Utah State University
ECE 6010
Stochastic Processes
Homework # 4 Solutions

1. Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

   (a) Show that $E[\mathbf{X}] = \boldsymbol{\mu}$ and $\text{cov}(\mathbf{X}, \mathbf{X}) = \Sigma$.
   Using characteristic functions:

   $$\phi_X(\mathbf{u}) = \exp[i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}^T\Sigma\mathbf{u}]$$

   Taking the gradient with respect to $\mathbf{u}$ we have

   $$\frac{\partial \phi_X(\mathbf{u})}{\partial \mathbf{u}} = \exp[i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}](i\boldsymbol{\mu} - \Sigma\mathbf{u})$$

   Now evaluating at $\mathbf{u} = \mathbf{0}$ and dividing by i we obtain

   $$E[X] = \boldsymbol{\mu}.$$

   Taking the gradient again with respect to $\mathbf{u}$ we have

   $$\exp[i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}](-\Sigma) + \exp[i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}](i\boldsymbol{\mu} - \Sigma\mathbf{u})(i\boldsymbol{\mu} - \Sigma\mathbf{u})^T$$

   and evaluating at $\mathbf{u} = \mathbf{0}$ we have

   $$-\Sigma - \boldsymbol{\mu}\boldsymbol{\mu}$$

   Dividing by i^{-2} we obtain $E[X^2] = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T$, from which it follows that $\text{cov}(\mathbf{X}, \mathbf{X}) = \Sigma$.

   (b) Show that $A\mathbf{X} + \mathbf{b} \sim \mathcal{N}(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T)$.
   We note that $\phi_X(\mathbf{u}) = E[e^{i\mathbf{u}^T\mathbf{X}}] = \exp[i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}]$. Then

   $$\phi_Y(\mathbf{u}) = E[e^{i\mathbf{u}^T\mathbf{Y}}] = E[e^{i\mathbf{u}^T(A\mathbf{X}+\mathbf{b})}] = e^{i\mathbf{u}^T\mathbf{b}}E[e^{i\mathbf{u}^T A\mathbf{X}}]$$

   But we can compute the expected value using what we know from $\phi_X(\mathbf{u})$:

   $$E[e^{i\mathbf{u}^T A\mathbf{X}}] = \phi_X(\mathbf{v})|_{\mathbf{v}=A^T\mathbf{u}} = \exp[i\mathbf{u}^T A\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}^T A\Sigma A^T\mathbf{u}]$$

So

$$\phi_Y(\mathbf{u}) = e^{i\mathbf{u}^T\mathbf{b}} \exp[i\mathbf{u}^T A\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}^T A\Sigma A^T\mathbf{u}] = \exp[i\mathbf{u}(A\boldsymbol{\mu} + \mathbf{b}) - \frac{1}{2}\mathbf{u}^T(A\Sigma A^T)\mathbf{u}].$$

From the form of the characteristic function we identify that $\mathbf{Y}$ is Gaussian with

$$E[\mathbf{Y}] = A\boldsymbol{\mu} + \mathbf{b} \qquad \operatorname{cov}(\mathbf{Y}, \mathbf{Y}) = A\Sigma A^T.$$

(c) Suppose $\Sigma > 0$ and write $\Sigma = CC^T$. Show that $C^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, I)$.
If $\mathbf{Y} = C^{-1}X - C^{-1}\boldsymbol{\mu}$ we have, by applying the previous result

$$E[\mathbf{Y}] = C^{-1}\boldsymbol{\mu} - C^{-1}\boldsymbol{\mu} = \mathbf{0}$$

$$\operatorname{cov}(\mathbf{Y}, \mathbf{Y}) = C^{-1}\Sigma C^{-T} = C^{-1}(CC^T)C^{-T} = I.$$

2. Suppose you have a random number generator which is capable of generating random numbers distributed as $X \sim \mathcal{N}(0, 1)$. Describe how to generate random vectors $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

   Generate n-dimensional random vectors $\mathbf{X}$ by generating n realizations of the scalar random variable X. Then $\operatorname{cov}(\mathbf{X}, \mathbf{X}) = I$.

   Write $\Sigma = CC^T$ (the Cholesky factorization). Let $\mathbf{Y} = C\mathbf{X} + \boldsymbol{\mu}$. Then

$$E[\mathbf{Y}] = \mathbf{m}$$

   and

$$\operatorname{cov}(\mathbf{Y}, \mathbf{Y}) = C \operatorname{cov}(\mathbf{X}, \mathbf{X})C^T = CIC^T = CC^T = \Sigma.$$

3. Suppose X and Y are r.v.s. Show that $E[(X - h(Y))^2]$ is minimized over all functions h be the function

$$h(y) = E[X|Y = y].$$

   Assume $E[X^2] < \infty$.

   (a) First approach: We will assume (only for convenience) that the r.v.s are continuous. We can write

$$E[(X - h(Y))^2] = E[E[(X - Y(Y))^2|Y = y]] = E\left[\int (X - h(y))^2 f_{X|Y}(X|Y = y)\, dx\right]$$
$$= \int \int (X - h(y))^2 f_{X|Y}(X|Y = y)\, dx f_Y(y)\, dy$$

To minimize this, we can minimize the inner integral

$$\int (X - h(y))^2 f_{X|Y}(X|Y = y)\, dx$$

for each value of $h(y)$. Since $Y = y$ is a fixed value for the inner integral, $h(y)$ is a fixed value for each value of y, and we can take the derivative with respect to $h(y)$ and equate to zero to minimize:

$$\frac{\partial}{\partial h(y)} \int (X - h(y))^2 f_{X|Y}(X|Y = y)\, dx = \int 2(X - h(y)) f_{X|Y}(X|Y = y)\, dx = 0$$

which leads to

$$h(y) \int f_{X|Y}(X|Y = y)\, dx = \int X f_{X|Y}(X|Y = y)\, dx$$

The integral on the left is equal to 1 (since it integrates over the entire set of X), and the integral on the right is equal to $E[X|Y = y]$.

(b) Approach 2: This one is more appealing, because it is not expressed in terms of integrals (making it immediately applicable to all types of distributions), it does not require interchanging limiting operations (i.e., taking derivatives inside integrals), and does not even require derivatives with respect to functions (which was only partially justified in the first approach). This presentation is due to Ross.

The proof is accomplished by establishing the following inequality:

$$E[(X - h(Y))^2] \geq E[(X - E[X|Y])^2]$$

This is accomplished as follows:

$$\begin{aligned} E[(X - h(Y))^2|Y] &= E[(X - E[X|Y] + E[X|Y] - h(Y))^2|Y] \\ &= E[(X - E[X|Y])^2|Y] + E[(E[X|Y] - h(Y))^2|Y] + \\ &\qquad 2E[(X - E[X|Y])\underbrace{(E[X|Y] - h(Y))}|Y] \end{aligned}$$

as may be verified by expanding out the second line. However, given Y the term $E[X|Y] - h(Y)$ (which has the underbrace), being a function of Y, can be treated as a constant and factored out of the expectation. Thus

$$\begin{aligned} E[(X - E[X|Y])(E[X|Y] - h(Y))|Y] &= (E[X|Y] - h(Y))E[(X - E[X|Y])|Y] \\ &= (E[X|Y] - h(Y))(E[X|Y] - E[X|Y]) \\ &= (E[X|Y] - h(Y))0 = 0 \end{aligned}$$

We thus obtain

$$\begin{aligned} E[(X-h(Y))^2|Y] &= E[(X-E[X|Y]+E[X|Y]-h(Y))^2|Y] \\ &= E[(X-E[X|Y])^2|Y]+E[(E[X|Y]-h(Y))^2|Y] \end{aligned}$$

or

$$E[(X-h(Y))^2|Y] \geq E[(X-E[X|Y])^2|Y]$$

Taking expectations of both sides we obtain

$$E[(X-h(Y))^2] \geq E[(X-E[X|Y])^2],$$

and equality holds if $h(Y)=E[X|Y]$.

4. Let $(X,Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$. Let $\mathbf{X} = \left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right] \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where

$$\Sigma = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}.$$

Determine the relationship between $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho$ and $\boldsymbol{\mu}$ and Σ.

First observe that $s_{21} = s_{12}$. Write down two expressions for the density, first in terms of the separate parameters,

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

and second in terms of the matrix parameters,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right].$$

Equating the constants in front we must have

$$\sigma_x^2\sigma_y^2(1-\rho^2) = s_{11}s_{22} - s_{21}^2.$$

It is clear that μ_x and μ_y are the first and second elements of $\boldsymbol{\mu}$. Let us assume only for convenience that $\boldsymbol{\mu} = \mathbf{0}$. Expanding the exponent of the second form of the pdf and equating to terms in the first form of the pdf we find

$$s_{11} - s_{21}^2/s_{22} = \sigma_x^2(1-\rho^2)$$

$$s_{22} - s_{21}^2/s_{11} = \sigma_y^2(1-\rho^2)$$

$$s_{11}s_{22}/s_{12} - s_{12} = \sigma_x\sigma_y/\rho$$

Comparing all of these, we have

$$\sigma_x^2 = s_{11} \qquad \sigma_y^2 = s_{22} \qquad \rho = \frac{s_{21}}{\sqrt{s_{11}s_{22}}}$$

5. Suppose $\mathbf{X} = \left[\begin{smallmatrix} X_1 \\ X_2 \\ X_3 \end{smallmatrix}\right] \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ where

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 6 & 3 \\ 1 & 3 & 8 \end{bmatrix}$$

(a) The value $X_1 = 1.5$ is measured. Determine the best estimate for (X_2, X_3).
We can write $\boldsymbol{\mu}_{2,3}$ for the mean of the unmeasured variables. The estimation formula is

$$\hat{\boldsymbol{\mu}}_{2,3} = \boldsymbol{\mu}_{2,3} + \Sigma_{[2,3],1}\Sigma_1^{-1}(x_1 - \mu_1)$$

where $\Sigma_{[2,3],1}$ is the covariance of the unmeasured values with the measured value,

$$\Sigma_{[2,3],1} = \begin{bmatrix} \mathrm{cov}(X_2, X_1) \\ \mathrm{cov}(X_3, X_1) \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and $\Sigma_1 = \Sigma_{1,1} = 4$. We obtain

$$\hat{\boldsymbol{\mu}}_{2,} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}(4)^{-1}(1.5 - 1) = \begin{bmatrix} 2.25 \\ 3.125 \end{bmatrix}.$$

(b) In a separate problem, the values $X_2 = 1$ and $X_3 = 5$ are measured. Determine the best estimate of X_1.

$$\hat{\mu}_1 = \mu_1 + \Sigma_{1,[2,3]}\Sigma_{[2,3]}^{-1}(\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} - \boldsymbol{\mu}_{2,3})$$

where

$$\Sigma_{1,[2,3]} = \begin{bmatrix} \mathrm{cov}(X_2, X_1) & \mathrm{cov}(X_3, X_1) \end{bmatrix} = \begin{bmatrix} 2 & 1 \end{bmatrix}$$

and

$$\Sigma_{[2,3]} = \begin{bmatrix} 6 & 3 \\ 3 & 8 \end{bmatrix}$$

Thus

$$\hat{\mu}_1 = 1 + \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 6 & 3 \\ 3 & 8 \end{bmatrix}^{-1} (\begin{bmatrix} 1 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix}) = 0.6667.$$

(c) Determine a random vector **Y** which is a whitened version of **X**.

Using MATLAB,

```
S = [4 2 1; 2 6 3; 1 3 8];
R = chol(S);
C = R';
C*C'   % check the result
```

we determine that

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2.2361 & 0 \\ 0.5 & 1.1180 & 2.5495 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2.2361 & 0 \\ 0.5 & 1.1180 & 2.5495 \end{bmatrix}^T = CC^T.$$

Now let $\mathbf{Y} = C(\mathbf{X} - \boldsymbol{\mu})$. Then $\mathbf{Y}$ is whitened.