1. (5 points) Let $X_1, X_2, X_3 \overset{iid}{\sim} N_2(\mu, \Sigma)$ where $\mu = (1, -2)'$ and $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$. Let $Y_1 = 2X_1 + 4X_2 + 3X_3$ and $Y_2 = 3X_1 + 5X_2 + 2X_3$. Find the distribution of $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$.

1.)

$a = \begin{pmatrix} 2 \\ 4 \\ 3 \end{pmatrix}$

$\sum_{i=1}^{3} a_i \mu_i = \sum_{i=1}^{3} a_i (1, -2)'$

$= 2(1,-2)' + 4(1,-2)' + 3(1,-2)'$

$= \begin{pmatrix} 2 \\ -4 \end{pmatrix} + \begin{pmatrix} 4 \\ -8 \end{pmatrix} + \begin{pmatrix} 3 \\ -6 \end{pmatrix}$

$= \begin{pmatrix} 9 \\ -18 \end{pmatrix}$

$(a'a)\Sigma = (2, 4, 3)\begin{pmatrix} 2 \\ 4 \\ 3 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= (4 + 16 + 9)\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= 29\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= \begin{pmatrix} 58 & 29 \\ 29 & 87 \end{pmatrix}$

$b = \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}$

$\sum_{i=1}^{3} b_i \mu_i = \sum_{i=1}^{3} b_i (1, -2)'$

$= 3(1,-2)' + 5(1,-2)' + 2(1,-2)'$

$= \begin{pmatrix} 3 \\ -6 \end{pmatrix} + \begin{pmatrix} 5 \\ -10 \end{pmatrix} + \begin{pmatrix} 2 \\ -4 \end{pmatrix}$

$= \begin{pmatrix} 10 \\ -20 \end{pmatrix}$

$(b'b)\Sigma = (3, 5, 2)\begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= (9 + 25 + 4)\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= 38\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= \begin{pmatrix} 76 & 38 \\ 38 & 114 \end{pmatrix}$

$(a'b)\Sigma = (2, 4, 3)\begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= (6 + 20 + 6)\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= 32\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$= \begin{pmatrix} 64 & 32 \\ 32 & 96 \end{pmatrix}$

$P = 2 \qquad 2p = 4$

☆
- Let $X_1, X_2, \ldots, X_n$ be mutually independent with $X_i \sim N_p(\mu_i, \Sigma)$ (each X_i has the same Σ), then
- $V_1 = a_1X_1 + a_2X_2 + \cdots + a_nX_n \sim N_p(\sum_{i=1}^{n} a_i\mu_i, (a'a)\Sigma)$.
- Moreover, if $V_2 = b_1X_1 + b_2X_2 + \cdots + b_nX_n$, then
- $\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \sum_{i=1}^{n} a_i\mu_i \\ \sum_{i=1}^{n} b_i\mu_i \end{pmatrix}, \begin{pmatrix} (a'a)\Sigma & (a'b)\Sigma \\ (a'b)\Sigma & (b'b)\Sigma \end{pmatrix}\right)$
- and $V_1 \perp V_2$ if $a'b = 0$.

$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \sum_{i=1}^{n} a_i\mu_i \\ \sum_{i=1}^{n} b_i\mu_i \end{pmatrix}, \begin{pmatrix} (a'a)\Sigma & (a'b)\Sigma \\ (a'b)\Sigma & (b'b)\Sigma \end{pmatrix}\right)$

so $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_4\left(\begin{pmatrix} (9, -18)' \\ (10, -20)' \end{pmatrix}, \begin{pmatrix} \begin{pmatrix} 58 & 29 \\ 29 & 87 \end{pmatrix} & \begin{pmatrix} 64 & 32 \\ 32 & 96 \end{pmatrix} \\ \begin{pmatrix} 64 & 32 \\ 32 & 96 \end{pmatrix} & \begin{pmatrix} 76 & 38 \\ 38 & 114 \end{pmatrix} \end{pmatrix}\right)$

from ☆

2.)

2. (4 points) (Invariance of Hotelling T^2)
Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N_p(\mu_X, \Sigma_X)$ and $Y_i = AX_i + b, i = 1, 2, \ldots, n$ where, $A \in R^{p\times p}$ is a non-singular matrix of constants and $b \in R^p$ is a vector of constants.
Let $T_X = n(\bar{X} - \mu_X)'S_X^{-1}(\bar{X} - \mu_X)$ and $T_Y = n(\bar{Y} - \mu_Y)'S_Y^{-1}(\bar{Y} - \mu_Y)$ where S_X is the sample covariance matrix of $X_1, X_2, \ldots, X_n$, S_Y is the sample covariance matrix of $Y_1, Y_2, \ldots, Y_n$ and $\bar{X}$ and $\bar{Y}$ are the corresponding sample means.
Show that $T_Y = T_X$.

- $X \sim N_p(\mu, \Sigma)$ and $d \in R^p$ is a vector of constants, then
- $X + d \sim N(\mu + d, \Sigma)$

- Result: If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N_p(\mu, \Sigma)$, then
- $n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) \sim T^2(p, n-1)$.

- Result 4.3 p157:
- If $A_{q\times p} \in R^{q\times p}$ and $d \in R^q$ are constants and $X \sim N_p(\mu, \Sigma)$, then $AX + d \sim N_q(A\mu + d, A\Sigma A')$.

: $T^2(p, n-1) \overset{d}{=} \left(\frac{(n-1)p}{n-1-p+1}\right)F_{p,n-1-p+1} = \left(\frac{(n-1)p}{n-p}\right)F_{p,n-p}$

- Exercise: Prove that $X \sim N_p(\mu, \Sigma) \Rightarrow Z = \Sigma^{-\frac{1}{2}}(X - \mu) \sim N_p(0, I)$.

Since $X_1, X_2 \ldots X_n \sim N_p(\mu_X, \Sigma_X)$ then
$T_X = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) \sim T_X^2(p, n-1)$
$Y_i = AX_i + b$ where $A_{p\times p} \in R^{p\times p}$ and $d \in R^p$ also so

$T_X = n(\bar{X}-\mu)' S^{-1} (\bar{X}-\mu) \sim T_X^2(p, n-1)$

$Y_i = AX_i + b$ where $A_{p\times p} \in R^{p\times p}$ and $\det \in R^p$ also so

$Y_i \sim N_p(A\mu_X + b, A\Sigma_X A')$

$T_Y = n(\bar{Y}-\mu_Y)' S_Y^{-1} (\bar{Y}-\mu_Y) \sim T_Y^2(p, n-1)$ and $T_X^2(p, n-1) = T_Y^2(p, n-1) \stackrel{d}{=} \left(\frac{(n-1)p}{n-p}\right) F_{p, n-p}$ ∎

3a.)

3. (a) (7 points) Let $X \sim N_p(\mu, \Sigma), \mu \in R^p$, $Y_1 = CX$ and $Y_2 = DX$ where $C \in R^{q_1\times p}, D \in R^{q_2\times p}$ are matrices of constants. Use Result 4.3 on page 9 of our lecture slides on Chap 3 and 4 (i.e. chap3_4_RandomSampling_MVN) to prove that $(Y_1, Y_2)'$ has a multivariate Normal distribution. Identify the mean and the covariance matrix of this multivariate Normal distribution. When are Y_1 and Y_2 independent? Explain.

$X \sim N_p(\mu, \Sigma)$ and $C \in R^{q_1\times p}$ and $D \in R^{q_2\times p}$ $\Rightarrow$

$Y_1 = CX \sim N_{q_1}(C\mu, C\Sigma C')$, $Y_2 = DX \sim N_{q_2}(D\mu, D\Sigma D')$

- Result 4.3 p157
- If $A_{q\times p} \in R^{q\times p}$ and $d \in R^q$ are constants and $X \sim N_p(\mu, \Sigma)$, then $AX + d \sim N_q(A\mu + d, A\Sigma A')$.

let $Y = \begin{pmatrix} X \\ X \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$ let $q_1 \geq q_2$

$C = \begin{bmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & \ddots & \vdots \\ c_{q_1 1} & \cdots & c_{q_1 p} \end{bmatrix} = \begin{bmatrix} C_1' \\ \vdots \\ C_{q_1}' \end{bmatrix}$ let $Q = \begin{bmatrix} C_1', & D_1' \\ \vdots & \vdots \\ C_{q_2}', & D_{q_2}' \\ C_{q_2+1}', & 0 \\ \vdots & \vdots \\ C_{q_1}', & 0 \end{bmatrix}$ so $Q \in R^{q_1 \times 2p}$ is a matrix of constants

$D = \begin{bmatrix} D_1' \\ \vdots \\ D_{q_2}' \end{bmatrix}$

then by 4.3 $QY \sim N_{q_1}\left(Q\begin{pmatrix} \mu \\ \mu \end{pmatrix}, Q\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} Q'\right)$

$Y_1 \perp Y_2$ when $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = diag(\Sigma, \Sigma)$

- Special case: $X \sim N_p(\mu, \Sigma)$ and $\Sigma = diag(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp})$
- $\Rightarrow X_1, X_2, \ldots, X_p$ are independent
- and $X_i \sim N(\mu_i, \sigma_{ii})$ for $i = 1, 2, \ldots, p$.

3b.)

(b) (7 points) Let $X \sim N_n(\mu 1, \sigma^2 I), \mu \in R, \sigma^2 \in R^+$. Let $Y_1 = \frac{1}{n}1'X$ and $Y_2 = C_n X$ where C_n is the centering matrix (i.e. $C_n = I - \frac{1}{n}J$). Use the result in part a above to find the distribution of $(Y_1, Y_2)'$.
Comment on your results. In particular this result is related to one very important result that we come across very often in statistics. What is that? Discuss.

let $D = \frac{1}{n}1' = \underbrace{\left[\frac{1}{n}, \frac{1}{n}, \ldots \frac{1}{n}\right]}_{1\times n}$ $C = C_n = I - \frac{1}{n}J = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}_{n\times n} - \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}_{n\times n}$

$= \begin{bmatrix} \frac{n-1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \ddots & \vdots \\ -\frac{1}{n} & \cdots & \frac{n-1}{n} \end{bmatrix}_{n\times n}$

$Q = \begin{bmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ -\frac{1}{n} & \ddots & & \vdots & 0 & \cdots & 0 \\ \vdots & & \ddots & -\frac{1}{n} & \vdots & \ddots & \vdots \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} & 0 & \cdots & 0 \end{bmatrix}_{n\times 2n}$

$Y = \begin{pmatrix} X \\ X \end{pmatrix} \sim N_{2n}\left(\begin{pmatrix} \mu 1 \\ \mu 1 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 I & \sigma_{12} I \\ \sigma_{21} I & \sigma_2^2 I \end{pmatrix}\right)$

thus from above $QY \sim N_n\left(Q\begin{pmatrix} \mu 1 \\ \mu 1 \end{pmatrix}, Q\begin{pmatrix} \sigma_1^2 I & \sigma_{12}^2 I \\ \sigma_{21}^2 I & \sigma_2^2 I \end{pmatrix} Q'\right)$

which is

3c.)

(c) (3 points) Use the result in part (a) above to show that if $X \sim N_p(\mu, \sigma^2 I), \sigma^2 \in R^+$ and G is a $q \times p$ matrix such that $GG' = I_q$, then GX and $(I - G'G)X$ are independent.

$cov(GX, (I-G'G)X) = E(G X (I-G'G)X) - E(G X) E((I-G'G)X)$

$= E(G X I - G G'GX) - G\sigma^2 I ((I - G'G)\sigma^2 I)$

$= E(GX - IGX) - G\sigma^2 I (\sigma^2 I - G'G\sigma^2 I)$

$= E(GX - GX) - G\sigma^2\sigma^2 I + GG'G\sigma^2\sigma^2 I$

$= 0 - G\sigma^2\sigma^2 I - G\sigma^2\sigma^2 I$

$= 0$

So since GX and $(I-G'G)X$ has a normal distribution (from a) and their covariance is 0, GX and $(I - GG')X$ are independent

4.)

4. (10 points) Let $Y_1, Y_2, \ldots, Y_{100} \overset{iid}{\sim} Uniform(14, 20)$. Define the bivariate random vectors $X_i = (Y_i, Z_i)'$ where $Z_i = Y_i^2, i = 1, 2, \ldots, 100$. Use the multivariate central limit theorem to find the limiting distribution of $\bar{X}_{100} = \frac{1}{100}\sum_{i=1}^{100} X_i$. Explain all your work in detail.

$\bar{X} = \frac{1}{100}\sum_{i=1}^{100} X_i = \frac{1}{100}\sum_{i=1}^{100}\begin{pmatrix} Y_i \\ Z_i\end{pmatrix} = \frac{1}{100}\left(\sum_{i=1}^{100} Y_i, \sum_{i=1}^{100} Y_i^2\right)' = (\bar{Y}, \overline{Y^2})'$

$E(Y_i) = \frac{14+20}{2} = 17$

$V(Y_i) = \frac{(20-14+1)^2 - 1}{12} = 4$

$E(Y_i^2) = V(Y_i) + E(Y_i)^2$

$= 4 + 17^2$

$= 293$

$\mu = \begin{pmatrix} 17 \\ 293 \end{pmatrix} \in R^{2\times 1}$ $\quad \sqrt{n} = \sqrt{100} = 10$

$\sqrt{n}(\bar{X} - \mu) \overset{app}{\sim} N_p(0, \Sigma)$

$(\bar{X} - \mu) \sim \frac{N_p(0, \Sigma)}{\sqrt{n}}$

$\bar{X} \sim \frac{N_p(0, \Sigma)}{\sqrt{n}} + \mu = \frac{1}{10} N_p(0, \Sigma) + \begin{pmatrix} 17 \\ 293 \end{pmatrix}$

- If If $X_1, X_2, \ldots, X_n$ are independent observations from population with mean μ and covariance Σ, then
- 1) $\sqrt{n}(\bar{X} - \mu) \overset{app}{\sim} N_p(0, \Sigma)$ (The central limit theorem)
- 2) $n(\bar{X} - \mu)' S^{-1}(\bar{X} - \mu) \overset{app}{\sim} \chi_p^2$
- for $n - p$ large

5.)

5. (14 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N_p(0, \Lambda)$ where Λ is a $p \times p$ diagonal matrix with diagonal elements $\sigma_1^2, \ldots, \sigma_p^2$. i.e. $\Lambda = diag(\sigma_1^2, \ldots, \sigma_p^2)$. Let $D = diag(S_1^2, \ldots, S_p^2)$, where S_i^2 is the ith diagonal element of the sample covariance matrix S. For testing the null hypothesis $H_0 : \mu = 0$ for this model, one statistician is considering the test statistic $T = \frac{n}{p}\bar{X}' D^{-1}\bar{X}$. Show that this statistic T can be expressed as $T = \frac{1}{p}\sum_{j=1}^{p} Y_j$ where $Y_1, Y_2, \ldots, Y_p \overset{iid}{\sim} F(1, n-1)$. Based on the central limit theorem, what is the limiting distribution of T as $p \to \infty$?
Note 1: This statistic T is not exactly Hotelling T^2 statistic that we normally use for teting $H_0 : \mu = 0$ in a general multivariate Normal model (i.e. the one that we discussed in class).
Hint 1: In your answer, you must first identify Y_j in terms of the given variables, then prove that $Y_i \sim F(1, n-1)$ and $Y_1, Y_2, \ldots, Y_p$ are independent.
Hint 2: For $Y \sim F(\nu_1, \nu_2)$, $E(Y) = \frac{\nu_2}{\nu_2 - 2}$ when $\nu_2 > 2$ and $Var(Y) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ when $\nu_2 > 4$ (page 707 Evans and Rosenthal).

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N_p(0, \Lambda) \Rightarrow \bar{X} \sim N_p(0, \frac{1}{n}\Lambda)$

$(n-1)D \sim W_{p,n-1}(\Lambda)$

$\bar{X} \perp S \Rightarrow \bar{X} \perp D$

Result: If $x \sim N_p(\mu, \Sigma)$, $M \sim W_{p,m}(\Sigma)$ and $x \perp M$, then $m(x-\mu)'M^{-1}(x-\mu) \sim T^2(p, m)$.

$\sqrt{n}\,\bar{X} \sim N_p(\sqrt{n}\,0, \Lambda) = N_p(0, \Lambda)$

$\mu = 0 \quad m = n-1$

- If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N_p(\mu, \Sigma)$, then
- $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$
- $(n-1)S \sim W_{p,n-1}(\Sigma)$
- $\bar{X} \perp S$

$(n-1)(\sqrt{n}\bar{X} - \sqrt{n}0)'((n-1)D)^{-1}(\sqrt{n}\bar{X} - \sqrt{n}0) \sim T^2(p, n-1)$

$\Rightarrow n\bar{X}'D^{-1}\bar{X} \sim T^2(p, n-1)$

$\Rightarrow \frac{n}{p}\bar{X}'D^{-1}\bar{X} \sim \frac{1}{p}T^2(p, n-1)$

- Note: $T^2(p, n-1) \overset{d}{=} \left(\frac{(n-1)p}{n-1-p+1}\right) F_{p,n-1-p+1} = \left(\frac{(n-1)p}{n-p}\right) F_{p,n-p}$

let $Y_j = \frac{Q_k}{\frac{n-p}{1}} = \frac{Q_k}{\frac{1}{n-p}} \sim F(1, n-1)$ where $Q_k \overset{iid}{\sim} \chi^2_{(1)}$ so Y_j are independent

$\Rightarrow \sum$ [illegible]

- Additive property: If $X \sim \chi^2_{(m)}$ and $Y \sim \chi^2_{(n)}$ then $X + Y \sim \chi^2_{(m+n)}$

let $Y_j = \frac{Q_k}{\sum_{l=1}^{n-p} \frac{1}{n-1} R_l} = \frac{Q_k}{\frac{1}{n-1}\sum_{l=1}^{n-p} R_l} \sim F(1, n-1)$

$k = 1, \ldots, p$

$l = 1, \ldots, n-p$

where $Q_k \overset{iid}{\sim} \chi^2_{(1)}$ so Y_j are independent

$\Rightarrow \sum_{k=1}^{p} Q_k \sim \chi^2_{(p)}$

- Additive property: If $X \sim \chi^2_{(m)}$ and $Y \sim \chi^2_{(n)}$ then $X + Y \sim \chi^2_{(m+n)}$

Let X and Y are two independent variables where $X \sim \chi^2_{(m)}$ and $Y \sim \chi^2_{(n)}$

Then $\frac{X/m}{Y/n} \sim F$ distribution with degrees of freedom (m, n)

and $R_l \overset{iid}{\sim} \chi^2_{(1)}$

$\Rightarrow \sum_{l=1}^{n-p} R_l \sim \chi^2_{(n-p)}$

$\Rightarrow \frac{\left(\frac{1}{p}\sum_{k=1}^{p} Q_k\right)}{\left(\frac{1}{n-p}\sum_{l=1}^{n-p} R_l\right)} \sim F_{p, n-p}$

$$\frac{1}{p} T^2(p, n-1) = \frac{1}{p} \frac{(n-1)p}{n-p} F_{p,n-p}$$

$$= \frac{1}{p}(n-1)\frac{p}{n-p} \frac{\left(\frac{1}{p}\sum_{k=1}^{p} Q_k\right)}{\left(\frac{1}{n-p}\sum_{l=1}^{n-p} R_l\right)}$$

$$= \frac{1}{p}\left(\frac{\sum_{k=1}^{p} Q_k}{\frac{1}{n-1}\sum_{l=1}^{n-p} R_l}\right)$$

$$= \frac{1}{p}\sum_{k=1}^{p}\left(\frac{Q_k}{\sum_{l=1}^{n-p}\frac{1}{n-1} R_l}\right)$$

since $Y_j = \frac{Q_k}{\sum_{l=1}^{n-p}\frac{1}{n-1} R_l} \overset{iid}{\sim} F(1, n-1)$

$$= \frac{1}{p}\sum_{j=1}^{p} Y_j$$

$$T = \frac{1}{p}\sum_{j=1}^{p} Y_j = \frac{1}{p}\left(\frac{p}{p}\right)\sum_{j=1}^{p} Y_j = \bar{Y}$$

by CLT $T = \bar{Y} \to N\left(-n+1, \frac{2(n-1)^2(n-2)}{p(n-3)^2(n-5)}\right)$ as $p \to \infty$, $n > 5$

$E(Y) = \frac{n-1}{1-2} = -n+1$ when $n-1 > 2 \Rightarrow n > 3$

$$Var(Y) = \frac{2(n-1)^2(1+n-1-2)}{1(n-1-2)^2(n-1-4)}$$

$$= \frac{2(n-1)^2(n-2)}{(n-3)^2(n-5)}$$ when $n-1 > 4 \Rightarrow n > 5$