Math 224 **Fall 2017**
Homework 3 **Drew Armstrong**

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

- Section 2.1, Exercises 6, 7, 8, 12.
- Section 2.3, Exercises 1, 3, 4, 12, 13, 14.
- Section 2.4, Exercises 12.

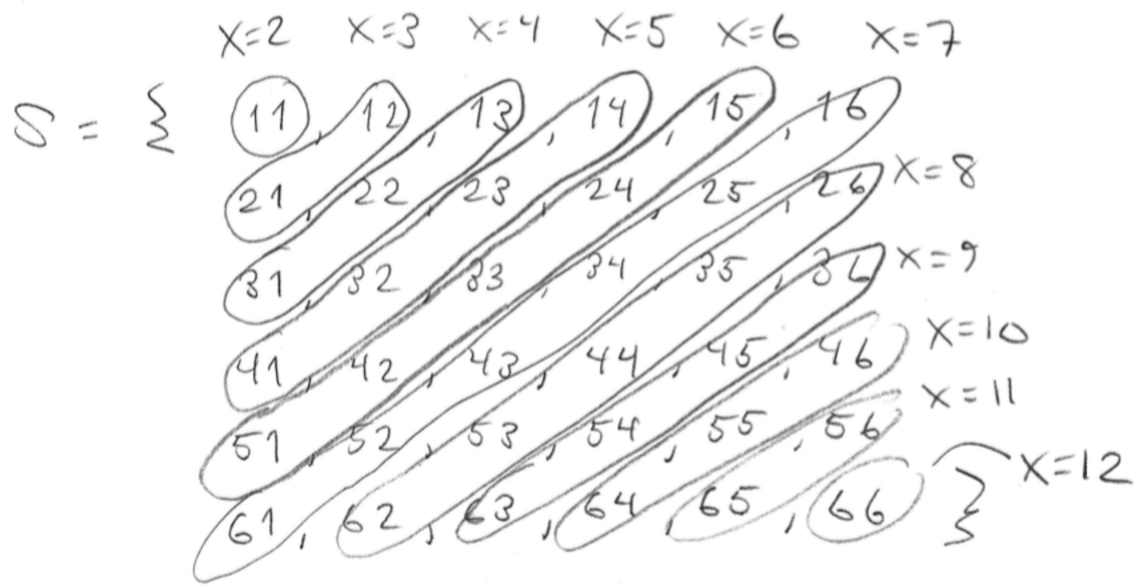
Solutions to Book Problems.

2.1-6. Throw a pair of fair 6-sided dice and let X be the sum of the two numbers that show up.

(a) The support of this random variable (i.e., the set of possible values) is

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

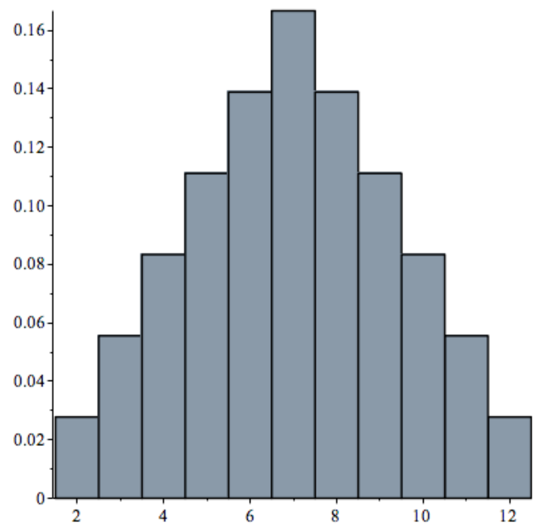
And here is the sample space S, with the events "$X = k$" for $k \in S_X$ circled:



Since the dice are fair we suppose that each of the $\#S = 6^2 = 36$ possible outcomes is equally likely. Therefore we obtain the following table showing the pmf of X:

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

(b) Here is a histogram for the probability mass function of X.



2.1-7. Roll two fair 6-sided dice and let X be the **minimum** of the two numbers that show up. Let Y be the **range** of the two outcomes, i.e., the absolute value of the difference of the two numbers that show up.

(a) The support of X is $S_X = \{1, 2, 3, 4, 5, 6\}$. Here is the sample space with the events "$X = k$" circled for each $k \in S_X$:

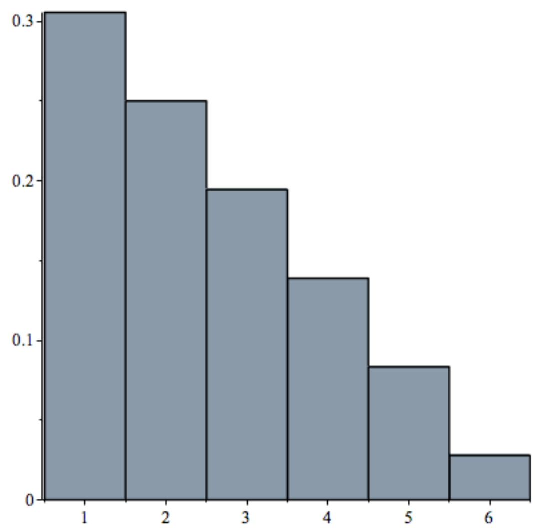
$S = \{$ 11, 12, 13, 14, 15, 16
21, 22, 23, 24, 25, 26
31, 32, 33, 34, 35, 36
41, 42, 43, 44, 45, 46
51, 52, 53, 54, 55, 56
61, 62, 63, 64, 65, 66 $\}$

X=1 X=2 X=3 X=4 X=5 X=6

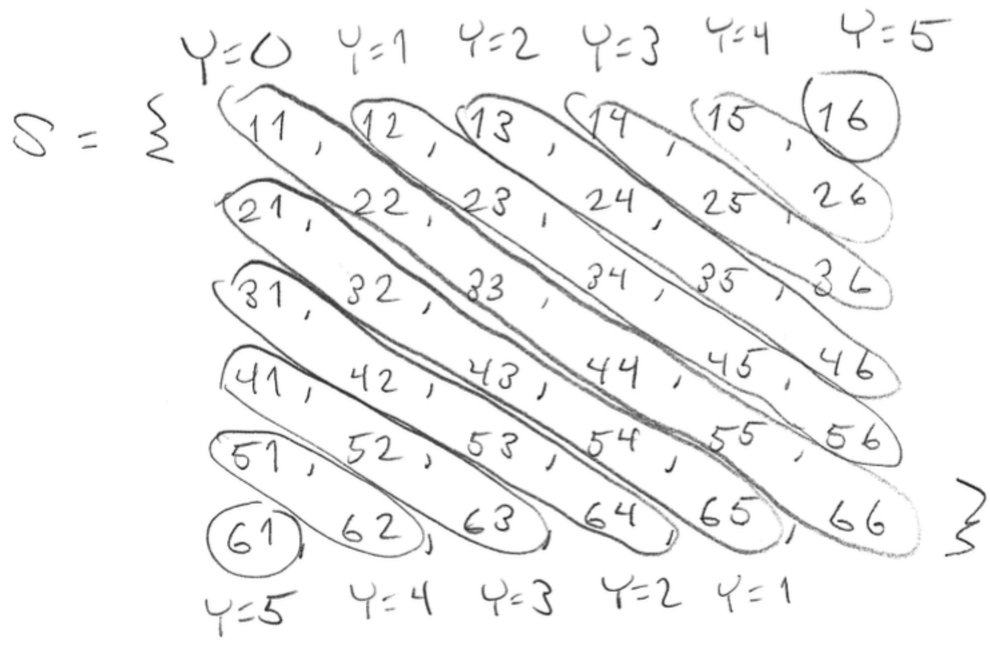
Since the $\#S = 36$ outcomes are equally likely we obtain the following table showing the pmf of X:

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

(b) And here is a histogram for the pmf of X:



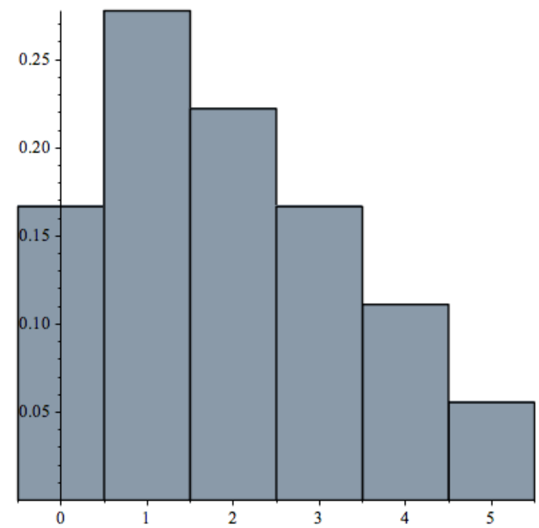
(c) The support of Y is $S_Y = \{0, 1, 2, 3, 4, 5\}$. Here is the sample space with the events "$Y = k$" circled for each $k \in S_Y$:



Since the $\#S = 36$ outcomes are equally likely we obtain the following table showing the pmf of Y:

k	0	1	2	3	4	5
$P(Y = k)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

(d) And here is a histogram for the pmf of Y.



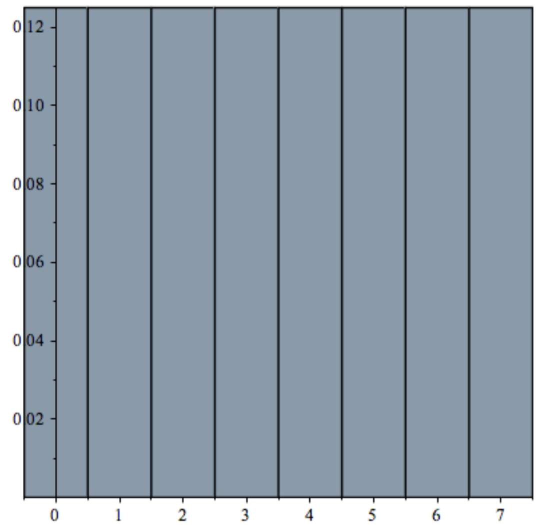
2.1-8. A fair 4-sided die has faces numbered $0, 0, 2, 2$. You roll the die and let X be the number that shows up. Another fair 4-sided die has faces numbered $0, 1, 4, 5$. You roll the die and let Y be the number that shows up. Let $W = X + Y$.

(a) The support of W is $S_W = \{0, 1, 2, 3, 4, 5, 6, 7\}$. Here is the sample space S with the events "$W = k$" circled for each $k \in S_W$:

Since the $\#S = 4^2 = 16$ outcomes are equally likely we obtain the following table showing the pmf of W:

k	0	1	2	3	4	5	6	7
$P(W = k)$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$

(b) And here is a histogram for the pmf of W:



2.1-12. Let X be the number of accidents per week in a factory and suppose that the pmf of X is given by

$$f_X(k) = P(X = k) = \frac{1}{(k+1)(k+2)} = \frac{1}{k+1} - \frac{1}{k+2} \qquad \text{for } k = 0, 1, 2, \ldots$$

Find the conditional probability of $X \geq 4$, given that $X \geq 1$.

Solution: Let $A =$ "$X \geq 4$" and $B =$ "$X \geq 1$" and note that $A \subseteq B$, which implies that $A \cap B = A$. Now we are looking for the conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{P(X \geq 4)}{P(X \geq 1)}.$$

To compute $P(X \geq 4)$ and $P(X \geq 1)$, let us first investigate why $P(S) = P(X \geq 0) = 1$. This is because we have a "telescoping" infinite series:

$$\begin{aligned}
P(X \geq 0) &= P(X = 0) + P(X = 1) + P(X = 2) + \cdots \\
&= \left(1 - \cancel{\frac{1}{2}}\right) + \left(\cancel{\frac{1}{2}} - \cancel{\frac{1}{3}}\right) + \left(\cancel{\frac{1}{3}} - \cancel{\frac{1}{4}}\right) + \cdots \\
&= 1 + 0 + 0 + 0 + \cdots \\
&= 1.
\end{aligned}$$

The same idea shows us that $P(X \geq n) = 1/(n+1)$ for any n. Indeed, we have

$$\begin{aligned}
P(X \geq n) &= P(X = n) + P(X = n+1) + P(X = n+2) + \cdots \\
&= \left(\frac{1}{n+1} - \cancel{\frac{1}{n+2}}\right) + \left(\cancel{\frac{1}{n+2}} - \cancel{\frac{1}{n+3}}\right) + \left(\cancel{\frac{1}{n+3}} - \cancel{\frac{1}{n+4}}\right) + \cdots \\
&= \frac{1}{n+1} + 0 + 0 + 0 + \cdots \\
&= \frac{1}{n+1}.
\end{aligned}$$

Finally, we conclude that

$$P(X \geq 4 \,|\, X \geq 1) = \frac{P(X \geq 4)}{P(X \geq 1)} = \frac{1/5}{1/2} = \frac{2}{5}.$$

///

2.3-1. Find the mean and variance of the following distributions:

(a) $f(k) = 1/5$ for $k = 5, 10, 15, 20, 25$. *Solution:* The mean is

$$\mu = \sum_k k \cdot f(k) = 5 \cdot \frac{1}{5} + 10 \cdot \frac{1}{5} + 15 \cdot \frac{1}{5} + 20 \cdot \frac{1}{5} + 25 \cdot \frac{1}{5} = \frac{75}{5} = 15.$$

To compute the variance we will use the following table:

k	5	10	15	20	25
μ	15	15	15	15	15
$(k-\mu)^2$	100	25	0	25	100
$f(k)$	1/5	1/5	1/5	1/5	1/5

Then we have

$$\sigma^2 = \sum_k (k-\mu)^2 \cdot f(k) = 100 \cdot \frac{1}{5} + 25 \cdot \frac{1}{5} + 0 \cdot \frac{1}{5} + 25 \cdot \frac{1}{5} + 100 \cdot \frac{1}{5} = \frac{250}{5} = 50.$$

(b) $f(x) = 1$ for $k = 5$. *Solution:* The mean is

$$\mu = \sum_k k \cdot f(k) = 5 \cdot 1 = 5$$

and the variance is

$$\sigma^2 = \sum_k (k-\mu)^2 \cdot f(k) = (5-5)^2 \cdot 1 = 0.$$

(c) $f(k) = (4-k)/6$ for $k = 1, 2, 3$. *Solution:* The mean is

$$\mu = \sum_k k \cdot f(k) = 1 \cdot \frac{4-1}{6} + 2 \cdot \frac{4-2}{6} + 3 \cdot \frac{4-3}{6} = \frac{13}{6} = \frac{10}{6} = \frac{5}{3}.$$

To compute the variance we use the following table:

k	1	2	3
μ	5/3	5/3	5/3
$(k-\mu)^2$	4/9	1/9	16/9
$f(k)$	3/6	2/6	1/6

Then we have

$$\sigma^2 = \sum_k (k-\mu)^2 \cdot f(k) = \frac{4}{9} \cdot \frac{3}{6} + \frac{1}{9} \cdot \frac{2}{6} + \frac{16}{9} \cdot \frac{1}{6} = \frac{30}{54} = \frac{5}{9}.$$

2.3-3. Given $E(X = 4) = 10$ and $E[(X+4)^2] = 116$, determine

(a) $\mathrm{Var}(X+4)$ *Solution:* Let $Y = X + 4$. Then

$$\begin{aligned} \mathrm{Var}(Y) &= E[Y^2] - E[Y]^2 \\ \mathrm{Var}(X+4) &= E[(X+4)^2] - E[X+4]^2 \\ &= 116 - 10^2 \\ &= 16. \end{aligned}$$

(b) $\mu = E[X]$ *Solution:* Since 4 is **constant** (i.e., it is not random) we have $E[4] = 4$ and then by linearity of expectation we have

$$\begin{aligned} E[X + 4] &= 10 \\ E[X] + E[4] &= 10 \\ E[X] + 4 &= 10 \\ E[X] &= 6. \end{aligned}$$

(c) $\sigma^2 = \text{Var}(X)$ *Solution:* We want to compute

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

and we already know that $E[X] = 6$, so it only remains to compute $E[X^2]$. To do this, we use linearity:

$$\begin{aligned} E[(X + 4)^2] &= 116 \\ E[X^2 + 8X + 16] &= 116 \\ E[X^2] + 8E[X] + 16 &= 116 \\ E[X^2] + 8 \cdot 6 &= 100 \\ E[X^2] + 48 &= 100 \\ E[X^2] &= 52. \end{aligned}$$

We conclude that

$$\sigma^2 = \text{Var}(X) = E[X^2] - E[X]^2 = 52 - 6^2 = 56 - 36 = 16,$$

and hence also $\sigma = \sqrt{16} = 4$.

2.3-4. Let μ and σ^2 be the mean and variance of the random variable X. Determine $E[(X - \mu)/\sigma]$ and $E\left[((X - \mu)/\sigma)^2\right]$

Solution: Since μ and σ are **constant** (i.e., not random) we use linearity to get

$$\begin{aligned} E\left[\frac{X - \mu}{\sigma}\right] &= E\left[\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right] \\ &= \frac{1}{\sigma}E[X] - E\left[\frac{\mu}{\sigma}\right] \\ &= \frac{1}{\sigma} \cdot \mu - \frac{\mu}{\sigma} = 0. \end{aligned}$$

Similarly, we have

$$E\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] = E\left[\frac{1}{\sigma^2}(X - \mu)^2\right] = \frac{1}{\sigma^2}E\left[(X - \mu)^2\right] = \frac{1}{\sigma^2} \cdot \sigma^2 = 1.$$

We conclude that the random variable $(X - \mu)/\sigma$ has mean 0 and variance 1 (hence also standard deviation 1). I wonder if that's useful for something. ///

2.3-12. Let X be the number of people selected at random that you must ask in order to find someone with the same birthday as yours. (Assume that each day of the year is equally likely and ignore February 29.)

(a) What is the pmf of X? *Answer:* Assuming that people's birthdays are independent, we can treat each person as a coin flip with

$$\begin{aligned} H &= \text{"same birthday as yours,"} \\ T &= \text{"different birthday."} \end{aligned}$$

Since all birthdays are equally likely we have $P(H) = 1/365$. Then the occurrence of the "first head" is a geometric random variable with pmf

$$P(X = k) = P(T)^{k-1}P(H) = \left(\frac{364}{365}\right)^{k-1}\left(\frac{1}{365}\right) = \frac{364^{k-1}}{365^k}.$$

(b) Give the values of the mean, variance, and standard deviation of X. *Answer:* If X is a geometric random variable with probability of heads $P(H) = p$ then the table in the front of the book tells us that

$$\mu = \frac{1}{p} \qquad \text{and} \qquad \sigma^2 = \frac{1-p}{p^2}.$$

In our case we have

$$\mu = \frac{1}{1/365} = 365 \qquad \text{and} \qquad \sigma^2 = \frac{364/365}{(1/365)^2} = 364 \cdot 365 = 132860,$$

hence also $\sigma = \sqrt{132860} \approx 364.5$.

(c) Find $P(X > 400)$ and $P(X < 300)$. *Answer:* In general, for a geometric random variable X and an integer n we have

$$\begin{aligned} P(X > n) &= \sum_{k>n} P(X = k) \\ &= \sum_{k=n+1}^{\infty} (1-p)^{k-1}p \\ &= (1-p)^n p + (1-p)^{n+1}p + (1-p)^{n+2}p + \cdots \\ &= (1-p)^n p\left(1 + (1-p) + (1-p)^2 + (1-p)^3 + \cdots\right) \\ &= (1-p)^n p \cdot \frac{1}{1-(1-p)} \\ &= (1-p)^n p \cdot \frac{1}{p} \\ &= (1-p)^n. \end{aligned}$$

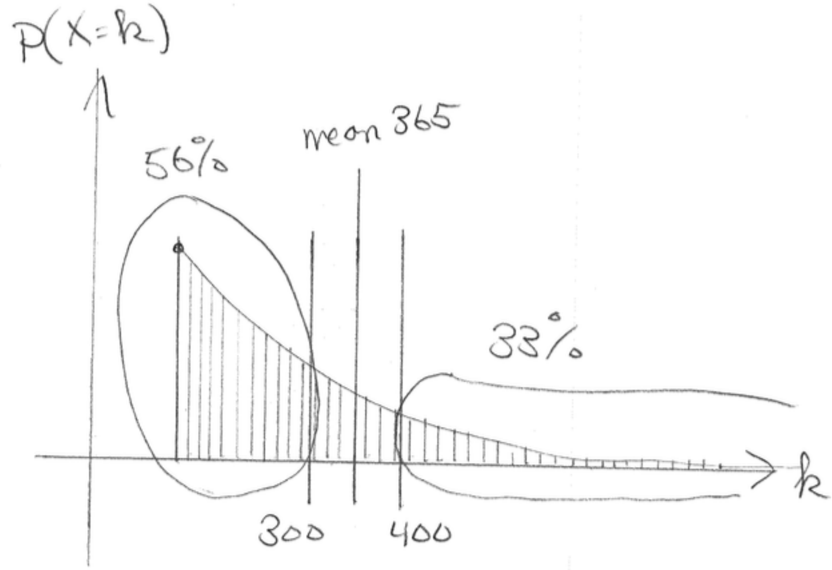
Therefore in our case we have

$$P(X > 400) = (1-p)^{400} = \left(\frac{364}{365}\right)^{400} \approx 33.37\%$$

and

$$P(X < 300) = 1 - P(X \geq 300) = 1 - P(X > 299) = 1 - \left(\frac{364}{365}\right)^{299} \approx 55.97\%.$$

Remark: Here is a picture illustrating the results of the previous problem:



The distribution is centered at $X = 365$, but we observe that there isn't much chance of being close to the mean. This is because the distribution is **very spread out**, as confirmed by the large standard deviation $\sigma \approx 364.5$.

2.3-13. For each question on a multiple-choice test, there are five possible answers, of which exactly one is correct. If a student selects answers at random, give the probability that the first question answered correctly is question 4.

Solution: We can think of each question as a coin flip with $P(H) = 1/5$. If X is the first question answered correctly then X is a geometric random variable with pmf

$$P(X = k) = (1-p)^{k-1}p = \left(\frac{4}{5}\right)^{k-1}\left(\frac{1}{5}\right) = \frac{4^{k-1}}{5^k}.$$

Thus the probability that the first correct answer occurs on question 4 is

$$P(X = 4) = \frac{4^3}{5^4} = 10.24\%.$$

2.3-14. The probability that that a machine produces a defective item is 0.01. Each item is checked as it is produced. Assume that these are independent trials, and compute the probability that at least 100 items must be checked to find one that is defective.

Solution: Let X be the occurence of the first defective item. This is a geometric random variable with $p = 0.01$. The probability we are looking for is

$$P(X \geq 100) = P(X > 99) = (1-p)^{99} = (0.99)^{99} \approx 36.97\%.$$

2.4-12. In a certain casino game, three fair 6-sided dice are rolled. We can think of each die as a "coin flip" with "heads"="a 5 shows up." Let Y be the number of heads that shows up, which has a binomial pmf:

$$P(Y=k) = \binom{3}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{3-k}.$$

The support of Y is $S_Y = \{0,1,2,3\}$ and your winnings in the game are determined by another random variable X, which is defined by

$$X = \begin{cases} -\$1 & \text{if } Y=0 \\ \$1 & \text{if } Y=1 \\ \$2 & \text{if } Y=2 \\ \$3 & \text{if } Y=3 \end{cases}$$

(a) The support of X is $S_X = \{-1,1,2,3\}$ and the following table shows the pmf of X:

k	-1	1	2	3
$P(X=k)$	$\frac{5^3}{6^3}$	$3\cdot\frac{5^2}{6^3}$	$3\cdot\frac{5}{6^3}$	$\frac{1}{6^3}$

(b) Therefore the mean of X (i.e., your expected winnings) is

$$\begin{aligned}\mu = E[X] &= \sum_{k\in S_X} kP(X=k) \\ &= -1\cdot\frac{5^3}{6^3} + 1\cdot 3\cdot\frac{5^2}{6^3} + 2\cdot 3\cdot\frac{5}{6^3} + 3\cdot\frac{1}{6^3} \\ &= \frac{-5^3+3\cdot 5^2+6\cdot 5+3}{6^3} = \frac{-17}{216} \approx -\$0.0787.\end{aligned}$$

That is, you should expect to lose about 8 cents on this game. To compute the variance we use the following table:

k	-1	1	2	3
$(k-\mu)^2$	$\frac{39601}{46656}$	$\frac{54289}{46656}$	$\frac{201601}{46656}$	$\frac{442225}{46656}$
$P(X=k)$	$\frac{5^3}{6^3}$	$3\cdot\frac{5^2}{6^3}$	$3\cdot\frac{5}{6^3}$	$\frac{1}{6^3}$

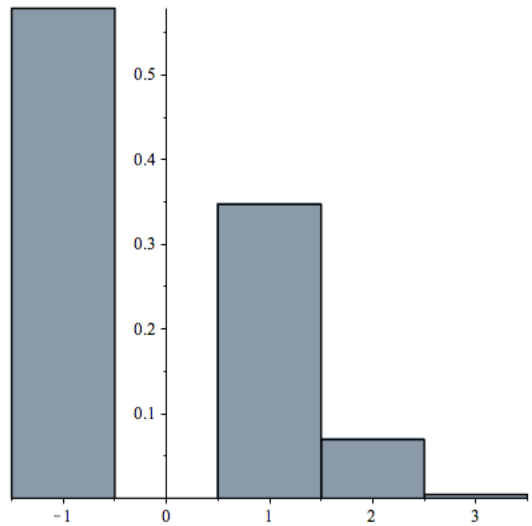
We find that

$$\sigma^2 = \mathrm{Var}(X) = \sum_{k\in S_X}(k-\mu)^2 P(X=k) = \frac{57815}{46656} \approx 1.239,$$

and hence the standard deviation is

$$\sigma = \sqrt{57815/46656} \approx 1.113.$$

(c) And here is a histogram for the pmf of your winnings:



Additional Problems.

1. Two Formulas for Expectation. Let S be the sample space of an experiment (assume that S is finite) and let $X : S \to \mathbf{R}$ be any random variable. Let $S_X \subseteq \mathbb{R}$ be the "support" of X, i.e., the set of possible values that X can take. Explain why the following formula is true:

$$\sum_{s\in S} X(s)\cdot P(s) = \sum_{k\in S_X} k\cdot P(X=k).$$

Solution: We can think of "$X = k$" $\subseteq S$ as the event consisting of all outcomes $s \in S$ with the property $X(s) = k$. That is, we have

$$\text{“}X=k\text{”} = \{s\in S : X(s)=k\}.$$

Then applying the probability measure to both sides gives

$$\boxed{P(X=k) = \sum_{s\in \text{“}X=k\text{”}} P(s).}$$

We will apply this boxed formula to prove the identity. The key idea is to replace the sum $\sum_{s\in S}$ with the equivalent double sum $\sum_{k\in S_X}\sum_{s\in \text{“}X=k\text{”}}$. That is, instead of summing over all $s \in S$, we first sum over all $k \in S_X$ and for each fixed k we sum over all $s \in$ "$X = k$", which amounts to the same thing. Here's the proof:[1]

$$\begin{aligned}\sum_{s\in S} X(s)P(s) &= \sum_{k\in S_X}\sum_{s\in \text{“}X=k\text{”}} X(s)P(s)\\ &= \sum_{k\in S_X}\sum_{s\in \text{“}X=k\text{”}} kP(s)\\ &= \sum_{k\in S_X} k\left(\sum_{s\in \text{“}X=k\text{”}} P(s)\right)\end{aligned}$$

[1] I will never ask you to write a proof like this on an exam. The only reason I assigned this problem is because I wanted you to remember the two formulas for expectation. Hopefully the extra exposure will increase your chances of recall.

$$= \sum_{k \in S_X} kP(X = k).$$

///

2. Expected Value of a Binomial Random Variable.

(a) Use the explicit formula for binomial coefficients to prove that

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

Solution 1: The right hand side is given by

$$n\binom{n-1}{k-1} = \frac{n(n-1)!}{(k-1)!\,[(n-1)-(k-1)]!} = \frac{n!}{(k-1)!(n-k)!}$$

and the left hand side is given by

$$k\binom{n}{k} = \frac{k}{k!} \cdot \frac{n!}{(n-k)!} = \frac{1}{(k-1)!} \cdot \frac{n!}{(n-k)!} = \frac{n!}{(k-1)!(n-k)!}.$$

Note that the formulas are the same. ///

Solution 2: A club of k people is selected from a classroom of n students. One member of the club is chosen to be the president. In how many ways can we choose the club? One the one hand, there are $\binom{n}{k}$ ways to choose the club members and then k ways to choose the president from the club members, for a total of

$$k\binom{n}{k} \qquad \text{choices.}$$

On the other hand, we could choose one student from the class to be the president of the club. There are n ways to do this. Then, we choose $k-1$ of the remaining $n-1$ students to be the other members of the club. There are $\binom{n-1}{k-1}$ ways to do this, for a total of

$$n\binom{n-1}{k-1} \qquad \text{choices.}$$

Since these two formulas count the same thing, they must be equal. ///

(b) Use part (a) to compute the expected value of a binomial random variable:

$$\begin{aligned}
E[X] &= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \boxed{k\binom{n}{k}} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \boxed{n\binom{n-1}{k-1}} p^k (1-p)^{n-k} \\
&= n\sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= n\sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-(\ell+1)}
\end{aligned}$$

$$
\begin{aligned}
&= np \cdot \boxed{\sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell}(1-p)^{(n-1)-\ell}} \\
&= np \cdot \boxed{(p + (1-p))^{n-1}} \\
&= np \cdot 1^{n-1} \\
&= np.
\end{aligned}
$$

///

Remark: In class we saw an alternate way to compute the expected value of a binomial random variable. Flip a coin n times and let X_i be the "number of heads" that occur on the ith coin flip. That is, let

$$
X_i = \begin{cases} 1 & \text{if the } i\text{th flip is } H, \\ 0 & \text{if the } i\text{th flip is } T. \end{cases}
$$

Each X_i has a so-called *Bernoulli distribution*, with pmf given by the following table:

k	0	1
$P(X_i = k)$	$1-p$	p

And hence each X_i has expected value

$$
E[X_i] = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = 0 \cdot (1-p) + 1 \cdot p = p.
$$

By adding the random variables $X_1, X_2, \ldots, X_n$ we obtain a binomial random variable:

$$
\begin{aligned}
(\text{total \# heads}) &= (\text{\# heads on 1st flip}) + \cdots + (\text{\# heads on } n\text{th flip}) \\
X &= X_1 + X_2 + \cdots + X_n.
\end{aligned}
$$

And then the linearity of expectation gives

$$
\begin{aligned}
E[X] &= E[X_1 + X_2 + \cdots + X_n] \\
&= E[X_1] + E[X_2] + \cdots + E[X_n] \\
&= p + p + \cdots + p \\
&= np.
\end{aligned}
$$

Which method do you like better?