Jaideep Singh Chawla
190110030
jdchawla@iitb.ac.in

Assignment 2
DS203: Programming for Data Sciences

August 22, 2021

Exercise 1

$X \sim Exp(\lambda_1),\ Y \sim Exp(\lambda_2)$

$$P(min(X,Y) > z) = P(X > z, Y > z) = P(X > z)P(Y > z) = e^{-z\lambda_1}e^{-z\lambda_2} = e^{-z(\lambda_1+\lambda_2)}$$

(Note that X and Y are independent random variables.)

$min(X,Y)$ is also exponentially distributed with parameter $\lambda = \lambda_1 + \lambda_2$

Let $Z = max(X,Y),\ Z \geq 0$

$$F_Z(z) = P(max(X,Y) \leq z) = P(X \leq z, Y \leq z) = (1 - e^{-z\lambda_1})(1 - e^{-z\lambda_2})$$

$$\begin{aligned} f_z(z) = \frac{dF_Z(z)}{dz} &= \lambda_2 e^{-z\lambda_2}(1 - e^{-z\lambda_1}) + \lambda_1 e^{-z\lambda_1}(1 - e^{-z\lambda_2}) \\ &= \lambda_1 e^{-z\lambda_1} + \lambda_2 e^{-z\lambda_2} - (\lambda_1 + \lambda_2)e^{-z(\lambda_1+\lambda_2)} \end{aligned}$$

Exercise 2

Let $P(X = x|Y = 3) = P(Z = x)$. Then, $Z \sim Bin\left(3, \frac{1}{3}\right)$

Explanation. For the remaining 3 non-blue selections, we can choose either a white ball or a red ball and the probability of getting a white ball is half the probability of red ball.

Thus, $E[Z] = E[X|Y = 3] = 3.\frac{1}{3} = 1$

(if $A \sim Bin(n,p)$, then, $E[A] = np$.)

Exercise 3

X_1 and X_2 are independent binomial random variables with respective parameters (n_1, p) and (n_2, p).

$$\begin{aligned} P(X_1 = k|X_1 + X_2 = m) &= \frac{P(X_1 = k, X_2 = m-k)}{P(X_1 + X_2 = m)} \\ &= \frac{P(X_1 = k)P(X_2 = m-k)}{P(X_1 + X_2 = m)} \\ &= \frac{\binom{n_1}{k}p^k(1-p)^{n_1-k}\binom{n_2}{m-k}p^{m-k}(1-p)^{n_2-m+k}}{\binom{n_1+n_2}{m}p^m(1-p)^{n_1+n_2-m}} \qquad [X_1 + X_2 \sim Bin(n_1+n_2, p)] \\ &= \frac{\binom{n_1}{k}\binom{n_2}{m-k}p^m(1-p)^{n_1+n_2-m}}{\binom{n_1+n_2}{m}p^m(1-p)^{n_1+n_2-m}} \\ &= \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}} \end{aligned}$$

Exercise 4

Let $X \sim Unif(-1,1)$ and $Y = X^2$; Clearly *dependent*. Also, $E[X] = 0$.

Correlation. $Cov(X,Y) = E[XY] - E[X]E[Y] = E[X^3] = \int_{-1}^{1} 0.5x^3 = 0$ (*uncorrelated*)

Exercise 5

$X \sim Poi(\lambda),\ \lambda \sim Exp(1)$ (The mean of an exponential random variable is the inverse of its parameter.)

$$\begin{aligned}
P(X=n) = E[P(X=n|\lambda)] &= \int_0^\infty e^{-\lambda}\frac{\lambda^n}{n!}f(\lambda)d\lambda \\
&= \int_0^\infty e^{-\lambda}\frac{\lambda^n}{n!}e^{-\lambda}d\lambda \\
&= \int_0^\infty e^{-2\lambda}\frac{\lambda^n}{n!}d\lambda \\
&= \frac{1}{2^{n+1}}\int_0^\infty e^{-x}\frac{x^n}{n!}dx \quad \text{[Change of Variable: } x = 2\lambda\text{]} \\
&= \frac{1}{2^{n+1}}\frac{\Gamma(n+1)}{n!} \\
&= \frac{1}{2^{n+1}} \quad \text{since, } n \text{ is a natural number.}
\end{aligned}$$

Exercise 6

$$f_{X,Y}(x,y) = \begin{cases} c(1+xy) & 2 \le x \le 3 \\ 0 & otherwise \end{cases}$$

1. For valid density function, $\iint_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy = 1$

$$\begin{aligned}
&\implies \int_1^2\int_2^3 c(1+xy)\,dx\,dy = 1 \\
&\implies c\int_1^2 x + y\frac{x^2}{2}\Big|_2^3\,dy = 1 \\
&\implies c\int_1^2 1 + \frac{5y}{2}\,dy = 1 \\
&\implies c\Big[y + \frac{5y^2}{4}\Big]_1^2 = 1 \\
&\implies \frac{19c}{4} = 1 \\
&\implies c = \frac{4}{19}
\end{aligned}$$

2.

$$\begin{aligned}
f_X(x) = \int f_{XY}(x,y)dy &= \int_1^2 \frac{4}{19}(1+xy)\,dy \\
&= \frac{4}{19}\Big[y + \frac{xy^2}{2}\Big]_1^2 \\
&= \frac{4}{19}\Big(1 + \frac{3}{2}x\Big)
\end{aligned}$$

$$\begin{aligned}
f_X(x) = \int f_{XY}(x,y)dx &= \int_2^3 \frac{4}{19}(1+xy)\,dx \\
&= \frac{4}{19}\Big[x + \frac{yx^2}{2}\Big]_2^3 \\
&= \frac{4}{19}\Big(1 + \frac{5}{2}x\Big)
\end{aligned}$$

Exercise 7

Let X = no. o f accidents a policyholder have in a year. *Given.* $X \sim Pois(\lambda),\ g(\lambda) = \lambda e^{-\lambda}$

$$\begin{aligned}
P(X=n) = E[P(X=n|\lambda)] &= \int_0^\infty e^{-\lambda}\frac{\lambda^n}{n!}g(\lambda)d\lambda \\
&= \int_0^\infty e^{-\lambda}\frac{\lambda^n}{n!}\lambda e^{-\lambda}d\lambda \\
&= \int_0^\infty e^{-2\lambda}\frac{\lambda^{n+1}}{n!}d\lambda \\
&= \frac{1}{2^{n+2}}\int_0^\infty e^{-x}\frac{x^{n+1}}{n!}dx \quad \text{[Change of Variable: } x = 2\lambda\text{]} \\
&= \frac{1}{2^{n+2}}\frac{\Gamma(n+2)}{n!} \\
&= \frac{n+1}{2^{n+2}} \quad \text{since, } n \text{ is a natural number.}
\end{aligned}$$

Exercise 8

Let X ad Y denotes the number of females and males respectively who visit the yoga studio. *Given.* $X + Y \sim Pois(\lambda)$

$$\begin{aligned}
P(X=n, Y=m) &= P(X=n, Y=m|X+Y=m+n).P(X+Y=m+n) \\
&= \binom{m+n}{n}p^n(1-p)^m e^{-\lambda}\frac{\lambda^{m+n}}{(m+n)!}
\end{aligned}$$

(Each person who visits is, independently, female with probability p or male with probability $(1-p)$ thus, X or Y follow binomial distribution for a fixed number of visitors)

Exercise 9

We will make use of the linearity properties of the $E[X]$ function i.e.,

$$E[X+Y] = E[X] + E[Y]$$
$$E[kX] = kE[X]$$

$$\begin{aligned}
Cov(aX_1+b, cX_2+b) &= E[(aX_1+b)(cX_2+b)] - E[aX_1+b]E[cX_2+b] \\
&= E[acX_1X_2 + abX_1 + bcX_2 + b^2] - (aE[X_1]+b)(cE[X_2]+b) \\
&= acE[X_1X_2] + abE[X_1] + bcE[X_2] + b^2 - acE[X_1]E[X_2] \\
&\quad - abE[X_1] - bcE[X_2] - b^2 \\
&= ac(E[X_1X_2] - E[X_1]E[X_2]) \\
&= acCov(X_1, X_2)
\end{aligned}$$

$$\begin{aligned}
Cov(X_1+X_2, X_3) &= E[(X_1+X_2)X_3] - E[X_1+X_2]E[X_3] \\
&= E[X_1X_3 + X_2X_3] - (E[X_1]+E[X_2])E_(X_3) \\
&= E[X_1X_3] + E[X_2X_3] - E[X_1]E[X_3] - E[X_2]E[X_3] \\
&= E[X_1X_3] - E[X_1]E[X_3] + E[X_2X_3] - E[X_2]E[X_3] \\
&= Cov(X_1, X_3) + Cov(X_2 + X_3)
\end{aligned}$$

Exercise 10

Given. $n = 100$, $\delta = 0.95$, $\hat{\mu} = 0.45$ Let the true mean by μ.

$$\begin{aligned} P(|\hat{\mu} - \mu| > \epsilon) &\leq 2e^{n\epsilon^2} = 0.05 \\ &\implies 2e^{100\epsilon^2} = 0.05 \\ &\implies \epsilon = 0.192 \end{aligned}$$

So, our confidence inteval is $(0.45 - 0.192, 0.45 + 0.192) = (0.258, 0.642)$

$$\epsilon_{new} = \epsilon/2 = 0.096$$

Now,

$$\begin{aligned} n_{new} &= \frac{1}{\epsilon_{new}^2} log\Big(\frac{2}{0.05}\Big) \\ &= \frac{1}{(0.096)^2} log\Big(\frac{2}{0.05}\Big) \\ &= 400 \end{aligned}$$

So, we need 300 more sample points.