1 Family Planning

Mr. and Mrs. Brown 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 Browns 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 Browns will have? What is the expected number of children that the Browns will have?

Solution:

(a) The sample space is the set of all possible sequences of children that the Browns 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 Will I Get My Package?

A delivery guy in some company is out delivering n packages to n customers, where $n\in\mathbb{N}$, $n>1$. Not only does he hand a random package to each customer, he opens the package before delivering it with probability $1/2$. Let X be the number of customers who receive their own packages unopened.

(a) Compute the expectation $\mathbb{E}(X)$.

(b) Compute the variance $\mathrm{var}(X)$.

Solution:

(a) Define

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th customer gets his/her package unopened,} \\ 0, & \text{otherwise.} \end{cases}$$

By linearity of expectation, $\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{E}(X_i)$. We have

$$\mathbb{E}(X_i) = \mathbb{P}[X_i = 1] = \frac{1}{2n},$$

since the ith customer will get his/her own package with probability $1/n$ and it will be unopened with probability $1/2$ and the delivery guy opens the packages independently.
Hence,

$$\mathbb{E}(X) = n \cdot \frac{1}{2n} = \boxed{\frac{1}{2}}.$$

(b) To calculate $\text{var}(X)$, we need to know $\mathbb{E}(X^2)$.
By linearity of expectation:

$$\mathbb{E}(X^2) = \mathbb{E}\big((X_1 + X_2 + \cdots + X_n)^2\big) = \mathbb{E}\Big(\sum_{i,j} X_i X_j\Big) = \sum_{i,j} \mathbb{E}(X_i X_j).$$

Then we consider two cases, either $i = j$ or $i \neq j$.
Hence $\sum\limits_{i,j} \mathbb{E}(X_i X_j) = \sum\limits_{i} \mathbb{E}(X_i^2) + \sum\limits_{i \neq j} \mathbb{E}(X_i X_j)$.

$$\mathbb{E}(X_i^2) = \mathbb{E}(X_i) = \frac{1}{2n}$$

for all i. To find $\mathbb{E}(X_i X_j)$, we need to calculate $\mathbb{P}[X_i X_j = 1]$.

$$\mathbb{P}[X_i X_j = 1] = \mathbb{P}[X_i = 1]\mathbb{P}[X_j = 1 \mid X_i = 1] = \frac{1}{2n} \cdot \frac{1}{2(n-1)}$$

since if customer i has received his/her own package, customer j has $n-1$ choices left.
Hence,

$$\mathbb{E}(X^2) = n \cdot \frac{1}{2n} + n \cdot (n-1) \cdot \frac{1}{2n} \cdot \frac{1}{2(n-1)} = \frac{3}{4},$$

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{3}{4} - \frac{1}{4} = \boxed{\frac{1}{2}}.$$

3 Double-Check Your Intuition Again

(a) You roll a fair six-sided die and record the result X. You roll the die again and record the result Y.

(i) What is $\operatorname{cov}(X+Y, X-Y)$?

(ii) Prove that $X+Y$ and $X-Y$ are not independent.

For each of the problems below, if you think the answer is "yes" then provide a proof. If you think the answer is "no", then provide a counterexample.

(b) If X is a random variable and $\operatorname{var}(X) = 0$, then must X be a constant?

(c) If X is a random variable and c is a constant, then is $\operatorname{var}(cX) = c\operatorname{var}(X)$?

(d) If A and B are random variables with nonzero standard deviations and $\operatorname{Corr}(A,B) = 0$, then are A and B independent?

(e) If X and Y are not necessarily independent random variables, but $\operatorname{Corr}(X,Y) = 0$, and X and Y have nonzero standard deviations, then is $\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y)$?

(f) If X and Y are random variables then is $\mathbb{E}(\max(X,Y)\min(X,Y)) = \mathbb{E}(XY)$?

(g) If X and Y are independent random variables with nonzero standard deviations, then is

$$\operatorname{Corr}(\max(X,Y), \min(X,Y)) = \operatorname{Corr}(X,Y)?$$

Solution:

(a) (i) $\operatorname{cov}(X+Y, X-Y) = \operatorname{cov}(X,X) + \operatorname{cov}(X,Y) - \operatorname{cov}(Y,X) - \operatorname{cov}(Y,Y) = \operatorname{cov}(X,X) - \operatorname{cov}(Y,Y) = 0$

(ii) Observe that $\mathbb{P}(X+Y = 7, X-Y = 0) = 0$ because if $X-Y = 0$, then the sum of our two dice rolls must be even. However both $\mathbb{P}(X+Y = 7)$ and $\mathbb{P}(X-Y = 0)$ are nonzero so $\mathbb{P}(X+Y = 7, X-Y = 0) \neq \mathbb{P}(X+Y = 7) \cdot \mathbb{P}(X-Y = 0)$

(b) Yes. If we write $\mu = \mathbb{E}[X]$, then $0 = \operatorname{var}(X) = \mathbb{E}[(X-\mu)^2]$ so $(X-\mu)^2$ must be identically 0 since perfect squares are non-negative. Thus $X = \mu$.

(c) No. We have $\operatorname{var}(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = c^2\,\mathbb{E}[(X - \mathbb{E}[X])^2] = c^2\operatorname{var}(X)$ so if $\operatorname{var}(X) \neq 0$ and $c \neq 0$ or $c \neq 1$ then $\operatorname{var}(cX) \neq c\operatorname{var}(X)$. This does prove that $\sigma(cX) = c\sigma(X)$ though.

(d) No. Let $A = X+Y$ and $B = X-Y$ from part (a). Since A and B are not constants then part (b) says they must have nonzero variances which means they also have nonzero standard deviations. Part (a) says that their covariance is 0 which means they are uncorrelated, and that they are not independent.

Recall from lecture that the converse is true though.

(e) Yes. If $\text{Corr}(X,Y) = 0$, then $\text{cov}(X,Y) = 0$. We have $\text{var}(X+Y) = \text{cov}(X+Y, X+Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X,Y) = \text{var}(X) + \text{var}(Y)$.

(f) Yes. For any values x, y we have $\max(x,y)\min(x,y) = xy$. Thus, $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$.

(g) No. You may be tempted to think that because $(\max(x,y), \min(x,y))$ is either (x,y) or (y,x), then $\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)$ because $\text{Corr}(X,Y) = \text{Corr}(Y,X)$. That reasoning is flawed because $(\max(X,Y), \min(X,Y))$ is not always equal to (X,Y) or always equal to (Y,X) and the inconsistency affects the correlation. It is possible for X and Y to be independent while $\max(X,Y)$ and $\min(X,Y)$ are not.

For a concrete example, suppose X is either 0 or 1 with probability $1/2$ each and Y is independently drawn from the same distribution. Then $\text{Corr}(X,Y) = 0$ because X and Y are independent. Even though X never gives information about Y, if you know $\max(X,Y) = 0$ then you know for sure $\min(X,Y) = 0$.

More formally, $\max(X,Y) = 1$ with probability $3/4$ and 0 with probability $1/4$, and $\min(X,Y) = 1$ with probability $1/4$ and 0 with probability $3/4$. This means

$$\mathbb{E}[\max(X,Y)] = 1 * 3/4 + 0 * 1/4 = 3/4$$

and

$$\mathbb{E}[\min(X,Y)] = 1 * 1/4 + 0 * 3/4 = 1/4.$$

Thus,

$$\text{cov}(\max(X,Y), \min(X,Y)) = \mathbb{E}[\max(X,Y)\min(X,Y)] - 3/16 = 1/4 - 3/16 = 1/16 \neq 0.$$

We conclude that $\text{Corr}(\max(X,Y), \min(X,Y)) \neq 0 = \text{Corr}(X,Y)$.