1. (25 pts) Sea U_1 y U_2 dos variables aleatorias independientes $\mathsf{U}(0,1)$. Suponga $\boldsymbol{X} = (X_1, X_2, X_3, X_4)^\top$ donde

$$X_1 = U_1, \qquad X_2 = U_2, \qquad X_3 = U_1 + U_2, \qquad X_4 = U_1 - U_2.$$

Calcule la matrix de correlación $\boldsymbol{R}$ de $\boldsymbol{X}$. ¿Cuántas componentes (PC) son de interés? Muestre que

$$\boldsymbol{t}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{t}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \\ 1 \end{pmatrix},$$

son vectores propios de $\boldsymbol{R}$ asociados a λ's no triviales. Interprete las dos primeras PC obtenidas.

Solución:

Notemos que E(X1) = E(U1) = 1/2, E(X2) = E(U2) = 1/2, E(X3) = E(U1 + U2) = 1, E(X4) = E(U1-U2) = 0. Como U1, U2 son independientes eso implica que Cov(U1,U2)=0, además se tiene que la varianza separa la suma, por ende Var(X1) = 1/12, Var(X2) = 1/12, Var(X3) = Var(U1+U2) = Var(U1)+Var(U2) = 1/6, Var(X4) = Var(U1-U2) = Var(U1) + Var(U2) = 1/6.

Objetivo: Encontrar la matriz de correlación.

Notemos que X normalizado viene dado por:

$\overline{X1}$ = (X1 - 1/2)/(1/12)^(1/2) = (U1 - 1/2)$\sqrt{12}$

$\overline{X2}$ = X1

$\overline{X3}$ = (X3 - 1)/(1/6) ^(1/2) = (U1 + U2 - 1)$\sqrt{6}$

$\overline{X4}$ = X4/(1/6)^(1/2) = (U1-U2)$\sqrt{6}$

Por contrucción todas las variables anteriores tienen varianza 1, luego resta calcular los restos de la terminos de la matriz de correlación.

Cov($\overline{X1}$,$\overline{X2}$) = 0, pues U1,U2 son independientes.

Cov($\overline{X1}$,$\overline{X3}$) =$\sqrt{12*6}$ Cov(U1-1/2,U1+U2-1) =$\sqrt{12 * 6}$ Var(U1) = 1/$\sqrt{2}$

Cov(X1,X4) = 1/$\sqrt{2}$

Cov(X2,X3) =$\sqrt{12*6}$*Var(U2) = 1/$\sqrt{2}$

Cov(X2,X4) =$\sqrt{12*6}$*(-1)Var(U2) = - 1/$\sqrt{2}$

Cov(X3,X4) = 6(Var(U1)-Var(U2))=0

De esa forma:

$$R = \begin{pmatrix} 1 & 0 & 1/2 & 1/\sqrt{2} \\ 0 & 1 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} & 1 & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 1 \end{pmatrix}$$

Computacionalmente se tiene que

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (2, 2, 0, 0)$$

$$T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & -1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} & -1/\sqrt{2} \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Por tanto las componentes principales son dadas por

$$y_1 = t_1 \cdot \bar{x} = \frac{1}{\sqrt{2}}\bar{x}_1 + \frac{1}{\sqrt{2}}\bar{x}_2 + \bar{x}_3$$
$$y_2 = t_2 \cdot \bar{x} = \frac{1}{\sqrt{2}}\bar{x}_1 + \frac{1}{\sqrt{2}}\bar{x}_2 + \bar{x}_4$$
$$y_3 = t_3 \cdot \bar{x} = -\frac{1}{\sqrt{2}}\bar{x}_1 + \frac{1}{\sqrt{2}}\bar{x}_2 + \bar{x}_4$$
$$y_4 = t_4 \cdot \bar{x} = -\frac{1}{\sqrt{2}}\bar{x}_1 - \frac{1}{\sqrt{2}}\bar{x}_2 + \bar{x}_3$$

Como la traza de R es 4 entonces tenemos que el porcentaje que explica la primera y segunda componente es de un 50% (2/4) mientras que las otras 2 no explican (0/4 = 0).

2. (25 pts) Considere un modelo de análisis factorial

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, n,$$

donde $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{z}_i \sim \mathsf{N}_q(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{\epsilon}_i \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Psi})$, y $\boldsymbol{\Psi} = \operatorname{diag}(\psi_1, \ldots, \psi_p)$. Sea $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}^\top + \widehat{\boldsymbol{\Psi}}$, con $\widehat{\boldsymbol{\Gamma}}$ y $\widehat{\boldsymbol{\Psi}}$ los MLE de $\boldsymbol{\Gamma}$ y $\boldsymbol{\Psi}$, respectivamente. Muestre que $\operatorname{tr}(\boldsymbol{S}\widehat{\boldsymbol{\Sigma}}^{-1}) = p$.

Recuerde que: por el Teorema de Sherman-Morrison-Woodbury, tenemos

$$(\boldsymbol{A} + \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{DA}^{-1}$$

Solución:

Consideremos: $A = \Psi$, $B = \Gamma$, $C = Id$ y $D = \Gamma^T$.

Por el Teorema de Sherman-Morrison-Woodbury se tiene que:

$$\begin{aligned}\Sigma^{-1} &= (\Psi + \Gamma\Gamma^T)^{-1} \\ &= \Psi^{-1} - \Psi^{-1}\Gamma(Id + \Gamma^T\Psi^{-1}\Gamma)^{-1}\Gamma^T\Psi^{-1}\end{aligned}$$

Por invarianza de los estimadores tenemos que:

$$\hat{\Sigma}^{-1} = \hat{\Psi}^{-1} - \hat{\Psi}^{-1}\hat{\Gamma}(Id + \hat{\Gamma}^T\hat{\Psi}^{-1}\hat{\Gamma})^{-1}\hat{\Gamma}^T\hat{\Psi}^{-1}$$

Por resultado visto en clases sabemos que los MLE presentados cumplen las siguientes igualdades

$$\begin{aligned}S\hat{\Psi}^{-1}\hat{\Gamma} &= \hat{\Gamma}(Id + \hat{\Gamma}^T\hat{\Psi}^{-1}\hat{\Gamma}) \\ \hat{\Psi} &= \operatorname{diag}(S - \hat{\Gamma}\hat{\Gamma}^T) \\ \hat{\Lambda} &= \hat{\Gamma}^T\hat{\Psi}^{-1}\hat{\Gamma}\end{aligned}$$

Reemplazando se tiene que:

$$\begin{aligned}S\hat{\Sigma}^{-1} &= S\hat{\Psi}^{-1} - \underbrace{S\hat{\Psi}^{-1}\hat{\Gamma}(Id + \hat{\Gamma}^T\hat{\Psi}^{-1}\hat{\Gamma})^{-1}}_{Id}\hat{\Gamma}^T\hat{\Psi}^{-1} \\ &= S\hat{\Psi}^{-1} - \hat{\Gamma}\hat{\Gamma}^T\hat{\Psi}^{-1} \\ &= (S - \hat{\Gamma}\hat{\Gamma}^T)\hat{\Psi}^{-1}\end{aligned}$$

Recordemos que:
* AB = diag(A)B, si B es diagonal.
* Tr(A) = Tr(diag(A))

Finalemente:

$$\begin{aligned}Tr(S\hat{\Sigma}^{-1}) &= Tr\{(S - \hat{\Gamma}\hat{\Gamma}^T)\hat{\Psi}^{-1}\} \\ &= Tr\{\operatorname{diag}(S - \hat{\Gamma}\hat{\Gamma}^T)\hat{\Psi}^{-1}\} \\ &= Tr\{\hat{\Psi}\hat{\Psi}^{-1}\} \\ &= Tr\{Id_{p\times p}\} \\ &= p\end{aligned}$$

3. (25 pts) Sea $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$, donde $\boldsymbol{X}$ es $n \times k$ de rango k, y las filas de $\boldsymbol{E}$ son IID $\mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Sea $\boldsymbol{X}\boldsymbol{B} = \boldsymbol{X}_1\boldsymbol{B}_1 + \boldsymbol{X}_2\boldsymbol{B}_2$, donde $\boldsymbol{X}_2$ es $n \times k_2$. Derive el test de razón de verosimilitudes para probar $H_0 : \boldsymbol{B}_2 = \boldsymbol{0}$

Por construcción se tiene que la Y sigue la ley normal multivariada con la siguiente estructura:

$$T \sim N_{n,p}(X_1B_1 + X_1B_2, I_n \otimes \Sigma)$$

con densidad conjunta dada por la siguiente ecuación:

$$(2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\mathrm{Tr}(Y - X_1B_1 - X_2B_2)^\top\Sigma^{-1}(Y - X_1B_1 - X_2B_2)\right)$$

Por tanto la log verosimilitud viene dada por:

$$\ell(B_1, B_2, \Sigma) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{Tr}\left(\Sigma^{-1}(Y - X_1B_1 - X_2B_2)(Y - X_1B_1 - X_2B_2)^\top\right)$$

En primer lugar trabajaremos en encontrar el estimador bajo H_0 es directo pues se tiene un modelo de regresión. Por ende

$$\begin{aligned}\hat{B}_1 &= \left(X_1^\top X_1\right)^{-1}X_1^\top y\\ \hat{\Sigma} &= \frac{1}{n}\left(Y - X_1\hat{B}_1\right)^\top\left(Y - X_1\hat{B}_1\right)\end{aligned} \tag{1}$$

Luego para H_1 debemos estimar sin restricciones. Note que:

$$\mathrm{d}_{B_1}\ell = -\frac{1}{2}\mathrm{Tr}\left(\Sigma^{-1}\left[(-X_1\mathrm{d}B_1)^\top(Y - X_1B_1 - X_2B_2) - (Y - X_1B_1 - X_2B_2)^\top(X_1\mathrm{d}B_1)\right]\right) = \mathrm{Tr}\left(\Sigma^{-1}(Y - X_1B_1 - X_2B_2)^\top X_1\mathrm{d}B_1\right)$$

Resolviendo $\mathrm{d}_{B_1}\ell = 0$ entonces:

$$\begin{aligned}X_1^\top(Y - X_1B_1 - X_2B_2) = 0 &\Longrightarrow X_1^\top X_1B_1 = X_1^\top Y - X_1\top X_2B_2\\ &\Longrightarrow B_1 = (X_1^\top X_1)^{-1}(X_1^\top Y - X_1^\top X_2B_2)\end{aligned}$$

Es fácil darse cuenta que con el mismo procedimiento se puede obtener que $B_2 = (X_2^\top X_2)^{-1}(X_2^\top Y - X_2^\top X_1B_1)$, y desde la Tarea 1 o visto en clases se puede sacar rápidamente que $\Sigma = \frac{1}{n}(Y - X_1B_1 - X_2B_2)^\top(Y - X_1B_1 - X_2B_2)$. Con lo que se obtiene un sistema de ecuaciones no lineal de estimación.

$$\begin{pmatrix}\tilde{B}_1\\ \tilde{B}_2\\ \tilde{\Sigma}\end{pmatrix} = \begin{pmatrix}(X_1^\top X_1)^{-1}(X_1^\top Y - X_1^\top X_2\tilde{B}_2)\\ (X_2^\top X_2)^{-1}(X_2^\top Y - X_2^\top X_1\tilde{B}_1)\\ \frac{1}{n}(Y - X_1\tilde{B}_1 - X_2\tilde{B}_2)^\top(Y - X_1\tilde{B}_1 - X_2\tilde{B}_2)\end{pmatrix}$$

Que se puede resolver numéricamente. Por simplicidad uno puede definir la función matricial $Q(B_1, B_2) = (Y - X_1B_1 - X_2B_2)^\top(Y - X_1B_1 - X_2B_2)$ y así darse cuenta de varias cosas, por ejemplo:

$$\begin{aligned}\max L(B_1, B_2, \Sigma) &= L(\tilde{B}_1, \tilde{B}_2, \tilde{\Sigma})\\ &= (2\pi)^{-\frac{np}{2}}|\tilde{\Sigma}|\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\left[\frac{1}{n}Q(\tilde{B}_1, \tilde{B}_2)\right]^{-1}Q(\tilde{B}_1, \tilde{B}_2)\right)\right)\\ &= (2\pi)^{-\frac{np}{2}}|\tilde{\Sigma}|\exp\left(-\frac{np}{2}\right)\end{aligned}$$

Y bajo H_0 la verosimilitud toma la forma:

$$\begin{aligned}\max L(B_1, 0, \Sigma) &= L(\hat{B}_1, 0, \hat{\Sigma})\\ &= (2\pi)^{-\frac{np}{2}}|\hat{\Sigma}|\exp\left(-\frac{np}{2}\right)\end{aligned}$$

Por tanto:

$$\Lambda = \frac{\max L(B_1, B_2, \Sigma)}{\max L(B_1, 0, \Sigma)} = \left\{\frac{|\tilde{\Sigma}|}{|\hat{\Sigma}|}\right\}^{\frac{-n}{2}}$$

Conclusión:

$$LR = \frac{|\tilde{\Sigma}|}{|\hat{\Sigma}|}$$

Y como criterio de rechazo es para p-valores (LR) pequeños.