Centro de Investigación en Matemáticas (CIMAT). Unidad Monterrey

Estadítica Multivariada
Tarea 1

Marcelo Alberto Sanchez Zaragoza

16 de febrero de 2021

1. Problema 1

Demuestre que la matriz de centrado $P = I - \frac{1}{n}11^t$ cumple las siguientes propiedades:

a) Tiene rango $(n-1)$, es decir, tienen $n-1$ columnas o renglones linealmente independientes.

b) Sus valores propios son 0 o 1.

Solución
Inciso a)

La matriz P es de la forma:

$$P = I - \frac{1}{n}11^t = \begin{pmatrix} \frac{n-1}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} \\ \frac{-1}{n} & \frac{n-1}{n} & \cdots & \frac{-1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \cdots & \frac{n-1}{n} \end{pmatrix}$$

Si realizamos la suma de los primeros (n-1) renglones nos da como resultado el último renglón multiplicado por un -1, de la matriz P. Hemos encontrado un reglón que es combinación lineal de los demás por lo que la matriz P no es de rango completo.
Ahora no tomemos en cuenta este renglón(el último) y trabajemos con los demás, para analizar cuantos de ellos son linealmente independientes. Nuestra matriz queda de la siguiente forma:

$$\begin{pmatrix} \frac{n-1}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} & \frac{-1}{n} \\ \frac{-1}{n} & \frac{n-1}{n} & \cdots & \frac{-1}{n} & \frac{-1}{n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \cdots & \frac{n-1}{n} & \frac{-1}{n} \end{pmatrix}$$

Ahora si a todos los renglones les restamos al primer renglón $[\frac{n-1}{n}, \frac{-1}{n}, \frac{-1}{n}, \cdots, \frac{-1}{n}]$, vamos a tener lo siguiente:

$$\begin{pmatrix} \frac{n-1}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} & \frac{-1}{n} \\ 1 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

El primer renglón lo multiplicamos por n

$$\begin{pmatrix} n-1 & -1 & \cdots & -1 & -1 \\ -1 & 1 & \cdots & 0 & 0 \\ -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

Finalmente le sumamos todos los renglones al primer renglón y obtenemos lo siguiente:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ -1 & 1 & \cdots & 0 & 0 \\ -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

Con esta última matriz podemos observar que todos los renglones son linealmente independientes por lo que nuestra matriz P tiene (n-1) renglones linealmente independientes y por lo tanto su rango es de (n-1).

Inciso b)
Sabemos que $Px = \lambda x$ y que $P^2 = P$, multiplicamos ambos lados por la matriz P y tenemos lo siguiente:

$$\begin{aligned} Px &= \lambda x \\ P(Px) &= P(\lambda x) \\ P^2 x &= \lambda P x \\ P^2 x &= \lambda^2 x \end{aligned}$$

llegamos a que el valor propio de la matriz P^2 es λ^2 y que para la matriz P es λ, también sabemos que $P^2 = P$, por lo que podemos hacer esta igualdad:

$$\begin{aligned} \lambda^2 x &= \lambda x \\ \lambda^2 x - \lambda x &= 0 \\ (\lambda^2 - \lambda) x &= 0 \\ \lambda(\lambda - 1) = 0 \quad x &\neq 0 \\ \Rightarrow \ \lambda = 0 \quad \ \lambda &= 1 \end{aligned}$$

Finalmente encontramos que los valores propios de la matriz centrado P son 0 y 1.

Problema 2

Dados los datos, resuelva:

a) Dibújese el diagrama de dispersión múltiple y coméntese el aspecto del gráfico

b) Para X_1 y X_2 calcúlense, respectivamente, las medias muestrales $\bar{x_1}$ y $\bar{x_2}$, las varianzas muestrales s_{11} y s_{22}, la covarianza entre X_1 y X_2 , s_{12}, y la correlación entre ambas, r_{12} . Interprétese el valor obtenido de r_{12}.

c) Utilizando la matriz de datos X y la de centrado P, calcúlense el vector de medias muestrales $\bar{x}$ y la matriz de covarianzas muestrales S. A partir de ésta obténgase la matriz de correlaciones R.

Solución

Inciso a)

En la figura 1.1 vemos que las variables en cada gráfico de dispersión se encuentran correlacionados de forma lineal, ya que si se dibuja una linea recta en algunos gráficos esta linea no se encontraria tan lejos de la mayoria de los puntos.

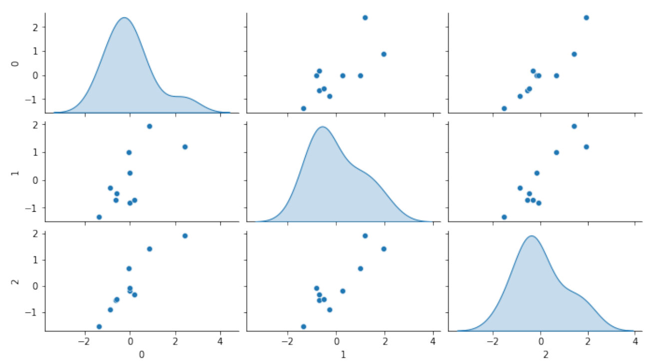

Figura 1.1: Gráfico de dispersión.

También es puede agregar que la unica dependencia que tenemos en la distribución normal bivariada es lineal por lo que si hay dicha dependencia solo seria de este tipo. Por lo anterior visto en los gráficos de dispersión se espera que las variables tenga cierta dependencia entre ellas, por ejemplo se esperaria que la variable X_3 tenga una correlación grande con la variable X_2 y también con la variable X_1, igual se espera cierta correlación entre la variable X_1 y X_2.
En la parte de la diagonal, los gráficos solo toman una de las variables y tomamos un kernel Gaussiano.

Código implementado en $PYTHON$ para hacer la figura 1.1.

```
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import matplotlib as mpl
import numpy as np
import seaborn as sns
import os
from sklearn.preprocessing import StandardScaler
tabla = {'x1': [8.7,14.3,18.9,19.0,20.5,14.7,18.8,37.3,12.6,25.7],
  'x2': [0.3,0.9,1.8,0.8,0.9,1.1,2.5,2.7,1.3,3.4],
  'x3': [3.1,7.4,9.0,9.4,8.3,7.6,12.6,18.1,5.9,15.9]}
data = pd.DataFrame(tabla)
ss = StandardScaler()
//se estandarizan los datos
scaled_df = ss.fit_transform(data)
scaled_df = pd.DataFrame(scaled_df)
data = pd.concat([scaled_df], axis=1)
pp2 = sns.pairplot(data, height=1.8, aspect=1.8,plot_kws=dict(edgecolor="w", linewidth
   =0.5),
diag_kind="kde", diag_kws=dict(shade=True),palette='flare',markers=['o','s','D'])

```

Inciso b)
Para poder encontrar el valor de $\bar{x}_1$, $\bar{x}_2$, s_{11} y s_{22}, vamos a hacer uso de

lo siguiente:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

$$s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

donde encontramos que nuestros valores son: $\bar{x}_1 = 19{,}05$, $\bar{x}_2 = 1{,}57$, $s_{11} = 56{,}9685$, $s_{22} = 0{,}8941$ y $s_{12} = 5{,}1705$.
Para encontrar el corficiente de correlación usaremos:

$$r_{jk} = \frac{s_{jk}}{s_j s_k}$$

donde encontramos que $r_{12} = 0{,}72447$, el valor nos dice que existe cierta relación entre las variables x_1(Duración media hipotetica) y x_2(Precio medio), es decir, dada la información de una de ellas podemos obtener información de la que no nos muestren. La relación que existe es entre las variables es lineal.

Inciso c)
Vamos a definir nuestras matrices X y P para encontrar los valores de la matriz S. Recordemos que $P = I - \frac{1}{n}11^t$ en este caso $n = 10$, por lo que nuestra matriz va a ser cuadrada y de dimensión 10.

$$X = \begin{pmatrix} 8{,}7 & 0{,}3 & 3{,}1 \\ 14{,}3 & 0{,}9 & 7{,}4 \\ 18{,}9 & 1{,}8 & 9{,}0 \\ 19{,}0 & 0{,}8 & 9{,}4 \\ 20{,}5 & 0{,}9 & 8{,}3 \\ 14{,}7 & 1{,}1 & 7{,}6 \\ 18{,}8 & 2{,}5 & 12{,}6 \\ 37{,}3 & 2{,}7 & 18{,}1 \\ 12{,}6 & 1{,}3 & 5{,}9 \\ 25{,}7 & 3{,}4 & 15{,}9 \end{pmatrix} \qquad P = \begin{pmatrix} 0{,}9 & -0{,}1 & ... & -0{,}1 \\ -0{,}1 & 0{,}9 & ... & -0{,}1 \\ ... & ... & ... & ... \\ -0{,}1 & -0{,}1 & ... & -0{,}1 \\ -0{,}1 & -0{,}1 & ... & 0{,}9 \end{pmatrix}$$

Para encontrar nuestro vector de media, lo calculamos con: $\bar{X} = \frac{1}{n}X^t 1$ y para la matriz de covarianzas tenemos $S = \frac{1}{n}X^t P X$, encontramos que las matrices son:

$$\bar{X} = \begin{pmatrix} 19{,}0{,}5 \\ 1{,}57 \\ 9{,}73 \end{pmatrix} \quad S = \begin{pmatrix} 56{,}9685 & 5{,}1705 & 30{,}4775 \\ 5{,}1705 & 0{,}8941 & 3{,}6479 \\ 30{,}4775 & 3{,}647 & 18{,}7641 \end{pmatrix}$$

Finalmente para la matriz $R = D^{1/2} S D^{-1/2}$, donde la matriz D es una matriz diagonal de tamaño 3x3, la matriz $D^{1/2}$ contiene las desviaciones típicas.

$$R = \begin{pmatrix} 1{,}0 & 0{,}7244 & 0{,}9323 \\ 0{,}7244 & 1{,}0 & 0{,}8908 \\ 0{,}9323 & 0{,}8908 & 1{,}0 \end{pmatrix}$$

Código implementado en $PYTHON$ para encontrar las matrices R, S y $\bar{X}$.

```
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import matplotlib as mpl
import numpy as np
import seaborn as sns
import os
x_1=np.array([(8.7,14.3,18.9,19.0,20.5,14.7,18.8,37.3,12.6,25.7)
    ,(0.3,0.9,1.8,0.8,0.9,1.1,2.5,2.7,1.3,3.4),
(3.1,7.4,9.0,9.4,8.3,7.6,12.6,18.1,5.9,15.9)])
x_m=[19.05,1.57,9.73]
I=np.identity(10)
a=np.zeros((10,1))
uno=np.ones_like(a)
p= I − 1/10*np.dot(uno,uno.T)
s_1=np.dot(x_1,p)
S=np.dot(s_1,y)
S=S/10
D=np.array([(7.5477,0,0),(0,0.9455,0),(0,0,4.331)])
D_inv=np.linalg.inv(D)
r_1=np.dot(D_inv,S)
R=np.dot(r_1,D_inv)

```

Problema 3

Considérese $x_1, ..., x_n$ de vectores de $\mathbb{R}^p$. Pruébese que la matriz de covarianzas $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t$, se puede expresar como $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^t - \bar{x}\bar{x}^t$.

Solución

Empezamos a desarrollar y obtenemos lo siguiente:

$$\begin{aligned} S &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i^t - \bar{x}^t) \\ &= \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \bar{x})x_i^t - \sum_{i=1}^{n}(x_i - \bar{x})\bar{x}^t\right\} \end{aligned}$$

Observe que $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

$$\begin{aligned} S &= \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \bar{x})x_i^t - \bar{x}^t(0)\right\} \\ &= \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \bar{x})x_i^t\right\} \\ &= \frac{1}{n}\left\{\sum_{i=1}^{n} x_i x_i^t - \sum_{i=1}^{n}\bar{x}x_i^t\right\} \end{aligned}$$

Note que $n\bar{x} = \sum_{i=1}^{n} x_i$, también observe que $n\bar{x}^t = \sum_{i=1}^{n} x_i^t$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n} x_i x_i^t - \bar{x}(n\bar{x})^t\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i x_i^t - \bar{x}\bar{x}^t$$

$$\Rightarrow S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^t - \bar{x}\bar{x}^t$$

Problema 4

Considere una población normal bivariada con $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ y $\rho_{12} = 0{,}5$.

a) Escriba la densidad normal bivariada explicitamente.

b) Escriba la expresión de distancia cuadrada generalizada $(x-\mu)^t\Sigma^-1(x-\mu)$ como función de x_1 y x_2.

c) Determine y grafique el contorno de densidad constante que contiene el 50 % de la probabilidad.

d) Especifique la distribución condicional de X_1 dado que $X_2 = x_2$.

Solución

Inciso a)

Primero se busco el valor de σ_{21}, se ocupo un despeje de la siguiente expresión para encontrar ese valor $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$, encontramos que el valor es $\sigma_{12} = \sqrt{2}/2$, así ya podemos expresar nuestras matrices μ y Σ:

$$\mu = \begin{pmatrix} 0 \\ 2 \end{pmatrix}; \quad \Sigma = \begin{pmatrix} 2 & \sqrt{2}/2 \\ \sqrt{2}/2 & 1 \end{pmatrix}$$

De donde podemos escribir nuestra distribución explicitamente de la distribución normal bivariada que nos proporcionan

$$f(x) = \frac{1}{(2\pi)\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} e^{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]}$$

Nuestra distribución bivariada es:

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)\sqrt{3/2}} e^{-\frac{1}{2(3/4)}\left[\left(\frac{x_1-0}{\sqrt{2}}\right)^2 + \left(\frac{x_2-2}{\sqrt{1}}\right)^2 - 2(1/2)\left(\frac{x_1-0}{\sqrt{2}}\right)\left(\frac{x_2-2}{\sqrt{1}}\right)\right]} \\ &= \frac{1}{(2\pi)\sqrt{3/2}} e^{-\frac{2}{3}\left[\left(\frac{x_1}{\sqrt{2}}\right)^2 + (x_2-2)^2 - \left(\frac{x_1}{\sqrt{2}}\right)(x_2-2)\right]} \end{aligned}$$

Inciso b)
Para poder obtener la expresión que nos pide el ejercicio vamos a realizar el producto de $(x-\mu)^t\Sigma^-1(x-\mu)$. Desarrollando tenemos:

$$\begin{aligned} (x-\mu)^t\Sigma^-1(x-\mu) &= \begin{pmatrix} x_1-0 \\ x_2-2 \end{pmatrix}^t \begin{pmatrix} 2/3 & -\sqrt{2}/3 \\ -\sqrt{2}/3 & 4/3 \end{pmatrix} \begin{pmatrix} x_1-0 \\ x_2-2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{2}{3}x_1 - \frac{\sqrt{2}}{3}(x_2-2) & -\frac{\sqrt{2}}{3}x_1 + \frac{4}{3}(x_2-2) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2-2 \end{pmatrix} \\ &= \frac{2}{3}x_1^2 + \frac{4}{3}x_2^2 - \frac{2\sqrt{2}}{3}x_1x_2 + \frac{4\sqrt{2}}{3}x_1 - \frac{16}{3}x_2 + \frac{16}{3} \end{aligned}$$

Finalmente tenemos lo siguiente, donde x es un vector:

$$f(x) = \frac{2}{3}x_1^2 + \frac{4}{3}x_2^2 - \frac{2\sqrt{2}}{3}x_1x_2 + \frac{4\sqrt{2}}{3}x_1 - \frac{16}{3}x_2 + \frac{16}{3}$$

Inciso c)
Para este inciso vamos a trabajar con lo siguiente:

$$(x-\mu)^t\Sigma^-1(x-\mu) \leq \mathcal{X}_2^2(\alpha)$$

donde

$$Pr[(x-\mu)^t\Sigma^-1(x-\mu) \leq \mathcal{X}_2^2(\alpha)] = 1-\alpha$$

que son contornos que contienen (1 - α)x100 % de los valores de x.
Para el ejercicio tenemos que $\alpha = 0{,}5$

$$\frac{2}{3}x_1^2 + \frac{4}{3}x_2^2 - \frac{2\sqrt{2}}{3}x_1x_2 + \frac{4\sqrt{2}}{3}x_1 - \frac{16}{3}x_2 + \frac{16}{3} \leq 1{,}3863$$

Ya que nos piden gráficar el contorno, se tienen que encontrar los valores y vectores propios de la matriz Σ.
Partimos de lo siguiente para encontrar lo antes mencionado:

$$\left|\Sigma - \lambda I\right| = 0$$

Donde encontramos que los valores propios son: $\lambda_1 = 2{,}3660$ y $\lambda_2 = 0{,}6339$, con vectores

$$U_1 = \begin{pmatrix} 0{,}8880 \\ -0{,}4597 \end{pmatrix} \text{ y } U_2 = \begin{pmatrix} 0{,}4597 \\ 0{,}8880 \end{pmatrix}$$

Una vez que hemos encontrado nuestros vectores y valores propios vamos a gráficar nuestro contorno, recordemos que los ejes están definidos por $\pm c\sqrt{\lambda_i}e_i$, donde $\Sigma_i e_i = \lambda_i e_i$ para $i = 1, 2$.
Tenemos que el centro de nuestro contorno se encuentra en $C = (0, 2)$, definido por (μ_1, μ_2).
En la figura 1.2 podemos observar la región que contiene el 50 % de la probabilidad.

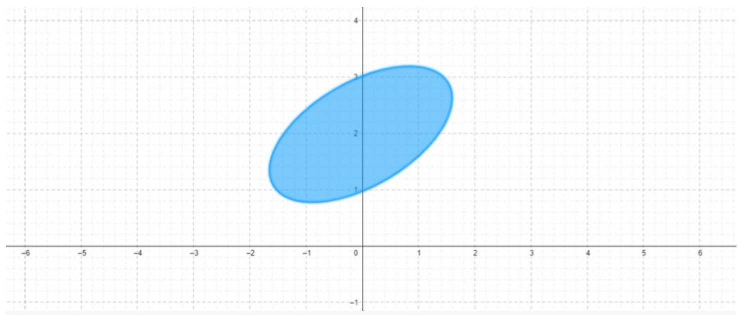

Figura 1.2: Contorno de densidad para $\alpha = 0{,}5$.

Código implementado en $PYTHON$ para encontrar los valores y vectores propios.

```
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import matplotlib as mpl
import numpy as np
import seaborn as sns
import os
vals,vecs=np.linalg.eig(x_1)
idx=vals.argsort()[::-1] // ordenamos nuestroa valores propios
eigvals = vals[idx]
eigvecs = vecs[:,idx]
eigvals
eigvecs
```

Inciso d)
Sabemos que la distribución condicional de x_1 dado x_2 es normal multivariada con:

$$Media = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$
$$Covarianza = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Para nuestro ejercicio tenemos:

$$\begin{pmatrix} x_1 \\ --- \\ x_2 \end{pmatrix}; \quad \mu = \begin{pmatrix} \mu_1 \\ --- \\ \mu - 2 \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \sigma_{11} & | & \sigma_{12} \\ - & - & - \\ \sigma_{21} & | & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & | & \Sigma_{12} \\ - & - & - \\ \Sigma_{21} & | & \Sigma_{22} \end{pmatrix}$$

Al sustituir los valores que nos proporcionan y encontrados tenemos lo siguiente:

$$Media = \frac{\sqrt{2}}{2}(x_2 - 2)$$
$$Covarianza = 2 - \frac{\sqrt{2}}{2} * \frac{\sqrt{2}}{2} = \frac{3}{2}$$

Note que no fue necesario buscar una matriz inversa para este inciso ya que Σ_{22} es solo un valor por lo que solo queda así: $\Sigma_{22}^{-1} = \frac{1}{1}$.
Hemos encontrado la distribución condicional de la forma:

$$f(X_1|X_2 = x_2) \sim N(\tfrac{\sqrt{2}}{2}(x_2 - 2), \tfrac{3}{2})$$

Problema 5

Sea X un vector aleatorio de distribución normal con media $\mu = (-1, 1, 0)^t$ y matriz de covarianzas

$$\Sigma = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

a) Hállese la distribución de $X_1 + 2X_2 - 3X_3$.

b) Hállese un vector a_{2x1} tal que las variables X_1 y $X_1 - a^t \begin{pmatrix} X_2 \\ X_3 \end{pmatrix}$ sean independientes.

c) Calcúlese la distribución de X_3 condicionada a $X_1 = x_1$ y $X_2 = x_2$.

Solución

Inciso a)

Sabemos que el vector X es de la forma: $X = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix}^t$ y nos piden encontrar la distribución de una combinación lineal de las variables X_1, X_2 y X_3. Por lo que nuestro vector C que nos ayudara a obtener dicha combinación lineal es: $C = \begin{pmatrix} 1 & 2 & -3 \end{pmatrix}^t$

Por medio de unas serie de propiedades vistas en clase sabemos lo siguiente:

$$a^t x = \sum_{i=1}^{p} a_i x_i \sim N(a^t \mu, a^t \Sigma a)$$

Entonces para encontrar la distribución de $C^t X = X_1 + 2X_2 - 3X_3$, tenemos:

$$X_1 + 2X_2 - 3X_3 \sim N(C^t \mu, C^t \Sigma C)$$

Realizando las respectivas multiplicaciones de matrices tenemos que $C^t \mu = 1$ y $C^t \Sigma C = 13$, finalmente:

$$X_1 + 2X_2 - 3X_3 \sim N(1, 13)$$

Inciso b)

Tomemos un vector $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ donde $y_1 = X_1$ y $y_2 = X_1 - a^t \begin{pmatrix} X_2 \\ X_3 \end{pmatrix}$, queremos que y_1 y y_2 sean independientes. Para obtener la independencia de estas dos variables vamos a buscar los valores del vector $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, de esta forma nuestro vector que al principio definimos quedaria de la siguiente manera:

$$Y = \begin{pmatrix} X_1 \\ X_1 - a_1 X_2 - a_2 X_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & a_1 & a_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = CX$$

Donde obtenemos la matriz C, que al multiplicarla por el vector X nos da como resultado el vector Y, $Y = CX$. Una vez que tenemos definido nuestro vector Y, el pedir independencia implica que la matriz de covarianzas sea de la siguiente forma:

$$\Sigma_y = \begin{pmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_{22} \end{pmatrix}$$

Donde sabemos que

$$\Sigma_y = Cov(X) = C\Sigma_x C^t$$

$$= \begin{pmatrix} 1 & 1 - a_2 \\ 1 - a_2 & 1 - 2a_2 + 2a_2a_1 + 3a_1^2 + 2a_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_{22} \end{pmatrix}$$

Donde solo debemos resolver que $1-a_2 = 0 \Rightarrow a_2 = 1$, como solo nos piden la independencia entre y_1 y y_2, el valor que puede tomar a_1 no afectaria dicha independencia por lo que nuestro vector a_{2x1} quedaria así:

$$a = \begin{pmatrix} a_1 \\ 1 \end{pmatrix}$$

donde a_1 puede tomar cualquier valor.

Inciso c)
Para encontrar la distribución condicional de X_3 dado X_1 y X_2 podemos inciar con:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ - - - - \\ X_3 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ - - - - \\ \mu_3 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & | & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & | & \sigma_{23} \\ - & - & - & - \\ \sigma_{31} & \sigma_{32} & | & \sigma_{33} \end{pmatrix}$$

Observe que ahora encontramos la distribución conjunta de X_1 y X_2, ya que encontramos las matrices $\mu_{1,2}$ y $\Sigma_{1,2}$, donde solo hay que sustituir valores.

$$X_{1,2} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \mu_{1,2} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma_{1,2} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Ya que nos piden encontrar la distribución condicional sabemos que dicha distribución condicional es normal multivariada con:

$$Media = \mu_1^* + \Sigma_{12}\Sigma_{22}^{-1}(x_2^* - \mu_2^*)$$
$$Covarianza = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Ahora reescribimos nuestro vector X, μ y Σ para poder encontrar los valores de la $Media$ y $Covarianza$ de la distribución condicional $f(X_3|X_1 = x_1, X_2 = x_2)$.

$$X^* = \begin{pmatrix} X_3 \\ ----- \\ X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix}; \quad \mu^* = \begin{pmatrix} \mu_3 \\ ----- \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix}$$

$$\Sigma^* = \begin{pmatrix} \sigma_{33} & | & \sigma_{31} & \sigma_{32} \\ - & - & - & - \\ \sigma_{13} & | & \sigma_{11} & \sigma_{12} \\ \sigma_{23} & | & \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix}$$

Al sustituir tenemos lo siguiente:

$$Media = 0 + \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/3 \end{pmatrix} \left(x_2^* \quad - \quad \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right)$$
$$= \begin{pmatrix} 1 & 1/3 \end{pmatrix} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad - \quad \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right) \quad = x_1 + 1 + \frac{x_2 - 1}{3}$$

$$Covarianza = (2) - \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2 - (1 + 1/3) = 2/3$$

Finalmente encontramos la distribución condicional:

$$f(X_3|X_1 = x_1, X_2 = x_2) \sim N(x_1 + 1 + \tfrac{x_2-1}{3}, \tfrac{2}{3})$$