Task 1

A. $E[X+Y+Z]$: denotes the expectation when the events 'X', 'Y' and 'Z' happen simultaneously.

To prove, $E[X+Y+Z] = E[X] + E[Y] + E[Z]$

$$\Rightarrow E[X+Y+Z] = \sum_{x,y,z} (x+y+z)\, P(X=x, Y=y, Z=z)$$

$$= \sum_{x,y,z} x\, P(X=x, Y=y, Z=z) +$$

$$\sum_{x,y,z} y\, P(X=x, Y=y, Z=z) +$$

$$\sum_{x,y,z} z\, P(X=x, Y=y, Z=z)$$

Consider the first term,

$$\sum_{x,y,z} x\, P(X=x, Y=y, Z=z) = \sum_{x} x \sum_{y}\sum_{z} P(X=x, Y=y, Z=z)$$

where $\sum_{y}\sum_{z} P(X=x, Y=y, Z=z) = P(X=x)$

$$\Rightarrow \sum_{x,y,z} x\, P(X=x, Y=y, Z=z) = \sum_{x} x\, P(X=x) = E[X]$$

(by definition)

Similarly, $\sum_{x,y,z} y\, P(X=x, Y=y, Z=z) = E[Y]$, and

$$\sum_{x,y,z} z\, P(X=x, Y=y, Z=z) = E[Z]$$

$$\therefore \quad E[X+Y+Z] = E[X] + E[Y] + E[Z] \quad \square$$

——x——

B.

(a) $\begin{bmatrix} 8 & -5 \\ -5 & 10 \end{bmatrix} \rightarrow$ Valid.

(b) $\begin{bmatrix} 3 & 7 \\ 7 & -5 \end{bmatrix} \rightarrow$ Invalid ; as the matrix is not positive semi-definite having one eigenvalue as -9.06

(c) $\begin{bmatrix} 8 & 5 \\ 3 & 3 \end{bmatrix} \rightarrow$ Invalid ; not symmetric

(d) $\begin{bmatrix} 3 & 7 \\ 7 & 4 \end{bmatrix} \rightarrow$ Invalid; matrix not positive semi-definite, one eigenvalue as - 3.52

(e) $\begin{bmatrix} 5 & 4 \\ 4 & 8 \end{bmatrix} \rightarrow$ Valid.

———— x ————

C.

Rick : 4points 'r'

$$\mu_r = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad \Sigma_{rr} = \begin{bmatrix} 5 & 3 \\ 3 & 7 \end{bmatrix}$$

Morty : 6 points 'm'

$$\mu_m = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \qquad \Sigma_{mm} = \begin{bmatrix} 8 & 4 \\ 4 & 3 \end{bmatrix}$$

By definition of covariance -

$$\Sigma = E[xx^T] - E[x]E[x^T]$$

$$\Rightarrow \Sigma = E[xx^T] - \mu\mu^T$$

$$\Rightarrow E[xx^T] = \Sigma + \mu\mu^T$$

$$\sum_x (xx^T) = N(\Sigma_x + \mu_x\mu_x^T) \quad - \quad ①$$

Using equation 1

(a) $\sum_r (xx^T) = ?$

$$\sum_r (xx^T) = 4\left(\Sigma_{rr} + \mu_r \mu_r^T\right)$$

$$= 4\left(\begin{bmatrix} 5 & 3 \\ 3 & 7 \end{bmatrix} + \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 36 & 36 \\ 36 & 64 \end{bmatrix} ,,$$

(b) $\sum_m (xx^T) = ?$

$$\sum_m (xx^T) = 6\left(\Sigma_{mm} + \mu_m \mu_m^T\right) = 6\left(\begin{bmatrix} 8 & 4 \\ 4 & 3 \end{bmatrix} + \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 72 & 0 \\ 0 & 42 \end{bmatrix} ,,$$

(c) $\mu_{r+m} = ?$; $\mu_{r+m} = \frac{1}{N_{r+m}}\left((x)_r + (x)_m\right)$

where $(x)_r = N_r \mu_r$, $(x)_m = N_m \mu_m$

$$\Rightarrow \mu_{r+m} = \frac{1}{N_{r+m}}\left(N_r \mu_r + N_m \mu_m\right)$$

$$= \frac{1}{10}\left(4\begin{bmatrix} 2 \\ 3 \end{bmatrix} + 6\begin{bmatrix} -2 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -0.4 \\ 2.4 \end{bmatrix} ,,$$

(d) $\Sigma_{m+r} = ?$, $\Sigma_{m+r} = E\left[(xx^T)\right]_{m+r} - \mu_{m+r}\mu_{m+r}^T$ (by the def. of covariance)

$$E\left[(xx^T)\right]_{m+r} = \frac{1}{N_{m+r}}\left((xx^T)_m + (xx^T)_r\right) = \frac{1}{6+4}\left(\begin{bmatrix} 36 & 36 \\ 36 & 64 \end{bmatrix} + \begin{bmatrix} 72 & 0 \\ 0 & 42 \end{bmatrix}\right)$$

$$= \frac{1}{10}\begin{bmatrix} 108 & 36 \\ 36 & 106 \end{bmatrix}$$

$$E\left[(xx^T)\right]_{m+r} = \begin{bmatrix} 10.8 & 3.6 \\ 3.6 & 10.6 \end{bmatrix}$$

$$\Rightarrow \quad \Sigma_{m+r} = \begin{bmatrix} 10.8 & 3.6 \\ 3.6 & 10.6 \end{bmatrix} - \begin{bmatrix} 0.16 & -0.96 \\ -0.96 & 5.76 \end{bmatrix}$$

$$\Sigma_{m+r} = \begin{bmatrix} 10.64 & 4.56 \\ 4.56 & 4.84 \end{bmatrix} //$$

D. (a) Given: A, B are independent events. $\therefore$ $P(A\cap B) = P(A)P(B)$

Since, $P(A|B) = P(A)$ if A is not dependent on B

$$\Rightarrow P(A|B) = P(A) = \frac{P(A\cap B)}{P(B)}$$

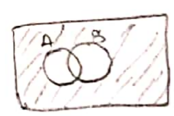
$$\Rightarrow \left[P(A\cap B) = P(A)\,P(B)\right]$$

To prove: A^c and B^c are also independent

$$\Rightarrow P(A^c\cap B^c) = P(A^c)\,P(B^c)$$

Proof:

$$P(A^c\cap B^c) = P(A\cup B)^c$$



$$= 1 - P(A\cup B)$$

$$= 1 - \left[P(A) + P(B) - P(A\cap B)\right]$$

Sub, $P(A) = 1 - P(A^c)$, $P(B) = 1 - P(B^c)$, $P(A\cap B) = P(A)P(B)$

$$\Rightarrow P(A^c\cap B^c) = 1 - \left[1 - P(A^c) + 1 - P(B^c) - \left[(1 - P(A^c))(1 - P(B^c))\right]\right]$$

$$= 1 - \left[1 - \cancel{P(A^c)} \cancel{+1} - \cancel{P(B^c)} \cancel{-1} + \cancel{P(A^c)} + \cancel{P(B^c)} - P(A^c)P(B^c)\right]$$

$$= 1 - \left[1 - P(A^c)P(B^c)\right]$$

$$\left[P(A^c\cap B^c) = P(A^c)\,P(B^c)\right] \quad \square$$

(b.) <u>Given:</u> $X \sim N(\mu_x, \Sigma_x)$, $Y = AX + b$

<u>To prove:</u> $Y \sim N(\mu_y, \Sigma_y)$

<u>Proof</u> (for μ_y)

$$Y = AX + b$$

$$E[Y] = E[AX + b]$$

$$= A\,E[X] + b$$

$$\Rightarrow \mu_y = A\mu_x + b$$

(for Σ_y)

$$\Sigma_y \triangleq E\left[(Y - \mu_y)(Y - \mu_y)^T\right]$$

$$= E\left\{\left[(AX + b) - (A\mu_x + b)\right]\left[(AX + b) - (A\mu_x + b)\right]^T\right\}$$

$$= E\left\{\left[A(x - \mu_x)\right]\left[A(x - \mu_x)\right]^T\right\}$$

$$= E\left\{A(x - \mu_x)(x - \mu_x)^T A^T\right\}$$

$$= A\,E\left\{(x - \mu_x)(x - \mu_x)^T\right\} A^T$$

$$= A\Sigma_x A^T$$

$$\Rightarrow \Sigma_y = A\Sigma_x A^T$$

$$\Rightarrow Y \sim N(A\mu_x + b, A\Sigma_x A^T) \quad \square$$

———— X ————

Task 2

A. Given → Cancer : C , Positive result : P

$$P(C) = 0.01 \quad , \quad P(\neg C) = 0.99$$

$P(P|\neg C)$ = false positive = 20% false negative = 10% = $P(\neg P|C)$

	P	¬P
C	0.9	0.1
¬C	0.2	0.8

to find $P(C|P) = ?$

$$\therefore \quad P(C|P) = \frac{P(P|C)\,P(C)}{P(P)} \qquad \text{(By Bayes' Theorem)}$$

$$= \frac{P(P|C)\,P(C)}{P(P|C)\,P(C) + P(P|\neg C)\,P(\neg C)}$$

$$= \frac{0.9 \times 0.1}{0.9 \times 0.1 + 0.2 \times 0.99} = 0.0434 \approx \underline{4.34\%}$$

Reducing the rate of false negatives is more beneficial in this case, as the damage from not knowing that a patient has cancer is more severe than falsely believing that they have cancer.

BUT, analytically the results show otherwise-

Case 1: Reducing false negative by 7% → 10% to 3%.

$$\therefore \quad P(C|P) = \frac{0.97 \times 0.01}{0.97 \times 0.01 + 0.2 \times 0.99} = 0.0467 = 4.67\%$$

<u>Case 2</u>: Reducing false positive by 7% → 20% to 13%

$$P(C/P) = \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.13 \times 0.99} = 0.0653 = 6.53\%$$

It is clear from the computations that given this data, reducing the rate of false positives offer the most bang for the buck.

——— × ———

B. "Bayes Filter"

action(u) := paint , state (x) := Coloured, blank , Sensor (z) := coloured, blank.

C: coloured
b: blank
p: paint

<u>Given</u>: Sensor model -

$$p(z=c, x=b) = 0.2$$
$$p(z=b, x=b) = 1-0.2 = 0.8$$
$$p(z=c, x=c) = 0.7$$
$$p(z=b, x=c) = 1-0.7 = 0.3$$

Action model -

$$p(x_{t+1}=c \mid x_t=b, u_{t+1}=p) = 0.9$$
$$p(x_{t+1}=b \mid x_t=b, u_{t+1}=p) = 1-0.9 = 0.1$$
$$p(x_{t+1}=c \mid x_t=c, u_{t+1}=p) = 1$$
$$p(x_{t+1}=b \mid x_t=c, u_{t+1}=p) = 0$$

} intuitively

Initial state -

$$p(x_0=c) = p(x_0=b) = 0.5$$

To find - $p(x_t=b \mid u_t=P, z_t=c) = ?$

formula used - $p(x_t \mid u_t, z_t) = \eta \, p(z_t \mid x_t) \sum p(x_t / u_t, x_{t-1}) P(x_{t-1})$

$$\Rightarrow P(x_t = c / u_t = P, z_t = c) = \eta\, P(z_t = c / x_t = c)\Big[P(x_t = c / u_t = P, x_{t-1} = c)\, P(x_{t-1} = c)$$

$$+ P(x_t = c / u_t = P, x_{t-1} = b)\, P(x_{t-1} = b)\Big] \quad \text{—— (1)}$$

$$= \eta\, 0.7\big[1(0.5) + 0.9(0.5)\big] = \eta\, 0.7 \times 0.95$$

$$= \eta\, 0.665$$

$$\Rightarrow P(x_t = b / u_t = P, z_t = c) = \eta\, P(z_t = c / x_t = b)\Big[P(x_t = b / u_t = P, x_{t-1} = c)\, P(x_{t-1} = c)$$

$$+ P(x_t = b / u_t = P, x_{t-1} = b)\, P(x_{t-1} = b)\Big] \quad \text{—— (2)}$$

$$= \eta\, 0.2\big[0(0.5) + 0.1 \times 0.5\big] = \eta\, 0.2 \times 0.05$$

$$= \eta\, 0.01$$

$$(1) + (2) = 1$$

$$\Rightarrow \eta(0.665 + 0.01) = 1$$

$$\Rightarrow \eta = \underline{1.48}$$

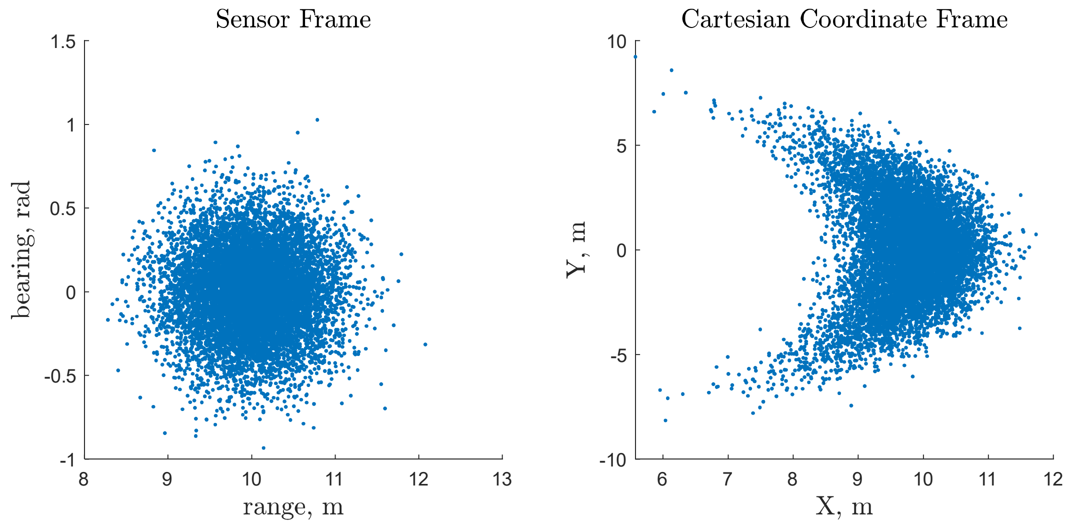
Substituting '$\eta = 1.48$' in (2)

$$\Rightarrow P(x_t = b / u_t = P, z_t = c) = \eta\, 0.01 = 0.0148 = \underline{\underline{1.48\%}}$$

—— x ——

Task 3

A.

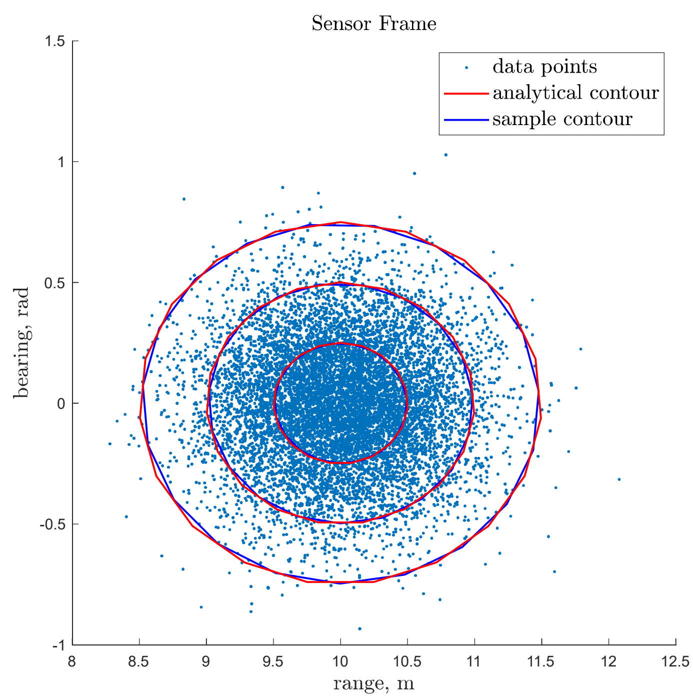


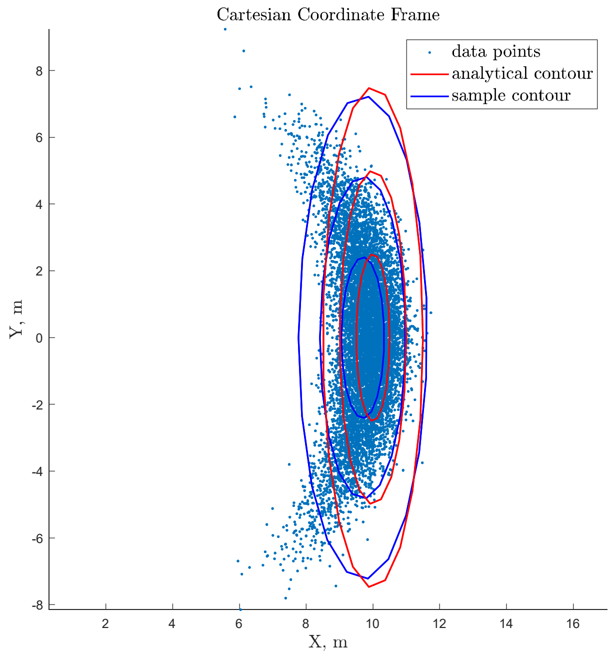
B.

$$Covariance\_cartesian = \begin{bmatrix} 0.25 & 0 \\ 0 & 6.25 \end{bmatrix}$$

C.

It is observed that when the analytical contour and the sample-based contours are superimposed for the sensor frame data points they match with a high certainty as can be confirmed by the Mahalanobis distance in part D. But the same is not true for the points in the cartesian coordinate system. The sample-based contours and the analytical contours fails to match because the calculated points for the cartesian coordinates are an approximation computed by linearizing the non-linear polar system.





D.

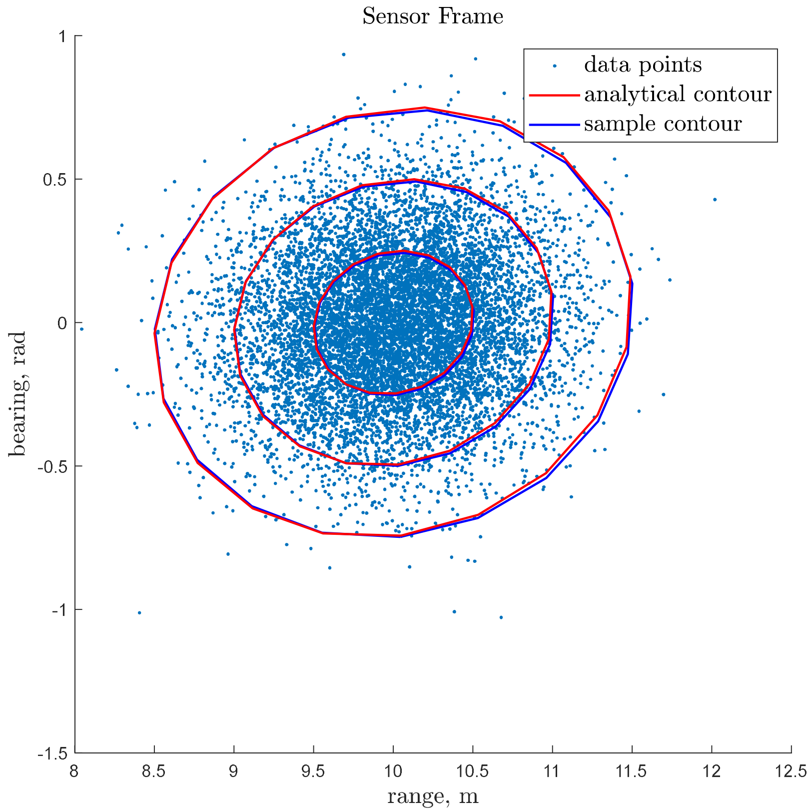
K	Percentage Samples			Mahalanobis Distance		
	Sensor Frame	Cartesian Frame	True Gaussian	Sensor Frame	Cartesian Frame	True Gaussian
1-sigma	39.72	38.80	39.35	1.02	1.09	1
2-sigma	86.62	79.75	86.47	4.07	3.61	4
3-sigma	98.98	93.06	98.89	9.16	7.78	9

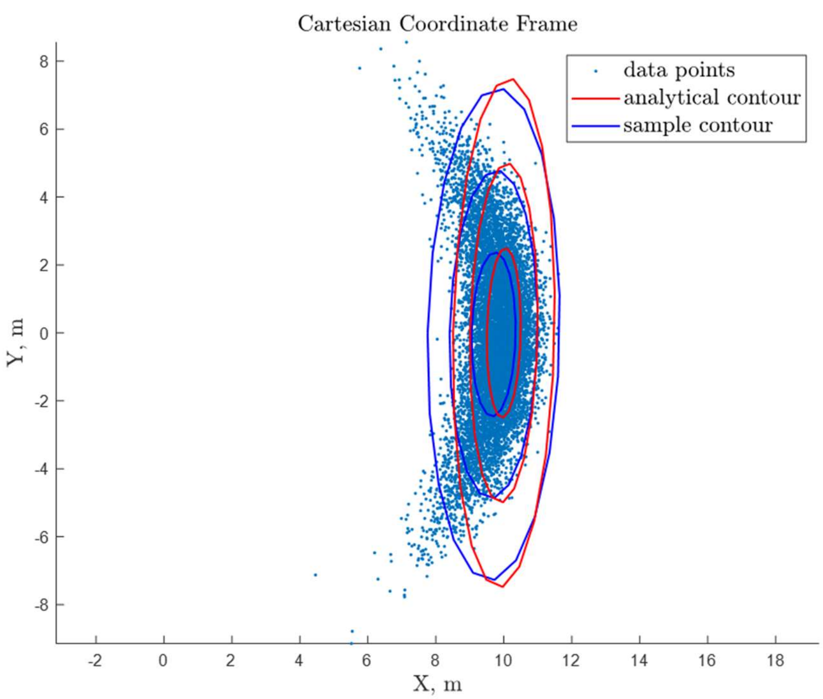
E. The table below shows the trend in the counts when the noise parameters, i.e. the standard deviation of the range and bearing are changed. It is seen that varying the bearing noise makes the counts come closer to the theoretical values, which is due to the fact that the bearing angle is the main source of non-linearity in the system. So, when the noise in bearing in decreased the accuracy of approximation in linearization increases. And, as expected decreasing the noise in both range and bearing produces the most favorable results. This proves that these errors play a major role when a system is linearized using these gaussian parameters, which only bloats while being used for assumptions for linearization.

K	sigma_r = 0.5 sigma_b = 0.25	sigma_r = 0.5 sigma_b = 0.01	sigma_r = 0.01 sigma_b = 0.25	sigma_r = 0.01 sigma_b = 0.01	True Gaussian
1-sigma	38.42	38.75	11.04	38.55	39.35
2-sigma	79.53	85.76	19.17	86.61	86.47
3-sigma	93.03	98.73	24.25	98.93	98.89

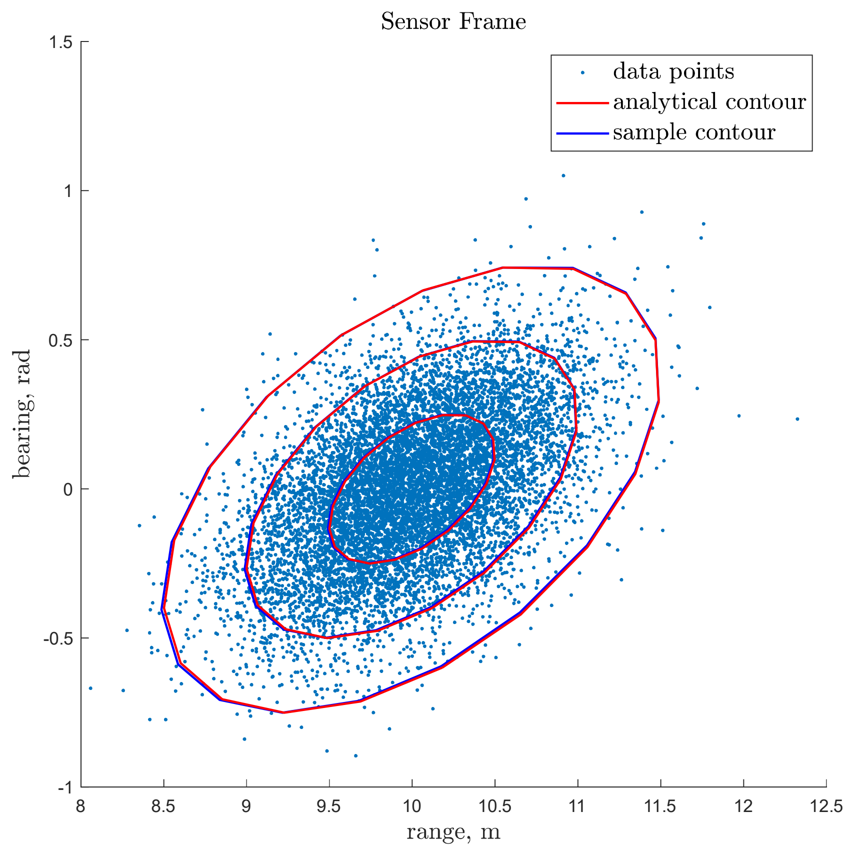
F.

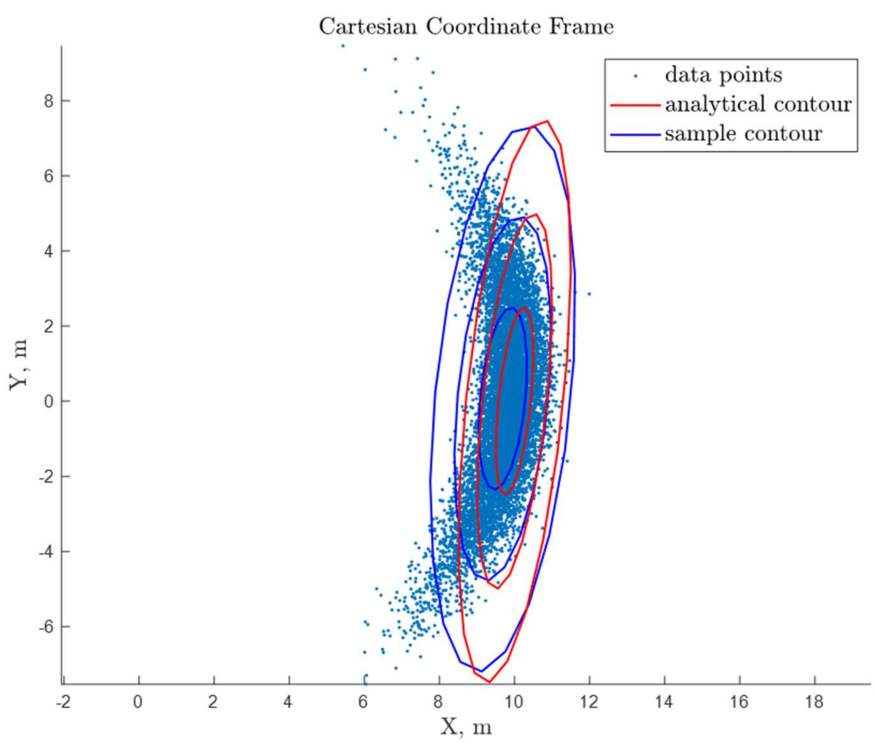
a. $\rho_{r\theta} = 0.1$



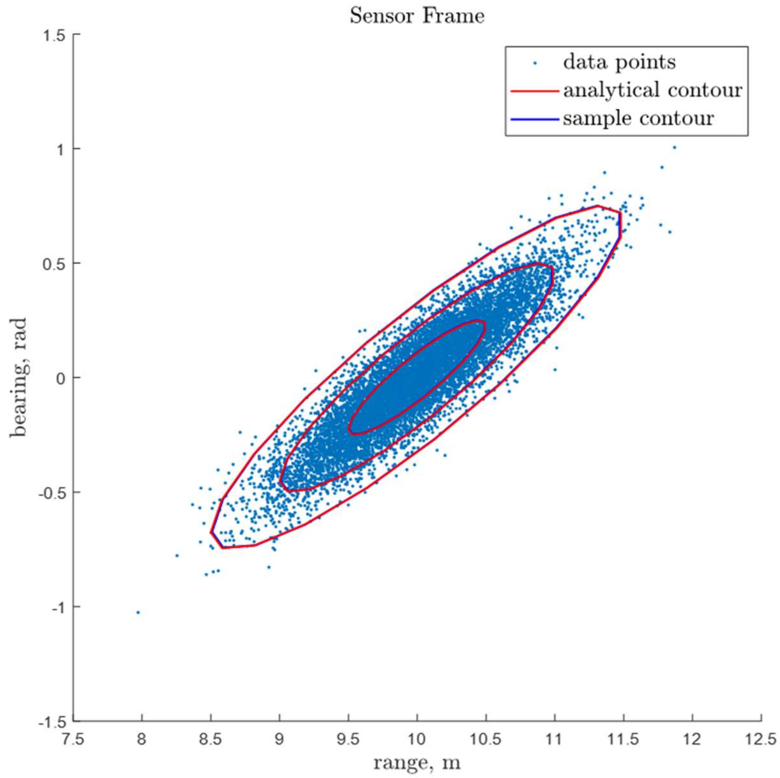


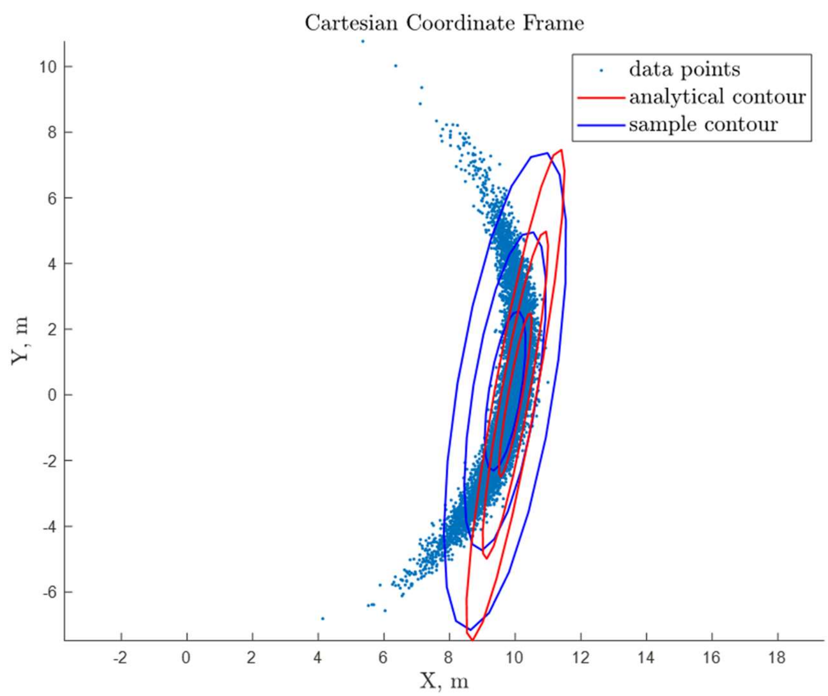
b. $\rho_{r\theta} = 0.5$





c. $\rho_{r\theta} = 0.9$





Task 4

A. Given: $\theta \sim N(\mu_0, \sigma_0^2)$, $p(x|\theta) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(\frac{-(\theta-\mu_x)^2}{2\sigma_x^2}\right)$ — (1)

To find: μ_θ for $\theta \sim N(\mu_\theta, \sigma_\theta^2)$ using MAP estimator

Proof:

$$p(\theta|x) = \frac{p(x|\theta)\,p(\theta)}{p(x)} \propto p(x|\theta)\,p(\theta)$$

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmin}} \; \underbrace{-\log(p(x|\theta)\,p(\theta))}_{\text{to minimize.}}$$

$$\Rightarrow \quad f: -\log(p(x|\theta)) - \log(p(\theta))$$

$$: -\log(\eta_1) + \frac{(\theta-\mu_x)^2}{2\sigma_x^2} - \log\eta_2 + \frac{(\theta-\mu_0)^2}{2\sigma_0^2} \qquad \text{using (1)}$$

(to find minimizer) $\frac{\partial f}{\partial \theta} = \frac{2\mu_0 - 2\theta}{2\sigma_0^2} + \frac{2\mu_x - 2\theta}{2\sigma_x^2} = 0$

Solving for $\theta \Rightarrow \quad \mu_\theta = \frac{\mu_0 \sigma_x^2 + \mu_x \sigma_0^2}{\sigma_x^2 + \sigma_0^2}$ //

——— x ———

B. To find: $\theta \sim N(\mu_\theta, \sigma_\theta^2)$

$$p(\theta|x) = \frac{p(x|\theta)\,p(\theta)}{p(x)} \propto p(x|\theta)\,p(\theta)$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left(\frac{-(\theta-\mu_x)^2}{2\sigma_x^2}\right) \cdot \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(\frac{-(\theta-\mu_0)^2}{2\sigma_0^2}\right)$$

$$= \eta \exp\left[\frac{-(\theta^2 + \mu_x^2 - 2\theta\mu_x)}{2\sigma_x^2} - \frac{(\theta^2 + \mu_0^2 - 2\theta\mu_0)}{2\sigma_0^2}\right]$$

$$\propto \exp\left\{\frac{-\theta^2\sigma_x^2 - \theta\sigma_0^2 + 2\theta\mu_x\sigma_0^2 + 2\theta\mu_0\sigma_x^2 - \mu_x^2\sigma_0^2 - \mu_0^2\sigma_x^2}{2\sigma_0^2\sigma_x^2}\right\}$$

$$\propto \exp\left\{\frac{-\theta^2(\sigma_x^2+\sigma_0^2) + 2\theta(\mu_x\sigma_0^2 + \mu_0\sigma_x^2) - (\mu_x^2\sigma_0^2 + \mu_0^2\sigma_x^2)}{2\sigma_0^2\sigma_x^2}\right\}$$

$$\propto \exp\left(\frac{-\theta^2 + 2\theta\left(\frac{\mu_x\sigma_0^2 + \mu_0\sigma_x^2}{\sigma_x^2+\sigma_0^2}\right) - \left(\frac{\mu_x + \sigma_0^2 + \mu_0\sigma_x^2}{\sigma_x^2+\sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma_x^2}{\sigma_0^2+\sigma_x^2}}\right)$$

$$\cdot \exp\left\{-\frac{(\mu_0^2\sigma_x^2 + \sigma_0^2\mu_x^2)}{2\sigma_0^2\sigma_x^2}\right\}$$

$$\propto \exp\left\{-\frac{\left(\theta - \left(\frac{\mu_x\sigma_0^2 + \mu_0\sigma_x^2}{\sigma_0^2+\sigma_x^2}\right)\right)^2}{2\cdot\frac{\sigma_x^2\sigma_0^2}{\sigma_x^2+\sigma_0^2}}\right\}$$

$$\therefore \mu_\theta = \frac{\mu_x\sigma_0^2 + \mu_0\sigma_x^2}{\sigma_0^2+\sigma_x^2}, \quad \sigma_\theta^2 = \frac{\sigma_x^2\sigma_0^2}{\sigma_x^2+\sigma_0^2}$$

$$\therefore \theta|x \sim N(\mu_\theta, \sigma_\theta^2) \qquad (\text{for sample size } 1)$$

(for sample size n) $\longrightarrow$

$$p(\theta|x) \propto p(\theta)p(x|\theta) = p(\theta)\,p(x_1|\theta)\,p(x_2|\theta)\ldots p(x_n|\theta)$$

$$= \frac{1}{\sigma_0^2\sqrt{2\pi}}\exp\left\{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right\} \times \prod_{i=1}^{n}\frac{1}{\sigma_x^2\sqrt{2\pi}}\exp\left\{\frac{-(\theta-\mu_{x_i})^2}{2\sigma_x^2}\right\}$$

$$= \text{Const.} \exp\left\{ \frac{-\theta^2 + 2\theta\mu_0 - \mu_0^2}{2\sigma_0^2} + \sum_{i=1}^{n} \frac{-\theta^2 + 2\theta\mu_{x_i} - \mu_{x_i}^2}{2\sigma_x^2} \right\}$$ (4.3)

$$\propto \exp\left\{ - \frac{\left(\mu - \left(\frac{\mu_0\sigma_x^2 + \sum_{i=1}^{n} \sigma_0^2 \mu_{x_i}}{\sigma_x^2 + n\sigma_0^2}\right)\right)^2}{2 \frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + n\sigma_0^2}} \right\}$$

$$\therefore \quad \mu_\theta = \frac{\mu_0\sigma_x^2 + \sum_{i=1}^{n} \sigma_0^2 \mu_{x_i}}{\sigma_x^2 + n\sigma_0^2} \; ; \quad \sigma_\theta^2 = \frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + n\sigma_0^2}$$

———— X ————