

Estadística Multivariante

LibreIM

Doble Grado en Ingeniería Informática y Matemáticas

Universidad de Granada

libreim.github.io/apuntesDGIIM



Este libro se distribuye bajo una licencia CC BY-NC-SA 4.0.

Eres libre de distribuir y adaptar el material siempre que reconozcas a los autores originales del documento, no lo utilices para fines comerciales y lo distribuyas bajo la misma licencia.

creativecommons.org/licenses/by-nc-sa/4.0/

Estadística Multivariante

LibreIM

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada

libreim.github.io/apuntesDGIIM

Índice

I. Teoría	6
1. Fundamentación probabilística de vectores aleatorios	6
1.1. Independencia	7
1.2. Función característica	8
1.2.1. Transformaciones lineales	8
1.2.2. Relación con la función de densidad	9
1.3. Aspectos generales sobre vectores aleatorios	9
1.3.1. Esperanza y covarianza	9
1.4. Caracterización de la distribución de un v.a. en términos de las distribuciones univariantes de c.l. de sus componentes	13
1.5. Momentos y cumulantes	14
1.5.1. Cambio de variables	14
2. Distribución normal multivariante (en el caso $\Sigma > 0$)	15
2.1. Vector de medias y matriz de covarianzas	16
2.1.1. Sobre independencia y condicionamiento	17
2.2. Función característica de la DNM	20
2.3. Algunas descomposiciones matriciales. Diagonalización.	22
2.3.1. Descomposición $U^T D U$ de matrices simétricas definidas no negativas.	22
2.3.2. Descomposición espectral de matrices simétricas definidas no negativas.	23
2.3.3. Matrices de intercambio de filas/columnas.	25
3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.	27
3.1. Definición y caracterizaciones.	27
3.2. Distribuciones esféricas y elípticas	30
3.2.1. Caracterización en términos de la densidad esférica estándar (C-IV)	30
3.3. Distribuciones esféricas y elípticas	32
3.3.1. Distribuciones esféricas	32
3.3.2. Distribuciones elípticas	32
3.4. Formas cuadráticas basadas en vectores aleatorios normales	34
3.5. Distribuciones χ^2 y F no centradas	34
3.5.1. Distribución χ^2 no centrada	34

Índice

3.5.2. Funciones hipergométricas generalizadas	35
3.5.3. Distribución F de Snedecor no centrada	36
3.5.4. Resultados sobre formas cuadráticas	37
4. Inferencia en la Distribución Normal Multivariante	42
4.1. Introducción	42
4.2. EMV de máxima verosimilitud de μ y Σ en una DNM	44
4.3. Estimación de los coeficientes de correlación	45
4.4. Estimadores de máxima verosimilitud de μ y Σ en una DNM . . .	47
4.5. Distribuciones de la media muestral y de la matriz de dispersiones	48
4.6. Distribuciones asintóticas de A_n y X_n para una población cualquiera	48
4.7. Distribución de Wishart	51
4.8. Marginalización y condicionamiento	53
4.9. Estadístico T^2 de Hotelling	54
4.9.1. Aplicaciones al problema de contraste sobre el vector de medias	55
II. Ejercicios	57

Parte I.

Teoría

1. Fundamentación probabilística de vectores aleatorios

En esta sección definimos algunos conceptos bien conocidos de la teoría de la probabilidad, a modo de breve repaso y para establecer la notación que utilizaremos en todo el libro.

Definición 1.1 (Vector y variable aleatorios). Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Un *vector aleatorio*, $X = (X_1, \dots, X_p)$ es una función medible

$$X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^p, \mathcal{B}^p).$$

Cuando $p = 1$, lo notamos X y decimos que es una *variable aleatoria*. La condición de medibilidad es: $X^{-1}(B) \in \mathcal{A}$ para cada $B \in \mathcal{B}^p$.

Definición 1.2 (Probabilidad inducida). La *probabilidad inducida por o distribución* de un vector aleatorio $X = (X_1, \dots, X_p)$ se define como la función

$$P_X[B] := P[X^{-1}(B)].$$

para todo $B \in \mathcal{B}^p$. Es decir, $P_X = P \circ X^{-1}|_{\mathcal{B}}$.

Definición 1.3. Cuando, para dos vectores aleatorios X y Y se tenga $P_X = P_Y$, diremos que son *iguales en distribución* y lo notaremos $X \stackrel{d}{=} Y$.

Notación. Cuando escribamos probabilidades de sucesos asociados a un vector aleatorio X , usaremos la siguiente notación:

- $P[X \in A] = P_X[A]$,
- $P[X_1 \leq x_1, \dots, X_p \leq x_p] = P_X[(-\infty, x_1] \times \dots \times (-\infty, x_p]]$,

y algunas formas análogas.

Definición 1.4 (Función de distribución). Se define la *función de distribución* asociada a P_X como

$$F_X : \mathbb{R}^p \rightarrow [0, 1] \\ (x_1, \dots, x_p) \mapsto P[X_1 \leq x_1, \dots, X_p \leq x_p].$$

Definición 1.5 (Función de densidad). Si existe una función $f_X : \mathbb{R}^p \rightarrow \mathbb{R}$ que sea integrable en el sentido de Lebesgue y tal que

$$F_X(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f_X(u_1, \dots, u_p) du_1 \dots du_p \quad \text{para todo } \mathbf{x} \in \mathbb{R}^p,$$

diremos que f_X es una *función de densidad* asociada a F_X . En los puntos de continuidad de f_X , la podemos escribir como:

$$f_X(\mathbf{x}) = \frac{\partial^p}{\partial x_1 \dots \partial x_p} F_X(\mathbf{x}).$$

1.1. Independencia

Veamos ahora algunas nociones relativas a la independencia de variables aleatorias.

Definición 1.6 (Conjunto rectangular). Se dice que $B \in \mathcal{B}^p$ es un *conjunto rectangular* si es un producto de p conjuntos de \mathcal{B} . Es decir, si

$$B = B_1 \times \cdots \times B_p, \quad \text{con } B_j \in \mathcal{B} \quad \text{para } j = 1, \dots, p.$$

Observamos que en general, dado un conjunto rectangular $B \in \mathcal{B}^p$ y un vector aleatorio X se tiene que

$$P_X[B] \neq P_{X_1}[B_1] \cdots P_{X_p}[B_p],$$

lo que motiva la siguiente definición.

Definición 1.7 (Independencia). Sea $X = (X_1, \dots, X_p)$ un vector aleatorio. Si para todo conjunto rectangular $B \in \mathcal{B}^p$ se verifica que

$$P_X[B] = P_{X_1}[B_1] \cdots P_{X_p}[B_p],$$

se dice que las variables aleatorias X_1, \dots, X_p son *independientes*.

Proposición 1.1 (Caracterizaciones de independencia). Dado un vector aleatorio $X = (X_1, \dots, X_p)$ sus componentes son independientes si, y solo si se da alguna de las siguientes condiciones:

- (i) La función de distribución verifica $F_X(\mathbf{x}) = F_{X_1}(x_1) \cdots F_{X_p}(x_p)$ para todo $\mathbf{x} \in \mathbb{R}^p$.
- (ii) X tiene una función de densidad y $f_X(\mathbf{x}) = f_{X_1}(x_1) \cdots f_{X_p}(x_p)$ para todo $\mathbf{x} \in \mathbb{R}^p$, salvo, a lo sumo, en un conjunto de medida de Lebesgue nula.

1.2. Función característica

Definición 1.8 (Función característica). La función característica de un vector aleatorio $X = (X_1, \dots, X_p)$ se define como la función

$$\psi_X(t) = E[e^{it^T X}], \quad \text{para } t \in \mathbb{R}^p.$$

Teorema 1.1 (Unicidad). La función característica de un vector aleatorio determina de forma única su distribución.

El siguiente resultado nos proporciona otra caracterización de la independencia de las componentes de un vector aleatorio, análoga a las descritas anteriormente pero ahora utilizando la función característica recién definida.

Proposición 1.2. Las componentes de $X = (X_1, \dots, X_p)^T$ son independientes si, y solo si

$$\psi_X(t) = \prod_{k=1}^p \psi_{X_k}(t_k).$$

Proposición 1.3. Si las componentes de $X = (X_1, \dots, X_p)$ son independientes, entonces la función característica de la variable $Y = \sum_{k=1}^p X_k$ es

$$\psi_Y(t) = \prod_{k=1}^p \psi_{X_k}(t).$$

1.2.1. Transformaciones lineales

A continuación estudiaremos qué forma tiene la función característica de un vector aleatorio cuando este es el resultado de someter otro vector aleatorio a transformaciones lineales.

Sea $X = (X_1, \dots, X_p)^T$ un vector aleatorio p -dimensional y sea $Y = (Y_1, \dots, Y_q)^T$ otro vector aleatorio q -dimensional definido como $Y = BX + b$, con $B \in \mathcal{M}_{q \times p}(\mathbb{R})$ una matriz constante y $b \in \mathbb{R}^q$ un vector también constante.

Proposición 1.4. La función característica de Y se obtiene como:

$$\psi_Y(t) = e^{it^T b} \psi_X(B^T t), \quad \text{para } t = (t_1, \dots, t_q)^T \in \mathbb{R}^q.$$

En particular, si

$$X = (X_{(1)} | X_{(2)})^T := (X_{(1),1}, \dots, X_{(1),k}, X_{(2),1}, \dots, X_{(2),p-k}),$$

con $X_{(1)}$ de dimensión $(k \times 1)$ y $X_{(2)}$ de dimensión $(p-k) \times 1$, se tendría que

$$X_{(1)} = (I_k | 0_{p-k})X + 0_{k \times 1},$$

1. Fundamentación probabilística de vectores aleatorios

por lo que, para $\mathbf{t}_{(1)} = (t_1, \dots, t_k)^T \in \mathbb{R}^k$ la función característica viene determinada por

$$\psi_{X_{(1)}}(\mathbf{t}_{(1)}) = \underbrace{e^{it_{(1)}^T \mathbf{0}}}_{=1} \psi_X((I_k | 0_{p-k}) \mathbf{t}_{(1)}) = \psi_X\left(\frac{\mathbf{t}_{(1)}}{0_{(p-k)}}\right).$$

1.2.2. Relación con la función de densidad

La función de densidad (en el caso continuo) de un vector aleatorio $\mathbf{X} = (X_1, \dots, X_p)^T$ y la correspondiente función característica constituyen un par de *transformadas de Fourier*, de forma que

$$\psi_X(\mathbf{t}) = \int_{\mathbb{R}^p} e^{it^T \mathbf{x}} f_X(\mathbf{x}) d\mathbf{x},$$

y por otro lado

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} e^{-it^T \mathbf{x}} \psi_X(t) dt.$$

1.3. Aspectos generales sobre vectores aleatorios

En este apartado definiremos y estudiaremos algunas de las propiedades de los vectores aleatorios con las que trabajaremos habitualmente.

1.3.1. Esperanza y covarianza

Para definir la esperanza de un vector aleatorio partimos de la definición de esperanza de una variable aleatoria y la extendemos de forma natural al caso multivariante.

Definición 1.9. Sea $\mathbf{X} = (X_1, \dots, X_n)^T$ un vector aleatorio. Se define el *vector de medias* de \mathbf{X} como:

$$\boldsymbol{\mu}_X = E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ \dots \\ E[X_p] \end{pmatrix} = \begin{bmatrix} \mu_1 \\ \dots \\ \mu_p \end{bmatrix}$$

siempre que existan *todas* las esperanzas unidimensionales.

Observamos a continuación que la propiedad de linealidad de la esperanza en las variables aleatorias se traslada trivialmente al caso multivariante.

Proposición 1.5 (Propiedad de linealidad). Sea $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$ con $\mathbf{B} \in M_{q \times p}$ constante y $\mathbf{b} \in \mathbb{R}^q$ constante. Entonces,

$$\boldsymbol{\mu}_Y = E[\mathbf{Y}] = \mathbf{B}E[\mathbf{X}] + \mathbf{b} = \mathbf{B}\boldsymbol{\mu}_X + \mathbf{b}.$$

Podemos incluso seguir iterando este proceso y extender esta propiedad al caso de matrices aleatorias. Sea $\mathbf{X} \in \mathcal{M}_{p \times q}$ es una matriz aleatoria, y $\mathbf{B} \in \mathcal{M}_{m \times p}$,

1. Fundamentación probabilística de vectores aleatorios

$C \in \mathcal{M}_{q \times n}$, $D \in \mathcal{M}_{m \times n}$ matrices constantes. Entonces, para $W = BXC + D$, una matriz aleatoria de dimensión $m \times n$, la matriz de medias viene dada por

$$E[W] = (E[W_{kl}])_{(kl)} = B \cdot E[X] \cdot C + D, \quad \text{con } E[X] = (E[X_{ij}])_{(ij)}.$$

Como comprobaremos a continuación, la definición de la matriz de covarianzas no es más que una extensión al caso multivariante de la definición de covarianza de dos variables aleatorias.

Definición 1.10. Se define la *matriz de covarianzas* del vector $X = (X_1, \dots, X_p)^T$ como:

$$\Sigma = \text{Cov}(X) = E[(X - \mu)(X - \mu)^T] = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{bmatrix},$$

donde $\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ji}$ es la covarianza de las variables aleatorias X_i y X_j . Solo puede definirse si existen todas las covarianzas.

Nota. Los elementos de la diagonal de la matriz de covarianzas se pueden calcular como

$$\sigma_{ii} = E[(X_i - \mu_i)^2] = \text{Var}(X_i) = \sigma_i^2,$$

siendo σ_i la *desviación típica*.

Nota. La desigualdad de Cauchy-Schwarz nos dice que existen todas las covarianzas si, y solo si existen las varianzas de cada componente.

Proposición 1.6 (Propiedades elementales). La matriz de covarianzas verifica las siguientes propiedades elementales:

1. Es simétrica.
2. Los elementos de su diagonal son no negativos.
3. La clase de matrices de covarianzas en dimensión $p \times p$ coincide con la clase de matrices simétricas definidas no negativas.

Demostración. La comprobación de las dos primeras propiedades es inmediata, por lo que vamos a ver únicamente la demostración de la propiedad 3.

3. \Rightarrow Supongamos que Σ es la matriz de covarianza de un vector aleatorio X , y μ_X es su vector de medias. Sea $\alpha \in \mathbb{R}^p$. Tenemos que probar que

1. Fundamentación probabilística de vectores aleatorios

$\alpha^T \Sigma \alpha \geq 0$. Consideramos la variable aleatoria $\alpha^T X$.

$$\begin{aligned} 0 \leq \text{Var}(\alpha^T X) &= E[\|\alpha^T X - E[\alpha^T X]\|^2] = E[\|\alpha^T X - \alpha^T \mu_X\|^2] \\ &= E[\|\alpha^T (X - \mu)\|^2] \\ &=^{(1)} E[\alpha^T (X - \mu)(X - \mu)^T \alpha] \\ &= \alpha^T E[(X - \mu)(X - \mu)^T] \alpha \\ &= \alpha^T \Sigma \alpha, \end{aligned}$$

donde en (1) hemos separado el cuadrado y hemos traspuesto uno de los términos (el resultado es el mismo, un escalar).

◀ Sea A una matriz simétrica, definida no negativa. Entonces, se puede encontrar una factorización $A = BB^T$, donde $C \in \mathcal{M}_p(\mathbb{R})$. Consideramos un vector aleatorio X cuyas entradas son p variables independientes e idénticamente distribuidas con distribución normal estándar. Entonces, $\Sigma_X = I_p$. Esto se puede comprobar usando que $E[XY] = E[X]E[Y]$ si X e Y son variables aleatorias independientes. Consideramos también $Y = BX$, entonces:

$$\Sigma_Y = B\Sigma_X B^T = BB^T = A.$$

□

Vamos a ver ahora diferentes **casos** que se pueden dar respecto a la matriz Σ .

Si Σ es definida positiva Se nota como $\Sigma > 0$. En este caso Σ es *no singular* pues ($|\Sigma| > 0$), y por tanto existe su inversa, Σ^{-1} . Esto es interesante pues nos permite normalizar el vector aleatorio de manera que tenga un vector de medias cero y una matriz de covarianzas que sea la identidad.

Nota. Usaremos la notación $X \sim (\mu, \Sigma)$ para denotar que la media y la matriz de covarianzas están bien definidas.

Definición 1.11 (Normalización). Dado un vector aleatorio $X \sim (\mu, \Sigma)$, con $\Sigma > 0$, se define su *normalización* en «origen» y «escala» como

$$Z = C^{-1}(X - \mu),$$

para cualquier matriz $C \in M_{p \times p}$ tal que $\Sigma = CC^T$. Esta matriz C no es única, pues existen infinitas descomposiciones de Σ de esta forma.

Veamos a continuación algunas de las propiedades de esta variable Z .

Proposición 1.7. Una normalización de un vector aleatorio tiene media 0 y matriz de covarianzas identidad.

1. Fundamentación probabilística de vectores aleatorios

Demostración.

Para la esperanza tenemos,

$$E[Z] = E[C^{-1}(X - \mu)] = C^{-1}E[(X - \mu)] = C^{-1}(E[X] - \mu) = C^{-1}(\mu - \mu) = 0,$$

y para la matriz de covarianzas,

$$\text{Cov}(Z) = E[ZZ^T] = E[C^{-1}(X - \mu)(X - \mu)^T(C^{-1})^T]$$

y por la propiedad de las matrices que nos dice que $(C^{-1})^T = (C^T)^{-1}$,

$$\begin{aligned} &= C^{-1}E[(X - \mu)(X - \mu)^T](C^T)^{-1} = C^{-1}\Sigma(C^T)^{-1} = C^{-1}CC^T(C^T)^{-1} = I \cdot I \\ &= I. \end{aligned} \quad \square$$

Definición 1.12. Se define la *distancia de Mahalanobis* de $X \sim (\mu, \Sigma)$ con $\Sigma > 0$ con respecto a su vector de medias como

$$\Delta(X, \mu) = \{(X - \mu)^T \Sigma (X - \mu)\}^{\frac{1}{2}},$$

que es una variable aleatoria.

Vamos a interpretar esta definición. Sea Z una normalización de X :

$$\begin{aligned} \Delta(X, \mu) &= \{(X - \mu)^T (CC^T)^{-1} (X - \mu)\}^{\frac{1}{2}} = \{(X - \mu)^T (C^T)^{-1} C^{-1} (X - \mu)\}^{\frac{1}{2}} \\ &= \{Z^T Z\}^{\frac{1}{2}} = \|Z\|, \end{aligned}$$

con norma la euclídea en \mathbb{R}^p . Así, esta distancia es la norma de cualquier normalización de X . Recordemos que $\|Z\|$ es una variable aleatoria, pues es una aplicación de una función medible a un vector aleatorio.

Proposición 1.8. Sea $X = (X_1, \dots, X_p)^T \sim (\mu, \Sigma)$ un vector aleatorio con $\Sigma > 0$. Entonces,

- (i) $E[\Delta^2(X, \mu)] = p$,
- (ii) la ecuación $\Delta(X, \mu) = k$ —con $k \geq 0$ constante— define la hipervariada de contorno correspondiente a aquellos puntos $x \in \mathbb{R}^p$ tales que, en el espacio transformado \mathbb{R}^p por

$$x \mapsto z = C^{-1}(x - \mu)$$

se sitúan en la esfera euclídea p -dimensional de radio k y centro en el origen.

Σ es **semidefinida positiva** (esto es, $\Sigma \geq 0$ con $|\Sigma| = 0$). En este caso Σ es singular, por lo que no existe Σ^{-1} . Por tanto, se tiene que $\text{rango}(\Sigma) = r < p$ y $\Sigma = CC^T$ con $C \in \mathcal{M}_{p \times r}(\mathbb{R})$ y cuyo rango es r .

1. Fundamentación probabilística de vectores aleatorios

Corolario 1.1. Con probabilidad 1, las componentes del vector $\mathbf{X} = (X_1, \dots, X_n)$ cumplirán una relación de dependencia lineal del tipo:

$$\alpha^T \mathbf{X} = k, \quad \alpha \in \mathbb{R}^p, \quad \alpha \neq 0, \quad k \in \mathbb{R}$$

Por tanto, toda la variabilidad de \mathbf{X} se sitúa en un hiperplano afín (casi seguramente con respecto a la medida P).

Demostración. Como $\det(\Sigma) = 0$, $\exists \alpha \neq 0$ tal que $\alpha^T \Sigma \alpha = 0$. Definimos la variable aleatoria $Y = \alpha^T \mathbf{X}$, y se tiene que:

$$E[Y] = E[\alpha^T \mathbf{X}] = \alpha^T E[\mathbf{X}] = \alpha^T \mu =: k,$$

y

$$\begin{aligned} \text{Var}(Y) &= E[\|Y - k\|^2] = E[\|\alpha^T \mathbf{X} - \alpha^T \mu\|^2] \\ &= E[\|\alpha^T (\mathbf{X} - \mu)\|^2] = E[\alpha^T (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T \alpha] \\ &= \alpha^T E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] \alpha = \alpha^T \Sigma \alpha = 0 \end{aligned}$$

con lo que Y es una variable aleatoria degenerada: $P[Y = k] = 1 \iff \alpha^T \mathbf{X} = k$ c.s. \square

Dicho esto, existen varias **medidas de variación global**. Podemos distinguir, dado Σ , y λ_j con $j = 1, \dots, p$ los autovalores de Σ :

- (i) $\det(\Sigma) = \prod_{j=1}^p \lambda_j$,
- (ii) $\text{tr}(\Sigma) = \sum_{j=1}^p \sigma_j^2 = \sum_{j=1}^p \lambda_j$.

1.4. Caracterización de la distribución de un v.a. en términos de las distribuciones univariantes de c.l. de sus componentes

Teorema 1.2. Sea $\mathbf{X} = (X_1, \dots, X_n)^T$ un vector aleatorio. La distribución de \mathbf{X} queda unívocamente determinada por las distribuciones univariantes de la forma

$$\alpha^T \mathbf{X}, \quad \alpha \in \mathbb{R}^p.$$

Demostración. Sea $Y_\alpha = \alpha^T \mathbf{X}$ (variable aleatoria unidimensional), para cada $\alpha \in \mathbb{R}^p$. La función característica de Y_α viene dada por

$$\psi_{Y_\alpha}(t) = E[e^{itY_\alpha}] = E[e^{it(\alpha^T \mathbf{X})}] \quad \text{para todo } t \in \mathbb{R}.$$

En particular, si $t = 1$

$$\psi_{Y_\alpha}(1) = E[e^{i\alpha^T \mathbf{X}}] = \psi_{\mathbf{X}}(\alpha)$$

Es decir:

$$\psi_{\mathbf{X}}(t) = \psi_{Y_t}(1), \quad \forall t \in \mathbb{R}^p$$

\square

1.5. Momentos y cumulantes

Sea $X = (X_1, \dots, X_p)^T$ un vector aleatorio con función característica ϕ_X .

Definición 1.13. Se define el **momento** (no centrado) p -dimensional de orden (r_1, \dots, r_p) de X como:

$$\mu_{r_1, \dots, r_p}^{1, \dots, p} = E[X_1^{r_1} \dots X_p^{r_p}].$$

Los momentos se pueden obtener a partir de la función característica derivándose de su expansión de Taylor (respecto al origen).

$$\phi_X(t) = E[e^{it^T X}] = E\left[\sum_{r=0}^{\infty} \frac{1}{r!} (it^T X)^r\right] = \sum_{r=0}^{\infty} \sum_{r_1 + \dots + r_p = r} \mu_{r_1, \dots, r_p}^{1, \dots, p} \frac{(it_1)^{r_1} \dots (it_p)^{r_p}}{r_1! \dots r_p!}$$

En particular, podemos enunciar este teorema:

Teorema 1.3. Si $E[|X_1|^{m_1} \dots |X_p|^{m_p}] < \infty$, entonces la función característica de X es (m_1, \dots, m_p) veces diferenciable y, si $m = m_1 + \dots + m_p$,

$$\frac{\partial^m}{\partial t_1^{m_1} \dots \partial t_p^{m_p}} \phi_X(t)|_{t=0} = i^m \mu_{m_1}^1 \dots \mu_{m_p}^p.$$

Consideremos el logaritmo de la función característica:

$$\log \phi(t).$$

Definición 1.14. Se definen los *cumulantes* (p -dimensionales) de orden (r_1, \dots, r_p) como los coeficientes de la correspondiente expansión:

$$\log \phi_X(t) = \sum_{r=0}^{\infty} \sum_{r_1 + \dots + r_p = r} \kappa_{r_1, \dots, r_p}^{1, \dots, p} \frac{(it_1)^{r_1} \dots (it_p)^{r_p}}{r_1! \dots r_p!}.$$

1.5.1. Cambio de variables

Teorema 1.4. Sea $X = (X_1, \dots, X_n)^T$ un vector aleatorio con función de densidad $f_X(x)$ positiva sobre $S \subseteq \mathbb{R}^d$ y continua. Sea $Y = (Y_1, \dots, Y_p)^T$ un vector aleatorio con:

$$Y = g(X) = (g_1(X), \dots, g_p(X))^T$$

con $g = (g_1, \dots, g_p) : S \rightarrow \mathcal{T}$ biyectiva, por lo que $\exists g^{-1}$, y denotamos $g^{-1} = h = (h_1, \dots, h_p)^T : \mathcal{T} \rightarrow S$.

Supongamos que existen las derivadas parciales:

$$\frac{\partial h_i(y)}{\partial y_j} \quad (i, j = 1, \dots, p)$$

y que son continuas sobre \mathcal{T} .

Entonces, la función de densidad $f_Y(y)$ del vector $Y = (Y_1, \dots, Y_p)^T = g(X)$ viene dada por:

$$f_Y(y) = f_X(g^{-1}(y)) |\det(J_{g^{-1}}(y))|.$$

Corolario 1.2 (Caso lineal). Sea $Y = BX + b$ con $B \in \mathcal{M}_p(\mathbb{R})$ y no singular. En este caso,

$$J_{g^{-1}}(g(x)) = [J_g(x)]^{-1} = \det(B)^{-1} = \det(B^{-1})$$

y se tiene:

$$f_Y(y) = f_X(B^{-1}(y - b)) |\det(B)^{-1}|.$$

Nota. Al factorizar una matriz definida positiva, siempre existe una factorización CC^T con $\det(C) > 0$.

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

En esta sección estudiamos la distribución normal multivariante, refiriéndonos siempre al caso en el que tiene matriz de covarianzas definida positiva. Algunos aspectos que justifican el estudio de esta distribución en particular:

- es la distribución multivariante más estudiada, y existen infinidad de resultados acerca de la misma,
- DNM es la base de muchas técnicas del análisis multivariante confirmatorio o inferencial,
- DNM es una extensión *natural* de la DNM univariante al caso multivariante,
- las distribuciones marginales de cualquier orden en una DNM son también normales,
- las distribuciones condicionadas (internamente) en una DNM también son normales y se obtienen de forma sencilla,
- la familia de DNM es cerrada bajo transformaciones lineales,
- la familia de DNM es cerrada bajo combinaciones lineales de vectores (mutuamente) independientes,
- la DNM viene determinada por los momentos de primer y segundo orden (media y matriz de covarianzas),
- si dos subvectores de un vector aleatorio con DNM tienen correlaciones cruzadas nulas, entonces dichos subvectores son (mutuamente) independientes.

Definición 2.1. Sea $X = (X_1, \dots, X_p)^T$ un vector aleatorio. Se dice que X tiene una distribución normal p -variante si su densidad es de la forma:

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

$$f_X(x) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

Siendo $\mu = (\mu_1, \dots, \mu_p)^T \in \mathbb{R}^p$ y Σ una matriz escalar $p \times p$ simétrica y definida positiva.

Veamos que, en efecto, f_X es una función de densidad

1. $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$,
2. $\int_{\mathbb{R}^p} f_X(x) dx = 1$. En efecto, como Σ es una matriz simétrica y definida positiva, podemos descomponerla de la forma $\Sigma = CC^T$ con $C \in M_p(\mathbb{R})$, $\det(C) > 0$. La forma de la densidad sugiere el cambio de variable $Z = C^{-1}(x - \mu)$. El determinante del jacobiano de la transformación es $\det(C^{-1})$. Por otro lado:

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = z^T z$$

Por tanto, podemos escribir:

$$\begin{aligned} \int_{\mathbb{R}^p} f_X(x) dx &= \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} dx \\ &= \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}} \det(C)^{\frac{1}{2}} \det(C^T)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} z^T z \right\} \det(C) dz \quad \left. \begin{array}{l} \text{cambio} \\ \text{de varia-} \\ \text{ble} \end{array} \right\} \\ &= \int_{\mathbb{R}^p} \frac{1}{(\sqrt{2\pi})^p} \exp \left\{ -\frac{1}{2} \sum_{j=1}^p z_j^2 \right\} dz_1 \dots dz_p \\ &= \prod_{j=1}^p \int_{\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} z_j^2 \right\}}_{\text{densidad de una } N(0, 1)} dz_j \\ &= \prod_{j=1}^p 1 = 1. \end{aligned}$$

2.1. Vector de medias y matriz de covarianzas

Del cálculo anterior se desprende que el vector aleatorio

$$Z = C^{-1}(X - \mu)$$

se distribuye con función de densidad

$$f_Z(z) = \prod_{j=1}^p \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} z_j^2 \right\} = \prod_{j=1}^p f_{Z_j}(z_j)$$

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

Es decir, para cada $j = 1, \dots, p$

$$Z_j \sim N(0, 1)$$

siendo Z_1, \dots, Z_p independientes (en particular, incorreladas), luego $\mathbf{Z} \sim N_p(0, I_p)$.

A partir de esto, usando 1.4, se obtiene que μ y Σ son, respectivamente, el vector de medias y la matriz de covarianzas del vector aleatorio $\mathbf{X} \sim N_p(\mu, \Sigma)$.

Proposición 2.1 (Transformación afín sobre una normal). Sea $\mathbf{X} \sim N_p(\mu, \Sigma)$, con $\Sigma > 0$ e $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$ con \mathbf{B} matriz escalar $p \times p$ no singular y $\mathbf{b} \in \mathbb{R}^p$. Entonces, se tiene que:

$$\mathbf{Y} \sim N_p(\mathbf{B}\mu + \mathbf{b}, \mathbf{B}\Sigma\mathbf{B}^T).$$

Demostración. De 1.2 se sigue que

$$\begin{aligned} f_Y(\mathbf{y}) &= f_X(\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b})) |\det(\mathbf{B})|^{-1} \\ &= \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2} |\det(\mathbf{B})|} \exp \left\{ -\frac{1}{2} (\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}) - \mu)^T \Sigma^{-1} (\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}) - \mu) \right\} \\ &= \frac{1}{(2\pi)^{p/2} \det(\mathbf{B}\Sigma\mathbf{B}^T)^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{b} - \mathbf{B}\mu)^T (\mathbf{B}\Sigma\mathbf{B}^T)^{-1} (\mathbf{y} - \mathbf{b} - \mathbf{B}\mu) \right\}. \end{aligned}$$

□

Teorema 2.1 (Caracterización de DNM). Sea \mathbf{X} un vector aleatorio p -dimensional, entonces \mathbf{X} tiene distribución normal multivariante, de vector de medias μ y matriz de covarianzas $\Sigma > 0$, si y solo si

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mu$$

con $\mathbf{A} \in \mathcal{M}_p(\mathbb{R})$ no singular, $\mathbf{A}\mathbf{A}^T = \Sigma$, $\mathbf{Z} \sim N_p(0, I_p)$.

Demostración. Supuesto que $\mathbf{X} = \mathbf{A}\mathbf{Z} + \mu$, de 2.1 se deduce que $\mathbf{X} \sim N_p(\mu, \Sigma)$. Supuesto que $\mathbf{X} \sim N_p(\mu, \Sigma)$, sin más que definir $\mathbf{Z} := \mathbf{A}^{-1}(\mathbf{X} - \mu)$ se tiene la igualdad, y claramente $\mathbf{Z} \sim N_p(0, I_p)$. □

2.1.1. Sobre independencia y condicionamiento

Proposición 2.2. Sea $\mathbf{X} = (X_1, \dots, X_p)^T$ un vector aleatorio con DNM $\mathbf{X} \sim N_p(\mu, \Sigma)$, $\Sigma > 0$. Si la matriz Σ es diagonal, entonces las variables aleatorias componentes del vector son independientes y tienen distribución normal univariante $X_i \sim N(\mu_i, \sigma_i^2)$.

Demostración. Por hipótesis, sabemos que:

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right\}$$

Y además:

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

- $\det(\Sigma)^{\frac{1}{2}} = \left(\prod_{i=1}^p \sigma_i^2\right)^{\frac{1}{2}} = \prod_{i=1}^p \sigma_i$,
- $\Sigma^{-1} = \text{diag}\left((\sigma_1^2)^{-1}, \dots, (\sigma_p^2)^{-1}\right)$,
- $(x-\mu)^T \text{diag}\left((\sigma_1^2)^{-1}, \dots, (\sigma_p^2)^{-1}\right)(x-\mu) = \sum_{i=1}^p (x_i - \mu_i)^2 \sigma_i^{-2} = \sum_{i=1}^p \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$.

Por tanto, reescribiendo:

$$\begin{aligned} f_X(x) &= \frac{1}{(2\pi)^{\frac{p}{2}} \prod \sigma_i} \exp \left\{ -\frac{1}{2} (x-\mu)^T \text{diag}\left((\sigma_1^2)^{-1}, \dots, (\sigma_p^2)^{-1}\right)(x-\mu) \right\} \\ &= \frac{1}{(2\pi)^{\frac{p}{2}} \prod \sigma_i} \exp \left\{ -\frac{1}{2} \sum_{i=1}^p \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \right\} \\ &= \prod_{i=1}^p \frac{1}{(2\pi)^{\frac{p}{2}} \prod \sigma_i} \exp \left\{ -\frac{1}{2} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \right\} \\ &= \prod_{i=1}^p f_{X_i}(x_i) \end{aligned}$$

□

El recíproco de este último resultado dado es cierto:

Proposición 2.3. Si $X = (X_1, \dots, X_p)^T$ es un vector aleatorio con componentes mutuamente independientes que tienen DNM, $X_i \sim N(\mu_i, \sigma_i)$, con $\sigma > 0$. Entonces, X tiene DNM

$$X \sim N_p(\mu, \Sigma)$$

con $\mu = (\mu_1, \dots, \mu_p)^T$ y $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) > 0$.

Nota. Exigimos $\sigma_i > 0$, $\forall i = 1, \dots, p$, y entonces $\Sigma > 0$.

Proposición 2.4. Sea X una v.a. $N(0, 1)$ y sea W una v.a. con distribución $U(\{-1, 1\})$, independiente de X . Consideramos $Y = WX$. Entonces, se verifica:

- $Y \sim N(0, 1)$,
- X e Y son incorreladas,
- X e Y no son independientes.

Esta proposición es un contraejemplo para ver que, en general, incorrelación no implica independencia (aunque alguna de las marginales sea una normal).

Teorema 2.2. Sea $X = (X_1, \dots, X_p)^T \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$. Supongamos que las componentes de X están ordenadas de tal modo que, por la partición del vector $X = \left(X_{(1)}^T | X_{(2)}^T\right)^T$ con $X_{(1)} = (X_1, \dots, X_q)^T$ y $X_{(2)} = (X_{q+1}, \dots, X_p)^T$, se tiene

$$\mu = \left(\mu_{(1)}^T | \mu_{(2)}^T\right)^T, \quad \Sigma = \begin{pmatrix} \Sigma_{(11)} & 0 \\ 0 & \Sigma_{(22)} \end{pmatrix}.$$

Entonces, $X_{(1)}$ y $X_{(2)}$ son mutuamente independientes y además, $X_{(1)} \sim N_q(\mu_{(1)}, \Sigma_{(11)})$ y $X_{(2)} \sim N_{p-q}(\mu_{(2)}, \Sigma_{(22)})$.

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

Demostración. Comenzamos notando que, como $\Sigma = \text{diag}(\Sigma_{(11)}, \Sigma_{(22)})$, entonces:

- $\det(\Sigma)^{\frac{1}{2}} = \det(\Sigma_{(11)})^{\frac{1}{2}} \det(\Sigma_{(22)})^{\frac{1}{2}}$
- $\Sigma^{-1} = \text{diag}(\Sigma_{(11)}^{-1}, \Sigma_{(22)}^{-1})$

De esta forma, tenemos:

$$\begin{aligned} (x - \mu)^T \Sigma^{-1} (x - \mu) &= ((x_{(1)} - \mu_{(1)})^T | (x_{(2)} - \mu_{(2)})^T) \begin{bmatrix} \Sigma_{(11)} & 0 \\ 0 & \Sigma_{(22)} \end{bmatrix} \begin{pmatrix} x_{(1)} - \mu_{(1)} \\ x_{(2)} - \mu_{(2)} \end{pmatrix} \\ &= (x_{(1)} - \mu_{(1)})^T \Sigma_{(11)}^{-1} (x_{(1)} - \mu_{(1)}) + (x_{(2)} - \mu_{(2)})^T \Sigma_{(22)}^{-1} (x_{(2)} - \mu_{(2)}) \end{aligned}$$

Podemos factorizar entonces la función de densidad de \mathbf{X} :

$$\begin{aligned} f_{\mathbf{X}}(x) &= \frac{1}{(2\pi)^{\frac{q}{2}} \det(\Sigma_{(11)})^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x_{(1)} - \mu_{(1)})^T \Sigma_{(11)}^{-1} (x_{(1)} - \mu_{(1)}) \right\} \\ &\quad \cdot \frac{1}{(2\pi)^{\frac{p-q}{2}} \det(\Sigma_{(22)})^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x_{(2)} - \mu_{(2)})^T \Sigma_{(22)}^{-1} (x_{(2)} - \mu_{(2)}) \right\} \end{aligned}$$

□

El recíproco también es cierto, en el sentido en el que lo era el recíproco del resultado anterior.

Teorema 2.3. Sea $\mathbf{X} = (X_1, \dots, X_p)^T \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$. Supongamos el particionamiento de \mathbf{X} como en el resultado anterior, pero ahora con

$$\Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}$$

Entonces, se tiene:

1. $\mathbf{X}_{(1)}$ y $\mathbf{X}_{(2)} - \Sigma_{(21)} \Sigma_{(11)}^{-1} \mathbf{X}_{(1)}$ son independientes,
- 2.

$$\begin{cases} \mathbf{X}_{(1)} \sim N_q(\mu_{(1)}, \Sigma_{(11)}) \\ \mathbf{X}_{(2)} - \Sigma_{(21)} \Sigma_{(11)}^{-1} \mathbf{X}_{(1)} \sim N_{p-q}(\mu_{(2)} - \Sigma_{(21)} \Sigma_{(11)}^{-1} \mu_{(1)}, \Sigma_{(22)} - \Sigma_{(21)} \Sigma_{(11)}^{-1} \Sigma_{(12)}) \end{cases}$$

3. La distribución condicionada de $\mathbf{X}_{(2)}$ dado $\mathbf{X}_{(1)} = x_{(1)}$ es una DNM

$$N_{p-q}(\mu_{(2)} + \Sigma_{(21)} \Sigma_{(11)}^{-1} (x_{(1)} - \mu_{(1)}), \Sigma_{(22)} - \Sigma_{(21)} \Sigma_{(11)}^{-1} \Sigma_{(12)}).$$

Demostración. Sea $C = \begin{pmatrix} I_1 & 0 \\ -\Sigma_{(21)} \Sigma_{(11)}^{-1} & I_2 \end{pmatrix}$. C es no singular, pues $\det(C) = 1$. Consideramos el cambio lineal $\mathbf{Y} = C\mathbf{X}$:

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

$$\mathbf{Y} = \begin{pmatrix} \mathbf{X}_{(1)} \\ \mathbf{X}_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mathbf{X}_{(1)} \end{pmatrix} := \begin{pmatrix} \mathbf{Y}_{(1)} \\ \mathbf{Y}_{(2)} \end{pmatrix}.$$

Por 2.1, $\mathbf{Y} \sim N_p(C\mu, C\Sigma C^T)$. Se tiene:

$$C\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mu_{(1)} \end{pmatrix}, \quad C\Sigma C^T = \begin{pmatrix} \Sigma_{(11)} & 0 \\ 0 & \Sigma_{(22.1)} \end{pmatrix},$$

donde $\Sigma_{(22.1)} := \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}$. Así, 1 y 2 se siguen de 2.2.

Para probar 3 escribimos, usando el teorema de Bayes, la expresión de la de función de densidad de la distribución condicionada:

$$\begin{aligned} f_{\mathbf{X}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1)}}(\mathbf{x}_{(2)}) &= \frac{f_{\mathbf{X}_{(1)},\mathbf{X}_{(2)}}(\mathbf{x}_{(1)},\mathbf{x}_{(2)})}{f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)})} \\ &= \frac{f_{\mathbf{Y}_{(1)},\mathbf{Y}_{(2)}}(\mathbf{x}_{(1)},\mathbf{x}_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mathbf{x}_{(1)})}{f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)})} \end{aligned} \quad \left. \begin{array}{l} \text{usando el teorema de} \\ \text{cambio de variable} \\ \text{1.2 con } \mathbf{X} = C^{-1}\mathbf{Y}, \\ \det(C) = 1 \end{array} \right\}$$

Veamos el valor de la forma cuadrática que aparecerá en $f_{\mathbf{Y}_{(1)},\mathbf{Y}_{(2)}}$:

$$\begin{aligned} \left(\frac{\mathbf{x}_{(1)}}{\mathbf{x}_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mathbf{x}_{(1)}} \right) - C\mu &= \begin{pmatrix} \mathbf{x}_{(1)} - \mu_{(1)} \\ \mathbf{x}_{(2)} - \mu_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}(\mathbf{x}_{(1)} - \mu_{(1)}) \end{pmatrix} := D(\mathbf{x}_{(2)}), \\ D(\mathbf{x}_{(2)})^T (C\Sigma C^T) D(\mathbf{x}_{(2)}) &= \overbrace{(\mathbf{x}_{(1)} - \mu_{(1)})^T \Sigma_{(11)}^{-1} (\mathbf{x}_{(1)} - \mu_{(1)})}^{:=A} \\ + \underbrace{(\mathbf{x}_{(2)} - \mu_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}(\mathbf{x}_{(1)} - \mu_{(1)}))^T \Sigma_{(22.1)}^{-1} (\mathbf{x}_{(2)} - \mu_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}(\mathbf{x}_{(1)} - \mu_{(1)}))}_{:=B(\mathbf{x}_{(2)})} \end{aligned}$$

Resultando:

$$\begin{aligned} f_{\mathbf{X}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1)}}(\mathbf{x}_{(2)}) &= \frac{\frac{1}{(2\pi)^{p/2}(\det(\Sigma_{(11)})\det(\Sigma_{(22.1)}))^{1/2}} \exp\left\{-\frac{1}{2}(A+B(\mathbf{x}_{(2)}))\right\}}{\frac{1}{(2\pi)^{q/2}\det(\Sigma_{(11)})^{1/2}} \exp\left\{-\frac{1}{2}A\right\}} \\ &= \frac{1}{(2\pi)^{\frac{p-1}{2}}\det(\Sigma_{(22.1)})^{1/2}} \exp\left\{-\frac{1}{2}B(\mathbf{x}_{(2)})\right\}. \end{aligned}$$

□

2.2. Función característica de la DNM

Cabe recordar en esta sección la definición 1.8.

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

Proposición 2.5 (Función característica de la DNM). Dado un vector aleatorio $X \sim N_p(\mu, \Sigma)$, su función característica viene dada por

$$\psi_X(t) = \exp \left\{ it^T \mu - \frac{1}{2} t^T \Sigma t \right\}.$$

Demostración. Tomamos una descomposición $\Sigma = CC^T$, con $C \in \mathcal{M}_p(\mathbb{R})$ no singular, $\det(C) > 0$, y aplicamos el cambio de variable $Y = C^{-1}(X - \mu)$.

$$\begin{aligned} \Psi_X(t) &= \int_{\mathbb{R}^p} e^{it^T x} f_X(x) dx \\ &= \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp \left\{ it^T x - \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} dx \\ &= \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp \left\{ it^T (Cy + \mu) - \frac{1}{2} y^T y \right\} \det(C) dy \\ &= e^{it^T \mu} \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}} \exp \left\{ \sum_{j=1}^p i \alpha_j y_j - \frac{1}{2} y_j^2 \right\} \\ &= \prod_{j=1}^p e^{it_j \mu_j} \Psi_{Y_j}(\alpha_j) \\ &=^{(1)} \exp \left\{ it^T \mu - \frac{1}{2} t^T \Sigma t \right\}, \end{aligned}$$

con $\alpha_j = t^T c_j$, y c_j la columna j -ésima de la matriz C , recordando que:

$$\Psi_{N(\mu, \sigma^2)}(t) = e^{it\mu - \frac{1}{2} \sigma^2 t^2}$$

y en (1) teniendo en cuenta:

- $\sum_{j=1}^p t_j \mu_j = t^T \mu$,
- $\sum_{j=1}^p \alpha_j^2 = \alpha^T \alpha = t^T C C^T t = t^T \Sigma t$.

□

Proposición 2.6. Sea $X = (X_1, \dots, X_p)^T \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$. Sea $Y = (Y_1, \dots, Y_q)^T$ definido como $Y = BX + b$, donde $B \in \mathcal{M}_{q \times p}(\mathbb{R})$ de rango $q \leq p$, y $b \in \mathbb{R}^q$. Entonces, se tiene que

$$Y \sim N_q(B\mu + b, B\Sigma B^T),$$

con $B\Sigma B^T > 0$.

Demostración. Para $t \in \mathbb{R}^q$, se tiene:

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

$$\begin{aligned}
 \psi_Y(t) &= E[e^{it^T(BX+b)}] \\
 &= e^{it^T b} E[e^{i(B^T t)^T X}] \\
 &= e^{it^T b} \psi_X(B^T t) \\
 &= e^{it^T b} \exp \left\{ i(B^T t)^T \mu - \frac{1}{2} (B^T t)^T \Sigma B^T t \right\} \\
 &= \exp \left\{ it^T (B\mu + b) - \frac{1}{2} t^T B \Sigma B^T t \right\},
 \end{aligned}$$

y la correspondencia biunívoca entre funciones características y distribuciones concluye la prueba de la primera afirmación. Además, se verifica que $B \Sigma B^T > 0$ por ser tanto $\Sigma > 0$ como B de rango máximo. \square

Corolario 2.1 (Marginalización). Sea $X = (X_1, \dots, X_p)^T \sim N_p(\mu, \Sigma)$ un vector aleatorio con $\Sigma > 0$. Entonces para todo subvector $X_r = (X_{r_1}, \dots, X_{r_q})^T$ —donde $r = (r_1, \dots, r_q)^T$ con $r_1, \dots, r_q \in \{1, \dots, p\}$ y $q \leq p$ — se tiene que $X_r \sim N_q(\mu_r, \Sigma_r)$, siendo

- μ_r el subvector de μ correspondiente a r , y
- Σ_r la submatriz de Σ definida por las filas y columnas correspondientes a r .

2.3. Algunas descomposiciones matriciales. Diagonalización.

Proposición 2.7. Sea A una matriz simétrica y definida positiva. Entonces, $\|x\|_A := (x^T A x)^{1/2}$ es una norma en \mathbb{R}^p .

2.3.1. Descomposición $U^T D U$ de matrices simétricas definidas no negativas.

Consideramos $A \in \mathcal{M}_p(\mathbb{R})$ simétrica y definida no negativa.

Teorema 2.4 (Descomposición $U^T D U$). Existe una matriz U triangular superior, con los elementos de la diagonal todos 1, tal que

$$A = U^T D U,$$

con D una matriz diagonal con elementos de la diagonal no negativos, única.

Corolario 2.2 (Descomposición de Cholesky). Sea r el rango de A , entonces existe una única matriz triangular superior T con r elementos diagonales positivos y $p - r$ filas nulas, tal que

$$A = T^T T,$$

y, de hecho, se tiene

$$T = D^{1/2} U$$

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

si D y U son las de la descomposición $U^T D U$.

2.3.2. Descomposición espectral de matrices simétricas definidas no negativas.

En esta sección no exigimos que A sea simétrica ni definida no negativa.

Definición 2.2 (Polinomio característico). Se llama *polinomio característico* de A al definido por:

$$p(\lambda) = \det(A - \lambda I).$$

Proposición 2.8. Las raíces reales del polinomio característico de A son los valores propios de A .

Definición 2.3 (Multiplicidades algebraica y geométrica). Sea $\lambda \in \mathbb{R}$ un valor propio de A . Se llama *multiplicidad algebraica* de λ a la multiplicidad de λ como raíz del polinomio característico. Se llama *multiplicidad geométrica* de λ a la dimensión del subespacio que generan los vectores propios asociados a λ .

Diagonalización.

Vamos a recordar algunos resultados relativos a la diagonalización de matrices, que nos llevarán a la descomposición espectral.

Definición 2.4. Se dice que $A \in \mathcal{M}_p(\mathbb{R})$ es diagonalizable si existe una matriz $Q \in \mathcal{M}_p(\mathbb{R})$ no singular tal que:

$$Q^{-1} A Q = D,$$

con D una matriz diagonal.

Teorema 2.5. Sea $A \in \mathcal{M}_p(\mathbb{R})$. Supongamos que A es diagonalizable por otra matriz $Q \in \mathcal{M}_p(\mathbb{R})$ no singular, esto es

$$Q^{-1} A Q = D,$$

con D una matriz diagonal. Denotemos $Q = (q_1, \dots, q_p)$, y $D = \text{diag}(d_1, \dots, d_p)$. Entonces:

- (i) $\text{rango}(A) = \text{número de elementos de la diagonal de } D \text{ que son distintos de } 0$,
- (ii) $\det(A) = \prod_{i=1}^p d_i$,
- (iii) $\text{tr}(A) = \sum_{i=1}^p d_i$,
- (iv) el polinomio característico de A es:

$$p(\lambda) = (-1)^p (\lambda - d_1) \dots (\lambda - d_p),$$

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

- (v) el espectro de A comprende los escalares distintos incluidos en la diagonal de D ,
- (vi) las multiplicidades algebraicas y geométricas de un autovalor λ de A coinciden y son iguales al número de elementos diagonales de D que igualan λ ,
- (vii) las columnas de Q son autovectores linealmente independientes de A (la columna q_i es un autovector correspondiente al autovalor d_i),

Definición 2.5. Se dice que A es *diagonalizable ortogonalmente* si es diagonalizable por una matriz ortogonal, esto es, $A = QDQ^T$, con $Q^T = Q^{-1}$, $QQ^T = Q^TQ = I$.

Corolario 2.3. Si A es diagonalizable ortogonalmente, debe ser simétrica, pues $A^T = (QDQ^T)^T = (Q^T)^T D^T Q^T = QDQ^T = A$.

Proposición 2.9.

- A es definida no negativa \iff los autovalores de A son no negativos,
- A es definida positiva \iff los autovalores de A son positivos.

Y, si A es **simétrica**:

- A es definida no negativa \iff los autovalores de A son no negativos,
- A es definida positiva \iff los autovalores de A son positivos,
- A es semidefinida positiva \iff los autovalores de A son no negativos y al menos uno es nulo.

El concepto de raíz cuadrada de una matriz nos será útil:

Proposición 2.10. Sea una matriz $A \in \mathcal{M}_p(\mathbb{R})$ simétrica y definida no negativa. Entonces, existe una matriz $R \in \mathcal{M}_p(\mathbb{R})$ simétrica y definida no negativa tal que

$$A = R^2 = RR$$

Además, R es única y puede expresarse como:

$$R = Q \operatorname{diag}(\sqrt{d_1}, \dots, \sqrt{d_p}) Q^T$$

A R se le llama raíz cuadrada de una matriz simétrica y definida no negativa. La notamos $A^{1/2}$.

Descomposición espectral de una matriz simétrica definida no negativa.

Dada $A \in \mathcal{M}_p(\mathbb{R})$ simétrica y definida no negativa, vamos a definir su *descomposición espectral*.

2. Distribución normal multivariante (en el caso $\Sigma > 0$)

Definición 2.6. Se llama *descomposición espectral* de A a la descomposición en la forma:

$$A = H\Lambda H^T,$$

toda vez que se verifique que:

- (i) $H \in \mathcal{M}_p(\mathbb{R})$ es ortogonal,
- (ii) $\Lambda \in \mathcal{M}_p(\mathbb{R})$ es diagonal,

es decir, que A sea diagonalizable ortogonalmente por H y Λ .

Y, en efecto, el recíproco del corolario 2.3 es cierto y siempre existe tal descomposición:

Teorema 2.6. A es diagonalizable ortogonalmente si y solo si A es simétrica.

Sea $r \leq p$ el rango de A . Si elegimos H y Λ de forma que Λ tiene las filas distintas de cero al principio:

$$\Lambda = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix},$$

con $D \in \mathcal{M}_r(\mathbb{R})$, y escribimos $H = (H_1|H_2)$, con $H_1 \in \mathcal{M}_{p \times r}(\mathbb{R})$, $H_2 \in \mathcal{M}_{p \times (p-r)}(\mathbb{R})$, entonces:

$$A = (H_1|H_2)^T \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} (H_1|H_2) = H_1 D H_1^T.$$

2.3.3. Matrices de intercambio de filas/columnas.

Sea $I_{(r,s)}$ la matriz obtenida intercambiando en la identidad I_p las filas r y s . Dada $A \in \mathcal{M}_p(\mathbb{R})$,

- $I_{(r,s)}A$ es la matriz obtenida intercambiando en A las filas r y s ,
- $AI_{(r,s)}$ es la matriz obtenida intercambiando en A las columnas r y s .

Proposición 2.11 (Propiedades).

- (i) Si $r \neq s$, $\det(I_{(r,s)}) = -1$,
- (ii) $I_{(r,s)}$ es simétrica y ortogonal.

En la descomposición espectral de una matriz simétrica y definida no negativa, tenemos unicidad de D salvo reordenación de los elementos de la diagonal y la correspondiente reordenación de las columnas de Q :

$$A = QDQ^T = (QI_{(r,s)})(I_{(r,s)}DI_{(r,s)})(I_{(r,s)}Q^T)$$

Inciso - interpretación geométrica de la distancia de Mahalanobis y la densidad de la DNM.

Ahora estamos en disposición de acometer este ejercicio:

Ejercicio 2.1. Sea $\mathbf{X} = (X_1, \dots, X_p)^T \sim (\boldsymbol{\mu}, \Sigma)$ un vector aleatorio con $\Sigma > 0$.

1. Estudiar el lugar geométrico de los puntos $\Delta(\mathbf{x}, \boldsymbol{\mu}) = k$, e interpretarlo en función de los autovalores de Σ .
2. Si además $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, estudiar el lugar geométrico de los puntos definidos por $f_{\mathbf{X}}(\mathbf{x}) = h$ —con $h \geq 0$ — y dar una interpretación geométrica espectral.

Solución. Comenzamos dando la solución del apartado (1). Para $k \geq 0$ consideramos

$$\Delta(\mathbf{x}, \boldsymbol{\mu}) = k \iff (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = k^2 \iff (\mathbf{x} - \boldsymbol{\mu})^T (k^2 \Sigma)^{-1} (\mathbf{x} - \boldsymbol{\mu}) = 1.$$

Visto así se tiene que los puntos \mathbf{x} solución de esta ecuación definen el *elipsoide* de centro $\boldsymbol{\mu}$ con ejes principales definidos por los autovectores de $(k^2 \Sigma)^{-1}$, siendo la longitud al cuadrado de los semiejes igual a los inversos de los autovalores. Descomponiendo $\Sigma = H \Lambda H^T$, podemos escribir

$$\begin{aligned} (k^2 \Sigma)^{-1} &= (k H \Lambda H^T k)^{-1} = k^{-1} (H \Lambda H^T)^{-1} = k^{-1} (H^T)^{-1} \Lambda^{-1} H^{-1} k^{-1} \\ &= (k^{-1} H) \Lambda^{-1} (k^{-1} H)^T = H (k^2 \Lambda)^{-1} H^T. \end{aligned}$$

Los autovectores vienen definidos por la matriz H —columnas—, dando las direcciones de los ejes del elipsoide. Por otro lado, los autovalores de $(k^2 \Sigma)^{-1}$ vienen dados por los elementos diagonales de la matriz diagonal $(k^2 \Lambda)^{-1}$, es decir, los inversos de la forma $(k^2 \lambda_j)^{-1}$ de los autovalores de la matriz $k^2 \Sigma$. Finalmente, las longitudes al cuadrado de los semiejes coinciden con los propios autovalores $k^2 \lambda_j$, por lo que las longitudes de los semiejes son los valores $k \lambda_j^{1/2}$.

Procedemos ahora a ver la solución del apartado (2). La ecuación es, por definición,

$$f_{\mathbf{X}}(\mathbf{X}) = h \iff \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} = h.$$

Escribimos entonces

$$(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = -2 \ln (2\pi^{p/2} |\Sigma|^{1/2} h).$$

El máximo de $f_{\mathbf{X}}(\mathbf{x})$ se alcanza en $\mathbf{x} = \boldsymbol{\mu}$, que llamaremos

$$M := f_{\mathbf{X}}(\boldsymbol{\mu}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}}.$$

Así, la ecuación tendrá solución si $0 < h \leq M$. En ese caso,

$$(2\pi)^{p/2} |\Sigma|^{1/2} h \leq 1,$$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

luego

$$\log((2\pi)^{p/2} |\Sigma|^{1/2} h) \leq 0.$$

3. Distribución normal multivariante en el caso general

$$\Sigma \geq 0.$$

3.1. Definición y caracterizaciones.

Dado un vector aleatorio $X = (X_1, \dots, X_p)^T$, 2.1 y 1.2 nos permiten deducir la siguiente caracterización de la DNM:

Proposición 3.1. *X tiene DNM si y solo si toda combinación lineal de las componentes X_1, \dots, X_p de la forma*

$$\alpha^T X$$

con $\alpha \in \mathbb{R}^p - \{0\}$, tiene DN univariante no degenerada, necesariamente de media $\alpha^T \mu_X$ y varianza $\alpha^T \Sigma_X \alpha$.

Hasta ahora hemos descrito tenemos las siguientes formulaciones equivalentes de DNM:

- [D]: definición en términos de la densidad f_X ,
- [C-I]: $X = AZ + \mu$ con A no singular, $\Sigma = AA^T$,
- [C-II]: $\Psi_X = \exp \left\{ it^T \mu - \frac{1}{2} t^T \Sigma t \right\}$, y
- [C-III]: $\alpha^T X \sim N(\alpha^T \mu, \alpha^T \Sigma \alpha)$, $\forall \alpha \in \mathbb{R}^p - \{0\}$.

Podemos usar [C-II] para definir la DNM en el caso general de la siguiente manera:

Definición 3.1. Dado $X = (X_1, \dots, X_p)^T \sim N_p(\mu, \Sigma)$ un vector aleatorio con $\Sigma \geq 0$, se dice que tiene una DNM si su función característica viene dada por:

$$\Psi_X(t) = E[e^{it^T X}] = \exp \left\{ it^T \mu - \frac{1}{2} t^T \Sigma t \right\} \text{ para todo } t \in \mathbb{R}^p.$$

Proposición 3.2 (Transformaciones lineales de rango no necesariamente máximo).

Dado un vector aleatorio $X = (X_1, \dots, X_p)^T \sim N_p(\mu, \Sigma)$ con $\Sigma \geq 0$. Sea $Y = (Y_1, \dots, Y_q)^T$ definido como $Y = BX + b$, con $B \in \mathcal{M}_{q \times p}(\mathbb{R})$ una matriz de rango $r \leq \min\{q, p\}$, y $b \in \mathbb{R}^q$.

Entonces, se tiene que

$$Y \sim N_q(B\mu + b, B\Sigma B^T).$$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

Nota. La única novedad que nos aporta esta nueva proposición respecto a lo que ya teníamos (2.6), es que ahora $B\Sigma B^T \geq 0$ (antes teníamos que tenía que se estrictamente mayor). La demostración es análoga a la del caso $\Sigma > 0$.

Vamos a extender la [C-I] a partir de la definición que acabamos de dar para el caso general: sea $\mathbf{X} = (X_1, \dots, X_p)^T \sim N_p(\mu, \Sigma)$, con $\Sigma \geq 0$, es decir,

$$\Psi_{\mathbf{X}}(t) = \exp \left\{ it^T \mu - \frac{1}{2} t^T \Sigma t \right\} \quad \forall t \in \mathbb{R}^p$$

Descomponiendo $\Sigma = H^T \Lambda H = H_1 D H_1^T$, podemos escribir:

$$\begin{aligned} \Psi_{\mathbf{X}}(t) &= \exp \left\{ it^T \mu - \frac{1}{2} t^T \Sigma t \right\} \\ &= \exp \left\{ i^T H H^T \mu - \frac{1}{2} t^T H_1 D H_1^T t \right\} \\ &= \exp \left\{ it^T H_1 H_1^T \mu - \frac{1}{2} t^T H_1 D H_1^T t \right\} \cdot \exp \{ it H_2 H_2^T \mu \}. \end{aligned}$$

La forma de la función característica obtenida sugiere considerar el cambio de variable $\mathbf{Y} = H^T \mathbf{X}$:

$$\mathbf{Y} = \begin{pmatrix} H_1^T \\ H_2^T \end{pmatrix} \mathbf{X} = \begin{pmatrix} H_1^T \mathbf{X} \\ H_2^T \mathbf{X} \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

Con esto, se tiene:

$$\begin{aligned} (i) \quad \mu_{\mathbf{Y}} &= H^T \mu = \begin{pmatrix} H_1^T \mu \\ H_2^T \mu \end{pmatrix} = \begin{pmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{pmatrix} \\ (ii) \quad \Sigma_{\mathbf{Y}} &= H^T \Sigma H = H^T H \Delta H^T H = \Delta = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \implies \begin{cases} \Sigma_{Y_1} = D \\ \Sigma_{Y_2} = 0 \end{cases}, \end{aligned}$$

Y, por tanto:

$$\begin{cases} Y_1 \sim N_r(H_1 \mu, D), D > 0, \\ Y_2 = H_2 \mu \quad c.s. \end{cases}$$

Observamos, por 1.4:

$$\Psi_{\mathbf{X}}(t) = \Psi_{\mathbf{Y}}(H^T t)$$

Denotamos

$$\mathbf{v} = H^T t = \begin{pmatrix} H_1^T \\ H_2^T \end{pmatrix} t = \begin{pmatrix} H_1^T t \\ H_2^T t \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

Con esto:

$$\Psi_{\mathbf{Y}}(\mathbf{v}) = \Psi_{\mathbf{X}}(t) = \exp \left\{ i v_1 \mu_{Y_1} - \frac{1}{2} v_1^T \Sigma_{Y_1} v_1 \right\} \cdot \exp \{ i v_2^T \mu_{Y_2} \}$$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

y en la última igualdad, el primer término es $\Psi_{Y_1}(\nu_1)$ y el segundo es $\Psi_{Y_2}(\nu_2)$, luego Y_1, Y_2 son independientes.

Aplicamos a Y_1 la [C-I]: tenemos que Y_1 puede representarse como

$$Y_1 = D^{\frac{1}{2}}Z + H_1^T \mu, \quad Z \sim N_r(0, I)$$

Multiplicando por H :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \equiv \begin{pmatrix} D^{\frac{1}{2}}Z + H_1^T \mu \\ 0Z + H_2^T \mu \end{pmatrix} = \begin{pmatrix} D^{\frac{1}{2}} \\ 0 \end{pmatrix} Z + H^T \mu$$

Vemos que:

$$X = HY = H \begin{pmatrix} D^{\frac{1}{2}} \\ 0 \end{pmatrix} Z + HH^T \mu = H_1 D^{\frac{1}{2}} Z + \mu$$

Denotando $A = H_1 D^{\frac{1}{2}}$ se tiene finalmente:

$$X = AZ + \mu$$

con A una matriz $p \times r$, con $r \leq p$, de rango r y $Z \sim N_r(0, I)$.

Recíprocamente, supongamos que se cumple lo anterior. Usando 1.4,

$$\Psi_X(t) = \Psi_{AZ+\mu}(t) = \exp\{it^T \mu\} \Psi_Z(A^T t)$$

como $\Psi_Z(t) = \exp\{-\frac{1}{2}t^T t\}$, entonces:

$$\Psi_X(t) = \exp\left\{-\frac{1}{2}(A^T t)^T (A^T t)\right\} \exp\{it^T \mu\} = \exp\left\{-\frac{1}{2}t^T A A^T t\right\} \exp\{it^T \mu\}$$

y, como tenemos:

- $\mu_X = E[X] = A * 0 + \mu = \mu$
- $\Sigma_X = \text{Cov}(X) = A I A^T = A A^T$

Nos queda:

$$\Psi_X(t) = \exp\left\{it\mu - \frac{1}{2}t^T (A A^T) t\right\} \implies X \sim N_p(\mu, A A^T), \quad A A^T \geq 0$$

Con todo este procedimiento, podemos enunciar la siguiente caracterización, que es una extensión de [C-I].

Proposición 3.3 (Caracterización I). *Un vector aleatorio $X = (X_1, \dots, X_p)^T$ tiene DNM de vector de medias μ y matriz de covarianzas Σ (con $\Sigma \geq 0$), si y solo si se puede representar:*

$$X = AZ + \mu$$

con $A \in \mathcal{M}_{p \times r}(\mathbb{R})$, de rango r , $A A^T = \Sigma$, $Z \sim N_r(0, I)$.

Enunciamos también una nueva versión de [C-III]:

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

Proposición 3.4 (Caracterización III). Un vector aleatorio $X = (X_1, \dots, X_p)^T$ tiene DNM (con $\Sigma \geq 0$) si y solo si toda combinación lineal de las componentes X_1, \dots, X_p de la forma:

$$\alpha^T X, \quad \alpha^T \in \mathbb{R}^p$$

tiene distribución normal (univariante, posiblemente degenerada)

Demostración. Es análoga al caso anterior. □

Definición 3.2. En el subespacio afín generado por los autovectores correspondientes a autovalores no nulos, se tiene que se puede formular la función de densidad normal multivariante.

Proposición 3.5 (Normalidad de C.L. de vectores con DNM independientes). Sean X_k con $k = 1, \dots, m$ v.a. p -dimensionales independientes con distribución $N_p(\mu_k, \Sigma_k)$ respectivamente. Entonces, para cualquier conjunto de matrices constantes A_k con $k = 1, \dots, m$ de dimensión $q \times p$, se verifica que:

$$Y := \sum_{k=1}^m A_k X_k \sim N_q \left(\sum_{k=1}^m (A_k \mu_k), \sum_{k=1}^m (A_k \Sigma_k A_k^T) \right).$$

Teorema 3.1 (Cramer). Sean X_1 y X_2 vectores aleatorios y p -dimensionales independientes, con $X_1 \sim N_p(\mu_1, \Sigma_1)$ y $X_2 \sim N_p(\mu_2, \Sigma_2)$. Entonces, se tiene que $X_1 + X_2 \sim N_p(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$.

La conjetura de Paul Lévy preguntaba si el recíproco de esta afirmación es cierta. Cramer dio una solución para $p = 1$ en el año 1936, de la que es sencillo inducir la solución para cualquier p .

Teorema 3.2 (Recíproco del teorema de Cramer). Sean X_1 y X_2 vectores aleatorios y p -dimensionales independientes, tales que $X_1 + X_2$ tiene DNM. Entonces, X_1 y X_2 también tienen DNM.

Ejercicio 3.1. Suponiendo que este último resultado es cierto para $p = 1$, probar que es cierto para cualquier p .

3.2. Distribuciones esféricas y elípticas

3.2.1. Caracterización en términos de la densidad esférica estándar (C-IV)

Vamos a comenzar razonando cómo podemos trabajar con coordenadas polares. Sea $X \sim N_p(\mu, \Sigma)$, con $\Sigma > 0$. Para cada $k > 0$ la ecuación

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = k^2$$

define un elipsoide en \mathbb{R}^p sobre el que la densidad f_X es constante.

Sea $\Sigma = CC^T$, con $C_{p \times p}$ no singular. Definimos $Z = C^{-1}(X - \mu)$, con $Z \sim N_p(0, I_p)$, a partir del cual vamos a definir el vector aleatorio —excepto para el

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

caso $\mathbf{Z} = \mathbf{0}$ —

$$\mathbf{U} = \frac{\mathbf{Z}}{\|\mathbf{Z}\|} = \frac{\mathbf{Z}}{(\mathbf{Z}^T \mathbf{Z})^{1/2}},$$

que tiene norma 1. Observamos que \mathbf{U} se distribuye sobre la esfera unidad en \mathbb{R}^p .

Definimos a continuación $H\mathbf{U}$, con H matriz ortogonal. Entonces,

$$H\mathbf{U} = H \frac{\mathbf{Z}}{(\mathbf{Z}^T \mathbf{Z})^{1/2}} = \frac{H\mathbf{Z}}{(\mathbf{Z}^T H^T H \mathbf{Z})^{1/2}} = \frac{H\mathbf{Z}}{((H\mathbf{Z})^T (H\mathbf{Z}))^{1/2}}.$$

Observamos que $H\mathbf{Z} \sim N_p(H\mathbf{0}, H\mathbf{I}_p H^T) = N_p(\mathbf{0}, \mathbf{I}_p)$. Por tanto tenemos que \mathbf{U} y $H\mathbf{U}$ se distribuyen de igual forma. Ahora, como la distribución de \mathbf{U} no se ve afectada bajo ninguna isometría —multiplicando por una matriz ortogonal—, concluimos que \mathbf{U} se distribuye uniformemente sobre la esfera unidad en \mathbb{R}^p , lo que llamamos *densidad esférica estándar*. Para recuperar la información de \mathbf{Z} a partir de \mathbf{U} necesitamos conocer el «radio» $R = (\mathbf{Z}^T \mathbf{Z})^{1/2} = \|\mathbf{Z}\|$.

Proposición 3.6. El vector aleatorio $\mathbf{U} = \frac{\mathbf{Z}}{(\mathbf{Z}^T \mathbf{Z})^{1/2}}$ y la variable aleatoria $R = (\mathbf{Z}^T \mathbf{Z})^{1/2}$ son independientes.

Proposición 3.7. Un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ tiene DNM $N_p(\boldsymbol{\mu}, \Sigma)$ con $\Sigma > 0$ si, y solo si, \mathbf{X} es un vector aleatorio correspondiente al siguiente «experimento»:

1. Obtener el valor de una variable aleatoria $R^2 := V \sim \chi_p^2$. Se identifica $V = v$.
2. Elegir un punto aleatoriamente —es decir, según la distribución uniforme— en la esfera p -dimensional de radio $r = \sqrt{v}$.

$$\mathbf{Z} = (Z_1, \dots, Z_p)^T = (z_1, \dots, z_p)^T = \mathbf{z}.$$

3. Dado el vector de medias $\boldsymbol{\mu}$ y la matriz de covarianzas $\Sigma > 0$, elegir C (no singular) tal que $\Sigma = CC^T$ y obtener el valor de \mathbf{X} mediante la expresión $\mathbf{X} := C\mathbf{Z} + \boldsymbol{\mu}$. De nuevo se identifica $\mathbf{X} = \mathbf{x} = C\mathbf{z} + \boldsymbol{\mu}$.

Demostración. Veamos la justificación de la implicación (\implies). Por la caracterización (C-I) sabemos que podemos escribir $\mathbf{X} = C\mathbf{Z} + \boldsymbol{\mu}$, con $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$.

Escribimos, para $\mathbf{Z} \neq \mathbf{0}$,

$$\mathbf{Z} = (\mathbf{Z}^T \mathbf{Z})^{1/2} \cdot \frac{\mathbf{Z}}{(\mathbf{Z}^T \mathbf{Z})^{1/2}} = R \cdot \mathbf{U},$$

siendo R y \mathbf{U} independientes.

Ahora tenemos que:

•

$$R^2 = \mathbf{Z}^T \mathbf{Z} = Z_1^2 + \dots + Z_p^2 \sim \chi_p^2$$

al ser Z_1, \dots, Z_p independientes con $Z_i \sim N_p(0, 1)$ para $i = 1, \dots, p$,

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

- U tiene distribución uniforme sobre la esfera unidad.

A partir de la expresión $X = C \cdot R \cdot U + \mu$ se tiene que para generar los valores de X se puede seguir el procedimiento del enunciado. \square

3.3. Distribuciones esféricas y elípticas

3.3.1. Distribuciones esféricas

Definición 3.3. Se dice que un vector aleatorio $X = (X_1, \dots, X_p)^T$ tiene una distribución *esférica* si para toda matriz ortogonal H los vectores aleatorios X y HX tienen la misma distribución.

Veamos a continuación dos caracterizaciones de las distribuciones esféricas.

Proposición 3.8 (Caracterización en términos de la densidad). Dado un X un vector aleatorio p -dimensional con función de densidad f_X tiene distribución esférica si y solo si f_X puede expresarse como $g(x^T x)$ para alguna función escalar $g : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$.

Proposición 3.9 (Caracterización en términos de la función característica). Dado un X un vector aleatorio p -dimensional con función de densidad f_X tiene distribución esférica si y solo si su función característica tiene la forma $\xi(t^T t)$ para alguna función escalar $\xi : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$.

3.3.2. Distribuciones elípticas

Definición 3.4 (Distribución elíptica). Se dice que un vector aleatorio $X = (X_1, \dots, X_p)^T$ tiene *distribución elíptica* de parámetros $\mu \in \mathbb{R}^p$ y $V \in \mathcal{M}_p(\mathbb{R})$ —simétrica y definida positiva— si se puede expresar de la forma $X = AZ + \mu$, con $V = AA^T$ y Z un vector aleatorio con distribución esférica. Se denota como $X \sim E_p(\mu, V)$.

Proposición 3.10 (Caracterización en función de la densidad). Dado un X un vector aleatorio p -dimensional con función de densidad f_X tiene distribución elítica si y solo si f_X puede expresarse como

$$|V|^{-1/2} g((x - \mu)V^{-1}(x - \mu))$$

para alguna función escalar $g : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$.

Proposición 3.11 (Caracterización en términos de la función característica). Dado un X un vector aleatorio p -dimensional con función de densidad f_X tiene distribución esférica si y solo si su función característica tiene la forma

$$e^{it^T \mu} \xi(t^T V t)$$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

para alguna función escalar $\xi : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$.

Proposición 3.12. Sea $X \sim E_p(\mu, V)$, con V diagonal. Si X_1, \dots, X_p son mutuamente independientes entonces X tiene distribución normal multivariante.

Proposición 3.13. Sea Z un vector aleatorio con distribución esférica tal que $P[Z = 0] = 0$. Entonces $U = Z/\|Z\|$ y $R = \|Z\|$ son independientes.

Demostración. Comenzamos viendo que dada una matriz H ortogonal cualquiera,

$$HU = \frac{HZ}{(Z^T Z)^{1/2}} = \frac{HZ}{(Z^T H^T H Z)^{1/2}} = \frac{HZ}{((HZ)^T (HZ))^{1/2}} = \frac{Z}{(Z^T Z)^{1/2}} = U,$$

luego U se distribuye uniformemente en la esfera unidad en \mathbb{R}^p .

A continuación probamos la independencia de U y R . Para cada $C \in \mathcal{B}(\mathbb{R}^+)$ tal que $P[R \in C] \neq 0$ queremos probar que $\mu = P[U \in B \mid R \in C] = P[U \in B]$ para todo $B \in \mathcal{B}(S_p)$. Por definición,

$$P[U \in B \mid R \in C] = \frac{P[(U \in B) \wedge (R \in C)]}{P[R \in C]} = \frac{P[Z \in B \cdot C]}{P[R \in C]},$$

donde $B \cdot C = \{z \in \mathbb{R}^p - \{0\} : z/\|z\| \in B, \|z\| \in C\}$. Luego,

$$\frac{P[Z \in B \cdot C]}{P[R \in C]} = \frac{P[H^T Z \in B \cdot C]}{P[R \in C]}$$

para cualquier matriz H ortogonal por ser la distribución de Z esférica. Vemos que $H^T Z \in B \cdot C$ si y solo si $Z \in (H(B)) \cdot C$. En efecto, $H^T Z \in B \cdot C$ si y solo si $H(H^T Z) \in H(B \cdot C)$, siendo

$$\begin{aligned} H(B \cdot C) &= \{H z : z \in B \cdot C\} \\ &= \left\{ H z : \frac{z}{\|z\|} \in B, \|z\| \in C \right\} \\ &= \left\{ H z : \frac{H z}{\|H z\|} \in H B, \|H z\| \in C \right\} \end{aligned}$$

y llamando $w = H z$,

$$\begin{aligned} &= \left\{ w : \frac{w}{\|w\|} \in H B, \|w\| \in C \right\} \\ &= H(B) \cdot C, \end{aligned}$$

probando entonces esta última afirmación. Concluimos entonces que

$$\begin{aligned} \frac{P[H^T Z \in B \cdot C]}{P[R \in C]} &= \frac{P[Z \in (H(B)) \cdot C]}{P[R \in C]} \\ &= \frac{P[(U \in H(B)) \wedge (R \in C)]}{P[R \in C]} \\ &= P[U \in H(B) \mid R \in C] \\ &= \mu(H(B)). \end{aligned}$$

□

3.4. Formas cuadráticas basadas en vectores aleatorios normales

Los resultados presentados en esta subsección nos serán de utilidad cuando nos adentremos en la inferencia en la sección siguiente.

El resultado que motiva esta subsección es el siguiente. Sea $X = (X_1, \dots, X_p)^T$ un vector aleatorio y sea A una matriz constante $p \times p$. Consideremos la forma cuadrática $X^T A X$ —que es un vector aleatorio—. En particular, si $X \sim N_p(\mu, \Sigma)$ bajo ciertas condiciones sobre A se tendrá que la distribución de $X^T A X$ es una distribución χ^2 no centrada.

3.5. Distribuciones χ^2 y F no centradas

3.5.1. Distribución χ^2 no centrada

Definición 3.5 (Distribución χ^2 centrada). Sea Y una variable aleatoria. Decimos que Y sigue una *distribución χ^2 centrada* si es la suma de los cuadrados de variables aleatorias independientes con distribución normal estándar $N(0, 1)$. Es decir, dado $Z = (Z_1, \dots, Z_n)^T \sim N_n(0, I_n)$ tal que $\|Z\|^2 = Z^T Z$,

$$Y = \sum_{i=1}^n Z_i^2 \sim \chi_n^2.$$

Su función de densidad es

$$f_Y(y) = \frac{1}{2^{n/2} \Gamma(n/2)} y^{(n/2)-2} e^{-y/2}, \quad \text{para } y > 0,$$

y su función de distribución,

$$F_Y(y) = \frac{\gamma(n/2, y/2)}{\Gamma(n/2)}, \quad \text{para } y > 0.$$

Recordemos que la función gamma está definida como:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \quad \text{para } z \in \mathbb{C} \text{ tales que } \operatorname{Re}(z) > 0.$$

Además, las funciones gamma incompletas son:

- $\Gamma(z, v) = \int_v^\infty t^{z-1} e^{-t} dt$.
- $\gamma(0, v) = \int_0^v t^{z-1} e^{-t} dt$ con $v > 0$.

Definición 3.6 (Distribución χ^2 no centrada). Sea $X = (X_1, \dots, X_n)^T \sim N_n(\mu, I_n)$. Entonces, la v.a. $Y = X X^T$ tiene la función de densidad:

$$f_Y(y) = e^{\frac{\delta}{2}} {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\delta y\right) \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}, \quad y > 0$$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

siendo $\delta = \mu^T \mu$. Se dice que Y tiene distribución χ^2 no centrada con n grados de libertad y parámetro de no centralidad δ , denotándose $\chi_n^2(\delta)$

Nota. ${}_0F_1$ es la función hipergeométrica generalizada de órdenes 0 y 1, también llamada función hipergeométrica confluyente límite.

Proposición 3.14 (Momentos de primer y segundo orden). • $E[Y] = n + \delta$
• $Var(Y) = 2n + 4\delta$

Nota. $\Psi_Y^{(k)}(0) = \frac{d^k \Psi_Y(t)}{dt^k} \Big|_{t=0} = t^k E[Y^k]$

Proposición 3.15. Si Z_1 y Z_2 son dos v.a. independientes con

$$Z_1 \sim \chi_{n_1}^2(\delta_1) \quad (1)$$

$$Z_2 \sim \chi_{n_2}^2(\delta_2) \quad (2)$$

entonces, se tiene que:

$$Z_1 + Z_2 \sim \chi_{n_1+n_2}^2(\delta_1 + \delta_2)$$

3.5.2. Funciones hipergométricas generalizadas

Definición 3.7 (Función hipergeométrica generalizada). Se denomina función (serie) hipergeométrica generalizada de órdenes p y q a

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k \dots (a_p)_k}{(b_1)_k \dots (b_q)_k} \frac{z^k}{k!}$$

donde $(a)_k = a(a+1)\dots(a+k-1)$ con $a_1, \dots, b_1, \dots, b_q$ parámetros (posiblemente complejos) y $z \in \mathbb{C}$ el argumento

Nota. • Ningún parámetro b_j puede ser cero o entero negativo
• Si algún parámetro en el numerador es cero o un entero negativo, los términos de la serie se anulan a partir de un k y queda un polinomio en z
• La serie converge para todo z si $p \leq q$, converge para $|z| < 1$ y diverge para $|z| > 1$ si $p = q + 1$ y diverge para todo $z \neq 0$ si $p > q + 1$
• Caso particular: ${}_2F_1$ es la función hipergeométrica clásica o gaussiana
• Caso particular: ${}_1F_1$ es la función hipergeométrica confluyente

${}_0F_1$ se define como ${}_0F_1(; b; z) = \lim_{a \rightarrow \infty} {}_1F_1(a; b; z)$

Función característica de la distribución $\chi_n^2(\delta)$:

$$\Psi_{\chi_n^2(\delta)}(t) = (1 - 2it)^{-\frac{n}{2}} e^{\frac{it\delta}{1-2it}} \quad \forall t \in \mathbb{R}$$

$$X \sim N(\mu, \sigma^2) \rightarrow \Psi_{X^2}(t) = \frac{1}{(1-2i\sigma^2 t)^{\frac{1}{2}}} e^{\frac{it\mu^2}{1-2i\sigma^2 t}}$$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

3.5.3. Distribución F de Snedecor no centrada

Definición 3.8 (Distribución F de Snedecor centrada). Si $Y_1 \sim \chi_{n_1}, Y_2 \sim \chi_{n_2}$ centradas e independientes, la distribución F centrada se define como la distribución del cociente:

$$F = \frac{Y_1/n_1}{Y_2/n_2}$$

Se denota F_{n_1, n_2} o $F(n_1, n_2)$

También puede verse como:

$$F = \frac{\sum_{k=1}^{n_1} Z_{1k}^2}{\sum_{l=1}^{n_2} Z_{2l}^2}$$

con

- $Z_1 = (Z_{11}, \dots, Z_{1n})^T \sim N_{n_1}(0, I_{n_1})$
- $Z_2 = (Z_{21}, \dots, Z_{2n})^T \sim N_{n_2}(0, I_{n_2})$
- Z_1, Z_2 independientes

o , equivalentemente:

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N_{n_1+n_2}(0, I_{n_1+n_2})$$

Su función de densidad viene dada por

$$g_F(f) = \frac{1}{f B(\frac{n_1}{2}, \frac{n_2}{2})} \left(\frac{n_1 f}{n_1 f + n_2} \right)^{\frac{n_1}{2}} \left(1 - \frac{n_1 f}{n_1 f + n_2} \right)^{\frac{n_2}{2}}$$

para $f \geq 0$ con $n_1, n_2 \in \mathbb{N} - \{0\}$.

La función de distribución viene dada por:

$$G_F(f) = I_{\frac{n_1 f}{n_1 f + n_2}} \left(\frac{n_1}{2}, \frac{n_2}{2} \right)$$

Donde esta I es la función beta incompleta regularizada, que definimos ahora.

Definición 3.9 (Función Beta). La función beta viene dada por:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 t^{x-1}(1-t)^{y-1} dt$$

donde $x, y \in \mathbb{C}, \text{Re}(x), \text{Re}(y) > 0$.

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

Definición 3.10 (Función Beta Regularizada). Esta función viene dada por:

$$B(x; a, b) = \int_0^x t^{a-1} (a-t)^{b-1} dt$$

con $x \in [0, 1]$, $a, b \in \mathbb{C}$, $\text{Re}(a), \text{Re}(b) > 0$.

Con estas dos funciones, definimos:

Definición 3.11 (Función beta incompleta regularizada).

$$I_X(a, b) = \frac{B(x; a, b)}{B(a, b)}$$

Proposición 3.16 (Función característica de F). La función característica viene dada por:

$$\Psi_{F_{n_1, n_2}}(t) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_2}{2})} U(\frac{n_1}{2}, 1 - \frac{n_2}{2}; -\frac{n_2}{n_2} it)$$

donde U es la función hipergeométrica confluyente de segunda especie:

$$U(a, b; z) = \frac{\Gamma(1-b)}{\Gamma(a+1-b)} {}_1F_1(a, b; z) + \frac{\Gamma(b-1)}{\Gamma(a)} z^{1-b} {}_1F_1(a+1-b, 2-b; z)$$

Proposición 3.17 (Momentos de primer y segundo orden). • $E[F] = \frac{n_2}{n_2-2}$ para $n_2 > 2$

• $\text{Var}(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ para $n_2 > 4$.

Si $n_2 \in [0, 2]$ no está definida. Para $n_2 \in (2, 4]$, la varianza es infinita.

Definición 3.12. Sea $Y_1 \sim \chi_{n_1}^2(\delta)$ e $Y_2 \sim \chi_{n_2}^2$ independientes. Entonces, la variable

$$F = \frac{Y_1/n_1}{Y_2/n_2}$$

tiene función de densidad:

$$g_F(t) = e^{-\frac{\delta}{2}} {}_1F_1(\frac{1}{2}(n_1+n_2); \frac{1}{2}n_1; \frac{\frac{1}{2}n_1}{2t})$$

3.5.4. Resultados sobre formas cuadráticas

Sea $X^T A X$ una forma cuadrática con $X \sim N_p(\mu, \Sigma)$. Podemos hacer un estudio directo sobre $E[X^T A X]$ y $\text{Var}(X^T A X)$.

Proposición 3.18. Sea $X \sim N(\mu, \Sigma)$, con $\Sigma > 0$ y A una matriz $p \times p$ simétrica. Entonces:

1. $E[X^T A X] = \text{tr}(A\Sigma) + \mu^T A \mu$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

$$2. \text{Var}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = 2 \text{tr}((\mathbf{A} \Sigma)^2) + 4 \mu^T \mathbf{A} \Sigma \mathbf{A} \mu$$

Demostración. Vamos a probar el primer elemento de los dos, pues puede que su demostración nos sera útil más adelante.

$$E[\mathbf{X}^T \mathbf{A} \mathbf{X}] = E[\text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X})] = E[\text{tr}(\mathbf{A} \mathbf{X} \mathbf{X}^T)] = \text{tr}(E[\mathbf{A} \mathbf{X} \mathbf{X}^T]) = \text{tr}(\mathbf{A} E[\mathbf{X} \mathbf{X}^T]) = (*)$$

Donde en la primera igualdad hemos usado que $\mathbf{X}^T \mathbf{A} \mathbf{X}$ es $D \times 1$ y en la segunda hemos usado una propiedad de las matrices que dice que si $C_{m \times n}$, $B_{n \times m}$, entonces $\text{tr}(BC) = \text{tr}(CB)$. En las siguientes, hemos usado la linealidad de la esperanza. Ahora como sabemos que:

$$\Sigma = E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] = \dots = E[\mathbf{X} \mathbf{X}^T] - \mu \mu^T$$

Y usando esto en la igualdad anterior, tenemos:

$$(*) = \text{tr}(\mathbf{A} E[\mu \mu^T]) = \text{tr}(\mathbf{A} \Sigma + \mathbf{A} \mu \mu^T) = \text{tr}(\mathbf{A} \Sigma) + \text{tr}(\mathbf{A} \mu \mu^T) = \quad (3)$$

$$= \text{tr}(\mathbf{A} \Sigma) + \text{tr}(\mu^T \mathbf{A} \mu) = \text{tr}(\mathbf{A} \Sigma) + \mu^T \mathbf{A} \mu \quad (4)$$

□

Llegamos pues al siguiente teorema:

Teorema 3.3. Sea $\mathbf{X} \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$. Entonces:

1. $(\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) \sim \chi_p^2$
2. $\mathbf{X}^T \Sigma^{-1} \mathbf{X} \sim \chi_p^2(\delta)$ con $\delta = \mu^T \Sigma^{-1} \mu$

Demostración. 1. Hacemos el cambio $\mathbf{Y} = C^{-1}(\mathbf{X} - \mu)$ con $\Sigma = CC^T$. Tenemos entonces que $\mathbf{Y} \sim N_p(0, I_p)$. Por tanto, podemos escribir:

$$(\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) = (\mathbf{X} - \mu)^T (CC^T)^{-1} (\mathbf{X} - \mu) = \dots = (C^{-1}(\mathbf{X} - \mu))^T (C^{-1}(\mathbf{X} - \mu)) = \mathbf{Y} \mathbf{Y}^T \sim \chi_p^2$$

2. En este caso, tomamos el cambio $\mathbf{V} = C^{-1} \mathbf{X}$ con $\Sigma = CC^T$. Entonces, se tiene que

$$\mathbf{V} \sim N_p(C^{-1} \mu, C^{-1} \Sigma (C^{-1})^T) \equiv N_p(C^{-1} \mu, C^{-1} CC^T (C^{-1})^T) \equiv N_p(C^{-1} \mu, I_p)$$

Y por tanto, escribimos:

$$\mathbf{X}^T \Sigma^{-1} \mathbf{X} = \mathbf{X}^T (CC^T)^{-1} \mathbf{X} = \dots = (C^{-1} \mathbf{X})^T (C^{-1} \mathbf{X}) = \mathbf{V}^T \mathbf{V} \sim \chi_p^2(\mu^T (C^{-1})^T (C^{-1} \mu)) \equiv \chi_p^2(\mu^T \Sigma^{-1} \mu)$$

□

Teorema 3.4. Sea $\mathbf{X} \sim N_p(\mu, I_p)$ y sea $B_{p \times p}$ simétrica. Entonces, $\mathbf{X}^T \mathbf{B} \mathbf{X}$ tiene una distribución χ^2 no centrada si y solo si \mathbf{B} es idempotente (i.e., $\mathbf{B}^2 = \mathbf{B}$), en cuyo

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

caso los grados de libertad y el parámetro de no centralidad son, respectivamente

$$k = \text{rango}(B) = \text{tr}(B) \quad (5)$$

$$\delta = \mu^T B \mu \quad (6)$$

Demostración. \Rightarrow Supongamos que $\mathbf{X}^T B \mathbf{X} \sim X_k(\delta)$. Sea $r = \text{rango}(B)$. Como B es simétrica, se puede descomponer como

$$H^T B H = D \mapsto B = H D H^T$$

con $H_{p \times p}$ ortogonal y $D_{p \times p}$ diagonal, que tiene en las r primeras filas, los $\lambda_i \neq 0$ con $i = 1 \dots r$. Hacemos entonces el cambio $V = H^T \mathbf{X}$. Entonces:

$$\nu \sim N_p(H^T \mu, H^T I_p H) \equiv N_p(H^T \mu, I_p) \equiv N_p(D, I_p)$$

donde hemos denotado $\nu = H^T \mu$. Podemos escribir:

$$\mathbf{X}^T B \mathbf{X} = \mathbf{X}^T H D H^T \mathbf{X} = (H^T \mathbf{X})^T D (H^T \mathbf{X}) = V^T D V = \sum_{j=1}^r \lambda_j \nu_j^2$$

donde $\nu_j^2 \sim \chi_1^2(\nu_j^2)$ y siendo ν_1^2, \dots, ν_r^2 independientes. La función característica de $\mathbf{X}^T B \mathbf{X}$ se obtiene como

$$\Psi_{\mathbf{X}^T B \mathbf{X}}(t) = \Psi_{\sum_{j=1}^r \lambda_j \nu_j^2}(t) = \prod_{j=1}^r \Psi_{\lambda_j \nu_j^2}(t) = \prod_j \Psi_{\nu_j^2}(\lambda_j t)$$

y, como cada función característica individual $\Psi_{\nu_j^2}(\cdot)$ tiene la forma

$$\Psi_{\chi_1^2}(s) = (1 - 2is)^{-\frac{1}{2}} \exp \left\{ \frac{is\delta}{1 - 2is} \right\}, \quad \delta = \nu_j^2$$

luego, conjuntamente, se tiene:

$$\begin{aligned} \Psi_{\mathbf{X}^T B \mathbf{X}}(t) &= \prod_{j=1}^r (1 - 2i\lambda_j t)^{-\frac{1}{2}} \exp \left\{ \frac{i\lambda_j t \nu_j^2}{1 - 2i\lambda_j t} \right\} \\ &= \exp \left\{ it \sum_{j=1}^r \frac{\lambda_j \nu_j^2}{1 - 2i\lambda_j t} \right\} \prod_{j=1}^r (1 - 2i\lambda_j t)^{\frac{1}{2}}, \quad \text{para todo } t \in \mathbb{R}. \end{aligned}$$

Por otra parte, por la hipótesis $\mathbf{X}^T B \mathbf{X} \sim \chi_k^2(\delta)$, con función característica:

$$\Psi_{\mathbf{X}^T B \mathbf{X}}(t) = \exp \left\{ \frac{it\delta}{1 - 2it} \right\} (1 - 2it)^{-\frac{k}{2}}, \quad \text{para todo } t \in \mathbb{R}.$$

Así que comparando ambas expresiones, puesto que deben ser iguales para todo $t \in \mathbb{R}$, debe tenerse que:

- $\lambda_j = 1$ con $j = 1, \dots, r$
- $\delta = \sum_{j=1}^r \lambda_j \nu_j^2 = \sum_{j=1}^r \nu_j^2 = \nu^T D \nu = (\mu^T H) D H^T \mu = \mu^T H D H^T \mu$

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

- $r = k$

En consecuencia, podemos escribir:

$$H^T B H = D = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$$

pues es una matriz idempotente, es decir:

$$(H^T B H)(H^T B H) = H^T B H,$$

pero multiplicando a la izquierda por H y a la derecha por H^T (con $H^T H = H H^T = I$), se tiene que la identidad anterior es:

$$H^T B^2 H = H^T B H \implies B^2 = B.$$

$\boxed{\Leftarrow}$ Supongamos ahora que B es idempotente. Se puede escribir:

$$H^T B H = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$$

para cierta matriz ortogonal H^T . Recordamos que una matriz simétrica es idempotente si y solo si sus autovalores son ceros o unos. Ahora, definimos $V = H^T X$. Se tiene que:

$$V \sim N_p(H^T \mu, H^T I_p H) \equiv N_p(H^T \mu, I_p)$$

por lo que

$$X^T B X = X^T H \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} H^T X = V^T \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} V = \sum_{j=1}^k V_j^2$$

y, puesto que V_j , con $j = 1, \dots, k$ son independientes, se tiene que

$$X^T B X \sim \chi_k^2(\delta)$$

con

$$\delta = \sum_{j=1}^k (E[V_j])^2 = E[V^T] \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} E[V] = \quad (7)$$

$$= E[X^T H] \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} E[H^T X] = \quad (8)$$

$$= E[X^T] H \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} H^T E[X] = \mu^T B \mu \quad (9)$$

□

Teorema 3.5. Sea $X \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$. Supongamos el particionamiento:

3. Distribución normal multivariante en el caso general $\Sigma \geq 0$.

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} ; \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} ; \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

con \mathbf{X}_1 y μ_1 son $q \times 1$, y Σ_{11} es $q \times q$. Entonces:

$$Q = (\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) - (\mathbf{X}_1 - \mu_1)^T \Sigma_{11}^{-1} (\mathbf{X}_1 - \mu_1) \sim \chi_{p-q}^2$$

Nota. Como vemos, no se está exigiendo ninguna independencia entre \mathbf{X}_1 y \mathbf{X}_2 .

Demostración. Sea $C_{p \times p}$ no singular tal que $\Sigma = CC^T$. Podemos escribir $C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$ con $C_1 q \times p$, $C_2 (p-q) \times p$. Dado que

$$CC^T = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} (C_1^T C_2^T) = \begin{pmatrix} C_1 C_1^T & C_1 C_2^T \\ C_2 C_1^T & C_2 C_2^T \end{pmatrix}$$

identificamos $\Sigma_{11} = C_1 C_1^T$ y hacemos el cambio $U = C^{-1}(\mathbf{X} - \mu)$, que es lo mismo que $\mathbf{X} = CU + \mu$. Así: $\mathbf{X}_1 = C_1 U + \mu_1$.

Sabemos también que $U \sim N_p(0, I_p)$, por lo que:

$$Q = U^T U - U^T C_1^T (C_1 C_1^T)^{-1} C_1 U = U^T [I - C_1^T (C_1 C_1^T)^{-1} C_1] U$$

Ahora, aplicando el teorema anterior al vector U y la matriz $B = I - C_1^T (C_1 C_1^T)^{-1} C_1$, tenemos que $U^T B U$ tiene una distribución χ^2 no centrada (con parámetro de no centralidad $\delta = 0^T B 0 = 0$, es decir, centrado), si y solo si B es idempotente.

$$\begin{aligned} BB &= [I - C_1^T (C_1 C_1^T)^{-1} C_1] [I - C_1^T (C_1 C_1^T)^{-1} C_1] = \\ &= I - 2C_1^T (C_1 C_1^T)^{-1} C_1 + C_1^T (C_1 C_1^T)^{-1} (C_1 C_1^T) (C_1 C_1^T)^{-1} C_1 = \\ &= I - 2C_1^T (C_1 C_1^T)^{-1} C_1 + C_1^T (C_1 C_1^T)^{-1} C_1 = I - C_1^T (C_1 C_1^T)^{-1} C_1 = B \end{aligned}$$

Por lo que B es idempotente. Ahora, $\text{rango}(B) = \text{tr}(B) = \text{tr}(I_p - C_1^T (C_1 C_1^T)^{-1} C_1)$ y vamos que:

$$\text{tr}(I_p) - \text{tr}(C_1^T (C_1 C_1^T)^{-1} C_1) = \text{tr}(I_p) - \text{tr}((C_1 C_1^T)^{-1} C_1 C_1^T) = \text{tr}(I_p) - \text{tr}(I_q) = p - q$$

Por lo que $Q \sim \chi_{p-q}^2$. □

Teorema 3.6. Sea $\mathbf{X} \sim N_p(\mu, \Sigma)$, con $\Sigma > 0$ y $B_{p \times p}$ simétrica. Entonces, $\mathbf{X}^T B \mathbf{X}$ tiene una distribución $\chi_k^2(\delta)$ donde $k = \text{rango}(B)$ y $\delta = \mu^T B \mu$ si y solo si $B\Sigma$ es idempotente.

Nota. $B\Sigma B\Sigma = B\Sigma \iff B\Sigma B = B$

Demostración. Sea $C_{p \times p}$ no singular tal que $\Sigma = CC^T$ y sea $Y = C^T \mathbf{X}$. Se tiene que

$$Y \sim N_p(C^{-1}\mu, C^{-1}(CC^T)(C^{-1})^T) \equiv N_p(C^{-1}\mu, I_p)$$

por lo que

$$\mathbf{X}^T B \mathbf{X} = (CY)^T B (CY) = Y^T C^T B C Y$$

Aplicando un teorema anterior, tenemos que $X^T B X$ tiene una distribución χ_k^2 no centrada si y solo si la matriz $C^T B C$ es idempotente. La prueba se reduce a probar que $B\Sigma$ es idempotente si y solo si $C^T B C$ es idempotente. Veamoslo:

\Rightarrow Supongamos que $B\Sigma$ es idempotente, es decir, $B\Sigma B = B$. Entonces:

$$(C^T B C)(C^T B C) = C^T B C C^T B C = C^T B \Sigma B C = C^T B C \Rightarrow \text{idempotente}$$

\Leftarrow Supongamos ahora que $C^T B C$ es idempotente. Entonces:

$$C^T B C = (C^T B C)(C^T B C) = C^T B \Sigma B C$$

Multiplicamos a la izquierda por $(C^T)^{-1}$ y a la derecha por C^{-1}

$$(C^T)^{-1} C^T B C C^{-1} = B = B \Sigma B = (C^T)^{-1} (C^T B \Sigma B C) C^{-1} \Rightarrow B \Sigma \text{ idempotente}$$

En cuanto a k y δ , por el mismo teorema anterior tenemos que

$$k = \text{rango}(C^T B C) = \text{tr}(C^T B C) = \text{tr}(B C C^T) = \text{tr}(B \Sigma) = \text{rango}(B \Sigma) = \text{rango}(B)$$

, donde en las dos últimas igualdades hemos usado que $B\Sigma$ es idempotente y Σ es no singular. Veamos δ

$$\delta = (C^{-1} \mu)(C^T B C)(C^{-1} \mu) = \mu(C^T)^{-1} C^T B C C^{-1} \mu = \mu^T B \mu^T$$

□

4. Inferencia en la Distribución Normal Multivariante

4.1. Introducción

Consideremos una población que sigue una distribución normal multivariante $N_p(\mu, \Sigma)$ con $\Sigma > 0$. Para poder realizar la inferencia será necesario tomar una muestra aleatoria simple de dicha población, esto es, observaciones independientes idénticamente distribuidas. La denotaremos por

$$\{X_\alpha : \alpha = 1, \dots, N\},$$

donde N es el tamaño muestral y X_{α_i} , la componente i -ésima correspondiente al vector X_i de la observación α -ésima.

Nuestro objetivo inicial será estimar μ y Σ . El enfoque que tomaremos será la maximización de la función de verosimilitud.

Definición 4.1 (Media muestral). Llamamos *media muestral* al vector de \mathbb{R}^p construido de la forma:

$$\bar{X} = \frac{1}{N} \sum_{\alpha=1}^N X_{\alpha} = \begin{pmatrix} \frac{1}{N} \sum_{\alpha=1}^N X_{\alpha_1} \\ \vdots \\ \frac{1}{N} \sum_{\alpha=1}^N X_{\alpha_p} \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{pmatrix}$$

Definición 4.2 (Matriz de dispersiones muestral). Llamamos *matriz de dispersiones muestral* con respecto a \bar{X} a la matriz

$$A = \sum_{\alpha=1}^N (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})^T.$$

A partir de la definición de matriz de dispersiones muestral podemos realizar las siguientes definiciones.

Definición 4.3. Definimos:

- La *matriz de covarianzas muestral*, $S_N = A/N$.
- La *matriz de quasi-covarianzas muestral*, $S_N = A/(N-1)$.
- La *matriz de correlaciones muestral*,

$$R = D^{-1/2} S_N D^{-1/2},$$

donde

$$D = \text{diag}(S_{11}, \dots, S_{pp}) \quad \text{y} \quad D^{-1/2} = \text{diag}\left(\frac{1}{\sqrt{S_{11}}}, \dots, \frac{1}{\sqrt{S_{pp}}}\right).$$

El siguiente resultado nos será útil en desarrollos posteriores y es independiente de la distribución multivariante que se esté considerando.

Lema 4.1. Sea $\{X_{\alpha} : \alpha = 1, \dots, N\}$ una muestra de una población p -dimensional y sea \bar{X} su media muestral. Entonces, para cualquier vector $b \in \mathbb{R}^p$ se verifica que

$$\sum_{\alpha=1}^N (X_{\alpha} - b)(X_{\alpha} - b)^T = \sum_{\alpha=1}^N (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})^T + N(\bar{X} - b)(\bar{X} - b)^T.$$

En particular, si tomamos

- el vector $b = \mu$ (si existe), entonces

$$\sum_{\alpha=1}^N (X_{\alpha} - \mu)(X_{\alpha} - \mu)^T = \sum_{\alpha=1}^N (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})^T + N(\bar{X} - \mu)(\bar{X} - \mu)^T.$$

4. Inferencia en la Distribución Normal Multivariante

- el vector $b = 0$, entonces

$$\sum_{\alpha=1}^N (X_{\alpha} - \mu)(X_{\alpha} - \mu)^T = \sum_{\alpha=1}^N (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})^T + N\bar{X}\bar{X}^T.$$

4.2. EMV de máxima verosimilitud de μ y Σ en una DNM

Definición 4.4 (Función de verosimilitud). La función de verosimilitud se expresa como la función de μ y Σ dada por:

$$L(\mu, \Sigma; x_1, \dots, x_N) = f_{(\mu, \Sigma)}(x_1, \dots, x_N)$$

Trabajamos ahora con esta función de verosimilitud:

$$\begin{aligned} L(\mu, \Sigma; x_1, \dots, x_N) &= f_{(\mu, \Sigma)}(x_1, \dots, x_N) = \\ &= \prod_{\alpha=1}^N f_{(\mu, \Sigma)}(X_{\alpha}) = \\ &= \prod_{\alpha=1}^N \frac{1}{(2\pi)^{pN/2} \det(\Sigma)^{N/2}} \exp\left\{-\frac{1}{2}(X_{\alpha} - \mu)^T \Sigma^{-1}(X_{\alpha} - \mu)\right\} = \\ &= \frac{1}{(2\pi)^{\frac{pN}{2}} \det(\Sigma)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2} \sum_{\alpha=1}^N (X_{\alpha} - \mu)^T \Sigma^{-1}(X_{\alpha} - \mu)\right\} \end{aligned}$$

Por conveniencia, expresamos la suma de forma cuadrática en el exponente de la forma siguiente:

$$\begin{aligned} \sum_{\alpha=1}^N (X_{\alpha} - \mu)^T \Sigma^{-1}(X_{\alpha} - \mu) &= \text{tr}\left(\sum_{\alpha=1}^N (X_{\alpha} - \mu)^T \Sigma^{-1}(X_{\alpha} - \mu)\right) \\ &= \sum_{\alpha=1}^N \text{tr}((X_{\alpha} - \mu)^T \Sigma^{-1}(X_{\alpha} - \mu)) \\ &= \sum_{\alpha=1}^N \text{tr}(\Sigma^{-1}(X_{\alpha} - \mu)(X_{\alpha} - \mu)^T) \\ &= \text{tr}\left(\sum_{\alpha=1}^N \Sigma^{-1}(X_{\alpha} - \mu)(X_{\alpha} - \mu)^T\right) \\ &= \text{tr}\left(\Sigma^{-1} \sum_{\alpha=1}^N (X_{\alpha} - \mu)(X_{\alpha} - \mu)^T\right) \\ &= \text{tr}(\Sigma^{-1}[A + N(X - \mu)(X - \mu)^T]) = \text{tr}(\Sigma^{-1}A) + N * \text{tr}(\Sigma^{-1}(X - \mu)(X - \mu)^T) \\ &= \text{tr}(\Sigma^{-1}A) + N * \text{tr}((X - \mu)^T \Sigma^{-1}(X - \mu)) = \\ &= \text{tr}(\Sigma^{-1}A) + N(X - \mu)^T \Sigma^{-1}(X - \mu) \end{aligned}$$

4. Inferencia en la Distribución Normal Multivariante

Así, tenemos que:

$$L(\mu, \Sigma) = \frac{1}{(2\Pi)^{\frac{pN}{2}} \det \Sigma^{\frac{N}{2}}} \exp\left\{-\frac{1}{2} \text{tr}(\Sigma^{-1}A) - \frac{N}{2}(\mathbf{X} - \mu)^T \Sigma^{-1}(\mathbf{X} - \mu)\right\}$$

por lo que

$$\ln L(\mu, \Sigma) = -\frac{pN}{2} \ln(2\Pi) - \frac{N}{2} \ln(\det \Sigma) - \frac{1}{2} \text{tr}(\Sigma^{-1}A) - \frac{N}{2}(\mathbf{X} - \mu)^T \Sigma^{-1}(\mathbf{X} - \mu)$$

Proposición 4.1. Sea $\mathbf{X} \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$, y sea $\{X_\alpha : \alpha = 1, \dots, N\}$ una m.a.s. de \mathbf{X} . Entonces, los EMV de μ y Σ son, respectivamente $\bar{\mathbf{X}}$ y $S_N = \frac{A}{N}$, bajo la condición de que A sea definida positiva.

Lema 4.2 (Watson). Sea $f(G) = -N \ln(|G|) - \text{tr}(G^{-1}D)$, G es una matriz simétrica definida positiva, D una matriz simétrica definida positiva. Entonces, el máximo de f se alcanza en $G = (\frac{1}{N}D)$ y su valor es $f(\frac{1}{N}D) = pN * \ln(N) - N * \ln(|D|) - pN$

La demostración de este resultado se puede encontrar en la página 50 del libro de Raón Gutiérrez.

Nos interesa ahora maximizar $L(\mu, \Sigma)$, que es lo mismo que maximizar $\ln[L(\mu, \Sigma)]$, con $\Sigma > 0$, para maximizar en función de μ solo atendemos al término:

$$-\frac{N}{2}(\mathbf{X} - \mu)^T \Sigma^{-1}(\mathbf{X} - \mu)$$

este es negativo, y en valor máximo sería 0, alcanzado en $\bar{\mathbf{X}} = \mu$

$$\ln(L(\bar{\mathbf{X}}, \Sigma)) = -\frac{pN}{2} \ln(2\Pi) - \frac{N}{2} \ln(|\Sigma|) - \frac{1}{2} \text{tr}(\Sigma^{-1}A) + 0 = -\frac{pN}{2} \ln(2\Pi) - \frac{1}{2} f(\Sigma)$$

donde en la última igualdad se ha usado el lema de Watson. Esto alcanza el mínimo en $\frac{D}{N} = \frac{A}{N} = S_N$

En principio A no tiene por qué ser definida positiva, pero tenemos el teorema de Dijkstra que dice lo siguiente:

Teorema 4.1 (Teorema de Dijkstra). Sea $\mathbf{X} \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$. $\{X_\alpha : \alpha = 1, \dots, N\}$ y sea $A = \sum_{k=1}^N (\mathbf{X}_\alpha - \bar{\mathbf{X}})(\mathbf{X}_\alpha - \text{vect} \bar{\mathbf{X}})^T$. Entonces, A es definida positiva con probabilidad 1 si y solo si $N > p$.

Nota. Con este teorema, queremos decir que basta tomar un número de muestras mayor al número de variables a estudiar

4.3. Estimación de los coeficientes de correlación

A continuación, veremos el enunciado del teorema de Zehna que será aplicado en numerosas situaciones:

Teorema 4.2 (Zehna). Sea $P = \{P_\theta : \theta \in \Theta\}$ una función de medidas de probabilidad sobre $(\mathcal{X}, \mathcal{B})$. Consideramos $g : \Theta \rightarrow \Omega$ una función de Θ sobre un intervalo Ω de un espacio euclídeo r -dimensional. Entonces, si \hat{g} es un estimador máximo verosímil de θ , $g(\hat{g})$ lo es de $g(\theta)$.

Podemos aplicar este resultado para calcular los estimadores de máxima verosimilitud de los diversos coeficientes de correlación. Consideremos $\mathbf{X} = (\mathbf{X}_{(1)}^T | \mathbf{X}_{(2)}^T)^T$ y las dos correspondientes particiones indicadas en μ y Σ :

$$\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \quad \Sigma = (\sigma_{ij}) = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}$$

particiones que vamos a considerar también en el vector de medias $\bar{\mathbf{X}}$ y en las matrices de dispersiones muestral A y de covarianzas muestral S , con $A = NS$, esto es:

$$\bar{\mathbf{X}} = \begin{pmatrix} \bar{\mathbf{X}}_{(1)} \\ \bar{\mathbf{X}}_{(2)} \end{pmatrix} \quad A = (a_{ij}) = \begin{pmatrix} A_{(11)} & A_{(12)} \\ A_{(21)} & A_{(22)} \end{pmatrix} \quad S = (s_{ij}) = \begin{pmatrix} S_{(11)} & S_{(12)} \\ S_{(21)} & S_{(22)} \end{pmatrix}$$

Sean $e_i = (0, \dots, \underset{(i)}{1}, \dots, 0)^T$ y $e_j = (0, \dots, \underset{(j)}{1}, \dots, 0)^T$. Entonces, $\mu_i = e_i^T \mu$ y $\Sigma_{ij} = e_i^T \Sigma e_j$. Por tanto, aplicando el th. de Zehna se verificará que:

- $\hat{\mu}_i = e_i^T \hat{\mu} = e_i^T \bar{\mathbf{X}} = \bar{X}_i$ para $i = 1, \dots, p$.
- $\hat{\sigma}_{ij} = e_i^T \hat{\Sigma} e_j = e_i^T \frac{A}{N} e_j = \frac{a_{ij}}{N} = s_{ij}$ para $i, j = 1, \dots, p$

Del mismo modo, si llamamos $E_1 = [I_q | O_{q \times (p-q)}]$ y $E_2 = [O_{(p-q) \times q} | I_{p-q}]$, tenemos:

$$\Sigma_{(11)} = E_1 \Sigma E_1^T, \quad \Sigma_{(22)} = E_2 \Sigma E_2^T, \quad \Sigma_{(12)} = E_1 \Sigma E_2^T, \quad \Sigma_{(21)} = E_2 \Sigma E_1^T$$

Y, aplicando el teorema de Zehna tenemos:

$$\begin{cases} \hat{\Sigma}_{(11)} = E_1 \Sigma E_1^T = E_1 \frac{A}{N} E_1^T = \frac{A_{11}}{N} = S_{11} \\ \hat{\Sigma}_{(22)} = E_2 \Sigma E_2^T = E_2 \frac{A}{N} E_2^T = \frac{A_{22}}{N} = S_{22} \\ \hat{\Sigma}_{(12)} = E_1 \Sigma E_2^T = E_1 \frac{A}{N} E_2^T = \frac{A_{12}}{N} = S_{12} \\ \hat{\Sigma}_{(21)} = E_2 \Sigma E_1^T = E_2 \frac{A}{N} E_1^T = \frac{A_{21}}{N} = S_{21} \end{cases}$$

Definición 4.5 (Coeficiente de correlación lineal de Pearson). El coeficiente de correlación lineal de Pearson, p_{ij} , es $p_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$, y su estimador máximo verosímil es:

$$\hat{p}_{ij} = \frac{\hat{\sigma}_{ij} \hat{\sigma}_i \hat{\sigma}_j}{s_i s_j} = r_{ij}$$

Además, si notamos por $B_{(i)}^T$ a la fila i -ésima de la matriz $\Sigma_{(21)}$, entonces $B_{(i)}^T = e_i^T \Sigma_{(21)}$, con lo cual, por el teorema de Zehna, $\hat{B}_{(i)} = e_i^T \hat{\Sigma}_{(21)} = e_i^T \frac{A}{N} = \frac{a_{(i)}^T}{N} = s_i^T$, siendo a_i^T y s_i^T las filas i -ésimas de A y S respectivamente.

4. Inferencia en la Distribución Normal Multivariante

Por tanto, dado el coeficiente de correlación lineal múltiple $R_{q+i|1,\dots,q} = \frac{\sqrt{B_{(i)}^T \Sigma_{(11)}^{-1} B_{(i)}}}{\sigma_{q+i}}$, se verificará:

$$R_{q+i|1,\dots,q}^{\wedge} = \frac{\sqrt{\hat{B}_{(i)}^T \hat{\Sigma}_{(11)}^{-1} \hat{B}_{(i)}}}{\hat{\sigma}_{q+i}} = \frac{\sqrt{\frac{a_i^T}{N} (\frac{A_{11}}{N})^{-1} \frac{a_i}{N}}}{\sqrt{\frac{a_{q+i,q+i}}{N}}} = \frac{\sqrt{a_i^T A_{11}^{-1} a_i}}{a_{q+i,q+i}} = \frac{\sqrt{S_i^T S_{11}^{-1} S_i}}{S_{q+i}}$$

4.4. Estimadores de máxima verosimilitud de μ y Σ en una DNM

Definición 4.6 (Función de verosimilitud). Sea el vector aleatorio $X \sim N_p(\mu, \Sigma)$, con $\Sigma > 0$ y sea $\{X_\alpha : \alpha = 1, \dots, N\}$ una muestra aleatoria simple. Para una realización muestral $\{x_\alpha : \alpha = 1, \dots, N\}$, definimos la *función de verosimilitud* como

$$L(\mu, \Sigma; x_1, \dots, x_N) = f_{(\mu, \Sigma)}(x_1, \dots, x_N) = \prod_{\alpha=1}^N f_{(\mu, \Sigma)}(x_\alpha).$$

Proposición 4.2 (Estadísticos máximo verosímiles del vector de medias y de la matriz de covarianzas). Sea el vector aleatorio $X \sim N_p(\mu, \Sigma)$, con $\Sigma > 0$ y sea $\{X_\alpha : \alpha = 1, \dots, N\}$ una muestra aleatoria simple de dicha población. Entonces, los estadísticos máximo verosímiles del vector de medias μ y de la matriz de covarianzas Σ son, respectivamente,

$$\hat{\mu} = \bar{X} \quad y \quad \hat{\Sigma} = \frac{A}{N} = S_N.$$

Ya estamos en condiciones de demostrar la proposición 4.2.

Demostración (Proposición 4.2). Consideremos primero el logaritmo de la función de verosimilitud, dado por

$$\log L(\mu, \Sigma) = -\frac{pN}{2} \log(2\pi) - \frac{N}{2} \log(\det(\Sigma)) - \frac{1}{2} \text{tr}(\Sigma^{-1}A) - \frac{N}{2} (\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu).$$

Ahora vayamos por partes. Al maximizar dicha función en μ , se maximizará donde se minimice la forma cuadrática

$$(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu).$$

Como $\Sigma > 0$, se tiene que $\Sigma^{-1} > 0$ así que $(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu) \geq 0$, que de hecho es igual a cero si y solo si $\mu = \bar{X}$. Luego el estadístico máximo verosímil de μ es \bar{X} , ya que el valor que maximiza a $L(\mu, \Sigma)$ es $\mu = \bar{X}$.

Por otro lado, para maximizar $\log L(\mu, \Sigma)$ en Σ hemos de maximizar la función

$$-\frac{pN}{2} \log(2\pi) + \frac{1}{2} f(\Sigma), \quad \text{con} \quad f(\Sigma) = -N \log(\det(\Sigma)) - \text{tr}(\Sigma^{-1}A).$$

Tomando $G = \Sigma$ y $D = A$ en el lema de Watson (??), el máximo se alcanza en $A/N = S_N$, que es la matriz de covarianzas muestral. \square

4.5. Distribuciones de la media muestral y de la matriz de dispersiones

A continuación vamos a obtener tanto la distribución exacta del vector media muestral, como de la matriz de dispersiones $A = NS$, bajo la hipótesis de normalidad. El resultado clave para hallar dichas distribuciones es el Teorema de Fisher, que nos proporcionará ambas distribuciones, además de la independencia entre ambas.

Conocer la forma en que se distribuye la matriz de dispersiones es fundamental para demostrar el carácter definido positivo de la misma. Esto, junto con las condiciones bajo las que se verifica estarán recogidas en el Teorema de Dykstra.

Teorema 4.3 (Teorema de Fisher multivariante). *Dada una muestra aleatoria $\{X_\alpha : \alpha = 1, \dots, N\}$ de una población $N_p(\mu, \Sigma)$, el vector de medias \bar{X} se distribuye según una distribución $N_p(\mu, \Sigma/N)$. La matriz de dispersiones A (EMV de Σ) se distribuye como lo haga*

$$\sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha^T,$$

siendo los vectores Z_α independientes e idénticamente distribuidos según $N_p(0, \Sigma)$. Ambas distribuciones, A y \bar{X} , son independientes.

Teorema 4.4 (Teorema de Dykstra). *Sea el vector aleatorio $X \sim N_p(\mu, \Sigma)$, con $\Sigma > 0$, y sea $\{X_\alpha : \alpha = 1, \dots, N\}$ una muestra aleatoria extraída de la población. La matriz de dispersiones muestral*

$$A = \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})^T$$

es definida positiva con probabilidad 1 si y solo si $p < N$.

4.6. Distribuciones asintóticas de A_n y X_n para una población cualquiera

Proposición 4.3. *Sea $\{X_\alpha\}_{\alpha \geq 1}$ una sucesión de vectores p -dimensionales independientes e idénticamente distribuidos con media μ desconocida y matriz de covarianzas Σ . Sea*

$$\bar{X}_N = \frac{1}{N} \sum_{\alpha=1}^N X_\alpha \quad \text{para todo } N \geq 1.$$

Entonces, cuando $N \rightarrow \infty$,

$$N^{1/2}(\bar{X}_N - \mu) = N^{-1/2} \sum_{\alpha=1}^N (\bar{X}_\alpha - \mu) \sim N_p(0, \Sigma).$$

4. Inferencia en la Distribución Normal Multivariante

Demostración. La demostración se basa en el teorema central del límite. Sea

$$Y_N = N^{-1/2} \sum_{\alpha=1}^N (X_\alpha - \mu).$$

Por el teorema de continuidad de las funciones características es suficiente demostrar que la función característica $\Psi_{Y_N}(t)$ converge a $\exp\{-1/2 t^T \Sigma t\}$. Sabemos que para todo $t \in \mathbb{R}$ se tiene que $\Psi_{Y_N}(t) = \Psi_{t^T Y_N}(1)$ y tenemos que ver que cómo es $\Psi_{t^T Y_N}(s)$ para $s \in \mathbb{R}$. Tenemos

$$t^T Y_N = N^{-1/2} \sum_{\alpha=1}^N (t^T X_\alpha - t^T \mu),$$

donde sabemos que $t^T X_\alpha - t^T \mu \sim (0, t^T \Sigma t)$ para todo $\alpha = 1, \dots, N$. Aplicando el teorema central del límite tenemos

$$\frac{\sum_{\alpha=1}^N (t^T X_\alpha - t^T \mu)}{N^{1/2} \sqrt{t^T \Sigma t}} = \frac{N^{-1/2} \sum_{\alpha=1}^N (t^T X_\alpha - t^T \mu)}{\sqrt{t^T \Sigma t}} = \frac{t^T Y_N}{\sqrt{t^T \Sigma t}} \sim N_p(0, 1).$$

Luego $t^T Y_N \sim N_p(0, t^T \Sigma t)$. Por el teorema de continuidad de las funciones características se tiene que para todo $s \in \mathbb{R}$, cuando N tiende a infinito, $\Psi_{t^T Y_N}(s)$ tiende a $\exp\{-1/2 s^2 (t^T \Sigma t)\}$. \square

Proposición 4.4 (Distribución asintótica de la matriz de dispersión). Sea $\{X_\alpha\}_{\alpha \geq 1}$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos, con vector de medias μ y matriz de covarianzas Σ . Además suponemos que los vectores X_α tienen momentos de cuarto orden finitos. Consideremos, para cada N fijo,

$$\bar{X}_N = \frac{1}{N} \sum_{\alpha=1}^N X_\alpha \quad \text{y} \quad A_N = \sum_{\alpha=1}^N (X_\alpha - \bar{X}_N)(X_\alpha - \bar{X}_N)^T.$$

Entonces,

$$N^{-1/2}(A_N - N\Sigma) \sim N_p(0, V),$$

cuando $N \rightarrow \infty$; en el sentido de que

$$N^{-1/2}(\text{Vec}(A_N) - N \text{Vec}(\Sigma)) \sim N_p(0, V),$$

de nuevo cuando $N \rightarrow \infty$ y con

$$V = \text{Cov}(\text{Vec}((X_{\alpha-\mu})(X_{\alpha-\mu})^T)).$$

Demostración. Por la fórmula de los momentos,

$$A_N = \sum_{\alpha=1}^N (X_\alpha - \mu)(X_\alpha - \mu)^T - N(\bar{X}_N - \mu)(\bar{X}_N - \mu)^T.$$

Llamamos

$$Z_\alpha = (X_\alpha - \mu)(X_\alpha - \mu)^T \quad \text{y} \quad B_N = (\bar{X}_N - \mu)(\bar{X}_N - \mu)^T,$$

4. Inferencia en la Distribución Normal Multivariante

de forma que

$$A = \sum_{\alpha=1}^N Z_{\alpha} - N B_N.$$

Vectorizando tenemos

$$\text{Vec}(A) = \sum_{\alpha=1}^N \text{Vec}(Z_{\alpha}) - N \text{Vec}(B_N).$$

Entonces,

$$N^{-1/2} (\text{Vec}(A) - N \text{Vec}(\Sigma)) = N^{-1/2} \left(\sum_{\alpha=1}^N (\text{Vec}(Z_{\alpha}) - \text{Vec}(\Sigma)) \right) - N^{1/2} \text{Vec}(B_N).$$

Vamos a estudiar el comportamiento asintótico de los dos sumandos de la última igualdad. Observamos primero que

$$E[Z_{\alpha}] = E[(X_{\alpha} - \mu)(X_{\alpha} - \mu)^T] = \Sigma,$$

es decir, $E[\text{Vec}(Z_{\alpha})] = \text{Vec}(\Sigma)$. Para el primer sumando, los vectores $\text{Vec}(Z_{\alpha})$, con $\alpha = 1, \dots, N$, constituyen una sucesión de vectores aleatorios independientes e idénticamente distribuidos (de dimensión $p^2 \times 1$) con vector de medias $\text{Vec}(\Sigma)$ y matriz de covarianzas V . Aplicando el resultado anterior vemos que

$$N^{-1/2} \left(\sum_{\alpha=1}^N (\text{Vec}(Z_{\alpha}) - \text{Vec}(\Sigma)) \right)$$

sigue una distribución $N_{p^2}(\mathbf{0}, V)$ cuando $N \rightarrow \infty$. Para el segundo sumando, podemos escribir

$$\begin{aligned} N^{1/2} \text{Vec}(B_N) &= N^{1/2} \text{Vec}((\bar{X}_N - \mu)(\bar{X}_N - \mu)^T) \\ &= \text{Vec} \left(\frac{1}{N^{1/4}} (N^{-1/2}(\bar{X}_N - \mu)) \frac{1}{N^{1/4}} (N^{-1/2}(\bar{X}_N - \mu))^T \right). \end{aligned}$$

Aplicando el resultado sobre el comportamiento asintótico de \bar{X}_N , se tiene que $N^{1/2}(\bar{X}_N - \mu) \sim N_p(\mathbf{0}, \Sigma)$ cuando $N \rightarrow \infty$. Tenemos entonces que tanto

$$\frac{1}{N^{1/4}} (N^{-1/2}(\bar{X}_N - \mu)) \quad \text{como} \quad \frac{1}{N^{1/4}} (N^{-1/2}(\bar{X}_N - \mu))^T$$

convergen en probabilidad a $\mathbf{0}$. Luego

$$\text{Vec} \left(\frac{1}{N^{1/4}} (N^{-1/2}(\bar{X}_N - \mu)) \frac{1}{N^{1/4}} (N^{-1/2}(\bar{X}_N - \mu))^T \right)$$

converge en probabilidad al vector $\mathbf{0}$. □

Proposición 4.5. Sea $\{X_{\alpha}\}_{\alpha \in \mathbb{N}}$ una sucesión de variables aleatorias tal que cuando α tiende a infinito, X_{α} converge a X en distribución. Sea $\{Y_{\alpha}\}_{\alpha \in \mathbb{N}}$ una sucesión

de variables aleatorias tal que

$$Y_\alpha \xrightarrow[\alpha \rightarrow \infty]{d} \mathbf{0} \left(\iff Y_\alpha \xrightarrow[\alpha \rightarrow \infty]{p} \mathbf{0} \right).$$

Entonces, $X_\alpha + Y_\alpha \xrightarrow[\alpha \rightarrow \infty]{d} X$.

4.7. Distribución de Wishart

Wishart, 1928.

Motivación. Según el Teorema de Fisher (multivariante) para una muestra aleatoria simple X_1, \dots, X_n de una distribución normal multivariante $N_p(\mu, \Sigma)$ con $\Sigma > 0$ se tiene que $A = \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})^T$ se distribuye como lo hace $\sum_{\alpha=1}^N Z_\alpha Z_\alpha^T$, con Z_α variables aleatorias independientes e idénticamente distribuidas según una distribución $N_p(0, \Sigma)$.

Definición 4.7. Sean Z_1, \dots, Z_n vectores aleatorios independientes e idénticamente distribuidos según una distribución $N_p(\mathbf{0}, \Sigma)$ con $\Sigma > 0$ y con $n \geq P$. Se define la *distribución de Wishart centrada con n grados de libertad* como la distribución de la matriz aleatoria $\sum_{\alpha=1}^N Z_\alpha Z_\alpha^T$. Se denota por $W_p(n, \Sigma)$.

Como consecuencia, dada una muestra aleatoria simple de una población que sigue una distribución $N_p(\mathbf{0}, \Sigma)$ con $\Sigma > 0$ tenemos que

$$A = NS_N = (N-1)S_{N-1} \sim W_p(n, \Sigma).$$

En el caso $\Sigma > 0$, la distribución de Wishart tiene la función de densidad:

$$f(A) = \frac{|A|^{n-p-1/2} \exp\{-\frac{1}{2} \text{tr}(\Sigma^{-1}A)\}}{2^{np/2} |\Sigma|^{n/2} \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(\frac{n+1-i}{2})}.$$

Para A matriz simétrica definida positiva, tener $n \geq P$ nos garantiza que $A > 0$. Usando la función gamma multivariante,

$$\Gamma_m(z) = \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(z - \frac{i-1}{2}\right)$$

para $z \in \mathbb{C}$ con $\text{Re}(z) \geq 0$,

$$f(A) = \frac{|A|^{n-p-1/2} \exp\{-\frac{1}{2} \text{tr}(\Sigma^{-1}A)\}}{2^{np/2} |\Sigma|^{n/2} \Gamma_p\left(\frac{n}{2}\right)}.$$

Proposición 4.6 (Ley de Wishart). Si $a \sim W_1(n, \sigma^2)$ entonces $\frac{a}{\sigma^2} \sim \chi_n^2$.

Ejercicio 4.1. Demostrar la proposición 4.6. Para ello hay que hacer un cambio de variable en a .

Proposición 4.7 (Función característica de una matriz aleatoria). La función característica de una matriz aleatoria $Y_{r \times s}$ se define como

$$\Psi_Y(\Theta) = E[\exp\{i \operatorname{tr}(\Theta^T Y)\}]$$

para toda Θ en el espacio de matrices de orden $r \times s$.

Demostración. Vemos inmediatamente que

$$\begin{aligned} \Psi_{\operatorname{vec}(Y)}(\operatorname{vec}(\Theta)) &= E[\exp\{i[\operatorname{vec}(\Theta)^T] \operatorname{vec}(Y)\}] \\ &= E\left[\exp\left\{i \sum_{k=1}^s \left(\sum_{l=1}^r \Theta_{lk} Y_{lk}\right)\right\}\right] \\ &= E[\exp\{i \operatorname{tr}(\Theta^T Y)\}]. \end{aligned}$$

□

Veamos a continuación algunas propiedades de la Ley de Wishart.

Proposición 4.8 (Momentos de primer y segundo orden). Sea $A \sim W_p(n, \Sigma)$. Entonces,

- $E[A] = A\Sigma$.
- $\operatorname{Cov}(a_{ij}, a_{kl}) = n(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$.

Proposición 4.9 (Reproductividad respecto de los grados de libertad). Sean A_1, \dots, A_q matrices aleatorias independientes, con $A_j \sim W_p(n_j, \Sigma)$, con $j = 1, \dots, q$. Entonces,

$$A = \sum_{j=1}^q A_j \sim W_p\left(\sum_{j=1}^q n_j, \Sigma\right).$$

Proposición 4.10 (Transformaciones lineales «rectangulares»). Sea A una matriz aleatoria tal que $A \sim W_q(n, \Sigma)$ y sea $M_{k \times p}$ una matriz no aleatoria de rango $k \leq p$. Entonces,

$$MAM^T \sim W_k(n, M\Sigma M^T).$$

Análogamente,

$$MA^{-1}M^T \sim W_k(n - m + k, (M\Sigma^{-1}M^T)^{-1}),$$

donde m es...

Demostración. Por definición $A = \sum_{\alpha=1}^n Z_\alpha Z_\alpha^T$, con $Z_\alpha \sim N_p(\mathbf{0}, \Sigma)$, independientes para cada α . Consideremos las transformaciones lineales

$$Y_\alpha = MZ_\alpha \quad \text{para } \alpha = 1, \dots, n$$

de tal forma que $Y_\alpha \sim N_k(\mathbf{0}, M\Sigma M^T)$. Entonces, por definición,

$$B = \sum_{\alpha=1}^n Y_\alpha Y_\alpha^T = \dots = MAM^T \sim W_k(n, M\Sigma M^T).$$

□

Proposición 4.11 (Distribución de la matriz de cuasi-covarianzas muestral). Sea $S_{n-1} = A/n$ (con $n = N - 1$) la matriz de cuasi-covarianzas muestral que procede de la muestra aleatoria simple con población $N_p(\mu, \Sigma)$. Entonces, $S_{n-1} \sim W_p(n, \Sigma/n)$ y además, $S_N \sim W_p(N - 1, \Sigma/N)$, con $S_N = A/N$.

Demostración. Sabemos que $A \sim W_p(n, \Sigma)$. Tomamos $M = 1/\sqrt{n}Z_p$ y tenemos que

$$S_n = MAM^T \quad \text{y} \quad M\Sigma M^T = \Sigma/n.$$

Entonces, $S_n \sim W_p(n, \Sigma/n)$. □

4.8. Marginalización y condicionamiento

Proposición 4.12. Sea $A \sim W_p(n, \Sigma)$. Entonces,

$$\frac{a_{ii}}{\sigma_i^2} \sim \chi_n^2 \quad \text{para } i = 1, \dots, p.$$

Demostración. Tomando $M = e_i^T = (0, \dots, 0, \underbrace{1}_i, 0, \dots, 0)$, podemos escribir $a_{ii} = MAM^T$ y $\sigma_i^2 = \sigma_{ii} = M\Sigma M^T$. Por lo tanto, $a_{ii} \sim W_1(n, \sigma^2)$, por lo que $a_{ii}/\sigma_i^2 \sim \chi_n^2$. □

Proposición 4.13. Sea $A \sim W_p(n, \Sigma)$, particionado de la forma

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

donde A_{11} tiene dimensión $(q \times q)$ y A_{22} , dimensión $(p - q) \times (p - q)$. Entonces,

1. $A_{11} \sim W_q(n, \Sigma_{11})$, $A_{22} \sim W_{p-q}(n, \Sigma_{22})$.
2. Si además $\Sigma_{12} = \Sigma_{21} = 0$, entoncess A_{11} y A_{21} son independientes.
3. Sean $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$, $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$, $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Entonces,
 - a) $A_{11.2} \sim W_q(n - (p - q), \Sigma_{11.2})$ y es independiente de A_{12} y de A_{22} .
 - b) $A_{22.1} \sim W_{p-q}(n - q, \Sigma_{22.1})$ y es independiente de A_{21} y de A_{11} .
 - c) $A_{12} \mid A_{22} \sim N_{(p-q) \times q}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{11.2} \otimes A_{22})$, donde \otimes es el producto de Kronecker.

Proposición 4.14 (Relacionando con la distribución de Hotelling). Sea $A \sim W_p(n, \Sigma)$ y sea Y un vector aleatorio p -dimensional independiente de A tal que $P[Y = 0] = 0$. Entonces,

$$\frac{Y^T A Y}{Y^T \Sigma Y} \sim \chi_n^2,$$

que además, es independiente de \mathbf{Y} . Análogamente,

$$\frac{\mathbf{Y}^T \mathbf{A}^{-1} \mathbf{Y}}{\mathbf{Y}^T \Sigma^{-1} \mathbf{Y}} \sim \chi_{n-p+1}^2,$$

que también es independiente de \mathbf{Y} .

Demostración. Tomamos una realización muestral \mathbf{y}^T y aplicamos la propiedad de las transformaciones lineales con $\mathbf{M} = \mathbf{y}^T$. Entonces, $\mathbf{y}^T \mathbf{A} \mathbf{y} \sim W_1(n, \mathbf{y}^T \Sigma \mathbf{y})$. Como χ_n^2 no depende de \mathbf{y} , también

$$\frac{\mathbf{Y}^T \mathbf{A} \mathbf{Y}}{\mathbf{Y}^T \Sigma \mathbf{Y}} \sim \chi_n^2.$$

□

Corolario 4.1. Sea $\{X_\alpha, \alpha = 1, \dots, N\}$ una muestra aleatoria simple de una población $N_p(\boldsymbol{\mu}, \Sigma)$. Entonces tenemos que, con $n = N - 1$,

$$n \frac{\bar{\mathbf{X}}^T S_n \bar{\mathbf{X}}}{\bar{\mathbf{X}}^T \Sigma \bar{\mathbf{X}}} \sim \chi_n^2,$$

siendo independiente de $\bar{\mathbf{X}}$.

4.9. Estadístico T^2 de Hotelling

Contraste de hipótesis sobre $\boldsymbol{\mu}$ cuando Σ es conocida.

Definición 4.8. Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, con $\Sigma > 0$. y sea \mathbf{B} una matriz aleatoria $(p \times q)$ -dimensional con $\mathbf{B} \sim W_p(n, \Sigma)$. Suponemos además que \mathbf{X} y \mathbf{B} son independientes. Definimos entonces

$$T^2 = n \mathbf{X}^T \mathbf{B} \mathbf{X}.$$

Proposición 4.15. Bajo las condiciones de la definición 4.8, se tiene que

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

con condición de no centralidad $\delta = \boldsymbol{\mu}^T \Sigma \boldsymbol{\mu}$.

Veamos el estadístico T^2 de Hotelling asociado al muestreo de una población con distribución normal multivariante.

Definición 4.9. Sea $\{X_\alpha : \alpha = 1, \dots, N\}$ una muestra aleatoria simple de una población con distribución $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, con $\Sigma > 0$. Sean $\bar{\mathbf{X}}, S_N, S_n$, donde $n = N - 1$. Se define el estadístico T^2 de Hotelling basado en la muestra anterior como

$$T^2 = n \bar{\mathbf{X}}^T S_n^{-1} \bar{\mathbf{X}} = N \bar{\mathbf{X}}^T S_N^{-1} \bar{\mathbf{X}} = n N \bar{\mathbf{X}}^T \mathbf{A}^{-1} \bar{\mathbf{X}}.$$

Proposición 4.16. Bajo las condiciones de la definición 4.9, se tiene que

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

con condición de no centralidad $\delta = \mu^T \Sigma \mu$.

Demostración. Análoga a la demostración anterior, teniendo en cuenta que $\bar{X} \sim N_p(\mu, \Sigma/N)$ y $S_N \sim W_p(n, \Sigma/N)$ son independientes. \square

4.9.1. Aplicaciones al problema de contraste sobre el vector de medias

Vamos a ver dos casos para este problema:

- Si Σ es conocida, entonces resolveremos con la distribución de χ^2
- Si Σ es desconocida, resolveremos con la distribución T^2

El planteamiento del problema es el siguiente:

Sea $X \sim N_p(\mu, \Sigma)$, con $\Sigma > 0$, y sea $\{X_1, \dots, X_N\}$ con $N > p$, una m.a.s. Denotamos X a esta m.a.s. El problema de contraste

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

con $\mu_0 \in \mathbb{R}^p$ dado.

Teorema 4.5. Si Σ es conocida, entonces $U : \mathbb{R}^{N-p} \rightarrow \mathbb{R}_0^+$ dada por $U(X) = N(\bar{X} - \mu_0)^T \Sigma^{-1}(\bar{X} - \mu_0) \sim \chi_p^2(\sigma)$, con $\sigma = N(\mu - \mu_0)^T \Sigma^{-1}(\mu - \mu_0)$, y la función test para resolver el problema de contraste es:

$$\phi(X) = \begin{cases} 1 & u(x) > \chi_{p,\alpha}^2 \\ 0 & u(x) \leq \chi_{p,\alpha}^2 \end{cases}$$

para todo $x \in \mathbb{R}^{Np}$.

Demostración. Como $\bar{X} \sim N_p(\mu, \Sigma/N)$, tenemos que $\sqrt{N}(\bar{X} - \mu_0) \sim N_p(\sqrt{N}(\mu - \mu_0), \Sigma)$. Bajo la hipótesis nula, H_0 , se tiene $\sigma = 0$ y $U \sim \chi_p^2(\sigma)$, es decir, la función test se formula como:

$$\phi(X) = \begin{cases} 1 & u > c \\ 0 & u \leq c \end{cases}$$

donde c se calcula para que el test tenga tamaño α . Así,

$$\alpha = \sup_{H_0} E[\phi(X)] = P_{H_0}[u > c] = P[\chi_p^2 > c]$$

por lo que $c = \chi_{p,\alpha}^2$. \square

4. Inferencia en la Distribución Normal Multivariante

Nota. El test construido coincide con el que se habría obtenido mediante el método de la razón de verosimilitudes.

Vamos a ver ahora el caso en el que Σ es desconocida:

Teorema 4.6. Si Σ es desconocida, el estadístico de contraste es $\mathbb{T}^2 = N(\bar{X} - \mu_0)^T S_{N-1}^{-1}(\bar{X} - \mu_0)$, verificándose

$$\frac{\mathbb{T}^2}{n} \frac{n-p+1}{p} \sim F_{p; n-p+1}(\sigma), \quad \sigma = N(\mu - \mu_0)^T \Sigma^{-1}(\mu - \mu_0)$$

y la función de test para resolver el problema de contraste es:

$$\phi(X) = \begin{cases} 1 & t > npF_{p; n-p+1, \alpha}/(n-p+1) \\ 0 & t \leq npF_{p; n-p+1, \alpha}/(n-p+1) \end{cases}$$

siendo $F_{p; n-p+1, \alpha}$ el valor de una distribución F de Snedecor con p grados de libertad que deja a su derecha una probabilidad α , X una realización de la muestra aleatoria y t el valor que toma el estadístico \mathbb{T}^2 .

Demostración. Como $\bar{X} \sim N_p(\mu, \Sigma/N)$, tenemos que $\sqrt{N}(\bar{X} - \mu_0) \sim N_p(\sqrt{N}(\mu - \mu_0), \Sigma)$. Por otro lado, $A \sim W_p(n, \Sigma)$ y es independiente de \bar{X} , por lo que podemos definir $\mathbb{T}^2 = N(\bar{X} - \mu_0)^T S_{N-1}^{-1}(\bar{X} - \mu_0)$, verificándose $\frac{\mathbb{T}^2}{n} \frac{n-p+1}{p} \sim F_{p; n-p+1}(\sigma)$.

En consecuencia, bajo la hipótesis nula, el estadístico se distribuye como una F centrada con p y $n-p+1$ grados de libertad y constituye el estadístico de referencia para realizar el contraste. Por lo tanto, la función test es:

$$\phi(X) = \begin{cases} 1 & t(n-p+1)/np > c \\ 0 & t(n-p+1)/np \leq c \end{cases}$$

Finalmente, $\alpha = \sup_{H_0} E[\phi(X)] = P_{H_0}(\frac{\mathbb{T}^2}{n} \frac{n-p+1}{p} > c) = P(F_{p; n-p+1} > c)$ por lo que $c = F_{p; n-p+1, \alpha}$. \square

Parte II.

Ejercicios