DS 203

Assignment2

Gyandev Satyaram Gupta, 190100051

1. Let X and Y be independent exponential random variables with respective parameters λ_1 and λ_2. Find the distribution of the following.
   • min(X, Y)
   • max(X, Y)

Solution:

$$f_X(x) = \begin{cases} \lambda_1 e^{-\lambda_1 x} & \text{if } 0 \leq x \\ 0 & \text{if } otherwise \end{cases}$$

$$f_Y(y) = \begin{cases} \lambda_2 e^{-\lambda_2 y} & \text{if } 0 \leq y \\ 0 & \text{if } otherwise \end{cases}$$

$$F_X(x) = \int_0^x f_X(x)dx = \int_0^x \lambda_1 e^{-\lambda_1 x} dx = 1 - e^{-\lambda_1 x}$$

$$F_Y(y) = \int_0^y f_Y(y)dy = \int_0^y \lambda_2 e^{-\lambda_2 y} dy = 1 - e^{-\lambda_2 y}$$

•Take Z = min(X,Y)

$$\begin{aligned} P(min(X,Y) > m) &= P(X > m, Y > m) \\ &= P(X > m)P(Y > m) \\ &= (1 - (1 - e^{-\lambda_1 m}))(1 - (1 - e^{-\lambda_2 m})) \\ &= e^{-(\lambda_1+\lambda_2)m} \end{aligned}$$

$\implies P(min(X,Y) \leq m) = 1 - e^{-(\lambda_1+\lambda_2)m}$
$\implies F_Z(z) = 1 - e^{-(\lambda_1+\lambda_2)z}$
$\implies f_Z(z) = \frac{dF}{dz} = e^{-(\lambda_1+\lambda_2)z}$

•Take Z = max(X,Y)

$$\begin{aligned} P(min(X,Y) \leq m) &= P(X \leq m, Y \leq m) \\ &= P(X \leq m)P(Y \leq m) \\ &= (1 - e^{-\lambda_1 m})(1 - e^{-\lambda_2 m}) \\ &= 1 - e^{-\lambda_1 m} - e^{-\lambda_2 m} + e^{-(\lambda_1+\lambda_2)m} \end{aligned}$$

$\implies F_Z(z) = 1 - e^{-\lambda_1 z} - e^{-\lambda_2 z} + e^{-(\lambda_1+\lambda_2)z}$
$\implies f_Z(z) = \frac{dF}{dz} = \lambda_1 e^{-\lambda_1 z} + \lambda_2 e^{-\lambda_2 z} - (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)z}$

2. A bag contains 3 white, 6 red and 5 blue balls. A ball is selected at random, it's color is noted and is then replaced in the bag before making the next selection. In all 6 selections are made. Let X = the number of white balls selected and Y = number of blue balls selected. Find E[X|Y = 3].

Solution: p = $\frac{3}{14}$ probability of white ball choosen
q = $\frac{6}{14}$ probability of red ball choosen
r = $\frac{5}{14}$ probability of blue ball choosen
Now so there are multiple possibilities of white ball like 0,1,2,3 ball to be drawn.
Suppose k white ball are drawn so remaining 3-k are red ball

$$\begin{aligned} P(X=k|Y=3) &= \frac{P(X=k \cap Y=3)}{P(Y=3)} \\ &= \frac{\binom{6}{3}\binom{3}{k}p^k q^{3-k} r^3}{\binom{6}{3} r^3 (1-r)^3} \\ &= \frac{\binom{3}{k} 3^k 6^{3-k} 14^3}{14^3 9^3} \\ &= \binom{3}{k}\frac{2^{3-k}}{27}, for(k=0,1,2,3) \end{aligned}$$

$$\begin{aligned} E(X|Y=3) &= \sum_0^3 xP(X=k|Y=3) \\ &= \sum_0^3 \binom{3}{k}\frac{2^{3-k}}{27} \\ &= \frac{(1+2)^3}{27} = 1 \end{aligned}$$

3. If X_1 and X_2 are independent binomial random variables with respective parameters (n_1, p) and (n_2, p). Calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$.

Solution: For Bernoulii $P(X=i) = \binom{n}{i}p^i(1-p)^{n-i}$, for $0 \leq i \leq n$
X_1 and X_2 are independent events

$$\begin{aligned} P(X_1=k|X_1+X_2=m) &= \frac{P(X_1=k \cap X_1+X_2=m)}{P(X_1+X_2=m)} \\ &= \frac{P(X_1=k)P(X_2=m-k)}{\binom{n_1+n_2}{m} * p^m * (1-p)^{n_1+n_2-m}} \\ &= \frac{\binom{n_1}{k}p^k(1-p)^{n_1-k}\binom{n_2}{m-k}p^{m-k}(1-p)^{n_1-m+k}}{\binom{n_1+n_2}{m} * p^m * (1-p)^{n_1+n_2-m}} \\ &= \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}} \end{aligned}$$

k $\leq n_1$ and also m $\leq n_1+n_2$
for $k \in \{1,2,3,...,m\}$ if m $< n_1$
for $k \in \{1,2,3,...,n_1\}$ if m $\geq n_1$

4. Give an example of two random variables X and Y that are uncorrelated but not independent.

Solution: We know correlation of X,Y is related as

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

So for uncorrelation Corr(X,Y) = 0
Choose X = Unif(-1,1) and $Y = X^2$ to make it dependent

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if} - \leq x \leq 1 \\ 0 & \text{if } otherwise \end{cases}$$

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= E(XX^2) - E(X)E(X^2) \\ &= E(X^3) - E(X)E(X^2) \\ &= \int_{-1}^{1} x^3 f_X(x)dx - \int_{-1}^{1} x f_X(x)dx \int_{-1}^{1} x^2 f_X(x)dx \\ &= 0 - 0 = 0 \end{aligned}$$

Hence X and Y are uncorrelated.
They are dependent because for a particular X you have an unique and fixed Y

$$P(X = a|Y = b) = \begin{cases} 1 & \text{if } b = a^2 \\ 0 & \text{if } otherwise \end{cases}$$

5. Suppose X is a Poisson random variable with mean λ. The parameter λ is itself a random variable whose distribution is exponential with mean 1.
Show that $P\{X = n\} = \frac{1}{2}^{n+1}$.

Solution: X $\sim$ Poi(λ) and $\lambda \sim Exp(1)$

$$\begin{aligned} P(X = n) &= \int_0^\infty e^{-\lambda} \frac{e^{-\lambda}\lambda^n}{n!} d\lambda \\ &= \frac{\int_0^\infty e^{-2\lambda}\lambda^n d\lambda}{n!} \\ &= \frac{\frac{\Gamma n+1}{2^{n+1}}}{n!} = \frac{1}{2^{n+1}} \end{aligned}$$

6. Suppose X and Y have joint density function $f_{X,Y}(x,y) = c(1 + xy)$ if $2 \leq x \leq 3$ and 1 ≤ y ≤ 2, and $f_{X,Y}(x,y) = 0$ otherwise.
1. Find c.
2. Find f_X and f_Y .

Solution: We know that:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$
$$\int_{1}^{2}\int_{2}^{3} c(1+xy)dxdy = 1$$
$$\int_{1}^{2} c(1+5y/2)dy = 1$$
$$c(1+5*3/4) = 1$$
$$\implies c = \frac{4}{19}$$

$$\begin{aligned} f_X(x) &= \int_1^2 \frac{4}{19}(1+xy)dy \\ &= \frac{4}{19}(1+3x/2), \forall 2 \le x \le 3 \\ &= 0 \text{ if otherwise} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_2^3 \frac{4}{19}(1+xy)dx \\ &= \frac{4}{19}(1+5x/2), \forall 1 \le y \le 2 \\ &= 0 \text{ if otherwise} \end{aligned}$$

7. An insurance company supposes that the number of accidents that each of its policyholders will have in a year is Poisson distributed, with the mean of the Poisson depending on the policyholder. If the Poisson mean of a randomly choosen policyholder has a gamma distribution with density function,

$$g(\lambda) = \lambda e^{-\lambda}, \lambda \ge 0$$

what is the probability that a randomly chosen policyholder has exactly n accidents next year?

Solution: X = no. of accidents of a particular policy holder
X $\sim$ Poi(λ)

$$P_X(n) = \frac{\lambda^n e^{-\lambda}}{n!}$$

$$\begin{aligned} P(X=n) &= \int_0^\infty \lambda e^{-\lambda}\frac{e^{-\lambda}\lambda^n}{n!}d\lambda \\ &= \frac{\int_0^\infty e^{-2\lambda}\lambda^{n+1}d\lambda}{n!} \\ &= \frac{\frac{\Gamma n+2}{2^{n+2}}}{n!} = \frac{n+1}{2^{n+2}} \end{aligned}$$

8. Suppose that the number of people who visit a yoga studio each day is a Poisson random variable with mean . Suppose further that each person who visits is, independently, female with probability p or male with probability 1 - p. Find the joint probability that exactly n women and m men visit the academy today.

Solution: X = no. of male visiting yoga studio each day
Y = no. of female visiting yoga studio each day
X + Y $\sim$ Poi(λ)

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

9. Let X_1, X_2, X_3 are RVs and a, b, c, d are constants. Show that
• Cov(aX_1 + b, cX_2 + b) = acCov(X_1, X_2)
• Cov($X_1 + X_2, X_3$) = Cov(X_1, X_3) + Cov(X_2, X_3)

Solution: We know $E(aX+b) = aE(X) + b$

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

Hence proved
WE know that E(X+Y) = E(X) + E(Y)

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

Hence proved

10. You are given n = 100 i.i.d. samples generated from a random experiment. Let the estimate of mean from these samples is $\hat{\mu} = 0.45$. We know that true mean lies somewhere around $\hat{\mu}$ and we would like to find an interval (around $\hat{\mu}$) such that the true value lies in the interval with probability at least 0.95.
• What would be your (confidence) interval? Specify the method you used to come up with the interval.

• If you want the your confidence interval to shrink by half, how many more samples would you need? (the estimate could be different now)

Solution: Take μ to be the actual mean

$$P(|\mu - \hat{\mu}| \leq \epsilon) \geq 0.95$$

We know that $P(|\mu - \hat{\mu}| > \epsilon) \leq \delta$, where $\delta = 2e^{-n\epsilon^2}$
Consider the marginal case,

$$\begin{aligned}
P(|\mu - \hat{\mu}| > \epsilon) &= \delta \\
P(|\mu - \hat{\mu}| \leq \epsilon) &= 1 - \delta \\
\implies 1 - \delta &\geq 0.95 \\
\implies 0.05 &\geq= 2e^{-100\epsilon^2} \\
\implies \epsilon &\geq \sqrt{\frac{-ln(0.025)}{100}} \\
\implies \epsilon &\geq 0.1921
\end{aligned}$$

So confidence interval (CI) = $[0.45 - 0.1921, 0.45 + 0.1921] = [0.2579, 0.6421]$
To make the interval half $\epsilon' = \frac{\epsilon}{2}$ and $\delta' = \delta$

$$\begin{aligned}
\delta' &= \delta \\
2e^{-n'\epsilon'^2} &= 2e^{-n\epsilon^2} \\
n'\epsilon'^2 &= n\epsilon^2 \\
n' &= n\frac{\epsilon^2}{\epsilon'^2} \\
n' &= 100 * 4 = 400
\end{aligned}$$

So we need 300 more samples to shrink the CI by half