

**1. The St Petersburg Paradox.** I am running a game. I will let you flip a fair coin until you get heads. If the first head shows up on the  $k$ th flip then I will give you  $r^k$  dollars.

- (a) Compute your expected winnings when  $r = 1$ .
- (b) Compute your expected winnings when  $r = 1.5$ .
- (c) Compute your expected winnings when  $r = 2$ . Does this make any sense? How much would you be willing to pay me to play this game?

[Hint: Use the geometric series.]

We will solve all three parts at the same time. Let  $X$  be the number of coin flips until you get heads. We recall that  $X$  has a geometric distribution with

$$P(X = k) = pq^{k-1} = (1/2)(1/2)^{k-1} = (1/2)^k = 1/2^k.$$

Your winnings are a function of  $X$ , namely  $g(X) = r^X$ . Therefore the expected value of your winnings is given by

$$E[g(X)] = \sum_{k=1}^{\infty} g(k)P(X = k) = \sum_{k=1}^{\infty} r^k \cdot 1/2^k = \sum_{k=1}^{\infty} (r/2)^k.$$

Now let us recall the geometric series. If  $|x| < 1$  then we have

$$\begin{aligned} 1 + x + x^2 + x^3 + \dots &= 1/(1 - x) \\ x + x^2 + x^3 + \dots &= 1/(1 - x) - 1 = x/(1 - x). \end{aligned}$$

Therefore, if  $0 \leq r < 2$  then  $|r/2| < 1$  and the expected value of your winnings is

$$E[g(X)] = \frac{r}{2} + \left(\frac{r}{2}\right)^2 + \left(\frac{r}{2}\right)^3 + \dots = \frac{r/2}{1 - r/2} = \frac{r}{2 - r}.$$

In particular, if  $r = 1$  then you expect to win<sup>1</sup>

$$\frac{r}{2 - r} = \frac{1}{2 - 1} = 1 \text{ dollar,}$$

and if  $r = 1.5$  then you expect to win

$$\frac{r}{2 - r} = \frac{1.5}{2 - 1.5} = 3 \text{ dollars.}$$

These formulas might suggest to me how much to charge to play this game. For example, if I charge 4 dollars per play when  $r = 1.5$  then on average I will make 1 dollar per play. In fact, the original interpretation of expected value had to do with rational behavior in games of chance.

Now, how much should I charge to play the game when  $r = 2$ ? In this case  $|r/2| = 1$  so the geometric series does not converge:

$$E[g(X)] = \frac{r}{2} + \left(\frac{r}{2}\right)^2 + \left(\frac{r}{2}\right)^3 + \dots = 1 + 1 + 1 + \dots = \infty.$$

But surely no rational person would pay an infinite amount of money to play this game!

---

<sup>1</sup>Indeed, if  $r = 1$  then you will win 1 dollar no matter what happens, so your winnings are constant.

Remark: This problem was discussed by Daniel Bernoulli in 1738. It is called a “paradox” because it seems to give the “wrong answer” based on certain intuitive interpretations of the expected value. The issue is that probability is an experimental science and our mathematical concepts are only models of the real world. They give the “right answer” to the kinds of problems they were designed for, but they need not apply to every kind of problem. The mathematics discussed in this course was designed to model dice rolling and urn problems. It has a good track record when applied to physical sciences. Its relevance to social science is more dubious.

2. Let  $X$  be a random variable satisfying  $E[X] = 1$  and  $E[X^2] = 2$ . Use this to compute

- (a)  $\text{Var}(X)$
- (b)  $E[(X + 1)^2]$
- (c)  $\text{Var}(2X + 3)$

(a):  $\text{Var}(X) = E[X^2] - E[X]^2 = 2 - 1^2 = 1.$

(b):  $E[(X + 1)^2] = E[X^2 + 2X + 1] = E[X^2] + 2E[X] + 1 = 2 + 2 \cdot 1 + 1 = 5.$

Alternatively, we can use the fact that  $\text{Var}(X + 1) = \text{Var}(X)$  and hence

$$\begin{aligned}\text{Var}(X + 1) &= E[(X + 1)^2] - E[X + 1]^2 \\ \text{Var}(X) &= E[(X + 1)^2] - (E[X] + 1)^2 \\ 1 &= E[(X + 1)^2] - (1 + 1)^2 \\ 5 &= E[(X + 1)^2].\end{aligned}$$

(c): If  $a$  is constant then we recall that  $\text{Var}(aX) = a^2\text{Var}(X)$  and  $\text{Var}(X + a) = \text{Var}(X)$ . It follows from these identities that

$$\text{Var}(2X + 3) = \text{Var}(2X) = 4 \cdot \text{Var}(X) = 4 \cdot 1 = 4.$$

Alternatively, we can use the bilinearity of covariance:

$$\begin{aligned}\text{Var}(2X + 3) &= \text{Cov}(2X + 3, 2X + 3) \\ &= \text{Cov}(2X, 2X) + 2 \cdot \text{Cov}(2X, 3) + \text{Cov}(3, 3) \\ &= 4 \cdot \text{Cov}(X, X) + 2 \cdot 0 + 0 \\ &= 4 \cdot \text{Var}(X) \\ &= 4.\end{aligned}$$

**3. Standardization.** Let  $X$  be a random variable with  $E[X] = \mu_X$  and  $\text{Var}(X) = \sigma_X^2$  and consider the random variable

$$Z = \frac{X - \mu_X}{\sigma_X}.$$

- (a) Use the linearity of expectation to compute  $E[Z]$ .
- (b) Use the general properties of variance to compute  $\text{Var}(Z)$ .

(a): We use the linearity of expected value:

$$\begin{aligned}
 E[Z] &= E\left[\frac{1}{\sigma_X}(X - \mu_X)\right] \\
 &= \frac{1}{\sigma_X}E[X - \mu_X] \\
 &= \frac{1}{\sigma_X}(E[X] - \mu_X) \\
 &= \frac{1}{\sigma_X}(\mu_X - \mu_X) \\
 &= 0.
 \end{aligned}$$

(b): We use the identities  $\text{Var}(aX) = a^2\text{Var}(X)$  and  $\text{Var}(X + a) = \text{Var}(X)$  to get

$$\begin{aligned}
 \text{Var}(Z) &= \text{Var}\left(\frac{1}{\sigma_X}(X - \mu_X)\right) \\
 &= \frac{1}{\sigma_X^2}\text{Var}(X - \mu_X) \\
 &= \frac{1}{\sigma_X^2}\text{Var}(X) \\
 &= \frac{1}{\sigma_X^2} \cdot \sigma_X^2 \\
 &= 1.
 \end{aligned}$$

Remark: In the next chapter we will use this procedure to study normal random variables. If  $X$  has a “normal distribution” then it will turn out that  $Z$  has a “standard normal distribution”, and we will be able to look up the probabilities  $P(a < Z < b)$  in a table.

4. Consider a fair six-sided die with sides labeled  $\{1, 2, 3, 4, 5, 6\}$ . Roll the die twice and let

$$\begin{aligned}
 X &= \text{the number you get on the first roll,} \\
 Y &= \text{the number you get on the second roll,} \\
 Z &= X + Y.
 \end{aligned}$$

Compute the variances  $\text{Var}(X)$ ,  $\text{Var}(Y)$ ,  $\text{Var}(Z)$  and the covariances  $\text{Cov}(X, Y)$ ,  $\text{Cov}(X, Z)$ .

Here are the pmf tables for  $X$  and  $Y$ :

$k$	1	2	3	4	5	6	$k$	1	2	3	4	5	6
$P(X = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$P(Y = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Since the distributions are identical we will have  $\text{Var}(X) = \text{Var}(Y)$ . Note that

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

and

$$E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6},$$

hence

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{105}{36} = \frac{35}{12}.$$

By assumption the random variables  $X$  and  $Y$  are independent, hence  $\text{Cov}(X, Y) = 0$ .

To compute  $\text{Var}(Z)$  we can use the formula

$$\begin{aligned}\text{Var}(Z) &= \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y) \\ &= 2 \cdot \text{Var}(X) + 2 \cdot 0 \\ &= 2 \cdot \frac{35}{12} = \frac{35}{6}.\end{aligned}$$

Alternatively, we can write down the pmf table for  $Z$ :<sup>2</sup>

$k$	2	3	4	5	6	7	8	9	10	11	12
$P(Z = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Then compute the variance from this.

Finally, we use the bilinearity of covariance to compute

$$\text{Cov}(X, Z) = \text{Cov}(X, X + Y) = \text{Cov}(X, X) + \text{Cov}(X, Y) = \text{Var}(X) + 0 = \frac{35}{12}.$$

Note that this is positive because an increase in  $X$  causes an increase in  $Z$ . Alternatively, you could write out the entire  $6 \times 11$  joint pmf table for  $X, Z$  and compute the covariance from there. Wouldn't that be fun?

**5.** Let  $X, Y : S \rightarrow \mathbb{R}$  be random variables with the following joint distribution table:

$X \setminus Y$	1	2	3	
1	$1/21$	$5/21$	$3/21$	$9/21$
2	$4/21$	$2/21$	$6/21$	$12/21$
	$5/21$	$7/21$	$9/21$	

How to read the table: We have  $S_X = \{1, 2\}$  and  $S_Y = \{1, 2, 3\}$ . The entries in the right column are  $P(X = k)$ , the entries in the bottom row are  $P(Y = \ell)$  and the entries inside the table are  $P(X = k, Y = \ell)$ .

- Use the table to compute  $P(X + Y \geq 4)$ .
- Use the table to compute  $E[X]$  and  $E[Y]$ .
- Use the table to compute  $E[XY]$  and  $\text{Cov}(X, Y)$ .

(a): There are three cells of the table corresponding to  $X + Y \geq 4$ . Adding them gives

$$\begin{aligned}P(X + Y \geq 4) &= P(X = 1, Y = 3) + P(X = 2, Y = 2) + P(X = 2, Y = 3) \\ &= 3/21 + 2/21 + 6/21 \\ &= 11/21.\end{aligned}$$

(b): Reading off the marginal probabilities gives

$$E[X] = \sum_k k \cdot P(X = k) = 1 \cdot \frac{9}{21} + 2 \cdot \frac{12}{21} = \frac{33}{21}$$

<sup>2</sup>To find this you should write out all 36 elements of the sample space and count the outcomes corresponding to each value of  $Z$ .

and

$$E[Y] = \sum_k k \cdot P(Y = k) = 1 \cdot \frac{5}{21} + 2 \cdot \frac{7}{21} + 3 \cdot \frac{9}{21} = \frac{46}{21}.$$

(c): Reading off the joint probabilities gives

$$\begin{aligned} E[XY] &= \sum_{k,\ell} k\ell \cdot P(X = k, Y = \ell) \\ &= 1 \cdot 1 \cdot (1/21) + 1 \cdot 2 \cdot (5/21) + 1 \cdot 3 \cdot (3/21) \\ &\quad + 2 \cdot 1 \cdot (4/21) + 2 \cdot 2 \cdot (2/21) + 2 \cdot 3 \cdot (6/21) \\ &= 72/21, \end{aligned}$$

and hence

$$\text{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y] = (72/21) - (33/21)(46/21) = -2/147.$$

**6. Uncorrelated Does Not Imply Independent.** We say that random variables  $X, Y : S \rightarrow \mathbb{R}$  are independent if  $P(X = k, Y = \ell) = P(X = k)P(Y = \ell)$  for all possible values  $k, \ell \in \mathbb{R}$ . This property implies that  $E[XY] = E[X]E[Y]$  and hence  $\text{Cov}(X, Y) = 0$ . On the other hand, the identity  $\text{Cov}(X, Y) = 0$  does not necessarily imply that  $X$  and  $Y$  are independent. Consider the following example:

$X \setminus Y$	-1	0	1	
-1	0	0	1/4	1/4
0	1/2	0	0	1/2
1	0	0	1/4	1/4
	1/2	0	1/2	

- (a) Explain why these  $X$  and  $Y$  are **not** independent.  
 (b) Use the table to show that  $\text{Cov}(X, Y) = 0$ .

(a): These  $X$  and  $Y$  are not independent because, for example,  $P(X = -1, Y = -1) = 0$ , while  $P(X = -1)P(Y = -1) = (1/4)(1/2) = 1/8 \neq 0$ .

(b): Reading off the marginal probabilities gives

$$\begin{aligned} E[X] &= \sum_k k \cdot P(X = k) = (-1) \cdot (1/4) + 0 \cdot (1/2) + 1 \cdot (1/4) = -1/4 + 1/4 = 0, \\ E[Y] &= \sum_k k \cdot P(Y = k) = (-1) \cdot (1/2) + 0 \cdot (0) + 1 \cdot (1/2) = -1/2 + 1/2 = 0, \end{aligned}$$

and reading off the joint probabilities gives

$$\begin{aligned}
 E[XY] &= \sum_{k,\ell} k\ell \cdot P(X = k, Y = \ell) \\
 &= (-1) \cdot (-1) \cdot (0) + (-1) \cdot 0 \cdot (0) + (-1) \cdot 1 \cdot (1/4) \\
 &\quad + 0 \cdot (-1) \cdot (1/2) + 0 \cdot 0 \cdot (0) + 0 \cdot 1 \cdot (0) \\
 &\quad + (-1) \cdot (0) + 1 \cdot 0 \cdot (0) + 1 \cdot 1 \cdot (1/4) \\
 &= -1/4 + 1/4 \\
 &= 0.
 \end{aligned}$$

Hence

$$\text{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y] = 0 - 0 \cdot 0 = 0.$$

**7. Multinomial Covariance.** Suppose that a fair  $s$ -sided die is rolled  $n$  times, and let  $X_i$  be the number of times that the  $i$ th face shows up.

- Compute  $\text{Var}(X_i)$  for any  $i$ . [Hint: Think of each roll as a coin flip with  $H =$  “you get side  $i$ ” and  $T =$  “you don’t get side  $i$ ”. Use the formula for variance of a binomial.]
- Compute  $\text{Var}(X_i + X_j)$  for any  $i \neq j$ . [Hint: Think of each roll as a coin flip with  $H =$  “you get side  $i$  or  $j$ ” and  $T =$  “you get some other side”.]
- Combine (a), (b) to compute  $\text{Cov}(X_i, X_j)$ . Simplify your formula as much as possible.

(a): We can think of  $X_i$  as a binomial random variable with  $H =$  “you get side  $i$ ” and  $T =$  “you don’t get side  $i$ ”, so that  $P(H) = 1/s$  and  $P(T) = (s - 1)/s$ . From the formula for the variance of a binomial random variable we obtain

$$\text{Var}(X_i) = n \cdot P(H) \cdot P(T) = n \cdot \frac{1}{s} \cdot \frac{s-1}{s} = \frac{n(s-1)}{s^2}.$$

(b): Note that  $X_i + X_j$  is the number of times that sides  $i$  or  $j$  show up. We can think of this as a binomial random variable with  $H =$  “you get side  $i$  or  $j$ ” and  $T =$  “you get some other side”, so that  $P(H) = 2/s$  and  $P(T) = (s - 2)/s$ . From the formula for the variance of a binomial random variable we obtain

$$\text{Var}(X_i + X_j) = n \cdot P(H) \cdot P(T) = n \cdot \frac{2}{s} \cdot \frac{s-2}{s} = \frac{2n(s-2)}{s^2}.$$

(c): Combining parts (a) and (b) gives

$$\begin{aligned}
 \text{Var}(X_i) + \text{Var}(X_j) + 2\text{Cov}(X_i, X_j) &= \text{Var}(X_i + X_j) \\
 2\text{Cov}(X_i, X_j) &= \text{Var}(X_i + X_j) - \text{Var}(X_i) - \text{Var}(X_j) \\
 2\text{Cov}(X_i, X_j) &= \text{Var}(X_i + X_j) - 2 \cdot \text{Var}(X_i) \\
 &= \frac{2n(s-2)}{s^2} - \frac{2n(s-1)}{s^2} \\
 &= \frac{2n}{s^2} [(s-2) - (s-1)] \\
 &= \frac{2n}{s^2} (-1),
 \end{aligned}$$

and hence  $\text{Cov}(X_i, X_j) = -n/s^2$ . Note that this is always negative.

Discussion: We can use exactly the same method to solve the problem in the case of a biased die. Suppose that  $P(\text{side } i) = p_i$ , so that  $p_1 + p_2 + \cdots + p_s = 1$ . In this case one can show that

$$\text{Var}(X_i) = np_i(1 - p_i) \quad \text{and} \quad \text{Var}(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j),$$

so that

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \frac{1}{2} [\text{Var}(X_i + X_j) - \text{Var}(X_i) - \text{Var}(X_j)] \\ &= \frac{1}{2} [n(p_i + p_j)(1 - p_i - p_j) - np_i(1 - p_i) - np_j(1 - p_j)] \\ &= \frac{n}{2} [p_i + p_j - p_i^2 - 2p_i p_j - p_j^2 - p_i + p_i^2 - p_j + p_j^2] \\ &= \frac{n}{2} [-2p_i p_j] \\ &= -np_i p_j. \end{aligned}$$

Wow, those cancellations seemed miraculous. There must be a more clever way to do this. As always, it is easier to deal with Bernoulli random variables. Let us write

$$B_{ik} = \begin{cases} 1 & \text{if side } i \text{ shows up on the } k\text{th roll,} \\ 0 & \text{otherwise,} \end{cases}$$

so that  $E[B_{ik}] = p_i$  and  $X_i = \sum_k B_{ik} = B_{i1} + B_{i2} + \cdots + B_{in}$ . Using the linearity of the expected value gives<sup>3</sup>

$$E[X_i] = E[B_{i1}] + E[B_{i2}] + \cdots + E[B_{in}] = p_i + p_i + \cdots + p_i = np_i.$$

Then since  $B_{ik}B_{jk} = 0$  for all  $i \neq j$  and since  $B_{ik}, B_{j\ell}$  are independent for all  $k \neq \ell$  we have

$$\begin{aligned} E[X_i X_j] &= E \left[ \left( \sum_k B_{ik} \right) \left( \sum_\ell B_{j\ell} \right) \right] \\ &= E \left[ \sum_{k=\ell} B_{ik} B_{j\ell} + \sum_{k \neq \ell} B_{ik} B_{j\ell} \right] \\ &= E \left[ 0 + \sum_{k \neq \ell} B_{ik} B_{j\ell} \right] \\ &= \sum_{k \neq \ell} E[B_{ik} B_{j\ell}] \\ &= \sum_{k \neq \ell} E[B_{ik}] E[B_{j\ell}] \\ &= \sum_{k \neq \ell} p_i p_j \\ &= (n^2 - n) p_i p_j. \end{aligned}$$

In the last step we used the fact that there are  $n^2 - n$  pairs  $(k, \ell)$  where  $k \neq \ell$ . Finally,

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i] E[X_j] = (n^2 - n) p_i p_j - (np_i)(np_j) = -np_i p_j.$$

That was a bit more intuitive I think; i.e., fewer tricks required.

---

<sup>3</sup>Here we are replicating the proof that the expected value of a binomial is  $np$ .