ANALISIS DE COMPONENTES PRINCIPALES (PCA)

Como estudiar datos $x_1, \dots, x_n$, cuya dimensión puede ser grande, no es una tarea inmediata, nos gustaría una forma de representar $x_1, \dots, x_n$ en dimensiones menores. En estadística hemos aprendido que un "buen" resumen de los datos es un promedio, pero una forma de hacerlo más "flexible" sería darle un peso a cada componente de x_i, la cual midiera una "importancia" en el promedio.

$x_1, \dots, x_n$ realizaciones del vect. aleatorio $\mathbb{X}$

$$\alpha' \mathbb{X} = \sum_{i=1}^{p} \alpha_i X_i, \quad \text{donde} \quad \sum_{i=1}^{p} \alpha_i^2 = 1 .$$

La forma en que $\mathbb{X}$ varía es una característica de interés, por lo cual nos planteamos el problema de encontrar $\alpha \in \mathbb{R}^p$, tal que $\sum_{j=1}^{p} \alpha_j^2 = 1$ y $\alpha' \mathbb{X}$ tiene varianza máxima. En otras palabras hay que encontrar $\alpha \in \mathbb{R}^p$ tal que $\alpha' \mathbb{X}$ alcance el máximo

$$\max \left\{ VAR(\alpha' \mathbb{X}) : \alpha \in \mathbb{R}^p \text{ y } \|\alpha\| = 1 \right\}$$

$$= \max \left\{ \alpha' \, VAR(\mathbb{X}) \, \alpha : \alpha \in \mathbb{R}^p \text{ y } \|\alpha\| = 1 \right\}$$

Si regresamos al Teorema C, específicamente al caso en que $\mathbb{B} = \mathbb{I}_p$, es claro que una solución al problema de optimización anterior, es tomar $\alpha = \gamma_1$, donde γ_1 es el vector propio correspondiente al más grande valor propio λ_1 de la matriz de covarianzas $\Sigma \equiv VAR(\mathbb{X})$.

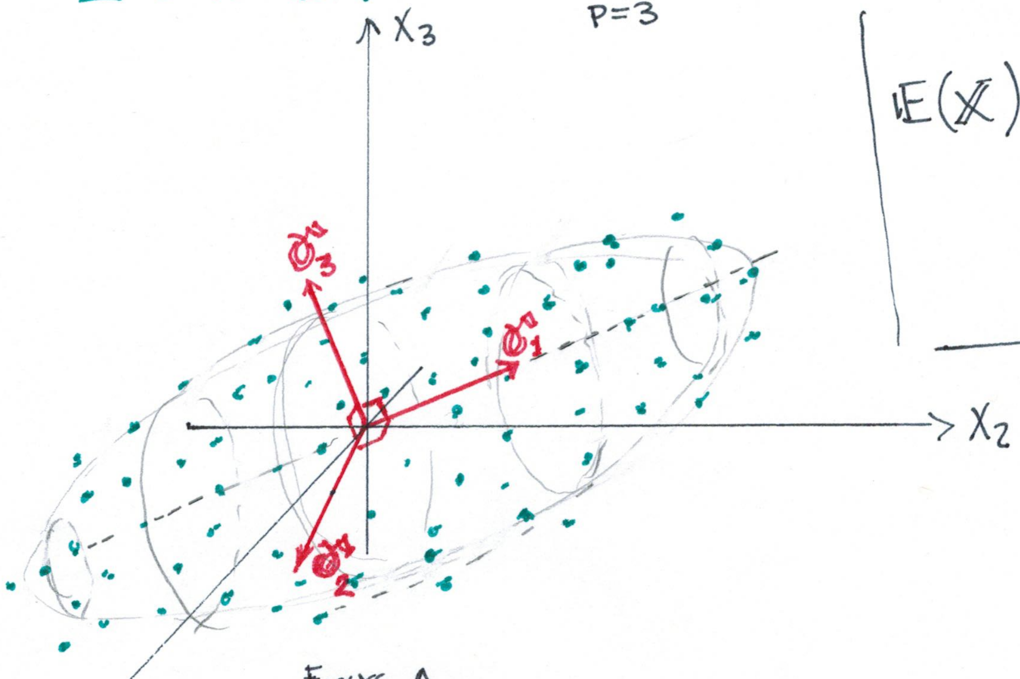


Figura A

$$\mathbb{E}(\mathbb{X}) = \begin{pmatrix}\mathbb{E}(X_1)\\ \mathbb{E}(X_2)\\ \mathbb{E}(X_3)\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

Si los datos tuvieran un comportamiento "ideal"(1) como en el dibujo, la dirección en $\mathbb{R}^3$ en la que existe mayor variabilidad está

(1) ver página 15

(1)

Asumiendo $\mathbb{x}_1, \ldots, \mathbb{x}_n$ (los datos) son una muestra aleatoria del vector aleatorio $\mathbb{X}$ tal que $\mathbb{X} \sim F$, donde la función de distribución multivariada F es tal que:

$$\mathbb{E}(\mathbb{X}) = \mathbb{O} \quad \text{y} \quad VAR(\mathbb{X}) = \Sigma > 0.$$

Además $F : \mathbb{R}^3 \to [0,1]$ con soporte como en el dibujo de la Figura A.

$$\text{Soporte}(F) = \{ \mathbb{x} \in \mathbb{R}^3 : \quad f(\mathbb{x}) > 0 \}$$

f = densidad de F.

dada por γ_1^u Pregunta: ¿Porqué es así? (*)

Por el Teorema C la elección $\alpha = \gamma_1^u$ garantiza que $VAR(\gamma_1^{u\prime} X) = \gamma_1^{u\prime} VAR(X) \gamma_1^u$ alcanza el máximo, el cual vale λ_1. $\left(VAR(\gamma_i^{u\prime} X) = \lambda_1 \right)$.

(PC1) .... $Y_1 \equiv \gamma_1^{u\prime} X$ ← La primera componente principal

Es una combinación lineal "estandarizada" (es decir convexa: $\sum_{j=1}^{p} \gamma_{j1}^2 = 1$; $\gamma_i^{u\prime} = (\gamma_{i1}, \ldots, \gamma_{ip})$) de las componentes de X. Está construída para tener varianza máxima.

(PC2) ... $Y_2 \equiv \gamma_2^{u\prime} X$ ← La segunda componente principal

donde γ_2^u es el vector propio de Σ asociado a $\lambda_2 (\leq \lambda_1)$, entonces $\Sigma \gamma_2^u = \lambda_2 \gamma_2^u$ por lo que $\gamma_2^{u\prime} \Sigma \gamma_2^u = \lambda_2$ $\left(VAR(\gamma_2^{u\prime} X) = \lambda_2 \leq \lambda_1 \right)$.

(PC3) ... $Y_3 \equiv \gamma_3^{u\prime} X$ ← La tercera componente principal

etc ...

Asumiendo que $\Sigma > 0$ y tomando en cuenta que $\gamma_1, \ldots, \gamma_p$ son una base ortonormal en $\mathbb{R}^p$, pensemos en la figura A, $p=3$.

Para un dato x en la nube de la figura A la proyección de x sobre γ_i está dada por

$$\mathbb{P}\text{roy}_{\gamma_i}(x) = (\gamma_i' x) \cdot \gamma_i = y_i \gamma_i \qquad i=1,2,3,$$

pero además

$$(\Phi) \cdots \quad x = \mathbb{P}\text{roy}_{\gamma_1}(x) + \mathbb{P}\text{roy}_{\gamma_2}(x) + \mathbb{P}\text{roy}_{\gamma_3}(x)$$

$$= y_1 \gamma_1 + y_2 \gamma_2 + y_3 \gamma_3,$$

y_i es el peso o el escalamiento que hay que darle al vector γ_i, $i=1,2,3$, para obtener x en la ecuación (Φ), pero además y_i es una muestra o realización de la v.a. $Y_i = \gamma_i' X$.

En particular y_1 es una muestra de Y_1 la primera componente principal

Para responder a la pregunta (*) en la pag 16

Supóngase ahora, que nunca vimos la Figura A, pero que tenemos los datos $x_1, \dots, x_n$. Entonces repetimos el proceso para calcular y_1, y_2 y y_3 (en la ecuación (♠)) para cada x_i; $i=1,2,\dots,n$. Con esto vamos a obtener una muestra de tamaño n para cada una de las componentes Y_1, Y_2 y Y_3, sean $\{y_1^i : i=1,\dots,n\}$, $\{y_2^i : i=1,\dots,n\}$ y $\{y_3^i : i=1,\dots,n\}$ estas muestras. Por construccion $VAR(Y_1) = \lambda_1 \geq \lambda_2 = VAR(Y_2) \geq \lambda_3 = VAR(Y_3)$, entonces es de esperarse que los valores $\{y_1^i\}_{i=1}^n$ se extiendan más (tengan mayor dispersión) que los valores $\{y_2^i\}_{i=1}^n$ y $\{y_3^i\}_{i=1}^n$ y que a su vez los valores $\{y_2^i\}_{i=1}^n$ tengan mayor dispersión que los valores $\{y_3^i\}_{i=1}^n$. Como $\{y_1^i\}_{i=1}^n$ son los escalamientos en la dirección de $\hat{\gamma}_1$, $\{y_2^i\}_{i=1}^n$ son los escalamientos en la dirección de $\hat{\gamma}_2$ y $\{y_3^i\}_{i=1}^n$ son los escalamientos en la dirección de γ''_3, entonces es de esperarse que la configuración de los datos sea como en la figura A.

NOTA: Para datos tales que $\mathbb{X} \sim F$ donde $\mathbb{E}(\mathbb{X}) \neq 0$, tendríamos que dibujar el elipsoide de la figura A en otra localización de $\mathbb{R}^3$. Lo que aparece como el origen $(0,0,0)$ tendría que convertirse en el vector $(\mathbb{E}(X_1), \mathbb{E}(X_2), \mathbb{E}(X_3))$ $p=3$.

En general los datos tendrán una distribución F para la cual la media no es el vector $\mathbb{O}$, es decir $x_1, \dots, x_n$ son una muestra de $\mathbb{X} \sim F$ con $\mathbb{E}(\mathbb{X}) \neq 0$.

En algunos libros se asume $\mathbb{E}(\mathbb{X}) = 0$ porque de no ser así, se puede hacer teoría con el vector aleatorio centrado $\tilde{\mathbb{X}}' = \mathbb{X} - \mathbb{E}(\mathbb{X})$

y en tal caso $\mathbb{E}(\tilde{\mathbb{X}}) = \mathbb{O}$ porque

$$\mathbb{E}(\tilde{\mathbb{X}}) = \mathbb{E}(\mathbb{X} - \mathbb{E}(\mathbb{X})) = \mathbb{E}(\mathbb{X}) - \mathbb{E}(\mathbb{E}(\mathbb{X}))$$

$$= \mathbb{E}(\mathbb{X}) - \mathbb{E}(\mathbb{X}) = \mathbb{O}$$

Vamos a escribir las relaciones (PC1), (PC2) (PC3), etc... que definen las componentes principales de un vector aleatorio $\mathbb{X}$ usando Σ y μ (la matriz de varianzas-covarianzas y el vector de medias) que son características de F, la distribución de $\mathbb{X}$.

$$\mu = \mathbb{E}(\mathbb{X})$$

descomposición de Jordan

de dim. $p \times p$ simétrica y positivo def. $\longrightarrow$ $\Sigma = VAR(\mathbb{X}) = \Gamma \Lambda \Gamma'$

$$\Lambda = DIAG(\lambda_1, \dots, \lambda_p)$$

$$\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)$$

Las relaciones (PC1), (PC2), (PC3), ... se pueden escribir en forma matricial como

$$\mathbb{Y} = \Gamma'(\mathbb{X} - \mu)$$

y en tal caso $\mathbb{E}(\tilde{\mathbb{X}}) = \mathbb{O}$ porque

$$\mathbb{E}(\tilde{\mathbb{X}}) = \mathbb{E}(\mathbb{X} - \mathbb{E}(\mathbb{X})) = \mathbb{E}(\mathbb{X}) - \mathbb{E}(\mathbb{E}(\mathbb{X}))$$

$$= \mathbb{E}(\mathbb{X}) - \mathbb{E}(\mathbb{X}) = \mathbb{O}$$

Vamos a escribir las relaciones (PC1), (PC2) (PC3), etc... que definen las componentes principales de un vector aleatorio $\mathbb{X}$ usando Σ y $\boldsymbol{\mu}$ (la matriz de varianzas-covarianzas y el vector de medias) que son características de F, la distribución de $\mathbb{X}$.

$$\boldsymbol{\mu} = \mathbb{E}(\mathbb{X})$$

descomposición de Jordan

de dim. pxp simétrica y positivo def. $\longrightarrow$ $\Sigma = VAR(\mathbb{X}) = \Gamma \Lambda \Gamma'$

$$\Lambda = DIAG(\lambda_1, \dots, \lambda_p)$$

$$\Gamma = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_p)$$

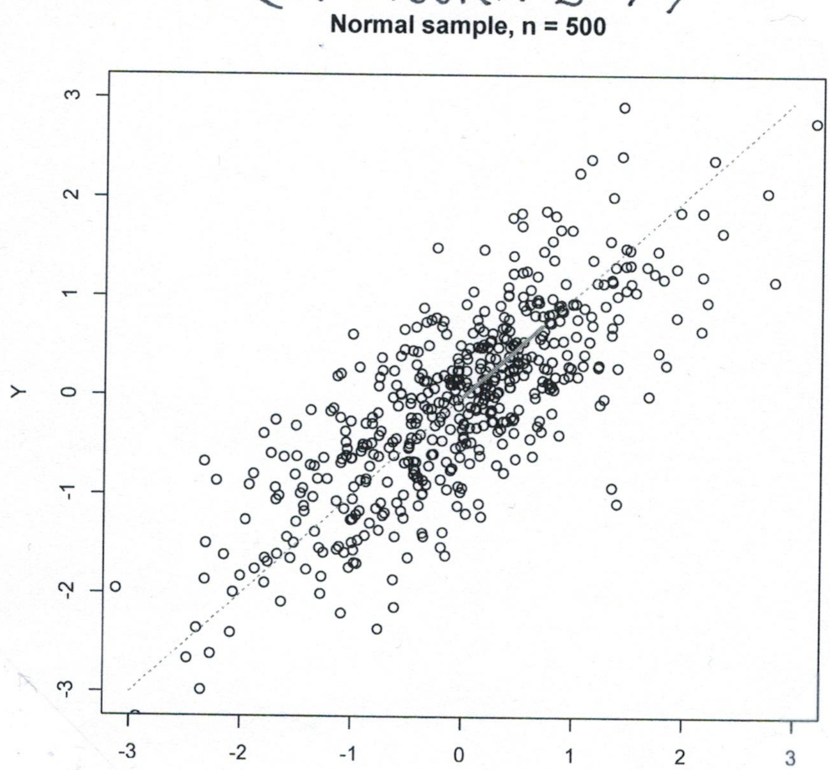
Las relaciones (PC1), (PC2), (PC3), ... se pueden escribir en forma matricial como

$$\mathbb{Y} = \Gamma'(\mathbb{X} - \boldsymbol{\mu}) \quad \dots \text{(CompPrin)}$$

Ejemplo: Supóngase $X \sim N_2(0, \Sigma)$; $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$

con $\rho > 0$

<< FIGURA B >>



Para encontrar los valores propios de Σ tenemos que resolver la ecuación

$$\begin{vmatrix} 1-\lambda & \rho \\ \rho & 1-\lambda \end{vmatrix} = 0 \qquad (1-\lambda)^2 - \rho^2 = 0$$

cuyas soluciones son $\rho_1 = 1+\rho$ y $\rho_2 = 1-\rho$, de donde

$$\Lambda = \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}.$$

El vector propio correspondiente a $\lambda_1 = 1+\rho$ se obtiene de

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (1+\rho)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{o bien}$$

$$\left.\begin{aligned} x_1 + \rho x_2 &= (1+\rho) x_1 \\ \rho x_1 + x_2 &= (1+\rho) x_2 \end{aligned}\right\} \text{ambas implican } x_1 = x_2$$

Todos los vectores $x = \begin{pmatrix} x_1 \\ x_1 \end{pmatrix}$ son propios para Σ, el vector x renormalizado asociado a λ_1 es

$$\gamma_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

La figura muestra γ_1 en color rojo (línea en rojo, gruesa) su dirección coincide con la dirección en la que los datos tienen su mayor variabilidad (línea punteada en rojo). El segundo vector renormalizado (asociado a $\lambda_2 = 1-\rho$) es

$$\gamma_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix},$$

de forma que

$$\Gamma = (\gamma_1, \gamma_2) = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

que satisface

$$\Sigma = \Gamma \Delta \Gamma'$$

<u>NOTA</u>: Si se ajustara una regresión lineal a datos (en donde esto tenga sentido), en general la dirección del vector propio asociado a λ_1 y la pendiente de la línea recta ajustada vía mínimos cuadrados serán diferentes. La razón es que la estimación vía mínimos cuadrados tiene por objetivo minimizar distancias (errores) verticales, este objetivo es

diferente al que se tiene al maximizar una forma cuadrática seleccionando un vector propio de la matriz de varianzas-covarianzas de los datos.

La transformacion de componentes principales en la ecuación (CompPrin), para el vector alectorio $\mathbb{X}$ queda como

$$\mathbb{Y} = \Gamma'(\mathbb{X} - \mu) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \mathbb{X},$$

es decir

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix}.$$

La primera componente principal es

$$Y_1 = \frac{1}{\sqrt{2}}(X_1 + X_2).$$

La segunda componente principal es

$$Y_2 = \frac{1}{\sqrt{2}}(X_1 + X_2).$$

A continuación, verificaremos que, como se establece en el Teorema C, $VAR(\mathbb{Y}_i) = \lambda_i$.

$$VAR(\mathbb{Y}_1) = VAR\left\{\frac{1}{\sqrt{2}}(X_1 + X_2)\right\} = \frac{1}{2} VAR(X_1 + X_2)$$

$$= \frac{1}{2}\left\{VAR(X_1) + VAR(X_2) + 2\,COV(X_1, X_2)\right\}$$

$$= \frac{1}{2}\left\{1 + 1 + 2\rho\right\} = 1 + \rho = \lambda_1$$

$$VAR(\mathbb{Y}_2) = VAR\left\{\frac{1}{\sqrt{2}}(X_1 - X_2)\right\} = \frac{1}{2} VAR(X_1 - X_2)$$

$$= \frac{1}{2}(1 + 1 - 2\rho) = 1 - \rho = \lambda_2$$

Teorema: Dado un vector aleatorio $\mathbb{X}$ tal que $\mathbb{E}(\mathbb{X}) = \mathbb{\mu}$ y $VAR(\mathbb{X}) = \Sigma$, sea $\mathbb{Y} = \Gamma'(\mathbb{X} - \mathbb{\mu})$ la transformación de Componentes Principales. Entonces:

(1) $\mathbb{E}(\mathbb{Y}_j) = 0 \; ; \; \forall j = 1, 2, \ldots, p.$

(2) $VAR(\mathbb{Y}_j) = \lambda_j \; ; \; \forall j = 1, 2, \ldots, p.$

(3) $COV(\mathbb{Y}_i, \mathbb{Y}_j) = 0 \; ; \; \forall i \neq j \;\; i, j \in \{1, \ldots, p\}.$

(4) $\mathrm{VAR}(Y_1) \geq \mathrm{VAR}(Y_2) \geq \cdots \geq \mathrm{VAR}(Y_p) \geq 0.$

(5) $\sum_{j=1}^{p} \mathrm{VAR}(Y_j) = \mathrm{traza}(\Sigma)$

(6) $\prod_{j=1}^{p} \mathrm{VAR}(Y_j) = |\Sigma|.$

Dem

1) $\mathbb{E}[Y_j] \underset{\uparrow}{=} \gamma_j' \mathbb{E}(X - \mu) = \gamma_j'(\mu - \mu) = 0$

$Y_j = \gamma_j'(X - \mu)$

2) $\mathrm{VAR}(Y_j) = \mathrm{VAR}(\gamma_j'(X-\mu)) = \gamma_j' \mathrm{VAR}(X-\mu)\gamma_j$

$= \gamma_j' \mathrm{VAR}(X)\gamma_j = \gamma_j' \Sigma \gamma_j$

Pero por el Teorema C (cuando $B = I$)

Si $\lambda_1, \dots, \lambda_p$ son los valores propios de $B^{-1}A = A$

entonces $\lambda_p \leq \frac{x'Ax}{x'x} \leq \lambda_1 \quad \forall x \neq 0;\ x \in \mathbb{R}^p.$

En particular si $\|x\| = 1$ $\lambda_p \leq x'Ax \leq \lambda_1$.

Si γ_j es el vector propio (normalizado) de $A = \Sigma$, asociado a λ_j entonces

$$\Sigma \gamma_j = \lambda_j \gamma_j$$

Lo que implica que $\gamma_j' \Sigma \gamma_j = \lambda_j \gamma_j' \gamma = \lambda_j$

$$\therefore \quad VAR(Y_j) = \lambda_j$$

3) $COV(Y_i, Y_j) = \mathbb{E}\Big[\big(Y_i - \mathbb{E}(Y_i)\big)\big(Y_j - \mathbb{E}(Y_j)\big)\Big]$

$$\underset{(1)}{=} \mathbb{E}[Y_i Y_j] = \mathbb{E}[\gamma_i'(X-\mu)(X-\mu)'\gamma_j]$$

$$= \gamma_i' \, \mathbb{E}\big[(X-\mu)(X-\mu)'\big] \gamma_j$$

$$= \gamma_i' \Sigma \gamma_j \underset{\text{Teor(DJ)}}{=} \gamma_i' \Gamma \Lambda \Gamma' \gamma_j .$$

Pero las columnas de Γ (los vectores propios de Σ) son ortogonales y con norma 1

$$\gamma_i' \Gamma = (0, 0, \ldots, \underset{\substack{\uparrow \\ \text{lugar } i}}{1}, 0, \ldots, 0)$$

$$\Gamma' \gamma_j = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \\ \\ \\ \text{— lugar } j \\ \\ \\ \\ \end{array} \quad , \text{ entonces}$$

$$\gamma_i' \Gamma \Lambda \Gamma' \gamma_j = \begin{cases} 0 & i \neq j \\ \lambda_i & i = j \end{cases} \qquad \text{de aquí se siguen (3) y (2)}$$

(4) Se sigue de que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$
y se tiene (2)

(5) Es resultado del siguiente teorema
para matrices

Teor: Sea A una matriz de dimensiones $p \times p$ con valores propios $\lambda_1, \dots, \lambda_p$, entonces

$$\text{traza}(A) = \sum_{i=1}^{p} \lambda_i$$

(+) véase por ejemplo: Magnus y Neudecker
Matrix differential calulus with Applications in Statistics and Econometrics Wiley (1999).

(6) Es resultado del siguiente resultado

Teor: Sea A una matriz de dimensiones $p \times p$ con valores propios $\lambda_1, \dots, \lambda_p$, entonces

$$|A| = \prod_{j=1}^{p} \lambda_j$$

véase (+)

Hasta ahora se han estudiado las componentes principales para **un vector** aleatorio $\mathbb{X}$ cuando sus primeros dos momentos existen y asumiendo $\Sigma > 0$. ¿Cómo se pueden construir estimadores de $\mathbb{Y}_1, \ldots, \mathbb{Y}_p$ usando los datos $\mathbb{x}_1, \ldots, \mathbb{x}_n$?

En la práctica se sugiere reemplazar las caracteristicas poblacionales $\mu \ (= \mathbb{E}(\mathbb{X}))$ y $\Sigma \ (= VAR(\mathbb{X}))$ por sus estimadores muestrales

$$\hat{\mu} \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbb{x}_i = \overline{\mathbb{x}}_{(n)} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} x_{i1} \\ \frac{1}{n} \sum_{i=1}^{n} x_{i2} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^{n} x_{ip} \end{pmatrix}$$

$\overline{x}_{(n)}$ has j-th row given by

$$\overline{x}_{\cdot j} \equiv \frac{1}{n} \sum_{i=1}^{n} x_{ij}$$

el promedio (sobre todos los individuos) de la j-ésima variable.

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \overbrace{(\mathbb{x}_i - \overline{\mathbb{x}}_{(n)})}^{p \times 1} \overbrace{(\mathbb{x}_i - \overline{x}_{(n)})'}^{1 \times p}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} x_{i1} - \overline{x}_{\cdot 1} \\ x_{i2} - \overline{x}_{\cdot 2} \\ \vdots \\ x_{ip} - \overline{x}_{\cdot p} \end{pmatrix} (x_{i1} - \overline{x}_{\cdot 1}, x_{i2} - \overline{x}_{\cdot 2}, \ldots, x_{ip} - \overline{x}_{\cdot p})$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} (x_{i1}-\bar{x}_{\cdot 1})^2 & (x_{i1}-\bar{x}_{\cdot 1})(x_{i2}-\bar{x}_{\cdot 2}) & \cdots & (x_{i1}-\bar{x}_{\cdot 1})(x_{ip}-\bar{x}_{\cdot p}) \\ (x_{i2}-\bar{x}_{\cdot 2})(x_{i1}-\bar{x}_{\cdot 1}) & (x_{i2}-\bar{x}_{\cdot 2})^2 & \cdots & (x_{i2}-\bar{x}_{\cdot 2})(x_{ip}-\bar{x}_{\cdot p}) \\ \vdots & & \ddots & \vdots \\ (x_{ip}-\bar{x}_{\cdot p})(x_{i1}-\bar{x}_{\cdot 1}) & \cdots & & (x_{ip}-\bar{x}_{\cdot p})^2 \end{pmatrix}$$

$\hat{\Sigma}$ tiene entrada j,k dada por $\longrightarrow$

$$\hat{\Sigma}_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x}_{\cdot j})(x_{ik}-\bar{x}_{\cdot k}) \qquad j=1,2,\dots,p, \quad k=1,2,\dots,p.$$

Nota: En algunas referencias se considera la "corrección de Bessel", definiendo $\hat{\Sigma}$ como la matriz de dimensiones $p \times p$ con entrada j,k dada por:

$$\hat{\Sigma}^{(c)}_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_{\cdot j})(x_{ik}-\bar{x}_{\cdot k}) \cdots (A)$$

Esta corrección tiene el fin de hacer de de (A) un estimador insesgado para $\mathrm{cov}(X_j, X_k)$. $\left(\hat{\Sigma}^{(c)} = \frac{n}{n-1}\hat{\Sigma} \right.$ es estimador insesgado para $\left.\Sigma\right)$

Sea $\mathcal{X}$ la matriz de datos dada por:

$$\mathcal{X} \equiv \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \text{de dimensiones } n \times p .$$

$\mathcal{X}$ contiene renglones dados por los individuos de la muestra (el i-ésimo renglón de $\mathcal{X}$ es x_i').

$$\mathbb{1}_n \equiv \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1} .$$

$$\overline{x}_{(n)} = \frac{1}{n} \mathcal{X}' \mathbb{1}_n .$$

(i) $\boxed{\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\overline{x}_{\cdot j})(x_{ik}-\overline{x}_{\cdot k}) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} x_{ij}x_{ik}\right) - \overline{x}_{\cdot j}\,\overline{x}_{\cdot k} \\ &= \text{entrada } j,k \text{ en } \hat{\Sigma} \end{aligned}}$

Con estas notaciones, podemos escribir

$$(B) \ldots \quad \hat{\Sigma} \overset{(i)}{=} \frac{1}{n}\mathcal{X}'\mathcal{X} - \overline{x}_{(n)}\overline{x}_{(n)}' = \frac{1}{n}\left(\mathcal{X}'\mathcal{X} - \frac{1}{n}\mathcal{X}'\mathbb{1}_n\mathbb{1}_n'\mathcal{X}\right).$$

Usando la "matriz de centrado"

$$\mathcal{H} = \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n', \qquad (\mathbb{I}_n \equiv \text{matriz identidad } n \times n)$$

también podemos escribir

$$(C) \ldots \quad \hat{\Sigma} = \frac{1}{n}\mathcal{X}'\mathcal{H}\mathcal{X} .$$

Claramente $\mathcal{H}$ es simétrica:

$$\mathcal{H}' = \left(\mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)' = \mathbb{I}_n' - \frac{1}{n}\left(\mathbb{1}_n\mathbb{1}_n'\right)'$$

$$= \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' = \mathcal{H},$$

e idempotente:

$$\mathcal{H}\mathcal{H} = \left(\mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)\left(\mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)$$

$$= \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' + \left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)\left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right).$$

Pero $\left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)\left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right) = \frac{1}{n^2}\left(\mathbb{1}_n\mathbb{1}_n'\right)\left(\mathbb{1}_n\mathbb{1}_n'\right)$

$$= \frac{1}{n^2}\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}}_{n\times n}\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}}_{n\times n} = \frac{1}{n^2}\begin{pmatrix} n & n & \cdots & n \\ n & n & \cdots & n \\ \vdots & \vdots & & \vdots \\ n & n & & n \end{pmatrix}$$

$$= \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} = \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'$$

luego

$$\mathcal{H}\mathcal{H} = \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' + \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'$$

$$= \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' = \mathcal{H}$$

Usaremos la ecuación (C) y las propiedades de $\mathcal{H}$ para ver que $\hat{\Sigma}$ es positivo semi-definida:

Sea $\mathbf{a} \in \mathbb{R}^p$, tal que $\mathbf{a} \neq \mathbb{O}$, entonces por (C)

$$\mathbf{a}'\hat{\Sigma}\mathbf{a} = \frac{1}{n}\mathbf{a}'\mathcal{X}'\mathcal{H}\mathcal{X}\mathbf{a} = \frac{1}{n}\mathbf{a}'\mathcal{X}'\mathcal{H}^2\mathcal{X}\mathbf{a}$$

$\mathcal{H} = \mathcal{H}\mathcal{H}$
$= \mathcal{H}'\mathcal{H}$

$$= \frac{1}{n}\mathbf{a}'\mathcal{X}'\mathcal{H}'\mathcal{H}\mathcal{X}\mathbf{a} = \frac{1}{n}\mathbb{b}'\mathbb{b},$$

donde $\mathbb{b} \equiv \mathcal{H}\mathcal{X}\mathbf{a}$ es un vector de dimensiones $n \times 1$.

Pero como $\mathbb{b}'\mathbb{b} = \sum_{i=1}^{n} b_i^2 \geq 0$, entonces

$$\mathbf{a}'\hat{\Sigma}\mathbf{a} \geq 0 \quad .$$

Usaremos $\hat{\mu}$ y $\hat{\Sigma}$ para calcular estimadores de $\gamma_1, .., \gamma_p$. Para este fin, usaremos la descomposición de Jordan para $\hat{\Sigma}$

$$\hat{\Sigma} = \mathcal{G}\mathcal{L}\mathcal{G}' \quad \ldots\ldots \text{(D)}$$

$\mathbb{L}$ es una matriz diagonal de dimensiones $p \times p$ y sobre su diagonal tiene los valores propios $l_1, \ldots, l_p$ de la matriz $\hat{\Sigma}$.

$\mathbb{G}$ es una matriz de dimensión $p \times p$ cuyas columnas son los vectores propios $g_1, \ldots g_p$ de $\hat{\Sigma}$

De la ecuación (CompPrin) en la página 20

$$(\mathbb{Y}_1, \ldots, \mathbb{Y}_p) = (\mathbb{X} - \mu)' \Gamma$$

Para la primera observación en la muestra $\mathbb{x}_1$ y reemplazando cantidades muestrales

$$(y_{11}, y_{21}, \ldots, y_{p1}) = (\mathbb{x}_1 - \bar{\mathbb{x}}_{(n)})' \mathbb{G}$$

Para $\mathbb{x}_2$

$$(y_{12}, y_{22}, \ldots, y_{p2}) = (\mathbb{x}_2 - \bar{\mathbb{x}}_{(n)})' \mathbb{G}$$

$$\vdots$$

Para $\mathbb{x}_n$

$$(y_{1n}, y_{2,n}, \ldots, y_{p,n}) = (\mathbb{x}_n - \bar{x}_{(n)})' \mathbb{G}$$

En el lado derecho escribimos todos los términos

$$\begin{pmatrix} x_{11}-\bar{x}_{\bullet 1} & x_{12}-\bar{x}_{\bullet 2}, \ldots, & x_{1p}-\bar{x}_{\bullet p} \\ x_{21}-\bar{x}_{\bullet 1} & x_{22}-\bar{x}_{\bullet 2}, \ldots, & x_{2p}-\bar{x}_{\bullet p} \\ \vdots & \vdots & \vdots \\ x_{n1}-\bar{x}_{\bullet 1} & x_{n2}-\bar{x}_{\bullet 2}, \ldots, & x_{np}-\bar{x}_{\bullet p} \end{pmatrix} \mathbb{G}$$

$$= \left[\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} - \begin{pmatrix} \bar{x}_{\bullet 1} & \bar{x}_{\bullet 2}, \ldots, & \bar{x}_{\bullet p} \\ \bar{x}_{\bullet 1} & \bar{x}_{\bullet 2}, \ldots, & \bar{x}_{\bullet p} \\ \vdots & \vdots & \vdots \\ \bar{x}_{\bullet 1} & \bar{x}_{\bullet 2} \cdots & \bar{x}_{\bullet p} \end{pmatrix} \right] \mathbb{G}$$

$$= \left(\mathcal{X} - \mathbb{1}_n \overline{\mathbf{x}}_{(n)}' \right) \mathbb{G} .$$

En el lado izquierdo tenemos

$$\mathbb{Y} \equiv \begin{pmatrix} y_{11} & y_{21} & \cdots & y_{p1} \\ y_{12} & y_{22} & \cdots & y_{p2} \\ \vdots & \vdots & & \vdots \\ y_{1n} & y_{2n} & \cdots & y_{pn} \end{pmatrix}$$

$\leftarrow$ La columna j tiene n observaciones de la j-ésima componente principal $\mathbb{Y}_j$

$$\mathbb{Y} = \left(\mathcal{X} - \mathbb{1}_n \overline{\mathbf{x}}_{(n)}' \right) \mathbb{G} \quad \cdots\cdots \text{(CP)}$$

En forma análoga a como se desarrollaron las igualdades (B) y (C) para $\hat{\Sigma}$ y $\mathcal{X}$, podemos probar que si $\Sigma_{\mathbb{Y}} = VAR(\mathbb{Y})$ ($\mathbb{Y}$ = vector de componentes principales definido en la ecuación (CompPrin) página 20), entonces

$$\hat{\Sigma}_{\mathbb{Y}} = \frac{1}{n} \mathcal{Y}' \mathcal{H} \mathcal{Y},$$

pero usando la definición de la matriz de centrado $\mathcal{H}$ y la forma de la matriz de componentes principales $\mathcal{Y}$ en la ecuación (CP)

$$\hat{\Sigma}_{\mathbb{Y}} = \frac{1}{n} \mathcal{G}' (\mathcal{X} - \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)})' \mathcal{H} (\mathcal{X} - \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)}) \mathcal{G}.$$

Pero $\mathcal{H}$ es tal que

$$\mathcal{H} \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)} = (\mathbb{I}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}'_n) \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)}$$

$$= \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)} - \frac{1}{n} \mathbb{1}_n (\mathbb{1}'_n \mathbb{1}_n) \overline{x}'_{\mathcal{G}(n)}$$

$$= \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)} - \frac{1}{n} \mathbb{1}_n (n) \overline{x}'_{\mathcal{G}(n)}$$

$$= \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)} - \mathbb{1}_n \overline{x}'_{\mathcal{G}(n)} = \mathbb{O} \leftarrow \text{matriz de } n \times p \text{ con entradas} = 0$$

y lo anterior implica que

matriz de pxn con entradas = 0 $\to$ $$\mathbb{O} = (\mathbb{1}_n \bar{x}'_{(n)})' \mathcal{H}' = (\mathbb{1}_n \bar{x}'_{(n)})' \mathcal{H}$$

(por simetría de $\mathcal{H}$).

Entonces

$$\hat{\Sigma}_{\mathbb{Y}} = \frac{1}{n} \mathcal{G}'(\mathcal{X})' \mathcal{H} \mathcal{X} \mathcal{G} = \mathcal{G}' \hat{\Sigma} \mathcal{G}$$

$$\overset{(D)}{=} \mathcal{G}' \mathcal{G} \mathcal{L} \mathcal{G}' \mathcal{G} = \mathcal{L}$$

$\therefore$ $\hat{\Sigma}_{\mathbb{Y}}$ es una matriz diagonal que sobre su diagonal tiene los elementos $l_1, \dots, l_p$, los valores propios de $\hat{\Sigma}$.

Como los elementos sobre la diagonal de $\hat{\Sigma}_{\mathbb{Y}}$ deben ser $\widehat{VAR}(\mathbb{Y}_1), \widehat{VAR}(\mathbb{Y}_2), \dots, \widehat{VAR}(\mathbb{Y}_p)$ entonces

$$\widehat{VAR}(\mathbb{Y}_i) = l_i \qquad i = 1, 2, \dots, p.$$