Question 1: (25 Points) Consider the following joint probability density function for random variables X and Y.

$$f(x,y) = 6x \qquad 0 \le x \le 1, \quad 0 \le y \le 1-x$$

a) Find the marginal pdf of X.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^{1-x} 6x\,dy$$

$$= 6xy \Big|_0^{1-x} = 6x - 6x^2$$

b) Calculate $E(XY)$.

$$E[XY] = \int_0^1 \int_0^{1-x} (x \cdot y \cdot 6x)\,dy\,dx$$

$$= \int_0^1 3x^2(1-x)^2\,dx$$

$$= x^3 - \frac{6}{4}x^4 + \frac{3}{5}x^5 \Big|_0^1$$

$$= \frac{1}{10}$$

c) (4 points) Find $\text{Cov}(X, Y)$, that is, the covariance of X and Y.
Hint: You can use the following information, $E(X) = 0.5$ and $E(Y) = 0.25$.

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

$$= \frac{1}{10} - \frac{1}{2} \cdot \frac{1}{4} = \frac{-1}{40}$$

d) Are X and Y independent?

$$f_{X,Y}(x,y) = 6x$$

$$f_X(x) = 6x - 6x^2$$

$$f_Y(y) = \int_0^{1-y} 6x\,dx = 3(1-y)^2$$

$$f_{X,Y}(x,y) \ne f_X(x) \cdot f_Y(y)$$

So they are not independent.

OR simply

$$\text{Cov}(X,Y) \ne 0$$

X and Y cannot be independent.

Question 2: (20 Points)

a) (15 points) Among the most famous of all meteor showers are the so-called Perseids, which occur each year in early August. In some areas, the frequency of visible Perseids can be as high as forty per hour. Assume that the number of Perseids observed in any given time interval has Poisson distribution. Calculate the probability that an observer sees at least 3 Perseids in 4 minutes, in early August.

$$\lambda = \frac{160}{60}$$

$$P(X \geqslant 3) = 1 - P(x=0) - P(x=1) - P(x=2)$$

$$= 1 - \sum_{k=0}^{2} \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

$$= 1 - 0.07 - 0.185 - 0.247$$

$$\approx 0.438$$

b) (5 points) Define $M = \max(X, Y)$ where $X \sim \text{Exp}(0.7)$ and $Y \sim \text{Exp}(1.2)$ are independent from each other. Derive the distribution of M.

Hint: For an exponential random variable with parameter λ, the cdf is given by $F(x) = 1 - e^{-\lambda x}$, $x > 0$.

$$F_M(m) = P(M \leq m) = P(\max(X, Y) \leq m)$$

$$= P(X \leq m, Y \leq m)$$

$$= P(X \leq m) \cdot P(Y \leq m)$$

$$= \left(1 - e^{(-0.7)m}\right)\left(1 - e^{(-1.2)m}\right)$$

$$f_M(m) = F_M'(m) = 0.7 \cdot e^{(-0.7)m} \cdot \left(1 - e^{(-1.2)m}\right)$$

$$+ \left(1 - e^{(-0.7)m}\right) \cdot (1.2) \cdot e^{(-1.2)m}$$

$$= (0.7) e^{(-0.7)m} + (1.2) \cdot e^{(-0.7)m} - (1.9) \cdot e^{(-1.9)m}$$

Question 3: (20 Points) Sir Francis Galton (1822-1911) is known for his groundbreaking work in using fingerprints for identification purposes, in addition to all his contributions in genetics. The number of ridge crossings, called the ridge count (simply the number of lines that appear in a finger print) in males can be approximated by a normal distribution with mean 146 and standard deviation 52.

a) (6 points) What is the probability that a randomly selected male has a ridge count between 70 and 170?

Let Y be ridge count

$P(70 \leq Y \leq 170) = P\left(\frac{63.5 - 146}{52} \leq \frac{Y - 146}{52} \leq \frac{170.5 - 146}{52} \right)$

$= P(-1.47 \leq Z \leq 0.47)$

$= 0.62$

b) If a sample of 50 males are to be selected randomly from the population, what is the approximate probability that their average ridge count will be larger than 130?

Let $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ $n = 50$

$P(\bar{Y} \geq 130) = P\left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} > \frac{130 - \mu}{\sigma/\sqrt{n}} \right)$

$= P\left(Z > \frac{130 - 146}{52/\sqrt{50}} \right)$

≈ 0.9852

c) Suppose a random sample of size 50 is actually drawn and it is found that the sample mean of ridge counts is 152. Construct a 95% confidence interval for the mean ridge count based on your sample. Does your interval include the population mean?

$\left(\bar{y} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \; \bar{y} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right)$

$= (137.53, 166.41)$

Yes, interval includes 146.

Question 4: (25 Points) a) (10 points) Derive the maximum likelihood estimator for the parameter of Poisson distribution. Is it unbiased? If not, suggest an unbiased estimator.

$$L(\lambda, x_1, \cdots x_n) = \prod_{j=1}^{\hat{n}} \frac{e^{-\lambda} \lambda^{x_j}}{x_j!}, \quad \ln L(\lambda, x_1, \cdots x_n) = -n\lambda -$$

$$\ln L(\lambda, x_1 \cdots x_n) = -n\lambda - \sum_{j=1}^{\hat{n}} \ln(x_j!) + \left(\sum_{j=1}^{\hat{n}} x_j\right) \cdot \ln \lambda$$

$$\frac{d \ln L(\lambda, x_1 \cdots x_n)}{d\lambda} = -n + \frac{1}{\lambda} \cdot \left(\sum_{i=1}^{\hat{n}} x_i\right) = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{\sum_{i=1}^{\hat{n}} x_j}{n}$$

$$E[\hat{\lambda}] = E[\bar{X}] = \frac{1}{n} \sum_{i=1}^{\hat{n}} E[X_i] = \lambda.$$

It is   unbiased.

b) (8 points) Derive the method of moments estimator for θ in the pdf
$$f(y; \theta) = (\theta^2 + \theta)y^{\theta-1}(1 - y), \quad 0 \le y \le 1.$$

$$E[Y] = \int_0^1 y \cdot f(y, \theta) \, dy = \int_0^1 (\theta^2 + \theta) \cdot y^{\theta}(1-y) \, dy$$

$$= (\theta^2 + \theta) \cdot \left(\frac{y^{\theta+1}}{\theta+1} - \frac{y^{\theta+2}}{\theta+2}\right)\Big|_0^1 = \frac{\theta^2 + \theta + \theta}{\theta+2} \cdot \frac{1}{\theta+2} = \frac{\theta}{\theta+2}$$

$$\Rightarrow \quad \frac{\theta_e}{\theta_e + 2} = \bar{y}$$

$$\Rightarrow \boxed{\theta_e = \frac{2\bar{y}}{1-\bar{y}}}$$

c) (7 points) What is the smallest sample size of a random sample from a Bernoulli distribution with parameter p, if the margin of error for estimating p should not exceed 3% at the 90% confidence level?

$$100(1-\alpha) = 90 \Rightarrow \alpha = 0.1$$

$$P(z \ge z_{\alpha/2}) = 0.05 \Rightarrow z_{\alpha/2} = 1.64$$

$$d = 3 \cdot 10^{-2}$$

$$n \ge \frac{z_{\alpha/2}^2}{4 \cdot d^2} = 747.11 \Rightarrow n = 748 \quad \text{is} \quad \text{smallest.}$$

Question 5: (20 Points) Let a random variable Y be equal to 1 if a chemical substance is radioactive, and 0 if not. On the other hand, its physical strength can be classified as weak, moderate, strong, and quantified with -1, 0, 1, respectively, and denoted by random variable X. Consider the joint probability distribution for radioactivity and strength of the substances found in a certain geographical region of the world.

	0	1
-2	0.20	0.25
0	0.10	0.30
2	0.10	0.05

a) (6 points) Let $W = XY$. Find the probability function of W.

$P_W(-2) = P(X=-2, Y=1) = 0.25$

$\cancel{P_W(0) = P(X=0) + P(Y=0) \sim P(X=0, Y=0) = 0.7}$

$P_W(+2) = P(X=2, Y=1) = 0.05$

b) (6 points) Find conditional probability $P(Y = 1 | X = 2)$.

$P(Y=1 | X=2) = \dfrac{P(X=2, Y=1)}{P(X=2)} = \dfrac{0.05}{0.15} = \dfrac{1}{3}$

c) (8 points) If 120 different substances are chosen at random from the subgroup of strong ones, what is the approximate probability that at most 45 of them are radioactive?

$A \sim Bern\left(\frac{1}{3}\right)$

$P\left(A_1 + \cdots \quad A_{120} \le 45\right) = P\left(\dfrac{A_1 + \cdots + A_{120} - n \cdot p}{\sqrt{n \cdot p \,(1-p)}} \le \dfrac{45.5 - np}{\sqrt{n p \,(1-p)}}\right)$

$= P\left(z \le \dfrac{45.5 - 120 \cdot \frac{1}{3}}{\sqrt{120 \cdot \frac{1}{3} \cdot \frac{2}{3}}}\right)$

$= P(z \le 1.07) = 0.8577$