1 Family Planning

Note 15

Mr. and Mrs. Johnson decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Johnsons have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

(b) Compute the joint distribution of G and C. Fill in the table below.

	$C=1$	$C=2$	$C=3$
$G=0$			
$G=1$			

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

$\mathbb{P}[G=0]$	
$\mathbb{P}[G=1]$	

$\mathbb{P}[C=1]$	$\mathbb{P}[C=2]$	$\mathbb{P}[C=3]$

(d) Are G and C independent?

(e) What is the expected number of girls the Johnsons will have? What is the expected number of children that the Johnsons will have?

Solution:

(a) The sample space is the set of all possible sequences of children that the Johnsons can have: $\Omega = \{g, bg, bbg, bbb\}$. The probabilities of these sample points are:

$$\mathbb{P}[g] = \frac{1}{2}$$
$$\mathbb{P}[bg] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$\mathbb{P}[bbg] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$
$$\mathbb{P}[bbb] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

(b)

	$C = 1$	$C = 2$	$C = 3$
$G = 0$	0	0	$\mathbb{P}[bbb] = 1/8$
$G = 1$	$\mathbb{P}[g] = 1/2$	$\mathbb{P}[bg] = 1/4$	$\mathbb{P}[bbg] = 1/8$

(c) Marginal distribution for G:

$$\mathbb{P}[G = 0] = 0 + 0 + \frac{1}{8} = \frac{1}{8}$$
$$\mathbb{P}[G = 1] = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$$

Marginal distribution for C:

$$\mathbb{P}[C = 1] = 0 + \frac{1}{2} = \frac{1}{2}$$
$$\mathbb{P}[C = 2] = 0 + \frac{1}{4} = \frac{1}{4}$$
$$\mathbb{P}[C = 3] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

(d) No, G and C are not independent. If two random variables are independent, then

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x]\mathbb{P}[Y = y].$$

To show this dependence, consider an entry in the joint distribution table, such as $\mathbb{P}[G = 0, C = 3] = 1/8$. This is not equal to $\mathbb{P}[G = 0]\mathbb{P}[C = 3] = (1/8) \cdot (1/4) = 1/32$, so the random variables are not independent.

(e) We can apply the definition of expectation directly for this problem, since we've computed the marginal distribution for both random variables.

$$\mathbb{E}[G] = 0 \cdot \mathbb{P}[G = 0] + 1 \cdot \mathbb{P}[G = 1] = 1 \cdot \frac{7}{8} = \frac{7}{8}$$
$$\mathbb{E}[C] = 1 \cdot \mathbb{P}[C = 1] + 2 \cdot \mathbb{P}[C = 2] + 3 \cdot \mathbb{P}[C = 3] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{7}{4}$$

2 Pullout Balls

Note 15

Suppose you have a bag containing four balls numbered $1, 2, 3, 4$.

(a) You perform the following experiment: pull out a single ball and record its number. What is the expected value of the number that you record?

(b) You repeat the experiment from part (a), except this time you pull out two balls together and record the product of their numbers. What is the expected value of the total that you record?

Solution:

(a) Let X be the number that you record. Each ball is equally likely to be chosen, so

$$\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}[X = x] = 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} = 2.5.$$

As demonstrated here, the expected value of a random variable need not, and often is not, a feasible value of that random variable (there is no outcome ω for which $X(\omega) = 2.5$).

(b) Let Y be the product of two numbers that you pull out. Then

$$\mathbb{E}[Y] = \frac{1}{\binom{4}{2}}(1\cdot 2 + 1\cdot 3 + 1\cdot 4 + 2\cdot 3 + 2\cdot 4 + 3\cdot 4) = \frac{2+3+4+6+8+12}{6} = \frac{35}{6}.$$

3 Balls in Bins

You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(a) What is $\mathbb{E}[X_i]$?

(b) What is the expected number of empty bins?

(c) Define a collision to occur when a ball lands in a nonempty bin (if there are n balls in a bin, count that as $n-1$ collisions). What is the expected number of collisions?

Solution:

(a) We will use linearity of expectation. Note that the expectation of an indicator variable is just the probability the indicator variable = 1. (Verify for yourself that is true).

$$\begin{aligned}\mathbb{E}[X_i] &= \mathbb{P}[\text{ball 1 falls into bin } i] + \mathbb{P}[\text{ball 2 falls into bin } i] + \cdots + \mathbb{P}[\text{ball } k \text{ falls into bin } i] \\ &= \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n}.\end{aligned}$$

(b) Let I_i be the indicator variable denoting whether bin i ends up empty. This can happen if and only if all the balls end in the remaining $n-1$ bins, and this happens with a probability of $\left(\frac{n-1}{n}\right)^k$. Hence the expected number of empty bins is

$$\mathbb{E}[I_1 + \ldots + I_n] = \mathbb{E}[I_1] + \ldots + \mathbb{E}[I_n] = n\left(\frac{n-1}{n}\right)^k$$

(c) The number of collisions is the number of balls minus the number of occupied bins, since the first ball of every occupied bin is not a collision.

$$\begin{aligned}\mathbb{E}[\text{collisions}] &= k - \mathbb{E}[\text{occupied bins}] = k - n + \mathbb{E}[\text{empty locations}] \\ &= k - n + n\left(1 - \frac{1}{n}\right)^k\end{aligned}$$

4 Linearity

Solve each of the following problems using linearity of expectation. Explain your methods clearly.

(a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability $1/3$ (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability $1/5$, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

(b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears? (*Hint*: Consider where the sequence "book" can appear in the string.)

Solution:

(a) Let A_i be the indicator you win the ith time you play game A and B_i be the same for game B. The expected value of A_i and B_i are

$$\begin{aligned}\mathbb{E}[A_i] &= 1\cdot\frac{1}{3}+0\cdot\frac{2}{3}=\frac{1}{3},\\ \mathbb{E}[B_i] &= 1\cdot\frac{1}{5}+0\cdot\frac{4}{5}=\frac{1}{5}.\end{aligned}$$

Then the expected total number of tickets you receive, by linearity of expectation, is

$$3\,\mathbb{E}[A_1]+\cdots+3\,\mathbb{E}[A_{10}]+4\,\mathbb{E}[B_1]+\cdots+4\,\mathbb{E}[B_{20}]=10\left(3\cdot\frac{1}{3}\right)+20\left(4\cdot\frac{1}{5}\right)=26.$$

Note that $10\left(3\cdot\frac{1}{3}\right)$ and $20\left(4\cdot\frac{1}{5}\right)$ matches the expression directly gotten using the expected value of a binomial random variable.

(b) There are $1,000,000-4+1=999,997$ places where "book" can appear, each with a (non-independent) probability of $1/26^4$ of happening. If A is the random variable that tells how many times "book" appears, and A_i is the indicator variable that is 1 if "book" appears starting at the ith letter, then

$$\begin{aligned}\mathbb{E}[A] &= \mathbb{E}[A_1+\cdots+A_{999,997}]\\ &= \mathbb{E}[A_1]+\cdots+\mathbb{E}[A_{999,997}]\\ &= \frac{999,997}{26^4}\approx 2.19.\end{aligned}$$