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

WEEK 4 EXERCISES

You are expected to do all these problems, but for **Homework 4** please turn in **only Problems 2, 3, 5, and 6** on **Thursday September 20 at the start of lecture**.

1. Flattened Die

Let $p \in (0, 1)$ and let X be the number of spots showing on a flattened die that shows its six faces according to the following chances:

- $P(X = 1) = P(X = 6)$
- $P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5)$
- $P(X = 1 \text{ or } 6) = p$

(a) Find $E(X)$.

(b) Find $SD(X)$. Explain algebraically and also by an intuitive argument why the answer is an increasing function of p.

2. Poisson Expectations

Let X have the Poisson (μ) distribution. Find

(a) $E(X + 1)$

(b) $E(1/(X + 1))$

3. Collecting Distinct Values

(a) A fair die is rolled n times. Find the expected number of times the face with six spots appears.

(b) A fair die is rolled n times. Find the expected number of faces that *do not* appear, and say what happens to this expectation as n increases.

(c) Use your answer to Part **b** to find the expected number of distinct faces that *do* appear in n rolls of a die.

(d) Find the expected number of times you have to roll a die till you have seen all of the faces. This is a version of what is known as the *collector's problem.* The collector is waiting to get a complete set.

4. Aces and Face Cards

A standard deck consists of 52 cards of which 4 are aces, 4 are kings, and 12 (including the four kings) are "face cards" (Jacks, Queens, and Kings).

Cards are dealt at random without replacement from a standard deck till all the cards have been dealt.

Find the expectation of the following. None of them requires a long calculation.

(a) the number of aces among the first 5 cards

(b) the number of face cards that *do not* appear among the first 13 cards

(c) the number of aces among the first 5 cards minus the number of kings among the last 5 cards

(d) the number of cards before the first ace

(e) the number of cards strictly in between the first ace and the last ace

(f) the number of face cards before the first ace

5. The Gamma Densities

In this problem you will start with some calculus exercises and then develop one of the fundamental families of densities.

(a) The *Gamma function* of mathematics is defined by

$$\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt, \qquad r > 0$$

That letter is the upper case Greek letter Gamma. You can assume that the integral converges and that therefore $\Gamma(r)$ is a positive number.

Use integration by parts to show that

$$\Gamma(r+1) = r\Gamma(r), \quad r > 0$$

(b) Use Part **a** and induction to show that if r is a positive integer then $\Gamma(r) = (r-1)!$. This is an indication that the Gamma function is a continuous extension of the factorial function.

(c) Let X have density given by

$$f_X(t) = \begin{cases} \frac{1}{\Gamma(r)}t^{r-1}e^{-t}, & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

We say that X has the *gamma* $(r, 1)$ density. In the case $r = 1$, this should be a density that you recognize. Provide its name and the appropriate parameters.

(d) Now fix $\lambda > 0$ and consider the function

$$f(t) = \begin{cases} \frac{\lambda^r}{\Gamma(r)}t^{r-1}e^{-\lambda t}, & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Show that this function is a density. It is called the gamma (r, λ) density.

(e) Use Part **d** to fill in the blank: For every $r > 0$,

$$\int_0^\infty t^{r-1}e^{-\lambda t}dt = \underline{\hspace{3cm}}$$

Now let Y have the gamma (r, λ) density. Use what you have shown in the previous parts to find $E(Y)$. Provide the numerical value of $E(Y)$ in the case $r = 2.2$ and $\lambda = 1.1$. You don't need a computer.

(f) Let Y have gamma (r, λ) density. Find $SD(Y)$.

6. Bounds

A random variable X, not necessarily non-negative, has $E(X) = 20$ and $SD(X) = 4$. For the bounds below, use 0 and 1 only if you feel you can't do better than those.

(a) Find upper and lower bounds for $P(0 < X < 40)$.

(b) Find upper and lower bounds for $P(10 < X < 40)$.

(c) Find an upper bound for $P(X \geq 40)$.

(d) Find an upper bound for $P(X^2 \geq 900)$.

7. Randomized Response

Survey respondents understandably don't like to answer questions about sensitive topics such as illegal drug use. If data scientists want to estimate the proportion of illegal drug users in a population, they have to devise methods of getting the information they need while maintaining the privacy of the individual respondents.

Randomized response schemes are often used in such situations. In one such scheme, each surveyed person is given a coin and asked to answer YES or NO after following these instructions out of sight of the surveyor:

- Toss the coin.
- If it lands heads, then truthfully answer, "Do you use illegal drugs?"
- If it lands tails, then toss it again and answer, "Did the second toss land heads?"

This way each respondent answers YES or NO but the surveyor doesn't know which question was answered. The data scientists then have to estimate the proportion of illegal drug users based on the overall proportion of YES answers, which includes the YES answers to the second question.

Let the unknown proportion of illegal drug users in a large population be p, and suppose a random sample of size n is surveyed using the scheme above. You can assume that the sampling is equivalent to drawing at random with replacement.

(a) Let X be the proportion of sampled people who answer YES. Find $E(X)$.

(b) Use X to construct an unbiased estimate of p.

8. Indicators and the Inclusion-Exclusion Formula

We guessed the general inclusion-exclusion formula (see Section 5.2 of the Prob 140 textbook) but we never proved it. Let's get that done.

(a) Let $x_1, x_2, \ldots, x_n$ be numbers. Expand the product $(1-x_1)(1-x_2)$ and then expand $(1-x_1)(1-x_2)(1-x_3)$ by using the expansion you got for $(1-x_1)(1-x_2)$. Now guess a formula for the expansion of the product

$$\prod_{i=1}^{n}(1 - x_i)$$

and use induction to prove it. The induction shouldn't take many steps. It consists of just two observations, both of which can be expressed in English without complicated notation.

[Parts **b** and **c** are on the next page.]

(b) Let $A_1, A_2, \ldots, A_n$ be events. For each i in the range 1 through n let I_i be the indicator of A_i. Let I be the indicator of $\cup_{i=1}^n A_i$. Explain why

$$I \;=\; 1 - \prod_{i=1}^{n} (1 - I_i)$$

(c) Use Parts **a** and **b** to establish the inclusion-exclusion formula.