y en tal caso $\mathbb{E}(\tilde{\mathbb{X}}) = \mathbb{0}$ porque

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

$$= \mathbb{E}(\mathbb{X}) - \mathbb{E}(\mathbb{X}) = \mathbb{0}$$

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

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

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

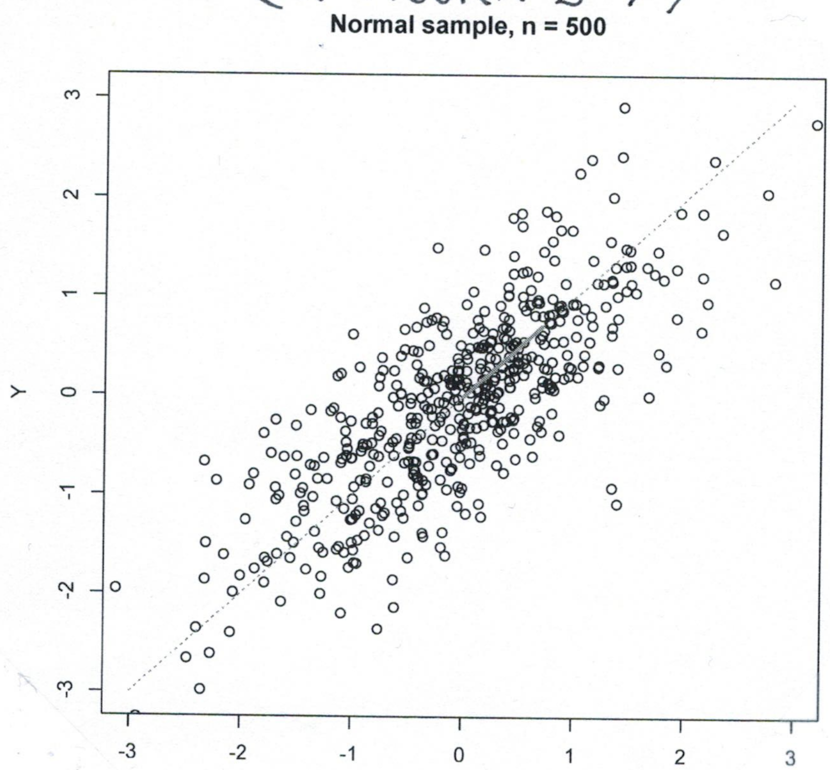
Las relaciones (PC1), (PC2), (PC3), ... se pueden escribir en forma matricial como

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

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

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

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

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

La figura muestra γ_1^{α} en color rojo (línea en rojo, gruesa) su dirección coincide con la dirección en la que los datos tienen su mayor variabilidad (linea 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 aleatorio $\mathbb{X}$ queda como

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

es decir

$$\begin{pmatrix} \mathbb{Y}_1 \\ \mathbb{Y}_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix}.$$

La primera componente principal es

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

La segunda componente principal es

$$\mathbb{Y}_2 = \frac{1}{\sqrt{2}}(X_1 + X_2).$$