Stat 200A, Fall 2018
A. Adhikari and Z. Bartha

WEEK 10 EXERCISES

You are expected to do all these problems, but for **Homework 10** please turn in **only Problems 3, 5, and 7** on **Thursday November 1 at the start of lecture**.

1. A Mixed Population

A population consists of men, women, and children. Here are some summary statistics on their heights.

- 100 women, average height 65 inches, SD of heights 3 inches
- 75 men, average height 69 inches, SD of heights 4 inches
- 25 children, average height 50 inches, SD of heights 2 inches

Let X be the height of a person picked at random from this population. Find $E(X)$ and $SD(X)$.

Let I be the random variable that encodes the type of the person that we pick: $I = 1$ for women, $I = 2$ for men, $I = 3$ for children. Then

$$E[X] = E[E[X \mid I]] = \frac{100}{200} \cdot 65 + \frac{75}{200} \cdot 69 + \frac{25}{200} \cdot 50 = 64.625,$$

and

$$\begin{aligned} Var[X] &= E[Var[X \mid I]] + Var[E[X \mid I]] \\ &= \left(\frac{100}{200} \cdot 3^2 + \frac{75}{200} \cdot 4^2 + \frac{25}{200} \cdot 2^2\right) + \left(\frac{100}{200} \cdot 65^2 + \frac{75}{200} \cdot 69^2 + \frac{25}{200} \cdot 50^2 - 64.625^2\right) \\ &= 44.984375 \end{aligned}$$

2. Heads in Tosses of a Random Coin

A random variable X with values in $(0, 1)$ has expectation μ and SD σ.

Fix a positive integer n. Given that $X = p$, toss an p-coin n times. Let H be the number of heads.

Find $E(H)$ and $Var(H)$.

The conditional distribution of H given X is Binomial(n, X), therefore

$$\begin{aligned} E[H] &= E[E[H \mid X]] = E[nX] = n\mu, \\ Var[H] &= E[Var[H \mid X]] + Var[E[H \mid X]] = E[nX(1 - X)] + Var[nX] = n\mu - nE[X^2] + n^2Var[X] \\ &= n\mu - n(Var[X] + E[X]^2) + n^2Var[X] = n\mu - n\sigma^2 - n\mu^2 + n^2\sigma^2. \end{aligned}$$

3. Two-Colored Die

A die that has 2 blue faces and 4 green faces is rolled repeatedly. Let R be the number of rolls till both colors have appeared. Find $E(R)$ and $Var(R)$.

Let F be the color of the first roll. Then R given $F = blue$ is equal in distribution to 1 plus a Geometric(4/6) random variable, whereas R given $F = green$ is 1 plus a Geometric(2/6) variable. Hence,

$$\begin{aligned}
E[R] &= E[E[R \mid F]] = \frac{2}{6}\left(1+\frac{6}{4}\right) + \frac{4}{6}\left(1+\frac{6}{2}\right) = \frac{21}{6}, \\
Var[R] &= E[Var[R \mid F]] + Var[E[R \mid F]] \\
&= \left(\frac{2}{6}\cdot\frac{2/6}{(4/6)^2} + \frac{4}{6}\cdot\frac{4/6}{(2/6)^2}\right) + \left(\frac{2}{6}\left(1+\frac{6}{4}\right)^2 + \frac{4}{6}\left(1+\frac{6}{2}\right)^2 - \left(\frac{21}{6}\right)^2\right) = 4.75
\end{aligned}$$

4. Comparing Waiting Times

A p-coin is tossed repeatedly. Let W_H be the number of tosses till the first H, W_{HH} the number of tosses till the first time the pattern HH appears, and W_{HT} the number of tosses till the first time the pattern HT appears.

(a) Let $p = 1/2$. Explain *without calculation* why $E(W_{HH})$ must be different from $E(W_{HT})$, and say which of the two values is smaller. [Think about what happens when the pattern is *not* completed on the toss following W_H.]

In both cases first we have to wait until the first H. Then both patterns get completed in the next toss with probability 1/2 each. However, if HH is not completed right away, that means we get a T, and have to start all over, whereas if HT is not completed, that means we get an H, and we again have a probability 1/2 of completing it in the next toss (and in every further toss as long as we keep tossing heads). This implies that $E[W_{HT}] < E[W_{HH}]$.

(b) Now let $0 < p < 1$. By conditioning appropriately, find $E(W_{HT})$ in terms of p.

Let F be the first toss. Given $F = heads$, W_{HT} is equal in distribution to $1 + W_T$, where $W_T \sim Geo(1-p)$. Given $F = tails$, W_{HT} is equal in distribution to $1 + W^*_{HT}$, the latter being an independent copy of W_{HT}. Then

$$E[W_{HT}] = E[E[W_{HT} \mid F]] = p\left(1 + \frac{1}{1-p}\right) + (1-p)(1 + E[W_{HT}]).$$

Solving this for $E[W_{HT}]$ we get

$$E[W_{HT}] = \frac{1}{p} + \frac{1}{1-p}.$$

(Alternatively, we could use the fact that $W_{HT} \stackrel{d}{=} W_H + W_T$, the sum of two geometric variables.)

(c) Find $E(W_{HHH})$ and simplify it as much as possible. Then guess an answer for the expected number of tosses till you get n heads in a row.

Let us use from lecture that $E[W_{HH}] = \frac{1+p}{p^2}$. Let N be the the first toss after the first appearance of HH. Given $N = heads$, $W_{HHH} = W_{HH} + 1$, and given $N = tails$, $W_{HHH} \stackrel{d}{=} W_{HH} + 1 + W^*_{HHH}$, the latter being an independent copy of W_{HHH}. (After an HHT, we have to start all over.) Then

$$E[W_{HHH}] = E[E[W_{HHH} \mid N]] = p(E[W_{HH}] + 1) + (1-p)(E[W_{HH}] + 1 + E[W_{HHH}]).$$

Solving this for $E[W_{HHH}]$, we get

$$E[W_{HHH}] = \frac{E[W_{HH}] + 1}{p} = \frac{1 + p + p^2}{p^3}.$$

Following this method, we get very similarly that

$$E[W_{(n+1)H}] = \frac{E[W_{nH}] + 1}{p},$$

where nH is the pattern of n H's in a row. Thus,

$$E[W_{nH}] = \frac{1 + p + p^2 + \ldots + p^{n-1}}{p^n}.$$

5. Overlapping Tosses

Consider a sequence of i.i.d. Bernoulli (p) trials. Consider the three variables X, Y, and Z defined by:

- X is the number of successes in trials 1 through 100
- Y is the number of successes in trials 51 through 100
- Z is the number of successes in trials 51 through 150

(a) For each of X, Y, and Z, say what the distribution is and provide the parameters.

$X \sim Bin(100, p)$, $Y \sim Bin(50, p)$, $Z \sim Bin(100, p)$.

(b) Fix k in the range $0, 1, \ldots, 100$ and find the conditional distribution of Y given $X = k$. Recognize this as a famous one and provide the parameters.

$$\mathbb{P}(Y = y \mid X = k) = \frac{\mathbb{P}(Y = y, X = k)}{\mathbb{P}(X = k)} = \frac{\binom{50}{y} p^y (1-p)^{50-y} \binom{50}{k-y} p^{k-y} (1-p)^{50-(k-y)}}{\binom{100}{k} p^k (1-p)^{100-k}} = \frac{\binom{50}{y}\binom{50}{k-y}}{\binom{100}{k}},$$

so the conditional distribution of Y given $X = k$ is Hypergeometric$(100, k, 50)$.

(c) Find the least squares predictor of Y based on X and say whether it is a linear function of X. (If it is, then the best linear predictor is in fact the best among all predictors.)

The least squares predictor of Y based on X, using the above, is

$$E[Y \mid X] = 50 \cdot \frac{X}{100} = \frac{X}{2},$$

which is a linear function of X.

(d) Find $Var(Y \mid X)$.

Based on part (b),

$$Var[Y \mid X] = 50 \cdot \frac{X}{100} \cdot \frac{100 - X}{100} \cdot \frac{100 - 50}{100 - 1}.$$

(e) Find $E(Z \mid X)$ and $Var(Z \mid X)$.

$Z = Y + W$, where W is the number of successes in trials 101 through 150, which has a Bin$(50, p)$ distribution, independently of X and Y. Hence,

$$E[Z \mid X] = E[Y + W \mid X] = E[Y \mid X] + E[W] = \frac{X}{2} + 50p,$$

and

$$Var[Z \mid X] = Var[Y + W \mid X] = Var[Y \mid X] + Var[W] = 50 \cdot \frac{X}{100} \cdot \frac{100 - X}{100} \cdot \frac{100 - 50}{100 - 1} + 50p(1 - p).$$

6. Predicting Max Based on Min

Let $U_1, U_2, \ldots, U_n$ be i.i.d. uniform $(0,1)$ variables. As usual, let $U_{(1)}$ be their minimum and $U_{(n)}$ their maximum.

(a) Draw n points on the unit interval, and use the diagram to guess the least squares predictor of $U_{(n)}$ based on $U_{(1)}$. Symmetry is your friend.

Given $U_{(1)}$, the joint distribution of the other $U_{(i)}$'s is that of the order statistics of $n-1$ uniforms on $(U_{(1)}, 1)$, so their conditional expectations are evenly spaced on that interval. This gives gaps of length $(1-U_{(1)})/n$ between the endpoint $U_{(1)}$, the conditional expectations in order and the other endpoint 1. In particular, $E[U_{(n)} \mid U_{(1)}] = 1 - (1-U_{(1)})/n$.

(b) Prove that your guess is correct.

The joint density of $(U_{(1)}, U_{(n)})$ is $f_{U_{(1)},U_{(n)}}(x,y) = n(n-1)(y-x)^{n-2}$, so the conditional density is

$$f_{U_{(n)}|U_{(n)}}(y \mid x) = \frac{f_{U_{(1)},U_{(n)}}(x,y)}{f_{U_{(1)}}(x)} = \frac{n(n-1)(y-x)^{n-2}}{n(1-x)^{n-1}} = (n-1)\cdot\frac{(y-x)^{n-2}}{(1-x)^{n-1}}.$$

This yields

$$E[U_{(n)} \mid U_{(1)} = x] = \int_x^1 y \cdot (n-1) \cdot \frac{(y-x)^{n-2}}{(1-x)^{n-1}}\, dy = \ldots = 1 - \frac{1-x}{n},$$

by integration by parts.

(c) Is the best predictor in Parts (a) and (b) a linear function of $U_{(1)}$? (If it is, then the best linear predictor is in fact the best among all predictors.)

Yes.

7. Mixtures and Moment Generating Functions

Start with two random variables X and Y, and construct the new random variable W as follows:

Toss a p-coin. If it lands heads, let $W = X$. If it lands tails, let $W = Y$.

(a) Write the cdf F_W of W in terms of the cdf F_X of X and the cdf F_Y of Y. The answer will show why the distribution of W is called a *mixture* of the distributions of X and Y. It is important to keep in mind that the value of W is either a value of X or a value of Y; it's not an average of the two. But the distribution of W is an average of the two distributions.

Let I be the result of the toss with $I = 1$ for heads, $I = 2$ for tails. Then for any z

$$\begin{aligned}
F_W(z) &= \mathbb{P}(W \le z) = \mathbb{P}(W \le z, I=1) + \mathbb{P}(W \le z, I=2) = \mathbb{P}(X \le z, I=1) + \mathbb{P}(Y \le z, I=2)\\
&= \mathbb{P}(X \le z)\mathbb{P}(I=1) + \mathbb{P}(Y \le z)\mathbb{P}(I=2) = p\mathbb{P}(X \le z) + (1-p)\mathbb{P}(Y \le z)\\
&= pF_X(z) + (1-p)F_Y(z)
\end{aligned}$$

(b) Write the mgf M_W of W in terms of the mgf M_X of X and the mgf M_Y of Y. You can assume that all the mgf's exist in an interval around 0, which is the condition needed for all the nice properties of mgf's to hold.

$$M_W(t) = E[e^{tW}] = E[E[e^{tW} \mid I]] = pE[e^{tX}] + (1-p)E[e^{tY}] = pM_X(t) + (1-p)M_Y(t)$$

(c) Suppose X is standard normal and Y is normal with mean $\mu > 0$ and variance 1. Could the distribution of W be normal? Why or why not?

The mgf of W by the above is

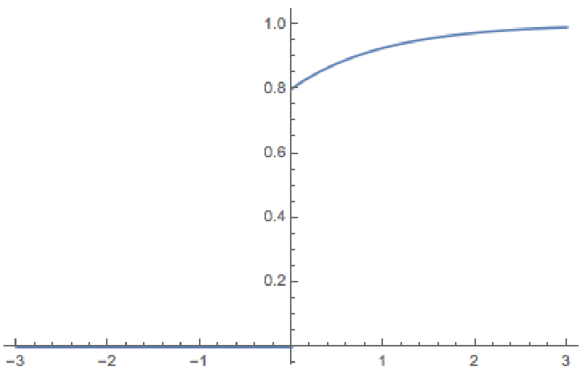
$$M_W(t) = pe^{t^2/2} + (1-p)e^{t^2/2+\mu t} = e^{t^2/2}(p + (1-p)e^{\mu t}). \tag{1}$$

On the other hand, $E[W] = (1-p)\mu$ and $Var[W] = 1 + \mu^2 p(1-p)$ by the laws of total expectation and variance. Hence, if W was normal, its mgf should be

$$\begin{aligned}\exp\left(\frac{Var[W]t^2}{2} + E[W]t\right) &= \exp\left(\frac{(1+\mu^2 p(1-p))t^2}{2} + (1-p)\mu t\right) \\ &= \exp\left(\frac{t^2}{2}\right) \cdot \exp\left(\frac{\mu^2 p(1-p)t^2}{2} + (1-p)\mu t\right),\end{aligned}$$

which is not the same function as that in (1), unless either p, $1-p$ or μ is 0. Thus, W cannot be normal apart from these extreme cases.

(d) Let V have cdf F_V defined by $F_V(v) = 1 - \frac{e^{-v}}{5}$ for $v \geq 0$ and 0 otherwise. Sketch (carefully) the graph of F_V. Then describe the distribution of V as a mixture of two distributions whose moment generating functions you know, and hence find M_V.



We can see that V has a mass of $\frac{4}{5}$ at 0, and behaves like an exponential variable besides that. Based on this and part (a), we can write F_V as a mixture in the following way:

$$1 - \frac{e^{-v}}{5} = \frac{4}{5} \cdot 1 + \frac{1}{5} \cdot (1 - e^{-v})$$

for $v \geq 0$, so with probability $p = 4/5$, $V = 0$ and with probability $1 - p = 1/5$, $V = Y$ with $Y \sim Exp(1)$. Hence, the mgf of V is

$$M_V(t) = pE[e^{t \cdot 0}] + (1-p)E[e^{tY}] = \frac{4}{5} + \frac{1}{5} \cdot \frac{1}{1-t}.$$

8. MGF of a Random Sum

Let N be a non-negative integer valued random variable, and let $X, X_1, X_2, \ldots$ be i.i.d. and independent of N. As in class, define the "random sum" S by

$$\begin{aligned} S &= 0 \ \text{ if } N = 0 \\ &= X_1 + X_2 + \cdots + X_n \ \text{ if } N = n > 0 \end{aligned}$$

(a) Let M be our usual notation for moment generating functions. By conditioning on N, show that

$$M_S(t) \quad = \quad M_N(\log M_X(t))$$

assuming that all the quantities above are well defined. [The identity $w = e^{\log(w)}$ might be handy.]

(b) Let N have the geometric (p) distribution on $\{1, 2, 3, \ldots\}$. Find the mgf of N. This doesn't use Part (a).

(c) Let $X_1, X_2, \ldots$ be i.i.d. exponential (λ) variables and let N be geometric as in Part (b). Use the results of Parts (a) and (b) to identify the distribution of S.

(d) Find the density of S by conditioning on N, and hence confirm the result of Part (c).

[Find $P(S \in ds)$ by conditioning on N.]

(e) Use a Poisson (λ) process and Poissonization (also known as "thinning" in the context of Poisson processes) to find yet another way of confirming the result of Part (c).