Problem set 4: More summarizing distributions

Due October 23, 2023, at 9pm

(Your name here)

NOTE: Start with the file `ps4_2023_more_summarizing_distributions.qmd` (available from the github repository at https://github.com/UChicago-pol-methods/IntroQSS-F23/tree/main/assignments). Modify that file to include your answers. Make sure you can "render" the file (e.g. in RStudio by clicking on the `Render` button). Submit both the qmd file and the PDF via Canvas.

The entire problem set refers to the following joint distribution of two random variables X and Y:

$$f(x,y) = \begin{cases} 1/2 & x = 0, y = 0 \\ 1/6 & x = 1, y = 0 \\ 1/3 & x = 1, y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Part 1: Covariance and correlation

(1a) Compute the covariance of X and Y. Show your work.

Answer: I will show the result using both formulations.

First, compute $E[X] = 0 \times 1/2 + 1 \times 1/6 + 1 \times 1/3 = 1/2$.

Next, compute $E[Y] = 0 \times 1/2 + 0 \times 1/6 + 1 \times 1/3 = 1/3$.

By the first formulation, $\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$. This works out to $\text{Cov}[X, Y] = (1/2)(-1/2)(-1/3) + (1/6)(1/2)(-1/3) + (1/3)(1/2)(2/3) = 1/6$.

By the second formulation, $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$. $E[XY]$ is $1/3$. So $\text{Cov}[X, Y] = 1/3 - 1/6 = 1/6$.

(1b) Compute the correlation of X and Y. Show your work. (There is no need to repeat calculations from (1a).)

Answer: To skip a little work we could note that both X and Y are Bernoulli random variables, so $V[X] = (1/2)(1-1/2) = 1/4$ and $V[Y] = (1/3)(1-1/3) = 2/9$. We could also compute them as usual with the second variance formulation: $V[X] = E[X^2] - E[X]^2 = (1/2) - (1/2)^2 = 1/4$ and $V[Y] = E[Y^2] - E[Y]^2 = 1/3 - 1/9 = 2/9$.

Then $\text{Corr}[X,Y] = \frac{\text{Cov}[X,Y]}{\sigma[X]\sigma[Y]} = \frac{1/6}{\sqrt{1/4}\sqrt{2/9}} = \frac{1}{\sqrt{2}} \approx .707$

(1c) Write a function that takes as arguments `x` (a vector of possible values of a random variable X), `y` (a vector of possible values of a random variable Y), and `fxy` (a vector of frequencies $f(x,y)$ for each combination of x and y) and returns $\text{Cov}[X,Y]$. Use it to confirm the covariance calculation you did above.

```
my_cov <- function(x, y, fxy){
  ex <- sum(x*fxy)
  ey <- sum(y*fxy)
  sum(fxy*(x - ex)*(y - ey))
}
my_cov(x = c(0,1,1), y = c(0, 0, 1), fxy = c(1/2, 1/6, 1/3))
```

```
[1] 0.1666667
```

(1d) Write a function with the same arguments that computes correlation between two random variables X and Y. Use it to confirm the correlation calculation you did above.

```
my_var <- function(x, fx){
  ex <- sum(x*fx)
  sum(fx*(x-ex)^2)
}
my_corr <- function(x, y, fxy){
  my_cov(x, y, fxy)/(sqrt(my_var(x, fxy))*sqrt(my_var(y, fxy)))
}
my_corr(x = c(0,1,1), y = c(0, 0, 1), fxy = c(1/2, 1/6, 1/3))
```

```
[1] 0.7071068
```

Part 2: Conditional expectations, LIE, and BLP

(2a) What is $E[Y \mid X = 0]$? What is $E[Y \mid X = 1]$?

Answer: $E[Y \mid X = 0] = 0$, because when $X = 0$, Y only takes on the value 0. More explicitly, $f_{Y|X}(y|x)$ where $x = 0$ is 1 for $y = 0$ and 0 otherwise, so $\sum_y y f_{Y|X}(y|x) = 0$.

$E[Y \mid X = 1] = 2/3$. More explicitly, $f_{Y|X}(y|x)$ where $x = 1$ is $2/3$ for $y = 1$, $1/3$ for $y = 0$, and 0 otherwise, so $\sum_y y f_{Y|X}(y|x) = 2/3$.

(2b) Show that $E[Y] = E[E[Y \mid X]]$, i.e. that the law of iterated expectations holds in this case.

Answer: Using the the conditional expectations in the previous response, $E[E[Y \mid X]]$ is $\frac{1}{2}0 + \frac{1}{2}\frac{2}{3} = \frac{1}{3}$. This is equal to $E[Y] = (1/2)0 + (1/6)0 + (1/3)1 = \frac{1}{3}$.

(2c) What is the best linear predictor (BLP) of Y given X? Express your answer in terms of an intercept α and a slope β.

Answer: By Theorem 2.2.21 of Aronow and Miller, $\beta = \frac{\text{Cov}[X,Y]}{\text{V}[X]}$ and $\alpha = E[Y] - \beta E[X]$. Using calculations above, $\beta = \frac{1/6}{1/4} = 2/3$ and $\alpha = \frac{1}{3} - \frac{2}{3}\frac{1}{2} = 0$.

(2d) How close does the BLP of Y given X come to the conditional expectation function (CEF) of Y given X?

Answer: In this case, the BLP *is* the CEF: the BLP at $X = 0$ and $X = 1$ is exactly the same as the conditional expectations at those points. Because X takes on only two values, it is possible to draw a line connecting $E[Y \mid X = x]$ for both values of x. Thus the CEF is a line, and the BLP matches that line.

Part 3: Law of total variance

(3a) What is $V[Y \mid X = 0]$? What is $V[Y \mid X = 1]$?

Answer: $V[Y \mid X = 0] = 0$, because Y takes only one value in that case. When $X = 1$, Y is a Bernoulli random variable with a mean of $2/3$. Therefore $V[Y \mid X = 1] = (2/3)(1 - 2/3) = 2/9$.

(3b) Compute the two components of the Law of Total Variance (i.e. Eve's law) and confirm that they add up to $V[Y]$.

Answer: The first component is $E[V[Y|X]]$. From the previous question we have $E[V[Y|X]] = \frac{1}{2}0 + \frac{1}{2}\frac{2}{9} = \frac{1}{9}$.

The second component is $V[E[Y|X]]$. From (2a) we have $E[Y \mid X = 0] = 0$ and $E[Y \mid X = 1] = 2/3$. So $E[E[Y|X]] = \frac{1}{2}0 + \frac{1}{2}\frac{2}{3} = \frac{1}{3}$. Therefore $V[E[Y|X]] = E[(E[Y|X] - E[E[Y|X]])^2] = \frac{1}{2}\frac{-1}{3}^2 + \frac{1}{2}\frac{1}{3}^2 = \frac{1}{9}$.

The sum is therefore $\frac{1}{9} + \frac{1}{9} = \frac{2}{9}$.

Y is a Bernoulli random variable that is 1 with probability 1/3, so $V[Y] = (1/3)(1 - 1/3) = 2/9$.