

①a) i) We know that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

$$\Rightarrow K \int_0^{\infty} \int_0^{\infty} \frac{e^{-x/y} e^{-y}}{y} dx dy = 1$$

$$\Rightarrow K \int_0^{\infty} \left[\frac{y}{y} (-e^{-x/y}) \right]_0^{\infty} e^{-y} dy = K \int_0^{\infty} e^{-y} dy = 1.$$

$$\Rightarrow K(1 - 0) = 1$$

$$\Rightarrow \boxed{K = 1}$$

ii) Marginal density of Y , $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$.

$$= \int_0^{\infty} \frac{e^{-x/y} e^{-y}}{y} dx = \left(\frac{e^{-y}}{y} \right) (y) (e^{-x}) \Big|_0^{\infty}$$

$$\Rightarrow \boxed{f_Y(y) = \frac{e^{-y}}{0.4}}$$

$$\text{iii) } \Pr(0 < X < 1, 0.2 < Y < 0.4) = \int_{0.2}^{0.4} \int_0^1 f(x, y) dx dy$$

$$= \int_{0.2}^{0.4} \int_0^1 \frac{e^{-x/y} e^{-y}}{y} dx dy = \int_{0.2}^{0.4} (1 - e^{-1/y}) e^{-y} dy$$

Numerically integrating using 'integral' in MATLAB,

$$\rightarrow \Pr(0 < X < 1, 0.2 < Y < 0.4) = \boxed{0.1429}$$

$$iv) f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}$$

From part (ii) we know the marginal density of Y ,

$$\Rightarrow f_{X|Y=y}(x) = \frac{e^{-x/y} e^{-y}}{y} \times \frac{1}{e^{-y}} = \frac{e^{-x/y}}{y}$$

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x) dx.$$

$$= \int_0^{\infty} \frac{x e^{-x/y}}{y} dx. \quad \text{Let } \frac{x}{y} = \beta \Rightarrow dx = y d\beta.$$

and $x \geq 0 \Rightarrow \beta \geq 0$,
 $x \rightarrow \infty \Rightarrow \beta \rightarrow \infty$

$$\Rightarrow E(X|Y=y) = \int_0^{\infty} \beta e^{-\beta} y d\beta = y \int_0^{\infty} \beta e^{-\beta} d\beta.$$

$$= y \left(-\beta e^{-\beta} \Big|_0^{\infty} + \int_0^{\infty} e^{-\beta} d\beta \right)$$

$$\Rightarrow E(X|Y=y) = y$$

b) For a bivariate Gaussian,

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

$$\Rightarrow |\Sigma| = \sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2 = (1 - \rho^2) \sigma_X^2 \sigma_Y^2$$

$$\text{e} \Sigma^{-1} = \begin{bmatrix} \sigma_Y^2 & -\sigma_{XY} \\ -\sigma_{XY} & \sigma_X^2 \end{bmatrix}$$

$$f(x, y) = \frac{1}{2\pi \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} \begin{bmatrix} x - \mu_x & y - \mu_y \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}\right)$$

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2\sigma_x^2\sigma_y^2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\frac{(x-\mu_x)(y-\mu_y)\rho}{\sigma_x\sigma_y}\right)\right)$$

Substituting the values of $|\Sigma|$ and Σ^{-1} and simplifying,

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left(\frac{(x-4x)^2}{\sigma_x^2} + \frac{(y-4y)^2}{\sigma_y^2} - \frac{2(x-4x)(y-4y)\rho}{\sigma_x\sigma_y}\right)\right)$$

$$\text{Let } \frac{x-4x}{\sigma_x} = \alpha \text{ and } \frac{y-4y}{\sigma_y} = \beta$$

$$\Rightarrow f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(\alpha^2 + \beta^2 - 2\alpha\beta\rho)\right)$$

$$E(Y|X) = \int_{-\infty}^{\infty} y f_{Y|X}(y) dy$$

$$= \int_{-\infty}^{\infty} \frac{f(x, y)}{f_X(x)} y dy$$

We need $f_X(x)$ (marginal distribution of X)

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(\alpha^2 + \beta^2 - 2\alpha\beta\rho)\right) dy$$

$$\text{completing squares, } \alpha^2 + \beta^2 - 2\rho\alpha\beta = \beta^2 - 2\rho\alpha\beta + \rho^2\alpha^2 + (1-\rho^2)\alpha^2$$

$$= (\beta - \rho\alpha)^2 + (1-\rho^2)\alpha^2$$

Also $\frac{y - \mu_y}{\sigma_y} = \beta \Rightarrow \frac{dy}{\sigma_y} = d\beta$.

~~$$\Rightarrow f_x(x) = \frac{1}{2\pi\sigma_x} \int \exp\left(-\frac{(x + \beta)^2}{2(1-\rho^2)}\right) d\beta$$~~

$$\begin{aligned} \Rightarrow f_x(x) &= \frac{1}{2\pi\sigma_x} \int_{-\infty}^{\infty} \exp\left(-\left[\frac{(\beta - \rho x)^2}{2(1-\rho^2)} + (1-\rho^2)x^2\right]\right) d\beta \\ &= \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\left[\frac{(\beta - \rho x)^2}{2(1-\rho^2)}\right]\right) d\beta \\ &\quad \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \end{aligned}$$

The given integral is the integral of a normal random variable with mean ρx & $\sigma = \sqrt{1-\rho^2}$.

$$\Rightarrow f_x(x) = \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}}$$

in terms of x , $f_x(x) = \frac{\exp\left(-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right)}{\sqrt{2\pi}\sigma_x}$

(A normal RV with mean μ_x and standard deviation σ_x)

$$\begin{aligned} \therefore E(Y|X) &= \int y \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left[\frac{x^2 + y^2 - 2\rho xy}{1-\rho^2}\right]\right) dy \\ &= \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}\sigma_x} \end{aligned}$$

$$\Rightarrow E(Y|X) = \int \frac{y \exp\left(-\frac{1}{2} \frac{\alpha^2 + \beta^2 - 2\alpha\beta\gamma}{1-\beta^2}\right) dy}{\sqrt{2\pi} \sigma_Y \sqrt{1-\beta^2}}$$

$$= \int \frac{y}{\sqrt{2\pi} \sigma_Y \sqrt{1-\beta^2}} \left(\exp\left(-\frac{1}{2} \frac{(\beta - \alpha\gamma)^2}{(1-\beta^2)}\right) \right) dy$$

Let $\frac{y - \mu_Y}{\sigma_Y} = \beta \Rightarrow \frac{dy}{\sigma_Y} = d\beta$ and $y \rightarrow \infty \Rightarrow \beta \rightarrow \infty$
 $y \rightarrow -\infty \Rightarrow \beta \rightarrow -\infty$

Adding and subtracting μ_Y in the Numerator,

$$= \int \left(\frac{y - \mu_Y}{\sigma_Y} \right) \frac{1}{\sqrt{2\pi} \sqrt{1-\beta^2}} \exp\left(-\frac{1}{2} \frac{(\beta - \alpha\gamma)^2}{(1-\beta^2)}\right) d\beta + \frac{\mu_Y}{\sigma_Y \sqrt{2\pi} \sqrt{1-\beta^2}} \exp\left(-\frac{1}{2} \frac{(\beta - \alpha\gamma)^2}{(1-\beta^2)}\right) d\beta$$

$$= \frac{\sigma_Y}{\sqrt{2\pi} \sqrt{1-\beta^2}} \left[\int_{-\infty}^{\infty} \beta \exp\left(-\frac{1}{2} \frac{(\beta - \alpha\gamma)^2}{(1-\beta^2)}\right) d\beta + \int_{-\infty}^{\infty} \frac{\mu_Y}{\sigma_Y} \exp\left(-\frac{1}{2} \frac{(\beta - \alpha\gamma)^2}{(1-\beta^2)}\right) d\beta \right]$$

first integral: $\frac{\sigma_Y}{\sqrt{2\pi} \sqrt{1-\beta^2}} \int_{-\infty}^{\infty} \beta \exp\left(-\frac{1}{2} \frac{(\beta - \alpha\gamma)^2}{(1-\beta^2)}\right) d\beta$

The integral is the first moment of a normal R.V having $E(X) = \beta\alpha\gamma$ and $\sigma^2 = (1-\beta^2)$

$$\Rightarrow \text{first integral} = \sigma_{XY} (\beta\alpha) = \sigma_Y \left(\frac{\sigma_{XY} \alpha}{\sigma_X} \right)$$

$$= \frac{\sigma_{XY}}{\sigma_X} \alpha.$$

$$\text{Second integral: } \frac{\mu_Y}{1} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} \exp \left(-\frac{1}{2} \frac{(\beta - \beta\alpha)^2}{(1-\rho^2)} \right) d\beta.$$

The integral is the integral of a ^{PDF of} normal
 rv with $E(X) = \beta\alpha$ & $\sigma = \sqrt{1-\rho^2}$ ^{throughout}
 its domain \Rightarrow value of the integral is 1

$$\rightarrow \text{Second integral} = \mu_Y$$

$$\therefore E(Y|X) = \frac{\sigma_{XY}}{\sigma_X} \alpha + \mu_Y$$

$$\text{But } \alpha = \frac{X - \mu_X}{\sigma_X}$$

$$\Rightarrow E(Y|X) = \frac{\sigma_{XY}}{\sigma_X^2} (X - \mu_X) + \mu_Y$$

$$\Rightarrow E(Y|X) = \left(\frac{\sigma_{XY}}{\sigma_X^2} \right) X - \frac{\mu_X \sigma_{XY}}{\sigma_X^2} + \mu_Y$$

Thus $E(Y|X)$ is a linear function of X

$$\text{slope} = \frac{\sigma_{XY}}{\sigma_X^2}, \text{ intercept} = \mu_Y - \frac{\mu_X \sigma_{XY}}{\sigma_X^2}$$

② → The given expression for sample covariance matrix has been used and a function for evaluating the variance-covariance matrix has been written in MATLAB.

→ We note that the `cov()` function in MATLAB normalises by $N-1$ rather than N by default.

So I have passed an additional parameter `cov([x y], 1)` to ensure that it normalises by N and it can be compared with the user defined function.

→ X has been declared using `randn` as

$$Y = \text{randn}(\text{size}(X, 1)) \times 2 + 1;$$

$$(\sigma = 2 \text{ and } \mu = 1)$$

$$Y \text{ is simply } 3X^2 + 5X$$

(elementwise multiplication and elementwise addition in case of vectors)

→ Theoretical expectation: $\sigma_{XY} = E(XY) - E(X)E(Y)$

$$= (E(X(3X^2 + 5X)) - E(X)E(3X^2 + 5X))$$

$$= 3E(X^3) + 5E(X^2) - (E(X))(3)(E(X^2)) - 5(E(X))^2$$

$$E(X) = 1 \times 2 = 2$$

$$E(X^2) = \sigma_X^2 + (E(X))^2 = 4 + 1 = 5$$

Q To find $E(X^3)$, lets find $E(X^3) = E(a^3)$ where $A \sim N(\mu, \sigma)$

$$E(a^3) = \int_{-\infty}^{\infty} a^3 f(a) da$$

$$= \int_{-\infty}^{\infty} a^3 f(a) da$$

Since Gaussian is symmetric, $f(a) = f(1-a)$

$\Rightarrow a^3 f(a)$ is an odd function

$$\Rightarrow E(X^3) = 0.$$

That means $E\left(\frac{(X - \mu)^3}{\sigma^3}\right) = 0$

$$\Rightarrow E((X - \mu)^3) = 0$$

$$\Rightarrow E(X^3) - 3E(X)\mu^2 + 3E(X)\mu^2 - (E(X^3)) = 0.$$

We know $E(X) = \mu = 1$, $E(X^2) = 5$

$$\Rightarrow E(X^3) = 13.$$

$\therefore \sigma_{XY} \text{ Theoretical} = 3 \times 13 + (5)(5) - 3 \times 5 \times 5 \times 1^2$

$= \boxed{44}$

To examine convergence:

- I generated samples of size $1-10^4$, although the convergence wasn't monotonic there was convergence
- I then decided to try higher values of n and then plot a curve of σ_{xy} for values between 10^5 and 10^7 in steps of 10^6 again it wasn't

This time I plot deviation from actual value
 i.e., $|\text{estimate} - \text{theoretical value}|$ vs Sample size.

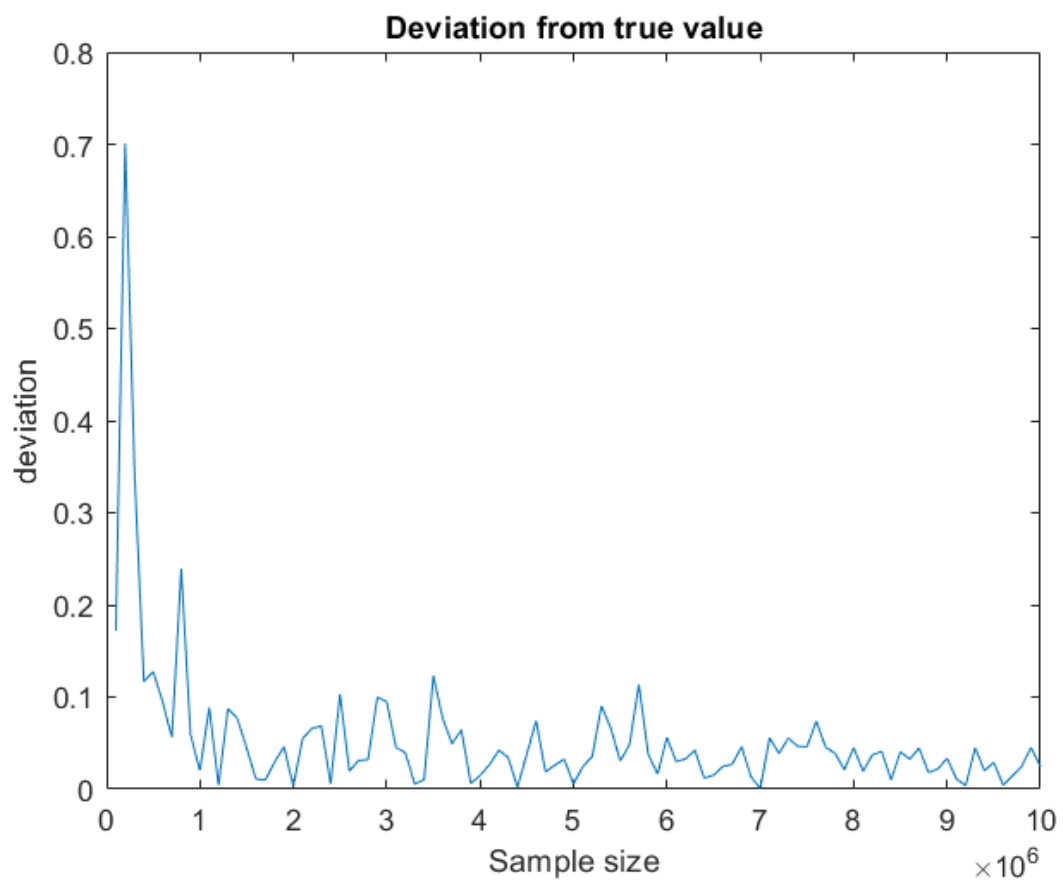
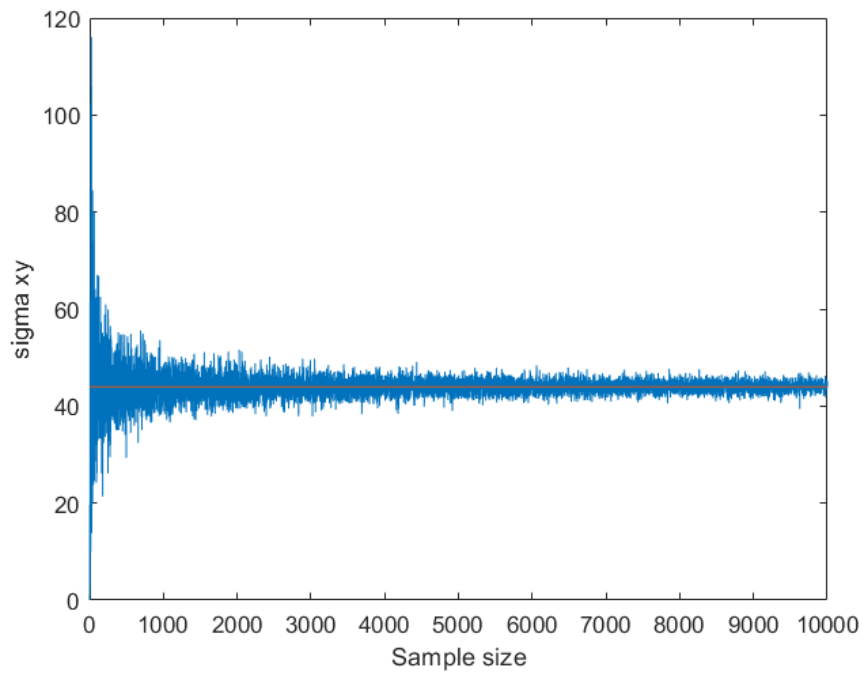
Again the convergence wasn't monotonic but there was convergence nonetheless (obs).

(these deviations went closer to zero as we increase n sample size)

| Sample size | $\sigma \times y$ | deviation |
|-----------------|-------------------|-----------|
| 10 | 51.055 | 7.055 |
| 10^2 | 41.177 | 2.823 |
| 10^3 | 44.079 | 0.0792 |
| 10^4 | 43.2184 | 0.7816 |
| 10^5 | 43.2596 | 0.2404 |
| 10^6 | 44.0896 | 0.0896 |
| 2×10^6 | 44.027 | 0.027 |
| 3×10^6 | 44.052 | 0.052 |

We can see that as we \uparrow n , the deviation bound
 closer and closer to zero.
 (or 'local mean deviation' goes to zero)

Question 2 Plots



③ a) $\Sigma = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix}$ Let $X_2 \rightarrow Y, X_3 = X_3$
 $Y = X_2$

variance-covariance matrix: $\begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 \end{bmatrix}$

to convert it into correlation

matrix: $\begin{bmatrix} \frac{\sigma_X^2}{\sigma_X^2} & \frac{\sigma_{XY}}{\sigma_X \sigma_Y} & \frac{\sigma_{XZ}}{\sigma_X \sigma_Z} \\ \frac{\sigma_{XY}}{\sigma_X \sigma_Y} & \frac{\sigma_Y^2}{\sigma_Y^2} & \frac{\sigma_{YZ}}{\sigma_Y \sigma_Z} \\ \frac{\sigma_{XZ}}{\sigma_X \sigma_Z} & \frac{\sigma_{YZ}}{\sigma_Y \sigma_Z} & \frac{\sigma_Z^2}{\sigma_Z^2} \end{bmatrix}$

(we need to normalise

the values as depicted above)

∴ ~~calculate~~ From Σ we can obtain,

$$\sigma_X = 2 \quad \sigma_Y = 3 \quad \sigma_Z = 5$$

$$\sigma_{XY} = 1 \quad \sigma_{XZ} = 2 \quad \sigma_{YZ} = -3$$

∴ correlation matrix ρ :

$$\begin{bmatrix} 1 & 1/6 & 1/5 \\ 1/6 & 1 & -1/5 \\ 1/5 & -1/5 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.1667 & 0.2 \\ 0.1667 & 1 & -0.2 \\ 0.2 & -0.2 & 1 \end{bmatrix}$$

$$\textcircled{3} \quad b) \quad \text{Cov}\left(X_1, \frac{X_2}{2} + \frac{X_3}{2}\right) = E\left(X_1 \left(\frac{X_2}{2} + \frac{X_3}{2}\right)\right) - E(X_1) E\left(\frac{X_2}{2} + \frac{X_3}{2}\right)$$

$\sigma_{X_1} \quad \sigma_{\left(\frac{X_2}{2} + \frac{X_3}{2}\right)}$

$$E\left(\frac{X_2}{2} + \frac{X_3}{2}\right) = E\left(\frac{X_2}{2}\right) + E\left(\frac{X_3}{2}\right) = \frac{E(X_2) + E(X_3)}{2} \quad \text{--- ①}$$

$$\begin{aligned} \text{var}\left(\frac{X_2 + X_3}{2}\right) &= \frac{1}{4} E\left(\left(\frac{X_2 + X_3}{2}\right)^2\right) - \left(E\left(\frac{X_2 + X_3}{2}\right)\right)^2 \\ &= \frac{1}{4} \left[E(X_2^2) + E(X_3^2) + 2E(X_2 X_3) - (E(X_2))^2 - (E(X_3))^2 - 2E(X_2)E(X_3) \right] \\ &= \frac{1}{4} \left[(E(X_2^2) - (E(X_2))^2) + (E(X_3^2) - (E(X_3))^2) + 2(E(X_2 X_3) - E(X_2)E(X_3)) \right] \\ &= \frac{1}{4} \left[\sigma_{X_2}^2 + \sigma_{X_3}^2 + 2\sigma_{X_2 X_3} \right] \end{aligned}$$

$$\therefore \sigma_{\left(\frac{X_2 + X_3}{2}\right)} = \frac{1}{2} \sqrt{\sigma_{X_2}^2 + \sigma_{X_3}^2 + 2\sigma_{X_2 X_3}} \quad \text{--- ②}$$

$$\begin{aligned} E\left(X_1 \left(\frac{X_2}{2} + \frac{X_3}{2}\right)\right) &= E(X_1) E\left(\frac{X_2 + X_3}{2}\right) \\ &= \frac{E(X_1 X_2)}{2} + \frac{E(X_1 X_3)}{2} - \frac{E(X_1)E(X_2)}{2} - \frac{E(X_1)E(X_3)}{2} \\ &= \frac{1}{2} (\sigma_{X_1 X_2} + \sigma_{X_1 X_3}) \quad \text{--- ③} \end{aligned}$$

From ①, ②, ③ we get,

$$\rho\left(X_1, \frac{X_2 + X_3}{2}\right) = \frac{\frac{1}{2}(\sigma_{X_1 X_2} + \sigma_{X_1 X_3})}{\sigma_{X_1} \left(\frac{1}{2} \sqrt{\sigma_{X_2}^2 + \sigma_{X_3}^2 + 2\sigma_{X_2 X_3}} \right)}$$

Substituting these values from variance & covariance matrix

$$\Rightarrow \rho\left(X_1, \frac{X_2 + X_3}{2}\right) = \frac{(2 + 1)}{2 \sqrt{(9 + 25 - 6)}}$$

$$\Rightarrow \boxed{\rho = +0.2835}$$

④

a). 200 records, each containing 10000 samples are generated from χ^2 distribution of dof 10.

o From the shape of the distribution the optimal MAE predictor X^* ($\min_x E(|X - \hat{X}|)$)

we can safely say that X^* should lie between 8 and 11

o A vector containing guess values of X^* is declared.

They are consecutive values with 0.001 spacing.

o The cost function for each record = $\text{mean } |X - \hat{X}^*|$, is found using 'sum' function and 'abs' function. (We use the 1-norm on the residual vector of each record)

o Each record is given 1/N weightage

$$J \text{ for a given record} = \sum_{i=1}^N |X(i) - \hat{X}| \times \frac{1}{N}$$

• The minimum cost is found using $\min()$

• $\min_J = 3.469 \Rightarrow \text{Average absolute error}$

$$\underline{X^* = 9.3360 = 9.336} = \boxed{3.469}$$

• We note that X^* varies in each run of the program.

• J vs \hat{X} curve is also plot.

b) μ_x is found using $\text{chi2stat}() = 10$

The required probabilities are found using the cdf of χ^2 distribution. $\Pr(0.9 X^* < X < 1.1 X^*)$

$$= F(X \leq 1.1 X^*)$$

where F is the cumulative distribution function. $- F(X \leq 0.9 X^*)$

We find that

$$\Pr(0.9 X^* < X < 1.1 X^*) = \underline{0.1724}$$

$$\Pr(0.9 \mu_x < X < 1.1 \mu_x) = \underline{0.1746}$$

$$\Rightarrow \Pr(0.9 \mu_x < X < 1.1 \mu_x) > \Pr(0.9 X^* < X < 1.1 X^*)$$

From the cdf plot we observe that,

μ_x is located nearer to the peak (in a more dense region)

than $X^* \quad \text{--- ①}$

Also $\mu_x > X^* \Rightarrow 0.2 \mu_x > X^*(0.2)$

\Rightarrow Range of values is greater in case of μ_x --- ②

Because of these 2 reasons we see

$$\Pr(0.94x < X < 1.14x) > \Pr(0.9x^* < X < 1.1x^*)$$

A more Qualitative reasoning:

→ x^* ~~is~~ can be considered an estimate of the median, since median minimises $\sum_{i=1}^n |x - x^*|$

Definition We can derive that $\int_{-\infty}^m f(x) dx = \int_m^{\infty} f(x) dx$.

where m is the ~~one~~⁰ median.

$$(\Rightarrow F(x \leq m) = 0.5)$$

So for X^2 distribution which has a sharper rise followed by a fall with tail values really low.

Compared to the ~~xx~~ left half of the peak, the median, by definition, being in the middle, will tend to the left side of the peak.

However $\mu_x = E(X) = \int_{-\infty}^{\infty} x f(x) dx$. Here, eventhough the tail probabilities are low, the ~~median~~ mean gives higher weightage to it than mean because of the fact that the ' x ' continuously increases in the right. This effect causes $E(X)$ to move ^{a bit} more to the right than the median.

So μ_x is closer to peak than x^*

→ μ_x is in a denser region, so the Probability of the RV taking ~~an~~ any value around it is more than $\Pr(0.9x^* < X < 1.1x^*)$ the RV taking value around the median.

Question 4 Cost function

