Math 309 Homework 6 — Fall 2014, Drew Armstrong

1. Expected Value. Let S be a **finite** sample space with probability function $P : \wp(S) \to \mathbb{R}$ and let $X : S \to \mathbb{R}$ be any random variable.

(a) For all real numbers $k \in \mathbb{R}$ we define the event $E_k = \{s \in S : X(s) = k\}$. This is the set of outcomes that take value k under X. Then we define $P(X = k) := P(E_k)$ and we call this the "probability that X equals k". Explain why

$$P(X = k) = \sum_{s \in E_k} P(s).$$

(The sum is over elements s of the set E_k and we use the notation $P(s) := P(\{s\})$.)

(b) Following Archimedes' "Law of the Lever", we define the **expected value** of the random variable X by

$$E(X) := \sum_k k \cdot P(X = k).$$

Since S is finite, there are only finitely many values k such that $P(X = k) \neq 0$, so we can interpret this as a **finite sum** (and hence avoid any complications with integrals or convergence). Use part (a) to explain why

$$E(X) = \sum_{s \in S} X(s) \cdot P(s),$$

where the sum is over all elements s of the sample space S.

(a) We first recall Kolmogorov's third axiom of probability:

$$P(E \sqcup F) = P(E) + P(F).$$

If $E \subseteq S$ is any event, then since S is finite we can write $E = \{s_1, s_2, \ldots, s_n\}$. Note that E is a disjoint union of its one-element subsets:

$$E = \{s_1\} \sqcup \{s_2\} \sqcup \cdots \sqcup \{s_n\}.$$

Then applying Kolmogorov's third axiom (and induction) gives

$$\begin{aligned} P(E) &= P(\{s_1\}) + P(\{s_2\}) + \cdots + P(\{s_n\}) \\ &= P(s_1) + P(s_2) + \cdots + P(s_n) \\ &= \sum_{i=1}^{n} P(s_i) \\ &= \sum_{s \in E} P(s). \end{aligned}$$

In other words, the probability of an event equals the sum of the probabilites of the outcomes it contains. Since this is true for any event E, it is true in particular for the event E_k.

(b) Recall that E_k is the set of outcomes $s \in S$ such that $X(s) = k$. Since X takes every element of S to **some** number, we observe that S is the disjoint union of the sets E_k over all the possible values of k. Thus we have

$$\begin{aligned}\sum_{s\in S} X(s)\cdot P(s) &= \sum_k \sum_{s\in E_k} X(s)\cdot P(s)\\ &= \sum_k \sum_{s\in E_k} k\cdot P(s)\\ &= \sum_k k \sum_{s\in E_k} P(s)\\ &= \sum_k k\cdot P(E_k)\\ &= \sum_k k\cdot P(X=k).\end{aligned}$$

The difficult part of this is just **remembering** what each piece of notation means. There is always some effort in setting up a mathematical notation, but once it's set up it can help us solve problems more easily by automating the process of reasoning.

2. Linearity of Expectation. Let X and Y be two random variables on a finite sample space S, and let a and b be constants. We define the random variable $aX+bY$ by $(aX+bY)(s) := aX(s)+bY(s)$ for all $s\in S$.

(a) Use the result of Problem 1 to prove that $E(aX+bY) = aE(X)+bE(Y)$.
(b) Use the result of part (a) to show that

$$E((X-E(X))^2) = E(X(X-1)) + E(X) - E(X)^2.$$

(a) We use the formula $E(X) = \sum_{s\in S} X(s)P(s)$ to obtain

$$\begin{aligned}E(aX+bY) &= \sum_{s\in S}(aX+bY)(s)P(s)\\ &= \sum_{s\in S}(aX(s)+bY(s))P(s)\\ &= \sum_{s\in S}(aX(s)P(s)+bY(s)P(s))\\ &= a\sum_{s\in S} X(s)P(s) + b\sum_{s\in S} Y(s)P(s)\\ &= aE(X)+bE(Y).\end{aligned}$$

What good is this? Well, it's a fundamental property of the expected value, and it makes many computations easier.

(b) Recall that the **variance** of a random variable X is defined by $\mathrm{Var}(X) := E((X-E(X))^2)$. In class we showed that $\mathrm{Var}(X) = E(X^2)-E(X)^2$. In this problem, we will prove a slightly different formula. We already used this formula in class to prove that the variance of the binomial random variable $B(n,p)$ equals $np(1-p)$. The formula is a bit funny, but it's the easiest proof that I know.

First recall (as we showed in class) that

$$\begin{aligned}
E((X-E(X))^2) &= E(X^2 - 2XE(X) + E(X)^2) \\
&= E(X^2) - E(2E(X)X) + E(E(X)^2) \\
&= E(X^2) - 2E(X)\cdot E(X) + E(X)^2 \\
&= E(X^2) - E(X)^2.
\end{aligned}$$

Here we used the fact that $2E(X)$ and $E(X)^2$ are constants. Next observe that

$$\begin{aligned}
E(X(X-1)) + E(X) - E(X)^2 &= E(X^2 - X) + E(X) - E(X)^2 \\
&= E(X^2) - E(X) + E(X) - E(X)^2 \\
&= E(X^2) - E(X)^2.
\end{aligned}$$

We conclude that the two formulas are equal, as desired.

3. An Urn Problem. An urn contains 6 red balls and 3 green balls. You reach in and grab 4 balls at random. (Assume that each outcome is equally likely.) Let X be the number of red balls that you get.

(a) Compute the probability $P(X=k)$ for each possible value of k.
(b) Compute the expected number of red balls, $E(X)$.
(c) Compute the variance $\mathrm{Var}(X)$. You can use the formula $E((X-E(X))^2)$ or the formula $E(X^2)-E(X)^2$. Recall that $E(X^2)$ can be expressed as $\sum_k k^2 \cdot P(X=k)$.

(a) There are $6+3=9$ balls in an urn. We reach in and grab 4 of them. The number of different outcomes is $\binom{9}{4} = 126$. We assume that every outcome is equally likely, with probability $1/126$. Thus to compute the probability of getting k red balls we just need to count the number of ways we can get k red balls. This number equals

$$\binom{6}{k}\binom{3}{4-k}$$

because there are $\binom{6}{k}$ ways to choose the k red balls and $\binom{3}{4-k}$ ways to choose the remaining $4-k$ green balls. Thus we have

$$P(X=k) = \frac{\binom{6}{k}\binom{3}{4-k}}{\binom{9}{4}} = \frac{\binom{6}{k}\binom{3}{4-k}}{126}.$$

We can list these probabilities in a table:

k	0	1	2	3	4
$P(X=k)$	$\frac{0}{126}$	$\frac{6}{126}$	$\frac{45}{126}$	$\frac{60}{126}$	$\frac{15}{126}$

Observe that the probabilities add to 1. [Remark: We could alternatively analyze this experiment by recording the order in which we choose the 4 balls. In that case the size of the sample space is $9\cdot 8\cdot 7\cdot 6 = 3024$. The calculations are a bit harder, but I guarantee you'll end up with the same probabilities. So why not do it the easy way?]

(b) The expected number of red balls is

$$
\begin{aligned}
E(X) &= \sum_k k \cdot P(X = k) \\
&= 0 \cdot \frac{0}{126} + 1 \cdot \frac{6}{126} + 2 \cdot \frac{45}{126} + 3 \cdot \frac{60}{126} + 4 \cdot \frac{15}{126} \\
&= \frac{0 \cdot 0 + 1 \cdot 6 + 2 \cdot 45 + 3 \cdot 60 + 4 \cdot 15}{126} \\
&= \frac{336}{126} \\
&= \frac{8}{3}.
\end{aligned}
$$

Does that make sense? Yes. The ratio of red balls in the urn is $\frac{6}{6+3} = \frac{6}{9} = \frac{2}{3}$. So if we reach in and grab 4 balls, we expect $\frac{2}{3} \cdot 4 = \frac{8}{3}$ of them to be red. This means we probably didn't make a mistake in our computation.

(c) To compute the variance we first need to know $E(X^2)$. We can think of this as the expected value of a different probability distribution:

k	0	1	4	9	16
$P(X^2 = k)$	$\frac{0}{126}$	$\frac{6}{126}$	$\frac{45}{126}$	$\frac{60}{126}$	$\frac{15}{126}$

We don't really care about the distribution of X^2; the only reason we computed it is because it gives us a shortcut to the variance. We have

$$E(X^2) = 0 \cdot \frac{0}{126} + 1 \cdot \frac{6}{126} + 4 \cdot \frac{45}{126} + 9 \cdot \frac{60}{126} + 16 \cdot \frac{15}{126} = \frac{996}{126} = \frac{23}{3}$$

and therefore

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{23}{3} - \left(\frac{8}{3}\right)^2 = \frac{5}{9} \approx 0.56.$$

It's harder to know if this is the correct answer because variance has less intuitive content than expected value. However, I looked up the formula for the variance of a hypergeometric distribution and verified that $\frac{5}{9}$ is indeed the correct answer. (Look it up and see.)

4. The Central Limit Theorem. Flip a fair coin $2n$ times and let X be the number of heads you get. In 1733, Abraham de Moivre observed that for **large** n and for **small** k, the probability $P(X = n + k)$ is approximately

$$\frac{1}{\sqrt{\pi n}} e^{-k^2/n}.$$

Use this to estimate the probability of getting heads between 2490 and 2510 times in 5000 flips of a fair coin. [Hint: Integrate.]

If we flip a fair coin 5000 times then the probability of getting heads $2500 + k$ times is exactly

$$\frac{\binom{5000}{2500+k}}{2^{5000}},$$

as you know. Thus the probability of getting between 2490 and 2510 heads (inclusive) is

$$\sum_{k=-10}^{10} \frac{\binom{5000}{2500+k}}{2^{5000}}$$

This sum is impossible to compute by hand because the numerators and denominators of these fractions all have more than 1500 digits! De Moivre's approximation tells us that

$$\frac{\binom{5000}{2500+k}}{2^{5000}} \approx \frac{1}{\sqrt{2500\pi}} e^{-k^2/2500},$$

so the probability is approximated by an integral:

$$\sum_{k=-10}^{10} \frac{\binom{5000}{2500+k}}{2^{5000}} \approx \int_{-10.5}^{10.5} \frac{1}{\sqrt{2500\pi}} e^{-k^2/2500}\, dk.$$

This may not seem helpful, but de Moivre was very clever and he was able to compute this integral by hand using Taylor series and other tricks. We won't compute it by hand. I'll put it in my computer and just tell you that the integral evaluates to

$$\int_{-10.5}^{10.5} \frac{1}{\sqrt{2500\pi}} e^{-k^2/2500}\, dk = 0.233522.$$

And that's pretty good because the exact value of the probability is

$$\sum_{k=-10}^{10} \frac{\binom{5000}{2500+k}}{2^{5000}} = 0.233518.$$

In summary, the probability of getting between 2490 and 2510 heads in 5000 flips of a fair coin is 23.35%.

Here is a picture of the (discrete) binomial distribution and the (continuous) normal distribution from this problem. (Outside of the displayed range the probability is essentially zero.) The distributions are so close together I can barely see the difference. Good job, de Moivre!