Section 2

Foundations of Statistical Inference (PLSC 503)

1/31/2019

Exercise 1

A researcher randomly picks one individual out of a population of 1000 people. In this population, 200 are Republicans, 400 are Democrats, and the remainder are Independents. The researcher records the Party ID of the randomly selected person.

1. Formally represent this random generative process as a probability space.
2. Formally define a random variable X that takes on the value 0 if the individual selected is a Democrat, 1 if the individual selected is an Independent, and 2 if the individual selected is a Republican.
3. What is the PMF of X?
4. Define or draw the CDF of X.
5. Compute (i) $\Pr(X \leq 1.34)$, and (ii) $\Pr(X < 1)$

Solution:

1.

Let R represent the outcome "Republican", D represent the outcome "Democrat", and I represent the outcome "Independent". Then we can define the probability space for the random generative process as the triple

$$(\Omega, S, P)$$

where the sample space is

$$\Omega = \{R, D, I\},$$

the event space is

$$S = \big\{\emptyset, \{R\}, \{D\}, \{I\}, \{R, D\}, \{R, I\}, \{D, I\}, \{R, D, I\}\big\},$$

and the probability measure consists of the following:

$$\begin{aligned}
&P(\emptyset) = 0,\\
&P\big(\{R\}\big) = 0.2,\\
&P\big(\{D\}\big) = 0.4,\\
&P\big(\{I\}\big) = 0.4,\\
&P\big(\{R, D\}\big) = P\big(\{R\}\big) + P\big(\{D\}\big) = 0.2 + 0.4 = 0.6,\\
&P\big(\{R, I\}\big) = P\big(\{R\}\big) + P\big(\{I\}\big) = 0.2 + 0.4 = 0.6,\\
&P\big(\{D, I\}\big) = P\big(\{D\}\big) + P\big(\{I\}\big) = 0.4 + 0.4 = 0.8,\\
&P\big(\{R, D, I\}\big) = P\big(\{R\}\big) + P\big(\{D\}\big) + P\big(\{I\}\big) = 0.2 + 0.4 + 0.4 = 1.
\end{aligned}$$

2.

Let X take on the following values to represent the outcome of the random generative process referenced in part 1, i.e. $\Omega = \{R, D, I\}$, and $X(\omega) = \omega, \forall \omega \in \Omega$. For $\omega \in \Omega$, the random variable X is

$$X(\omega) = \begin{cases} 0 & \text{if } \omega = D \\ 1 & \text{if } \omega = I \\ 2 & \text{if } \omega = R. \end{cases}$$

3.

The PMF of X is

$$f(x) = \begin{cases} 0.4 & : x = 0 \\ 0.4 & : x = 1 \\ 0.2 & : x = 2 \\ 0 & : \text{otherwise.} \end{cases}$$

4.

The CDF of X is

$$F(x) = \begin{cases} 0.4 & : 0 \leq x < 1 \\ 0.8 & : 1 \leq x < 2 \\ 1 & : x \geq 2 \\ 0 & : \text{otherwise.} \end{cases}$$

5.

(i) $\Pr[X < 1.34] = \sum_{x=0}^{1} f(x) = 0.4 + 0.4 = 0.8.$

(ii) $\Pr[X \leq 1] = \sum_{x=0}^{1} f(x) = 0.4 + 0.4 = 0.8.$

(iii) $\Pr[X < 1] = f(0) = 0.4.$

Exercise 2

Consider two events A and B such that $\Pr(A) = 1/2$ and $\Pr(B) = 1/3$. Find $\Pr(A \cap B)$ for each of these cases:

1. A and B are disjoint.
2. A and B are independent.
3. $B \subset A$.
4. $\Pr(A^C \cap B) = 1/7$.
5. $\Pr(B|A) = 1/2$.

Solution:

1. $P(A \cap B) = P(\emptyset) = 0$. (Definition of empty sets: $P(\emptyset) = 0$.)
2. $P(A \cap B) = P(A)P(B) = \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{1}{6}$. (See A&M, Definition 1.1.15: Events $A, B \in S$ are independent if $P(A \cap B) = P(A)P(B)$.)
3. If $B \subset A, A \cap B = B$. Thus $P(A \cap B) = P(B) = \frac{1}{3}$.

4.
$$\begin{aligned}
P(B) &= P(A \cap B) + P(A^C \cap B) \\
\frac{1}{3} &= P(A \cap B) + \frac{1}{7}.
\end{aligned}$$

Rearranging,

$$P(A \cap B) = \frac{4}{21}.$$

5. Recall the definition of conditional probability (A&M, Definition 1.1.8):

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Rearranging and substituting with given values, we get:

$$\begin{aligned}
P(A \cap B) &= P(B|A)P(A) \\
&= \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \\
&= \frac{1}{4}.
\end{aligned}$$

(Aronow and Miller refer to the rearranged version of conditional probability as the *Multiplicative Law of Probability*, Theorem 1.1.9.)

Exercise 3

Consider the following function:

$$f_X(x) = \begin{cases} \alpha \exp(-2x) & : x > 0 \\ 0 & : x \leq 0 \end{cases}$$

where α is unknown and $\exp(x)$, calculates the value of e to the power of x, where e is the base of the natural logarithm. Derive the only value of α such that the function can be a PDF (Hint: Chapter 1.2.16 in Aronow and Miller).

Solution:

In order to be a valid PDF, the function must satisfy two criteria:

1. $\int_{-\infty}^{\infty} f(x)dx = 1$
2. $\forall x \in \mathbb{R}, f(x) \geq 0$

Let's start with the first criteria, by finding the definite integral for $x > 0$:

$$\begin{aligned} \int \alpha \exp(-2x)dx &= -\frac{\alpha}{2} \int \exp(u)du \quad \text{Let } u = -2x; du = -2dx. \\ &= -\frac{\alpha}{2} \exp(u) + C \\ &= -\frac{\alpha}{2} \exp(-2x) + C \end{aligned}$$

Check your work (dropping the constant of integration):

$$\frac{d}{dx}\left[-\frac{\alpha}{2}\exp(-2x)\right] = -\frac{\alpha}{2}\frac{d}{dx}\left[\exp(-2x)\right] = \alpha \exp(-2x)$$

Now let's break apart the function and solve for α (ingoring constant of integration):

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx \\ &= 0 + -\frac{\alpha}{2}\exp(-2x)|_0^\infty \\ &= \left[-\frac{\alpha}{2}\exp(-\infty)\right] - \left[-\frac{\alpha}{2}\exp(0)\right] \\ &= \frac{\alpha}{2} \\ \implies \int_{-\infty}^{\infty} f(x)dx &= 1 \text{ if } \alpha = 2 \end{aligned}$$

Let's use the `integrate()` function in `R` to check our answer:

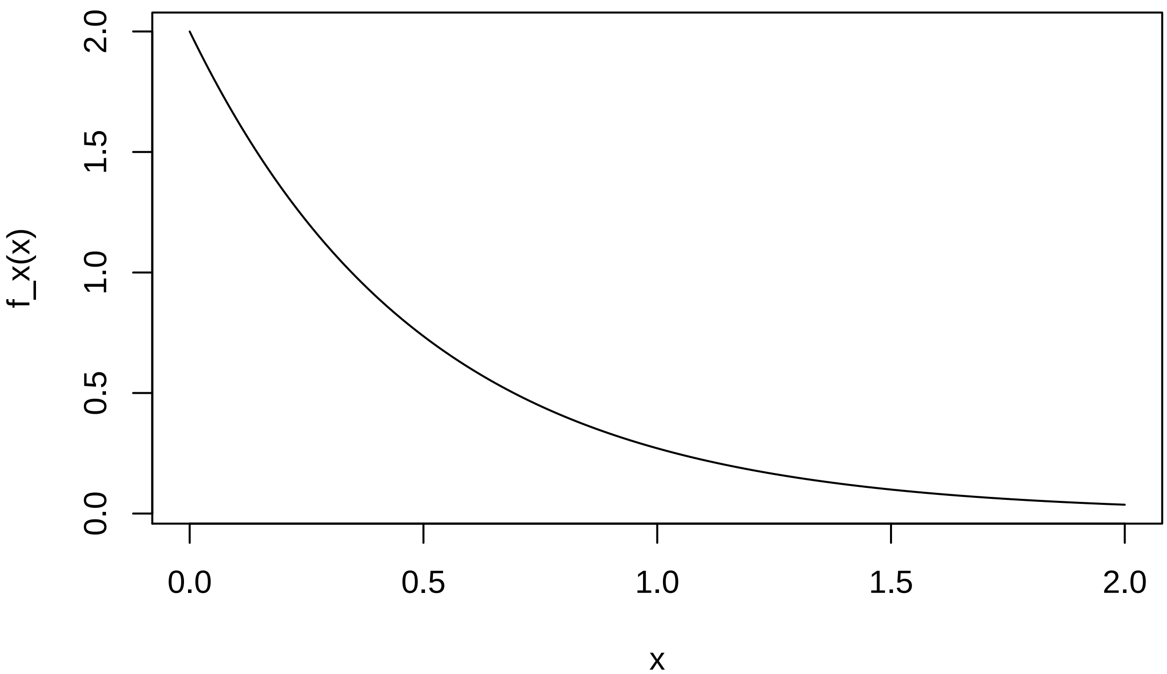
```
# Define the function for x > 0
f_x <- function(x){2*exp(-2*x)}

# Integrate over x > 0
integrate(f_x, lower = 0, upper = Inf)
```

```
## 1 with absolute error < 5e-07
```

The exponential function is strictly positive for all $x \in \mathbb{R}$, therefore, $c\exp(x)$ is strictly positive for $c > 0$. Then $f(x) = 2\exp(-2x)$ is strictly positive for $x \in \mathbb{R}$. We can also use `R` to plot the function to see what it looks like:

```
# Sequence from 0 to 2 by 0.001 increments
x <- seq(0, 2, by = 0.001)
# Plot the function
plot(x, f_x(x), type = "l")
```



Exercise 4

A fair four-sided die is rolled, and then a biased coin with probability p of heads is flipped as many times as the die roll says, e.g., if the result of the die roll is a 3, then the coin is flipped 3 times. Let X be the result of the die roll and Y be the number of times the coin comes up heads.

1. Find the joint PMF of X and Y.
2. Find the marginal PMF of X.
3. Find the marginal PMF of Y.
4. Find the conditional PMFs of Y given $X = x, \forall x \in \{1, 2, 3, 4\}$.
5. Find the conditional PMFs of X given $Y = y, \forall y \in \{0, 1\}$.
6. Show that $X \not\perp\!\!\!\perp Y$, i.e. X and Y are not independent.

Let's start by noting that the die roll follows a uniform distribution, i.e. $X \sim U\{1, 2, 3, 4\}$,

$$f_X(x) = \begin{cases} 1/4 & : x \in \{1, 2, 3, 4\} \\ 0 & : o.w. \end{cases}$$

Then, note that Y conditional on X is distributed binomial, i.e. $Y|X \sim \text{Bin}(X, p)$,

$$f_{Y|X}(y|x) = \binom{x}{y} p^y (1-p)^{(x-y)} \quad \text{for } y \leq x$$

Then the joint PMF for X and Y is,

$$\begin{aligned} f_{X,Y}(x, y) &= f_{Y|X}(y|x) f_X(x) \\ &= \frac{1}{4} \binom{x}{y} p^y (1-p)^{(x-y)} \end{aligned}$$

It's useful to write this as a table,

Joint PMF Table

x\y	0	1	2	3	4
1	$\frac{(1-p)}{4}$	$\frac{p}{4}$	-	-	-
2	$\frac{(1-p)^2}{4}$	$\frac{p(1-p)}{2}$	$\frac{p^2}{4}$	-	-
3	$\frac{(1-p)^3}{4}$	$\frac{3p(1-p)^2}{4}$	$\frac{3p^2(1-p)}{4}$	$\frac{p^3}{4}$	-
4	$\frac{(1-p)^4}{4}$	$p(1-p)^3$	$\frac{3p^2(1-p)^2}{2}$	$p^3(1-p)$	$\frac{p^4}{4}$

Note that $f(x = 1, y = 2)$ is undefined. Why? The event, "observing 2 heads from 1 coin flip" is undefined. Now, it's easy to calculate the marginal of Y by simply summing along the columns in the table above:

$$f_Y(y) = \begin{cases} \frac{1}{4}\left[(1-p) + (1-p)^2 + (1-p)^3 + (1-p)^4\right] & : y = 0 \\ \frac{1}{4}\left[p + 2p(1-p) + 3p(1-p)^2 + 4p(1-p)^3\right] & : y = 1 \\ \frac{1}{4}\left[p^2 + 3p^2(1-p) + 6p2(1-p)^2\right] & : y = 2 \\ \frac{p^3}{4} + p^3(1-p) & : y = 3 \\ \frac{p^4}{4} & : y = 4 \\ 0 & : o.w. \end{cases}$$

This can also be expressed as,

$$f_Y(y) = \frac{1}{4}\sum_{k=1}^{4}\binom{k}{y}p^y(1-p)^{k-y} \quad \text{for} \quad y \in \{0,1,2,3,4\}$$

Next we can use the joint PMF table, and the fact that $f(x|y) = \frac{f(x,y)}{f(y)}$, to derive the distribution of X conditional on Y, for $y \in \{0,1\}$,

$$f_{X|Y}(x|y=0) = \begin{cases} (1-p)\left[\sum_{k=1}^{4}(1-p)^k\right]^{-1} & : x = 1 \\ (1-p)^2\left[\sum_{k=1}^{4}(1-p)^k\right]^{-1} & : x = 2 \\ (1-p)^3\left[\sum_{k=1}^{4}(1-p)^k\right]^{-1} & : x = 3 \\ (1-p)^4\left[\sum_{k=1}^{4}(1-p)^k\right]^{-1} & : x = 4 \\ 0 & : o.w. \end{cases}$$

$$f_{X|Y}(x|y=1) = \begin{cases} p\left[p+2p(1-p)+3p(1-p)^2+4p(1-p)^3\right]^{-1} & : x = 1 \\ 2p(1-p)\left[p+2p(1-p)+3p(1-p)^2+4p(1-p)^3\right]^{-1} & : x = 2 \\ 3p(1-p)^2\left[p+2p(1-p)+3p(1-p)^2+4p(1-p)^3\right]^{-1} & : x = 3 \\ 4p(1-p)^3\left[p+2p(1-p)+3p(1-p)^2+4p(1-p)^3\right]^{-1} & : x = 4 \\ 0 & : o.w. \end{cases}$$

Finally, let's show $X \not\perp\!\!\!\perp Y$ using a proof by contradiction:

- $X \perp\!\!\!\perp Y \implies f_X(x)f_Y(y) = f_{X,Y}(x,y)$.
- Let $x = 1, y = 0$. Then, $f_X(x) = 1/4$, $f_Y(y) = \frac{1}{4}\left[\sum_{k=1}^{4}(1-p)^k\right]$, and $f_{X,Y}(x,y) = \frac{1}{4}(1-p)$.
- However, this yields a contradiction since $\frac{1}{16}\left[\sum_{k=1}^{4}(1-p)^k\right] \neq \frac{1}{4}(1-p)$.
- Therefore $X \not\perp\!\!\!\perp Y$.

Note that we do not need to check for all pairs (x,y) in the support of $f_{X,Y}(x,y)$.

We can also write the joint PMF as a function in R,

```
f_xy <- function(x, y, p = 1/2){
  if(x == 1 & y == 0){
    out <- (1-p)/4
  } else if(x == 2 & y == 0){
    out <- ((1-p)^2)/4
  } else if(x == 3 & y == 0){
    out <- ((1-p)^3)/4
  } else if(x == 4 & y == 0){
    out <- ((1-p)^4)/4
  } else if(x == 1 & y == 1){
    out <- p/4
  } else if(x == 2 & y == 1){
    out <- (p*(1-p))/2
  } else if(x == 3 & y == 1){
    out <- (3*p*(1-p)^2)/4
  } else if (x == 4 &  y == 1){
    out <- p*(1-p)^3
  } else if (x == 2 & y == 2){
    out <- p^2/4
  } else if (x == 3 & y == 2){
    out <- (3*p^2*(1-p))/4
  } else if (x == 4 & y == 2){
    out <- (3*p^2*(1-p)^2)/2
  } else if (x == 3 &  y == 3){
    out <- p^3/4
  } else if (x == 4 & y == 3) {
    out <- p^3*(1-p)
  } else if (x == 4 & y == 4){
    out <- p^4/4
  }
  return(out)
}

f_xy(x = 3, y = 0, p = 0.25)
```

```
## [1] 0.1054688
```

Suppose we want to calculate probabilities for $x \in \{1, 2, 3, 4\}$, fixing $y = 0$. We could write a for-loop:

```
for(i in 1:4){
  print(f_xy(x = i, y = 0, p = 0.25))
}
```

```
## [1] 0.1875
## [1] 0.140625
## [1] 0.1054688
## [1] 0.07910156
```

Unfortunately, the function is not vectorized so we cannot pass in a vector for x:

```
f_xy(x = 1:4, y =  0, p = 0.25)
```

```
## Warning in if (x == 1 & y == 0) {: the condition has length > 1 and only
## the first element will be used
```

```
## [1] 0.1875
```

Fortunately, it's easy to vectorize a function in R with the Vectorize() function:

```
f_xy2 <- Vectorize(f_xy)
f_xy2(x = 1:4, y = 0, p = 0.25)
```

```
## [1] 0.18750000 0.14062500 0.10546875 0.07910156
```

Then it's easy to veryify this is a valid PMF by summing up accross the support of (x, y) pairs,

```
# Calculate probabilities for all event pairs (x,y)
probs <- c(f_xy2(x = 1:4, y = 0, p = 0.25),
           f_xy2(x = 1:4, y = 1, p = 0.25),
           f_xy2(x = 2:4, y = 2, p = 0.25),
           f_xy2(x = 3:4, y = 3, p = 0.25),
           f_xy2(x = 4, y = 4, p = 0.25))
probs
```

```
##  [1] 0.1875000000 0.1406250000 0.1054687500 0.0791015625 0.0625000000
##  [6] 0.0937500000 0.1054687500 0.1054687500 0.0156250000 0.0351562500
## [11] 0.0527343750 0.0039062500 0.0117187500 0.0009765625
```

```
# Confirm these sum to 1
sum(probs)
```

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