Midsem: Probability and Statistics (50 Marks)

[Instruction: Please state reasons wherever applicable.]

1 5 Marks

1. For a continuous non-negative random variable X prove that $E[X^2] = 2\int_x x\bar{F}_X(x)dx$ where $\bar{F}_X(x) = 1 - F_X(x)$

   ### Solution:

   Let $f(x)$ denote the density of the random variable X. Therefore by definition:

   $F_X(x) = \int_{-\infty}^{x} f(u)du$ *(1 mark)*

   Therefore:

   $1 - F_X(x) = \int_{x}^{\infty} f(u)du$ *(1 mark)*

   Substituting this is the given statement

   $2\int_0^\infty (x(\int_x^\infty f(u)du)dx$ (1 mark)

   Changing the order of integration: *(By Fubini's Theorem)*

   $2\int_0^\infty x \int_x^\infty f(u)dudx$

   $=\int_0^\infty \int_0^u 2xf(u)dudx$

   $= \int_0^\infty u^2 f(u)du = E[X^2]$ *(2 marks)*

Refer to these links for Fubini's theorem

- https://mathinsight.org/double$_i$*ntegral*$_i$*terated*

- https://sites.math.washington.edu/ morrow/$335_1 2/fubini.pdf$

Note: There are other possible solutions to the problem which will be marked on the basis of the presented arguments

2. Let X be a continuous random variable with distribution $F_X(\cdot)$ and density $f_X(x)$. Find the density and distribution for $Z = \sqrt{X}$.

Solution

(3 marks)

The Cumulative distribution function (CDF) is given as

$F_{\mathrm{Z}}(z) = P_{\mathrm{Z}}(Z \leq z)$

$F_{\mathrm{Z}}(z) = P_{\mathrm{X}}(\sqrt{X} \leq z)$

$F_{\mathrm{Z}}(z) = P_{\mathrm{X}}(X \leq z^2)$

$F_{\mathrm{Z}}(z) = F_{\mathrm{X}}(z^2)$

(2 marks)

For getting Probability Density function (PDF), we need to differentiate CDF :-

$f_{\mathrm{Z}}(z) = 2z f_{\mathrm{X}}(z^2)$

3. Consider two exponential random variables X and Y with parameters λ_1 and λ_2 respectively. Consider $Z = min(X, Y)$ (min stands for minimum). Find the probability density and cumulative distribution of Z.

Solution:

Given $Z = min(X, Y)$. As $X \geq 0, Y \geq 0 \implies Z \geq 0$.

Consider some $a \geq 0$,

$$\begin{aligned} P(Z > a) &= P(min(X, Y) > a) \\ &= P(X > a, Y > a) \\ &= P(X > a)P(Y > a) \quad (X \text{ and } Y \text{ are independent}) \end{aligned}$$

We know that $P(X > a) = 1 - F_X(a) = 1 - (1 - e^{-\lambda_1 a}) = e^{-\lambda_1 a}$. Similarly, $P(Y > a) = e^{-\lambda_2 a}$.

Substituting above,

$$P(Z > a) = e^{-\lambda_1 a} e^{-\lambda_2 a} = e^{-(\lambda_1 + \lambda_2)a}$$

$\therefore P(Z \leq a) = 1 - P(Z > a) = 1 - e^{-(\lambda_1+\lambda_2)a}$ *(3 marks).*

$$\begin{aligned} &\implies Z \sim Exponential(\lambda_1 + \lambda_2) \quad \textit{(1 mark)} \\ &\implies f_Z(z) = (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)z} \quad \textit{(1 mark)} \end{aligned}$$

Marks Division

- 3M for deriving CDF of Z
- 1M for interpreting $Z \sim Exponential(\lambda_1 + \lambda_2)$ (from CDF of Z).
- 1M for writing PDF of Z (after derivation)

Note: Simple stating of final answers without any logical approach will be given 0.

4. Let X and Y denote Gaussian random variables with mean μ_1 and μ_2 and standard deviation σ_1 and σ_2 respectively. Consider $Z = X+Y$. Using Moment generating functions, show that Z is also a Gaussian random variable, with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

Solution:

Moment Generating Function of Normal Distribution:

From the definition of the Gaussian distribution, X has probability density function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

From the definition of a moment generating function:

$$M_X(t) = \mathsf{E}\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, \mathrm{d}x$$

So:

$$\begin{aligned}
M_X(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(tx - \frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x \\
&= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\left(\sqrt{2}\sigma u + \mu\right) t - u^2\right) \mathrm{d}u && \text{substituting } u = \frac{x-\mu}{\sqrt{2}\sigma} \\
&= \frac{\exp \mu t}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(-\left(u^2 - \sqrt{2}\sigma u t\right)\right) \mathrm{d}u \\
&= \frac{\exp \mu t}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(-\left(u - \frac{\sqrt{2}}{2}\sigma t\right)^2 + \frac{1}{2}\sigma^2 t^2\right) \mathrm{d}u \\
&= \frac{\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(-v^2\right) \mathrm{d}v && \text{substituting } v = u - \frac{\sqrt{2}}{2}\sigma t \\
&= \frac{\sqrt{\pi} \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)}{\sqrt{\pi}} && \text{Gaussian Integral} \\
&= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)
\end{aligned}$$

Moment generating functions are a linear combination of of independent random. variables, hence:

$Z = Y + X$

$M_Z(s) = E[e^{sZ}]$

$M_Z(s) = E[e^{s(X+Y)}]$

$M_Z(s) = E[e^{sX}e^{sY}]$

$M_Z(s) = E[e^{sX}] * E[e^{sY}])$ (since X and Y are independent)

$M_Z(s) = M_X(s) * M_Y(s)$

Using this Property:

$M_Z(t) = M_X(t) * M_Y(t)$

If X $N(\mu, \sigma^2)$, then
$M_X(t) = exp(\mu t + \frac{\sigma^2 t^2}{2})$
Therefore the moment generating function for Z is:
$M_Z(t) = M_X(t) * M_Y(t) = exp(\mu_x t + \frac{\sigma_x^2 t^2}{2})exp(\mu_y t + \frac{\sigma_y^2 t^2}{2})$
Evaluating the product
$M_Z(t) = exp(\mu_x t)exp(\mu_y t)exp(\frac{\sigma_y^2 t^2}{2})exp(\frac{\sigma_x^2 t^2}{2})$
Re-arranging the product using properties of exponentiation
$M_Z(t) = exp[t(\mu_x + \mu_y) + \frac{t^2}{2}(\sigma_x^2 + \sigma_y^2)]$
Therefore we have just shown that the moment generating function of the Random variable Z is same as that of a RV with
mean $= \mu_x + \mu_y$
variance $= \sigma_x^2 + \sigma_y^2$
Therefore by the **Uniqueness property of Moment Generating function,** Z must be normally distributed with the given mean and variance.

Marks Division:

- Deriving Expression for MGF of Gaussian Distribution *2 marks*
- Proving $M_z(s) = M_X(s) * M_Y(s)$ *0.5 Marks*
- Proof of the required expression *2Marks*

- Uniqueness of MGF *0.5 Marks*

5. Let U_1 and U_2 be two independent Uniform random variables with support [0,1]. Then find the cdf or pdf of $U_1 + U_2$.

   Solution:
   Let $Z = U_1 + U_2$, then we can use the convolution formula:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{U_1}(u_1) f_{U_2}(z - u_1) du_1$$

   Since U_1 and U_2 are both uniform random variables with support [0,1]:

$$f_{U_1}(u) = f_{U_2}(u) = \begin{cases} 1 & \text{if } 0 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   Thus, support for Z = [0,2]. For the product of pdfs to be non-zero in the convolution equation we must have:

$$\begin{aligned} 0 \leq u_1 \leq 1 \\ \text{and, } 0 \leq z - u_1 \leq 1 \implies z - 1 \leq u_1 \leq z \end{aligned}$$

   To aid this calculation we divide support of Z into 2 intervals.

   Case 1: $0 \leq z \leq 1$
   Since $0 \leq z \leq 1$, the 2 conditions coalesce to $0 \leq u_1 \leq z$. Applying this on the convolution equation we get:

$$f_Z(z) = \int_0^z f_{U_1}(u_1) f_{U_2}(z - u_1) du_1 = \int_0^z 1 \cdot du_1 = z$$

   Case 2: $1 \leq z \leq 2$
   Since $1 \leq z \leq 2$, the 2 conditions coalesce to $z - 1 \leq u_1 \leq 1$. Applying this on the convolution equation we get:

$$f_Z(z) = \int_{z-1}^1 f_{U_1}(u_1) f_{U_2}(z - u_1) du_1 = \int_{z-1}^1 1 \cdot du_1 = 2 - z$$

$$f_Z(z) = \begin{cases} z & \text{if } 0 \leq z \leq 1 \\ 2 - z & \text{if } 1 \leq z \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

   Marking Scheme: 2 + 1.5 * 2 (one for each case)
   Note: This is one method to solve this problem. There are other acceptable methods too and those will be marked on case by case basis.

6. If X and Y are independent random variables, prove that $Var(X+Y) = Var(X)$ + $Var(Y)$. (Recall that $Var(X) = E[(X - E[X])^2]$))

   Solution: For Independent Random Variables X and Y, we have,

$$E[XY] = E[X]E[Y] \qquad \textit{(1.5 mark)} \tag{1}$$

   We can expand $Var(X+Y)$ as, *(3 mark)*
   $Var(X+Y) = E[(X+Y)^2] - (E[X+Y])^2$
   $= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$
   $= E[X^2] + 2E[XY] + E[Y^2] - [(E[X])^2 + 2E[X]E[Y] + (E[Y])^2]$
   $= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2E[XY] - 2E[X]E[Y]$
   $= Var(X) + Var(Y) + 2E[XY] - 2E[X]E[Y]$
   Since, X and Y are independent, based on above, we substitute equation (1) and finally get:
   $Var(X+Y) = Var(X) + Var(Y)$ *(0.5 mark)*

2 10 Marks

1. $X_1, X_2, ..., X_n$ are independent and identically distributed $Bernoulli(p)$ random variables (i.e., they take the value 1 with probability p and 0 otherwise). Consider $S_n = \sum_{i=1}^{n} X_n$ Find the PMF, MGF, mean and variance of Sn.

Solution:

PMF (5 marks)

Given,

$$\begin{aligned}
S_n &= \sum_{i=1}^{n} X_i \\
\implies M_{S_n}(t) &= E[e^{tS_n}] \\
&= E[e^{t\sum X_i}] \\
&= E[e^{tX_1}e^{tX_2}\cdots e^{tX_n}] \\
&= E[e^{tX_1}]E[e^{tX_2}]\cdots E[e^{tX_n}] \quad \text{(Since } X_i\text{'s are independent)} \\
&= \left(E[e^{tX_1}]\right)^n \quad \text{(Since } X_i\text{'s are identically distributed)} \\
&= (e^0(1-p) + e^t p)^n \\
&= (1 - p + e^t p)^n
\end{aligned}$$

which is the MGF of a Binomial distribution with parameters n and p (By the uniqueness property of MGFs).

$$\therefore S_n \sim Binomial(n, p) \quad \textit{(2 mark)}$$

Note: *2 marks* for correctly identifying and proving that S_n is a Binomial Distribution using either combinatorial proof, MGF-uniqueness, convolution arguments or other valid arguments.

$\therefore$ Range of S_n, $R = \{0, 1, 2, 3, \ldots, n\}$ *(1 mark)*

PMF is defined as follows:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where } k \in R \quad \textit{(2 marks)}$$

MGF (2 marks)

MGF has already been computed in the previous subpart. An alternate approach is as follows:

$$
\begin{aligned}
M_{S_n}(t) &= E[e^{tS_n}] \\
&= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} e^{tk} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k (1-p)^{n-k} \\
&= (1-p+e^t p)^n \quad \text{(By Binomial Theorem)}
\end{aligned}
$$

Mean of S_n (1.5 mark)

$$
\begin{aligned}
E[S_n] &= \left[M'_{S_n}(t)\right]_{t=0} \\
&= \left[np \cdot e^t \cdot (1-p+e^t p)^{n-1}\right]_{t=0} \\
&= np
\end{aligned}
$$

1 mark for working and *0.5 mark* for final answer.

Variance of S_n (1.5 mark)

$$
\begin{aligned}
Var(S_n) &= E[S_n^2] - (E[S_n])^2 \\
&= \left[M''_{S_n}(t)\right]_{t=0} - (np)^2 \\
&= \left[np \cdot e^t \ (1-p+e^t p)^{n-1} + n(n-1)p^2 \cdot e^{2t} \ (1-p+e^t p)^{n-2}\right]_{t=0} - n^2p^2 \\
&= np + n(n-1)p^2 - n^2p^2 \\
&= np + n^2p^2 - np^2 - n^2p^2 \\
&= np - np^2 \text{ or } np(1-p)
\end{aligned}
$$

1 mark for working and *0.5 mark* for final answer.

Note: PMF is used in all the other parts. So, if the PMF is not correct, then marks will be deducted for the other subparts. There are other acceptable approaches to this problem which will be marked based on the arguments/working presented.

2. (a) (5 marks) The joint probability mass function of the discrete random variables X and Y are given by $p_{X,Y}(x,y) = \frac{1}{2^{x+y}}$, $x = 1, 2., ...$ and $y = 1, 2, ...$

   i. Find the expression for marginal pmf $p_X(x)$ and $p_Y(y)$ and the conditional pmf $p_{X|Y}(x|y)$.

   ii. Find $E[XY]$ and determine if the RV X and Y are independent.

   (b) (5 Marks) The joint pdf of random variables X and Y is given by $f_(X,Y)(x,y) = \lambda\mu e^{-\lambda x-\mu y}, x \geq 0, y \geq 0, \lambda > 0, \mu > 0$.

   i. Find the expression for the marginal pdf's $f_X(x)$ and $f_Y(y)$ and the joint CDF $F_{X,Y}(x,y)$

   ii. Are the RV X and Y independent? Give reasons.

Solution

1. (a) i. The marginal PMF of x, $p_X(x)$ is defined as

$$p_X(x) = \sum_{y_j \in R_Y} p_{XY}(x, y_j), \text{ for any } x \in R_X$$

Therefore,

$$\begin{aligned}
p_X(x) &= \sum_{y_j \in R_Y} p_{XY}(x, y_j) \\
&= \sum_{y_j \in N} \frac{1}{2^{x+y_j}} \\
&= \frac{1}{2^x} \times \sum_{y_j \in N} \frac{1}{2^{y_j}} \\
&= \frac{1}{2^x} \times \frac{\frac{1}{2}}{1-\frac{1}{2}} \\
&= \frac{1}{2^x} \times \frac{\frac{1}{2}}{\frac{1}{2}} \\
&= \frac{1}{2^x} \quad \text{(1 Mark)}
\end{aligned}$$

By symmetry,

$$p_Y(y) = \frac{1}{2^y} \quad \text{(1 Mark)}$$

The Conditional PMF $p(X|Y)$ can be given as:

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p_{XY}(x,y)}{p_Y(y)} \\ &= \frac{\frac{1}{2^{x+y}}}{\frac{1}{2^y}} \\ &= \frac{1}{2^x} \quad \text{(1 Mark)} \end{aligned}$$

ii. Clearly,

$$\begin{aligned} p_{XY}(x,y) &= \frac{1}{2^{x+y}} \\ &= \frac{1}{2^x} \times \frac{1}{2^y} \\ &= p_X(x) \times p_Y(y) \end{aligned}$$

Thus, X and Y are **independent** random variables. (1 Mark)

The Expectation $E[XY]$ of two Independent Joint Random Variables X and Y is given as:

$$E[XY] = E[X]E[Y]$$

$$\begin{aligned} E[X] &= \sum_{x \in N} x \times p_X(x) \\ &= \sum_{x \in N} \frac{x}{2^x} \\ &= 2 \quad \text{(Using infinite Arithmetic Geometric Progression Sum or Different Proof} \end{aligned}$$

Proof for sum
Similarly by symmetry,

$$E[Y] = 2$$

Therefore,

$$E[XY] = E[X]E[Y] = 2 \times 2 = 4 \quad \text{(1 Mark)}$$

(b) i. The marginal PDF of X (continuous variable) is defined as :

$$f_X(x) = \int_y f_{XY}(x, y)$$

Therefore,

$$\begin{aligned} f_X(x) &= \int_0^{\infty} \lambda\mu e^{-\lambda x - \mu y} dy \\ &= -\lambda e^{-\lambda x} \Big[e^{-\mu y} \Big]_0^{\infty} \\ &= \lambda e^{-\lambda x} \quad \text{(1.25 marks)} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_0^{\infty} \lambda\mu e^{-\lambda x - \mu y} dx \\ &= \mu e^{-\mu y} \quad \text{(1.25 marks)} \end{aligned}$$

The joint CDF is defined as follows:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = \int_0^x \int_0^y f_{XY}(x, y)\, dy\, dx$$

Therefore,

$$\begin{aligned} F_{X,Y}(x, y) &= \int_0^x \int_0^y \lambda\mu e^{-\lambda x - \mu y} dy\, dx \\ &= \int_0^x -\lambda e^{-\lambda x} \Big[e^{-\mu y} \Big]_0^y dx \\ &= -\lambda\big(e^{-\mu y} - 1\big) \int_0^x e^{-\lambda x} dx \\ &= \big(e^{-\mu y} - 1\big) \Big[e^{-\lambda x} \Big]_0^x \\ &= \big(e^{-\lambda x} - 1\big)\big(e^{-\mu y} - 1\big) \\ &= (1 - e^{-\lambda x})(1 - e^{-\mu y}) \quad \text{(1.5 marks)} \end{aligned}$$

From the next question, X and Y are independent.
The joint CDF of independent variables is as follows:

$$F_{XY}(x, y) = F_X(x) F_Y(y)$$

Therefore,

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda x} dx \\ &= -\left[e^{-\lambda x} \right]_0^x \\ &= (1 - e^{-\lambda x}) \end{aligned}$$

Similarly,

$$\begin{aligned} F_Y(y) &= \int_0^x \mu e^{-\mu y} dy \\ &= (1 - e^{-\mu y}) \end{aligned}$$

Therefore,

$$F_{XY}(x, y) = (1 - e^{-\lambda x})(1 - e^{-\mu y})$$

ii. Two random variables are independent when:

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

Applying this,

$$\begin{aligned} f_X(x) f_Y(y) &= \mu e^{-\lambda x} \lambda e^{-\mu y} \\ &= \lambda \mu e^{-\lambda x - \mu y} \\ &= f_{XY}(x, y) \end{aligned}$$

Therefore, X and Y are independent. (1 mark)