Homework 2 KEY

Joint, Marginal and Conditional

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

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{4}{3-x-y}}{\binom{9}{3}}, \quad x = 0,1,2,3, \ y = 0,1,2 \ , x+y \leq 3$$

In this expression we have that in the numerator $\binom{3}{x}$ denotes the number of ways we can draw x white chips from the 3 white chips in the urn. Similarly $\binom{2}{y}$ is the number of ways to choose y blue chips from the urn. The last term $\binom{4}{3-x-y}$ is how many ways to draw the red chips from the remaining 3 draws that are not blue or white. The numerator thus denotes the total number of ways to draw x white chips, y blue chips, and 3-x-y red chips from the urn. The denominator represents the total number of different ways to draw 3 chips from the 9 chips in the urn. Thus dividing the two quantities gives us our PMF.

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.

   We have that each $U_i \sim Unif(0,1)$. By properties of the Uniform distribution we know that $P(U_i < u) = u$ for Uniform(0,1) random variables.

   Thus the probability that a given U_i is under .331 is .331. Similarly, $P(.331 < x < .820) = .820 - .331 = 0.489$. And $P(U_i > .820) = 1 - P(U_i < .82) = .18$.

   Thus for each U_i, we know it can fall in one of those three categories. Additionally we have that the probability of being in any of those 3 categories is equal across the 1029 variables. Since 1029 is a fixed number, and each of the RV's is independent, we can thus model this as a multinomial random variable. More specifically we have:

$$(X_1, X_2, X_3) \sim Multinom(n = 1029, \pi = (.331, .489, .18))$$

To find $P(X_1 = 354, X_2 = 492, X_3 = 183)$ we can use the dmultinom function in R.

```
p2_prob <- dmultinom(c(354,492,183), size=1029,prob = c(.331,.489,.18))
p2_prob
```

```
## [1] 0.0006064605
```

So $P(X_1 = 354, X_2 = 492, X_3 = 183) = 6.0646049 \times 10^{-4}$

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.

We have by properties of Multinomial distributions that:

$$X_2 \sim Binom(n = 1029, \pi = .489)$$

So we have that $E(X_2) = n\pi = 503.181$ and $SD(X_2) = \sqrt{n\pi(1-\pi)} = 16.035133$

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, $x, y = 0, 1, 2, 3, \ldots$ 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.

$$\begin{aligned}
f_2(y) &= \sum_{x=1}^{y} (\frac{\lambda}{2})^y \frac{e^{-y}}{x!(y-x)!} \\
&= (\frac{\lambda}{2})^y e^{-\lambda} \sum_{x=1}^{y} \frac{1}{x!(y-x)!} \\
&= (\frac{\lambda}{2})^y \frac{e^{-\lambda}}{y!} \sum_{x=1}^{y} \frac{y!}{x!(y-x)!} \\
&= (\frac{\lambda}{2})^y \frac{e^{-\lambda}}{y!} \sum_{x=1}^{y} \binom{y}{x} 1^x 1^{y-x} \\
&= (\frac{\lambda}{2})^y \frac{e^{-\lambda}}{y!} \times (1+1)^y \text{using the binomial theorem} \\
&= (\frac{\lambda}{2})^y \frac{e^{-\lambda}}{y!} \times (2)^y \\
&= \frac{\lambda^y e^{-\lambda}}{y!} \text{ for y=0,1,2,3,\ldots}
\end{aligned}$$

We see that this a poisson distribution with rate parameter λ.

```
\emph{Hint: you will need to use the Binomial theorem with a = 1 and b = 1 to perform the summation o
```

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			0.108045				
$f_2(y)$			0.299145				

We have that $f(x, y) = f(y|x) \times f(x)$

Thus for the first 3 entries in the second column (i, iv, v) we are simply calculating

$$f(x, y = 2) = f(y = 2|x) \times \frac{1}{3}$$

For x=4,5,6.

For entry ii we are doing the same calculation with x=4, and y=4. For entry iii, we see that it is impossible to get y=5 hits with x=4 at bats, so the probability is 0. For the final entry vi, we simply add the entries in that column (i,iv,v) to find the marginal distribution of y at y=2.

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.

We have by properties of continuous distributions:

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

Thus using the values from the table in part a we have that $f(x|y = 2)$ is as follows:

	x=4	x=5	x=6
f(x|y=2)	0.2948403	0.3439803	0.3611794