PARAMETER AND STATE ESTIMATION

ASSIGNMENT 1

G PRASHANT (BS17B011)

G. PRASHANT
BS17B011

CH5115: PARAMETER AND STATE ESTIMATION

ASSIGNMENT 1

Question 1

Part (a)

Given:

$$f(x,y) = \begin{cases} K\dfrac{e^{-x/y}e^{-y}}{y} & ; \quad x>0, y>0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

(i) As $f(x,y)$ is a probability density function, it follows the property,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

$$\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{0}^{\infty}\int_{0}^{\infty} K\frac{e^{-x/y}e^{-y}}{y}\,dx\,dy = 1$$

[As $f(x,y)=0$ when $x,y\leq 0$]

$$\Rightarrow K\int_{0}^{\infty}\frac{e^{-y}}{y}\int_{0}^{\infty} e^{-x/y}\,dx\,dy = K\int_{0}^{+\infty} -e^{-y}\left[e^{-x/y}\right]_{0}^{\infty} dy = 1$$

$$\Rightarrow K\int_{0}^{\infty} e^{-y}\,dy = 1$$

Hence,

$$\boxed{K = 1}$$

(ii) Marginal density of Y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

$$= \int_0^{\infty} \frac{e^{-x/y} e^{-y}}{y}\,dx$$

$$= \frac{e^{-y}}{y}\left(\frac{-y}{1}\right) \cdot e^{-x/y}\Big|_0^{\infty}$$

$$= e^{-y}$$

Hence, $f_Y(y) = e^{-y}$, when $y>0$

More formally,

$$\boxed{f_Y(y) = \begin{cases} e^{-y} & , y>0 \\ 0 & , \text{otherwise} \end{cases}}$$

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

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

$$= \int_{0.2}^{0.4} -e^{-y}\left[e^{-x/y}\right]_0^1 dy = \int_{0.2}^{0.4} -e^{-y}(e^{-1/y}-1)\,dy$$

$$= \int_{0.2}^{0.4} e^{-y}\,dy - \int_{0.2}^{0.4} e^{-y} e^{-1/y}\,dy$$

$$= -e^{-y}\Big|_{0.2}^{0.4} - \int_{0.2}^{0.4} e^{-y} e^{-1/y}\,dy$$

$$= 0.14841 - \int_{0.2}^{0.4} e^{-y} e^{-1/y}\,dy$$

Integrating numerically using "integral" in MATLB, we get

$$\int_{0.2}^{0.4} e^{-y} e^{-1/y} dy = 0.00554$$

Hence, $\boxed{\begin{aligned} Pr(0<X<1, 0.2<Y<0.4) &= 0.14841 - 0.00554 \\ &= 0.14287 \end{aligned}}$

iv) Conditional expectation

$$E(X/Y) = \int_{-\infty}^{\infty} x f(x|y)\, dx$$

$$f(x|y) = \frac{f(x,y)}{f(y)} = \begin{cases} \frac{e^{-x/y}}{y} & , x,y>0 \\ 0 & , \text{otherwise} \end{cases}$$

$$\Rightarrow E(X/Y=y) = \int_{0}^{\infty} \frac{x e^{-x/y}}{y} dx$$

$$= \frac{1}{y}\left[-xy\, e^{-x/y} + y\int e^{-x/y} dx\right]_0^{\infty}$$

$$= \frac{1}{y}\left[-xy e^{-x/y} - y^2 e^{-x/y}\right]_0^{\infty}$$

$$= \left[-e^{-x/y}(x+y)\right]_0^{\infty}$$

$$= y$$

Hence, $\boxed{E(X/Y=y) = y}$

Part (b)

The joint Gaussian distribution of two random variables X and Y ($Z = [x \quad y]$) can be expressed as:

$$f_{X,Y}(x,y) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(Z-\mu)\cdot\Sigma^{-1}\cdot(Z-\mu)^T\right)$$

where $\mu = [\mu_x \quad \mu_y] \rightarrow$ mean vector

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \rightarrow \text{covariance matrix.}$$

$\rho \rightarrow$ correlation coefficient.

Expanding the equation,

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

Let $Q(x,y) = \left(\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\left(\frac{-1}{2(1-\rho^2)}\right)$

Adding and subtracting $\frac{\rho^2(x-\mu_x)^2}{\sigma_x^2}$,

$$Q(x,y) = -\left[\frac{\left(\frac{y-\mu_y}{\sigma_y} - \rho\frac{(x-\mu_x)}{\sigma_x}\right)^2}{2(1-\rho^2)} + \frac{(x-\mu_x)^2}{2\sigma_x^2}\right]$$

$$= -\left[\frac{\left(y-\left(\mu_y + \frac{\sigma_y}{\sigma_x}\rho(x-\mu_x)\right)\right)^2}{2\sigma_y^2(1-\rho^2)} + \frac{(x-\mu_x)^2}{2\sigma_x^2}\right]$$

Let $\tilde{\mu}_y(x) = \mu_y + \frac{\sigma_y}{\sigma_x}\rho(x-\mu_x)$

$\tilde{\sigma}_y = \sigma_y\sqrt{1-\rho^2}$

This gives

$$f(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma_X} \exp\left(\frac{-1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\right) \cdot \frac{1}{\sqrt{2\pi}\,\tilde{\sigma}_Y} \exp\left(\frac{-1}{2}\frac{(y-\tilde{\mu}_Y)^2}{\tilde{\sigma}_Y^2}\right)$$

Marginal distribution

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

$$= N(\mu_X, \sigma_X^2) \underbrace{\int_{-\infty}^{\infty} N(\tilde{\mu}_Y, \tilde{\sigma}_Y^2)\,dy}_{=1 \text{ (property of density functions)}}$$

Hence

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_X} \exp\left(\frac{-1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\right)$$

Now,

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{1}{\sqrt{2\pi}\,\tilde{\sigma}_Y} \exp\left(\frac{-1}{2}\frac{(y-\tilde{\mu}_Y)^2}{\tilde{\sigma}_Y^2}\right)$$

$$E[Y|X] = \text{mean of } f(y|x) = \tilde{\mu}_Y$$

which gives

$$\boxed{E[Y|X] = \mu_Y + \frac{\sigma_Y}{\sigma_X} e (x-\mu_X)}$$

Therefore, $E[Y|X]$ is a linear function of X with slope $\frac{\sigma_Y}{\sigma_X} e$ and intercept $\mu_Y - \frac{\sigma_Y e}{\sigma_X}\mu_X$.

Question 2

Given $X \sim N(1,2)$ and $Y = 3X^2 + 5X$

For the first part, we sample 100 points from the distribution of X and generate Y accordingly.

We then calculate/estimate the sample covariance matrix, given by:

$$\hat{\sigma}_{yx} = \frac{1}{N} \sum_{k=1}^{N} (y[k] - \bar{y})(x[k] - \bar{x})$$

By implementing a MATLAB function to ~~cal~~ estimate covariance matrix based on the above formula, we get

$$\hat{\sigma} = \begin{array}{cc} & \begin{array}{cc} x & y \end{array} \\ \begin{array}{c} x \\ y \end{array} & \begin{bmatrix} 4.47 & 48.1551 \\ 48.1551 & 926.4462 \end{bmatrix} \end{array}$$

Hence $\hat{\sigma}_{xx} = 4.47$, $\hat{\sigma}_{yy} = 926.4462$,

$\hat{\sigma}_{xy} = \hat{\sigma}_{yx} = 48.1551$

By estimating sample covariance matrix using cov function in MATLAB, we get.

$$\underset{(\text{MATLAB})}{\hat{\sigma}} = \begin{bmatrix} 4.5151 & 48.6415 \\ 48.6415 & 935.8042 \end{bmatrix}$$

It can be noticed that the values in the covariances matrices obtained by both methods did not exactly match. This is because MATLAB's implementation of cov function accounts for Bessel's correction, which uses $N-1$ instead of N in the denominator of the formula for calculating covariance of two random variables. Bessel's correction is known to correct bias in the estimation of ~~covar~~ sample covariance.

The formula after ~~Bessel~~ Bessel's correction is

$$\hat{\sigma}_{yx} = \frac{1}{N-1}\sum_{k=1}^{N}(y[k]-\bar{y})(x[k]-\bar{x})$$

To calculate the theoretical value of $\hat{\sigma}_{yx}$, we need to determine

$$\sigma_{yx} = E[YX] - E[Y]E[X]$$

$$= E[3X^3+5X^2] - E[3X^2+5X]\mu_x$$

$$= 3E[X^3]+5E[X^2] - (3E[X^2]+5E[X])(1)$$

$$= 3E[X^3]+2E[X^2]-5 \qquad \text{(As } E[X]=\mu_x=1)$$

$$= 3E[X^3]+2(Var(X)+\mu_x^2)-5$$

$$= 3E[X^3]+2(2+1)-5$$

$$= 3E[X^3]+5$$

$$E[X^3] = \int_{-\infty}^{+\infty} x^3 f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x^3 \left(\frac{1}{\sqrt{2\pi}(2)} e^{-\frac{(x-1)^2}{2(2)^2}} \right) dx$$

$$= 13 \qquad \text{(Numerical integration)}$$

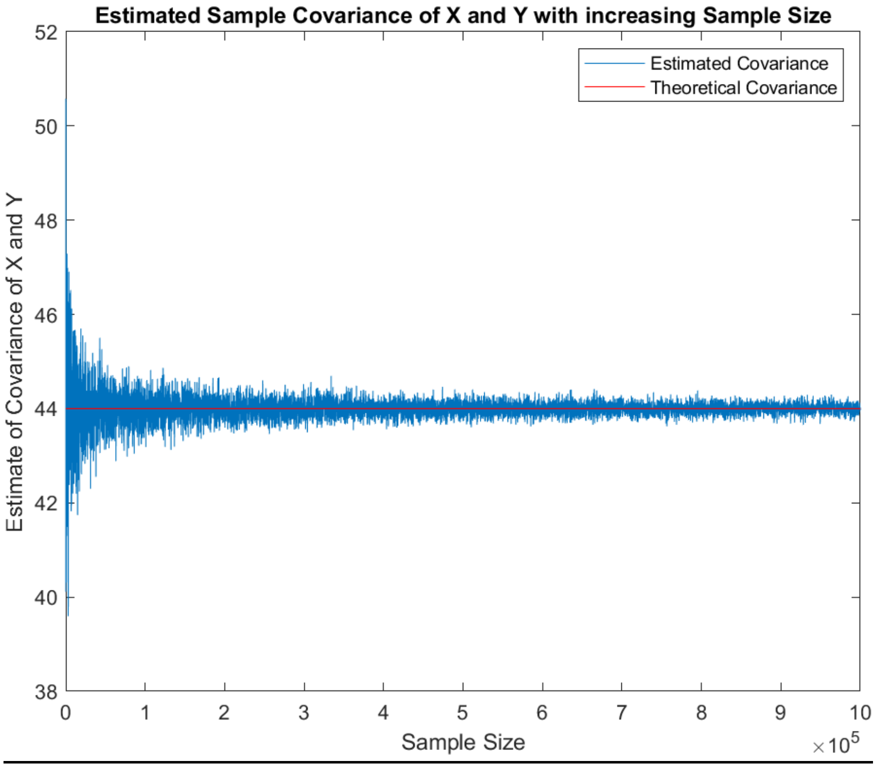
Hence,

$$\boxed{\sigma_{yx} = 3(13)+5 = 44}$$

Using Simulation to show the convergence of Covariance

The sample covariances of Random Variables X and Y were estimated with varying sample sizes N, from 10^2 to 10^6, in steps of 100. As calculated in the previous page, the theoretical value of the covariance is 44.

The plot of the estimated sample covariance against the sample size, along with the marked theoretical value is shown below:



From the plot, it can be observed that as the sample size increases, the sample covariance estimate (blue) of random variables X and Y tends to the theoretical value of 44 (red line), however with some fluctuations. Hence, as we keep increasing N, or when $N \to \infty$, $\widehat{\sigma_{yx}} \to 44$.

Question 3

Part (a)

Given:

$$\Sigma = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \sigma_{x_1x_3} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 & \sigma_{x_2x_3} \\ \sigma_{x_1x_3} & \sigma_{x_2x_3} & \sigma_{x_3}^2 \end{bmatrix} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix}$$

where $\sigma_{x_i}^2 \rightarrow$ variance of X_i

$\sigma_{x_ix_j} \rightarrow$ covariance of X_i and X_j, $i \neq j$

Correlation matrix $\rho_{ij} = \dfrac{\Sigma_{ij}}{\sigma_{x_i}\sigma_{x_j}}$

$$\sigma_{x_1} = \sqrt{4} = 2$$
$$\sigma_{x_2} = \sqrt{9} = 3$$
$$\sigma_{x_3} = \sqrt{25} = 5$$

Hence,

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

Part (b)

$$\text{Cov}\left(X_1, \tfrac{1}{2}X_2 + \tfrac{1}{2}X_3\right) = E\left[X_1\left(\frac{X_2}{2} + \frac{X_3}{2}\right)\right] - \bar{X}_1 E\left(\frac{X_2}{2} + \frac{X_3}{2}\right)$$

$$= \frac{1}{2}\left[E[X_1X_2 + X_1X_3] - \bar{X}_1(\bar{X}_2 + \bar{X}_3)\right)$$

$$= \frac{1}{2}\left(E[X_1X_2] - \bar{X}_1\bar{X}_2 + E[X_1X_3] - \bar{X}_1\bar{X}_3\right)$$

$$= \frac{1}{2}\left(\sigma_{x_1 x_2} + \sigma_{x_1 x_3}\right)$$

$$= \frac{1}{2}(1+2) = \frac{3}{2} = 1.5$$

$$Cov\left(X_1, \tfrac{1}{2}X_2 + \tfrac{1}{2}X_3\right) = 1.5$$

$$Var\left(\frac{X_2}{2} + \frac{X_3}{2}\right) = E\left[\left(\frac{X_2}{2}+\frac{X_3}{2}\right)^2\right] - E\left[\frac{X_2}{2} + \frac{X_3}{2}\right]^2$$

$$= E\left[\frac{X_2^2}{4} + \frac{X_3^2}{4} + \frac{X_2X_3}{2}\right] - \left(E\left[\frac{X_2}{2}\right]^2 + E\left[\frac{X_3}{2}\right]^2 + 2E\left[\frac{X_2X_3}{4}\right]\right)$$

$$= \frac{1}{4}\left(E[(X_2^2)] - E[X_2]^2\right) + \frac{1}{4}\left(E[X_3^2] - E[X_3]^2\right)$$

$$= \frac{\sigma_{x_2}^2 + \sigma_{x_3}^2}{4}$$

$$std\left(\frac{X_2}{2} + \frac{X_3}{2}\right) = \sigma_{\frac{X_2}{2}+\frac{X_3}{2}} = \sqrt{\frac{\sigma_{x_2}^2 + \sigma_{x_3}^2}{4}} = \frac{\sqrt{34}}{2}$$

$$Corr\left(X_1, \frac{X_2}{2} + \frac{X_3}{2}\right) = \frac{Cov\left(X_1; \frac{X_2}{2} + \frac{X_3}{2}\right)}{\sigma_{X_1}\, \sigma_{\frac{X_2}{2}+\frac{X_3}{2}}}$$

$$= \frac{1.5}{2\cdot\sqrt{34}/2} = \frac{1.5}{\sqrt{34}} \simeq 0.2572$$

$\boxed{\text{Hence, the correlation between } X_1 \text{ and } \tfrac{1}{2}X_2 + \tfrac{1}{2}X_3 \text{ is } 0.2572}$

Question 4

Part (a)

A χ^2 distribution is given by:

$$\chi^2(K) = f(x;K) = \begin{cases} \dfrac{x^{K/2-1} \cdot e^{-x/2}}{2^{K/2} \cdot \Gamma(K/2)}, & \text{when } x>0 \\ 0, & \text{otherwise} \end{cases}$$

Given $K=10$, the optimal MAE predictor of $X \sim \chi^2(10)$ is given by:

$$X^* = \underset{\hat{X}}{\operatorname{argmin}}\, E[|X-\hat{X}|]$$

The expression of $E[|X-\hat{X}|]$ is given by:

$$E[|X-\hat{X}|] = \int_0^\infty |x-\hat{x}| f(x)\,dx \quad [\text{As } f(x)=0 \text{ when } x\leq 0]$$

$$= \int_0^{\hat{x}} (\hat{x}-x) f(x)\,dx + \int_{\hat{x}}^{\infty} (x-\hat{x}) f(x)\,dx$$

$$= \hat{x}\int_0^{\hat{x}} f(x)\,dx - \hat{x}\int_{\hat{x}}^{\infty} f(x)\,dx - \int_0^{\hat{x}} x f(x)\,dx + \int_{\hat{x}}^{\infty} x f(x)\,dx$$

$$= \hat{x}F(\hat{x}) - \hat{x}(1-F(\hat{x})) - \int_0^{\hat{x}} x f(x)\,dx + \int_{\hat{x}}^{\infty} x f(x)\,dx$$

where $F(x)$ is the cumulative distribution function (CDF) of χ^2 distribution.

Hence, by making use of chi2pdf and chi2cdf functions in MATLAB, we implement "fminsearch" to minimize

$$E[|X-\hat{x}|] = 2\hat{x}F(\hat{x}) - \hat{x} - \int_0^{\hat{x}} x f(x)\,dx + \int_{\hat{x}}^{\infty} x f(x)\,dx.$$

to determine the optimal value ~~of~~ $\hat{x} = x^*$, and the expected absolute error at x^*.

Results:

$x^* = 9.3418$

$E[|X-x^*|] = 3.4698$

We also notice that,

$F(x^*) = 0.5$,

implying x^* is the median of the distribution $f(x)$.

Part (b)

$$\Pr(0.9x^* < X < 1.1x^*) = F(1.1x^*) - F(0.9x^*)$$
$$= F(10.27598) - F(8.40762)$$
$$= 0.5834 - 0.4109 = 0.1725$$

For a χ^2 distribution with degrees of freedom K,

$$\mu_x = E[X] = \int_0^{\infty} x\, f(x)dx = K = 10$$

$\Rightarrow \mu_x = 10.$

$$Pr(0.9\mu_x < X < 1.1\mu_x) = F(1.1\mu_x) - F(0.9\mu_x)$$
$$= F(11) - F(9)$$
$$= 0.6425 - 0.4679$$
$$= 0.1746$$

Hence, it can be observed that

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

making $P(0.9X^* < X < 1.1X^*)$ lower.

The χ^2 distribution is a positively skewed density function with median $<$ mean, or $X^* < \mu_x$. Hence, the region in the neighborhood of the mean μ_x covers more area than the region in the neighborhood of the median, given that $(1.1\mu_x - 0.9\mu_x) > (1.1X^* - 0.9X^*)$, i.e., the interval size is higher when the mean.

Moreover, with the degrees of freedom $K = 10$, the χ^2 distribution decays at a very gentle rate, and a broader area is covered in the interval $(0.9\mu_x, 1.1\mu_x)$ when compared to $(0.9X^*, 1.1X^*)$.

Due to these reasons, $Pr(0.9X^* < X < 1.1X^*)$ is lower than $Pr(0.9\mu_x < X < 1.1\mu_x)$