Ben Juarez
UID: 2134561

Problem Set 3

1 Problem 1

We want to find the conditional expectation $\mathbb{E}[X|Y=y]$. So, let us first find the marginal PDF of Y from the given joint PDF as follows:

$$f_Y(y) = \int_{-y}^{y} f(x,y)dx = \int_{-y}^{y} \frac{y^2-x^2}{8} e^{-y} dx = \frac{e^{-y}}{8} \int_{-y}^{y} y^2 - x^2 dx$$

$$\implies f_Y(y) = \frac{e^{-y}}{8}(y^2 x - \frac{x^3}{3})|_{-y}^{y} = \frac{y^3 e^{-y}}{6}$$

Let us now find the conditional PDF of X, given that $Y=y$ and knowing that $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$:

$$f_{X|Y}(x|y) = \frac{\frac{y^2-x^2}{8}e^{-y}}{\frac{y^3 e^{-y}}{6}} = \frac{y^2-x^2}{8} \cdot \frac{6}{y^3} = \frac{3(x^2-y^2)}{4y^3}$$

Let us now find the conditional expectation knowing that $\mathbb{E}[X|Y=y] = \int_{-y}^{y} x f_{X|Y}(x|y)dx$:

$$\mathbb{E}[X|Y=y] = \int_{-y}^{y} x \cdot \frac{3(x^2-y^2)}{4y^3} dx = 0$$

Thus, we have $\mathbb{E}[X|Y=y] = 0$.

2 Problem 2

2.1 Part a

We have the level of exposure $x \in [0,1]$ to an influenze virus that is contracted with probability $q(x) = x^\gamma$. Exposure level of randomly selected student is modeled by $X \sim Beta(\alpha, \beta)$ with density $f_X(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$.

We want to find the conditional density of the exposure level of that student given that the student has influenza, in other words, let us find $f(x|I)$ such that I implies a case of influenza. We will utilize Baye's formula as follows:

$$f(x|I) = \mathbb{P}(X=x|I) = \frac{\mathbb{P}(I|X=x)\cdot\mathbb{P}(X=x)}{\mathbb{P}(I)} = \frac{q(x)\cdot f_X(x)}{\mathbb{E}[\mathbb{E}[q|X=x]]} = \frac{q(x)\cdot f_X(x)}{\mathbb{E}[\mathbb{P}(I|X=x)]}$$

$$= \frac{q(x)\cdot f_X(x)}{\int_0^1 \mathbb{P}(I|X=x)\cdot f_X(x)dx} = \frac{q(x)\cdot f_X(x)}{\int_0^1 q(x)\cdot f_X(x)dx} \propto \frac{x^\gamma \cdot x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^\gamma \cdot x^{\alpha-1}(1-x)^{\beta-1}dx}$$

$$\propto \frac{x^{(\alpha+\gamma)-1}(1-x)^{\beta-1}}{\int_0^1 x^{(\alpha+\gamma)-1}(1-x)^{\beta-1}dx} \propto \frac{x^{(\alpha+\gamma)-1}(1-x)^{\beta-1}}{Beta(\alpha+\gamma,\beta)}$$

Clearly, this is beta distribution $Beta(\alpha+\gamma, \beta)$.

2.2 Part b

Let I denote the case that the student has influenza. Since we ended up with $Beta(\alpha+\gamma, \beta)$, we can directly calculate the expected value as follows:

$$\mathbb{E}[X|I] = \frac{\alpha+\gamma}{\alpha+\gamma+\beta}$$

2.3 Part c

See corresponding MATLAB section.

3 Problem 3

We have a box containing n shoelaces such that we randomly choose two free ends and tie them up, repeating until there are no free ends. We want to find the number of expected total loops, starting with $n = 100$ shoelaces. *My initial guess for the expected number of loops with 100 shoelaces is 10.*

So, with n shoelaces, we know that before the first iteration of choosing two free ends, there are $2 \times n$ possible ends to select from the box. Let L_i denote the number of loops are the end of the ith iteration. With this, within the 1st iteration, let us imagine the situation just before selection of the second free end. At this point, there is only 1 free end that would create a loop and $2n-1$ free ends that would not create a loop. So, the probability of creating a loop after the 1st iteration is $\frac{1}{2n-1}$:

$$\mathbb{E}[L_i] = \frac{1}{2n-1}(1) + \frac{2n-2}{2n-1}(0)$$

For the second iteration, we essentially have $n-1$ shoelaces in the box because we either created a loop with a single shoelace or we tied two separate shoelaces together (which can still be tied into a larger loop). This distinction that a loop is not limited to just 1 shoelace is important, because it tells that for any iteration i in this process, there will be $n-i$ laces or series of laces with available free ends. To further clarify this, let us show the first 3 iterations with n shoelaces:

$$\mathbb{E}[L_{i=1}] = \frac{1}{2n-1}$$
$$\mathbb{E}[L_{i=2}] = \frac{1}{2n-3}$$
$$\mathbb{E}[L_{i=3}] = \frac{1}{2n-5}$$

$$\implies \mathbb{E}[L_i] = \frac{1}{2n-2i+1}$$

Also, this process repeats until no free ends are left, and considering that we start with $2n$ free ends and reduce the number of free ends by 2 each iteration, we know that this process repeats n times. So, we know that the total expected number is

$$\mathbb{E}[L] = \mathbb{E}[\sum_{i=1}^{n} L_i] = \sum_{i=1}^{n} \mathbb{E}[L_i] = \sum_{i=1}^{n} \frac{1}{2n-2i+1}$$

due to linearity of expectation. However, we can see that we can rewrite this sum in a more efficient (and equivalent) way as follows:

$$\implies \mathbb{E}[L] = \sum_{i=1}^{n} \frac{1}{2i-1}$$

Also, for $n = 100$, we have $\mathbb{E}[L] = \mathbb{E}[\sum_{i=1}^{100} L_i] = 3.3$. So, the expected number of loops at the end of this process with 100 shoelaces is about 3 loops, which is significantly less than my initial estimate. I underestimated how rare it would be to create a loop throughout this whole process.

4 Problem 4

4.1 Part a

Let us suppose that for independent trials, each of which is equally likely to have any of m possible outcomes are performed until the same outcome occurs k consecutive times. With N_k denoted as the number of trials, let us show that $\mathbb{E}[N_k] = 1 + m + ... + m^{k-1} = \sum_{i=1}^{k} m^{i-1}$.

Thinking inductively and using the law of total expectation, we arrive at the following under the assumption that the claim is true for $\mathbb{E}[N_{k-1}]$:

$$\mathbb{E}[N_k] = \mathbb{E}[\mathbb{E}[N_k|N_{k-1}]]$$

$$\implies \mathbb{E}[N_k] = \mathbb{E}[N_{k-1}] + \frac{1}{m} + (1 - \frac{1}{m})\mathbb{E}[N_k] = \mathbb{E}[N_{k-1}] + \frac{1}{m} + \mathbb{E}[N_k] - \frac{1}{m}\mathbb{E}[N_k]$$

$$\implies \frac{1}{m}\mathbb{E}[N_k] = \frac{1}{m} + \mathbb{E}[N_{k-1}] \implies \mathbb{E}[N_k] = 1 + m \cdot \mathbb{E}[N_{k-1}]$$

$$\implies \mathbb{E}[N_k] = 1 + m \cdot (1 + m + ... + m^{k-2}) = 1 + m + ... + m^{k-1} = \sum_{i=1}^{k} m^{i-1} \checkmark$$

We can see that this makes sense because we logically if we have a sequence of $k-1$, we get the same number again with probability $p = \frac{1}{m}$, otherwise the sequence needs to restart and get back up to k (with probability $1-p$). Thus, by law of total expectation, we have inductively shown that the claim holds for $\mathbb{E}[N_k]$.

4.2 Part b

Let us asses whether the evidence is weak or not using $\mathbb{V}[N_k] \approx \mathbb{E}[N_k]^2$ with $m = 10$. Let us evaluate $\mathbb{E}[N_9]$ and the standard deviation σ_{N_9} as follows:

$$\mathbb{E}[N_9] = 1+10+...+10^{9-1} = \sum_{i=1}^{9} m^{i-1} = 111,111,111 \qquad \sigma_{N_9} = \sqrt{\mathbb{V}[N_9]} \approx \sqrt{\mathbb{E}[N_9]^2} = \mathbb{E}[N_9]$$

$$\implies \mathbb{E}[N_9] - n_9 = 111,111,111 - 24,658,609 = 85,452,502 < \sigma_{N_9}$$

Thus, we have that the evidence is weak.

5 Problem 5

We want to find the expected value of X and its variance. Let us use the law of total expectation (treating Q as special case of continuous beta distribution) and the law of total variance as follows:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Q = q]] = \int_0^1 \mathbb{E}[X|Q = q] f_Q(q) dq$$

Since we are given $X|Q = q \sim Bin(n, q)$, we know $\mathbb{E}[X|Q = q] = nq$. Also, $f_Q(q) = 1$ for the given $Q \sim U[0, 1]$.

$$\implies \int_0^1 nqdq = n \int_0^1 qdq = \frac{n}{2}$$

So, we have $\mathbb{E}[X] = \frac{n}{2}$. Let us now evaluate $\mathbb{V}[X]$ as follows:

$$\mathbb{V}[X] = \mathbb{E}[\mathbb{V}[X|Q]] + \mathbb{V}[\mathbb{E}[X|Q]]$$

For $Bin(n, q)$, we know the expectation and variance as nq and $nq(1 - q)$, respectively:

$$\implies \mathbb{V}[X] = \mathbb{E}[nq(1 - q)] + \mathbb{V}[nq] = \int_0^1 nq(1 - q) f_Q(q) dq + (n^2)\mathbb{V}[q]$$

$$\implies \mathbb{V}[X] = n \cdot \int_0^1 q - q^2 dq + (n^2)(\frac{1}{12}) = \frac{n}{2} - \frac{n}{3} + \frac{n^2}{12} = \frac{n}{6} + \frac{n^2}{12}$$

Thus, we have $\mathbb{E}[X] = \frac{n}{2}$ and $\mathbb{V}[X] = \frac{n}{6} + \frac{n^2}{12}$.

6 MATLAB

```
% Ben Juarez    PS3Q2c
clear;
alpha = 2;
beta = 6;
gamma = 2;
n = 10^4;

x = betarnd(alpha, beta, 1, n);
infected = 0;
exp = 0;

for i = 1:n
    q = x(i)^gamma;
    if q > rand() % if disease is contracted
        exp = exp + x(i); % tracks exposure levels
        infected = infected + 1; % tracks number of infected
    end
end
avg_x = exp/infected % calculates sample average of exposure level of those
 infected
E = (alpha + gamma) / (alpha + gamma + beta) % calculates result from (b)
%
% Result confirmed!
%
```

```
avg_x =

    0.3933


E =

    0.4000
```

Published with MATLAB® R2021b
