Shengjie Li
RUID: 188008047

Homework 2

Problem 1: Let $\mathcal{X}_1, \mathcal{X}_2$ be two jointly Gaussian vectors with means μ_1, μ_2 covariance matrices Σ_{11}, Σ_{22} and cross covariance matrix $\Sigma_{12} = \mathbb{E}[(\mathcal{X}_1 - \mu_1)(\mathcal{X}_2 - \mu_2)^t]$. By computing the conditional probability density prove that $\mathcal{X}_1$ given $\mathcal{X}_2$ continuous to be Gaussian with mean that depends on $\mathcal{X}_2$ but with a covariance matrix which is independent of $\mathcal{X}_2$.

Proof.

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \mathbb{E}[\begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \end{bmatrix} \begin{bmatrix} \mathcal{X}_1^t & \mathcal{X}_2^t \end{bmatrix}]$$
$$\Sigma_{11} = \mathbb{E}[\mathcal{X}_1, \mathcal{X}_1^t] \qquad \Sigma_{12} = \mathbb{E}[\mathcal{X}_1, \mathcal{X}_2^t]$$
$$\Sigma_{21} = \mathbb{E}[\mathcal{X}_2, \mathcal{X}_1^t] \qquad \Sigma_{22} = \mathbb{E}[\mathcal{X}_2, \mathcal{X}_2^t]$$

Suppose $\mathcal{X}$ has zero mean, we can get the joint probability density function

$$f(\mathcal{X}_1, \mathcal{X}_2) = \frac{e^{-\frac{1}{2}\begin{bmatrix} \mathcal{X}_1^t & \mathcal{X}_2^t \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \end{bmatrix}}}{\sqrt{(2\pi)^{d_1+d_2}|\Sigma|}}$$
$$f(\mathcal{X}_2) = \frac{e^{-\frac{1}{2}\mathcal{X}_2^t \Sigma_{22}^t \mathcal{X}_2}}{\sqrt{(2\pi)^{d_2}|\Sigma_{22}|}}$$
$$f(\mathcal{X}_1|\mathcal{X}_2) = \frac{f(\mathcal{X}_1, \mathcal{X}_2)}{f(\mathcal{X}_2)}$$
$$= \frac{\frac{e^{-\frac{1}{2}\begin{bmatrix} \mathcal{X}_1^t & \mathcal{X}_2^t \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \end{bmatrix}}}{\sqrt{(2\pi)^{d_1+d_2}|\Sigma|}}}{\frac{e^{-\frac{1}{2}\mathcal{X}_2^t \Sigma_{22}^t \mathcal{X}_2}}{\sqrt{(2\pi)^{d_2}|\Sigma_{22}|}}}$$
$$= \frac{e^{-\frac{1}{2}(\begin{bmatrix} \mathcal{X}_1^t & \mathcal{X}_2^t \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \end{bmatrix} - \mathcal{X}_2^t \Sigma_{22}^t \mathcal{X}_2)}}{\sqrt{(2\pi)^{d_1}\frac{|\Sigma|}{|\Sigma_{22}|}}}$$

By using Schur's Inversion Formula, we can get:

$$\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}^{-1} = \begin{bmatrix}0 & 0\\ 0 & \Sigma_{22}^{-1}\end{bmatrix} + \begin{bmatrix}I\\ -E\end{bmatrix}\Delta^{-1}\begin{bmatrix}I & -F\end{bmatrix}$$

$$E = \Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{22}^{-1}\Sigma_{12}^{t}$$

$$F = \Sigma_{12}\Sigma_{22}^{-1}$$

$$\Delta = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{t}$$

$$\begin{bmatrix}\mathcal{X}_1^t & \mathcal{X}_2^t\end{bmatrix}\Sigma^{-1}\begin{bmatrix}\mathcal{X}_1\\ \mathcal{X}_2\end{bmatrix} - \mathcal{X}_2^t\Sigma_{22}^t\mathcal{X}_2 = \begin{bmatrix}\mathcal{X}_1^t & \mathcal{X}_2^t\end{bmatrix}\begin{bmatrix}0 & 0\\ 0 & \Sigma_{22}^{-1}\end{bmatrix}\begin{bmatrix}\mathcal{X}_1\\ \mathcal{X}_2\end{bmatrix} + \begin{bmatrix}\mathcal{X}_1^t & \mathcal{X}_2^t\end{bmatrix}\begin{bmatrix}I\\ -E\end{bmatrix}\Delta^{-1}\begin{bmatrix}I & -F\end{bmatrix}\begin{bmatrix}\mathcal{X}_1\\ \mathcal{X}_2\end{bmatrix} - \mathcal{X}_2^t\Sigma_{22}^t\mathcal{X}_2$$

$$= (\mathcal{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathcal{X}_2)^t(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t)(\mathcal{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathcal{X}_2)$$

$$\det\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{11}^t & \Sigma_{22}\end{bmatrix} = \det(\Sigma_{22})\det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t)$$

$$\frac{|\Sigma|}{|\Sigma_{22}|} = |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t|$$

$$f(\mathcal{X}_1|\mathcal{X}_2) = \frac{e^{-\frac{1}{2}(\mathcal{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathcal{X}_2)^t(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t)(\mathcal{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathcal{X}_2)}}{\sqrt{(2\pi)^{d_1}|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t|}}$$

Thus,

$$\mathcal{X}_1 \sim \mathcal{N}(\Sigma_{12}\Sigma_{22}^{-1}\mathcal{X}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t)$$

When $\mathcal{X}_1$ and $\mathcal{X}_2$ do not have zero mean,

$$\mathcal{X}_1 - \mu_1 \sim \mathcal{N}(\Sigma_{12}\Sigma_{22}^{-1}(\mathcal{X}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t)$$

□

Problem 2: Consider a Bernoulli random variable χ that takes the value a_1 with probability p and the value $a_2 (a_2 \neq a_1)$ with probability $1-p$.

a) Compute the the average and the variance of χ.

$$\mathbb{E}(\chi) = a_1 p + a_2(1-p)$$

$$Var(\chi) = \mathbb{E}(\chi^2) - \mathbb{E}(\chi)^2 = a_1^2 p + a_2^2(1-p) - (a_1 p + a_2(1-p))^2$$

$$= p(1-p)(a_1^2 + a_2^2 - 2a_1a_2)$$

$$= p(1-p)(a_1 - a_2)^2$$

b) Suppose now that you generate N independent realizations of χ. Propose a way to estimate $p = \mathbb{P}(\chi = a_1)$.

Suppose $\chi_1, \chi_2, \ldots, \chi_N$ are N observations.
Suppose in these N obervations, there are N_1 of them are value a_1. Then we can estimate the probablity $\hat{p} = \frac{N_1}{N}$.
If we use the indicator function, $\hat{p} = \frac{1}{N}\mathbb{1}\{\chi_i = a_1\}$.

c) Compute the mean and variance of your estimate. What can you conclude from this computation when you consider $N \to \infty$?

$$\mathbb{E}(\hat{p}) = \frac{1}{N}\sum_{i=1}^{N} p = p$$

$$Var(\hat{p}) = \mathbb{E}[(\hat{p}-p)^2] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\overline{(\mathbb{1}\{\chi_i = a_1\} - p)(\mathbb{1}\{\chi_j = a_1\} - p)}$$

When $i \neq j$,
$\overline{(\mathbb{1}\{\chi_i = a_1\} - p)(\mathbb{1}\{\chi_j = a_1\} - p)} = \overline{(\mathbb{1}\{\chi_i = a_1\} - p)}\ \overline{(\mathbb{1}\{\chi_j = a_1\} - p)} = 0$
Therefore,

$$\begin{aligned}\mathbb{E}[(\hat{p}-p)^2] &= \frac{1}{N^2}\sum_{i=1}^{N}\overline{(\mathbb{1}\{\chi_i = a_1\} - p)^2} \\ &= \frac{p(1-p)}{N}\end{aligned}$$

Problem 3: $\mathcal{X}$ is a random vector and there are K different possibilities that can generate realizations of this vector. Let $f_1(X), ..., f_K(X)$ the corresponding pdfs and $p_1, ..., p_K$ the corresponding prior probabilities that each case can occur of each possibility (with $p_1 + \cdots + p_K = 1$). Using total probability and the trick that relates a pdf to the probability of a differential event, show that the pdf $f(X)$ of $\mathcal{X}$ satisfies

$$f(X) = p_1 f_1(X) + \cdots + p_k f_K(X).$$

Let now χ_1, χ_2 be two random variables which 99% of the time are independent and Normally (Gaussian) distributed, both with mean 0 and variance 1 and 1% of the time they are independent and Normally distributed both with mean 0 and variance $\sigma_2 \neq 1$.

a) Compute the joint pdf of the two random variables.

$$\begin{aligned} f(X)dx &= f_1(X)dx \cdot p_1 + \cdots + f_K(X)dx \cdot p_K \\ f(X) &= f_1(X)p_1 + \cdots + f_K(X)p_K \end{aligned}$$

Because 99% of the time are independent and Normally distributed, both with mean 0 and variance 1 and 1% of the time they are independent and Normally distributed both with mean 0 and variance $\sigma_2 \neq 1$, we can get:

$$\begin{bmatrix}\chi_1 \\ \chi_2\end{bmatrix} \sim \mathcal{N}(0, \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}), 0.99$$

$$\begin{bmatrix}\chi_1 \\ \chi_2\end{bmatrix} \sim \mathcal{N}(0, \sigma_2\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}), 0.01$$

Then the joint pdf is:

$$f(X) = 0.99\frac{e^{-\frac{1}{2}(\chi_1^2+\chi_2^2)}}{\sqrt{(2\pi)^2}} + 0.01\frac{e^{-\frac{1}{2\sigma^2}(\chi_1^2+\chi_2^2)}}{\sqrt{(2\pi)^2\sigma^4}}$$

b) Examine if the two random variables are *independent*.

$$f(\chi_1) = 0.99\frac{e^{-\frac{1}{2}(\chi_1^2)}}{\sqrt{(2\pi)^2}} + 0.01\frac{e^{-\frac{1}{2\sigma^2}(\chi_1^2)}}{\sqrt{(2\pi)^2\sigma^4}}$$
$$f(\chi_2) = 0.99\frac{e^{-\frac{1}{2}(\chi_2^2)}}{\sqrt{(2\pi)^2}} + 0.01\frac{e^{-\frac{1}{2\sigma^2}(\chi_2^2)}}{\sqrt{(2\pi)^2\sigma^4}}$$
$$f(X) \neq f(\chi_1) \times f(\chi_2)$$

Thus, they are not independent.

c) Give an example of two random variables that are *uncorrelated* but not independent.

Let $X \sim U(-1, 1)$.
Let $Y = X^2$.
They are not independent.

$$\begin{aligned} cov(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] \\ &= 0 \end{aligned}$$

Thus, they are uncorrelated.

Problem 4: Let χ, ζ be random variables that are related through the equality

$$\zeta = |\chi + s|.$$

a) If the pdf of χ is $f_\chi(x)$ compute the pdf of ζ when s is a deterministic quantity.

Denote $F_\chi(x)$ as the cdf of χ.
Then we have $\mathbb{P}(\zeta \le z) = \mathbb{P}(|\chi + s| \le z) = \mathbb{P}(-z \le \chi + s \le z) = F_\chi(z - s) - F_\chi(-z - s)$ for $z \ge 0$.
Then the pdf is:

$$\begin{aligned} f_g(z) &= \frac{dF_g(z)}{dz} \\ &= f_\chi(z - s) - f_\chi(-z - s). \end{aligned}$$

b) Repeat the previous question when s is a random variable independent from χ and takes only the two values 0 and 1 with probabilities 0.2 and 0.8 respectively.

$$f_g(z) = f_g(z|s = 0)\mathbb{P}(s = 0) + f_g(z|s = 1)\mathbb{P}(s = 1)$$

c) Under the assumptions of question b) compute the posterior probability $\mathbb{P}(s = 0|\zeta = z)$. *Hint: For the computation of the pdf of a random variable the simplest way is to start with the computation of the cdf and then take the derivative. For b) use total probability.*

Problem 5: Consider the space of all scalar random variables.

a) Show that this is a vector space by defining properly the operation of addition and multiplication.

b) For any two random variables χ, ψ we define the mapping $< \chi, \psi >= \mathbb{E}[\chi\psi]$. Show that this mapping is an inner product in our vector space.

c) What particular form do you obtain when you apply the general Schwarz inequality?

d) How would you extend the previous definitions if you want a vector space comprised of *random vectors* of length d? Define properly the inner product and find the new form of the Schwartz inequality.