Homework 2

Joint, Marginal and Conditional

Jaiden Atterbury

Instructions

Please answer the following questions in the order in which they are posed. Add a few empty lines below each and write your answers there. Focus on answering in complete sentences. You will also need scratch paper/pen to work out the answers before typing it.

For help with formatting documents in RMarkdown, please consult R Markdown: The Definitive Guide. Another option is to search using Google.

Exercises

1. (Joint PMF) An urn contains four red chips, three white chips, and two blue chips. A random sample of size 3 is drawn without replacement. Let X denote the number of white chips in the sample and Y the number of blue chips. Write a formula for the joint PMF of X and Y. Be sure to explain the components of your formula and why it makes sense.

   Hint: You will first want to make a table on scratch paper to understand the possible values for x and y and their corresponding probabilities. Then the formula for the PMF will be evident.

Let X denote the number of white chips in the sample and Y denote the number of blue chips in the sample. In order to find a formula for the joint PMF of X and Y we will first want to make a table to understand the possible values for x and y and their corresponding probabilities.

Joint PMF of X and Y (table):

	x			
y	0	1	2	3
0	0.048	0.214	0.143	0.012
1	0.143	0.286	0.036	0
2	0.048	0.036	0	0

Explanation/Derivation of the joint PMF formula
In each cell of the above table there are 4 components to the probability calculation. Since the urn contains four red chips, three white chips, and two blue chips, if we sample 3 chips at random without replacement, we know that X and Y can possibly take on the values: $x = 0, 1, 2, 3$ and $y = 0, 1, 2$. Now, when calculating the probability for each (x, y) pair, there are 4 things we must keep track of: 1.) how many ways we can select x white chips from the 3 in the urn, 2.) how many ways we can select y blue chips from the 2 in the urn, 3.) how many ways we can select $4 - x - y$ red chips from the 4 in the urn, and lastly 4.) how many ways we can select 3 chips from the 9 total chips in the urn. It turns out that we can calculate 1.) - 4.) using the binomial coefficient in each case.

Joint PMF of X and Y (formula):
Using the above table and derivation, the explicit formula for the joint PMF of X and Y turns out to be:

$$\boxed{f(x,y) = \frac{\binom{4}{4-x-y}\binom{3}{x}\binom{2}{y}}{\binom{9}{3}},\ x = 0,1,2,3 \text{ and } y = 0,1,2}.$$

2. (Multinomial) Let $U_1, U_2, ..., U_{1029}$ be independent uniformly distributed random variables. Let X_1 equal the number of U_i less than .331, X_2 equal the number between .331 and .820, and X_3 equal the number greater than .820.

   a. Calculate, using, R the probability of observing $X_1 = 354, X_2 = 492, X_3 = 183$. Be sure to show your code. State the joint distribution before launching into calculations.

Setup:
Since $U_1, U_2, ..., U_{1029}$ are independent uniformly distributed random variables on the interval $[0, 1)$, and X_1 equals the number of U_i less than .331, X_2 equals the number between .331 and .820, and X_3 equals the number greater than .820, we want to find the probability of landing into each of these categories. In order to find these probabilities we must use the fact that $U_i \sim Unif(0,1)$ and that the PDF of a Uniform random variable on the interval $[a, b)$ is $f(x) = \frac{1}{b-a},\ a \le x < b$.

Finding Probabilities:
Let π_1, π_2, and π_3 be the probability of landing in the ranges represented by X_1, X_2, and X_3. Then it follows from the definition of the PDF that the PDF of U_i is $f(u) = 1,\ 0 \le u < 1$. Thus it follows that π_1, π_2, and π_3 are:

$$\begin{aligned}
\pi_1 &= \int_a^b f(u)du \\
&= \int_0^{0.331} 1du \\
&= [u]_0^{0.331} \\
&= 0.331 - 0 \\
&= \boxed{0.331}
\end{aligned}$$

$$\begin{aligned}
\pi_2 &= \int_a^b f(u)du \\
&= \int_{0.331}^{0.820} 1du \\
&= [u]_{0.331}^{0.820} \\
&= 0.820 - 0.331 \\
&= \boxed{0.489}
\end{aligned}$$

$$
\begin{aligned}
\pi_3 &= \int_a^b f(u)du \\
&= \int_{0.820}^{1} 1du \\
&= [u]_{0.820}^{1} \\
&= 1 - 0.820 \\
&= \boxed{0.180}
\end{aligned}
$$

Assumptions
Since we are looking for the probability of finding a certain combination of the X_1, X_2, and X_3 values, it follows that the multinomial distribution is probably the best fit for this problem. However, before we blindly use this distribution to model our situation, let's check our assumptions first:

1.) There are a fixed number, $n = 1029$, of "trials." Where each trial represents a draw from $Unif(0,1)$.

2.) Probabilities are fixed for each X_i. In particular, $\pi_1 = 0.331$, $\pi_2 = 0.489$, and $\pi_3 = 0.180$.

3.) There are a fixed number of outcomes. In particular, we have k = 3 outcomes, where X_1 = draw in $[0, 0.331)$, X_2 = draw in $[0.331, 0.820]$, and X_3 = draw in $(0.820, 1)$.

4.) Trials are independent of each other. In our situation, since it was previously stated that the U_i's are independently distributed, each draw of the $Unif(0,1)$ has no impact on any previous or future draws of the $Unif(0,1)$.

Since $< X_1, X_2, X_3 >$ satisfies all of the assumptions of using the multinomial distribution, it follows that: $< X_1, X_2, X_3 > \sim Multinomial(n = 1029, \pi =< 0.331, 0.489, 0.120 >)$. The PDF of $< X_1, X_2, X_3 >$ is $f(x_1, x_2, x_3) = \frac{1029!}{x_1!x_2!x_3!}(0.331)^{x_1}(0.489)^{x_2}(0.120)^{x_3}$ where $x_1 + x_2 + x_3 = n$ and $\pi_1 + \pi_2 + \pi_3 = 1$.

Calculate the probability of observing $X_1 = 354, X_2 = 492, X_3 = 183$:

```
prob_of_event <- dmultinom(x = c(354, 492, 183),
                           size = 1029,
                           prob = c(0.331, 0.489, 0.180))
```

Thus, the probability of observing the event $X_1 = 354, X_2 = 492, X_3 = 183$ is 6.0646049×10^{-4}.

b. Calculate, using R as a calculator, the expected values and standard deviation of X_2? Be sure to show your code. State the marginal distribution before launching into calculations.

Since Theorem 14.1 tells us that the marginal distribution of X_i is $Binom(n, \pi_i)$, it follows that $X_2 \sim Binom(n = 1029, \pi_2 = 0.489)$. Thus to find $E[X_2]$ and $SD[X_2]$, we can use the facts that when $X \sim Binom(n, \pi)$, $E[X] = n\pi$ and $Var[X] = n\pi(1-\pi)$, as well as the fact that $SD[X] = +\sqrt{Var[X]}$.

Calculating $E[X_2]$ and $SD[X_2]$:

```
ex_val <- 1029 * 0.489
var <- 1029 * 0.489 * (1- 0.489)
sd <- sqrt(var)
```

Thus, from the above R calculations, $E[X_2] = 503.181$ and $SD[X_2] = 16.035133$.

3. (Marginal PMF) Let X and Y be discrete random variables with joint PMF

$$f(x, y) = \left(\frac{\lambda}{2}\right)^y \frac{e^{-\lambda}}{x!\,(y-x)!}$$

where x and y are (non-negative) integers and $0 \le x \le y$. That is, $x, y = 0, 1, 2, 3, \dots$ but with the constraint $0 \le x \le y$.

Determine $f_2(y)$, the marginal distribution of Y. Is this a familiar distribution? Show the steps.

Hint: you will need to use the Binomial theorem with $a = 1$ and $b = 1$ to perform the summation over x.

Finding $f_2(y)$ of $f(x, y)$:

$$\begin{aligned}
f_2(y) &= \sum_x f(x,y) \\
&= \sum_{x=0}^{y} \left(\frac{\lambda}{2}\right)^y \frac{e^{-\lambda}}{x!\,(y-x)!} \quad \text{Since } 0 \le x \le y \\
&= e^{-\lambda}(\frac{\lambda}{2})^y \sum_{x=0}^{y} \frac{1}{x!(y-x)!} \\
&= \frac{1}{y!} e^{-\lambda}(\frac{\lambda}{2})^y \sum_{x=0}^{y} \frac{y!}{x!(y-x)!} \\
&= \frac{e^{-\lambda}(\frac{\lambda}{2})^y}{y!} \sum_{x=0}^{y} \binom{y}{x}
\end{aligned}$$

Notice $\sum_{x=0}^{y} \binom{y}{x}$ can be solved using the binomial theorem with $n = y$, $a = 1$, and $b = 1$. Thus $(a+b)^n$ is equivalent to $(1+1)^y$ in our case, which equals 2^y.

$$\begin{aligned}
&= \frac{e^{-\lambda}(\frac{\lambda}{2})^y 2^y}{y!} \\
&= \frac{e^{-\lambda}(\frac{\lambda}{2} \cdot 2)^y}{y!} \\
&= \boxed{\frac{e^{-\lambda}\lambda^y}{y!}}
\end{aligned}$$

Thus, we can see that $\boxed{f_2(y) \sim Poisson(\lambda)}$, which is indeed a familiar distribution.

4. (Hierarchical model) Suppose a player is equally likely to have 4, 5 or 6 at-bats (opportunities to bat) in a baseball game. If X is the number of opportunities to bat, then we are assuming that

$$f_1(x) = P(X = x) = \frac{1}{3}, \quad x = 4, 5, 6.$$

Suppose Y, the number of hits, is a Binomial random variable with size $X = x$ and probability of success $\pi = 0.3$. That is

$$f(y|x) = P(Y = y|X = x) = Binom(x, 0.3).$$

a. Fill in the numbers for the joint PMF, $f(x, y) = P(X = x, Y = y)$ in the cells indicated by (i) − (v). Also fill in the number for the marginal PMF $f_2(y)$ in the cell indicated by (vi). Show work below the table so we know you are not just guessing.

	y						
x	0	1	2	3	4	5	6
4			0.0882		0.0027	0	
5			0.1029				
6			0.0108045				
$f_2(y)$			0.299145				

Facts used to calculate probabilities:

1.) $f(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{f(x,y)}{\frac{1}{3}} = 3f(x,y)$.

2.) $f(x,y) = \frac{f(y|x)}{3}$.

3.) $f(y|x) = Binom(x, 0.3)$.

4.) $f(x,y) = \frac{Binom(x,0.3)}{3}$.

5.) If $Y \sim Binom(x, 0.3)$, then $f(y) = \binom{x}{y}(0.3)^y(0.7)^{x-y},\ y = 0, 1, 2, 3, 4, 5, 6$.

Calculating probabilties:

Using facts 4.) and 5.) we can easily calculate the probabilities in the table:

1.) $f(4,2) = \frac{\binom{4}{2}(0.3)^2(0.7)^2}{3} = 0.0882$.

2.) $f(5,2) = \frac{\binom{5}{2}(0.3)^2(0.7)^3}{3} = 0.1029$.

3.) $f(6,2) = \frac{\binom{6}{2}(0.3)^2(0.7)^4}{3} = 0.108045$.

4.) $f_2(2) = f(4,2) + f(5,2) + f(6,2) = 0.0882 + 0.1029 + 0.108045 = 0.299145$.

5.) $f(4,4) = \frac{\binom{4}{4}(0.3)^4(0.7)^0}{3} = 0.0081$.

6.) $f(4,5) = 0$ since you can't have 5 hits on only 4 at-bats.

b. Write the conditional distribution of X given $Y = 2$. That is

$$f(x|y = 2) = P(X = x|Y = 2).$$

Again, be sure to state any formulas you are plugging into so we know you are not guessing.

Facts used to calculate conditional distribution:

1.) $f(x|y = 2) = \frac{f(x,y)}{f_2(y)} = \frac{f(x,2)}{f_2(2)} = \frac{f(x,2)}{0.299146}$.

2.) Facts 1-5 from part a.).

Calculating the conditional distribution:

$$\begin{aligned} f(x|y=2) &= \frac{f(x,2)}{0.299146} \\ &= \frac{\binom{x}{2}(0.3)^2(0.7)^{x-2}}{3(0.299145)} \\ &= 0.100286\binom{x}{2}(0.7)^{x-2} \\ &= \frac{0.100286\binom{x}{2}(0.7)^{x}}{0.7^2} \\ &= 0.204665\binom{x}{2}(0.7)^{x} \\ &= \frac{200000}{977207}\binom{x}{2}(0.7)^{x} \end{aligned}$$

Thus, $\boxed{f(x|y=2) = \dfrac{200000}{977207}\dbinom{x}{2}(0.7)^x,\ x=4,5,6}$.

Check to see if conditional distribution is correct:
Since $f(x=4|y=2)+f(x=5|y=2)+f(x=6|y=2) = \frac{120}{407}+\frac{140}{407}+\frac{147}{407} = 1$, we can see that $f(x|y=2)$ does in fact equal $\frac{200000}{977207}\binom{x}{2}(0.7)^x,\ x=4,5,6$.