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

WEEK 3 EXERCISES

You are expected to do all these problems, but for **Homework 3** please turn in **only Problems 2, 5, 9, and 10** on **Thursday September 13 at the start of lecture**.

1. Tossing a Random Coin

This week some calculations were left unfinished in lecture. Here are some for you to finish. You'll need a calculator, but it doesn't have to be a fancy one.

I have a bag of coins, of which 25% are fair and the rest land heads with chance 0.9. I pick a coin at random and toss it twice. Let R be the chance that the chosen coin lands heads, and let H be the number of heads in the two tosses. In class we saw the joint distribution of R and H, along with the marginals. We also saw the conditional distribution of R given each different value of H. These are shown in the tables below.

Joint Distribution and Marginals

	$R = 0.5$	$R = 0.9$	
$H = 2$	0.0625	0.6075	0.67
$H = 1$	0.1250	0.1350	0.26
$H = 0$	0.0625	0.0075	0.07
	0.2500	0.7500	

Conditional Distributions of R given values of H

	$R = 0.5$	$R = 0.9$
Dist of R given $H = 2$	0.093284	0.906716
Dist of R given $H = 1$	0.480769	0.519231
Dist of R given $H = 0$	0.892857	0.107143

(a) The top table shows $P(R = 0.9, H = 1) = 0.1350$. Show the calculation that leads to this.

$\mathbb{P}(R = 0.9, H = 1) = \mathbb{P}(R = 0.9) \cdot \mathbb{P}(H = 1 \mid R = 0.9) = 0.75 \cdot \binom{2}{1} \cdot 0.9 \cdot 0.1 = 0.135$

(b) The bottom table shows $P(R = 0.9 \mid H = 2) = 0.906716$. Show how this results from values in the top table.

$\mathbb{P}(R = 0.9 \mid H = 2) = \frac{\mathbb{P}(R=0.9,H=2)}{\mathbb{P}(H=2)} = \frac{0.6075}{0.67} = 0.906716$

(c) The chosen coin is going to be tossed a third time. Let A be the event that the third toss lands heads. Find $P(A \mid H = 2)$ and also $P(A \mid H = 0)$.

$\mathbb{P}(A \mid H = 2) = \frac{\mathbb{P}(A,H=2)}{\mathbb{P}(H=2)} = \frac{\mathbb{P}(A,H=2 \mid R=0.5)\cdot\mathbb{P}(R=0.5)+\mathbb{P}(A,H=2 \mid R=0.9)\cdot\mathbb{P}(R=0.9)}{\mathbb{P}(H=2)} = \frac{0.5^3\cdot 0.25+0.9^3\cdot 0.75}{0.67} = 0.862687$

$\mathbb{P}(A \mid H = 0) = \frac{\mathbb{P}(A,H=0)}{\mathbb{P}(H=0)} = \frac{\mathbb{P}(A,H=0 \mid R=0.5)\cdot\mathbb{P}(R=0.5)+\mathbb{P}(A,H=0 \mid R=0.9)\cdot\mathbb{P}(R=0.9)}{\mathbb{P}(H=0)} = \frac{0.5\cdot 0.5^2\cdot 0.25+0.9\cdot 0.1^2\cdot 0.75}{0.07} = 0.542857$

(d) Are the three tosses mutually independent?

No, as $\mathbb{P}(A \mid H = 2) > \mathbb{P}(A \mid H = 0)$, that is, tossing heads increases the chance that we will toss heads later. This is because tossing heads increases the chance that we picked a coin with $R = 0.9$. (The tosses are only conditionally independent given R.)

2. Joint, Marginal, and Conditional Densities

And here's more from lecture for you to finish.

The random variables X and Y have joint density given by

$$f(x,y) \;=\; \begin{cases} 30(y-x)^4, & 0<x<y<1 \\ 0 \;\text{ elsewhere} \end{cases}$$

(a) Let f_X be the density of X and f_Y the density of Y. Find f_X and f_Y. Check that the answers you get are non-negative and integrate to 1.

$f_X(x) = \int_x^1 30(y-x)^4 dy = 6(y-x)^5\big|_{y=x}^1 = 6(1-x)^5$ for $0 \le x \le 1$.

$f_Y(y) = \int_0^y 30(y-x)^4 dx = -6(y-x)^5\big|_{x=0}^y = 6y^5$ for $0 \le y \le 1$.

(b) Find the cdf of X by using f_X from Part (a).

$F_X(z) = \int_0^z f_X(x)dx = \int_0^z 6(1-x)^5 dx = -(1-x)^6\big|_{x=0}^z = 1-(1-z)^6$

(c) Find the cdf of X again, this time by using the joint density f directly to find the relevant probabilities.

$F_X(z) = \mathbb{P}(X \le z) = \int_0^z \int_x^1 30(y-x)^4 dydx = \int_0^1 \int_0^{\min(y,z)} 30(y-x)^4 dxdy$, both double integrals give the formula that we got in part (b).

(d) Find the conditional density of Y given $X = 0.4$. In class we denoted this as $f_{Y|X=0.4}$.

$f_{Y|X=0.4}(y) = \frac{f(0.4,y)}{f_X(0.4)} = \frac{30(y-0.4)^4}{6 \cdot 0.6^5}$ for $0.4 < y < 1$.

(f) Find the conditional density of X given $Y = 0.4$.

$f_{X|Y=0.4}(x) = \frac{f(x,0.4)}{f_Y(0.4)} = \frac{30(0.4-x)^4}{6 \cdot 0.4^5}$ for $0 < x < 0.4$.

3. Recognizing Densities

The following observation can save a lot of calculation.

(a) Let f and g be two probability densities such that $f(x) = cg(x)$ for all x for some positive constant c. What must the value of c be, and why?

$1 = \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} cg(x)dx = c\int_{-\infty}^{\infty} g(x)dx = c \cdot 1 = c$, therefore $c = 1$.

(b) The joint density of X and Y is

$$f(x,y) \;=\; \begin{cases} ce^{-2x-3y}, \; x>0, \; y>0 \\ 0 \;\text{ elsewhere} \end{cases}$$

Without any calculation, answer the following questions. You'll need some facts from previous lectures.

(i) Are X and Y independent?

Yes, since $f(x,y) = ce^{-2x-3y}\mathbf{1}\{x>0, y>0\}$ can be written as a product of a function of x and a function of y, say $ce^{-2x}\mathbf{1}\{x>0\}$ and $e^{-3y}\mathbf{1}\{y>0\}$.

(ii) What are the marginal densities of X and Y?

By the above product form of $f(x,y)$, the marginal density of X is a constant multiple of $ce^{-2x}\mathbf{1}\{x>0\}$, hence it has to be (by part (a)) $2e^{-2x}\mathbf{1}\{x>0\}$, so $X \sim Exp(2)$. Similarly, the marginal density of Y is a constant multiple of $e^{-3y}\mathbf{1}\{y>0\}$, hence it is $3e^{-3y}\mathbf{1}\{y>0\}$, so $Y \sim Exp(3)$.

(iii) What is the value of c?

By the results of parts (i) and (ii), $f(x,y) = 2e^{-2x}\mathbf{1}\{x>0\} \cdot 3e^{-3y}\mathbf{1}\{y>0\} = 6e^{-2x-3y}\mathbf{1}\{x>0, y>0\}$, so $c = 6$.

4. Split in Roulette

If you bet a dollar on a "split" at roulette, you have a 2/38 chance of winning. If you win the bet, your net gain is \$17. If you lose the bet, you lose your dollar.

Suppose you make 100 \$1 bets on a "split", and assume the bets are independent of each other. Write a numerical formula for the chance that you come out ahead, that is, the chance that you end up with more money than you had at the start.

You come out ahead exactly if you win at least 6 times out of the 100 bets. The number of bets that you win (X) has Binomial(100, 2/38) distribution. Hence

$$\mathbb{P}(X \geq 6) = 1 - \mathbb{P}(0 \leq X \leq 5) = 1 - \sum_{k=0}^{5} \binom{100}{k} \left(\frac{2}{38}\right)^k \left(\frac{36}{38}\right)^{100-k}.$$

5. Categories in a Sample

In a population, 30% of the individuals are red, 50% are green, and 20% are blue. I draw a random sample of size 20 with replacement from this population. Let N_r be the number of red individuals drawn, N_g the number green, and N_b the number blue, so that $N_r + N_g + N_b = 20$.

In each part below, provide the name of the distribution as well as all the numerical parameters.

(a) What is the distribution of N_r?

Binomial(20, 0.3)

(b) What is the distribution of $N_r + N_g$?

Binomial(20, 0.8), since the chance of drawing a red or green individual is 0.8.

(c) What is the joint distribution of N_r, N_g, and N_b?

Multinomial(20; 0.3, 0.5, 0.2)

(d) What is the conditional distribution of N_r given $N_r + N_g = 15$?

$$\begin{aligned}\mathbb{P}(N_r = n_r \mid N_r + N_g = 15) &= \frac{\mathbb{P}(N_r = n_r, N_g = 15 - n_r, N_b = 5)}{\mathbb{P}(N_r + N_g = 15)} = \frac{\binom{20}{5}\binom{15}{n_r}0.3^{n_r} \cdot 0.5^{15-n_r} \cdot 0.2^5}{\binom{20}{15}0.8^{15} \cdot 0.2^5} \\ &= \binom{15}{n_r}\left(\frac{0.3}{0.8}\right)^{n_r}\left(\frac{0.5}{0.8}\right)^{15-n_r},\end{aligned}$$

hence the conditional distribution of N_r given $N_r + N_g = 15$ is Binomial(15, 3/8).

6. Randomized Experiments

(a) There are 100 men and 100 women who have agreed to participate in a randomized controlled experiment. The researchers are going to pick 100 participants at random without replacement to form the treatment group. Write an expression for the chance that the treatment group contains at least 45 people of each of the two genders.

$\sum_{k=45}^{55} \frac{\binom{100}{k}\binom{100}{100-k}}{\binom{200}{100}}$

(b) Of 45 patients in a randomized controlled experiment on the effectiveness of a new medication, 25 are assigned randomly to the treatment group and the remaining 20 to the control group. At the end of the experiment, 22 of the 25 patients in the treatment group are cured, compared to 9 of the 20 patients in the control group.

If you randomly pick 25 of the 45 patients, what is the chance that you will get 22 or more of the 31 patients who were cured?

$\sum_{k=22}^{25} \frac{\binom{31}{k}\binom{14}{25-k}}{\binom{45}{25}}$

(c) You don't have to calculate the decimal value of the chance in Part (b), but if you do, you will see that it is very small. What does this tell you about the medication? (This way of making inferences is called *Fisher's Exact Test* for the hypothesis that the medication performs the same as the control.)

This tells us that the medication is efficient with a high significance.

7. Combinatorics

A standard deck consists of 52 cards.

• There are 13 cards in each of 4 *suits*: hearts, diamonds, spades, and clubs. Hearts and diamonds are red and clubs and spades are black.

• In each suit, there is one card of each of 13 *ranks*: Ace, 2, 3, 4, . . . 10, Jack, Queen, King.

A *poker hand* consists of 5 cards dealt at random without replacement from the deck. Recall from Week 1 Exercises that the hand is called a simple random sample of cards.

(a) How many poker hands are there?

$\binom{52}{5}$

(b) Find the chance of *four of a kind*: ranks a, a, a, a, b

$\frac{13 \cdot 48}{\binom{52}{5}}$

(c) Find the chance of a *full house*: ranks a, a, b, b, b.

$\frac{13 \cdot \binom{4}{2} \cdot 12 \cdot \binom{4}{3}}{\binom{52}{5}}$

(d) Find the chance (carefully!) of *two pair*: ranks a, a, b, b, c

$\frac{\binom{13}{2}\binom{4}{2}\binom{4}{2} \cdot 44}{\binom{52}{5}}$

8. Lie Detector Tests

[Pitman 2.rev.10] According to a newspaper report, in 2 million lie detector tests, 300,000 were estimated to have produced erroneous results. Assuming these figures to be correct, answer the following:

(a) If ten tests were picked at random from these 2 million tests, what would be the chance that at least one of them produced an erroneous result? Sketch the histogram of the distribution of the number of erroneous results among these ten tests.

Let X be the number of tests among the 10 that we picked that produced an erroneous result. Then

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(0 \text{ erroneous results among the 10 tests}) = 1 - \frac{\binom{1700000}{10}}{\binom{2000000}{10}}.$$

In general, $X \sim HyperGeo(2000000, 300000, 10)$. Since 10 is negligible compared to 300,000, for a good a approximation we can assume that we sample the ten tests with replacement. This gives that approximately $X \sim Binomial(10, 3/20)$.

(b) Suppose these 2 million tests were done on a variety of machines. If a machine were picked at random, then ten tests picked at random from these tests performed on that machine, would it be reasonable to suppose that the chance that at least one of them produced an erroneous result would be the same as in a)?

No, since different machines could produce erroneous results with very different frequencies.

9. Accurate Test

A test for a disease produces a correct result with chance 0.99. Suppose the test is run on 300 patients and that the results of the patients are independent of each other.

(a) Find a formula for the exact chance that the result is correct for at least 295 patients.

$\sum_{k=295}^{300} \binom{300}{k} 0.99^k \cdot 0.01^{300-k}$

(b) Justify a Poisson approximation for the chance in (a), and find the value of the approximation.

The number of incorrect results (X) follows a Binomial(300, 0.01) distribution. Since 300 is large, 0.01 is small, and $300 \cdot 0.01 = 3$, a good approximation for the distribution of X (for small values) is Poi(3). Hence, the chance in (a) is

$$\mathbb{P}(X \leq 5) \approx \sum_{k=0}^{5} \frac{3^k}{k!} e^{-3} \approx 0.916082.$$

10. Poisson Facts

Let X_1 and X_2 be independent random variables such that X_i has the Poisson (μ_i) distribution for $i = 1, 2$.

(a) Use consecutive odds ratios to find the mode of the distribution of X_1. Be careful about the case where μ_1 is an integer.

$$R_k := \frac{\mathbb{P}(X_1 = k)}{\mathbb{P}(X_1 = k-1)} = \frac{\frac{\mu_1^k}{k!} e^{-\mu_1}}{\frac{\mu_1^{k-1}}{(k-1)!} e^{-\mu_1}} = \frac{\mu_1}{k}$$

When μ_1 is not an integer, $R_k > 1$ if $\mu_1 > k$ and $R_k < 1$ otherwise. Hence, $P(X_1 = k)$ attains its maximum at $k = \lfloor \mu_1 \rfloor$.

When μ_1 is an integer, $R_k > 1$ if $\mu_1 > k$, $R_k = 1$ if $\mu_1 = k$ and $R_k < 1$ if $\mu_1 < k$. Hence, $\mathbb{P}(X_1 = k)$ attains its maximum at two places: $k = \mu_1$ and $k = \mu_1 - 1$.

(b) For $m > 0$, find the conditional distribution of X_1 given that $X_1 + X_2 = m$. Recognize this as one of the famous distributions and provide its name and parameters.

Using that $X_1 + X_2 \sim Poi(\mu_1 + \mu_2)$,

$$\begin{aligned}
\mathbb{P}(X_1 = k \mid X_1 + X_2 = m) &= \frac{\mathbb{P}(X_1 = k, X_2 = m-k)}{\mathbb{P}(X_1 + X_2 = m)} = \frac{\mathbb{P}(X_1 = k)\mathbb{P}(X_2 = m-k)}{\mathbb{P}(X_1 + X_2 = m)} \\
&= \frac{\frac{\mu_1^k}{k!} e^{-\mu_1} \frac{\mu_2^{m-k}}{(m-k)!} e^{-\mu_2}}{\frac{(\mu_1+\mu_2)^m}{m!} e^{-(\mu_1+\mu_2)}} = \frac{m!}{k!(m-k)!} \cdot \frac{\mu_1^k \mu_2^{m-k}}{(\mu_1 + \mu_2)^m} \\
&= \binom{m}{k} \left(\frac{\mu_1}{\mu_1 + \mu_2} \right)^k \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^{m-k}
\end{aligned}$$

for $0 \leq k \leq m$. Hence, the conditional distribution of X_1 given that $X_1 + X_2 = m$ is $Binomial(m, \frac{\mu_1}{\mu_1+\mu_2})$.