APMA 1650

Homework 4 - Solutions

Instructions: Homework is due by 11:59pm EST in Gradescope on the day listed on the course webpage. You can use calculators for this assignment. Solutions must be written independently and *cannot* be shared with any other students.

You must *show all work* and *explain your answers thoroughly* to get full credit. **You will be graded partly on how well you explain the answers**.

There will be a 5pt penalty for homework submitted with problems incorrectly assigned to a page. A 10pt penalty will be applied for homework submitted during the late window.

---

1. (18 pts) Consider the following CDF of a random variable X

$$F_X(x) = \begin{cases} 0 & \text{for } x < -2 \\ 1/3 & \text{for } -2 \leq x < -1 \\ 1/2 & \text{for } -1 \leq x < 1 \\ 5/6 & \text{for } 1 \leq x < 2 \\ 1 & \text{for } 2 \leq x \end{cases}.$$

Find the following

a. (4 pts) $P(X > -2)$

b. (4 pts) $P(-1 \leq X \leq 2)$

c. (4 pts) $P(2X \leq -1)$

d. (6 ots) $E[X]$

Soltuion: From the CDF we can tell that this random variable is discrete with range

$$R_X = \{-2, -1, 1, 2\}$$

and looking at the size of the jumps we can readily conclude that the PMF is

$$P_X(x) = \begin{cases} 1/3 & \text{for } x = -2 \\ 1/6 & \text{for } x = -1 \\ 1/3 & \text{for } x = 1 \\ 1/6 & \text{for } x = 2 \\ 0 & \text{otherwise} \end{cases}.$$

a. Here we have that $P(X > -2) = 1 - P(X \le -2) = 1 - F_X(-2) = 2/3$.

b. In this case we find

$$\begin{aligned} P(-1 \le X \le 2) &= P(X \le 2) - P(X < -1) \\ &= P(X \le 2) - P(X \le -2) \\ &= F_X(2) - F_X(-2) = 1 - 1/3 = 2/3 \end{aligned}$$

c. Note that $P(2X \le -1) = P(X \le -1/2) = F_X(-1/2) = 1/2$

d. To calculate the expectation, we can use the PMF we worked out above giving

$$EX = -2(1/3) - 1(1/6) + 1(1/3) + 2(1/6) = 1/6 - 1/3 = -1/6.$$

2. (20 pts) A TV manufacturer makes a display that has standard 1920×1080 pixel array. Suppose that each pixel independently has a one in a million chance of being defective. Use the Poisson distribution to answer the following problems. Hint you may find the textbook Poisson CDF calculator very useful.

a. (10 pts) In order for a display to meet the ISO 9241-305 Class II standard, the manufacturer can't have more than 2 defective pixels per one million pixels in that display. Approximate the probability that a given TV does not meet the Class II standard.

b. (10 pts) Now suppose that the manufacturer needs to make 100 TVs that meet the Class II standard. Approximate the minimal number of TVs they should manufacture to be 99% certain they have at least 100 Class II TVs produced.

Solution:

a. A given TV has $1920 \times 1080 = 2,073,600$ pixels. Therefore the number of faulty pixels in a TV is expected to follow a Binomial distribution $X \sim \text{Binomial}(n, p)$, where $n = 2,073,600$ and $p = 1/1,000,000$. Note that since p is so small (i.e. rare) and n is so large, we can easily justify the Poisson distribution approximation and estimate the number of defective pixels per TV as a Poisson random variable

$$X \sim \text{Poisson}(\lambda)$$

where rate $\lambda = np = 2,073,600/1,000,000 = 2.0736$ faulty pixels per TV. In order to meet Class II, we need to have no more than $2 \times 2.0736 \approx 4.1472$ pixels per TV. Therefore the probability that a given TV does not meet the Class II standard is

$$P(X \ge 4.1472) = P(X > 4) = 1 - F_X(4).$$

Using the Poisson CDF calculator in the text with $\lambda = 2.0736$, we obtain $F_X(4) = 0.94046$ and therefore

$$P(X \ge 4.1472) = 1 - 0.94046 = 0.05954.$$

b. Now we consider how many TVs the manufacturer will need to produce to be 99% certain that they can meet the 100 Class II TV production demand. We know from part a that each TV has a probability $p = 0.05954$ of not meeting the Class II demand. Therefore if we make $n \geq 100$ TVs, the number of TVs that *fail to meet the Class II specification* is $Y_n \sim \text{Binomial}(n, p)$. Again since n is large and p is small, we will approximate with a Poisson distribution,

$$Y_n \sim \text{Poisson}(\lambda_n), \quad \lambda_n = np = n(0.05954).$$

Our goal is to find the smallest n such that the probability of the number of class II TVs $n - Y_n$ being bigger than 100 is less than 99%. In mathematical terms, we want to find n such that

$$P(n - Y_n \geq 100) = P(Y_n \leq n - 100) = F_{Y_n}(n - 100) \geq .99.$$

Note that the Poisson rate λ_n here depends on n.

To find such an n, we will guess an check using the CDF calculator for Poisson. It help to have a good educated guess. Lets use that fact that for many values of $n \approx 100$, $np \approx 7$ (rounding up for good measure) we should expect to loose 7 or so TVs in the process of making 100. Naturally this suggests we might want to make 107 TVs. Setting $n = 107$, however we find

$$n = 107, \lambda_{107} = 107p = 6.37078, \quad F_{Y_{107}}(107 - 100) = F_{Y_{107}}(7) = 0.69155.$$

This is clearly not high enough to meet the .99% benchmark. Indeed, there is quite a lot of variance in Y_n given by $\text{Var}(Y_n) = np \approx 7$. Therefore perhaps we want to account for the variance in the defective TVs and instead produce that is 2 standard deviations away $100 + 7 + 2\sqrt{7} \approx 113$ TVs. Setting $n = 113$, we find

$$n = 113, \lambda_{113} = 113p = 6.72802, \quad F_{Y_{113}}(113 - 100) = F_{Y_{113}}(13) = 0.99062$$

Hooray! This meets our goal with 99% certainty! Now lets see if this is the smallest number. Suppose we made one less, instead choosing $n = 112$, we find

$$n = 112, \lambda_{112} = 112p = 6.66848, \quad F_{Y_{112}}(112 - 100) = F_{Y_{112}}(12) = 0.98074$$

Which is just a smidge under .99, so lets stick with the safer bet of

$$n = 113 \text{ TVs}.$$

3. (18 pts)

a. (6 pts) Let $X \sim \text{Geometric}(p)$. Calculate $\text{Var}(X)$.

b. (6 pts) Let $Y \sim \text{Pascal}(m, p)$. Calculate $\text{Var}(Y)$.

c. (6 pts) Let $X \sim \text{Poisson}(\lambda)$. Calculate $\text{Var}(X)$.

Solution:

a. To calculate the variance of X, we first recall that we can calculate this using the formula

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

We know that $EX = 1/p$, and so whats left is to compute

$$EX^2 = \sum_{k=1}^{\infty} k^2 P_X(k) = p \sum_{k=1}^{\infty} k^2 q^{k-1},$$

where $q = 1 - p$. To determine this sum, we realize that

$$\sum_{k=1}^{\infty} kq^{k-1} = \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{1}{1-q}\right) = \frac{1}{(1-q)^2}$$

Therefore using the above formula

$$\begin{aligned}
\sum_{k=1}^{\infty} k^2 q^{k-1} &= \frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{k=1}^{\infty} kq^k\right) = \frac{\mathrm{d}}{\mathrm{d}q}\left(q\sum_{k=1}^{\infty} kq^{k-1}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{(1-q)^2}\right) = \frac{1}{(1-q)^2} + \frac{2q}{(1-q)^3} \\
&= \frac{1+q}{(1-q)^3}
\end{aligned}$$

It follows that

$$EX^2 = p\frac{1+q}{(1-q)^3} = \frac{2-p}{p^2},$$

and therefore

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

Alternate solution. An alternate solution to this is to instead use the formula

$$\text{Var}(X) = E[X(X-1)] - \mu_X(\mu_X - 1).$$

This is nicer because the derivatives workout better, giving

$$\begin{aligned}
E[X(X-1)] &= pq\sum_{k=1}^{\infty} k(k-1)q^{k-2} = pq\frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\sum_{k=1}^{\infty} q^k\right) \\
&= pq\frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\frac{1}{1-q}\right) = \frac{2pq}{(1-q)^3} \\
&= \frac{2q}{p^2},
\end{aligned}$$

and so

$$\mathrm{Var}(X) = \frac{2q}{p^2} - \frac{1}{p}\left(\frac{1}{p} - 1\right) = \frac{q}{p^2}$$

b. To calculate this, we recall that any Pascal(m, p) random variable can be written as a sum of m independent geometric random variables

$$Y = Y_1 + Y_2 + \ldots Y_m, \quad Y_i \sim \text{Geometric}(p).$$

Using the independence of Y_i, the properties of the variance of sums of independent random variables, and the variance we worked out in part a,

$$\mathrm{Var}(Y) = \mathrm{Var}(Y_1 + Y_2 + \ldots + Y_m) = m\mathrm{Var}(Y_i) = \frac{m(1-p)}{p^2}.$$

c. For $X \sim$ Poisson(λ) we know $\mu_X = EX = \lambda$. In this problem as we saw in the alternate solution to part a, we will find it useful to use the formula (although your do it many ways)

$$\mathrm{Var}(X) = E[X(X-1)] - \mu_X(\mu_X - 1).$$

We find that using the Taylor series for e^{λ}

$$\begin{aligned}
E[X(X-1)] &= e^{-\lambda}\sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} \\
&= e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{Var}(X) &= E[X(X-1)] - \mu_X(\mu_X - 1) \\
&= \lambda^2 - \lambda(\lambda - 1) \\
&= \lambda.
\end{aligned}$$

4. (14 pts)

a. (4 pts) Suppose X is a random variable with $E[X] = 1$ and $E[X(X-2)] = 3$. What is $\text{Var}(X)$?

b. (4 pts) Suppose X is a random variable with mean 1. Show that $E[X^2] \geq 1$.

c. (6 pts) Let $X = I_A$ be the indicator Bernoulli random variable for a probability zero event A. What is $\text{Var}(X)$?

Solution:

a. We know that
$$3 = E[X(X-2)] = E[X^2] - 2EX = E[X^2] - 2.$$
Solving for $E[X^2]$ gives $E[X^2] = 5$ and therefore
$$\text{Var}(X) = E[X^2] - (EX)^2 = 5 - 1 = 4.$$

b. We know that $\text{Var}(X) \geq 0$ and therefore
$$E[X^2] = \text{Var}(X) + (EX)^2 \geq (EX)^2 = 1.$$

c. If $X = I_A$, then $X \sim \text{Bernoulli}(0)$, since $P(A) = 0$ it follows that
$$\text{Var}(X) = pq = 0(1) = 0.$$

5. (30 pts) Suppose that for a (very easy) homework assignment students are tasked with flipping a fair coin 100 times and recording the outcomes. One student does the work and writes down the results of the 100 flips, another student is lazy and makes up the data. Here are the outcomes from the two students (in no particular order):

Student 1:
$THHHTHTTTTHTTHTTTHHTHTTHT$
$HHHTHTHHTHTTHHTTTTHTTTHTH$
$TTHHTTTTTTTTHTHHHHHTHTHTH$
$THTHTHHHHHTHHTTTTTHTTHHTH$

Student 2:
$HTTHTTHTHHTTHTHTHTTHHTHTT$
$HTTHHHTTHTTHTHTHTHHTTHTTH$
$THTHTHTHHHTTHTHTHTHHTHTTT$
$HTHHTHTHTHTHHTTHTHTHTTHHT$

Your goal is to figure out which student was most likely to have fudged the data. Your strategy is to count the number of TT pairs in each sample. (Here counting TT pairs involves counting how many times TT shows up out of *all* 99 *neighboring* pairs. For instance for the 8 tosses $TTTHTHTT$ has 3 TT pairs.)

a. (12 pts) Suppose you flip a fair coin n times. What is the expected number of TT pairs? (Hint: Use sums of Bernoulli random variables to count the number of occurrences.)

b. (3 pts) Based on your answers to a, which student do you think most likely fudged their data.

c. (15 pts) Repeat a and b for the number of TTT triples (i.e. $TTTTT$ has 3 triples).

Solution:

Intuition: Qualitatively, we see that the two sequences are quite different. One of them has long stretches (runs) of H's and T's, while the other has no runs of H's or T's longer than 3. While getting a long stretch of H's or T's, is rare since were flipping the coin so many times, we expect these rare events to show up a few times (think Poisson's distribution). However the data from the second student is suspiciously missing all "rare" events involving runs longer than 3.

a. To count the number of TT pairs in a sequence of n flips, we divy up the n flips into $n-1$ consecutive pairs. For each $i = 1, \ldots n-1$, let X_i be the following Bernoulli random variable

$$X_i = \begin{cases} 1 & \text{if the } i\text{th pair is a } TT \\ 0 & \text{otherwise} \end{cases}$$

Note that the probability that given pair is a TT is just $(1/2)(1/2) = 1/4$ and therefore

$$X_i \sim \text{Bernoulli}(1/4).$$

The number of TT pairs is then given by

$$X = X_1 + X_2 + \ldots X_{n-1}.$$

Note that X_i *are not independent* since the pairs overlap and therefore one being a TT pair affects the probability of the next being a TT pair, therefore X is *not* an Binomial distribution. None-the-less we can compute its expected value using linearity of expectation

$$EX = EX_1 + EX_2 + \ldots EX_{n-1} = (n-1)EX_i = \frac{n-1}{4}.$$

b. For $n = 100$ we then typically expect there to be $99/4 \approx 25$ TT pairs. If we go ahead and count the number of TT pairs in the above sample, we find

$$\text{Student } 1 : 25 \quad TT \text{ pairs.}$$

$$\text{Student } 2 : 15 \quad TT \text{ pairs}$$

Indeed Student 1 displays exactly what we expect to be the number of TT pairs, while Sudent 2 is significantly below the expected value. Based on this, I'd say it is most likely that Student 2 fudged the data.

c. If we instead count TTT triples, we follow a similar strategy divying up n flips into $n-2$ triples and assigning a Bernoulli random variable X_i to each triple with probability $p = (1/2)(1/2)(1/2) = 1/8$ of occuring. Using linearity of expected value, the expected number of TTT triples $X = X_1 + X_2 + \dots X_{n-2}$ is then

$$EX = \frac{n-2}{8}.$$

For the case $n = 100$, we see that we expect $98/8 \approx 12$ TTT triples. Counting the number of TTT triples in the above sample, we find

$$\text{Student } 1 : 15 \quad TTT \text{ triples}$$

$$\text{Student } 2 : 1 \quad TTT \text{ triple.}$$

While student 1 might be a little higher than expected in the number of TTT triples, student 2 is WAY off, with only 1 TTT triple out of the expected 12. Again this indicates that Student 2 is the one who is more likely to have fudged the data.