Collecting Premiums in Cereal Boxes

Your favorite breakfast cereal, in an effort to urge you to buy more cereal, encloses a toy or a premium in each box. How many boxes must you buy in order to collect all the premiums? This problem is also often called the *coupon collector's problem* in the literature on probability theory. Of course we cannot be *certain* to collect all the premiums, given finite resources, but we could think about the average, or expected number, of boxes to be purchased.

To make matters specific, suppose there are 6 premiums. The first box gives us a premium we did not have before. The probability the next box will not duplicate the premium we already have is 5/6. This waiting time for the next premium not already collected is a geometric random variable, with probability 5/6. The expected waiting time for the second premium is then 1/(5/6). Now we have two premiums, so the probability the next box contains a new premium is 4/6. This is again a geometric variable and our waiting time for collecting the third premium is 1/(4/6). This process continues. Since the expectation of a sum is the sum of the expectations of the summands and if we let X denote the total number of boxes purchased in order to secure all the premiums, we conclude that

$$E(X) = 1 + \frac{1}{\frac{5}{6}} + \frac{1}{\frac{4}{6}} + \frac{1}{\frac{3}{6}} + \frac{1}{\frac{2}{6}} + \frac{1}{\frac{1}{6}}$$

$$E(X) = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1}$$

$$E(X) = 1 + 1.2 + 1.5 + 2 + 3 + 6 = 14.7 \text{ boxes.}$$

Clearly the cereal company knows what it is doing! An exercise will ask the reader to show that the variance of X is 38.99, so unlucky cereal eaters could be in for buying many more boxes than the expectation would indicate.

This is an example of a series of trials, analogous to Poisson's trials, in which the probabilities vary. Since the total number of trials, X, can be regarded as a sum of geometric variables (plus 1 for the first box), and since the probability generating function for a geometric variable is $\frac{qt}{1-pt}$, the probability generating function of X is

$$P_X(t) = \frac{\frac{5}{6}t}{1-\frac{1}{6}t} \cdot \frac{\frac{4}{6}t}{1-\frac{2}{6}t} \cdot \frac{\frac{3}{6}t}{1-\frac{3}{6}t} \cdot \frac{\frac{2}{6}t}{1-\frac{4}{6}t} \cdot \frac{\frac{1}{6}t}{1-\frac{5}{6}t}.$$

This can be written as

$$P_X(t) = \frac{5!t^5}{(6-t)(6-2t)(6-3t)(6-4t)(6-5t)}.$$

The first few terms in a power series expansion of $P_X(t)$ are as follows:

$$P_X(t) = \frac{5t^5}{324} + \frac{25t^6}{648} + \frac{175t^7}{2916} + \frac{875t^8}{11664} + \frac{11585t^9}{139968} + \frac{875t^{10}}{10368} + \frac{616825t^{11}}{7558272} + \cdots$$

Probabilities can be found from $P_X(t)$, but not at all easily without a computer algebra system. The series above shows that the probability it takes 9 boxes in total to collect all 6 premiums is $875/11664 = 0.075$.

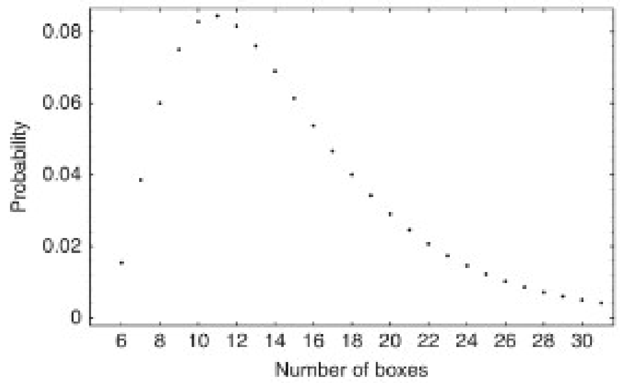

Figure 4.7 Probabilities for the cereal box problem.

A graph of the probability distribution function is shown in Figure 4.7. The probabilities shown there are the probabilities it takes n boxes to collect all 6 premiums. We will return to this problem and some of its variants in Chapter 7.

EXERCISES 4.7

1. Use the generating function for the binomial random variable with $p = 2/3$ to verify μ and σ^2.

2. In the cereal box problem, find σ^2 using a generating function.

3. **(a)** Find the probability generating function for a Poisson random variable with parameter λ.

(b) Use the generating function in part (a) to find the mean and variance of a Poisson random variable.

4. Use probability generating functions to show that the sum of independent Poisson variables, with parameters $\lambda_x$and λ_y, respectively, has a Poisson distribution with parameter $\lambda_x + \lambda_y$.

5. A discrete random variable, X, has probability distribution function $f(x) = k/2^x, x = 0, 1, 2, 3, 4$.

(a) Find k.

(b) Find $P_X(t)$, the probability generating function.

(c) Use $P_X(t)$ to find the mean and variance of X.

6. Use the probability generating function to find the mean and variance of a negative binomial variable with parameters r and p.

7. A fair coin is tossed eight times followed by 12 tosses of a coin loaded so as to come up heads with probability 3/4. What is the probability that

(a) exactly 10 heads occur?

(b) at least 10 heads occur?

8. Use the probability generating function of a Bernoulli random variable to show that the sum of independent Bernoulli variables is a binomial random variable.