Math 224 **Spring 2022**
Homework 3 **Drew Armstrong**

---

1. Functions of a Random Variable. Let X be the number of heads obtained in 3 flips of a fair coin. Compute the following expected values:

(a) $E[X^2]$,
(b) $E[2^X]$,
(c) $E[e^{tX}]$, where t is some constant.

[Hint: The general formula is $E[g(X)] = \sum_k g(k) \cdot P(X = k)$.]

First we compile a table of the pmf:

k	0	1	2	3
$P(X = k)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Then we use the formula for $E[g(X)]$ to compute the expected values.

(a): In this case we have $g(X) = X^2$:

$$E[X^2] = \sum_k k^2 \cdot P(X = k) = 0^2 \left(\frac{1}{8}\right) + 1^2 \left(\frac{3}{8}\right) + 2^2 \left(\frac{3}{8}\right) + 3^2 \left(\frac{1}{8}\right) = 3$$

(b): In this case we have $g(X) = 2^X$:

$$E[2^X] = \sum_k 2^k \cdot P(X = k) = 2^0 \left(\frac{1}{8}\right) + 2^1 \left(\frac{3}{8}\right) + 2^2 \left(\frac{3}{8}\right) + 2^3 \left(\frac{1}{8}\right) = \frac{27}{8}$$

(b): In this case we have $g(X) = e^{tX}$ where t is an unknown constant:

$$\begin{aligned} E[e^{tX}] &= \sum_k e^{tk} \cdot P(X = k) = e^{0t} \left(\frac{1}{8}\right) + e^t \left(\frac{3}{8}\right) + e^{2t} \left(\frac{3}{8}\right) + e^{3t} \left(\frac{1}{8}\right) \\ &= \frac{1}{8}\left(1 + 3e^t + 3e^{2t} + e^{3t}\right) \end{aligned}$$

This formula does not simplify any further.

Remark: The expression $E[e^{tX}]$ might seem random, but in fact this is an important concept called the *moment generating function* of the random variable X:

$$M_X(t) = E[e^{tX}] = \sum_k e^{tk} P(X = k).$$

The reason for the name is as follows. You may recall the power series expansion of the exponential function:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots + \frac{x^n}{n!} + \cdots .$$

If we substitute tX for t then this becomes

$$e^{tX} = 1 + Xt + \frac{X^2}{2}t^2 + \frac{X^3}{6}t^3 + \cdots + \frac{X^n}{n!}t^n + \cdots .$$

Then using the linearity of expectation gives

$$E[e^{tX}] = 1 + E[X]t + \frac{E[X^2]}{2}t^2 + \frac{E[X^3]}{6}t^3 + \cdots + \frac{E[X^n]}{n!}t^n + \cdots .$$

Thus the coefficients of t in the expansion of $E[e^{tX}]$ are essentially just the moments of X:

$$E[X], E[X^2], E[X^3], \ldots.$$

Sometimes this is the easiest way to compute the moments, especially when X is related to the normal distribution. We will return to this in Chapter 3.

The next problem uses a different kind of generating function, called the *probability generating function*. It is particularly useful for studying binomial random variables.

2. Tricks With the Binomial Theorem. Let X be a binomial random variable with parameters (n, p) and let z be any real number. The following function $G(z)$ is called the *probability generating function of* X:

$$G(z) = \sum_{k=0}^{n} P(X = k) \cdot z^k.$$

(a) Prove that $G'(1) = E[X]$. [This is the derivative of $G(z)$ evaluated at $z = 1$.]
(b) Use the Binomial Theorem to prove that $G(z) = (pz + q)^n$.
(c) Combine parts (a) and (b) to prove that $E[X] = np$.

(a): The derivative with respect to z is

$$G'(z) = \sum_{k=0}^{n} P(X = k) \cdot kz^{k-1}.$$

Then substituting $z = 1$ gives

$$G'(1) = \sum_{k=0}^{n} P(X = k) \cdot k1^{k-1} = \sum_{k=0}^{n} k \cdot P(X = k) = E[X].$$

(b): Since X is binomial we know that $P(X = k) = \binom{n}{k} p^k q^{n-k}$. And the Binomial Theorem says that for any numbers a, b we have

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Hence substituting $a = pz$ and $b = q$ gives

$$\begin{aligned} (pz + q)^n &= \sum_{k=0}^{n} \binom{n}{k} (pz)^k q^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \cdot z^k \\ &= \sum_{k=0}^{n} P(X = k) \cdot z^k \\ &= G(z). \end{aligned}$$

(c): First we use (b) to compute the derivative of $G(z)$ with respect to z:

$$G'(z) = n(pq + q)^{n-1} \cdot (p1 + 0) = np(pz + q)^{n-1}.$$

Then we use (a) to compute the expected value:

$$E[X] = G'(1) = np(p1 + q)^{n-1} = np(1)^{n-1} = np.$$

Remark: This is often the proof given in probability textbooks. Personally I find it too tricky,[1] and the concept of the probability generating functions turns out not to be very useful for other random variables. The moment generating function, discussed after Problem 1, is a much more useful concept.

3. Sums of Random Variables. A fair coin has sides labeled $\{1,3\}$. Suppose that you flip the coin three times and consider the random variables X_1, X_2, X_3 defined by

$$X_i = \text{the number that shows up on the } i\text{th coin flip.}$$

(a) Compute $E[X_1]$, $E[X_2]$ and $E[X_3]$.
(b) Let $X = X_1 + X_2 + X_3$ be the sum of the three numbers you get. Compute the probability mass function $P(X = k)$ and draw the probability histogram of X. [Hint: There is no shortcut. You must write down all 8 elements of the sample space.]
(c) Compute $E[X]$ using the pmf for X from part (b).
(d) Compute $E[X]$ using linearity and your answer from part (a).

(a): For any i we have $P(X_i = 1) = 1/2$ and $P(X_i = 3) = 1/2$, hence

$$E[X_i] = 1P(X_i = 1) + 3P(X_i = 3) = 1(1/2) + 3(1/2) = 2.$$

On average, we expect to see the number 2 on any particular coin flip. We will return to this in part (d).

(b): Consider the 8 elements of the sample space:

$$S = \{111, 113, 131, 311, 133, 313, 331, 333\}.$$

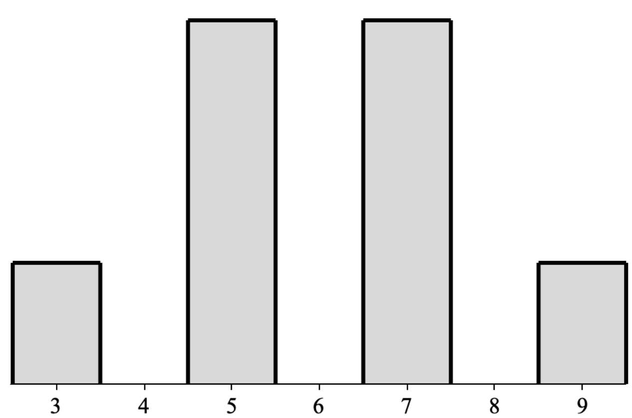
Here are the corresponding values of X:

s	111	113	131	311	133	313	331	333
$X(s)$	3	5	5	5	7	7	7	9

Since the outcomes are equally likely (the coin is fair) we obtain the pmf table of X:

k	3	5	7	9
$P(X = k)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Here is the histogram:



[1] The proof using linearity and expressing X as a sum of Bernoulli random variables is much more intuitive and much more useful for other situations.

(c): We can see from the symmetry of the histogram that $E[X] = 6$, but let's check it anyway:

$$\begin{aligned} E[X] &= 3P(X = 3) + 5P(X = 5) + 7P(X = 7) + 9P(X = 9) \\ &= 3(1/8) + 5(3/8) + 7(3/8) + 9(1/8) \\ &= 48/8 \\ &= 6. \end{aligned}$$

(d): Instead of going through all the work in part (b) we could have computed the expected value directly from (a) using linearity. Recall that $X = X_1 + X_2 + X_3$ and $E[X_1] = E[X_2] = E[X_3] = 2$, hence

$$E[X] = E[X_1] + E[X_2] + E[X_3] = 2 + 2 + 2 = 6.$$

Remark: The advantage of linearity is that it can be extended to any number of coin flips. Suppose you flip the coin 100 times. Let X_i be the number you obtain on the ith flip and let $X = X_1 + \cdots + X_{100}$ be the sum. Then the expected value of X is

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_{100}] = 2 + 2 + \cdots + 2 = 2 \cdots 100 = 200.$$

On the other hand, the method in part (b) and (c) is quite hard to generalize. Indeed, the sample space for 100 coin flips has 2^{100} elements! You would need to be really clever to work out the pmf of X. And this becomes even harder with dice instead of coins.

Moral: Linearity of expectation is a powerful tool that allows us to ignore many details.

4. Roulette. A European roulette wheel has 37 pockets: 1 colored green, 18 colored red and 18 colored black. To play the game you pay \$1 and pick a color from $\{r, b, g\}$ (red, black, green). Then the croupier[2] spins the wheel and observes which pocket a marble falls into. Your winnings are decided by the following rules:

- If you pick r and the marble lands in a red pocket you win \$2.
- If you pick b and the marble lands in a black pocket you win \$2.
- If you pick g and the marble lands in the green pocket you win \$35.
- Otherwise you get nothing.

Suppose that the 37 pockets are equally likely and let W be your winnings.

(a) Compute $E[W]$ if you pick r. [You get the same answer for color b.]
(b) Compute $E[W]$ if you pick g.
(c) Compute $E[W]$ if you pick from $\{r, b, g\}$ at random, each with probability 1/3.

[Hint: In each case the random variable W has support $\{0, 2, 35\}$, hence you need to compute the probabilities $P(W = 0)$, $P(W = 2)$ and $P(W = 35)$.]

(a): Suppose that you pick red. Then the value of your winnings W is either 0 or 2. The probabilities are

$$P(W = 0) = 19/37 \quad \text{and} \quad P(W = 2) = 18/37$$

since 18 of the 37 slots are red and 19 slots are "not red". Hence your expected winnings are

$$E[W] = 0P(W = 0) + 2P(W = 2) = 0(19/37) + 2(18/37) = 36/37 = \$0.973.$$

Remark: Your **profit** is winnings minus the cost to play: $W - 1$. So your expected profit is

$$E[W - 1] = E[W] - 1 = 1 - 0.973 = -\$0.027.$$

[2] The person who spins the wheel.

We could also say that the **house edge** is 2.7%. On average, the casino will take 2.7% of your money each time you bet red.

(b): Suppose you pick green. Then the value of your winning W is either 0 or 35. The probabilities are

$$P(W = 0) = 36/37 \quad \text{and} \quad P(W = 35) = 1/37,$$

since 1 slot is green and 36 slots are "not green". Your expected winnings are

$$E[W] = 0(36/37) + 35(1/37) = 35/37 = \$0.946.$$

Note that you win less money on average when you pick green than when you pick red or black. Don't be fooled!

(c): This one is trickier than I intended it to be. The goal is to compute the probabilities $P(W = 0)$, $P(W = 2)$ and $P(W = 35)$, since then we will obtain the expected value

$$E[W] = 0P(W = 0) + 2P(W = 2) + 35P(W = 35).$$

So how can we compute the probabilities? Let's name three events:

$$\begin{aligned} R &= \text{you bet on red,} \\ B &= \text{you bet on black,} \\ G &= \text{you bet on green.} \end{aligned}$$

We have assumed that $P(R) = P(B) = P(G) = 1/3$. We also know the values of the various conditional probabilities $P(\{W = k\}|R)$, $P(\{W = k\}|B)$, $P(\{W = k\}|G)$. These are just the numbers we computed in parts (a) and (b). Then since R, B, G divide the sample space we can use the Law of Total Probability:

$$\begin{aligned} P(W = k) &= P(\{W = k\} \cap R) + P(\{W = k\} \cap B) + P(\{W = k\} \cap G) \\ &= P(R)P(\{W = k\}|R) + P(B)P(\{W = k\}|B) + P(G)P(\{W = k\}|G). \end{aligned}$$

For the three possible values of k we obtain

$$\begin{aligned} P(W = 0) &= (1/3)(19/37) + (1/3)(19/37) + (1/3)(36/37) = 74/111, \\ P(W = 2) &= (1/3)(18/37) + (1/3)(18/37) + (1/3)(0/37) = 36/111, \\ P(W = 35) &= (1/3)(0/37) + (1/3)(0/37) + (1/3)(1/37) = 1/111. \end{aligned}$$

Hence the expected of your winnings is

$$E[W] = 0(74/111) + 2(36/111) + 35(1/111) = 107/111 = \$0.964.$$

Remark: You might have noticed that this answer can be obtained more easily by combining the answers from (a) and (b) in the following reasonable way:

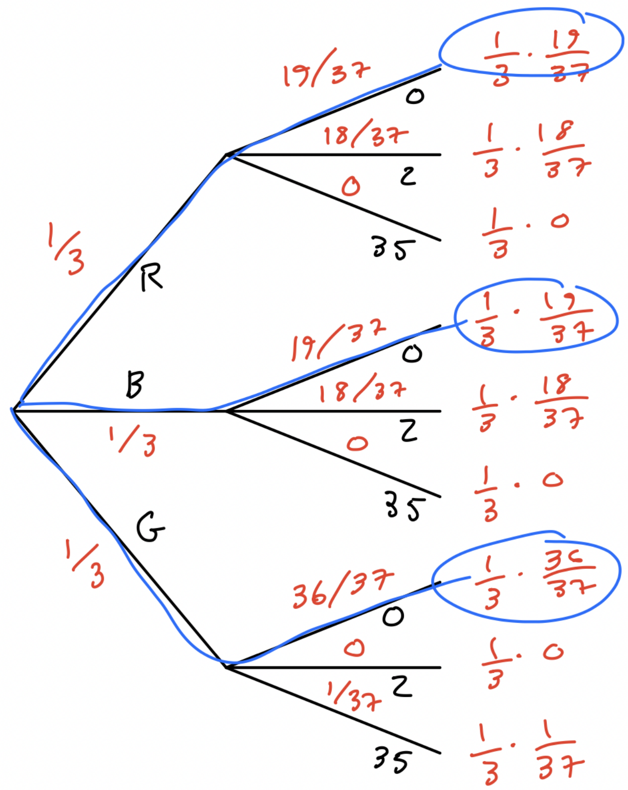
$$(1/3)(36/37) + (1/3)(36/37) + (1/3)(35/37) = 107/111.$$

To explain why this is correct requires the concept of *conditional expectation.* Given a random variable X and an event A, we let $E[X|A]$ denote the expected value of X, assuming that A happens. Then it follows from a general property of conditional expectation that

$$E[W] = P(R)E[W|R] + P(B)E[W|B] + P(G)E[W|G].$$

We won't study conditional expectation in this course.

Remark: You might find it easier to organize the information from part (c) in a tree diagram:



Here I have circled the branches corresponding to the event $W = 0$. To obtain the probability $P(W = 0)$ we add up the probabilities of these three branches:

$$P(W = 0) = (1/3)(19/37) + (1/3)(19/37) + (1/3)(36/37) = 74/111.$$

5. Geometric Random Variables. Let X be a geometric random variable with parameters p and $q = 1 - p$, so that $P(X = k) = pq^{k-1}$.[3] We can interpret X as the number of coin flips until we see H for the first time, where $P(H) = p$. Let's assume that $p \neq 0$ so that $q < 1$.

(a) Use the geometric series to verify that $\sum_k P(X = k) = 1$.
(b) If $k \geq 1$, use the geometric series to prove that $P(X \geq k) = q^{k-1}$.
(c) Differentiate the geometric series to prove that $E[X] = 1/p$.

For any number $|q| < 1$ the geometric series tells us that

$$1 + q + q^3 + q^4 + \cdots = 1/(1 - q).$$

If we view both sides as functions of q then we can also differentiate to obtain[4]

$$0 + 1 + 2q + 3q^2 + \cdots = 1/(1 - q)^2.$$

In particular, both of these identities hold when q is the probability of tails and $q < 1$.

(a): The sum of the probabilities is

$$\sum_k P(X = k) = P(X = 1) + P(X = 2) + P(X = 3) + \cdots$$

[3]We can think of X as the number of coin flips until we see heads.
[4]The derivative of $(1 - q)^{-1}$ is $(-1)(1 - q)^{-2}(0 - 1) = (1 - q)^{-2}$.

$$
\begin{aligned}
&= p + pq + pq^2 + \cdots \\
&= p(1 + q + q^2 + \cdots) \\
&= p \cdot \frac{1}{1-q} = p \cdot \frac{1}{p} = 1.
\end{aligned}
$$

(b): If $k \geq 1$ then we have

$$
\begin{aligned}
P(X \geq k) &= P(X = k) + P(X = k+1) + P(X = k+2) + \cdots \\
&= pq^{k-1} + pq^k + pq^{k+1} + \cdots \\
&= pq^{k-1}(1 + q + q^2 + \cdots) \\
&= pq^{k-1} \cdot \frac{1}{1-q} = pq^{k-1} \cdot \frac{1}{p} = q^{k-1}.
\end{aligned}
$$

Note that part (a) is a special case of (b) when $k = 1$.

(c): The expected value of X is

$$
\begin{aligned}
E[X] &= \sum_k k \cdot P(X = k) \\
&= 1P(X = 1) + 2P(X = 2) + 3P(X = 3) + \cdots \\
&= 1p + 2pq + 3pq^2 + \cdots \\
&= p(1 + 2q + 3q^2 + \cdots) \\
&= p \cdot \frac{1}{(1-q)2} = p \cdot \frac{1}{p^p} = \frac{1}{p}.
\end{aligned}
$$

6. The Coupon Collector Problem. Suppose that you roll a fair n-sided die until you see all n sides, and let X be the number of rolls that you did. In this problem I will guide you through a method to compute the expected number of rolls $E[X]$.

(a) Fix some $0 \leq k \leq n-1$ and suppose that you have already seen k sides of the die. Let X_k be the number of die rolls until you see one of the remaining $n-k$ sides. Compute $E[X_k]$. [Hint: This is a geometric random variable. Think of the die as a coin with $T =$"you see one of the k sides that you've already seen" and $H =$"you see one of the $n-k$ sides that you haven't seen yet". Use Problem 5(c).]

(b) We observe that $X = X_0 + X_1 + \cdots + X_{n-1}$ is the total number of rolls until you see all n sides of the die. Use part (a) and linearity of expectation to compute $E[X]$.

(c) Example: Suppose that you roll a fair 6-sided die until you see all six sides. On average, how many rolls do you expect to make?

Idea: We express X as a sum $X = X_0 + X_1 + \cdots + X_{n-1}$, where X_k is how long we have to wait between seeing k faces and $k+1$ faces of the die. It turns out that the expected value of X_k is easy to compute.

(a): Assuming we have already seen k faces of the die, let X_k be the number of die rolls until we see one of the remaining $n-k$ faces. We can think of each roll as a coin flip with $T =$ we see one the k faces we've already seen and $H =$ we see one of the $n-k$ faces that we haven't seen yet. Since $P(H) = (n-k)/n$, Problem 5(c) tells us that

$$
E[X_k] = \frac{1}{P(H)} = \frac{1}{(n-k)/n} = \frac{n}{n-k}.
$$

(b): Then we use the linearity of expectation to obtain

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

Apart from factoring out n, this formula does not simplify.

(c): Let X be the number of die rolls until you see every side of a fair 6-sided die. Then

$$\begin{aligned} E[X] &= \frac{6}{6} + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} \\ &= 1 + 1.2 + 1.5 + 2 + 3 + 6 \\ &= 14.7. \end{aligned}$$

This is a testable prediction.