DS 203: Assignment 2

~ Shubham Lohiya, 18D100020

Ex-1 $X \sim \exp(\lambda_1)$ & $Y \sim \exp(\lambda_2)$

1) Let the r.v. $Z = \min(X,Y)$. Let's find the CDF of Z

$$P(Z \leq z) = P(\min(X,Y) \leq z) = 1 - P(\min(X,Y) > z)$$

We know that if $\min(a,b) > c$, then $a > c$ & $b > c$

Also for r.v. $X \sim \exp(\lambda)$, $P(X > x) = e^{-\lambda x}$

Thus, $P(Z \leq z) = 1 - P((X > z) \cap (Y > z)) = 1 - \underbrace{P(X>z)P(Y>z)}_{x \,\&\, y \text{ are independent}}$

$$= 1 - e^{-\lambda_1 z} \cdot e^{-\lambda_2 z}$$

$$= 1 - e^{-(\lambda_1+\lambda_2)z} \rightarrow \text{s.t. } z > 0 \,\&\, 0 \text{ o/w}$$

Since the p.d.f. $f_Z(z) = \frac{\partial}{\partial z}(P(Z \leq z))$, we have:

$$f_Z(z) = \frac{\partial}{\partial z}\left(1 - e^{-(\lambda_1+\lambda_2)z}\right) = \boxed{(\lambda_1+\lambda_2)e^{-(\lambda_1+\lambda_2)z}}$$

Thus, $Z \sim \exp(\lambda_1+\lambda_2)$

2) Let the r.v. $Z = \max(X,Y)$. Let's find the CDF of Z

$$P(Z \leq z) = P(\max(X,Y) \leq z) = P(X \leq z)P(Y \leq z)$$

Using same args. as before, & result $P(X \leq x) = 1 - e^{-\lambda x}$ for $X \sim \exp(\lambda)$

$$P(Z \leq z) = (1 - e^{-\lambda_1 z})(1 - e^{-\lambda_2 z}) = 1 - e^{-\lambda_1 z} - e^{-\lambda_2 z} + e^{-(\lambda_1+\lambda_2)z}$$

$$f_Z(z) = \frac{\partial}{\partial z}(P(Z \leq z)) = \boxed{\lambda_1 e^{\lambda_1 z} + \lambda_2 e^{\lambda_2 z} - (\lambda_1+\lambda_2)e^{-(\lambda_1+\lambda_2)z}}$$

$\hookrightarrow$ s.t. $z > 0$, 0 o/w

Ex-2 Given probabilities $p_W = \frac{3}{14}$, $p_R = \frac{6}{14}$ & $p_B = \frac{5}{14}$, & 6 draws.

Since it's given that $Y = 3$, other three draws could have only W or R with probabilities: $p'_W = \frac{3}{3+6} = \frac{1}{3}$ & $p'_R = \frac{6}{3+6} = \frac{2}{3}$. So for the remaining 3 draws, $P(X = i) \sim \text{Binomial}(3, \frac{1}{3})$

$$\Rightarrow E[X \mid Y=3] = E[\text{Binomial}(3, 1/3)] = \boxed{1} \quad \text{as } E[\text{bin}(n,p)] = np$$

Ex-3 Since $X_1 \sim bin(n_1, p)$ & $X_2 \sim bin(n_2, p)$ share same param p, $(X_1 + X_2) \sim (n_1 + n_2, p)$. We need to find $P(X_1 = i \mid X_1 + X_2 = m)$

If $X_1 = i$ & $X_1 + X_2 = m$, then $X_2 = m - i$

$$\Rightarrow P(X_1 = i \mid X_1 + X_2 = m) = \frac{P(X_1 = i)\, P(X_2 = m-i)}{P(X_1 + X_2 = m)}$$

$$= \frac{\binom{n_1}{i} p^i (1-p)^{n_1 - i} \cdot \binom{n_2}{m-i} p^{m-i} (1-p)^{n_2 - m + i}}{\binom{n_1+n_2}{m} p^m (1-p)^{n_1+n_2-m}} = \boxed{\frac{\binom{n_1}{i}\binom{n_2}{m-i}}{\binom{n_1+n_2}{m}}}$$

Ex-4 Consider $X \sim uniform(-1, 1) \Rightarrow f_X(x) = \begin{cases} 1/2 & x \in [-1, 1] \\ 0 & o/w \end{cases}$

Now define $Y = X^2$. Clearly Y & X are dependent

Also, $Cov(X, Y) = E[XY] - E[X]E[Y]$

$$= E[X^3] - E[x]E[x^2] = \int_{-1}^{1} \cancel{\frac{x^3}{2}\,dx} - \int_{-1}^{1} \cancel{\frac{x}{2}\,dx} \int_{-1}^{1} \frac{x^2}{2}\,dx$$

$$= \boxed{0}$$

(As x^3 & x are odd funcs)

Thus, X & Y are uncorrelated but not independent.

Ex-5 Given $X \sim Pois(\lambda)$ & $\lambda \sim exp(1)$

Thus $P(X = n \mid \lambda = x) = \dfrac{e^{-x} x^n}{n!}$

Since $P(X = n) = \int_0^\infty P(X = n \mid \lambda = x) f_\lambda(x)\,dx$

$$= \int_0^\infty \frac{e^{-2x} x^n}{n!}\,dx$$

Substitute $t = 2x$
$\Rightarrow dt = 2dx$

$$= \frac{1}{2^{n+1} \cdot n!} \int_0^\infty e^{-t} t^n\,dt$$

We know $\Gamma(z) = \int_0^\infty e^{-x} x^{z-1}\,dx$

$$= \frac{1}{2^{n+1}\, n!} \Gamma(n+1) = \frac{1}{2^{n+1} n!} \cdot n!$$

Since for +ve integer n, $\Gamma(n) = (n-1)!$

$$= \boxed{\frac{1}{2^{n+1}}}$$

Ex-6 Given $f_{X,Y}(x,y) = \begin{cases} c(1+xy), & (x,y) \in [2,3]\times[1\ 2] \\ 0, & \text{otherwise} \end{cases}$

1) We know that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$

$$\Rightarrow 1 = c\int_1^2\int_2^3 1+xy\ dx\ dy = c\int_1^2 \left[x + y\frac{x^2}{2}\right]_2^3 dy$$

$$= c\int_1^2 1 + \frac{5y}{2}\,dy = c\left[y + \frac{5y^2}{4}\right]_1^2$$

$$\Rightarrow 1 = c\left(1 + \frac{15}{4}\right) \Rightarrow \boxed{c = \frac{4}{19}}$$

2) $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \frac{4}{19}\int_1^2 (1+xy)\,dy$$

$$= \boxed{\frac{4}{19}\left(1 + \frac{3}{2}x\right)}$$

$$f_Y(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \frac{4}{19}\int_2^3 (1+xy)\,dy = \boxed{\frac{4}{19}\left(1+\frac{5y}{2}\right)}$$

Ex-7 Let a r.v. N denote no. of accidents per year

Given, $(N\,|\,\lambda) \sim \text{Pois}(\lambda)$ & $\lambda \sim g(\lambda) = \text{gamma}(\lambda)$, $\lambda \geq 0$

Thus, $$P(N=n) = \int_0^{\infty} P(N=n\,|\,\lambda)\,g(\lambda)\,d\lambda$$

$$\Rightarrow P(N=n) = \int_0^{\infty} \frac{e^{-\lambda}\lambda^n}{n!}\lambda e^{-\lambda}\,d\lambda = \int_0^{\infty} \frac{e^{-2\lambda}\lambda^{n+1}}{n!}\,d\lambda$$

Substitute $t = 2\lambda$
$\Rightarrow dt = 2d\lambda$

$$= \int_0^{\infty} \frac{e^{-t}\,t^{n+1}}{2^{n+2}\,n!}\,dt = \frac{\Gamma(n+2)}{2^{n+2}\,n!}$$

$$= \frac{(n+1)!}{2^{n+2}\,n!} = \boxed{\frac{n+1}{2^{n+2}}}$$

Ex-8 Let T be total no. of people visiting a yoga studio on a day

Also, let M be no. of men & N be no. of women s.t. $M+N=T$

Given, $T \sim \text{Poisson}(\lambda)$ & $N \sim \text{Binomial}(T, p)$

Thus
$$P(M=m, N=n) = P(T=m+n) \cdot P(N=n \mid T=m+n)$$
$$= \frac{e^{-\lambda}\lambda^{m+n}}{(m+n)!} \cdot \binom{m+n}{n} p^n (1-p)^m$$
$$\Rightarrow P(M=m, N=n) = \boxed{\frac{e^{-\lambda} \cdot \lambda^{m+n} \cdot p^n (1-p)^m}{m!\, n!}}$$

Ex-9

1) Show $\text{Cov}(aX_1+b, cX_2+b) = ac\,\text{Cov}(X_1, X_2)$

LHS: $\text{Cov}(aX_1+b, cX_2+b) = E[(aX_1+b) - E[aX_1+b]] \cdot E[cX_2+b - E[cX_2+b]]$

$$= E[a(X_1 - E[X_1])] \cdot E[c(X_2 - E[X_2]]$$
$$= ac\; E[X_1 - E[X_1]] \cdot E[X_2 - E[X_2]] = \boxed{ac\,\text{Cov}(X_1, X_2) = \text{RHS}}$$

2) Show $\text{Cov}(X_1+X_2, X_3) = \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_3]$

LHS: $\text{Cov}(X_1+X_2, X_3) = E[X_1+X_2 - E[X_1+X_2]] \cdot E[X_3 - E[X_3]]$

$$= E[(X_1 - E[X_1]) + (X_2 - E[X_2])] \cdot E[X_3 - E[X_3]]$$
$$= (E[X_1 - E[X_1]] + E[X_2 - E[X_2]]) \cdot E[X_3 - E[X_3]]$$
$$= E[X_1 - E[X_1]]\, E[X_3 - E[X_3]] + E[X_2 - E[X_2]] \cdot E[X_3 - E[X_3]]$$
$$= \boxed{\text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_3) = \text{RHS}}$$

Ex-10 Given $n = 100$ samples of iid. r.v. X & $\hat{\mu} = 0.45$

1) We have the result $P(|\hat{\mu}-\mu| > \varepsilon) \leq 2e^{-2n\varepsilon^2}$, $\varepsilon > 0$ Hoeffding's inequality

for confidence interval $\mu \in [\hat{\mu}-\varepsilon, \hat{\mu}+\varepsilon]$

for > 0.95 confidence, $P(|\hat{\mu}-\mu| > \varepsilon) \lesssim 0.05$

$\Rightarrow \quad 2e^{-2n\varepsilon^2} \leq 0.05 \Rightarrow 2n\varepsilon^2 \geq -\log(0.025)$

Thus, $\varepsilon^2 \geq \dfrac{-\log(0.025)}{200} = 0.0184$

$\Rightarrow \varepsilon \geq 0.136$. We have tightest interval with $\boxed{\varepsilon = 0.136}$

Thus, $\mu \in (0.314, 0.586)$ with 95% confidence

2) for interval of half length, $\varepsilon' = \frac{\varepsilon}{2} = 0.068$

Also $2n\varepsilon'^2 \geq -\log(0.025)$

$\Rightarrow \quad n \geq \dfrac{-\log(0.025)}{2\varepsilon'^2} = 4 \cdot \left(\dfrac{-\log(0.025)}{2\varepsilon^2}\right) = 400$

$\Rightarrow \quad \boxed{n \geq 400}$ Thus at least 400 samples are needed.

Since we already have 100, we need 300 more samples.

Source: Hoeffding's inequality for confidence intervals

1st part with looser bound from the slides

Ex-10 Given $n = 100$ samples of iid. r.v. X & $\hat{\mu} = 0.45$

1) We have the result $P(|\hat{\mu}-\mu| > \varepsilon) \leq 2e^{-n\varepsilon^2}$, $\varepsilon > 0$

for confidence interval $\mu \in [\hat{\mu}-\varepsilon, \hat{\mu}+\varepsilon]$

for > 0.95 confidence, $P(|\hat{\mu}-\mu| > \varepsilon) \lesssim 0.05$

$\Rightarrow \quad 2e^{-n\varepsilon^2} \leq 0.05 \Rightarrow n\varepsilon^2 \geq -\log(0.025)$

Thus, $\varepsilon^2 \geq \dfrac{-\log(0.025)}{100} = 0.0369$

$\Rightarrow \varepsilon \geq 0.192$. We have tightest interval with $\boxed{\varepsilon = 0.192}$

Thus, $\mu \in (0.258, 0.642)$ with 95% confidence