CHAPTER

SEVENTYTHREE

PROBABILITY

73.1 Prologue

It was a slow day and Bob said he was bored. It was just after lunch, and he complained that there was nothing to do. Nobody really seemed to be listening, although Carlos said that Bob might considered studying, even reading ahead in the chapter. Undeterred, Bob said “Hey Alice, how about we play a game. We could take turns tossing a coin, with the other person calling heads or tails. We could keep score with the first one to a hundred being the winner.” Alice rolled her eyes at such a lame idea. Sensing Alice’s lack of interest, Bob countered “OK, how about a hundred games of Rock, Paper or Scissors?” Xing said “Why play a hundred times? If that’s what you’re going to do, just play a single game.”

Now it was Alice’s turn. “If you want to play a game, I’ve got a good one for you. Just as you wanted, first one to score a hundred wins. You roll a pair of dice. If you roll doubles, I win 2 points. If the two dice have a difference of one, I win 1 point. If the difference is 2, then it’s a tie. If the difference is 3, you win one point; if the difference is 4, you win two points; and if the difference is 5, you win three points. Xing interrupted to say “In other words, if the difference is d, then Bob wins $d - 2$ points.” Alice continues “Right! And there are three ways Bob can win, with one of them being the biggest prize of all. Also, rolling doubles is rare, so this has to be a good game for Bob.”

Dave ears perked up with Alice’s description. He had a gut feeling that this game wasn’t really in Bob’s favor and that Alice knew what the real situation was. Carlos was scribbling on a piece of paper, then said quietly that Bob really should be reading ahead in the chapter.

So what do you think? Is this a fair game? What does it mean for a game to be fair? Should Bob play—independent of the question of whether such silly stuff

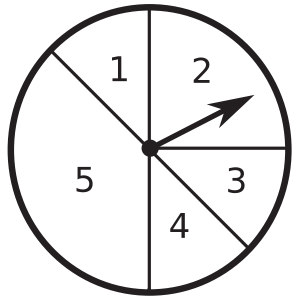

Figure 73.1: A Spinner for Games of Chance

should occupy one's time? And what does any of this conversation have to do with combinatorics?

73.2 An Introduction to Probability

We continue with an informal discussion intended to motivate the more structured development that will follow. Consider the "spinner" shown in Figure 10.1. Suppose we give it a good thwack so that the arrow goes round and round. We then record the number of the region in which the pointer comes to rest. Then observers, none of whom have studied combinatorics, might make the following comments:

1. The odds of landing in region 1 are the same as those for landing in region 3.
2. You are twice as likely to land in region 2 as in region 4.
3. When you land in an odd numbered region, then 60% of the time, it will be in region 5.

We will now develop a more formal framework that will enable us to make such discussions far more precise. We will also see whether Alice is being entirely fair to Bob in her proposed game to one hundred.

We begin by defining a *probability space* as a pair (S, P) where S is a finite set and P is a function that whose domain is the family of all subsets of S and whose range is the set $[0, 1]$ of all real numbers which are non-negative and at most one. Furthermore, the following two key properties must be satisfied:

1. $P(\emptyset) = 0$ and $P(S) = 1$.
2. If A and B are subsets of S, and $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

When (S, P) is a probability space, the function P is called a *probability measure*, the subsets of S are called *events*, and when $E \subseteq S$, the quantity $P(E)$ is referred to as the *probability* of the event E.

Note that we can consider P to be extended to a mapping from S to $[0,1]$ by setting $P(x) = P(\{x\})$ for each element $x \in S$. We call the elements of S *outcomes* (some people prefer to say the elements are *elementary outcomes*) and the quantity $P(x)$ is called the *probability* of x. It is important to realize that if you know $P(x)$ for each $x \in S$, then you can calculate $P(E)$ for any event E, since (by the second property), $P(E) = \sum_{x \in X} P(x)$.

Example 73.1. For the spinner, we can take $S = \{1, 2, 3, 4, 5\}$, with $P(1) = P(3) = P(4) = 1/8$, $P(2) = 2/8 = 1/4$ and $P(5) = 3/8$. So $P(\{2, 3\}) = 1/8 + 2/8 = 3/8$.

Example 73.2. Let S be a finite, nonempty set and let $n = |S|$. For each $E \subseteq S$, set $P(E) = |E|/n$. In particular, $P(x) = 1/n$ for each element $x \in S$. In this trivial example, all outcomes are equally likely.

Example 73.3. If a single six sided die is rolled and the number of dots on the top face is recorded, then the ground set is $S = \{1, 2, 3, 4, 5, 6\}$ and $P(i) = 1/6$ for each $i \in S$. On the other hand, if a pair of dice are rolled and the sum of the dots on the two top faces is recorded, then $S = \{2, 3, 4, \ldots, 11, 12\}$ with $P(2) = P(12) = 1/36$, $P(3) = P(11) = 2/36$, $P(4) = P(10) = 3/36$, $P(5) = P(9) = 4/36$, $P(6) = P(8) = 5/36$ and $P(7) = 6/36$. To see this, consider the two die as distinguished, one die red and the other one blue. Then each of the pairs (i, j) with $1 \leq i, j \leq 6$, the red die showing i spots and the blue die showing j spots is equally likely. So each has probability $1/36$. Then, for example, there are three pairs that yield a total of four, namely $(3, 1)$, $(2, 2)$ and $(1, 3)$. So the probability of rolling a four is $3/36 = 1/12$.

Example 73.4. In Alice's game as described above, the set S can be taken as $\{0, 1, 2, 3, 4, 5\}$, the set of possible differences when a pair of dice are rolled. In this game, we will see that the correct definition of the function P will set $P(0) = 6/36$; $P(1) = 10/36$; $P(2) = 8/36$; $P(3) = 6/36$; $P(4) = 4/36$; and $P(5) = 2/36$. Using Xing's more compact notation, we could say that $P(0) = 1/6$ and $P(d) = 2(6 - d)/36$ when $d > 0$.

Example 73.5. A jar contains twenty marbles, of which six are red, nine are blue and the remaining five are green. Three of the twenty marbles are selected at random.[1] Let $X = \{0, 1, 2, 3, 4, 5\}$, and for each $x \in X$, let $P(x)$ denote the probability that the number of blue marbles among the three marbles selected is x. Then $P(i) = C(9, i)C(11, 3 - i)/C(20, 3)$ for $i = 0, 1, 2, 3$, while $P(4) = P(5) = 0$. Bob says that it doesn't make sense to have outcomes with probability zero, but Carlos says that it does.

Example 73.6. In some cards games, each player receives five cards from a standard deck of 52 cards—four suits (spades, hearts, diamonds and clubs) with 13 cards, ace though king in each suit. A player has a *full house* if there are two values x and y for

[1] This is sometimes called *sampling without replacement*. You should imagine a jar with opaque sides—so you can't see through them. The marbles are stirred/shaken, and you reach into the jar blind folded and draw out three marbles.

which he has three of the four x's and two of the four y's, e.g. three kings and two eights. If five cards are drawn at random from a standard deck, the probability of a full house is

$$\frac{\binom{13}{1}\binom{12}{1}\binom{4}{3}\binom{4}{2}}{\binom{52}{5}} \approx 0.00144.$$

73.3 Conditional Probability and Independent Events

A jar contains twenty marbles of which six are red, nine are blue and the remaining five are green. While blindfolded, Xing selects two of the twenty marbles random (without replacement) and puts one in his left pocket and one in his right pocket. He then takes off the blindfold.

The probability that the marble in his left pocket is red is $6/20$. But Xing first reaches into his right pocket, takes this marble out and discovers that it is is blue. Is the probability that the marble in his left pocket is red still $6/20$? Intuition says that it's slightly higher than that. Here's a more formal framework for answering such questions.

Let (S, P) be a probability space and let B be an event for which $P(B) > 0$. Then for every event $A \subseteq S$, we define the *probability of A, given B*, denoted $P(A|B)$, by setting $P(A|B) = P(A \cap B)/P(B)$.

Discussion 73.7. Returning to the question raised at the beginning of the section, Bob says that this is just conditional probability. He says let B be the event that the marble in the right pocket is blue and let A be the event that the marble in the left pocket is red. Then $P(B) = 9/20$, $P(A) = 6/20$ and $P(A \cap B) = (9 \cdot 6)/380$, so that $P(A|B) = \frac{54}{380}\frac{20}{9} = 6/19$, which is of course slightly larger than $6/20$. Alice is impressed.

Example 73.8. Consider the jar of twenty marbles from the preceding example. A second jar of marbles is introduced. This jar has eighteen marbles: nine red, five blue and four green. A jar is selected at random and from this jar, two marbles are chosen at random. What is the probability that both are green? Bob is on a roll. He says "Let G be the event that both marbles are green, and let J_1 and J_2 be the event that the marbles come from the first jar and the second jar, respectively. Then $G = (G \cap J_1) \cup (G \cap J_2)$, and $(G \cap J_1) + (G \cap J_2) = \emptyset$. Furthermore, $P(G|J_1) = \binom{5}{2}/\binom{20}{2}$ and $P(G|J_2) = \binom{4}{2}/\binom{18}{2}$, while $P(J_1) = P(J_2) = 1/2$. Also $P(G \cap J_i) = P(J_i)P(G|J_i)$ for each $i = 1, 2$. Therefore,

$$P(G) = \frac{1}{2}\frac{\binom{5}{2}}{\binom{20}{2}} + \frac{1}{2}\frac{\binom{4}{2}}{\binom{18}{2}} = \frac{1}{2}\left(\frac{20}{380} + \frac{12}{306}\right)."$$

Now Alice is speechless.

73.3.1 Independent Events

Let A and B be events in a proability space (S, P). We say A and B are *independent* if $P(A \cap B) = P(A)P(B)$. Note that when $P(B) \neq 0$, A and B are independent if and only if $P(A) = P(A|B)$. Two events that are not independent are said to be *dependent*. Returning to our earlier example, the two events (A: the marble in Xing's left pocket is red and B: the marble in his right pocket is blue) are dependent.

Example 73.9. Consider the two jars of marbles from Example 10.8. One of the two jars is chosen at random and a single marble is drawn from that jar. Let A be the event that the second jar is chosen, and let B be the event that the marble chosen turns out to be green. Then $P(A) = 1/2$ and $P(B) = \frac{1}{2}\frac{5}{20} + \frac{1}{2}\frac{4}{18}$. On the other hand, $P(A \cap B) = \frac{1}{2}\frac{4}{18}$, so $P(A \cap B) \neq P(A)P(B)$, and the two events are not independent. Intuitively, this should be clear, since once you know that the marble is green, it is more likely that you actually chose the first jar.

Example 73.10. A pair of dice are rolled, one red and one blue. Let A be the event that the red die shows either a 3 or a 5, and let B be the event that you get doubles, i.e., the red die and the blue die show the same number. Then $P(A) = 2/6$, $P(B) = 6/36$, and $P(A \cap B) = 2/36$. So A and B are independent.

73.4 Bernoulli Trials

Suppose we have a jar with 7 marbles, four of which are red and three are blue. A marble is drawn at random and we record whether it is red or blue. The probability p of getting a red marble is $4/7$; and the probability of getting a blue is $1 - p = 3/7$.

Now suppose the marble is put back in the jar, the marbles in the jar are stirred, and the experiment is repeated. Then the probability of getting a red marble on the second trial is again $4/7$, and this pattern holds regardless of the number of times the experiment is repeated.

It is customary to call this situation a series of *Bernoulli trials*. More formally, we have an experiment with only two outcomes: *success* and *failure*. The probability of success is p and the probability of failure is $1 - p$. Most importantly, when the experiment is repeated, then the probability of success on any individual test is exactly p.

We fix a positive integer n and consider the case that the experiment is repeated n times. The outcomes are then the binary strings of length n from the two-letter alphabet $\{S, F\}$, for success and failure, respectively. If x is a string with i sucesses and $n - i$ failures, then $P(x) = \binom{n}{i} p^i (1-p)^{n-i}$. Of course, in applications, success and failure may be replaced by: head/tails, up/down, good/bad, forwards/backwards, red/blue, etc.

Example 73.11. When a die is rolled, let's say that we have a success if the result is a two or a five. Then the probability p of success is $2/6 = 1/3$ and the probability of failure is $2/3$. If the die is rolled ten times in succession, then the probability that we get exactly exactly four successes is $C(10,4)(1/3)^4(2/3)^6$.

Example 73.12. A fair coin is tossed 100 times and the outcome (heads or tails) is recorded. Then the probability of getting heads 40 times and tails the other 60 times is

$$\binom{100}{40}\left(\frac{1}{2}\right)^{40}\left(\frac{1}{2}\right)^{60} = \frac{\binom{100}{40}}{2^{100}}.$$

Discussion 73.13. Bob says that if a fair coin is tossed 100 times, it is fairly likely that you will get exactly 50 heads and 50 tails. Dave is not so certain this is right. Carlos fires up his computer and in few second, he reports that the probability of getting exactly 50 heads when a fair coin is tossed 100 times is

$$\frac{12611418068195524166851562157}{158456325028528675187087900672}$$

which is .079589, to six decimal places. In other words, not very likely at all. Xing is doing a modestly more complicated calculation, and he reports that you have a 99% chance that the number of heads is at least 20 and at most 80. Carlos adds that when n is very large, then it is increasingly certain that the number of heads in n tosses will be close to $n/2$. Dave asks what do you mean by close, and what do you mean by very large?

73.5 Discrete Random Variables

Let (S, P) be a probability space and let $X : S \longrightarrow \mathbb{R}$ be any function that maps the outcomes in S to real numbers (all values allowed, positive, negative and zero). We call[2] X a *random variable*. The quantity $\sum_{x \in S} X(x)P(x)$, denoted $E(X)$, is called the *expectation* (also called the *mean* or *expected value*) of the random variable X. As the suggestive name reflects, this is what one should expect to be the average behavior of the result of repeated Bernoulli trials.

Note that since we are dealing only with probability spaces (S, P) where S is a finite set, the range of the probability measure P is actually a finite set. Accordingly, we can rewrite the formula for $E(X)$ as $\sum_y y \cdot \text{prob}(X(x) = y)$, where the summation extends over a finite range of values for y.

Example 73.14. For the spinner shown in Figure 10.1, let $X(i) = i^2$ where i is the number of the region. Then

$$E(X) = \sum_{i \in S} i^2 P(i) = 1^2\frac{1}{8} + 2^2\frac{2}{8} + 3^2\frac{1}{8} + 4^2\frac{1}{8} + 5^2\frac{3}{8} = \frac{109}{8}.$$

[2]For historical reasons, capital letters, like X and Y are used to denote random variables. They are just functions, so letters like f, g and h might more seem more natural—but maybe not.

Note that $109/8 = 13.625$. The significance of this quantity is captured in the following statement. If we record the result from the spinner n times in succession as $(i_1, i_2, \ldots, i_n)$ and Xing receives a prize worth i_j^2 for each $j = 1, 2, \ldots, n)$, then Xing should "expect" to receive a total prize worth $109n/8 = 13.625n$. Bob asks how this statement can possibly be correct, since $13.625n$ may not even be an integer, and any prize Xing receives will have integral value. Carlos goes on to explain that the concept of expected value provides a formal definition for what is meant by a fair game. If Xing pays 13.625 cents to play the game and is then paid i^2 pennies where i is the number of the region where the spinner stops, then the game is fair. If he pays less, he has an unfair advantage, and if he pays more, the game is biased against him. Bob says "How can Xing pay 13.625 pennies?" Brushing aside Bob's question, Carlos says that one can prove that for every $\epsilon > 0$, there is some n_0 (which depends on ϵ) so that if $n > n_0$, then the probability that Xing's total winnings minus $13.625n$, divided by n is within ϵ of 13.625 is at least $1 - \epsilon$. Carlos turns to Dave and explains politely that this statement gives a precise meaning of what is meant by "close" and "large".

Example 73.15. For Alice's game as detailed at the start of the chapter, $S = \{0, 1, 2, 3, 4, 5\}$, we could take X to be the function defined by $X(d) = 2 - d$. Then $X(d)$ records the amount that Bob wins when the difference is d (a negative win for Bob is just a win for Alice in the same amount). We calculate the expectation of X as follows:

$$E(X) = \sum_{d=0}^{5} X(d)p(d) = -2\frac{1}{6} - 1\frac{10}{36} + 0\frac{8}{36} + 1\frac{6}{36} + 2\frac{4}{36} + 3236 = \frac{-2}{36}.$$

Note that $-2/36 = -.055555\ldots$. So if points were dollars, each time the game is played, Bob should expect to lose slightly more than a nickel. Needless to say, Alice likes to play this game and the more times Bob can be tricked into playing, the more she likes it. On the other hand, by this time in the chapter, Bob should be getting the message and telling Alice to go suck a lemon.

73.5.1 The Linearity of Expectation

The following fundamental property of expectation is an immmediate consequence of the definition, but we state it formally because it is so important to discussions to follow.

Proposition 73.16. *Let (S, P) be a probability space and let $X_1, X_2, \ldots, X_n$ be random variables. Then*

$$E(X_1 + X_2 + \cdots + X_t) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

73.5.2 Implications for Bernoulli Trials

Example 73.17. Consider a series of n Bernoulli trials with p, the probability of success, and let X count the number of successes. Then, we claim that

$$E(X) = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = np$$

To see this, consider the function $f(x) = [px + (1-p)]^n$. Taking the derivative by the chain rule, we find that $f'(x) = np[px + (1-p)]^{n-1}$. Now when $x = 1$, the derivative has value np.

On the other hand, we can use the binomial theorem to expand the function f.

$$f(x) = \sum_{i=0}^{n} \binom{n}{i} x^i p^i (1-p)^{n-i}$$

It follows that

$$f'(x) = \sum_{i=0}^{n} i \binom{n}{i} x^{i-} p^i (1-p)^{n-i}$$

And now the claim follows by again setting $x = 1$. Who says calculus isn't useful!

Example 73.18. Many states have lotteries to finance college scholarships or other public enterprises judged to have value to the public at large. Although far from a scientific investigation, it seems on the basis of our investigation that many of the games have an expected value of approximately fifty cents when one dollar is invested. So the games are far from fair, and no one should play them unless they have an intrinsic desire to support the various causes for which the lottery profits are targeted.

By contrast, various games of chance played in gambling centers have an expected return of slightly less than ninety cents for every dollar wagered. In this setting, we can only say that one has to place a dollar value on the enjoyment derived from the casino environment. From a mathematical standpoint, you are going to lose. That's how they get the money to build those exotic buildings.

73.6 Central Tendency

Consider the following two situations.

Situation 1. A small town decides to hold a lottery to raise funds for charitable purposes. A total of 10,001 tickets are sold, and the tickets are labeled with numbers from the set $\{0, 1, 2, \ldots, 10,000\}$. At a public ceremony, duplicate tickets are placed in a big box, and the mayor draws the winning ticket from out of the box. Just to heighten the suspense as to who has actually won the prize, the mayor reports that the winning number is at least $7,500$. The citizens ooh and aah and they can't wait to see who among them will be the final winner.

Situation 2. Behind a curtain, a fair coin is tossed $10,000$ times, and the number of heads is recorded by an observer, who is reputed to be honest and impartial. Again, the outcome is an integer in the set $\{0, 1, 2, \ldots, 10,000\}$. The observer then emerges from behind the curtain and announces that the number of heads is at least than $7,500$. There is a pause and then someone says "What? Are you out of your mind?"

So we have two probability spaces, both with sample space $S = \{0, 1, 2, \ldots, 10,000\}$. For each, we have a random variable X, the winning ticket number in the first situation, and the number of heads in the second. In each case, the expected value, $E(X)$, of the random variable X is $5,000$. In the first case, we are not all that surprised at an outcome far from the expected value, while in the second, it seems intuitively clear that this is an extraordinary occurrence. The mathematical concept here is referred to as *central tendency*, and it helps us to understand just how likely a random variable is to stray from its expected value.

For starters, we have the following elementary result which is called Markov's inequality.

Theorem 73.19. *Let X be a random variable in a probability space (S, P). Then for every $k > 0$,*

$$P(|X| \geq k) \leq E(|X|)/k.$$

Proof. Of course, the inequality holds trivially unless $k > E(|X|)$. For k in this range, we establish the equivalent inequality: $kP(|X| \geq k) \leq E(|X|)$.

$$\begin{aligned}
kP(|X| \geq k) &= \sum_{r \geq k} kP(|X| = r) \\
&\leq \sum_{r \geq k} rP(|X| = r) \\
&\leq \sum_{r > 0} rP(|X| = r) \\
&= E(|X|).
\end{aligned}$$

$\square$

To make Markov's inequality more concrete, we see that on the basis of this trivial result, the probability that either the winning lottery ticket or the number of heads is at least $7,500$ is at most $5000/7500 = 2/3$. So nothing alarming here in either case. Since we still feel that the two cases are quite different, a more subtle measure will be required.

73.6.1 Variance and Standard Deviation

Again, let (S, P) be a probability space and let X be a random variable. The quantity $E((X - E(X))^2)$ is called the *variance* of X and is denoted $\mathrm{var}(X)$. Evidently, the variance of X is a non-negative number. The *standard deviation* of X, denoted σ_X is then defined as the quantity $\sqrt{\mathrm{var}(x)}$, i.e., $\sigma_X^2 = \mathrm{var}(X)$.

Example 73.20. For the spinner shown at the beginning of the chapter, let $X(i) = i^2$ when the pointer stops in region i. Then we have already noted that the expectation $E(X)$ of the random variable X is 109/8. It follows that the variance $\mathrm{var}(X)$ is:

$$\begin{aligned}\mathrm{var}(X) &= (1^2 - \frac{109}{8})^2\frac{1}{8} + (2^2 - \frac{109}{8})^2\frac{1}{4} + (3^2 - \frac{109}{8})^2\frac{1}{8} + (4^2 - \frac{109}{8})^2\frac{1}{8} + (5^2 - \frac{109}{8})^2\frac{3}{8}\\ &= (108^2 + 105^2 + 100^2 + 93^2 + 84^2)/512\\ &= 48394/512\end{aligned}$$

It follows that the standard deviation σ_X of X is then $\sqrt{48394/512} \approx 9.722$.

Example 73.21. Suppose that $0 < p < 1$ and consider a series of n Bernoulli trials with the probability of success being p, and let X count the number of successes. We have already noted that $E(X) = np$. Now we claim the the variance of X is given by:

$$\mathrm{var}(X) = \sum_{i=0}^{n}(i - np)^2\binom{n}{i}p^i(1-p)^{n-i} = np(1-p)$$

There are several ways to establish this claim. One way is to proceed directly from the definition, using the same method we used previously to obtain the expectation. But now you need also to calculate the second derivative. Here is a second approach, one that capitalizes on the fact that separate trials in a Bernoulli series are independent.

Let $\mathcal{F} = \{X_1, X_2, \ldots, X_n\}$ be a family of random variables in a probability space (S, P). We say the family $\mathcal{F}$ is *independent* if for each i and j with $1 \leq i < j \leq n$, and for each pair a, b of real numbers with $0 \leq a, b \leq 1$, the follwing two events are independent: $\{x \in S : X_i(x) \leq a\}$ and $\{x \in S : X_j(x) \leq b\}$. When the family is independent, it is straightforward to verify that

$$\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n).$$

With the aid of this observation, the calculation of the variance of the random variable X which counts the number of successes becomes a trivial calculation. But in fact, the entire treatment we have outlined here is just a small part of a more complex subject which can be treated more elegantly and ultimately much more compactly—provided you first develop additional background material on families of random variables. For this we will refer you to suitable probability and statistics texts, such as those given in our references.

Proposition 73.22. *Let X be a random variable in a probability space (S, P). Then* $\operatorname{var}(X) = E(X^2) - E^2(X)$.

Proof. Let $E(X) = \mu$. From its definition, we note that

$$\begin{aligned}
\operatorname{var}(X) &= \sum_r (r-\mu)^2 \operatorname{prob}(X=r) \\
&= \sum_r (r^2 - 2r\mu + \mu^2) \operatorname{prob}(X=r) \\
&= \sum_r r^2 \operatorname{prob}(X=r) - 2\mu \sum_r r \operatorname{prob}(X=r) + \mu^2 \sum_r \operatorname{prob}(X=r) \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - E^2(X).
\end{aligned}$$

□

Variance (and standard deviation) are quite useful tools in discussions of just how likely a random variable is to be near its expected value. This is reflected in the following theorem, known as Chebychev's inequality.

Theorem 73.23. *Let X be a random variable in a probability space (S, P), and let $k > 0$ be a positive real number. If the expectation $E(X)$ of X is μ and the standard deviation is σ_X, then*

$$\operatorname{prob}(|X - E(X)| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

Proof. Let $A = \{r \in \mathbb{R} : |r - \mu| > k\sigma_X\}$.
Then we have:

$$\begin{aligned}
\operatorname{var}(X) &= E((X-\mu)^2) \\
&= \sum_{r \in \mathbb{R}} (r-\mu)^2 \operatorname{prob}(X=r) \\
&\geq \sum_{r \in A} (r-\mu)^2 \operatorname{prob}(X=r) \\
&\geq k^2\sigma_X^2 \sum_{r \in A} \operatorname{prob}(X=r) \\
&\geq k^2\sigma_X^2 \operatorname{prob}(|X-\mu| > k\sigma_X).
\end{aligned}$$

Since $\operatorname{var}(X) = \sigma_X^2$, we may now deduce that $1/k^2 \geq \operatorname{prob}(|X-\mu|) > k\sigma_X)$. Therefore, since $\operatorname{prob}(|X-\mu| \leq k\sigma_X) = 1 - \operatorname{prob}(|X-\mu| > k\sigma_X)$, we conclude that

$$\operatorname{prob}(|X-\mu| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

□

Example 73.24. Here's an example of how this inequality can be applied. Consider n tosses of a fair coin with X counting the number of heads. As noted before, $\mu = E(X) = n/2$ and $\text{var}(X) = n/4$, so $\sigma_X = \sqrt{n}/2$. When $n = 10,000$ and $\mu = 5,000$ and $\sigma_X = 50$. Setting $k = 50$ so that $k\sigma_X = 2500$, we see that the probability that X is within 2500 of the expected value of 5000 is at least 0.9996. So it seems very unlikely indeed that the number of heads is at least $7,500$.

Going back to lottery tickets, if we make the rational assumption that all ticket numbers are equally likely, then the probability that the winning number is at least $7,500$ is exactly $2501/100001$, which is very close to $1/4$.

Example 73.25. In the case of Bernoulli trials, we can use basic properties of binomial coefficients to make even more accurate estimates. Clearly, in the case of coin tossing, the probability that the number of heads in $10,000$ tosses is at least $7,500$ is given by

$$\sum_{i=7,500}^{10,000} \binom{10,000}{i} / 2^{10,000}$$

Now a computer algebra system can make this calculation exactly, and you are encouraged to check it out just to see how truly small this quantity actually is.

73.7 Probability Spaces with Infinitely Many Outcomes

To this point, we have focused entirely on probability spaces (S, P) with S a finite set. More generally, probability spaces are defined where S is an infinite set. When S is countably infinite, we can still define P on the members of S, and now $\sum_{x \in S} P(x)$ is an infinite sum which converges absolutely (since all terms are non-negative) to 1. When S is uncountable, P is not defined on S. Instead, the probabilty function is defined on a family of subsets of S. Given our emphasis on finite sets and combinatorics, we will discuss the first case briefly and refer students to texts that focus on general concepts from probability and statistics for the second.

Example 73.26. Consider the following game. Yolanda rolls a single die. She wins if she rolls a six. If she rolls any other number, she then rolls again and again until the first time that one of the following two situations occurs: (1) she rolls a six, which now this results in a loss or (2) she rolls the same number as she got on her first roll, which results in a win. As an example, here are some sequences of rolls that this game might take:

1. $(4, 2, 3, 5, 1, 1, 1, 4)$. Yolanda wins!
2. (6). Yolanda wins!

3. $(5, 2, 3, 2, 1, 6)$. Yolanda loses. Ouch.

So what is the probability that Yolanda will win this game?

Yolanda can win with a six on the first roll. That has probability $1/6$. Then she might win on round n where $n \geq 2$. To accomplish this, she has a $5/6$ chance of rolling a number other than six on the first roll; a $4/6$ chance of rolling something that avoids a win/loss decision on each of the rolls, 2 through $n-1$ and then a $1/6$ chance of rolling the matching number on round n. So the probability of a win is given by:

$$\frac{1}{6} + \sum_{n\geq 2} \frac{5}{6}\left(\frac{4}{6}\right)^{n-2}\frac{1}{6} = \frac{7}{12}.$$

Example 73.27. You might think that something slightly more general is lurking in the background of the preceding example—and it is. Suppose we have two disjoint events A and B in a probability space (S, P) and that $P(A) + P(B) < 1$. Then suppose we make repeated samples from this space with each sample independent of all previous ones. Call it a win if event A holds and a loss if event B holds. Otherwise, it's a tie and we sample again. Now the probability of a win is:

$$P(A) + P(A)\sum_{n\geq 1}(1 - P(A) - P(B))^n = \frac{P(A)}{P(A) + P(B)}.$$

73.8 Exercises

1. The club of seven (Alice, Bob, Carlos, Dave, Xing, Yolanda and Zori) are students in a class with a total enrolment of 35. The professor chooses three students at random to go to the board to work challenge problems.
   a) What is the probability that Yolanda is chosen?
   b) What is the probability that Yolanda is chosen and Zori is not?
   c) What is the probability that exactly two members of the club are chosen?
   d) What is the probability that none of the seven members of club are chosen?
2. Bob says to no one in particular, "Did you know that the probability that you will get at least one "7" in three rolls of a pair of dice is slightly less than 1/2. On the other hand, the probability that you'll get at least one "5" in six rolls of the dice is just over 1/2." Is Bob on target, or out to lunch?
3. Consider the spinner shown in Figure 10.1 at the beginning of the chapter.
   a) What is the probability of getting at least one "5" in three spins?
   b) What is the probability of getting at least one "3" in three spins?

c) If you keep spinning until you get either a "2" or a "5", what is the probability that you get a "2" first?

d) If you receive i dollars when the spinner halts in region i, what is the expected value? Since three is right in the middle of the possible outcomes, is it reasonable to pay three dollars to play this game?

4. Alice proposes to Bob the following game. Bob pays one dollar to play. Fifty balls marked $1, 2, \ldots, 50$ are placed in a big jar, stirred around, and then drawn out one by one by Zori, who is wearing a blindfold. The result is a random permutation σ of the integers 1, $2, \ldots, 50$. Bob wins with a payout of two dollars and fifty cents if the permutation σ is a derangement, i.e., $\sigma(i) \neq i$ for all $i = 1, 2, \ldots, n$. Is this a fair game for Bob? If not how should the payoff be adjusted to make it fair?

5. A random graph with vertex set $\{1, 2, \ldots, 10\}$ is constructed by the following method. For each two element subset $\{i, j\}$ from $\{1, 2, \ldots, 10\}$, a fair coin is tossed and the edge $\{i, j\}$ then belongs to the graph when the result is "heads." For each 3-element subset $S \subseteq \{1, 2, \ldots, n\}$, let E_S be the event that S is a complete subgraph in our random graph.

   a) Explain why $P(E_S) = 1/8$ for each 3-element subset S.

   b) Explain why E_S and E_T are independent when $|S \cap T| \leq 1$.

   c) Let $S = \{1, 2, 3\}$, $T = \{2, 3, 4\}$ and $U = \{3, 4, 5\}$. Show that $P(E_S|E_T) \neq P(E_S|E_T E_U)$.

6. Ten marbles labeled $1, 2, \ldots, 10$ are placed in a big jar and then stirred up. Zori, wearing a blindfold, pulls them out of the jar two at a time. Players are allowed to place bets as to whether the sum of the two marbles in a pair is 11. There are $C(10, 2) = 45$ different pairs and exactly 5 of these pairs sums to eleven.

   Suppose Zori draws out a pair; the results are observed; then she returns the two balls to the jar and all ten balls are stirred before the next sample is taken. Since the probability that the sum is an "11" is $5/45 = 1/9$, then it would be fair to pay one dollar to play the game if the payoff for an "11" is nine dollars. Similarly, the payoff for a wager of one hundred dollars should be nine hundred dollars.

   Now consider an alternative way to play the game. Now Zori draws out a pair; the results are observed; and the marbles are set aside. Next, she draws another pair from the remaining eight marbles, followed by a pair selected from the remaining six, etc. Finally, the fifth pair is just the pair that remains after the fourth pair has been selected. Now players may be free to wager on the outcome of any or all or just some of the five rounds. Explain why either everyone should or no one should wager on the fifth round. Accordingly, the last round is skipped and all marbles are returned to the jar and we start over again.

Also explain why an observant player can make lots of money with a payout ratio of nine to one. Now for a more challenging problem, what is the minimum payout ratio above which a player has a winning strategy?