Math 224 **Fall 2017**
Homework 4 **Drew Armstrong**

Problems from 9th edition of *Probability and Statistical Inference* by Hogg, Tanis and Zimmerman:

- Section 2.3, Exercises 16(a,d),18.
- Section 2.4, Exercises 13, 14.
- Section 4.1, Exercises 3, 4.
- Section 4.2, Exercises 3(a).
- Section 5.3, Exercises 2, 5.

Solutions to Book Problems.

2.3-16. Let X be the number of flips of a fair coin that are required to observe the same face on consecutive flips.

(a) Find the pmf of X. *Solution:* The event "$X = 1$" is empty, so that $P(X = 1) = 0$. The event "$X = 2$" consists of the sequences TT and HH so that

$$\begin{aligned} \text{“}X = 2\text{”} &= \{TT, HH\} \\ P(X = 2) &= P(TT) + P(HH) \\ &= 1/4 + 1/4 \\ &= 1/2. \end{aligned}$$

The event "$X = 3$" consists of the sequences HTT and THH so that

$$\begin{aligned} \text{“}X = 3\text{”} &= \{HTT, THH\} \\ P(X = 3) &= P(HTT) + P(THH) \\ &= 1/8 + 1/8 \\ &= 1/4. \end{aligned}$$

The event "$X = 4$" consists of the sequences $THTT$ and $HTHH$ so that

$$\begin{aligned} \text{“}X = 4\text{”} &= \{THTT, HTHH\} \\ P(X = 4) &= P(THTT) + P(HTHH) \\ &= 1/16 + 1/16 \\ &= 1/8. \end{aligned}$$

In general, the event "$X = k$" consists of exactly two sequences:

$$\underbrace{\cdots HTH}_{k-2 \text{ flips}}TT \qquad \text{and} \qquad \underbrace{\cdots THT}_{k-2 \text{ flips}}HH.$$

Since the coin is fair, each of these sequences has probability $1/2^k$, so that

$$P(X = k) = \frac{1}{2^k} + \frac{1}{2^k} = \frac{2}{2^k} = \frac{1}{2^{k-1}}.$$

The geometric series guarantees that this is, indeed, a probability mass function:

$$P(X \geq 2) = \sum_{k=2}^{\infty} \frac{1}{2^{k-1}} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1.$$

(d) Find the values of $P(X \leq 3)$, $P(X \geq 5)$ and $P(X = 3)$. *Solution:* We already saw that $P(X = 3) = 1/4$. To find $P(X \leq 3)$ we add up all the ways this can happen:

$$P(X \leq 3) = \sum_{k \leq 3} P(X = k) = P(X = 2) + P(X = 3) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

We can compute $P(X \geq 5)$ by summing a geometric series:

$$\begin{aligned} P(X \geq 5) &= \sum_{k=5}^{\infty} 1/2^{k-1} \\ &= 1/2^4 + 1/2^5 + 1/2^6 + 1/2^7 + \cdots \\ &= 1/2^4 \cdot [1 + 1/2 + 1/4 + 1/8 + \cdots] \\ &= 1/2^4 \cdot 2 \\ &= 1/2^3 \\ &= 1/8. \end{aligned}$$

Alternatively, we can compute the probability of the complement:

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 4) \\ &= 1 - [P(X = 2) + P(X = 3) + P(X = 4)] \\ &= 1 - [1/2 + 1/4 + 1/8] \\ &= 1 - 7/8 \\ &= 1/8. \end{aligned}$$

///

Remark: I didn't ask you to solve 2.3-16 (b) and (c) because we didn't talk enough about moment generating functions in class. Here are the solutions anyway. To compute the mgf of X we use another geometric series:

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \sum_{k=2}^{\infty} e^{tk} P(X = k) \\ &= \sum_{k=2}^{\infty} e^{tk} 1/2^{k-1} \\ &= e^t \cdot \sum_{k=2}^{\infty} (e^t/2)^{k-1} \\ &= e^t \cdot \left[e^t/2 + (e^t/2)^2 + (e^t/2)^3 + \cdots\right] \\ &= e^t \cdot (e^t/2) \cdot \left[1 + (e^t/2)^1 + (e^2/2)^2 + \cdots\right] \\ &= e^t \cdot (e^t/2) \cdot \frac{1}{1 - e^t/2} \\ &= \frac{e^{2t}}{2 - e^t}. \end{aligned}$$

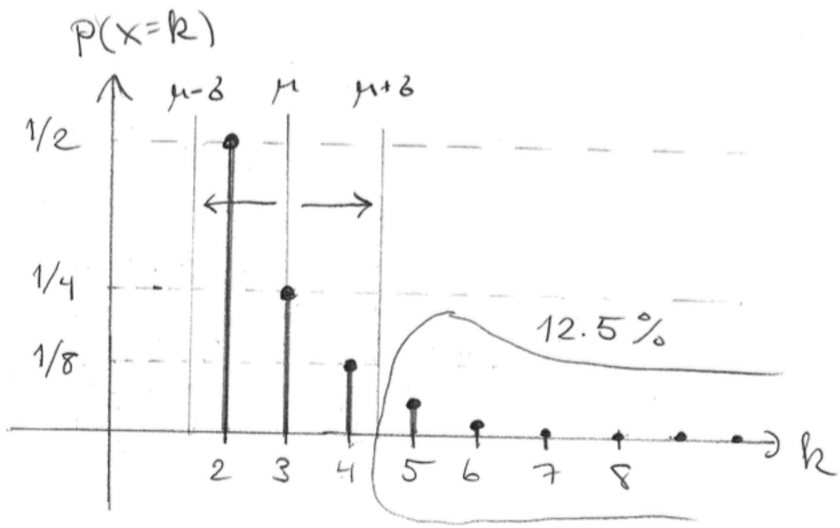
Now we can use this to compute the mean and variance. The only trick is to remember the quotient rule for derivatives:

$$\begin{aligned}
\mu = E[X] &= \frac{d}{dt} M_X(t)\Big|_{t=0} \\
&= \frac{d}{dt}\left(\frac{e^{2t}}{2-e^t}\right)\Big|_{t=0} \\
&= \frac{(2-e^t)(e^{2t})' - (e^{2t})(2-e^t)'}{(2-e^t)^2}\Big|_{t=0} \\
&= \frac{(2-e^t)(2e^{2t}) - (e^{2t})(-e^t)}{(2-e^t)^2}\Big|_{t=0} \\
&= \frac{(2-e^0)(2e^0) - (e^0)(-e^0)}{(2-e^0)^2} \\
&= \frac{(2-1)(2) - (1)(-1)}{(2-1)^2} \\
&= 3.
\end{aligned}$$

Alright, that was enough fun. My computer did the rest of the work:

$$\begin{aligned}
E[X^2] = \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} &= 11, \\
\sigma^2 = \mathrm{Var}(X) = E[X^2] - E[X]^2 &= 11 - (3)^2 = 2 \\
\sigma = \sqrt{2} &\approx 1.414.
\end{aligned}$$

Here is a picture summarizing the results of Exercise 2.3-16:



2.3-18. Let X have a geometric distribution, i.e., $P(X = k) = p(1-p)^{k-1}$. Show that for any non-negative integers j and k we have

$$P(X > k + j \mid X > k) = P(X > j).$$

Proof: We recall from HW3 that for all non-negative integers ℓ we have

$$\begin{aligned} P(X > \ell) &= \sum_{k=\ell+1}^{\infty} p(1-p)^{\ell-1} \\ &= p(1-p)^{\ell} + p(1-p)^{\ell+1} + p(1-p)^{\ell+2} + \cdots \\ &= p(1-p)^{\ell} \cdot \left[1 + (1-p)^1 + (1-p)^2 + \cdots\right] \\ &= p(1-p)^{\ell} \cdot \frac{1}{1-(1-p)} \\ &= (1-p)^{\ell}. \end{aligned}$$

Next we note that the event "$X > k + j$ and $X > k$" is the same as "$X > k + j$." Finally, we use the definition of conditional probability:

$$\begin{aligned} P(X > k + j \,|\, X > k) &= P(X > k + j \text{ and } X > k)/P(X > k) \\ &= P(X > k + j)/P(X > k) \\ &= (1-p)^{k+j}/(1-p)^k \\ &= (1-p)^j \\ &= P(X > j). \end{aligned}$$

///

What does it mean? A geometric random variable means we are waiting for something to happen. The number $P(X > j)$ is the probability that it will take at least j units of time for the thing to happen. Now suppose that we have been waiting for k units of time and the thing still hasn't happened. What is the chance that we will have to wait **at least** j **more** units of time? Answer: $P(X > j)$. Reason: A geometric random variable doesn't know how long we've been waiting because it has no memory. This is why we model it with a coin flip.

2.4-13. It is claimed that in a particular lottery, 1/10 of the 50 million tickets will win a prize. What is the probability of winning at least one prize if you purchase

(a) 10 tickets? *Solution:* Let X_i be the event defined by

$$X_i = \begin{cases} 1 & \text{if your } i\text{th ticket wins a prize,} \\ 0 & \text{if your } i\text{th ticket does not win a prize.} \end{cases}$$

These events are **not** independent. (For example, if your first ticket wins a prize, then your second ticket is slightly less likely to win a prize.) However, they are **approximately** independent because the number $50,000,000$ is so big. Therefore we will assume that $P(X_i = 1) = 1/10$ and $P(X_i = 0) = 9/10$ for all i. Under these assumptions, the total number of prizes

$$X = X_1 + X_2 + \cdots + X_{10}$$

is approximately binomial with $n = 10$ and $p = 1/10$. Therefore the probability of winning at least one prize is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1-p)^{10} = 1 - (9/10)^{10} \approx 65.13\%.$$

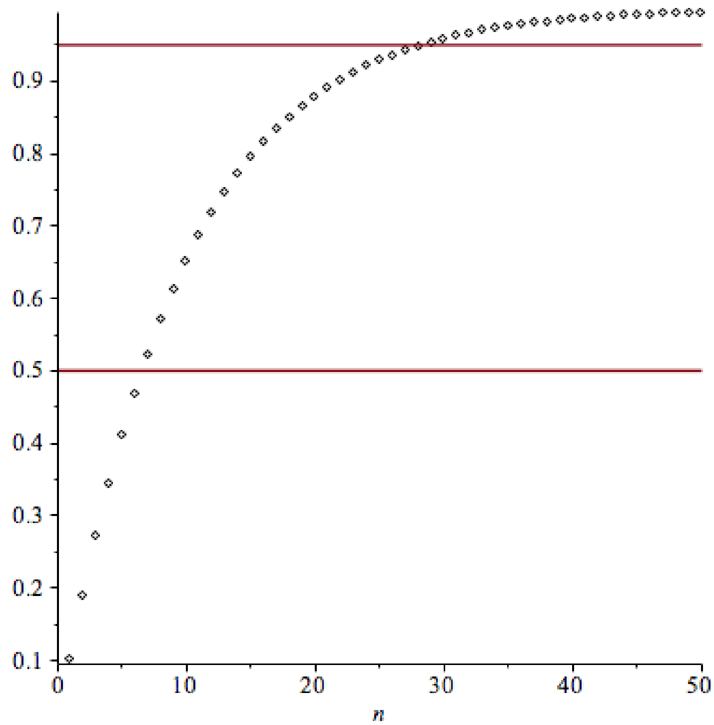
(b) 15 tickets? *Solution:* Using the same simplifying assumptions, the number of prizes X that we win is approximately binomial with $n = 15$ and $p = 1/10$. Therefore the probability of winning at least one prize is

$$P(X \geq 1) = 1 - (1-p)^{15} = 1 - (9/10)^{15} \approx 79.41\%.$$

2.4-14. Continuing from the previous problem, suppose that we buy n tickets. Then the number of prizes X that we win is approximately binomial with $p = 1/10$. (In reality it is hypergeometric.) Therefore the probability of winning at least one prize is approximately

$$P(X \geq 1) \approx 1 - (1-p)^n = 1 - (9/10)^n.$$

Here is a plot of the probability $P(X \geq 1)$ for values of n from 1 to 50:



It looks like the probability crosses 0.5 between $n = 6$ and $n = 7$, and the probability crosses 0.95 when n is around 30. To be precise, we have

$$\begin{aligned} n = 6 \quad &\to \quad P(X \geq 1) \approx 46.86\% \\ n = 7 \quad &\to \quad P(X \geq 1) \approx 52.17\% \\ n = 28 \quad &\to \quad P(X \geq 1) \approx 94.77\% \\ n = 29 \quad &\to \quad P(X \geq 1) \approx 95.29\%. \end{aligned}$$

///

Remark: In the previous two exercises we approximated the number of prizes X by a **binomial distribution** where n is the number of tickets we buy and $p = 1/10$ is the proportion of tickets that are winners. In reality X has a **hypergeometric distribution**. To see this, note that there are $5,000,000$ winning tickets and $45,000,000$ losing tickets in an urn. We reach in and grab n tickets at random. The probability of getting **exactly** k winning tickets is

$$P(X = k) = \binom{5,000,000}{k}\binom{45,000,000}{n-k} \Big/ \binom{50,000,000}{n},$$

and the probability of getting **at least one** winning ticket is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{45,000,000}{n} \Big/ \binom{50,000,000}{n}.$$

Therefore we are assuming for simplicity that

$$\binom{45,000,000}{n} \Big/ \binom{50,000,000}{n} \approx (9/10)^n.$$

It turns out that this approximation[1] is quite good for small values of n. Indeed, I ran all the calculations again with the exact formula and I got the same answers up to several decimal places.

4.1-3. Let X and Y be random variables with $S_X = \{1,2\}$ and $S_Y = \{1,2,3,4\}$ and with joint pmf given by the formula

$$f_{XY}(x,y) = \frac{x+y}{32}.$$

Solution: For (a) and (b) we draw the joint pmf as a table and then we sum the rows and columns to get the marginal pmfs:

X \ Y	1	2	3	4	
1	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{4}{32}$	$\frac{5}{32}$	$\frac{14}{32}$
2	$\frac{3}{32}$	$\frac{4}{32}$	$\frac{5}{32}$	$\frac{6}{32}$	$\frac{18}{32}$
	$\frac{5}{32}$	$\frac{7}{32}$	$\frac{9}{32}$	$\frac{11}{32}$	

For (c) through (f) we add the probabilities in the relevant cells of the table:

$$\begin{aligned}
P(X > Y) &= 3/32 \\
P(Y = 2X) &= 3/32 + 6/32 = 9/32 \\
P(X + Y = 3) &= 3/32 + 3/32 = 6/32 \\
P(X + Y \leq 3) = P(X \leq 3 - Y) &= 2/32 + 3/32 + 3/32 = 8/32.
\end{aligned}$$

(g): We note that X and Y are **not independent** because, for example, the joint probability $P(X = 1, Y = 1) = 2/32$ is not equal to the product of the marginal probabilities $P(X = 1)P(Y = 1) = (14/32)(5/32)$.

(h): We use tables to compute the 1st and 2nd moments of X and Y. Here is the table for X:

$P(X=k)$	$14/32$	$18/32$
k	1	2
k^2	1	4

$\rightarrow \quad E[X] = 1(14/32) + 2(18/32) = 25/16$

$\rightarrow \quad E[X^2] = 1(14/32) + 4(18/32) = 43/16$

[1]We'll talk more about these ideas after Exam2.

And here is the table for Y:

$P(Y=\ell)$	$\frac{5}{32}$	$\frac{7}{32}$	$\frac{9}{32}$	$\frac{11}{32}$
ℓ	1	2	3	4
ℓ^2	1	4	9	25

$$\rightarrow \quad E[Y] = 1\tfrac{5}{32} + 2\tfrac{7}{32} + 3\tfrac{9}{32} + 4\tfrac{11}{32} = 45/16$$
$$\rightarrow \quad E[Y^2] = 1\tfrac{5}{32} + 4\tfrac{7}{32} + 9\tfrac{9}{32} + 16\tfrac{11}{32} = 145/16$$

Finally, we compute the variances:

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2] - E[X]^2 = (43/16) - (25/16)^2 = 63/256, \\ \mathrm{Var}(Y) &= E[Y^2] - E[Y]^2 = (145/16) - (45/16)^2 = 295/256.\end{aligned}$$

4.1-4. Let X be a random number from the set $\{0,2,4,6,8\}$ and let Z be a random number from the set $\{0,1,2,3,4\}$. We obseve that X and Z are independent and that each possible pair of numbers has equal probability $1/5^2 = 1/25$.

Now let $Y = X + Z$. We expect that X and Y are **not independent**. To verify this we will compute the joint pmf of X and Y. First note that

$$S_Y = \{0,1,2,3,4,5,6,7,8,9,10,11,12\}.$$

We observe that each possible value of Y either has probability 0 (because it is impossible) or 1/25 (because there is exactly one way it can happen). Thus we obtain the following table showing the joint and marginal pmfs of X and Y (to save space we write $P = 1/25$):

$x \setminus y$	0	1	2	3	4	5	6	7	8	9	10	11	12	
0	P	P	P	P	P	0	0	0	0	0	0	0	0	$5P$
2	0	0	P	P	P	P	P	0	0	0	0	0	0	$5P$
4	0	0	0	0	P	P	P	P	P	0	0	0	0	$5P$
6	0	0	0	0	0	0	P	P	P	P	P	0	0	$5P$
8	0	0	0	0	0	0	0	0	P	P	P	P	P	$5P$
	P	P	$2P$	$2P$	$3P$	$2P$	$3P$	$2P$	$3P$	$2P$	$2P$	P	P	

To see that X and Y are not independent, we only need to observe, for example, that the joint probability

$$P(X=2, Y=0) = 0$$

is not equal to the product of the marginal probabilities:

$$P(X=2)P(Y=0) = (5/25)(1/25) \neq 0.$$

4.2-3(a). Roll a fair 4-sided die twice. Let X equal the outcome on the first roll and let Y equal the sum of the two rolls.

Here is a table showing the marginal and joint pmfs of X and Y (to save space we write $P = 1/16$):

$x \setminus y$	2	3	4	5	6	7	8	
1	P	P	P	P	0	0	0	$4P$
2	0	P	P	P	P	0	0	$4P$
3	0	0	P	P	P	P	0	$4P$
4	0	0	0	P	P	P	P	$4P$
	P	$2P$	$3P$	$4P$	$3P$	$2P$	P	

To compute μ_X and σ_x^2 we use the marginal distribution of X:

$$E[X] = 1(4P) + 2(4P) + 3(4P) + 4(4P) = 5/2,$$

$$E[X^2] = 1^2(4P) + 2^2(4P) + 3^2(4P) + 4^2(4P) = 15/2,$$
$$\sigma_X^2 = E[X^2] - E[X]^2 = (15/2) - (5/2)^2 = 5/4.$$

To compute μ_Y and σ_Y^2 we use the marginal distribution of Y:

$$\begin{aligned}
E[Y] &= 2(P) + 3(2P) + 4(3P) + 5(4P) + 6(3P) + 7(2P) + 8(P) = 5,\\
E[Y^2] &= 2^2(P) + 3^2(2P) + 4^2(3P) + 5^2(4P) + 6^2(3P) + 7^2(2P) + 8^2(P) = 55/2\\
\sigma_Y^2 &= E[Y^2] - E[Y]^2 = (55/2) - (5)^2 = 5/2.
\end{aligned}$$

To compute $\mathrm{Cov}(X, Y)$ we **could** use the joint pmf table to find $E[XY]$ and then compute $\mathrm{Cov}(X, Y) = E[XY] - \mu_X\mu_Y$, but there's a better way:

We will use the fact that $Y = X + Z$ where X is the number that shows up on the first roll and Z is the number that shows up on the second roll. Since Z is identically distributed with X we know that $\mathrm{Var}(Z) = \mathrm{Var}(X) = \sigma_X^2 = 5/4$, as shown above. Then we can use the fact that X and Z are **independent** to compute

$$\begin{aligned}
\mathrm{Var}(X + Y) &= \mathrm{Var}(X + X + Z)\\
&= \mathrm{Var}(2X + Z)\\
&= \mathrm{Var}(2X) + \mathrm{Var}(Z)\\
&= 2^2\mathrm{Var}(X) + \mathrm{Var}(Z)\\
&= 4(5/4) + (5/4)\\
&= 25/4.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Var}(X + Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2 \cdot \mathrm{Cov}(X, Y)\\
25/4 &= 5/4 + 5/2 + 2 \cdot \mathrm{Cov}(X, Y)\\
10/4 &= 2 \cdot \mathrm{Cov}(X, Y)\\
5/4 &= \mathrm{Cov}(X, Y),
\end{aligned}$$

and hence

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X\sigma_Y} = \frac{5/4}{\sqrt{5/4} \cdot \sqrt{5/2}} = \frac{\sqrt{2}}{2} \approx 0.707.$$

///

What does it mean? If you flip the pmf table upside-down (so it looks like a typical x, y-plane) then the diagonal cluster of P's is reasonably close to a straight line with positive slope. That's why the correlation $\rho_{XY} \approx 0.707$ is reasonably close to $+1$.

5.3-2. Let X_1 and X_2 be independent random variables with binomial distributions $b(3, 1/2)$ and $b(5, 1/2)$, respectively. Determine

(a) $P(X_1 = 2, X_2 = 4)$. *Answer:* Since X_1 and X_2 are **independent** we have

$$\begin{aligned}
P(X_1 = 2, X_2 = 4) &= P(X_1 = 2)P(X_2 = 4)\\
&= \binom{3}{2}(1/2)^2(1/2)^1\binom{5}{4}(1/2)^4(1/2)^1\\
&= (3/2^3)(5/2^3)\\
&= 15/2^6 \approx 5.86\%.
\end{aligned}$$

(b) $P(X_1+X_2=7)$. *Answer:* For this one it's helpful to draw the full pmf table. Since X_1 and X_2 are independent we compute their marginal distributions and then multiply them as follows:

$X_1 \backslash X_2$	0	1	2	3	4	5	
0	$\frac{1}{256}$	$\frac{5}{256}$	$\frac{10}{256}$	$\frac{10}{256}$	$\frac{5}{256}$	$\frac{1}{256}$	$1/8$
1	$\frac{3}{256}$	$\frac{15}{256}$	$\frac{30}{256}$	$\frac{30}{256}$	$\frac{15}{256}$	$\frac{3}{256}$	$3/8$
2	$\frac{3}{256}$	$\frac{15}{256}$	$\frac{30}{256}$	$\frac{30}{256}$	$\frac{15}{256}$	$\frac{3}{256}$	$3/8$
3	$\frac{1}{256}$	$\frac{5}{256}$	$\frac{10}{256}$	$\frac{10}{256}$	$\frac{5}{256}$	$\frac{1}{256}$	$1/8$
	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{3}$	

The event "$X_1+X_2=7$" corresponds to the two circled entries, so that

$$P(X_1+X_2=7)=\frac{3}{256}+\frac{5}{256}=\frac{8}{256}=\frac{1}{32}.$$

2.3-5. Let X_1 and X_2 be observations of a random sample of size $n=2$ from a distribution with pmf $f(x)=x/6$ and support $x=1,2,3$. Find the pmf of $Y=X_1+X_2$. Determine the mean and variance of Y in two different ways.

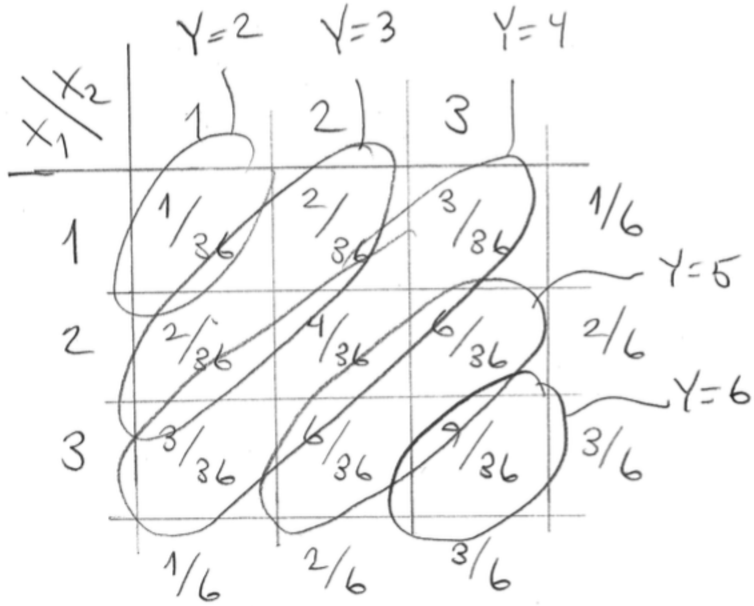
Solution: We assume that X_1 and X_2 are independent and that each has the marginal distribution

x	1	2	3
$P(X_i=x)$	$1/6$	$2/6$	$3/6$

Thus their joint pmf is given by the following table:

$X_1 \backslash X_2$	1	2	3	
1	$1/36$	$2/36$	$3/36$	$1/6$
2	$2/36$	$4/36$	$6/36$	$2/6$
3	$3/36$	$6/36$	$9/36$	$3/6$
	$1/6$	$2/6$	$3/6$	

To compute the pmf of $Y = X_1 + X_2$ we circle the event "$Y = y$" for each possible value of y:



And them we add up the probabilities in each blob to obtain:

$P(Y=y)$	1/36	4/36	10/36	12/36	9/36
y	2	3	4	5	6
y^2	4	9	16	25	36

Thus we have

$$
\begin{aligned}
E[Y] &= 2\frac{1}{36} + 3\frac{4}{36} + 4\frac{10}{36} + 5\frac{12}{36} + 6\frac{9}{36} = \frac{14}{3} \\
E[Y^2] &= 2^2\frac{1}{36} + 3^2\frac{4}{36} + 4^2\frac{10}{36} + 5^2\frac{12}{36} + 6^2\frac{9}{36} = \frac{206}{9} \\
\mathrm{Var}(Y) &= E[Y^2] - E[Y]^2 = (206/9) - (14/3)^2 = \frac{10}{9}.
\end{aligned}
$$

Alternatively, we can first compute the mean and variance of X_1 and X_2:

$$
\begin{aligned}
E[X_i] &= 1\frac{1}{6} + 2\frac{2}{6} + 3\frac{3}{6} = \frac{7}{3}, \\
E[X_i^2] &= 1^2\frac{1}{6} + 2^2\frac{2}{6} + 3^2\frac{3}{6} = 6, \\
\mathrm{Var}(X_i) &= E[X_i^2] - E[X_i]^2 = 6 - (7/3)^2 = \frac{5}{9}.
\end{aligned}
$$

And then we can use the algebraic properties of mean and variance to obtain

$$
\begin{aligned}
E[Y] &= E[X_1 + X_2] = E[X_1] + E[X_2] = \frac{7}{3} + \frac{7}{3} = \frac{14}{3}, \\
\mathrm{Var}(Y) &= \mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \frac{5}{9} + \frac{5}{9} = \frac{10}{9}.
\end{aligned}
$$

In the second equation we used the fact that X_1, X_2 are independent, which implies that $\mathrm{Cov}(X_1, X_2) = 0$. The fact that we got the same answer both times means that our answer is correct. Also, it agrees with the answer in the back of the book.

Additional Problems.

1. "Collecting Coupons." Each box of a certain brand of cereal comes with a toy. If there are n possible toys and if they are distributed randomly, how many boxes of cereal do you expect to buy before you get them all?

(a) Let X be a geometric random variable with pmf $P(X = k) = p(1-p)^{k-1}$. Use a geometric series to compute the moment generating function:

$$M(t) = E[e^{tX}] = \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} = e^t p \cdot \sum_{k=1}^{\infty} \left[e^t(1-p)\right]^{k-1} = ?$$

(b) Compute the derivative of $M(t)$ to find the expected value of X:

$$E[X] = M'(0) = ?$$

(b) Assuming that you already have ℓ of the toys, let X_ℓ be the number of boxes of cereal that you buy until you get a new toy. Observe that X_ℓ is geometric and use this fact to compute $E[X_\ell]$.

(d) Let X be the number of boxes that you buy until you see all n toys. Then we have

$$X = X_0 + X_1 + \cdots + X_{n-1}.$$

Use this to compute the expected value $E[X]$. [Hint: See Example 2.5-5 in the textbook for the case $n = 6$.]

Solution: Suppose X has pmf $P(X = k) = p(1-p)^{k-1}$ for $k \geq 1$. Then the mgf is

$$\begin{aligned}
M_X(t) = E[e^{tX}] &= \sum_{k\geq 1} kP(X = k) \\
&= \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} \\
&= e^t p \cdot \sum_{k=1}^{\infty} \left[e^t(1-p)\right]^{k-1} \\
&= e^t p \cdot \left[1 + e^t(1-p) + (e^t(1-p))^2 + \cdots\right] \\
&= e^t p \cdot \left[\frac{1}{1 - e^t(1-p)}\right] \\
&= \frac{e^t p}{1 - e^t(1-p)}.
\end{aligned}$$

To compute the expected value, we use the quotient rule to differentiate the mgf and then we substitute $t = 0$:

$$\begin{aligned}
E[X] = \left.\frac{d}{dt} M_X(t)\right|_{t=0} &= \left.\frac{d}{dt}\left(\frac{e^t p}{1 - e^t(1-p)}\right)\right|_{t=0} \\
&= \left.\frac{(1 - e^t(1-p))(e^t p)' - (e^t p)(1 - e^t(1-p))'}{(1 - e^t(1-p))^2}\right|_{t=0}
\end{aligned}$$

$$
\begin{aligned}
&= \left.\frac{(1-e^t(1-p))(e^tp)-(e^tp)(-e^t(1-p))}{(1-e^t(1-p))^2}\right|_{t=0} \\
&= \frac{(1-e^0(1-p))(e^0p)-(e^0p)(-e^0(1-p))}{(1-e^0(1-p))^2} \\
&= \frac{(1-1(1-p))(1p)-(1p)(-1(1-p))}{(1-1(1-p))^2} \\
&= \frac{(p)(p)-p(p-1)}{(p)^2} \\
&= \frac{p(p-(p-1))}{p^2} \\
&= p/p^2 \\
&= 1/p.
\end{aligned}
$$

What does it mean? Suppose you have a coin with $P(H) = p$. If you continue to flip the coin then you are most likely to see the **first head** on the $(1/p)$-th flip.

Now suppose you are collecting n random toys from cereal boxes and you already have ℓ of the toys. Let X_ℓ be the number of boxes you buy before you see a new toy. In this situation we can think of each cereal box as a "coin flip" with H="new toy" and T="old toy." Since the toys are randomly distributed this means that $P(H) = (n-\ell)/n$ and $P(T) = \ell/n$. Thus X_ℓ is a geometric random variable with $p = (n-\ell)/n$ and we conclude that

$$E[X_\ell] = \frac{1}{p} = \frac{n}{n-\ell}.$$

Finally, let X be the number of cereal boxes we buy before we see all n toys. This random variable is **not** geometric, but it is a **sum** of geometric random variables. Since X is the total number of boxes and since X_ℓ is the number of boxes from the ℓ-th toy to the $(\ell+1)$-st toy we conclude that

$$
\begin{aligned}
X &= X_0 + X_1 + X_2 + \cdots + X_{n-1} \\
E[X] &= E[X_0] + E[X_1] + E[X_2] + \cdots + E[X_{n-1}] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{n-(n-1)} \\
&= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1}.
\end{aligned}
$$

For example, suppose we continue to roll a fair $n = 6$ sided die and let X be the number of rolls until we see all six faces. Then on average we will perform

$$E[X] = \frac{6}{6} + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = 14.7 \text{ rolls}.$$

That was a fun problem.