A) & \text{otherwise.} \end{cases}$$

Note. When A is not empty (or more generally, when $P(A) \neq 0$), $P(B|A)$ may be viewed as the probability of G given that A is certain (or has already happened). For example, if a card is drawn from a standard deck then the probability that the card is a spade is $P(\spadesuit) = 13/53 = 1/4$. However, if the card drawn is known to be a black suit, then $P(\spadesuit|\text{black suit}) = 13/26 = 12$. Conditional probabilities are related to other probabilities as follows.

Theorem 1.3.11. Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then

$$P(A \text{ and } B) = P(A)P(B|A).$$

Corollary 1.3.12. Bayes' First Rule.
Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then $P(A)P(B|A) = P(B)P(A|B)$.

Note. For an interesting application of Bayes' (more general) Theorem, see my online notes for Mathematical Statistics 1 (STAT 4047/5047) on Section 1.4. Conditional Probability and Independence. In Example 1.4.A of those notes, an example is given concerning the possibility of getting a "false positive" for a rare disease in a medical test that administered to a large population (one could draw parallels between this story and wide-spread drug testing). Those notes also include biographical information on Thomas Bayes (1702-1761). Those notes include a detailed analysis of "The Monty Hall Problem," another application of Bayes' Theorem (see Exercise 1.4.30 in the notes, and the explanation and history following it).

Definition 1.3.13. Suppose E is a fixed but arbitrary sample space. Let A and B be subsets of E. If $P(B|A) = P(B)$, then A and B are *independent*.

Note. If A and B are independent and nonempty, then $P(B|A) = P(B)$, so $P(A)P(B|A) = P(A)P(B)$. By Corollary 1.3.12, $P(A)P(B|A) = P(B)P(A|B)$, so that we have $P(B)P(A|B) = P(A)P(B)$, or $P(A|B) = P(A)$. That is, B and A are independent. So A and B are independent if and only if B and A are independent. When this is the case, we have from Theorem 1.3.11 that $P(A \text{ and } B) = P(A)P(B)$ if and only if A and B are independent.

Example 1.3.15. We now mention a technique of George Pólya (1887–1985; we explore P'olya Theory of enumeration in Chapter 3) that gives an estimate of the number of typographical errors in a manuscript, based on the number of typographical errors found by two editors. Suppose one editor X finds x typographical errors, while a second editor Y finds y errors. Let z be the number of typos discovered by both editors, so that the total number of errors found by editors X and Y are $x + y - z$. If the manuscript, in fact, has t errors, then the empirical probability that X finds a randomly chosen error is $P(X) = x/t$ (and similarly, $P(Y) = y/t$). Now if one of the errors found by Y is chosen at random, then the empirical probability that X found it is $P(X|Y) = z/y$. Assuming that the editors are equally productive in detecting typographical errors, then we should have $z/y \approx x/t$; that is, we should have $P(X|Y) = P(X)$ (the detection of an error by X should be independent of its detection by Y). This then gives $t \approx xy/z$.

Definition. In an experiment with two outcomes, which we label a "success" and a "failure," where the probability of a success is p and the probability of a failure is q (so that $p + q = 1$), is a *binomial experiment*. If the experiment is repeated n, the the probability of r successes is $P(r) = C(n, r)p^r q^{n-r}$ for $0 \leq r \leq n$. This probability function is associated with the *binomial probability distribution*.

Note. The binomial distribution is covered in Foundations of Probability and Statistics-Calculus Based (MATH 2050) in Section 4.2. The Binomial Distribution, and in Mathematical Statistics 1 (STAT 4047/5047) in Section 3.1. The Binomial and Related Distributions.

Revised: 10/14/2022