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 $SD(X)$.

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 $Var(H)$.

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 $Var(R)$.

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.]

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

(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.

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.

(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.

(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.)

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

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

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.

(b) Prove that your guess is correct.

(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.)

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.

(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.

(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?

(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.

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).