Homework #2

1. $\chi_1|\chi_2$

$$f(\chi_1,\chi_2)=\frac{e^{-\frac{1}{2}(\chi-\mu)^T\Sigma^{-1}(\chi-\mu)}}{\sqrt{(2\pi)^d|\Sigma|}}$$

$$\chi=\begin{bmatrix}\chi_1\\ \chi_2\end{bmatrix},\mu=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\Sigma=\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{bmatrix},d=d_1+d_2$$

$$f(\chi_2)=\frac{e^{-\frac{1}{2}(\chi_2-\mu_2)^T\Sigma_{22}^{-1}(\chi_2-\mu_2)}}{\sqrt{(2\pi)^{d_2}|\Sigma_{22}|}}$$

$$f(\chi_1|\chi_2)=\frac{e^{-\frac{1}{2}((\chi-\mu)^T\Sigma^{-1}(\chi-\mu)-(\chi_2-\mu_2)^T\Sigma_{22}^{-1}(\chi_2-\mu_2))}}{\sqrt{(2\pi)^{d_1}\frac{|\Sigma|}{|\Sigma_{22}|}}}$$

Let $M=(\chi-\mu)^T\Sigma^{-1}(\chi-\mu)-(\chi_2-\mu_2)^T\Sigma_{22}^{-1}(\chi_2-\mu_2),N=\frac{|\Sigma|}{|\Sigma_{22}|}$

Consider $\Sigma^{-1}=\begin{bmatrix}0 & 0\\ 0 & \Sigma_{22}\end{bmatrix}+\begin{bmatrix}I\\ -E\end{bmatrix}\Delta^{-1}\begin{bmatrix}I & -F\end{bmatrix}$ by Schur's Inversion Formula, where $E=\Sigma_{22}^{-1}\Sigma_{12}^T,F=\Sigma_{12}\Sigma_{22}^{-1},\Delta=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$.

$M=(\chi_2-\mu_2)^T\Sigma_{22}^{-1}(\chi_2-\mu_2)+((\chi_1-\mu_1)^T-(\chi_2-\mu_2)^T\Sigma_{22}^{-1}\Sigma_{12}^T)(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1}((\chi_1-\mu_1)-\Sigma_{12}\Sigma_{22}^{-1}(\chi_2-\mu_2))-(\chi_2-\mu_2)^T\Sigma_{22}^{-1}(\chi_2-\mu_2)$

Consider $\Sigma_{22}=\Sigma_{22}^T$:
$M=((\chi_1-\mu_1)-\Sigma_{12}\Sigma_{22}^{-1}(\chi_2-\mu_2))^T(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1}((\chi_1-\mu_1)-\Sigma_{12}\Sigma_{22}^{-1}(\chi_2-\mu_2))$

Consider Σ_{22} is invertable.
$det(\Sigma)=det(\Sigma\begin{bmatrix}\Sigma_{22} & 0\\ 0 & \Sigma_{22}^{-1}\end{bmatrix})=det(\begin{bmatrix}\Sigma_{11}\Sigma_{22} & \Sigma_{12}\Sigma_{22}^{-1}\\ \Sigma_{12}^T\Sigma_{22} & I\end{bmatrix})=det(\Sigma_{22})det(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)$
$N=det(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)$

Denote $\mu_{1|2}=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(\chi_2-\mu_2),\Sigma_{1|2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$

$$f(\chi_1|\chi_2)=\frac{e^{-\frac{1}{2}(\chi_1-\mu_{1|2})^T\Sigma_{1|2}^{-1}(\chi_1-\mu_{1|2})}}{\sqrt{(2\pi)^{d_1}|\Sigma_{1|2}|}}$$

Therefore, $\chi_1|\chi_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\chi_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)$

2. Bernoulli random variable χ

a) The average and the variance of χ

$$E(\chi) = p \times a_1 + (1-p) \times a_2 = a_2 + p(a_1 - a_2)$$
$$var(\chi) = p \times (a_1 - E(\chi))^2 + (1-p) \times (a_2 - E(\chi))^2 = p(1-p) \times (a_1 - a_2)^2$$

b) Propose a way to estimate p = P(χ = a1)

The estimation of p is the number of realizations that $\chi = a_1$ divided by the totall number of realizations.

$\hat{p} = \frac{1}{N}\sum_{k=1}^{N} \mathbb{I}\{x_k = a_1\}$, where $\mathbb{I}\{x_k = a_1\} = \begin{cases} 1 & , x_k = a_1 \\ 0 & , x_k \neq a_1 \end{cases}$

c) the mean and variance of estimate.

Notice $\mathbb{I}\{x = a_1\}$ is a Bernoulli random variable that takes the value 1 with probability p and the value 0 with probability 1−p.

$$E(\hat{p}) = \frac{1}{N}\sum_{k=1}^{N} E(\mathbb{I}\{x_k = a_1\}) = p$$
$$var(\hat{p}) = E((\hat{p} - p)^2) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbb{I}\{x_i = a_1\} - p)(\mathbb{I}\{x_j = a_1\} - p)$$

Notice that when $i \neq j$:
$\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbb{I}\{x_i = a_1\} - p)(\mathbb{I}\{x_j = a_1\} - p) = (\sum_{i=1}^{N}\mathbb{I}\{x_i = a_1\} - Np)(\sum_{j=1}^{N}\mathbb{I}\{x_j = a_1\} - Np) = 0$

Therefore, $var(\hat{p}) = \frac{1}{N^2}\sum_{k=1}^{N}(\mathbb{I}\{x_i = a_1\} - p)^2 = \frac{1}{N}(\mathbb{I}\{x_i = a_1\} - p)^2$

Notices that $\mathbb{I}^2\{x_i = a_1\} = \mathbb{I}\{x_i = a_1\}$:
$var(\hat{p}) = \frac{1}{N}(p - 2p + p^2) = \frac{1}{N}(1-p)p$

Therefore, when $N \to \infty$, the error of $\hat{p}$ will be smaller and smaller.

3 X is a random vector and there are K different possibilities that can generate realizations of this vector.

a) f(X) = p1f1(X) +····+ pkfK(X).

$$P(X \le \chi \le X + dX) = \sum_{k=1}^{K} p_k P_k(X \le \chi \le X + dX)$$

Consider $P(X \le \chi \le X + dX) = f_\chi(X)dX$

$$f(X)dX = \sum_{k=1}^{K} p_k f_k(X)dX$$

$$f(X) = \sum_{k=1}^{K} p_k f_k(X)$$

(In case of ambiguity, we can denote $X \le \chi$ when each coordinate of X is smaller than or equal to that of χ.)

b) the joint pdf of the two random variables.

$$f_{12}(X_1, X_2) = 0.99\frac{e^{-\frac{1}{2}(X_1^2+X_2^2)}}{\sqrt{(2\pi)^2}} + 0.01\frac{e^{-\frac{1}{2}\frac{X_1^2+X_2^2}{\sigma^2}}}{\sqrt{(2\pi)^2\sigma^4}} = 0.99\frac{e^{-\frac{1}{2}(X_1^2+X_2^2)}}{2\pi} + 0.01\frac{e^{-\frac{1}{2}\frac{X_1^2+X_2^2}{\sigma^2}}}{2\pi\sigma^2}$$

c) independent

$$f1(X_1) = 0.99\frac{e^{-\frac{1}{2}X_1^2}}{\sqrt{2\pi}} + 0.01\frac{e^{-\frac{1}{2}\frac{X_1^2}{\sigma^2}}}{\sqrt{(2\pi)}\sigma}$$

$$f2(X_2) = 0.99\frac{e^{-\frac{1}{2}X_2^2}}{\sqrt{2\pi}} + 0.01\frac{e^{-\frac{1}{2}\frac{X_2^2}{\sigma^2}}}{\sqrt{(2\pi)}\sigma}$$

Because $f_{12}(X_1, X_2) \neq f1(X_1)f2(X_2)$, χ_1, χ_2 are not independent.

d) uncorrelated but not independent

Consider $cor(\chi_1, \chi_2) = E(\chi_1\chi_2) - E(\chi_1)E(\chi_2) = 0 - 0 = 0$:
Therefore, χ_1, χ_2 are 2 random variables that are uncorrelated but not independent.

4 ζ = |χ + s|.

a) s is a deterministic quantity.

$$P(\zeta \leq z) = P(-z - s \leq \zeta \leq z - s) = F_\chi(z - s) - F_\chi(-z - s)$$
$$f_\zeta(z) = f_\zeta(z - s) + f_\zeta(-z - s)$$

b) s is a random variable independent from χ.

$$P(\zeta \leq z) = P(-z \leq \zeta \leq z|s = 0)P(s = 0) + P(-z - 1 \leq \zeta \leq z - 1|s = 1)P(s = 1)$$
$$f_\zeta(z) = 0.2(f_\zeta(z) + f_\zeta(-z)) + 0.8(f_\zeta(z - 1) + f_\zeta(-z - 1))$$

c) P(s = 0|ζ = z)

$$P(s = 0|\zeta = z) = \frac{P(\zeta=z,s=0)}{P(\zeta=z)} = \frac{0.2(f_\zeta(z)+f_\zeta(-z))}{0.2(f_\zeta(z)+f_\zeta(-z))+0.8(f_\zeta(z-1)+f_\zeta(-z-1))}$$

5 the space of all scalar random variables

a) a vector space.

χ, ψ, ζ are scalar random varibales, and m, n, k are real numbers.

$\oplus$: $\chi \oplus \psi = \chi + \psi$
Commutativity:
$\chi \oplus \psi = \chi + \psi = \psi + \chi = \psi \oplus \chi$
Associativity:
$(\chi \oplus \psi) \oplus \zeta = (\chi + \psi) + \zeta = \chi + (\psi + \zeta) = \chi \oplus (\psi \oplus \zeta)$
Identity element: 0

$\odot : k \odot \chi = k\chi$
Associativity:
$(mn) \odot \chi = (mn)\chi = m(n\chi) = m \odot (n \odot \chi)$
Distributivity:
$k \odot (\chi \oplus \psi) = k(\chi + \psi) = k\chi + k\psi = (k \odot \chi) \oplus (k \odot \psi)$

$(m+n) \odot \chi = (m+n)\chi = m\chi + n\chi = (m \odot \chi) \oplus (n \odot \chi)$
Identity element: 1

b)< χ,ψ >= E[χψ]

symmetry:
$< \chi, \psi >= E(\chi\psi) = E(\psi\chi) =< \psi, \chi >$
Linearity:
$< k\chi, \psi >= E(k\chi\psi) = kE(\chi\psi) = k < \chi, \psi >$
$< \chi + \psi, \zeta >= E((\chi + \psi)\zeta) = E(\chi\zeta + \psi\zeta) = E(\chi\zeta) + E(\psi\zeta) =< \chi, \zeta > + < \psi, \zeta >$
Positive-definiteness:
$< \chi, \chi >= E(\chi^2)$
$\chi^2 \geq 0, \chi^2 = 0 \Leftrightarrow \chi = 0$
Therefore, $< \chi, \chi >= E(\chi^2) \geq 0$
$< \chi, \chi >= 0 \Leftrightarrow \chi = 0$

c) the general Schwarz inequality

$| < \chi, \psi > | \leq ||\chi|| \, ||\psi||$
$|E(\chi\psi)| \leq \sqrt{E(\chi^2)}\sqrt{E(\psi^2)}$

d) random vectors of length d

$\chi = (\chi_1, \chi_2, ..., \chi_d)^T$
$\psi = (\psi_1, \psi_2, ..., \psi_d)^T$

$\oplus$: $\chi \oplus \psi = (\chi_1 + \psi_1, \chi_2 + \psi_2, ..., \chi_d + \psi_d)^T$
$\odot$: $k \odot \chi = (k\chi_1, k\chi_2, ..., k\chi_d)^T$

For each coordinate of random vectors, use the previous defination of $\oplus$, $\odot$. It is easy to show the new defination still keeps all the properties.

$< \chi, \psi >= E(\chi^T\psi)$
symmetry:
$< \chi, \psi >= E(\chi^T\psi) = E(\chi_1\psi_1 + \chi_2\psi_2 + ... + \chi_d\psi_d) = E(\psi^T\chi) =< \psi, \chi >$
Linearity:
$< k\chi, \psi >= E(k\chi^T\psi) = kE(\chi^T\psi) = k < \chi, \psi >$
$< \chi + \psi, \zeta >= E((\chi + \psi)^T\zeta) = E((\chi_1 + \psi_1)\zeta_1 + (\chi_2 + \psi_2)\zeta_2 + ... + (\chi_d +$

$\psi_d)\zeta_2) = E((\chi_1\zeta_1) + (\psi_1\zeta_1) + (\chi_2\zeta_2) + (\psi_2\zeta_2) + ... + (\chi_d\zeta_d) + (\psi_d\zeta_2)) =$
$E(\chi^T\zeta + \psi^T\zeta) = E(\chi^T\zeta) + E(\psi^T\zeta) =< \chi, \zeta > + < \psi, \zeta >$

Positive-definiteness:

$< \chi, \chi >= E(\chi^T\chi) = E(\chi_1^2 + \chi_2^2 + ... + \chi_d^2) = E(||\chi||)$
$||\chi|| \geq 0$, $\chi^2 = 0 \Leftrightarrow \chi = 0$
Therefore, $< \chi, \chi >= E(||\chi||) \geq 0$
$< \chi, \chi >= 0 \Leftrightarrow \chi = 0$

The Schwartz inequality:

$|E(\chi^T\psi)| \leq \sqrt{E(||\chi||)}\sqrt{E(||\psi||)}$