MAT-269: Distribución Wishart

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Definición 1 (Distribución Wishart):

Si $\boldsymbol{Q} = \boldsymbol{Z}^\top \boldsymbol{Z}$, donde la matriz $n \times p$ satisface que $\boldsymbol{Z} \sim \mathsf{N}_{n,p}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$. Entonces se dice que $\boldsymbol{Q}$ tiene distribución Wishart con n grados de libertad y matriz de covarianza $\boldsymbol{\Sigma}$, en cuyo caso escribimos $\boldsymbol{Q} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$.

Observación:

Note que $\boldsymbol{Q} = \boldsymbol{Z}^\top \boldsymbol{Z}$, de ahí que podemos escribir la matriz de covarianza muestral $\boldsymbol{S}$ como:

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{Q} = \frac{1}{n-1}\boldsymbol{Z}^\top \boldsymbol{Z} = \widetilde{\boldsymbol{Z}}^\top \widetilde{\boldsymbol{Z}},$$

donde

$$\widetilde{\boldsymbol{Z}} = \frac{1}{\sqrt{n-1}}\boldsymbol{Z} \sim \mathsf{N}_{n,p}\Big(\boldsymbol{0}, \boldsymbol{I}_n \otimes \frac{1}{n-1}\boldsymbol{\Sigma}\Big).$$

Es decir,

$$\boldsymbol{S} \sim \mathsf{W}_p\Big(n, \frac{1}{n-1}\boldsymbol{\Sigma}\Big).$$

Definición 2 (Densidad Wishart):

Sea $\boldsymbol{A}$ matriz simétrica $k \times k$ y sea $\boldsymbol{\Sigma}$ matriz definida positiva $k \times k$. Si n es un entero tal que $n \geq k$. Entonces se dice que $\boldsymbol{A}$ tiene distribución Wishart no singular con n grados de libertad si la densidad conjunta de los $k(k+1)/2$ elementos distintos de $\boldsymbol{A}$ es dada por

$$f(\boldsymbol{A}) = \frac{1}{2^{nk}\Gamma_k(n/2)} |\boldsymbol{\Sigma}|^{-n/2} \exp\big(-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\boldsymbol{A}\big) |\boldsymbol{A}|^{(n-k-1)/2},$$

donde

$$\Gamma_k(a) = \int_{A>0} e^{-\operatorname{tr} \boldsymbol{A}} |\boldsymbol{A}|^{a-(k+1)/2} \, \mathrm{d}\boldsymbol{A},$$

es llamada función gamma multivariada y escribimos $\boldsymbol{A} \sim \mathsf{W}_k(n, \boldsymbol{\Sigma})$.

Observación:

Suponga que $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ son vectores aleatorios IID desde $\mathsf{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. Entonces

$$\boldsymbol{A} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^\top \sim \mathsf{W}_p(n, \boldsymbol{\Sigma}).$$

Si $\boldsymbol{\Sigma} > 0$ y $n \geq p$ entonces $\boldsymbol{A} > 0$ con probabilidad 1. Sino, por definición de $\boldsymbol{A}$, tendremos $\boldsymbol{A} \geq 0$ con probabilidad no nula de que $|\boldsymbol{A}| = 0$, en cuyo caso decimos que $\boldsymbol{A}$ tiene distribución Wishart singular.

Resultado 1 (Función característica)

Si $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ entonces la función característica de las $p(p+1)/2$ variables a_{ij}, $i, j = 1, \dots, p$ es:[1]

$$\varphi(\boldsymbol{\Theta}) = \mathsf{E}\left\{ \exp\left(i \sum_{j \leq k}^{p} \theta_{jk} a_{jk} \right) \right\} = |\boldsymbol{I}_p - i\boldsymbol{\Gamma}\boldsymbol{\Sigma}|^{-n/2},$$

donde $\boldsymbol{\Gamma} = (\gamma_{ij})$, para $i, j = 1, \dots, p$ con $\gamma_{ij} = (1 + \delta_{ij})\theta_{ij}$ ($\boldsymbol{\Theta} = (\theta_{ij})$ es simétrica) y δ_{ij} es el delta de Kronecker

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

[1] Función característica de $\chi^2(n) : \varphi(t) = (1 - 2it)^{-n/2}$.

Demostración:

La función característica $\varphi(\mathbf{\Theta})$ puede ser escrita como

$$\begin{aligned}\varphi(\mathbf{\Theta}) &= \mathsf{E}\left\{\exp\left(\frac{i}{2}\sum_{j,k=1}^{p}(1+\delta_{jk})\theta_{jk}a_{jk}\right)\right\}\\ &= \mathsf{E}\left\{\exp\left(\tfrac{i}{2}\operatorname{tr}\boldsymbol{A}\mathbf{\Gamma}\right)\right\}.\end{aligned}$$

Considere los siguientes casos:

(a) Suponga que n es un entero positivo. En este caso podemos escribir

$$\boldsymbol{A} = \boldsymbol{Z}^\top\boldsymbol{Z} = \sum_{j=1}^{n}\boldsymbol{z}_j\boldsymbol{z}_j^\top, \qquad \boldsymbol{z}_j \overset{\mathsf{IID}}{\sim} \mathsf{N}_p(\mathbf{0},\mathbf{\Sigma}),$$

o bien $\boldsymbol{Z} \sim \mathsf{N}_{n,p}(\mathbf{0}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$, luego

$$\varphi(\mathbf{\Theta}) = \mathsf{E}\left\{\exp\left(\tfrac{i}{2}\operatorname{tr}\boldsymbol{Z}^\top\boldsymbol{Z}\mathbf{\Gamma}\right)\right\}.$$

$$\begin{aligned}\varphi(\mathbf{\Theta}) &= \mathsf{E}\left\{\exp\left(\tfrac{i}{2}\operatorname{tr}\boldsymbol{Z}^\top\boldsymbol{Z}\mathbf{\Gamma}\right)\right\} = \mathsf{E}\left\{\exp\left(\tfrac{i}{2}\operatorname{tr}\textstyle\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^\top\mathbf{\Gamma}\right)\right\}\\
&= \mathsf{E}\left\{\exp\left(\frac{i}{2}\sum_{j=1}^n \operatorname{tr}\boldsymbol{z}_j\boldsymbol{z}_j^\top\mathbf{\Gamma}\right)\right\} = \mathsf{E}\left\{\exp\left(\frac{i}{2}\sum_{j=1}^n \boldsymbol{z}_j^\top\mathbf{\Gamma}\boldsymbol{z}_j\right)\right\}\\
&\overset{\mathsf{IND}}{=} \prod_{j=1}^n \mathsf{E}\left\{\exp\left(\tfrac{i}{2}\boldsymbol{z}_j\mathbf{\Gamma}\boldsymbol{z}_j\right)\right\} \overset{\mathsf{ID}}{=} \left(\mathsf{E}\left(\tfrac{i}{2}\boldsymbol{z}_1^\top\mathbf{\Gamma}\boldsymbol{z}_1\right)\right\}\right)^n.\end{aligned}$$

Haciendo $\boldsymbol{x} = \mathbf{\Sigma}^{-1/2}\boldsymbol{z}_1$, tenemos $\boldsymbol{x} \sim \mathsf{N}_p(\mathbf{0}, \boldsymbol{I})$, y

$$\varphi(\mathbf{\Theta}) = \left(\mathsf{E}\left\{\exp\left(\tfrac{i}{2}\boldsymbol{x}^\top\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}\mathbf{\Sigma}^{1/2}\boldsymbol{x}\right)\right\}\right)^n,$$

como $\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}\mathbf{\Sigma}^{1/2}$ es matriz simétrica, entonces existe una matriz ortogonal $\boldsymbol{H}$ tal que

$$\boldsymbol{H}\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}\mathbf{\Sigma}^{1/2}\boldsymbol{H}^\top = \mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_p),$$

con $\lambda_1, \dots, \lambda_p$ los valores propios de $\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}\mathbf{\Sigma}^{1/2}$. Haciendo $\boldsymbol{u} = \boldsymbol{H}\boldsymbol{x}$, sigue que $\boldsymbol{u} \sim \mathsf{N}_p(\mathbf{0}, \boldsymbol{I})$.

De este modo,

$$\begin{aligned}\varphi(\boldsymbol{\Theta}) &= \big(\mathsf{E}\,\{\exp\big(\tfrac{i}{2}\boldsymbol{u}^\top\boldsymbol{\Lambda}\boldsymbol{u}\big)\}\big)^n = \Big(\mathsf{E}\,\Big\{\exp\Big(\tfrac{i}{2}\textstyle\sum_{j=1}^p \lambda_j u_j^2\Big)\Big\}\Big)^n \\ &= \Big(\prod_{j=1}^p \mathsf{E}\,\{\exp\big(\tfrac{i}{2}\lambda_j u_j^2\big)\}\Big)^n,\end{aligned}$$

como cada $u_j^2 \sim \chi^2(1)$ independientes, para $j = 1, \dots, p$. Tenemos

$$\varphi(\boldsymbol{\Theta}) = \prod_{j=1}^p (1 - i\lambda_j)^{-n/2}.$$

Note que

$$\begin{aligned}\prod_{j=1}^p (1 - i\lambda_j) &= \det(\boldsymbol{I}_p - i\boldsymbol{\Lambda}) = \det(\boldsymbol{I}_p - i\boldsymbol{H}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{1/2}\boldsymbol{H}^\top) \\ &= \det(\boldsymbol{I}_p - i\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{1/2}) = \det(\boldsymbol{I}_p - i\boldsymbol{\Gamma}\boldsymbol{\Sigma}).\end{aligned}$$

(b) Suponga que n es un real tal que $n > p - 1$. Entonces $\boldsymbol{A}$ tiene densidad Wishart y por tanto,

$$\begin{aligned}\varphi(\boldsymbol{\Theta}) &= \mathsf{E}\left\{\exp\left(\tfrac{i}{2}\operatorname{tr}\boldsymbol{A\Gamma}\right)\right\}\\ &= \frac{1}{2^{np/2}\Gamma_p(n/2)}|\boldsymbol{\Sigma}|^{-n/2}\int_{A>0}\exp\left\{\operatorname{tr}\left(-\tfrac{1}{2}\boldsymbol{A}(\boldsymbol{\Sigma}^{-1} - i\boldsymbol{\Gamma})\right)\right\}|\boldsymbol{A}|^{(n-p-1)/2}\,\mathrm{d}\boldsymbol{A}\end{aligned}$$

Sabemos que

$$\int_{A>0}\exp\left\{\operatorname{tr}\left(-\tfrac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)\right\}|\boldsymbol{A}|^{a-(r+1)/2}\,\mathrm{d}\boldsymbol{A} = \Gamma_r(a)|\boldsymbol{\Sigma}|^{a}2^{ra}.$$

De este modo,

$$\begin{aligned}\int_{A>0}\exp\left\{\operatorname{tr}\left(-\tfrac{1}{2}\boldsymbol{A}(\boldsymbol{\Sigma}^{-1} - i\boldsymbol{\Gamma})\right)\right\}|\boldsymbol{A}|^{(n-p-1)/2}\,\mathrm{d}\boldsymbol{A}\\ = \Gamma_p(n/2)2^{np/2}|(\boldsymbol{\Sigma}^{-1} - i\boldsymbol{\Gamma})^{-1}|^{n/2}.\end{aligned}$$

Es decir,

$$\varphi(\boldsymbol{\Theta}) = |\boldsymbol{\Sigma}^{-1} - i\boldsymbol{\Gamma}|^{-n/2}|\boldsymbol{\Sigma}|^{-n/2} = |(\boldsymbol{\Sigma}^{-1} - i\boldsymbol{\Gamma})\boldsymbol{\Sigma}|^{-n/2} = |\boldsymbol{I} - i\boldsymbol{\Gamma\Sigma}|^{-n/2},$$

como deseado.

Observación:

Los momentos de una matriz Wishart $\boldsymbol{A}$ pueden ser hallados desde la función característica mediante diferenciación. (Tarea)

Alternativamente podemos considerar $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ muestra aleatoria desde $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, es decir

$$\mathsf{E}(\boldsymbol{y}_i) = \boldsymbol{\mu}, \qquad \mathsf{Cov}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}, \qquad i = 1, \dots, n.$$

Sabemos que

$$\boldsymbol{Q} = \sum_{i=1}^{n} (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^\top = \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})^\top - n(\overline{\boldsymbol{y}} - \boldsymbol{\mu})(\overline{\boldsymbol{y}} - \boldsymbol{\mu})^\top,$$

como $\overline{\boldsymbol{y}} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$, tenemos que

$$\begin{aligned}
\mathsf{E}(\boldsymbol{Q}) &= \sum_{i=1}^{n} \mathsf{E}\{(\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})^\top\} - n\,\mathsf{E}\{(\overline{\boldsymbol{y}} - \boldsymbol{\mu})(\overline{\boldsymbol{y}} - \boldsymbol{\mu})^\top\} \\
&= \sum_{i=1}^{n} \mathsf{Cov}(\boldsymbol{y}_i) - n\,\mathsf{Cov}(\overline{\boldsymbol{y}}) = n\boldsymbol{\Sigma} - n\frac{1}{n}\,\boldsymbol{\Sigma} \\
&= (n-1)\boldsymbol{\Sigma} = m\boldsymbol{\Sigma}, \qquad m = n - 1.
\end{aligned}$$

Por otro lado, para introducir la covarianza de una matriz Wishart, considere $\boldsymbol{z} \sim \mathsf{N}_p(\mathbf{0}, \boldsymbol{I}_p)$ y sea $\boldsymbol{W} = \boldsymbol{z}\boldsymbol{z}^\top$. Sabemos que

$$\mathsf{E}(\boldsymbol{z}) = \mathbf{0}, \qquad \mathsf{Cov}(\boldsymbol{z}) = \boldsymbol{I}.$$

Es decir, sabemos que $\mathsf{E}(\boldsymbol{W}) = \mathsf{E}(\boldsymbol{z}\boldsymbol{z}^\top) = \boldsymbol{I}$. Además, tenemos que $z_i \perp z_j$, $i \neq j$, y

$$\mathsf{E}(z_i) = \mathsf{E}(z_i^3) = 0, \qquad \mathsf{E}(z_i^2) = 1, \qquad \mathsf{E}(z_i^4) = 3.$$

Por otro lado, as fácil notar que

$$\boldsymbol{W} = \boldsymbol{z}\boldsymbol{z}^\top = (z_1\boldsymbol{z}, \dots, z_p\boldsymbol{z}).$$

Luego

$$\mathsf{E}(z_i\boldsymbol{z}) = \mathsf{E}\begin{pmatrix} z_i z_1 \\ \vdots \\ z_i^2 \\ \vdots \\ z_i z_p \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \boldsymbol{e}_i$$

De este modo,

$$\mathsf{E}(z_i z_j \boldsymbol{z}\boldsymbol{z}^\top) = \delta_{ij}\boldsymbol{I} + \boldsymbol{e}_i\boldsymbol{e}_j^\top + \boldsymbol{e}_j\boldsymbol{e}_i^\top,$$

esto lleva a notar que

$$\mathsf{Cov}(z_i\boldsymbol{z}, z_j\boldsymbol{z}) = \mathsf{E}(z_i\boldsymbol{z}\, z_j\boldsymbol{z}^\top) - \mathsf{E}(z_i\boldsymbol{z})\,\mathsf{E}(z_j\boldsymbol{z}^\top) = \delta_{ij}\boldsymbol{I} + \boldsymbol{e}_i\boldsymbol{e}_j^\top + \boldsymbol{e}_j\boldsymbol{e}_i^\top - \boldsymbol{e}_i\boldsymbol{e}_j^\top.$$

Sea $\boldsymbol{E}_{ij} = \boldsymbol{e}_i\boldsymbol{e}_j^\top$ una matriz de ceros salvo en el elemento ij, y defina

$$\boldsymbol{K}_p = \sum_{i=1}^p \sum_{j=1}^p (\boldsymbol{E}_{ij} \otimes \boldsymbol{E}_{ij}^\top) \in \mathbb{R}^{p^2 \times p^2},$$

dado que $\boldsymbol{K}_p$ posee la propiedad,

$$\boldsymbol{K}_p \operatorname{vec}(\boldsymbol{A}) = \operatorname{vec}(\boldsymbol{A}^\top).$$

es conocida como matriz de conmutación. De este modo,

$$\mathsf{Cov}(\boldsymbol{W}) = \mathsf{Cov}(\operatorname{vec}(\boldsymbol{W})) = \boldsymbol{I} + \boldsymbol{K}_p$$

Note también que para $\boldsymbol{y} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ y sea $\boldsymbol{W} = \boldsymbol{y}\boldsymbol{y}^\top$. Como

$$\boldsymbol{W} = \boldsymbol{y}\boldsymbol{y}^\top \stackrel{\mathsf{d}}{=} \boldsymbol{B}\boldsymbol{z}\boldsymbol{z}^\top\boldsymbol{B}, \qquad \boldsymbol{z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

y $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$. De este modo,[2]

$$\begin{aligned} \mathsf{Cov}(\boldsymbol{W}) &= \mathsf{Cov}(\boldsymbol{B}\boldsymbol{z}\boldsymbol{z}^\top\boldsymbol{B}^\top) = \mathsf{Cov}(\operatorname{vec}(\boldsymbol{B}\boldsymbol{z}\boldsymbol{z}^\top\boldsymbol{B}^\top)) \\ &= \mathsf{Cov}((\boldsymbol{B}\otimes\boldsymbol{B})\operatorname{vec}(\boldsymbol{z}\boldsymbol{z}^\top)) = (\boldsymbol{B}\otimes\boldsymbol{B})\,\mathsf{Cov}(\operatorname{vec}(\boldsymbol{z}\boldsymbol{z}^\top))(\boldsymbol{B}\otimes\boldsymbol{B})^\top \\ &= (\boldsymbol{B}\otimes\boldsymbol{B})(\boldsymbol{I}+\boldsymbol{K}_p)(\boldsymbol{B}\otimes\boldsymbol{B})^\top = (\boldsymbol{I}+\boldsymbol{K}_p)(\boldsymbol{B}\boldsymbol{B}^\top\otimes\boldsymbol{B}\boldsymbol{B}^\top) \qquad (1) \\ &= (\boldsymbol{I}+\boldsymbol{K}_p)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) = 2\boldsymbol{N}_p(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}). \end{aligned}$$

Para notar la Ecuación (1), sea $\boldsymbol{N}_p = \frac{1}{2}(\boldsymbol{I}+\boldsymbol{K}_p)$ llamada matriz “simetrizadora”. Entonces $\boldsymbol{N}_p\boldsymbol{K}_p = \boldsymbol{N}_p = \boldsymbol{K}_p\boldsymbol{N}_p$, de donde sigue que:

$$\boldsymbol{N}_p(\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{N}_p = \boldsymbol{N}_p(\boldsymbol{A}\otimes\boldsymbol{A}) = (\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{N}_p.$$

[2] Recuerde que $\operatorname{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{C}^\top\otimes\boldsymbol{A})\operatorname{vec}\boldsymbol{B}$.

Resultado 2 (Teorema de Cochran)

Si las matrices aleatorias $\boldsymbol{A}_1, \dots, \boldsymbol{A}_r$ de orden $p \times p$ son todas independientes y $\boldsymbol{A}_i \sim \mathsf{W}_p(n_i, \boldsymbol{\Sigma})$, $i = 1, \dots, r$. Entonces

$$\sum_{i=1}^{r} \boldsymbol{A}_i \sim \mathsf{W}_p(n, \boldsymbol{\Sigma}), \qquad n = \sum_{i=1}^{r} n_i.$$

Demostración:

La función característica de $\boldsymbol{A} = \sum_{i=1}^{r} \boldsymbol{A}_i$ es el producto de las funciones características de $\boldsymbol{A}_1, \dots, \boldsymbol{A}_r$ y de ahí que

$$\varphi_A(\boldsymbol{\Theta}) = \prod_{j=1}^{r} \det(\boldsymbol{I}_p - i\boldsymbol{\Gamma\Sigma})^{-n_j/2} = \det(\boldsymbol{I}_p - i\boldsymbol{\Gamma\Sigma})^{-n/2}.$$

Resultado 3

Si $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ y $\boldsymbol{M}$ es matriz $k \times p$ de rango completo. Entonces

$$\boldsymbol{M}\boldsymbol{A}\boldsymbol{M}^\top \sim \mathsf{W}_k(n, \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^\top).$$

Demostración:

La función característica de $\boldsymbol{M}\boldsymbol{A}\boldsymbol{M}^\top$ es:[3]

$$\begin{aligned}
\mathsf{E}\left\{\exp\left(\tfrac{i}{2}\operatorname{tr}\boldsymbol{M}\boldsymbol{A}\boldsymbol{M}^\top\boldsymbol{\Gamma}\right)\right\} &= \mathsf{E}\left\{\exp\left(\tfrac{i}{2}\operatorname{tr}\boldsymbol{A}\boldsymbol{M}^\top\boldsymbol{\Gamma}\boldsymbol{M}\right)\right\} \\
&= \det(\boldsymbol{I}_p - i\boldsymbol{M}^\top\boldsymbol{\Gamma}\boldsymbol{M}\boldsymbol{\Sigma})^{-n/2} \\
&= \det(\boldsymbol{I}_k - i\boldsymbol{\Gamma}\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^\top)^{-n/2},
\end{aligned}$$

que es la función característica de $\mathsf{W}_k(n, \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^\top)$.

[3] Hemos usado el resultado $|\boldsymbol{I}_p + \boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{I}_k + \boldsymbol{B}\boldsymbol{A}|$.

Resultado 4

Si $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ y $\boldsymbol{A}, \boldsymbol{\Sigma}$ son particionadas como:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

donde $\boldsymbol{A}_{11}$ y $\boldsymbol{\Sigma}_{11}$ son $k \times k$. Entonces $\boldsymbol{A}_{11} \sim \mathsf{W}_k(n, \boldsymbol{\Sigma}_{11})$.

Demostración:

Basta considerar $\boldsymbol{M} = (\boldsymbol{I}_k, \boldsymbol{0})$ en el Resultado 3 entonces $\boldsymbol{M}\boldsymbol{A}\boldsymbol{M}^\top = \boldsymbol{A}_{11}$ y $\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^\top = \boldsymbol{\Sigma}_{11}$ y el resultado sigue.

Resultado 5

Si $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ donde $\boldsymbol{A}$ y $\boldsymbol{\Sigma}$ son particionadas como en el Resultado 4 y $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, entonces $\boldsymbol{A}_{11}$ y $\boldsymbol{A}_{22}$ son independientes y sus distribuciones son respectivamente, $\mathsf{W}_k(n, \boldsymbol{\Sigma}_{11})$ y $\mathsf{W}_{p-k}(n, \boldsymbol{\Sigma}_{22})$.

Una prueba de este resultado puede ser consstruído observando que cuando $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ la función característica de $\boldsymbol{A}_{11}$ y $\boldsymbol{A}_{22}$ es el producto de las funciones características de $\boldsymbol{A}_{11}$ y $\boldsymbol{A}_{22}$.

Observación:

Note que cuando $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_{11}, \dots, \sigma_{pp})$, entonces los elementos diagonales de $\boldsymbol{A}$, son todos independientes, y

$$a_{ii} \sim \mathsf{W}_1(n, \sigma_{ii}), \qquad i = 1, \dots, p.$$

Es decir, $a_{ii}/\sigma_{ii} \sim \chi^2(n)$, para $i = 1, \dots, p$.

Resultado 6

Si $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ donde n es un entero positivo y $\boldsymbol{y}$ es un vector aleatorio p-dimensional el cual es independiente de $\boldsymbol{A}$ con $\mathsf{P}(\boldsymbol{y} = \boldsymbol{0}) = 0$, entonces $\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}/\boldsymbol{y}^\top \boldsymbol{\Sigma}\boldsymbol{y} \sim \chi^2(n)$ y es independiente de $\boldsymbol{y}$.

Demostración:

En el Resultado 3 considere $\boldsymbol{M} = \boldsymbol{y}^\top \ (1 \times p)$. Entonces, condicional a $\boldsymbol{y}$, $\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y} \sim \mathsf{W}_1(n, \boldsymbol{y}^\top \boldsymbol{\Sigma}\boldsymbol{y})$. Esto es,

$$\frac{\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{\Sigma}\boldsymbol{y}} \sim \chi^2(n).$$

Dado que la distribución no depende de $\boldsymbol{y}$, ésta también corresponde a la distribución incondicional de $\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}/\boldsymbol{y}^\top \boldsymbol{\Sigma}\boldsymbol{y}$.

Resultado 7

Si $\overline{\boldsymbol{x}}$ y $\boldsymbol{S}$ son la media y la matriz de covarianza muestral desde $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces

$$(n-1)\frac{\overline{\boldsymbol{x}}^\top \boldsymbol{S}\,\overline{\boldsymbol{x}}}{\overline{\boldsymbol{x}}^\top \boldsymbol{\Sigma}\,\overline{\boldsymbol{x}}} \sim \chi^2(n),$$

y es independiente de $\overline{\boldsymbol{x}}$.

Demostración:

Sabemos que $\overline{\boldsymbol{x}}$ y $\boldsymbol{S}$ son independientes y $\boldsymbol{S} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma}/(n-1))$. Aplicando el resultado anterior se completa la prueba.

Resultado 8

Suponga que $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ donde $\boldsymbol{A}$ y $\boldsymbol{\Sigma}$ tienen forma particionada:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

y sea $\boldsymbol{A}_{11\cdot 2} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ y $\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. Entonces,

(i) $\boldsymbol{A}_{11\cdot 2} \sim \mathsf{W}_k(n - p + k, \boldsymbol{\Sigma}_{11\cdot 2})$ y es independiente de $\boldsymbol{A}_{12}$ y $\boldsymbol{A}_{22}$.

(ii) La distribución condicional de $\boldsymbol{A}_{12}$ dado $\boldsymbol{A}_{22}$ es $\mathsf{N}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{A}_{22}, \boldsymbol{\Sigma}_{11\cdot 2} \otimes \boldsymbol{A}_{22})$.

(iii) $\boldsymbol{A}_{22} \sim \mathsf{W}_{p-k}(n, \boldsymbol{\Sigma}_{22})$.

Resultado 9

Si $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ y $\boldsymbol{M}$ es $k \times p$ de rango k, entonces

$$(\boldsymbol{M}\boldsymbol{A}^{-1}\boldsymbol{M}^\top)^{-1} \sim \mathsf{W}_k(n-p+k, (\boldsymbol{M}\boldsymbol{\Sigma}^{-1}\boldsymbol{M}^\top)^{-1}).$$

Demostración:

Sea $\boldsymbol{B} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{-1/2}$, donde $\boldsymbol{\Sigma}^{1/2}$ es una matriz raíz cuadrada de $\boldsymbol{\Sigma}$. Por el Resultado 3, sigue que

$$\boldsymbol{B} \sim \mathsf{W}_p(n, \boldsymbol{I}_p).$$

Haciendo $\boldsymbol{R} = \boldsymbol{M}\boldsymbol{\Sigma}^{-1/2}$, tenemos que

$$(\boldsymbol{M}\boldsymbol{A}^{-1}\boldsymbol{M}^\top)^{-1} = (\boldsymbol{R}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{B}^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{R}^\top)^{-1} = (\boldsymbol{R}\boldsymbol{B}^{-1}\boldsymbol{R}^\top)^{-1}$$

y $(\boldsymbol{M}\boldsymbol{\Sigma}^{-1}\boldsymbol{M}^\top)^{-1} = (\boldsymbol{R}\boldsymbol{R}^\top)^{-1}$ por tanto basta probar que

$$(\boldsymbol{R}\boldsymbol{B}^{-1}\boldsymbol{R}^\top)^{-1} \sim \mathsf{W}_k(n-p+k, (\boldsymbol{R}\boldsymbol{R}^\top)^{-1}).$$

EX UMBRA IN SOLEM

Considere $\boldsymbol{R} = \boldsymbol{L}(\boldsymbol{I}_k, \boldsymbol{0})\boldsymbol{H}$, donde $\boldsymbol{L}$ es matrix no singular $k \times k$ y $\boldsymbol{H}$ es ortogonal $p \times p$, entonces

$$
\begin{aligned}
(\boldsymbol{R}\boldsymbol{B}^{-1}\boldsymbol{R}^\top)^{-1} &= \left(\boldsymbol{L}(\boldsymbol{I}_k, \boldsymbol{0})\boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{H}^\top \begin{pmatrix} \boldsymbol{I}_k \\ \boldsymbol{0} \end{pmatrix} \boldsymbol{L}^\top\right)^{-1} \\
&= \boldsymbol{L}^{-\top}\left((\boldsymbol{I}_k, \boldsymbol{0})(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^\top)^{-1} \begin{pmatrix} \boldsymbol{I}_k \\ \boldsymbol{0} \end{pmatrix}\right)^{-1} \boldsymbol{L}^{-1} \\
&= \boldsymbol{L}^{-\top}\left((\boldsymbol{I}_k, \boldsymbol{0})\boldsymbol{C}^{-1} \begin{pmatrix} \boldsymbol{I}_k \\ \boldsymbol{0} \end{pmatrix}\right)^{-1} \boldsymbol{L}^{-1},
\end{aligned}
$$

donde $\boldsymbol{C} = \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^\top \sim \mathsf{W}_p(n, \boldsymbol{I})$. Ahora, sea

$$
\boldsymbol{D} = \boldsymbol{C}^{-1} = \begin{pmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{pmatrix}, \qquad \boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix},
$$

donde $\boldsymbol{D}_{11}$ y $\boldsymbol{C}_{11}$ son $k \times k$.

De este modo,
$$(\boldsymbol{R}\boldsymbol{B}^{-1}\boldsymbol{R}^\top)^{-1} = \boldsymbol{L}^{-\top}\boldsymbol{D}_{11}^{-1}\boldsymbol{L}^{-1},$$
y dado que $\boldsymbol{D}_{11}^{-1} = \boldsymbol{C}_{11} - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}$. Usando el Resultado 8 en su parte (i) sigue que $\boldsymbol{D}_{11}^{-1} \sim \mathsf{W}_k(n-p+k, \boldsymbol{I}_k)$.[4] De ahí que
$$\boldsymbol{L}^{-\top}\boldsymbol{D}_{11}^{-1}\boldsymbol{L}^{-1} \sim \mathsf{W}_k(n-p+k, \boldsymbol{L}^{-\top}\boldsymbol{L}^{-1}),$$
y notando que $\boldsymbol{L}^{-\top}\boldsymbol{L}^{-1} = (\boldsymbol{L}\boldsymbol{L}^\top)^{-1} = (\boldsymbol{R}\boldsymbol{R}^\top)^{-1}$ se completa la prueba.

[4] $\boldsymbol{C} \sim \mathsf{W}_p(n, \boldsymbol{I}_p)$

Resultado 10

Si $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ donde n es un entero positivo, $n > p - 1$ y $\boldsymbol{y}$ es un vector aleatorio p-dimensional distribuído independiente de $\boldsymbol{A}$ con $\mathsf{P}(\boldsymbol{y} = \boldsymbol{0}) = 0$, entonces

$$\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y}}{\boldsymbol{y}^\top \boldsymbol{A}^{-1} \boldsymbol{y}} \sim \chi^2(n - p + 1),$$

independiente de $\boldsymbol{y}$.

Demostración:

En el Resultado 9 hacer $\boldsymbol{M} = \boldsymbol{y}^\top$ entonces, condicional a $\boldsymbol{y}$ tenemos

$$(\boldsymbol{y}^\top \boldsymbol{A}^{-1} \boldsymbol{y})^{-1} \sim \mathsf{W}_1(n - p + 1, (\boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y})^{-1}),$$

esto es, $\boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y} / \boldsymbol{y}^\top \boldsymbol{A}^{-1} \boldsymbol{y} \sim \chi^2(n - p + 1)$ y como la distribución no depende de $\boldsymbol{y}$, esta también corresponde a la disstribución incondicional.

Para notar la utilidad del resultado anterior, considere $\boldsymbol{A} \sim \mathsf{W}_p(n, \boldsymbol{\Sigma})$ entonces la disstribución de $\boldsymbol{A}^{-1}$ es llamada Wishart inversa.

Usando el Resultado 10 tenemos que, para cualquier $\boldsymbol{a} \in \mathbb{R}^p$,

$$\mathsf{E}(\boldsymbol{a}^\top \boldsymbol{A}^{-1} \boldsymbol{a}) = \boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}\, \mathsf{E}\left\{\frac{1}{\chi^2(n-p+1)}\right\} = \frac{1}{n-p-1} \boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}.$$

De ahí que

$$\boldsymbol{a}^\top \mathsf{E}(\boldsymbol{A}^{-1}) \boldsymbol{a} = \frac{1}{n-p-1} \boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}, \qquad \forall\, \boldsymbol{a}.$$

Lo que implica que

$$\mathsf{E}(\boldsymbol{A}) = \frac{1}{n-p-1} \boldsymbol{\Sigma}^{-1}$$

Suponga que $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ son vectores aleatorios independientes $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\overline{\boldsymbol{x}}$ y $\boldsymbol{S}$ el vector de medias y la matriz de covarianza muestral y que se desea probar hipótesis sobre $\boldsymbol{\mu}$ con $\boldsymbol{\Sigma}$ desconocido.

Hotelling (1931) definió la estadística T^2 dada por

$$T^2 = n\overline{\boldsymbol{x}}^\top \boldsymbol{S}^{-1}\overline{\boldsymbol{x}},$$

cuando $p = 1$, T^2 es el cuadrado del estadístico t usado para probar $H_0 : \mu = 0$. Note que $T^2 \geq 0$ y si $\boldsymbol{\mu} = \mathbf{0}$ entonces $\overline{\boldsymbol{x}}$ debería ser cercano a 0.

De este modo, parece razonable rechazar la hipótesis nula $H_0 : \boldsymbol{\mu} = \mathbf{0}$ si el valor de T^2 es suficientemente grande.