

ECS764P -
Applied Statistics

Examination, January 2024

Student details:

Name: Pranav Narendra
Gopalkrishna

Number: 231052045

Question 1

a) $\int_{-\infty}^{\infty} f(x) dx = 1$

must be true for f if it is a probability density function

$$\begin{aligned}
 b) \quad & \int_{-\infty}^{\infty} f(x) dx = \int_0^1 (ax - 2)^2 dx \\
 &= \int_0^1 (a^2x^2 + 4 - 4ax) dx \\
 &= \left[\frac{a^2x^3}{3} + 4x - \frac{4ax^2}{2} \right]_0^1 \\
 &= \frac{a^2}{3} + 4 - 2a
 \end{aligned}$$

* Put the above as $= 1$

$$\Rightarrow \frac{a^2}{3} + 4 - 2a = 1$$

$$\begin{aligned}
 & \Rightarrow a^2 + -6a + 12 = 0 \\
 & \Rightarrow a = 3 + 1.73205 i \\
 & \text{or } a = 3 - 1.73205 i
 \end{aligned}$$

$$\Rightarrow a^2 + 12 - 6a = 3$$

$$\Rightarrow a^2 - 6a + 9 = 0$$

$$\Rightarrow a - 3a - 3a + 9 = 0$$

$$\Rightarrow a(a-3) - 3(a-3) = 0$$

$$\Rightarrow \underline{\underline{a = 3}}$$

$$c) \quad \mathbb{P}([-1, \frac{1}{2}]) \\ = \int_{-1}^{1/2} (3x - 2)^2 dx$$

$$= \int_0^1 (3x - 2)^2 dx \quad (\text{since support is } [0, 1])$$

$$= \left[\frac{3^2 x^3}{3} + 4x - \frac{5 \cdot 3x^2}{2} \right]_0^{1/2}$$

$$= [3x^3 + 4x - 6x^2]_0^{1/2}$$

$$= \frac{3}{8} + \frac{5}{2} * -\frac{6}{5}$$

$$= \frac{1}{8}$$

d) \Leftrightarrow Let F be the CDF.

$$F(t) = \int_{-\infty}^t f(x) dx$$

$$= \int_{-\infty}^t (3x - 2)^2 dx$$

If $t \in [0, 1]$,

$$F(t) = 3t^3 + 4t - 6t^2$$

If $t < 0$ or $t > 1$,

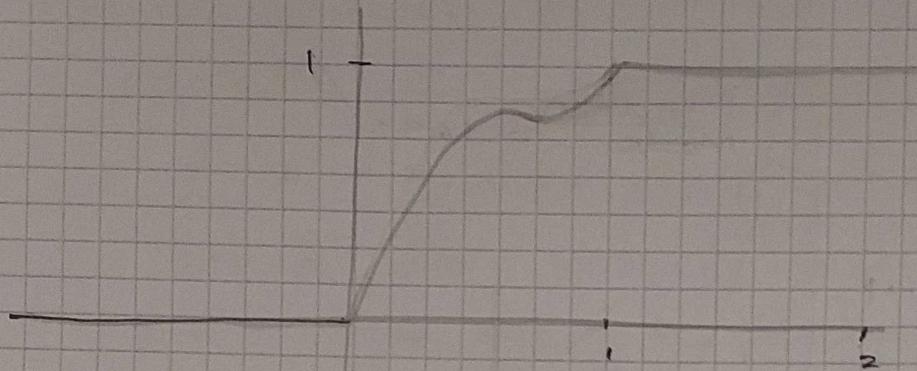
$$F(t) = 0$$

If $t > 1$,

$$F(t) = 1$$

Hence,

$$F(t) = \begin{cases} 3t^3 + 4t - 6t^2 & \text{if } t \in [0, 1] \\ 0 & \text{if } t < 0 \\ 1 & \text{if } t > 1 \end{cases}$$



e) Mean = $\mu(P)$

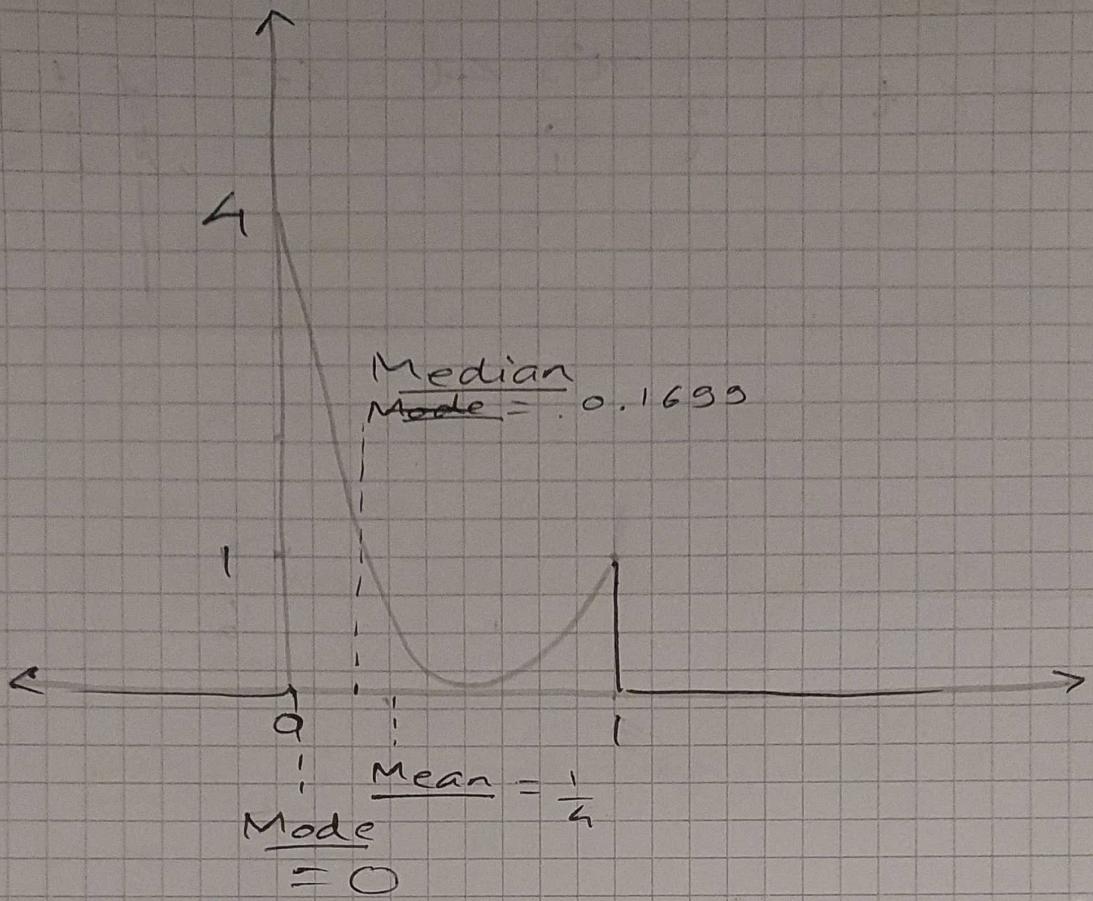
$$= \int_0^1 (3x-2)^2 \cdot x \, dx$$

$$= \int_0^1 (9x^3 + 4x - 12x^2) \, dx$$

$$= \left[\frac{9}{4}x^4 + \frac{4}{2}x^2 - \frac{12}{3}x^3 \right]_0^1$$

$$= \frac{9}{4} + 2 - 4$$

$$= \frac{1}{4}$$



$\text{Mode} < \text{Mean. } (\text{Mode} = 0)$
 \Rightarrow Positive skewness

Mod

I would not choose mode as the measure of centrality since the ~~data is b~~ points are highly skewed, thus the mode would not reflect an accurate central tendency.

Question 2

Part a

i) Support of $IP + IP + IP$ is exactly all possible sums

$$\{x_1 + x_2 + x_3 \mid x_1, x_2, x_3 \in \{-1, 1\}\}$$

Hence, we get

$\{-3, -1, 1, 3\}$ as the support.

ii) PMF (Probability mass function)

We know that

$$IP + IP + IP = + * (IP \otimes IP \otimes IP)$$

where $+ : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ s.t.

$$(x_1, x_2, x_3) \rightarrow x_1 + x_2 + x_3$$

Support was obtained as

$$\{-3, -1, 1, 3\}$$

$$IP + IP + IP (-3)$$

$$= + * (IP \otimes IP \otimes IP) (-3)$$

$$= (IP \otimes IP \otimes IP) (\{x_1, x_2, x_3\} \mid$$

By definition of pushforward $x_1 + x_2 + x_3 = -3$,
 $x_1, x_2, x_3 \in \{-1, 1\}\}$

$$= (IP \otimes IP \otimes IP) ((-1, -1, -1))$$

$$= IP(-1) IP(-1) IP(-1)$$

(by definition of product measure)

$$= \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{64}$$

Similarly we obtain

$$\begin{aligned}
 & (\bar{P} + P + \bar{P})(-1) \\
 &= (\bar{P} \otimes P \otimes \bar{P})(\{(-1, -1, -1), \\
 &\quad (-1, 1, -1), \\
 &\quad (-1, -1, 1)\}) \\
 &= \cancel{P}(\bar{P}(-1)P(-1)\bar{P}(-1)} \\
 &\quad + \bar{P}(-1)P(1)\bar{P}(-1) \\
 &\quad + P(-1)\bar{P}(-1)P(1) \\
 &= 3 \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \\
 &= \frac{9}{64}
 \end{aligned}$$

$$\begin{aligned}
 & (\bar{P} + P + \bar{P})(1) \\
 &= (\bar{P} \otimes P \otimes \bar{P})(\{(-1, 1, 1), \\
 &\quad (1, -1, 1), \\
 &\quad (1, 1, -1)\})
 \end{aligned}$$

$$\begin{aligned}
 &= 3 \cdot \bar{P}(-1)P(1)\bar{P}(1) \\
 &= 3 \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{27}{64}
 \end{aligned}$$

Hence, $(\bar{P} + P + \bar{P})(3)$

$$= 1 - \cancel{\frac{27}{64}} - \frac{9}{64} - \frac{1}{64}$$

(since probability measure of the support must be 1)

$$= \frac{27}{64}$$

$$\left(\frac{IP + (P+IP)}{3} \right) \left(-\frac{1}{3} \right) = (IP + IP + IP)(-1) \\ = \frac{9}{64}$$

$$\left(\frac{IP + (P+IP)}{3} \right) \left(\frac{1}{3} \right) = (IP + IP + IP)(1) \\ = \frac{27}{64}$$

$$\left(\frac{IP + (P+IP)}{3} \right) \left(\frac{1}{4} \right) = (IP + IP + IP)(3) \\ = \frac{27}{64}$$

iii) Let \bar{P}_{25} model the distribution of the average step size of a walk with 25 steps (i.e. sample size $= n = 25$).

CLT states:

$$\frac{\sqrt{n} (\bar{P}_{25} - \mu(IP))}{\sqrt{\text{Var}(IP)}} \approx \text{Norm}(0, 1)$$

as

(\bar{P}_n = distribution of the average step size of n random walk steps)

$$\Rightarrow \bar{P}_n \approx \text{Norm}\left(\mu(IP), \frac{\text{Var}(IP)}{n}\right)$$

Now, we have that

$$\mu(IP) = (-1) \frac{1}{4} + 1 \cdot \frac{3}{4} = \frac{1}{2}$$

$$\begin{aligned} \text{Var}(IP) &= \left(-1 - \frac{1}{2}\right)^2 \frac{1}{4} + \left(1 - \frac{1}{2}\right)^2 \frac{3}{4} \\ &= \frac{9}{16} + \frac{3}{16} = \frac{3}{4} \end{aligned}$$

$$n = 25$$

iii) This random walk is modelled by $\frac{1}{3}P + \frac{1}{3}P + \frac{1}{3}P$.

Mean of $\frac{1}{3}P + \frac{1}{3}P + \frac{1}{3}P$

$$= \sum_{x \in \{-3, -1, 1, 3\}} x \cdot (\frac{1}{3}P + \frac{1}{3}P + \frac{1}{3}P)(x)$$

$$= (-3) \cdot \frac{1}{64}$$

$$+ (-1) \cdot \frac{9}{64}$$

$$+ (1) \cdot \frac{27}{64}$$

$$+ (3) \cdot \frac{27}{64}$$

$$= 1.96875$$

Hence, mean position after 3 steps = 1.96875.

Part b

i) Support of $\frac{1}{3}(P + P + P)$

$$= \frac{1}{3} \cdot \text{Support of } P + P + P$$

$$= \left\{ -1, -\frac{1}{3}, \frac{1}{3}, 1 \right\}$$

ii) PMF of Note that

$$\frac{P + P + P}{3} = g_* (P + P + P)$$

where $g: \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \frac{x}{3}$

Hence, the PMF of $\frac{P + P + P}{3}$ is

($R(X \geq 1)$)

$$\left(\frac{P + P + P}{3} \right)(-1) = (P + P + P)(-3) = \frac{1}{64}$$

$$\therefore \bar{P}_{25} \approx \text{Norm}\left(\frac{1}{2}, \frac{3/4}{25}\right)$$

Probability of average step size being negative is then

$$\text{Norm}\left(\frac{1}{2}, \frac{3/4}{25}\right)((-\infty, 0))$$

$$= 0.0019462$$

$$= 0.19462 \% \text{ probability.}$$

Part c

Weak law of large numbers:

$$\lim_{N \rightarrow \infty} \bar{P}_N([u(P) - \epsilon, u(P) + \epsilon]) = 1 \quad (\text{for any } \epsilon > 0)$$

Hence, from this,

(i) & (v) follow.

Part d

Cauchy dist

Cauchy distribution.

Reason: the probability measure under this distribution has no defined mean.

Question 3

a) MLE (maximum likelihood estimator) is an estimator for a parameter (\hat{v} in this case) which maximises the probability of observing a given set of observations under the assumption that they are sampled from the given distribution.

$$\text{MLE}(\hat{v}) = \underset{\hat{v}}{\operatorname{argmax}} \left(\prod_{i=1}^n f_v(x_i) \right)$$

The joint probability of observing each observation x_i ($i = 1, 2, \dots, n$) must be maximised by optimising \hat{v} .

Difficulty: To find the maximum of the above function (i.e. joint probability), we must partially differentiate it with respect to \hat{v} , then find \hat{v} for which this partial derivative is zero.

However, differentiation of a product is complex.

To solve this, we transform

the product into sum using
using logarithm (which
preserves the maxima). Hence,

$$\text{MLE}(\hat{v}) = \underset{\hat{v}}{\operatorname{argmax}} \sum_{i=1}^n \log f_{\hat{v}}(x_i)$$

($\log \Rightarrow$ natural log)

NOTE: $\log \prod f_{\hat{v}}(x_i) = \sum \log f_{\hat{v}}(x_i)$

\hat{v} is the computed by solving

$$\frac{\delta \sum_{i=1}^n \log f_{\hat{v}}(x_i)}{\delta v} = 0$$

s.t. $\frac{\delta^2 \sum_{i=1}^n \log f_{\hat{v}}(x_i)}{\delta v^2} < 0$

to ensure it is
maximum not minimum.

b) MLE (\hat{v}) is calculated by solving:

$$\frac{\delta \sum_{i=1}^n \log f_{\hat{v}}(x_i)}{\delta v} = 0$$

$$\Rightarrow \frac{\delta \sum_{i=1}^n \log (2\hat{v} x_i e^{-\hat{v} x_i^2})}{\delta \hat{v}} = 0 \dots (1)$$

Now, note:

$$\begin{aligned} & \log (2\hat{v} x_i e^{-\hat{v} x_i^2}) \\ &= \log 2 + \log \hat{v} + \log x_i + \log e^{-\hat{v} x_i^2} \\ &= \log 2 + \log \hat{v} + \log x_i - \hat{v} x_i^2 \end{aligned}$$

NOTE: Log here
is natural log
(base e)

$\therefore (1)$ becomes

$$\nabla \sum_{i=1}^n (\hat{x}_i - x_i)^2 \neq 0$$

$$\nabla \frac{\sum_{i=1}^n \delta(\log 2 + \log \hat{v} + \log x_i - \hat{x}_i^2)}{\delta v} = 0$$

$$\Rightarrow \sum_{i=1}^n (\hat{x}_i - x_i^2) = 0$$

$$\Rightarrow \frac{n}{\hat{x}_i} - \sum_{i=1}^n x_i^2 = 0$$

$$\Rightarrow \hat{x}_i = \frac{n}{\sum_{i=1}^n x_i^2}$$

c) The law states:

$$\mu(g * \text{IP}) = \int g(x) f(x) dx$$

$$\text{Here, } g(x) = x^2$$

$$\therefore \mu(\text{IP}) = \int x^2 f(x) dx$$

$$= \int 2x^2 x^2 2\sqrt{x} e^{-\sqrt{x}} dx$$

$$= \int 2\sqrt{x}^3 e^{-\sqrt{x}} dx$$

$$= 2\sqrt{x} \int x^3 e^{-\sqrt{x}} dx$$

Now,

$$\frac{1}{\sqrt{v}}$$

$$= \frac{n}{\sum_{i=1}^n x_i^2}$$

$$\frac{n}{\sum_{i=1}^n x_i^2}$$

$$g * (\text{IP} \otimes \text{R} \dots \text{P})$$

$$\text{where } g: \mathbb{R}^n \rightarrow \mathbb{R}, (x_1, x_2, \dots, x_n) \mapsto \frac{\sum x_i^2}{n}$$

(done after exam)

Background for question 3c

Change of variable formula

Let X_1 & X_2 be 2 sets, & let f be a bijective function $f: X_1 \rightarrow X_2$.

Let g be some real valued function.

Also, define: (given P is the probability measure)

- p_1 as the density of dist. $P(X_1)$
- p_2 as the density of dist. $P(X_2)$

Then, the foll

Let P be ~~a~~ the probability measure on ~~X_2~~ . X_1 . Then, the probability measure on X_2 is defined by the pushforward measure

$f_* P$. Define:

- p_1 as the density of P
- p_2 as the density of $f_* P$

Then, the following holds:

$$\int g(x) p_2(x) dx = \int g(f(x)) p_1(x) dx$$

Change of variable formula

Now, put $g(x) = x$.

$$\Rightarrow \int x p_2(x) dx = \int f(x) p_1(x) dx$$

But $\int x p_2(x) dx = \mu(f_* P)$
(by definition of mean)

$$\Rightarrow \mu(f_* P) = \int f(x) p_i(x) dx$$

(Law of the unconscious statistician (LUS))

Question 3c.: (done after exam)

NOTE: There's a change in notation:

$f \Rightarrow$ the density of P

Hence, to restate LUS:

$$\mu(g_* P) = \int g(x) f(x) dx$$

(g does not refer to anything previously used, it is just an arbitrary function).

Now, we are given:

$$\mu'_2 = \int x^2 f(x) dx$$

(2nd order moment)

~~the RHS~~ The RHS matches the RHS of LUS as formulated, with $g(x) = x^2$.

$$\text{Hence, } \mu(g_* P) = \int x^2 f(x) dx$$

$$\text{where } g(x) = x^2 \dots \textcircled{1}$$

Now, consider the MLE of $\hat{\nu}$, i.e.

$$\hat{\nu} = n / \sum_{i=1}^n x_i^2$$

$$\Rightarrow \frac{1}{\hat{\nu}} = \frac{\sum_{i=1}^n x_i^2}{n}$$

Each x_i^2 is an independent sample from P , which means each x_i^2 is an independent sample of $g_* P$ (where $g(x) = x^2$ as defined)

$$\Rightarrow \frac{1}{n} = \frac{\sum x_i^2}{n} \text{ follows the distribution}$$

$$\frac{1}{n} \underbrace{(g * P) \otimes (g * P) \dots (g * P)}_{n \text{ times}} \\ = \overline{g * P} (\overline{g * P})_n$$

Now, we know that for any arbitrary arbitrary distribution

K , with mean μ ,

$$\mu(\overline{K}_n) = \mu(K) \quad (\text{i.e. sample mean is unbiased estimator of mean})$$

$$\Rightarrow \mu((\overline{g * P})_n) = \mu(g * P)$$

$$= \int x^2 f(x) dx \quad (\text{from } \textcircled{1})$$

$$= \mu'_2$$

$\Rightarrow \frac{1}{n}$ is an unbiased estimator of μ'_2 .

Question 4

Part a

i) We estimate β_1 as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

where

$$x_i \in x, y_i \in y$$

\bar{x} = sample mean of x

\bar{y} = sample mean of y

$n = 12$ (sample size)

We get:

$$\bar{x} = 8$$

$$\bar{y} = -4.5$$

Upon computation,

$$\hat{\beta}_1 = -2.95238$$

ii) We estimate β_0 as

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Using previously obtained values,
we get

$$\hat{\beta}_0 = 19.1190476$$

iii) In the case where we model the joint distribution of two variables X & Y (from which we have sampled ~~$x \in X^{12}$~~ $x \in X^{12}$ & $y \in Y^{12}$) as a linearly dependent distribution where Y is a linear function of X , i.e.

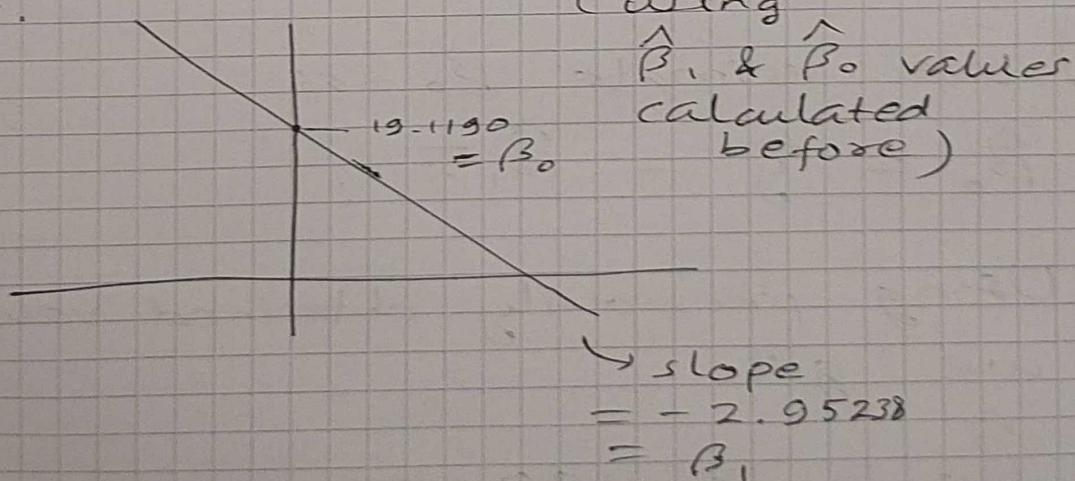
$$Y = \beta_1 X + \beta_0$$

Here, (which is a linear regression model), we have that

β_1 = slope of the regression

β_0 = Y -axis intercept of the regression

ex.



$$\Rightarrow p\text{-value} = 0.380$$

NOTE: t-test is 2 sided, so we compare as follows with the confidence level α :

$$0.380 > \frac{1-\alpha}{2} \quad (\alpha = 95\%)$$

i.e., $0.380 > \frac{0.05}{2}$

Hence, we fail to reject the null hypothesis that the samples were drawn from the joint distribution $P(X+Y)$ wherein

$$Y = \beta_1 X + \beta_0 + \epsilon$$

(with the given β_1 value that was tested & the assumed β_0 value)

ii) Similarly as (i), we do student's t-test for β_0 assuming $\beta_1 = -2 - 3$.

Test statistic for β_0 :

$$\frac{(\hat{\beta}_0 - \beta_0) \sqrt{n(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \sum_{i=1}^n x_i^2}} \sim t(n-2)$$

$$\text{Test statistic} = \star - 0.6275$$

$$\Rightarrow p\text{-value} = 0.27218875 > \frac{1-\alpha}{2}$$

Hence, we can fail to reject the ~~null~~ null hypothesis

iii) A joint distribution is a probability measure applied to a product space.

In our case, the above model does define a joint distribution over the product space

$$\{(x, \beta_0 + \beta_1 x + \varepsilon) \mid x \in X, \varepsilon \sim \text{Norm}(0, \sigma)\}$$

$\beta_0 + \beta_1 x + \varepsilon$ is also normally distributed, & x may be considered uniformly distributed.
(since ε is ~~a~~ normally distributed and x is a uniform random variable).

Thus, the joint distribution is also normal, hence a valid probability distribution

iv) β_1 is the rate of change in the target variable y with a unit change in the independent variable x . we got

$$\hat{\beta}_1 = -2.95238 < 0$$

\Rightarrow There is inverse relation between x & y

$\Rightarrow \rho_{xy}$ (i.e. correlation coefficient between x & y) is less than 0

$$\text{i.e. } \rho_{x,y} < 0$$

Part c (done after exam)

We have a joint density function

$$f(x, y) = \mathbb{1}_{[0,1]}(x) \mathbb{1}_{[x-\frac{1}{2}, x+\frac{1}{2}]}(y)$$

Consider: what does marginal density mean? Formally, it is the density of the pushforward $\pi_i * \mu_P$, where μ_P is the joint distribution (defined by f in our case), & π_i is the projection of the i th coordinate, i.e.

$$\pi_i : \mathbb{R}^n \rightarrow \mathbb{R}, (x_1, x_2, \dots, x_n) \mapsto x_i$$

(i.e. takes a tuple from the joint disto. & * returns the i th coordinate).

In our case,

$$\pi_1 : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto x$$

$$\pi_2 : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto y$$

ii) Consider $\pi_1 * \mu_P$. This defines the distribution of x based on the ~~distribution of each value~~ joint occurrence of each x with every possible y .

Hence, density of $\pi_1 * \mu_P$ (call it f_1) for each value of x is the ~~* density~~ of that value taken with every possible y .

$$\therefore \pi_1 * \mu_P f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

(i.e. take x as constant).

Now, note that if $x < 0$ or $x > 1$, $f(x, y) = 0$. Hence,

$f_1(x) = 0$ if $x < 0$ or $x > 1$

If $x \in [0, 1]$,

$$\begin{aligned} f_1(x) &= \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} f(x, y) dy \quad (\text{since } f(x, y) \\ &= \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} \mathbb{1}_{[x-\frac{1}{2}, x+\frac{1}{2}]}(y) dy \\ &= \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} 1 dy \quad (\text{since } \mathbb{1}_{[x-\frac{1}{2}, x+\frac{1}{2}]}(y) = 1 \\ &= \boxed{\cancel{[y]}} \\ &= [y]_{x-\frac{1}{2}}^{x+\frac{1}{2}} \\ &= x + \frac{1}{2} - x + \frac{1}{2} \\ &= 1 \end{aligned}$$

$$\therefore f_1(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{if } x < 0 \text{ or } x > 1 \end{cases}$$

iii) ii) Similarly,

$$\begin{aligned} f_2(\cancel{x})(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ \boxed{\substack{\downarrow \text{density of} \\ \mathcal{P}}} &= \int_0^1 f(x, y) dx \quad (\text{since } \mathbb{1}_{[0, 1]}(x) = 1 \text{ for } x \notin [0, 1]) \end{aligned}$$

If $y > 1 + \frac{1}{2} = \frac{3}{2}$, then

$y \notin [x - \frac{1}{2}, x + \frac{1}{2}] \nabla x \in [0, 1]$

Similarly, if $y < 0 - \frac{1}{2} = -\frac{1}{2}$,

$y \notin [x - \frac{1}{2}, x + \frac{1}{2}] \nabla x \in [0, 1]$

$\therefore f_2(y) = 0 \text{ if } y > \frac{3}{2} \text{ or } y < -\frac{1}{2}$

If $y \in [-\frac{1}{2}, \frac{3}{2}]$,

$$f_2(y) = \int_{-\infty}^{\infty} f_1(x) dx, \quad \int_0^1 1 dx$$
$$= [x]_0^1 = 1$$

$$\therefore f_2(y) = \begin{cases} 1 & \text{if } y \in [-\frac{1}{2}, \frac{3}{2}] \\ 0 & \text{if } y < -\frac{1}{2} \text{ or } y > \frac{3}{2} \end{cases}$$

iii) Mean of $\pi_1 * P$ is

$$\mu(\pi_1 * P) = \int_{-\infty}^{\infty} x f_1(x) dx$$

$$= \int_0^1 x f_1(x) dx$$

$$= \int_0^1 x dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

mean of $\pi_2 * P$ is

$$\mu(\pi_2 * P) = \int_{-\infty}^{\infty} y f_2(y) dy$$

$$= \int_{-1/2}^{3/2} y f_2(y) dy$$

$$= \int_{-1/2}^{3/2} y dy = \left[\frac{y^2}{2} \right]_{-1/2}^{3/2}$$

$$= \frac{9}{8} - \frac{1}{8} = 1$$

$$\text{Cov}(P) = \iint (x - \frac{1}{2})(y - 1) f(x, y) dx dy$$