1 Probabilities and Expected Values

1.1 Random Variables

Informally, a *random variable* is a number that takes on different values with different probabilities.[1] If X is a random variable, and x is a number, we write $\rho_X(x)$, or $Pr(X = x)$, for the probability that the value of X is x. In this course, most of our random variables will be *discrete*, meaning that there is a set of values D, the *support* of X, such that

1. For any $x \in D$, $\rho_X(x) > 0$, and
2. $\sum_{x \in D} \rho_X(x) = 1$

1.2 Independent Random Variables

Informally, we say that random variables X and Y are *independent* if knowledge of X gives no clue about the knowledge of Y. This imples that, for any given x and y, the probability that $X = x$ and $Y = y$ is $\rho_X(x) \cdot \rho_Y(y)$.

1.3 Expected Value

If X is a discrete random variable, then the *expected value* of X, written $E(X)$, is defined to be

$$E(X) = \sum_x \rho_X(x) \cdot x$$

If X is not discrete, then the formula for $E(X)$ involves integration.

Exercise 1.1 *The probability that a certain salesman makes a sale in a given day is $\frac{1}{3}$. What is the expected number of sales he makes in one day?*

Answer: Let X be the number of sales that the salesman makes in one day. Then the support of X is $D = \{0, 1\}$. $\rho_X(0) = \frac{2}{3}$, and $\rho_X(1) = \frac{1}{3}$. Thus $E(X) = \frac{2}{3} \cdot 0 + \frac{1}{3} \cdot 1 = \frac{1}{3}$.

Exercise 1.2 *The probability that a certain salesman makes a sale in a given day is $\frac{1}{3}$. What is the expected number of sales he makes in n days?*

Answer: Let X_i be the number of sales that the salesman makes in the i^{th} day. Then $X = X_1 + X_2 + \cdots + X_n$ is the number of sales that he makes in n days. The expected value of each X_i is given by the answer to Exercise 1.1, and expected values add. Thus

$$E(X) = \sum_{i=1}^{n} E(X_i) = n \cdot \tfrac{1}{3} = \tfrac{n}{3}$$

Exercise 1.3 *The probability that a second salesman makes one sale in a given day is $\frac{1}{2}$, and the probability that he makes two sales in that day is $\frac{1}{6}$. What is the expected number of sales the second salesman makes in a week (5 days)?*

[1] For a formal definition of a random variable, see, for example, `http://en.wikipedia.org/wiki/Random_variable`

1.4 Waiting Time

Given a process consisting of steps, the *waiting time* of an event is the number of steps that elapse before that event occurs.

Exercise 1.4 *There are two children waiting for candy. At step 1, one of the children gets candy. At step 2, the other child gets candy. What is the waiting time for each child to get candy?*

Answer: The waiting time for the first child is 1; the waiting time for the second child is 2.

Exercise 1.5 *There are two children waiting for candy. At step 1, one of the children gets candy. At step 2, the other child gets candy. What is the expected waiting time for a randomly chosen child?*

Answer: $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 2 = \frac{3}{2}$.

Exercise 1.6 *There are five children waiting for candy. At each step, one of the children gets candy, and no child gets candy twice. What is the expected waiting time for a randomly chosen child to get candy?*

Exercise 1.7 *Consider the first salesman, from Exercise 1.1. He will work every day until he makes a sale, then go on vacation. Assume that each day is independent of all previous days,* i.e., *if X_i is the number of sales he makes on the i^{th} day, then X_i and X_j are independent if $i \neq j$.*

(a) *What is the expected number of days he will work before he goes on vacation?*

(b) *What is the probability that he will eventually get to go on vacation?*

Answer:

(b): Because of independence, the probability that he will not make a sale in the first n days is $\left(\frac{2}{3}\right)^n$. Thus, the probability that he will never make a sale is $\lim_{n\to\infty} \left(\frac{2}{3}\right)^n = 0$. The probability that we will eventually go on vacation is thus 1.

(a): There are two somewhat different ways to work this problem.

Direct method: Let W be the waiting time to the first sale, that is, the number of days he works, which is a random variable. Then $\rho_W(1) = \frac{1}{3}$, since the probability is $\frac{1}{3}$ that he works only one day. Since the days events are independent of each other, the probability he makes a sale on the second day and not on the first day is $\frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$; thus, $\rho_W(2) = \frac{2}{9}$. Similarly, for any given integer $t \geq 1$, $\rho_W(t) = \frac{2^{t-1}}{3^t}$. Finally

$$E(W) = \sum_{t=1}^{\infty} \rho_W(t) \cdot t = \sum_{t=1}^{\infty} \frac{2^{t-1}}{3^t} \cdot t$$

But, do you remember how to sum that series?

Recursive method: Since each day is independent, the expected waiting time on the first day is the same as the expected waiting time on the second day if he didn't make a sale the first day. Thus

$$E(W) = 1 + \frac{2}{3} \cdot E(W)$$

The logic of this is that he must wait one day, after which he has the probability of $\frac{2}{3}$ of having made no progress. We conclude that $E(W) = 3$.

Exercise 1.8 *Referring to Exercise 1.2, let X_i be the number of sales the first salesman makes in the i^{th} day. Prove that X_1 and $X_1 + X_2$ are not independent.*

Exercise 1.9 *A deck of n cards contains one joker. The deck is shuffled. You then turn cards face up, one at a time, until you see the joker. What is the expected number of cards you turn?*

The following exercise is describes an actual drawing that is held every week at the local Rotary Club, whose meeting I attended last Thursday.

Exercise 1.10 *A deck of n cards contains two jokers. The deck is shuffled. Every week, a contenstant draws one card. If the card is a joker, the contestant wins, and the contest starts over the following week with a fresh deck. Otherwise, the drawn card is removed from the deck, increasing the probability that the next contestant wins.*
What is the expected number of weeks that will elapse before someone wins?

Exercise 1.11 *Suppose that $E_1, E_2, \ldots$ is an infinite sequence of independent and identically distributed Boolean variables, where each E_i has the value 1 (*i.e.,*true) with probability p, where $0 < p \leq 1$.*
Let $W = \min\{i : E_i\}$, the waiting time until some E_i holds, which is a random variable of integer type. Prove that $E(W) = 1/p$. (Hint: use a recurrence relation.)

Harmonic Numbers

The harmonic numbers are sums of reciprocals of integers. For any $n \geq 1$, define the n^{th} harmonic number H_n as follows:

$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$

Thus, for example, $H_1 = 1$, $H_2 = \frac{3}{2}$, and $H_3 = \frac{5}{6}$.

Exercise 1.12 *Prove that $H_n = \Theta(\log n)$.*

The Coupon Collector's Problem

Each day that you shop at a certain store, they give you a coupon that has one symbol on it. There are a total of n different symbols. When you have at least one coupon with each symbol, you win a prize.

Exercise 1.13 *Assuming that you shop at the store every day, and the coupon you get on each day is independently chosen, what is the expected waiting time before you get your prize?* (Hint: use a recurrence relation, together with the result given in Exercise 1.11.)