Universidad de Granada

Estadística Multivariante

Pedro Ramos Suárez

Doble Grado de Ingeniería Informática y Matemáticas

23 de noviembre de 2022

Índice

1. Distribución Normal Multivariante

1.1. Motivación sobre la DNM

1.1.1. General

- La DNM es una de las distribuciones más conocidas en la estadística, jugando un papel predominante en muchas áreas de aplicación.
- En el Análisis Multivariante, por ejemplo, la mayor parte de los procedimientos de inferencia para el análisis de datos vector-valuados se han desarrollado bajo la suposición de normalidad.

1.1.2. Aspectos específicos

- La DNM constituye la extensión natural, al contexto multivariante, de la distribución normal univariante. En particular, las distribuciones marginales univariantes de la DNM son normales.

  Más aún, las distribuciones marginales de cualquier dimensión de la DNM son también DNMs.
- Muchos procedimientos basados en la suposición de normalidad (multivariante, en general), particularmente procedimientos basados en los momentos de primer y segundo orden, cumplen ciertas propiedades de optimalidad.

  En determinados casos, los procedimientos presentan cierta robustez frente a desviaciones de la normalidad en la distribución teórica subyacente o en los propios datos. Una justificación viene dada, por ejemplo, por el TCL acerca de la DNM asintótica del vector de medias muestrales, de modo que, para grandes muestras, la DNM puede usarse como aproximación.

  En otros casos, la DNM sirve como modelo de referencia para evaluar posibles desviaciones de la normalidad.
- Técnicamente, la DNM ha sido estudiada en profundidad y ofrece ciertas ventajas en su tratamiento analítico, también desde el punto de vista estadístico-probabilístico:
  - La DNM viene determinada de forma única por los momentos de primer y segundo orden (μ, Σ).
  - Si dos subvectores de un vector aleatorio con DNM tienen correlaciones curzadas nulas, entonces dichos subvectores son mutuamente independientes.
  - La familia de DNMs es cerrada bajo transformaciones lineales (...).
  - La familia de DNMs es cerrada bajo combinaciones lineales de vectores independientes (...).
  - Las distribuciones condicionadas (internamente) en una DNM sin también DNMs. Sus momentos de primer y segundo orden se obtienen de forma sencilla a partir de los correspondientes a la DNM de referencia.
  - ...

1.1.3. Algunas notas históricas

- La DN bivariante comenzó a estudiarse a mediados del s. XIX. Tuvo un gran impacto a partir de una publicación de Francias Galton (1988), sobre análisis de correlación en genética.
- Karl Pearson (1892), a partir del trabajo de Galton, estableció su formulación matemática de forma rigurosa.
- Francis Edgeworth (1892) introdujo, en el estudio de la regresión múltiple y el análisis de correlación parcial y múltiple, elementos fundamentales para el desarrollo de la DNM.
- Durante el s. XX se sucedieron importantes avances en relación con la fundamentación teórica de la DNM, la inferencia estadística con referencia a la DNM (incluyendo el desarrollo y estudio de

otras distribuciones relacionadas, e.g. T^2 de Hotelling, Wishart), así como su contextualización en familias de distribuciones más generales (e.g. distribuciones simétricas de contornos elípticos).

1.2. Fundamentación probablilística (recordatorio)

1.2.1. Elementos

- $(\Omega, \mathcal{F}, P)$ espacio de probabilidad.
  - Ω: Espacio muestral.
  - $\mathcal{F}$: σ-álgebra.
  - P: Medida de probabilidad.
- $X = (X_1, \ldots, X_p)'$, vector aleatorio; i.e., aplicación medible

$$\begin{aligned} X: \quad (\Omega, \mathcal{F}) &\to \quad (\mathbb{R}^p, \mathcal{B}^p) \\ w &\mapsto \quad X(w) = (X_1(w), \ldots, X_p(w))', \end{aligned}$$

$$\forall B \in \mathcal{B}^p, X^{-1}(B) := \{w \in \Omega : X(w) \in B\} \in \mathcal{F}$$

- P_X, distribución de probabilidad inducida por X sobre $(\mathbb{R}^p, \mathcal{B}^p)$:

$$P_X[B] := P[X^{-1}(B)], \text{ para cada } B \in \mathcal{B}^p$$

- Nota: $X = (X_1, \ldots, X_p)' = (X_1, \ldots, X_p)^t = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$.

1.2.2. Función de distribución

- Se define la función de distribución asociada a P_X como

$$F_X(x) = P_X[X_1 \leq x_1, \ldots, X_p \leq x_p], \forall (x_1, \ldots, x_p) \in \mathbb{R}^p$$

- Si existe f_X integrable (en el sentido de Lebesgue) tal que

$$F_X(x) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f_X(u_1, \ldots, u_p) du_1 \ldots du_p, \forall x \in \mathbb{R}^p$$

  f_X se denomina función de distribución, estando definida salvo conjuntos de medida de Lebesgue nula. (Se dice entonces que la distribución de X es "continua").

  Si f_X es una función continua, se puede escribir:

$$f_X(x) = \frac{\delta^n}{\delta x_1 \ldots \delta x_p} F_X(x), \forall x \in \mathbb{R}^p$$

  (en otro caso, se podrá escribir la expresión únicamente para los puntos de continuidad de f_X) y f_X estará definida de forma unívoca bajo esa condición.

1.2.3. Independencia

- En general, para un conjunto producto de la forma $B = B_1 \times \cdots \times B_p$, con $B_j \in \mathcal{B}(j = 1, \ldots, p)$, se tendrá que

$$P_X[B] \neq P_{X_1}[B_1] \cdot ... \cdot P_{X_p}[B_p]$$

- Cuando se da la igualdad para todo conjunto producto $B = B_1 \times \cdots \times B_p \in \mathcal{B}$ se dice que las variables $X_1, \ldots, X_p$, componentes del vector aleatorio X, son independientes entre sí (o mutuamente independientes).
- Equivalentemente, esto ocurre cuando

$$F_X(x) = F_{X_1}(x_1) \cdot ... \cdot F_{X_p}(x_p), \forall x = (x_1, \ldots, x_p)' \in \mathbb{R}^p$$

  y en el caso continuo, si y sólo si

$$f_X(x) = f_{X_1}(x_1) \cdot ... \cdot f_{X_p}(x_p), \forall x = (x_1, \ldots, x_p)' \in \mathbb{R}^p$$

- De forma más general, se habla de independencia entre cualquier número finito de vectores aleatorios (por ejemplo, entre subvectores de un vector aleatorio dado), bajo condiciones de factorización similares en términos de los factores correspondientes a éstos.
- $X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix} \qquad X_{(1)} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \qquad X_{(2)} = \begin{pmatrix} X_{k+1} \\ \vdots \\ X_p \end{pmatrix}.$

  Si se cumple que

$$f_X(x) = f_{X_{(1)}}(x_{(1)}) \cdot f_{X_{(2)}}(x_{(2)}), \forall x \in \mathbb{R}^p$$

$$x = \begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix} \iff X_{(1)} \text{ y } X_{(2)} \text{ son independientes}$$

- Si X, Y son independientes, entonces $g(X), h(Y)$ son independientes para aplicaciones medibles g y h.

$$g : (\mathbb{R}^p, \mathcal{B}^p) \to (\mathbb{R}^{p'}, \mathcal{B}^{p'}) \qquad h : (\mathbb{R}^q, \mathcal{B}^q) \to (\mathbb{R}^{q'}, \mathcal{B}^{q'})$$

$$(\Omega, F) \xrightarrow{X} (\mathbb{R}^p, \mathcal{B}^p) \xrightarrow{g} (\mathbb{R}^{p'}, \mathcal{B}^{p'}) \qquad (\Omega, F) \xrightarrow{Y} (\mathbb{R}^q, \mathcal{B}^q) \xrightarrow{h} (\mathbb{R}^{q'}, \mathcal{B}^{q'})$$

$$\underset{g(X)}{\longrightarrow} \qquad \underset{h(Y)}{\longrightarrow}$$

1.2.4. Función característica

- Dado un vector aleatorio $X = (X_1, \ldots, X_p)'$, se define su función característica como

$$\phi_X(t) = E[e^{it'X}] = \int_\Omega e^{it'X(w)} P(dw) = \int_{\mathbb{R}^p} e^{it'x} P_X(dx), \forall t = (t_1, \ldots, t_p)' \in \mathbb{R}^p$$

- Recordatorio: $e^{ia} = \cos a + i \sin a$. $\qquad \mathbb{C} \equiv \mathbb{R}^2$.
- Teorema (existencia y unicidad): La función característica de un vector aleatorio siempre existe, y determina de forma única su distribución.
- La f.c. tiene gran utilidad como herramienta en la resolución de aspectos analíticos sobre distribuciones de vectores aleatorios.

1.2.5. Función característica e independencia

Dos resultados importantes:

- Resultado 1: Las componentes del vector aleatorio $X = (X_1, \ldots, X_p)'$ son independientes si y sólo si su función característica se factoriza de la forma

$$\phi_X(t) = E[e^{it'X}] = \prod_{k=1}^{p} E[e^{it_k X_k}] = \phi_{X_1}(t_1) \cdot ... \cdot \phi_{X_p}(t_p)$$

$$\forall t = (t_1, \ldots, t_p)' \in \mathbb{R}^p$$

  (i.e., como el producto de las funciones características de las distribuciones univariantes marginales).

- Resultado 2: Si las componentes del vector aleatorio $X = (X_1, \dots, X_p)'$ son independientes, entonces la función característica de la variable aleatoria

$$Y = \sum_{k=1}^{p} X_k$$

se factoriza como

$$\phi_Y(t) = E[e^{itY}] = \prod_{k=1}^{p} E[e^{itX_k}] = \phi_{X_1}(t) \cdot ... \cdot \phi_{X_p}(t), \forall t \in \mathbb{R}$$

1.2.6. Función característica y transformaciones lineales

- Sea $X = (X_1, \dots, X_p)'$ un vector p-dimensional, y sea $Y = (Y_1, \dots, Y_q)'$ un vector q-dimensional definido como

$$Y = BX + b$$

con B matriz $q \times p$ (cte) y b vector $q \times 1$ (cte).

Entonces, la f.c. de Y se obtiene, para cada $t = (t_1, \dots, t_q)' \in \mathbb{R}^q$, como

$$\phi_Y(t) = e^{it'b}\phi_X(B't) \qquad (1)$$

En particular, si el vector aleatorio X se particiona en dos subvectores como $X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}$, con $X_{(1)}$ de dimensión $k \times 1$ y X_2 de dimensión $(p-k) \times 1$, se tiene que

$$\phi_{X_{(1)}}(t_{(1)}) = \phi_X \begin{pmatrix} t_{(1)} \\ 0_{(2)} \end{pmatrix} \qquad (2)$$

con $t_{(1)} = (t_1, \dots, t_k)' \in \mathbb{R}^k$ y $0_{(2)} = (0, \dots, 0)' \in \mathbb{R}^{p-k}$.

- (1)

$$\phi_Y(t)E[e^{it'Y}] = E[e^{it'(BX+b)}] =$$
$$E[e^{it'b} + e^{it'Bx}] = e^{it'b}E[e^{i'Bx}] = e^{it'b}\phi_X(Bt)$$

- (2)

$$B = (I_k, 0_{p-k}) \qquad\qquad b = 0_{k\times 1}$$
$$X_{(1)} = (I_k, 0_{p-k})X + 0_{k\times 1}$$
$$X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}$$

Aplicamos (1).

$$\phi_{X_{(1)}} = e^{it'_{(1)}b}\phi_X(\begin{pmatrix} I_k \\ 0_{p-k} \end{pmatrix} t_{(1)}) \overset{t'_{(1)}b=0}{=} \phi_X \begin{pmatrix} t_{(1)} \\ 0_{p-k} \end{pmatrix} = \phi_X \begin{pmatrix} t_{(1)} \\ 0_{(2)} \end{pmatrix}$$

1.2.7. Función característica y función de densidad (caso continuo)

- La función de densidad de un vector aleatorio $X = (X_1, \dots, X_p)'$ con distribución continua se puede obtener como la transformada de Fourier de la función característica:

$$f_X(x) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} e^{-it'x}\phi_X(t)dt$$

Recordemos que, en este caso, la función característica también se escribe como

$$\phi_X(t) = \int_{\mathbb{R}^p} e^{it'x} f_X(x)dx$$

(Es decir, $f_X(\cdot)$ y $\phi_X(\cdot)$ constituyen un "par de transformadas de Fourier")

1.2.8. Observaciones finales

- Desde el punto de vista puramente estadístico, especialmente en el contexto de la Inferencia Estadística, interesa acerca de un vector aleatorio, X, el conocimiento sobre su distribución, P_X, puesto que los sucesos de interés vendrán expresados en términos de condiciones relativas a los valores que toma X.
- Son importantes, en particular, las familias de distribuciones multivariantes (modelos) que, de forma flexible, puedan dar una representación exacta o aproximada de la variabilidad o el comportamiento de poblaciones o fenómenos objeto de investigaciones aplicadas.
- En este contexto, la distribución normal multivariante (DNM) constituye un modelo central en el desarrollo y aplicación de la Estadística Multivariante.

1.3. Aspectos generales sobre vectores aleatorios

1.3.1. Esperanza y covarianza ("operadores")

- Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio.
- Se define el vector de medias de X como

$$\mu_X := E[X] := \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_p] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

  (siempre que existan las esperanzas unidimensionales)

- Propiedad de linealidad: Sea $Y = (Y_1, \dots, Y_1)'$ definido como

$$Y = BX + b$$

  con B matriz $q \times p$ (cte) y b vector $q \times 1$ (cte).

  Entonces,

$$\mu_Y = BE[X] + b = B\mu_X + b \qquad (1)$$

  (El resultado se extiende convenientemente al caso de transformaciones lineales de matrices aleatorias (2)).

- (1)

$$B = (b_{ij})_{\substack{1 \le i \le q \\ 1 \le j \le p}} \qquad b = (b_i)_{1 \le i \le p}$$

$$E(\sum_{j=1}^{p} b_{ij} X_i + b_i) = \sum_{j=1}^{p} b_{ij} E(X_i) + b_i$$

$$E[BX + b] = BE[X] + b$$

- (2) X matriz aleatoria $p \times q$. Sean B una matriz $m \times p$ (cte.), C una matriz $q \times n$ (cte.) y D una matriz $m \times n$ (cte.). Construimos:

$$W = BXC + D$$

  W es una matriz aleatoria $m \times n$.

  Definimos

$$E[X] = (E[X_{ij}])_{\substack{i=1,\dots,p \\ j=1,\dots,q}}$$

$$E[W] = (E[W_{kl}])_{\substack{k=1,\dots,m \\ l=1,\dots,n}} = BE[X]C + D$$

- Se define la matriz de covarianzas de X como

$$\Sigma_X = Cov(X) := E[(X-\mu_X)(X-\mu_X)'] = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix}$$

con

$$\sigma_{ij} = E[(X_i-\mu_i)(X_j-\mu_j)] \qquad (=\mu_{ji})$$

(siempre que existan las esperanzas unidimensionales).

En particular,

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

- Propiedades:
  1. Σ_X es simétrica.
  2. Los elementos de la diagonal de Σ_X son no negativos.
  3. La clase de matrices de covarianzas (dim. $p \times p$) coincide con la clase de matrices simétricas definidas no negativas (dim. $p \times p$) (3).

- (3)

$\Rightarrow$) $\forall \alpha \in \mathbb{R}^p$.

$$0 \le Var(\alpha' X) = E[(\alpha' X - \underbrace{E[\alpha' X]}_{=\alpha'\mu})^2] = E[(\alpha'(X-\mu))^2] = E[\alpha'(X-\mu)(X-\mu)'\alpha] =$$

$$= \alpha' E[(X-\mu)(X-\mu)']\alpha = \alpha'\Sigma_X\alpha$$

$\Leftarrow$) $r = rango(\Sigma_{p\times p}), r \le p, \Sigma = C_{p\times r}C', rango(C) = r$.
Definimos Y un v.a. $r \times 1$ con $\mu_y = 0$ y $Cov(Y) = I_{r\times r}$ tal que $X = CY$.

$$E[X] = C - E[Y] = 0$$

$$Cov(X) = E[XX'] = E[(CY)(CY)'] = E[CYY'C'] = CE[YY']C' = CC' = \Sigma$$

Hemos encontrado un v.a. que tiene a Σ como matriz de covarianzas.

- En relación con la propiedad 3, para cualquier matriz de covarianzas, Σ, se distinguen los casos siguientes:

  (A) Σ definida positiva (notación: $\Sigma > 0$).

  En este caso, Σ es no singular, con $|\Sigma| > 0$, y $\exists\Sigma^{-1}$.

  "Normalización" y distancia de Mahalanobis: Dado un vector aleatorio p-dimensional $X \sim (\mu, \Sigma)$, con $\Sigma > 0$, para cualquier elección de una matriz C de dimensión $p \times p$ tal que $\Sigma = CC'$ se obtiene una "normalización" (en origen y escala multidimensionales) del vector mediante la transformación

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

  En efecto, se tiene que $Z \sim (0_p, I_{p\times p})$. (4)

  - (4) Veamos que $E[Z] = 0_p, Cov(Z) = \Sigma_Z = I_{p\times p}$

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

$$Cov(Z) = \Sigma_Z = E[ZZ'] = E[C^{-1}(X-\mu)(X-\mu)'(C^{-1})'] =$$
$$= C^{-1}E[(X-\mu)(X-\mu)'](C^{-1})' = C^{-1}CC'(C-1)' = I_{p\times p}$$

- Nota: A matriz ortogonal.

$$AA' = A'A = I$$

$$\Sigma = CC' = CIC' = CAA'C' = CA(CA)'$$

Es decir, la transformación $\Sigma = CC'$ no es única.

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

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

Interpretación y observaciones:

- Se comprueba fácilmente que $\Delta(X, \mu) = ||Z||_p$, para cualquier normalización Z de X ($|| \cdot ||_p$ denota la norma euclídea en el espacio $\mathbb{R}^p$) (5).

- $\Delta(X, \mu)$ es una variable aleatoria, cumpliéndose que

$$E[\Delta^2(X, \mu)] = p \qquad (6)$$

- La ecuación

$$\Delta(x, \mu) = k$$

para $k \geq 0$ (cte.) y $x \in \mathbb{R}^p$, define un hiperelipsoide en $\mathbb{R}^p$, de tal modo que los puntos transformados por normalización se corresponden con la esfera euclídea p-dimensional de radio k con centro en el origen 0 (7).

Nota: Esfera para Z, elipsoide para X.

Los ejes del elipsoide vienen determinados por los autovectores. La longitud de los semiejes son los inversos de las raíces cuadradas de los autovalores de la matriz $(k^2\Sigma)^{-1}$, es decir, $k\sqrt{\lambda_j}$.

- (5)

$$\Delta(X, \mu) = ((X - \mu)'\Sigma^{-1}(X - \mu))^{\frac{1}{2}} = ((X - \mu)'(CC')^{-1}(X - \mu))^{\frac{1}{2}} =$$
$$= ((X - \mu)(C')^{-1}C^{-1}(X - \mu))^{\frac{1}{2}} = ((X - \mu)'(C^{-1})'C^{-1}(X - \mu))^{\frac{1}{2}} =$$
$$= ((C^{-1}(X - \mu))'C^{-1}(X - \mu))^{\frac{1}{2}} = (Z'Z)^{\frac{1}{2}} = ||Z||_2$$

- (6) $E[\Delta^2(X, \mu)] = p$.

$$Z \sim (0_p, I_{p\times p}) \qquad Z = C^{-1}(X - \mu) \qquad \Sigma = CC'$$

$$E[\Delta^2(X, \mu)] = E[(X - \mu)'(CC')^{-1}(X - \mu)] = E[Z'Z] =$$
$$= \sum_{i=1}^{p} E[Z_i^2] - \underbrace{E[Z_i]E[Z_i]}_{=0} = \sum_{i=1}^{p} Var(Z_i) = \sum_{i=1}^{p} 1 = p$$

- (7) $\Delta(x, \mu) = k, k \geq 0, x \in \mathbb{R}^p$.

$$\Delta(x, \mu) = ||z|| \qquad z = C^{-1}(x - \mu) \qquad \Sigma = CC'$$

$$||z|| = k$$

$$\Delta(x, \mu) = k \Rightarrow \{(x - \mu)'\Sigma^{-1}(x - \mu)\}^{\frac{1}{2}} = k \Rightarrow (x - \mu)'\Sigma^{-1}(x - \mu) = k^2 \Rightarrow$$
$$\Rightarrow (x - \mu)'(k^2\Sigma)^{-1}(x - \mu) = 1$$

Como Σ definida positiva, $k^2\Sigma$ es definida positiva y $(k^2\Sigma)^{-1}$ también lo es, por lo que la última ecuación representa un elipsoide.

Si los autovalores de Σ son λ_j, los de $k^2\Sigma$ son $k^2\lambda_j$ y los de $(k^2\Sigma)^{-1}$ son $(k^2\lambda_j)^{-1}$.

(B) Σ semidefinida positiva (notación $\Sigma \geq 0$ indicará, en general, "definida no negativa").

En este caso, Σ es singular, es decir, $|\Sigma| = 0$ y $\nexists \Sigma^{-1}$.

Por tanto, no se puede realizar una normalización en todo el espacio $\mathbb{R}^p$, ni definir la distancia de Mahalanobis de $X \sim (\mu, \Sigma)$ con respecto a su vector de medias μ a nivel p-dimesional.

Observaciones:

- Se tendrá que, siendo $rango(\Sigma) = r < p$, se puede escribir

$$\Sigma = CC', \text{ con } C \text{ matriz } p \times r \text{ de rango } r$$

- Como consecuencia: Con probabilidad 1, las componentes del vector aleatorio $X = (X_1, \ldots, X_p)'$ cumplirán (al menos) una relación de dependencia lineal del tipo

$$\alpha' X = k, \text{ con } \alpha \neq 0 \qquad (8)$$

(es decir, la variabilidad de X se sitúa (P_X−c.s.) en un hiperplano afín en $\mathbb{R}^p$)

- (8) Σ_X semidefinida positiva $\Rightarrow \exists \alpha$ autovector asociado autovalor 0.

$$Y = \alpha' X$$

$$\mu_Y = E[Y] = E[\alpha' X] = \alpha' E[X] = \alpha' \mu_X = k$$

$$Var[Y] = E[(Y - \mu_y)^2] = E[(\alpha' X - \alpha' \mu_X)^2] = E[(\alpha'[X - \mu_X])^2] =$$

$$= E[\alpha'(X - \mu_X)(X - \mu_X)(\alpha')'] = \alpha' E[(X - \mu_X)'(X - \mu_x)]\alpha = \alpha' \underbrace{\Sigma_X \alpha}_{0_p} = 0 \Rightarrow$$

$$\Rightarrow P[Y = k] = 1$$

- Transformaciones lineales: Sean $X \sim (\mu_X, \Sigma_X)$ vec. a. p-dimensional, $Y = BX + b$ vec. a. q-dimensional, con B matriz $q \times p$ (cte.) y b vector $q \times 1$ (cte.).

  Entonces, se tiene que

$$Y \sim (\mu_Y, \Sigma_Y) = (B\mu_X + b, B\Sigma_X B') \qquad (9)$$

- (9)

$$E[Y] = E[BX + b] = BE[X] + b = B\mu_X + b$$

$$Var[Y] = E[(Y - \mu_y)(Y - \mu_Y)'] = E[(Bx + b - B\mu_X - b)(Bx + b - B\mu_X - b)'] =$$

$$= E[(B(X - \mu_X))(B(X - \mu_X))'] = E[B(X - \mu_X)(X - \mu_X)B'] =$$

$$= BE[(X - \mu_X)(X - \mu_X)']B' = B\Sigma_X B'$$

- Algunas medias globales de variación:
  - $|\Sigma| = \prod_{j=1}^{p} \lambda_j$ (determinante).
  - $tr(\Sigma) = \sum_{j=1}^{p} \sigma_j^2 = \sum_{j=1}^{p} \lambda_j$ (traza).

  ($\lambda_1, \ldots, \lambda_p$ autovalores de Σ)

1.3.2. Función característica

- Recordatorio: Dado un vector aleatorio $X = (X_1, \dots, X_p)'$, se define su función característica como:

$$\phi_X(t) = E[e^{it'X}] = \int_\Omega e^{it'X(w)} P(dw) = \int_{\mathbb{R}^d} e^{it'x} P_X(dx), \forall t = (t_1, \dots, t_p)' \in \mathbb{R}^p$$

- Teorema: Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio. Se tiene que la distribución (multivariante) de X queda unívocamente determinada por el conjunto de todas las distribuciones (univariantes) de variables aleatorias de la forma

$$\alpha' X, \forall \alpha \in \mathbb{R}^p \qquad (10)$$

- (10) Sabemos que $\phi_X(t) = E[e^{it'X}]$. Tomamos $Y_n = \alpha' X = \sum_{j=1}^p \alpha_j X_j \Rightarrow \phi_{Y_\alpha}(t) = E[e^{it(\alpha' X)}]$.

  Tomando $t = 1 \Rightarrow \phi_{Y_\alpha}(1) = E[e^{i\alpha' X}] = \phi_X(\alpha)$.
  Tomando $Y_t = t'X \Rightarrow \phi_{Y_t}(1) = \phi_X(t)$.

- Conocer todas las distribuciones de $X_1, \dots, X_p$ no nos permiten conocer la de X excepto en el caso de independencia, mientras que conocer todas las combinaciones de la forma $\alpha_1 X_1 + \dots + \alpha_p X_p$ si nos lo permite.

1.3.3. Momentos y cumulantes

Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con función característica $\phi_X(t), t \in \mathbb{R}^p$.

- Momentos: Se define el momento (no centrado) p-dimensional de orden $(r_1, \dots, r_p)$ de X como

$$\mu^{1\dots p}_{r_1 \dots r_p} = E[X_1^{r_1} \dots X_p^{r_p}]$$

  Los momentos pueden obtenerse a partir de la función característica, derivándose de su expansión de Taylor (respecto al origen):

$$\phi(t) = E[e^{it'x}] = E[\sum_{r=0}^\infty \frac{1}{r!}(it'X)^r] = \sum_{r=0}^\infty \sum_{r_1+\dots+r_p=r} \mu^{1\dots p}_{r_1 \dots r_p} \frac{(it_1)^{r_1} \dots (it_p)^{r_p}}{r_1! \dots r_p!}$$

En particular, se tiene el siguiente resultado:

- Teorema: 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 continuamente diferenciable, y

$$\frac{\delta^m}{\delta^{m_1}_{t_1} \dots \delta^{m_p}_{t_p}} \phi(t) \mid_{t=0} = i^m E[X_1^{m_1} \dots X_p^{m_p}] = i^m \mu^{1\dots p}_{m_1 \dots m_p}$$

  $(m = m_1 + \dots + m_p)$

Cumulantes: Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con función característica $\phi_X(t)$, $t \in \mathbb{R}^p$. Consideramos la función

$$\log \phi(t)$$

- Se definen los cumulantes p-dimensional de orden $(r_1, \dots, r_p)$ de X como los coeficientes de la correspondiente expansión (respecto al origen),

$$\log \phi(t) = \sum_{r=0}^\infty \sum_{r_1+\dots+r_p=r} \kappa^{1\dots p}_{r_1 \dots r_p} \frac{(it_1)^{r_1} \dots (it_p)^{r_p}}{r_1! \dots r_p!}$$

1.3.4. Cambio de variables

- Teorema: Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con función de densidad $f_X(x), x \in \mathbb{R}^p$, positiva sobre $S \subseteq \mathbb{R}^p$ y continua-

  Sea $Y = (Y_1, \dots, Y_p)'$ un vector aleatorio con

  $$Y = g(x) = (g_1(X), \dots, g_p(X))',$$

  siendo la restricción

  $$g = (g_1, \dots, g_p)' : S \to T \equiv g(S) \subseteq \mathbb{R}^p$$

  una aplicación biyectiva, y sea $g^{-1} =: h = (h_1, \dots, h_p)'$. Supongamos que existen las derivadas parciales

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

  y son continuas sobre T.

  Entonces, la función de densidad $f_Y(y)$, $y \in \mathbb{R}^p$, del vector aleatorio $Y = g(X)$ viene dada por

  $$f_Y(y) = f_X(g^{-1}(y)) \cdot abs(J_{g^{-1}}(y))$$

  siendo $J_{g^{-1}}(y) = [J_g(g^{-1}(y))]^{-1}$ el determinante de la matriz jacobiana de la transformación g^{-1}.

- Caso lineal: Sean X vec.a. p-dimensional, Y vec.a. p-dimensional, definido por la transformación

  $$Y = g(X) := BX + b$$

  con B matriz $p \times p$ (cte.) no singular, b vector $p \times 1$ (cte.).

  Entonces, en este caso, se tiene que

  $$X = B^{-1}(Y - b)$$

  $$J_{g^{-1}}(\cdot) \equiv |B^{-1}| = |B|^{-1}$$

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

1.4. Descomposición Matriciales Diagonalización

A matriz cuadrada $p \times p$, simétrica, definida no negativa

$$A = QDQ' \iff Q'AQ = D$$

- Q matriz ortogonal, $QQ' = Q'Q = I$.
- D matriz diagonal (autovalores).

Como es definida no negativa, podemos escribirlo como:

$$A = QD^{\frac{1}{2}}D^{\frac{1}{2}}Q' = (QD^{\frac{1}{2}})(QD^{\frac{1}{2}})'$$

(ya que $D^{\frac{1}{2}} = (D^{\frac{1}{2}})'$)

Llamando $C = D^{\frac{1}{2}}$, podemos tomar otra matriz H otrogonal tal que $HH' = H'H = I$ y tenemos:

$$A = CC' = CIC' = CHH'C' = (CH)(CH)' = EE'$$

Tomando $H = Q'$ tenemos:

$$A = QD^{\frac{1}{2}}Q'QD^{\frac{1}{2}}Q' = (QD^{\frac{1}{2}}Q')(QD^{\frac{1}{2}}Q') = RR = R^2$$

Esta R se interpreta como la raíz cuadrada de A ($R = A^{\frac{1}{2}}$).

1.4.1. Descomposición de Cholesky

Es de la forma

$$A = TT'$$

donde T es una matriz triangular superior (y por tanto T' es triangular inferior), con todos los elementos de la diagonal no negativos.

Si $rango(A) = r$, entonces T tiene $p - r$ filas nulas.

1.5. Caso $\Sigma > 0$ definiciones, propiedades y caracterizaciones

1.5.1. Definición en términos de la densidad

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

$$f_X(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}, \forall x \in \mathbb{R}^p$$

siendo $\mu = (\mu_1, \ldots, \mu_p)' \in \mathbb{R}^p$ y Σ una matriz escalar $(p \times p)$-dimensional simétrica y definida positiva.

Se denota $X \sim N_p(\mu, \Sigma)$, representando μ y Σ los "parámetros" de la distribución.

Se demuestra que f_X está bien definida, es decir, que cumple las condiciones para ser una función de densidad:

1. $f_X(x) \geq 0, \forall x \in \mathbb{R}^p$.
2. $\int_{\mathbb{R}^p} f_X(x)dx = 1 \qquad (1).$

(1) $\int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\}dX.$
Al ser $\Sigma > 0$, tomamos una descomposición $\Sigma = CC'$, con C matriz cuadrada $p \times p$ no singular. Hacemos el siguiente "cambio de variable":

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

$$X = CZ + \mu$$

Elegimos C tal que $|C| > 0$. Supongamos que $|C| < 0$, tomamos A tal que $\Sigma = (CA)(CA)' = CAA'C'$ con $AA' = A'A = I$ de forma que $|CA| > 0$, y tomamos CA como C.

El determinante del Jacobiano de la transformación:

$$|C^{-1}| = |C|^{-1} \Rightarrow dX = |C|^{-1}dZ \Rightarrow dZ = |C|dX$$

$$\Sigma^{-1} = (CC')^{-1} = (C^{-1})'C^{-1}$$

$$|\Sigma| = |C||C'| = |C|^2 \Rightarrow |\Sigma|^{\frac{1}{2}} = |C|$$

Con todo esto, nos queda:

$$\int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\}dX = \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}|C|} exp\{-\frac{1}{2}Z'Z\}|C|dZ =$$

$$= \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}} exp\{-\frac{1}{2}\sum_{j=1}^{p} z_j^2\}dZ = \prod_{j=1}^{p}\int_{\mathbb{R}} \frac{1}{(2\pi)^{\frac{1}{2}}} exp\{-\frac{1}{2}z_j^2\}dz_j = \prod_{j=1}^{p} 1 = 1$$

- Nota: Esto corresponde a

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})}$$

1.5.2. Vector de medias y matriz de covarianzas

- A partir de cualquier factorización de la forma

$$\Sigma = CC'$$

para alguna matriz C de dimensión $p \times p$ no singular (por ser $\Sigma > 0$), se tiene que el vec.a. $Z = (Z_1, \ldots, Z_p)'$ definido por:

$$Z = C^{-1}(X - \mu) \qquad \text{(es decir, con } X = CZ + \mu\text{)}$$

se distribuye con función de densidad

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

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

$$Z_j \sim N_1(0, 1)$$

siendo $Z_1, \ldots, Z_p$ independientes (por tanto, incorreladas), y

$$Z \sim N_p(0, I_p)$$

- Se comprueba de forma inmediata que μ y Σ representan, respectivamente, el vector de medias y la matriz de covarianzas del vector aleatorio X, es decir:

$$\mu_X = \mu \qquad (2)$$

$$\Sigma_X = \Sigma \qquad (3)$$

- (2) $CC' = \Sigma$.

$$\mu_X = E[X] = E[CZ + \mu] = CE[Z] + \mu = \mu$$

- (3)

$$\Sigma_X = E[(X - \mu)(X - \mu)'] = E[(CZ + \mu - \mu)(CZ + \mu - \mu)'] = CE[ZZ']C' = CC'$$

Donde hemos utilizado que $Z_1, \ldots, Zp$ son independientes:

$$E[Z_1 Z_2] = E[Z_1]E[Z_2] = 0$$

$$E[ZZ'] = \begin{pmatrix} z_1^2 & z_2 z_1 & \cdots \\ z_1 z_2 & z_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} z_1^2 & 0 & \cdots \\ 0 & z_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = I$$

1.5.3. Algunas propiedades (caso $\Sigma > 0$)

1 Cambio de variables lineal: Sean

$$X \sim N_p(\mu_X, \Sigma_X) \qquad (\Sigma > 0)$$

$$Y = BX + b$$

con B matriz $p \times p$ (cte.) no singular y b vector $p \times 1$ (cte.).

Entonces, se tiene que:

$$Y \sim N_p(\mu_Y, \Sigma_Y) = N_p(B\mu_X + b, B\Sigma_X B') \qquad (4)$$

- (4) $X \sim N_p(\mu_X, \Sigma_X), Y = BX + b$

$$f_Y(y) = F_X(B^{-1}(y-b))abs(|B|^{-1}) =$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\{-\frac{1}{2}(B^{-1}(y-b)-\mu)'\Sigma^{-1}(B^{-1}(y-\mu)-b)\}abs(|B|^{-1}) =$$

Utilizamos que $I = B^{-1}B$, y que $abs(|B^{-1}|) = \sqrt{|B|^{-2}}$.

$$= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}|B|^{\frac{1}{2}}|B|^{\frac{1}{2}}} exp\{-\frac{1}{2}(y-B\mu-b)'(B\Sigma B')^{-1}(y-B\mu-b)\}$$

$(|\Sigma|^{\frac{1}{2}}|B|^{\frac{1}{2}}|B|^{\frac{1}{2}} = |B\Sigma B'|^{\frac{1}{2}})$

$$Y \sim N_p(B\mu + b, B\Sigma B')$$

- Caracterización 1: Un vector aleatorio p-dimensional X tiene una DNM no singular, $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$, si y sólo si

$$X \stackrel{d}{=} AZ + \mu$$

con A matriz $p \times p$ (cte.) no singular, $AA' = \Sigma$ y $Z \sim N_p(0, I_p)$ (5).

($\stackrel{d}{=}$ igual en distribución, tienen la misma distribución)

- (5)

$\Leftarrow$ Supongamos $X \stackrel{d}{=} AZ + \mu \Rightarrow X \sim N_p(A \cdot 0 + \mu, AI_pA') = N_p(\mu, AA') = N_p(\mu, \Sigma)$.

$\Rightarrow$ Supongamos $X \sim N_p(\mu, \Sigma), \Sigma > 0$.
Sea $A_{p\times p}$ no singular tal que $\Sigma = AA'$. Definimos $Z = A^{-1}(X-\mu) \Rightarrow X = AZ + \mu$.

$$Z = A^{-1}X + (A^{-1}\mu) \Rightarrow Z \sim N_p(A^{-1}\mu + (-A^{-1}\mu), A^{-1}\Sigma(A^{-1})') \equiv$$

$$\equiv N_p(0, A^{-1}AA'(A^{-1})') \equiv N_p(0, I_p)$$

2 Independencia y condicionamiento: Vemos a continuación algunos resultados fundamentales (dos sobre independencia y uno sobre condicionamiento) que, entre sobre posibilidades, pueden tratarse a partir de la definición de la DNM no singular en términos de su densidad.

- Resultado 1: Sea $X = (X_1, \ldots, X_p)'$ un vector aleatorio con DNM no singular,

$$X \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$

Si la matriz Σ es diagonal,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_p^2 \end{pmatrix} = diag(\sigma_1^2, \ldots, \sigma_p^2)$$

entonces las variables aleatorios componentes del vector, $X_i, i = 1, \ldots, p$, son independientes y tienen un DN univariante (no degenerada),

$$X_i \sim N(\mu_i, \sigma_i^2) \qquad (\sigma_i^2 > 0), \qquad i = 1, \ldots, p \qquad (6)$$

- (6)

$$f_X(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\} =$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}|\prod_i^p \sigma_i^2|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu)' \begin{pmatrix} \sigma_1^{-2} & 0 & \dots & 0 \\ 0 & \sigma_2^{-2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^{-2} \end{pmatrix} (x-\mu)\} =$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}|\prod_i^p \sigma_i^2|^{\frac{1}{2}}} exp\{-\frac{1}{2}(\sigma_1^{-2}(x_1-\mu_1)^2 \cdot \ldots \cdot \sigma_p^{-2}(x_p-\mu_p)^2)\} =$$

$$= \prod_{i=1}^{p} \frac{1}{(2\pi)^{\frac{1}{2}}|\sigma_i^2|^{\frac{1}{2}}} exp\{-\frac{1}{2}\sigma_i^{-2}(x_i-\mu_i)^2\} = \prod_{i=1}^{p} \frac{1}{(2\pi)^{\frac{1}{2}}|\sigma_i^2|^{\frac{1}{2}}} exp\{-\frac{1}{2}(\frac{x_i-\mu_i}{\sigma_i})^2\} =$$

$$= \prod_{i=1}^{p} f_{X_i}(x) \qquad\qquad X_i \sim N(\mu_i, \sigma_i^2)$$

- El recíproco de este resultado es cierto en el siguiente sentido:

  Si $X = (X_1, \dots, X_p)'$ es un vector aleatorio con componentes (mutuamente) independientes, cada una de ellas con DN univariantes (no degenerada),

  $$X_i \sim N(\mu_i, \sigma_i^2) \qquad (\sigma_i^2), \qquad i = 1, \dots, p$$

  entonces X tiene DNM no singular,

  $$X \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$

  (Observación: El resultado no es válido si solo se supone la incorrelación, o incluso la "independencia dos a dos", entre las componentes del vector).

- El Resultado 1 se puede generalizar al caso en que la matriz Σ es una matriz diagonal por cajas, estableciéndose las conclusiones en términos de los correspondientes subvectores del vector X (se enunciará, por simplicidad, en el caso de dos subvectores):

- Resultado 2: Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con DNM no singular,

  $$X \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$

  Supongamos que las componentes de X están ordenadas de tal modo que para la partición del vector

  $$X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}$$

  con $X_{(1)} = (X_1, \dots, X_q)', X_{(2)} = (X_{1+q}, \dots, X_p)'$, se tiene

  $$\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix} = \begin{pmatrix} \Sigma_{(11)} & 0 \\ 0 & \Sigma_{(22)} \end{pmatrix}$$

  (i. e., $\Sigma = diag(\Sigma_{(11)}, \Sigma_{(22)})$). Entonces, los vectores aleatorios $X_{(1)}$ y $X_{(2)}$ son (mutuamente) independientes y tienen cada uno DNM, de dimensiones respectivas q y $p-q$,

  $$\begin{matrix} X_{(1)} \sim N_q(\mu_{(1)}, \Sigma_{(11)}) \\ X_{(2)} \sim N_{p-q}(\mu_{(2)}, \Sigma_{(22)}) \end{matrix} \qquad (7)$$

- (7)

  $$f_X(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu)'\Sigma(x-\mu)\} = (*) \qquad X \sim N_p(\mu, \Sigma)$$

  - $|\Sigma| = |\Sigma_{(11)}||\Sigma_{(22)}| \qquad |\Sigma|^{\frac{1}{2}} = |\Sigma_{(11)}|^{\frac{1}{2}}|\Sigma_{(22)}|^{\frac{1}{2}} \qquad \Sigma^{-1} = \begin{pmatrix} (\Sigma_{(11)})^{-1} & 0 \\ 0 & (\Sigma_{(22)})^{-1} \end{pmatrix}$

- $(x-\mu)'\Sigma(x-\mu) = (X_{(1)}-\mu_{(1)})(X_{(2)}-\mu_{(2)}) \begin{pmatrix} (\Sigma_{(11)})^{-1} & 0 \\ 0 & (\Sigma_{(22)})^{-1} \end{pmatrix} \begin{pmatrix} X_{(1)}-\mu_{(1)} \\ X_{(2)}-\mu_{(2)} \end{pmatrix} =$
  $= (X_{(1)}-\mu_{(1)})'(\Sigma_{(11)})^{-1}(X_{(1)}-\mu_{(1)}) + (X_{(2)}-\mu_{(2)})'(\Sigma_{(22)})^{-1}(X_{(2)}-\mu_{(2)})$

$$(*) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma_{(11)}|^{\frac{1}{2}}|\Sigma_{(22)}|^{\frac{1}{2}}} exp\{-\frac{1}{2}(X_{(1)}-\mu_{(1)})'(\Sigma_{(11)})^{-1}(X_{(1)}-\mu_{(1)})\}exp\{-\frac{1}{2}(X_{(2)}-\mu_{(2)})'(\Sigma_{(22)})^{-1}(X_{(2)}-\mu_{(2)})\} =$$

$$= f_{X_{(1)}}(x_{(1)})f_{X_{(2)}}(x_{(2)}) \qquad (\frac{p}{2} = \frac{q+p-q}{2})$$

- Resultado 3: Sea (como en el Resultado 2)

$$X \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$

con el particionamiento

$$X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}$$

Entonces, se tiene que:

1. Los vectores $X_{(1)}$ y $X_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}X_{(1)}$ son independientes.
2. Dichos vectores se distribuyen según
   - $X_{(1)} \sim N_q(\mu_{(1)}, \Sigma_{(11)})$.
   - $X_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}X_{(1)} \sim N_{p-q}(\mu_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mu_{(1)}, \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)})$.
3. La distribución condicionada de $X_{(2)}$ dado $X_{(1)} = x_{(1)}$ es una DNM, de la forma

$$N_{p-q}(\mu_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(x_{(1)} - \mu_{(1)}), \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}) \qquad (8)$$

(En adelante, se denotará $\Sigma_{22\cdot 1} = \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}$)

- (8) Definimos la matriz

$$C = \begin{pmatrix} I_{(1)} & 0 \\ -\Sigma_{(21)}\Sigma_{(11)}^{-1} & I_{(2)} \end{pmatrix} \qquad |C| = 1 \Rightarrow C \text{ no singular}$$

Consideramos el cambio de variables dado por la transformación:

$$Y = CX = C\begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix} = \begin{pmatrix} X_{(1)} \\ X_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}X_{(1)} \end{pmatrix} = \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix} = Y \qquad X = C^{-1}Y$$

$$Y \sim N_p(C\mu, C\Sigma C') = N_p(C\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, C\begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix} C^{-1}) =$$

$$= N_p(\begin{pmatrix} \mu_{(1)} \\ \mu_2 - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mu_{(1)} \end{pmatrix}, \begin{pmatrix} \Sigma_{(11)} & 0 \\ 0 & \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)} = \Sigma_{22\cdot 1} \end{pmatrix})$$

Por el Resultado 2, los vectores $X_{(1)}$ y $X_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}X_{(1)}$ son independientes.

Nos queda ver el punto 3. Al tratarse de una distribución continua

$$f_{X_{(2)}|X_{(1)}=x_{(1)}}(x_{(2)}) = \frac{f_{X_{(1)}X_{(2)}}(x_{(1)}, x_{(2)})}{f_{X_{(1)}}(x_{(1)})} \overset{\text{c.v.}}{=} \frac{f_{Y_{(1)}Y_{(2)}}(x_{(1)}, x_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}x_{(1)}) \cdot 1}{f_{X_{(1)}}(x_{(1)})} =$$

$$= \frac{\frac{1}{(2\pi)^{\frac{p}{2}}(|\Sigma_{11}||\Sigma_{22\cdot 1}|)^{\frac{1}{2}}} exp\{-\frac{1}{2}[A + B(x_{(2)})]\}}{\frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma_{(11)}|^{\frac{1}{2}}} exp\{-\frac{1}{2}A\}} = \frac{1}{(2\pi)^{\frac{(p-1)}{2}}|\Sigma_{22\cdot 1}|^{\frac{1}{2}}} exp\{-\frac{1}{2}Bx_{(2)}\}$$

Nota: El cambio de variable lo vimos en la sección 1.3.4. Cambio de variables.

$$Cx = \begin{pmatrix} I_{(1)} & 0 \\ -\Sigma_{(21)}\Sigma_{(11)}^{-1} & I_{(2)} \end{pmatrix} \begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix} = \begin{pmatrix} x_{(1)} \\ x_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}x_{(1)} \end{pmatrix}$$

$$\begin{pmatrix} X_{(1)} \\ X_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}x_{(1)} \end{pmatrix} - \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} - \Sigma_{(21)}\Sigma_{(21)}\Sigma_{(11)}^{-1}\mu_{(1)} \end{pmatrix} = \begin{pmatrix} X_{(1)} - \mu_{(1)} \\ X_{(2)} - (\mu_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(x_1 - \mu_{(1)})) \end{pmatrix} \Rightarrow$$

$$\Rightarrow \begin{pmatrix} X_{(1)} - \mu_{(1)} \\ X_{(2)} - (\mu_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(x_1 - \mu_{(1)})) \end{pmatrix}' \begin{pmatrix} \Sigma_{(11)}^{-1} & 0 \\ 0 & \Sigma_{22\cdot 1}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X_{(1)} - \mu_{(1)} \\ X_{(2)} - (\mu_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(x_1 - \mu_{(1)})) \end{pmatrix} =$$

$$= (X_{(1)} - \mu_{(1)})'\Sigma_{(11)}^{-1}(X_{(1)} - \mu_{(1)}) + (x_{(2)} - (\mu_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(X_{(1)} - \mu_{(1)}))'\Sigma_{22\cdot 1}^{-1}(x_{(2)} - (\mu_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(x_{(1)} - \mu_{(1)}))) =$$

$$= A + B(x_{(2)})$$

Por lo que tenemos:

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

1.5.4. Función característica de la DNM (caso $\Sigma > 0$)

La función característica de la DNM, en el caso no singular ($\Sigma > 0$), puede obtenerse de forma directa a partir de la función de densidad.

- Caracterización 2: Un vector aleatorio p-dimensional X con vector de medias μ y matriz de covarianzas $\Sigma > 0$ tiene una DNM no singular, $X \sim N_p(\mu, \Sigma)$, si y sólo si su función característica tiene la forma

  $$\phi_X(t) = exp\{it'\mu - \frac{1}{2}t'\Sigma t\} \qquad t \in \mathbb{R}^p \qquad (9)$$

  (Observación: Se verá más adelante que en el casos singular, con Σ semidefinida positiva, la función característica tiene la misma expresión formal).

- (9)

  $$\phi_X(t) = E[e^{it'X}] \qquad \forall t \in \mathbb{R}^p$$

  $$\phi_X(t) = \int_{\mathbb{R}^p} e^{it'x} f_X(x)dx = \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\{it'x - \frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}dx =$$

  Hacemos un cambio de variable: $\Sigma = CC', Y = C^{-1}(X - \mu) \Rightarrow X = CY + \mu$:

  $$= \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}(|C||C'|)^{\frac{1}{2}}} exp\{it'(CY + \mu) - \frac{1}{2}Y'Y\}|C|dy =$$

  $$= e^{it'\mu} \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{p}{2}}} exp\{i(t'C)Y - \frac{1}{2}Y'Y\}dy =$$

  Llamamos $t'C = \alpha = (\alpha_1, \ldots, \alpha_p)$.

  $$= e^{it'\mu} \int_{\mathbb{R}^p} \frac{1}{(2\pi)^{\frac{1}{2}}} exp\{\sum_{j=1}^{p}[i\alpha_j y_j - \frac{1}{2}y_j^2]\}dy =$$

  $$= \prod_{j=1}^{p} e^{it_j\mu_j}\phi_{Y_j}(\alpha_j) =$$

  donde $\phi_{N(\mu,\sigma^2)}(t) = e^{it\mu - \frac{1}{2}\sigma^2t^2}$, por lo que $\phi_{N(0,1)}(t) = e^{-\frac{1}{2}t^2}$, y como $Y \sim N(0, I_{p\times p}) \Rightarrow Y_j \sim N(0,1)$.

  $$= \prod_{j=1}^{p} e^{it_j\mu_j - \frac{1}{2}\alpha_j^2} =$$

$$= exp\{\sum_{j=1}^{p}(it_j\mu_j - \frac{1}{2}\alpha_j^2)\} = e^{it'\mu - \frac{1}{2}t't\Sigma t}$$

donde hemos usado $\sum_{j=1}^{p} t_j\mu_j = t'\mu$, $\sum_{j=1}^{p} \alpha_j^2 = \alpha'\alpha = t'CC't = t'\Sigma t$.

- Recordatorio: X v.a. (multidimensional) $\sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma^2 > 0$. Función característica:

$$\phi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2} \qquad t \in \mathbb{R}$$

En el caso $Z \sim N(0,1)$, tenemos:

$$\phi_Z(t) = e^{\frac{-1}{2}t^2} \qquad t \in \mathbb{R}$$

- Frente al uso directo de la función de densidad (caso no singular), la función característica es muy útil en la derivación de diversos resultados, algunos de los cuales se ven a continuación.

- Resultado 4: Sean

$$X \sim N_p(\mu_X, \Sigma_X) \qquad (\Sigma_X > 0)$$
$$Y = BX + b$$

con B matriz $q \times p$ (cte.) de rango q ($\Rightarrow q \leq p$) y b vector $q \times 1$ (cte.).

Entonces, se tiene que

$$Y \sim N_q(\mu_Y, \Sigma_Y) = N_q(B\mu_X + b, B\Sigma_X B')$$

(siendo $rango(B\Sigma_X B') = q$, es decir, $B\Sigma_X B' > 0$) (10)

- (10)

$$\phi_Y(t) = E[e^{it'(BX+b)}] = E[e^{it'BX}e^{it'b}] = e^{it'b}E[e^{it'BX}] =$$
$$= e^{it'b}\phi_X(B't) = e^{it'b}exp\{i(B't)'\mu_X - \frac{1}{2}(B't)'\Sigma_X(B't)\} =$$
$$= exp\{it'(B\mu_X + b) - \frac{1}{2}t'B\Sigma_X B't\} = exp\{it'\mu_y - \frac{1}{2}t'\Sigma_Y t\} \Rightarrow$$
$$\Rightarrow Y \sim N_q(\mu_Y, \Sigma_Y) = N_q(B\mu_X + b, B\Sigma_X B')$$

Sea A una matriz cuadrada no singular ($p \times p$).

- Si B matriz ($m \times p$), entonces $rango(BA) = rango(B)$.
- Si C matriz ($p \times n$), entonces $rango(AC) = rango(C)$.

Si D es una matriz ($m \times n$), entonces $rango(DD') = rango(D'D) = rango(D) = rango(D')$.

En este caso, tenemos $\Sigma_X = CC'$, por lo que:

$$B\Sigma_X B' = BCC'B' = (BC)(BC)'$$

con B no singular, y $rango(C) = q$, entonces $rango(BC) = rango(B) = q$, y por lo tanto, $rango((BC)(BC)') = rango(BC) = q$.

1.5.5. Marginalización (caso $\Sigma > 0$)

En particular, del resultado anterior, se obtiene el siguiente, sobre normalidad de las distribuciones marginales en una DNM

- Resultado 5: Sea
$$X = (X_1, \ldots, X_p)' \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$
Entonces, para todo subvector
$$X_r = (X_{r_1}, \ldots, X_{r_q})', \qquad r = (r_1, \ldots, r_1)', \qquad q \leq p$$
(con $r_1, \ldots, r_1$ distintos entre sí) se tiene que
$$X_r \sim N_q(\mu_r, \Sigma_r) \qquad (\Sigma_r > 0)$$
siendo
$$\mu_r \text{ el subvector de } \mu \text{ correspondiente a } r$$
$$\Sigma_r \text{ la submatriz principal de } \Sigma \text{ correspondiente a } r$$
(11)

- (11) Tomamos $B_{q\times p} = (b_{ij})_{\substack{i=1,\ldots,q \\ j=1,\ldots,p}}$ con $b_{ij} = \begin{cases} 1 & \text{si } j = r_i \\ 0 & \text{si } j \neq r_i \end{cases}$.
$$X_r = BX + b \sim N_q(B\mu + b, B\Sigma_X B')$$
con $B\mu + b = B\mu = \mu_r$ y $B\Sigma B' = \Sigma_r$, por lo que $X_r \sim N_q(\mu_r, \Sigma_r)$.

1.5.6. Caracterización de la DNM en términos de normalidad de combinaciones lineales de las componentes (caso $\Sigma > 0$)

- Caracterización 3: Un vector aleatorio p-dimensional $X = (X_1, \ldots, X_p)'$ tiene una DNM no singular si y solo si toda combinación de la forma
$$\alpha' X, \qquad \text{con } \alpha \in \mathbb{R}^p \setminus \{0\}$$
tiene una DN univariante no degenerada (i. e., con varianza no nula) (12).

- (12)

  $\Rightarrow$ $X \sim N_p(\mu, \Sigma), \Sigma > 0$, $\alpha = (\alpha_1, \ldots, \alpha_p) \in \mathbb{R}^p \setminus \{0\}$.
  $Y_\alpha = a'X \Rightarrow Y_\alpha \sim N(\alpha'\mu, \alpha'\Sigma\alpha)$

  $\Leftarrow$ $\alpha \in \mathbb{R}^p \setminus \{0\}$, $Y_\alpha = \alpha' X \sim N(\mu_\alpha, \sigma^2_\alpha), \sigma^2_\alpha > 0$.
  $\mu_\alpha = \alpha'\mu_X, \sigma^2_\alpha = \alpha'\Sigma_X\alpha > 0$.
  $\phi_X(t) = \phi_{Y_t}(1), t \in \mathbb{R}^p$.
  $\phi_{Y_t}(s) = exp\{is\mu_t - \frac{1}{2}\sigma^2_t s^2\}$.
  $\phi_{Y_t}(1) = exp\{it'\mu_t - \frac{1}{2}t'\Sigma_X t\} \Rightarrow X \sim N_p\{\mu_X, \Sigma'_X\}, \Sigma_X > 0$.

1.5.7. Caracterización de la DNM en términos de la densidad esférica estándar (caso $\Sigma > 0$)

- Sea $X = (X_1, \ldots, X_p)'$ un vector aleatorio con DNM no singular,
$$X \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$
Para alguna elección de una matriz C no singular con $\Sigma = CC'$, consideramos la normalización de X dada por el vector aleatorio
$$Z := C^{-1}(X - \mu) \sim N_p(0, I_p)$$

Definimos ahora el vector aleatorio

$$U := \frac{Z}{||Z||} = \frac{Z}{(Z'Z)^{\frac{1}{2}}}$$

Puesto que $||U|| = ||\frac{Z}{||Z||}|| = \frac{||Z||}{||Z||} = 1$, se tiene que U se distribuye sobre la esfera unidad en $\mathbb{R}^p, \mathcal{S}_p$. Además, se cumple que

$$U \stackrel{d}{=} HU$$

para toda matriz ortogonal H de dimensión p (i.e., $HH' = H'H = I_p$); es decir, U ha de tener la distribución uniforme sobre $\mathcal{S}_p$ (distribución esférica sobre la esfera unidad) (13).

(13)

$$HU = H\frac{Z}{||Z||} = H\frac{Z}{(Z'Z)^{\frac{1}{2}}} = H\frac{Z}{(Z'H'HZ)^{\frac{1}{2}}} = \frac{HZ}{((HZ)(HZ)')^{\frac{1}{2}}} =$$

$$= \frac{HZ}{||HZ||} \stackrel{*}{=} \frac{HZ}{||Z||} \stackrel{d}{=} \frac{Z}{||Z||} \Rightarrow HU \stackrel{d}{=} U$$

$*$ se debe a:

$$Z \sim N_p(0, I_p)$$

$$HZ \sim N_p(H \cdot 0, HI_pH') = N_p(0, I_p)$$

- Un resultado importante: El vector aleatorio

$$U = \frac{Z}{(Z'Z)^{\frac{1}{2}}}$$

y la variable aleatoria

$$R = (Z'Z)^{\frac{1}{2}}$$

son independientes.

(Este resultado se demuestra, en general, para vectores aleatorios con distribución esférica)

- Caracterización 4: Un vector aleatorio p-dimensional $X = (X_1, \ldots, X_p)'$ tiene una DNM no singular,

$$X \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$

si y sólo si las realizaciones de X pueden generarse a partir del siguiente procedimiento (experimento aleatorio):

  1. Observar el valor de una variable aleatoria $R^2 := V \sim \mathcal{X}_p^2$ (Se identifica $V = v$).
  2. Elegir aleatoriamente (i.e., según la distribución uniforme), y de forma independiente, un punto en la esfera p-dimensional de radio $r = \sqrt{v}$ (Se identifica como $Z = (Z_1, \ldots, Z_p)' = (z_1, \ldots, z_p)' = z$).
  3. Dados un vector de medias μ y una matriz de covarianzas $\Sigma > 0$, eligiendo C no singular tal que $\Sigma = CC'$, obtener el valor de X mediante la expresión $X := CZ + \mu$ (Se identifica $X = x = Cz + \mu$).

1.6. Extensión al caso general, $\Sigma \geq 0$

1.6.1. Introducción

- En el apartado anterior se ha introducido la DNM en el caso $\Sigma > 0$, dándose varias formulaciones equivalentes, una de las cuales se ha adoptado como definición y las otras como caracterizaciones:

  D En términos de la densidad asociada (se requiere formalmente la condición $|\Sigma| \neq 0$ y, por tanto, $\exists \Sigma^{-1}$).

C-I Como transformación lineal (no singular) de un vector de componentes independientes con DN univariante estándar.

C-II En términos de la forma de la función característica.

C-III Bajo la condición de normalidad (no degenerada) de combinaciones lineales de las componentes.

C-IV Mediante transformaciones lineales (no singulares) a partir de la densidad esférica estándar y radio aleatorio $\sqrt{\chi_p^2}$ (donde también se requiere explícitamente la existencia de Σ^{-1})

- Con el fin de dar una formulación de la DNM extendida al caso general en que $\Sigma \geq 0$ (pudiendose ser singular, es decir, semidefinida positiva), se observa que [C-I], [C-II] y [C-III] ofrecen una viabilidad más directa, especialmente en el caso de [C-II].

- En este apartado, se trata de introducir dicha formulación extendida partiendo de la identificación de la DNM por su función característica (enfoque [C-II]) y discutiendo las implicaciones con respecto a las otras formas de interpretación.

1.6.2. Sobre la descomposición espectral de la matriz de covarianzas, $\Sigma \geq 0$

- Sea Σ la matriz de covarianzas de algún vector aleatorio $p-$dimensional. Por ser $\Sigma \geq 0$ y simétrica, puede factorizarse de la forma:

$$\Sigma = H\Lambda H'$$

con H una matriz $p \times p$ ortogonal ($HH' = H'H = I_p$), Λ una matriz $p \times p$ diagonal, cuyos elementos diagonales distintos son los autovalores de Σ.

Sea $r = rango(\Sigma), r \leq p$. Supongamos, por conveniencia, que las matrices se eligen de modo que

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

con D una matriz $r \times r$ no singular, escribiendo

$$H = \begin{pmatrix} H_1 & H_2 \end{pmatrix}$$

con H_1 una matriz $p \times r$ y H_2 una matriz $p \times (p - r)$.

Se tiene entonces que

$$\Sigma = \begin{pmatrix} H_1 & H_2 \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} = H_1 D H_1'$$

1.6.3. Función característica de la DNM (caso $\Sigma \geq 0$)

Introducimos formalmente la DNM en el caso general ($\Sigma \geq 0$) mediante la identificación de su función característica:

- Definición (extensión de [C-II]): Un vector aleatorio $p-$dimensional X con vector de medias μ y matriz de covarianzas $\Sigma \geq 0$ tiene una DNM, $X \sim N_p(\mu, \Sigma)$, su y sólo si su función característica tiene la forma:

$$\phi_X(t) = exp\{it'\mu - \frac{1}{2}t'\Sigma t\}$$

- Antes de tratar las extensiones de [C-I] y [C-II] a partir de esta definición, veremos un resultado general sobre transformaciones lineales.

1.6.4. Transformaciones lineales de rango no necesariamente máximo (caso $\Sigma \geq 0$)

- Resultado: Sean
$$X \sim N_p(\mu_X, \Sigma_X) \qquad (\Sigma_X \geq 0)$$
$$Y = BX + b$$
con B matriz $q \times p$ (cte.) ($\Rightarrow rango(B) \leq \text{mín}\{q, p\}$), b vector $q \times 1$ (cte.).

  Entonces, se tiene que
$$Y \sim N_q(\mu_Y, \Sigma_Y) = N_q(B\mu_X + b, B\Sigma_X B')$$
(siendo $rango(B\Sigma_X B') \leq \text{mín}\{rango(B), rango(\Sigma_X)\}$) (1).

(1) $t \in \mathbb{R}^q$,
$$\phi_Y(t) = E[e^{it'Y}] = E[e^{it'(BX+b)}] = e^{it'b}E[e^{it'BX}] = e^{it'b}E[e^{i(B't)'X}] = e^{it'b}\phi_X(B't) =$$
$$= e^{it'b}[e^{i(B't)\mu_X - \frac{1}{2}(B't)'\Sigma_X(B't)}] = e^{it'(B\mu_X+b)-\frac{1}{2}t'(B\Sigma_X B')t} = e^{it'\mu_Y - \frac{1}{2}t'\Sigma_y t} \Rightarrow$$
$$\Rightarrow Y \sim N_q(B\mu_X + b, B\Sigma_X B')$$

Para ver el rango:
$$rango(B(\Sigma_X B')) \leq \text{mín}\{rango(B), rango(\Sigma_X B')\}$$

Como:
$$rango(\Sigma_X B') \leq \text{mín}\{rango(\Sigma_X), rango(B)\}$$

tenemos:
$$rango(B\Sigma_X B') \leq \text{mín}\{rango(B), rango(\Sigma_X)\}$$

1.6.5. Extensión de la caracterización [C-I] (caso $\Sigma \geq 0$)

- Intuitivamente, en el caso en que la matriz Σ sea singular, la variabilidad del vector aleatorio X, con DNM (versión extendida) podrá ser explicada (con probabilidad 1) en términos de un número de variables aleatorias, con DN univariante e independientes, igual al rango de Σ. A continuación, vemos la argumentación formal para llegar al enunciado de la correspondiente caracterización, como extensión de [C-I].
- A partir de la descomposición espectral de Σ, vista al principio, de la forma:
$$\Sigma = H_1 D H_1'$$
y teniendo en cuenta que
$$HH' = \begin{pmatrix} H_1 & H_2 \end{pmatrix} \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} = H_1 H_1' + H_2 H_2'$$
podemos escribir la función característica de X como:
$$\phi_X(t) = e^{it'\mu - \frac{1}{2}t'\Sigma t} = e^{it'HH'\mu - \frac{1}{2}t'H_1 D H_1' t} = e^{it'H_1 H_1'\mu - \frac{1}{2}t'H_1 D H_1 t} e^{it'H_2 H_2'\mu}$$

- Esta expresión sugiere considerar el cambio de variables lineal
$$Y := H'X = \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} X = \begin{pmatrix} H_1'X \\ H_2'X \end{pmatrix} =: \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$
de modo que, por ser H ortogonal, se tiene también que $X = HY$.

  Los momentos de primer y segundo orden de Y vienen dados por
$$\mu_y = H'\mu = \begin{pmatrix} H_1'\mu \\ H_2'\mu \end{pmatrix} = \begin{pmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{pmatrix}$$

$$\Sigma_Y = H'\Sigma H = H'H\Lambda H'H = \Lambda = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} =: \begin{pmatrix} \Sigma_{Y_1} & 0 \\ 0 & \Sigma_{Y_2} \end{pmatrix}$$

Es decir,

$$Y_1 \sim N_r(H_1'\mu, D)$$
$$Y_2 \sim N_{p-r}(H_2'\mu, 0)$$

(es decir, la distribución degenerada en $H_2'\mu : P[Y_2 = H_2'\mu] = 1$)

- Por tanto, tenemos que la función característica de X se puede interpretar como

$$\phi_X(t) = E[e^{it'X}] = E[E^{it'HY}] = \phi_Y(H't)$$

Con el cambio de variables

$$v := H't = \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} t = \begin{pmatrix} H_1't \\ H_2't \end{pmatrix} =: \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

podemos escribir

$$\phi_Y(v) = \phi_X(t) = e^{it'H_1H_1'\mu - \frac{1}{2}t'H_1DH_1't} e^{it'H_2H_2'\mu} = e^{iv_1'\mu_{Y_1} - \frac{1}{2}v_1'\Sigma_{Y_1}v_1} e^{iv_2'\mu_{Y_2}} = \phi_{Y_1}(v_1)\phi_{Y_2}(v_2)$$

Es decir, Y_1 e Y_2 son independientes.

- Ahora, puesto que $Y_1 \sim N_r(H_1'\mu, D)$, con $D > 0$, se tiene por [C-I] que Y_1 puede representarse en distribución como

$$Y_1 \stackrel{d}{=} D^{\frac{1}{2}}Z + H_1'\mu$$

con $Z \sim N_R(0, I_r)$.

Para Y_2 se puede escribir

$$Y_2 \stackrel{P-c.s.}{=} H_2'\mu \equiv 0Z + H_2'\mu$$

siendo aquí 0 una matriz $(p-r) \times r$.

Conjuntamente (teniendo en cuenta que $\stackrel{P-c.s.}{=} \Rightarrow \stackrel{d}{=}$), podemos escribir

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

Se deshace el cambio de variables multiplicando por H,

$$X = NY \stackrel{d}{=} H \begin{pmatrix} D^{\frac{1}{2}} \\ 0 \end{pmatrix} Z + HH'\mu = H_1D^{\frac{1}{2}}Z + \mu$$

Finalmente, denotando $A = H_1D^{\frac{1}{2}}$, se tiene que

$$X \stackrel{d}{=} AZ + \mu$$

con A matriz $p \times r$ (cte.) de rango r, $Z \sim N_r(0, I_r)$.

- Recíprocamente, es inmediato que si se cumple que

$$X \stackrel{d}{=} AZ + \mu$$

con A matriz $p \times r$ (cte.) de rango r, $Z \sim N_r(0, I_r)$, teniendo en cuenta que la correspondencia biunívoca entre distribuciones y funciones características (teorema de inversión de Paul Lévy), ha de ser

$$\phi_X \equiv \phi_{AZ+\mu}$$

Es decir, para todo $t \in \mathbb{R}^p$

$$\phi_X(t) = \phi_{AZ+\mu}(t) = E[e^{it'(AZ+\mu)}] = e^{it'\mu}E[e^{i(t'A)Z}] = e^{it'\mu}\phi_Z(A't) =$$

$$= e^{it'\mu}e^{-\frac{1}{2}(t'A)(A't)} = e^{it'\mu-\frac{1}{2}t'(AA')t}$$

Por tanto, según la definición general de la DNM dada a partir de la extensión de [C-II],

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

con $\Sigma_X = AA' \geq 0$.

- Se puede enunciar, como conclusión, la caracterización siguiente:

  Caracterización 1 (caso general): Un vector aleatorio p-dimensional X tiene una DNM, $X \sim N_p(\mu, \Sigma)$ $(\Sigma \geq 0)$, si y solo si
  $$X \stackrel{d}{=} AZ + \mu$$
  con A matriz $p \times r$ (cte.), $rango(A) = r$, $AA' = \Sigma$, $Z \sim N_r(0, I_r)$.

1.6.6. Extensión de la caracterización [C-III] (caso $\Sigma \geq 0$)

En relación con la caracterización [C-III], el resultado extendido al caso general puede enunciarse de la forma siguiente:

- Caracterización 3 (caso general): Un vector aleatorio p-dimensional $X = (X_1, \ldots, X_p)'$ tiene una DNM (versión extendida) si y solo si toda combinación de la forma

  $$\alpha' X \qquad \text{con } \alpha \in \mathbb{R}^p$$

  tiene una DN univariante (posiblemente degenerada, i.e. con varianza nula) (2).

  (La demostración es similar a la correspondiente al caso no singular, con mínimas modificaciones)

(2) $\Rightarrow$ Supongamos $X \sim N_p(\mu, \Sigma)$ $(\Sigma \geq 0)$.
Sea $\alpha \in \mathbb{R}^p$ y sea $Y = \alpha' X$.
Tomamos $B_{1\times p} = \alpha'$, $b_{1\times 1} = 0$. Aplicando el resultado de transformaciones lineales de rango no necesariamente máximo, tenemos:

$$Y_\alpha \sim N_1(\alpha'\mu, \alpha'\Sigma\alpha) \qquad \text{con } \alpha'\Sigma\alpha \geq 0$$

en el caso en el que $\alpha'\Sigma\alpha = 0 \Rightarrow$ degenerada.

$\Leftarrow$ Supongamos $Y_\alpha = \alpha' X \sim N_1(\mu_{Y_\alpha}, \sigma^2_{Y_\alpha})$, con $\alpha \in \mathbb{R}^p$.
Necesariamente se tiene $\mu_{Y_\alpha} = \alpha'\mu$ y $\sigma_{Y_\alpha} = \alpha'\Sigma\alpha$, por lo que $X \sim (\mu, \Sigma)$.
Utilizamos el Teorema de la página 8 de vectores aleatorios: La distribución de X queda unívocamente definida por el conjunto de todas las distribuciones univariantes de la forma $\alpha' X$ del siguiente modo:

$$\phi_{Y_t}(1) = \phi_X(t) \qquad \text{con } Y_t = t'X \text{ (hemos cambiado } \alpha \text{ por } t)$$

$$\phi_{Y_t}(s) = exp\{is't'X - \frac{1}{2}s't'\Sigma ts\}$$

para $s = 1$:

$$\phi_{Y_t}(1) = exp\{it'X - \frac{1}{2}t'\Sigma t\} = \phi_X(t)$$

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

1.6.7. Discusión en relación con la posible no existencia de densidad en todo $\mathbb{R}^p$ (caso $\Sigma \geq 0$)

- En referencia a la definición basada en la función de densidad, dada para el caso no singular ($\Sigma > 0$), en el caso general que nos ocupa ($\Sigma \geq 0$) sólo existirá propiamente la densidad para un subvector, con dimensión igual al rango de la matriz Σ, de un vector obtenido a partir de X mediante un cambio de variables conveniente.

  Dicho cambio estará asociado a la descomposición espectral de Σ, y el subvector corresponderá a las componentes generadas por los autovalores estrictamente positivos de Σ.

  Es decir, se tiene que la distribución es no singular en un subespacio afín de dimensión $r = rango(\Sigma)$. Para el resto de dimensiones, la distribución es degenerada.

- Análogamente, se podrá dar una representación como distribución simétrica de contornos elípticos restringida al subespacio afín de dimensión $r = rango(\Sigma)$ mencionado.

1.6.8. Normalidad de combinaciones lineales de vectores aleatorios con DNM (versión general) independientes

- Resultado: Sean X_k, $k = 1, \dots, m$, vectores aleatorios p-dimensionales con DNM

$$X_k \sim N_p(\mu_k, \Sigma_k) \qquad \text{respectivamente)}$$

  e independientes. Entonces, para cualquier conjunto de matrices A_k, $k = 1, \dots, m$ (ctes.), de dimensión $q \times p$, se verifica que

$$Y := \sum_{k=1}^{m} A_k X_k \sim N_q(\sum_{k=1}^{m} A_k \mu_k, \sum_{k=1}^{m} A_k \Sigma_k A_k') \qquad (3)$$

- (3) $X_k \sim N_p(\mu_k, \Sigma_k)$ indep., $A_k \in \mathcal{M}_{q \times p}$ (ctes.) $\Rightarrow$
  $\Rightarrow Y = \sum_{k=1}^{m} A_k X_k \sim N_q(\sum_{i=1}^{m} A_k \mu_k, \sum_{k=1}^{m} A_k \Sigma_k A_k')$.

  Demostración:

$$\phi_Y(t) = E[e^{it'Y}] = E[e^{it'(\sum_{k=1}^{m} A_k X_k)}] = E[e^{\sum_{k=1}^{m} it' A_k X_k}] = E[e^{\sum_{i=1}^{m} i(A_k' t)' X_k}] =$$

$$= \prod_{k=1}^{m} E[e^{i(A_k' t)' X_k}] = \prod_{k=1}^{m} \phi_{X_k}(A_k t X_k) = \prod_{k=1}^{m} e^{it' A_x \mu_X - \frac{1}{2} t' A_k \Sigma_k A_k' t} = e^{\sum_{k=1}^{m}(it' A_k \mu_k - \frac{1}{2} t' A_k \Sigma_k A_k' t)} =$$

$$= e^{it' \sum_{k=1}^{m} A_k \mu_k - \frac{1}{2} t' \sum_{k=1}^{m} (A_k \Sigma_k A_k)' t}$$

1.6.9. [Complemento] El "problema de Lévy-Cramér" (teorema de Cramér)

- Como consecuencia del resultado anterior, se tiene, en particular, el siguiente caso simple:

  Sean X_1 y X_2 vectores aleatorios p-dimensionales con

$$X_1 \sim N_p(\mu_1, \Sigma_1) \qquad X_2 \sim N_p(\mu_2, \Sigma_2)$$

  e independientes. Entonces

$$X_1 + X_2 \sim N_p(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$$

- En este caso (i.e. bajo normalidad), se tiene que el recíproco también es cierto:

  Sean X_1 y X_2 vectores aleatorios p-dimensionales independientes tales que $X_1 + X_2$ tiene DNM. Entonces, X_1 y X_2 tiene también, cada uno, DNM.

- Históricamente, el resultado fue conjeturado por Paul Lévy, aunque no fue demostrado formalmente, por Harald Cramér, hasta 1936, mediante herramientas de la teoría de funciones analíticas, para el caso $p = 1$ (4).

- (4) $X = X_1 + X_2$ tiene una DNM $\overset{C-III}{\Rightarrow} \forall \alpha \in \mathbb{R}^n, \alpha' X$ tiene una DN u.
$\alpha' X = \alpha' X_1 + \alpha' X_2$ tiene DN u.
X_1 y X_2 son independientes $\Rightarrow \alpha X_1$ y αX_2 son independientes.
$\Rightarrow \alpha' X_1$ y $\alpha' X_2$ tienen DN u. $\overset{C-III}{\Rightarrow} X_1$ y X_2 tienen DNM.

- No veremos aquí la demostración. No obstante, es fácil probar (por ejemplo, mediante la caracterización [C-III]) que, partiendo de que el resultado es cierto para $p = 1$, lo es inmediatamente para cualquier p.

1.7. Complemento: Distribuciones esféricas y elípticas

1.7.1. Distribuciones esféricas

- Definición: Se dice que un vector aleatorio $X = (X_1, \ldots, X_p)'$ tiene una distribución esférica si X y HX tienen la misma distribución, para toda matriz ortogonal H de dimensión $p \times p$.

- Caracterización 1 (en términos de la densidad, si existe): Un vector aleatorio $X = (X_1, \ldots, X_p)'$ con distribución de tipo continuo tiene una distribución esférica si y sólo si su función de densidad puede expresarse como
$$f_X(x) = g(x'x)$$
para alguna función escalar $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$.

- Caracterización 2 (en términos de la función característica): Un vector aleatorio
$X = (X_1, \ldots, X_p)'$ tiene una distribución esférica si y sólo si su función característica es de la forma
$$\phi_X(t) = \xi(t't)$$
para alguna función escalar $\xi : \mathbb{R}_0^+ \to \mathbb{R}$.

- Ejemplo 1: $Z \sim N_p(0, \sigma^2 I_p)$ tiene una distribución esférica.

  (Observación: Bajo la caracterización 1 se requerirá que la distribución sea no degenerada, i.e. $\sigma^2 > 0$) (1).

- (1) $Z \sim (0, \sigma^2 I_p)$
Opción 1:
$$f_Z(z) = \frac{1}{(2\pi)^{\frac{p}{2}} |\sigma|^p} exp\{-\frac{1}{2}(z-\mu)' \Sigma^{-1} (z-\mu)\} = \frac{1}{(2\pi)^{\frac{p}{2}} |\sigma|^p} exp\{-\frac{1}{2\sigma^2} Z'Z\}$$
$$Z'Z = Y$$
$$g(Y) = \frac{1}{(2\pi)^{\frac{p}{2}} |\sigma|^p} exp\{-\frac{1}{2\sigma^2} Y\}, g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$$
Opción 2:
$$\phi_Z(t) = \exp\{it'\mu_Z - \frac{1}{2} t' \Sigma_Z t\} = exp\{it' \cdot 0 - \frac{1}{2} t' \sigma^2 I_p t\} = exp\{-\frac{1}{2} \sigma^2 t't\}$$
$$\xi(t't) = \xi(s) = e^{-\frac{1}{2}\sigma^2 s}, \xi : \mathbb{R}_0^+ \to \mathbb{R}$$

- Ejemplo 2: Sea X un vector aleatorio con distribución esférica y tal $P[X = 0] = 0$. Se define $U := \frac{X}{||X||}$, para $X \neq 0$, y con cualquier asignación cuando $X = 0$ (por ejemplo, en la propia esfera unidad, $\mathcal{S}_p \subset \mathbb{R}^p$). Entonces U también tiene una distribución esférica (2).

- Resultado auxiliar: Sean X e Y vectores aleatorios p-dimensionales. Sea $g : \mathbb{R}^p \to \mathbb{R}^q$ una aplicación medible (Borel). Entonces:

$$X \stackrel{d}{=} Y \Rightarrow g(X) \stackrel{d}{=} g(Y)$$

- Ejemplo 3: Más generalmente, si $X = (X_1, \ldots, X_p)'$ es un vector aleatorio p-dimensional con distribución esférica, se tiene entonces que, dada cualquier transformación radial "isotrópica" Borel-medible $h : \mathbb{R}^p \to \mathbb{R}^p$, de la forma

$$h(x) = g(||x||)x, \qquad \text{con } g : \mathbb{R}_0^+ \to \mathbb{R}_0^+ \text{ Borel-medible}$$

el vector aleatorio transformado $h(X) = (h_1(X), \ldots, h_p(X))'$ también tendrá una distribución esférica.

- Resultado (extensión de un resultado visto anteriormente):

Si X tiene una distribución esférica con $P[X = 0] = 0$, y se definen

$$R = ||X|| = (X'X)^{\frac{1}{2}}$$

$$U = \frac{X}{(X'X)^{\frac{1}{2}}} = R^{-1}X, \qquad \text{para } X \neq 0$$

(con cualquier asignación para U cuando $X = 0$; por ejemplo, en la propia esfera unicidad $\mathcal{S}_p \subset \mathbb{R}^p$), entonces U tiene también una distribución esférica, y R y U son independientes.

- Este resultado significa que la distribución de X, en el caso de una distribución esférica, viene en realidad determinada por la distribución de la variable aleatorio R.

1.7.2. Definiciones elípticas

- Definición: Se dice que un vector aleatorio $X = (X_1, \ldots, X_p)'$ tiene una distribución elíptica, con parámetros dados por un vecor p-dimensional μ y una matriz $(p \times p)$-dimensional V simétrica y definida positiva, si se puede expresar de la forma

$$X = AZ + \mu$$

con $V = AA'$, A matriz $p \times p$ no singular, Z vector aleatorio con distribución esférica.

Se denotará, en este caso, $X \sim E_p(\mu, V)$.

- Caracterización 1 (en términos de la densidad, si existe): Un vector aleatorio $X = (X_1, \ldots, X_p)'$ con distribución de tipo continuo tiene una distribución elíptica si y solo si su función de densidad puede expresarse como

$$f_X(x) = |V|^{-\frac{1}{2}} g((x-\mu)'V^{-1}(x-\mu))$$

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

- Caracterización 2 (en términos de la función característica): Un vector aleatorio $X = (X_1, \ldots, X_p)'$ tiene una distribución elíptica su y sólo si su función característica es de la forma

$$\phi_X(t) = e^{it'\mu}\xi(t'Vt)$$

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

- Ejemplo: $X \sim N_p(\mu, \Sigma)$, $(\Sigma > 0)$ tiene una distribución elíptica, $X \sim E_p(\mu, \Sigma)$. En este caso, por tanto, se tiene que $V = \Sigma$.

- Observaciones:
  - Si $X \sim E_p(0, \sigma^2 I_p)$, con $\sigma^2 > 0$, entonces X tiene una distribución esférica (y recíprocamente).

- Si para $X \sim E_p(\mu, V)$ existen los momentos $E[X]$ y $Cov(X)$, entonces se tiene que

$$E[X] = \mu$$

$$Cov(X) = \alpha V, \qquad \text{con } \alpha = -2\xi'(0)$$

- Las distribuciones marginales de distribuciones elípticas son elípticas.
- Las distribuciones condicionadas de distribuciones elípticas son elípticas.
- Independencia y normalidad: Sea $X \sim E_p(\mu, V)$, con V una matriz diagonal. Si $X_1, \ldots, X_p$ son (mutuamente) independientes, entonces $X \sim N_p(\mu, V)$.

1.8. Formas cuadráticas basadas en vectores aleatorios con DNM

1.8.1. Motivación

- Sea $X = (X_1, \ldots, X_p)'$ un vector aleatorio y sea A una matriz (cte.) $(p \times p)$-dimensional. Se considera la variable aleatoria dada por la forma cuadrática

$$X'AX$$

  Se plantea, en general, el problema de estudiar la distribución de $X'AX$ a partir del conocimiento de la distribución de X.

  (Como extensión, si X se reemplaza por una matriz aleatoria de dimensión $p \times q$, la forma cuadrática resultante constituirá una matriz aleatoria de dimensión $q \times q$).

- En particular, se considera el caso en que $X \sim N_p(\mu, \Sigma)$, que conduce a la distribución χ^2 no centrada.

  La distribución χ^2 no centrada y, definida a partir de ésta, la distribución F no centrada, son fundamentales, como distribuciones de diversos estadísticos de interés, en relación con la inferencia basada en la DNM.

1.8.2. Distribuciones χ^2 y F no centradas

DISTRIBUCIÓN χ^2 CENTRADA.

- Recordemos que la distribución χ^2 centrada con n grados de libertad se define como la distribución de la suma de cuadrados de n variables aleatorias independientes con distribución normal estándar:

$$Z = (Z_1, \ldots, Z_n)' \sim N_n(0, I_n) \to Y = Z'Z = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

  Función de densidad: $f_Y(y) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, y > 0.$

  Función de distribución: $F_Y(y) = \frac{\gamma(\frac{n}{2}, \frac{y}{2})}{\Gamma(\frac{n}{2})}, y > 0.$

- Nota: Se definen, $\forall z \in \mathbb{C}, \mathcal{R}(z) > 0$,
  - Función gamma: $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt.$
  - Funciones gamma incompletas: para $v > 0$,

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

Función característica: $\phi_Y(t) = (1-2it)^{-\frac{n}{2}}, \forall t \in \mathbb{R}$, de donde se obtienen, en particular, los momentos

$$E[Y] = n$$
$$Var(Y) = 2n$$

DISTRIBUCIÓN χ^2 NO CENTRADA.

- Definición (resultado): Sea $X = (X_1, \ldots, X_n)' \sim N_n(\mu, I_n)$. Entonces, la variable aleatoria $Y = X'X$ tiene función de densidad
$$f_Y(y) = e^{-\frac{\delta}{2}} {}_0F_1(;\frac{1}{2}n; \frac{1}{4}\delta y) \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}, \text{ para } y > 0$$
siendo $\delta = \mu'\mu$. Se dice que la variable Y tiene distribución χ^2 no centrada con n grados de libertad y parámetro de no centralidad δ, denotándose $\chi^2_n(\delta)$.

- Nota: En la expresión anterior, ${}_0F_1$ representa la 'función hipergeométrica generalizada' de órdenes 0 y 1, también llamada función hipergeométrica confluente límite.

  (Cuando $\mu = 0$, se tiene la distribución χ^2_n centrada)

  Función característica: $\phi_Y(t) = (1-2it)^{-\frac{n}{2}} e^{\frac{it\delta}{1-2it}}, \forall t \in \mathbb{R}$, de donde se obtienen, en particular, los momentos
$$\begin{cases} E[Y] = n + \delta \\ Var(Y) = 2n + 4\delta \end{cases} \quad (1)$$

- (1)
$$\Phi_Y(t) = (1-2it)^{\frac{n}{2}} e^{\frac{it\delta}{1-2it}}$$
$$iE[Y] = \frac{\delta}{\delta t}\Phi_Y(t) \mid_{t=0}=$$
$$= -\frac{n}{2}(1-2it)^{-\frac{n}{2}-1}(-2i)e^{\frac{it\delta}{1-2it}} + (1-2i)^{-\frac{n}{2}} \frac{i\delta(1-2it) + 2i(it\delta)}{(1-2it)^2} e^{\frac{it\delta}{1-2it}} \mid_{t=0}= i(n+\delta) \Rightarrow$$
$$\Rightarrow E[Y] = n + \delta$$

$$i^2E[Y^2] = \frac{\delta^2}{\delta t^2}\Phi_Y(t) \mid_{t=0}= \frac{\delta}{\delta t}[ni(1-2it)^{-\frac{n}{2}-1}e^{\frac{it\delta}{1-2it}} + (1-2it)^{-\frac{n}{2}-2}i\delta e^{\frac{it\delta}{1-2it}}] \mid_{t=0}=$$
$$= \frac{\delta}{\delta t}[(ni(1-2it)^{-1} + \delta i(1-2it)^{-2})\Phi_y(t)]_{t=0} = [....] = -2n - 4\delta + i^2(n+\delta)^2 =$$
$$= -(n^2 + 2n + 2n\delta + 4\delta + \delta^2) \Rightarrow E[Y^2] = n^2 + 2n + 2n\delta$$
$$Var[Y] = E[Y^2] - E[Y]^2 = n^2 + 2n + 2n\delta + 4\delta + \delta^2 - (n+\delta)^2 = 2n + 4\delta$$

- Nota: $X \sim N(\mu, \sigma^2) \Rightarrow \phi_{X^2}(t) = \frac{1}{(1-2i\sigma^2 t)} exp\{\frac{it\mu}{1-2i\sigma^2 t}\}$.

- Un resultado de interés: Si Y_1 y Y_2 son variables aleatorias independientes con
$$Y_1 \sim \chi^2_{n_1}(\delta_1), \qquad Y_2 \sim \chi^2_{n_2}(\delta_2)$$
entonces se tiene que
$$Y_1 + Y_2 \sim \chi^2_{n_1+n_2}(\delta_1 + \delta_2) \qquad (2)$$

- (2) $Y = Y_1 + Y_2$.
$$\Phi_Y(t) = E[e^{it(Y_1+Y_2)}] = E[E^{itY_1+itY_2}] \overset{Y_1,Y_2 \text{ ind}}{=} E[e^{itY_1}]E[e^{itY_2}] = \Phi_{Y_1}(t)\Phi_{Y_2}(t)$$
$$Y_1 \sim \chi^2_{n_1}(\sigma_1), Y_2 \sim \chi^2_{n_2}(\sigma_2)$$
$$Y_1 + Y_2 \sim \chi^2_{n_1+n_2}(\sigma 1 + \sigma 2)$$

DISTRIBUCIÓN F CENTRADA

- Recordemos que la distribución F centrada con (n_1, n_2) grados de libertad se define como la distribución del cociente:
$$F = \frac{\frac{Y_1}{n_1}}{\frac{Y_2}{n_2}}$$
con $Y_1 \sim \chi^2_{n_1}$ e $Y_2 \sim \chi^2_{n_2}$ (ambas centradas), independientes. Se denota F_{n_1,n_2}.

  Puede verse como la distribución del cociente
$$\frac{\frac{1}{n_1} Z'_{(1)} Z_{(1)}}{\frac{1}{n_2} Z'_{(2)} Z_{(2)}} = \frac{\frac{1}{n_1} \sum_{k=1}^{n_1} Z^2_{1k}}{\frac{1}{n_2} \sum_{l=1}^{n_2} Z^2_{2k}}$$
con: $Z_{(1)} = (Z_{11}, \ldots, Z_{1n_1})' \sim N_{n_1}(0, I_{n_1})$, $Z_{(2)} = (Z_{21}, \ldots, Z_{2n_1})' \sim N_{n_2}(0, I_{n_2})$, $Z_{(1)}$ y $Z_{(2)}$ independientes; equivalentemente,
$$Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix} \sim N_{n_1+n_2}(0, U_{n_1+n_2})$$

  Función de densidad: con $n_1, n_2 \in \mathbb{N} \setminus \{0\}$,
$$g_F(f) = \frac{1}{fB(\frac{n_1}{2}, \frac{n_2}{2})} (\frac{n_1 f}{n_1 f + n_2})^{\frac{n_1}{2}} (1 - \frac{n_1 f}{n_1 f + n_2})^{\frac{n_2}{2}}$$

  Función de distribución: $G_F(f) = I_{\frac{n_1 f}{n_1 f + n_2}}(\frac{n_1}{2}, \frac{n_2}{2}), f \geq 0$.

- Nota: Se definen, $\forall a, b \in \mathbb{C}$, con $\mathcal{R}(a), \mathbb{R}(b) > 0$,
  - Función beta: $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1} dt$.
  - Función beta incompleta: para $x \in [0, 1]$, $B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$, (para $x = 1$ se tiene la función beta completa).
  - Función beta incompleta regularizada: $I_x(a, b) = \frac{B(x;a,b)}{B(a,b)}$.

  Función característica (versión corregida de Philips (1982)):
$$\phi_F(t) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_2}{2})} U(\frac{n_1}{2}, 1 - \frac{n_2}{2}; -\frac{n_2}{n_1} it)$$

- Nota: En la expresión anterior, U representa la función hipergeométrica confluente de segunda especie,
$$U(a, b; z) = \frac{\Gamma(1-b)}{\Gamma(a+1-b)} {}_1F_1(a; b; z) + \frac{\Gamma(b-1)}{\Gamma(a)} z^{1-b} {}_1F_1(a+1-b; 2-b, z)$$
con ${}_1F_1$ la 'función hipergeométrica generalizada de órdenes 1 y 1', también llamada función de Kimar de primera especie).

  Se obtienen, en particular, los momentos
$$E[F] = \frac{n_2}{n_2 - 2}, \text{ para } n_2 > 2 (= \infty \text{ si } n_2 \in (0, 2])$$
$$Var(F) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \text{ para } n_2 > 4 (= \infty \text{ si } n_2 \in (2, 4])$$

DISTRIBUCIÓN F NO CENTRADA

- Definición (resultado): Sean $Y_1 \sim \chi^2_{n_1}(\delta)$ e $Y_2 \sim \chi^2_{n_2}$, independientes. Entonces, la variable

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

tiene función de densidad

$$\begin{aligned} g_F(f) =& e^{-\frac{\delta}{2}} {}_1F_1(\frac{1}{2}(n_1+n_2); \frac{1}{2}n_1; \frac{-\frac{1}{2}\frac{n_1}{n_2}\delta f}{1+\frac{n_1}{n_2}f}) \\ &\times \frac{\Gamma(\frac{1}{2}(n_1+n_2))}{\Gamma(\frac{1}{2}n_1)\Gamma(\frac{1}{2}n_2)} \frac{f^{\frac{n_1}{2}-1}(\frac{n_1}{n_2})^{\frac{n_1}{2}}}{(1+\frac{n_1}{n_2}f)^{\frac{(n_1+n_2)}{2}}}, \quad \text{para } f > 0 \end{aligned}$$

Se dice que F tiene distribución F no centrada con n_1 y n_2 grados de libertad y parámetro de no centralidad σ, denotándose $F_{n_1,n_2}(\delta)$.

Función característica (versión corregida de Philips (1982)):

$$\phi_F(t) = \frac{e^{-\frac{1}{2}\delta}}{\Gamma(\frac{1}{2}n_2)} \sum_{j=0}^{\infty} \frac{(\frac{\delta}{2})^j}{j!} \Gamma(\frac{1}{2}n_1 + \frac{1}{2}n_2 + j) U(\frac{n_1}{2} + j, 1 - \frac{n_2}{2}; -\frac{n_1}{n_2}it)$$

Se obtienen, en particular, los momentos

$$E[F] = \frac{n_2(n_1+\delta)}{n_1(n_2-2)}, \text{ para } n_2 > 2 (= \infty \text{ si } n_2 \in (0,2])$$

$$Var(F) = 2(\frac{n_2}{n_1})^2 \frac{(n_1+\delta)^2 + (n_1+2\delta)(n_2-2)}{(n_2-2)^2(n_2-4)}, \text{ para } n_2 > 4 (= \infty \text{ si } n_2 \in (2,4])$$

1.8.3. Formas cuadráticas $X'AX$, con $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$

A continuación veremos algunos resultados relativos a la distribución de formas cuadráticas del tipo

$$X'AX$$

con $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$ y A una matriz (cte.) $(p \times p)$-dimensional simétrica.

- De forma preliminar, se puede hacer un estudio directo de los momentos de primer y segundo orden:

  Sea $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$ y A una matriz (cte.) $(p \times p)$-dimensional simétrica. Entonces,

  (a) $E[X'AX] = tr(A\Sigma) + \mu' A \mu$ (3).

  (b) $Var(X'AX) = 2tr((A\Sigma)^2) + 4\sigma' A\Sigma A\mu$.

  (Observación: El resultado (a) no requiere la hipótesis de normalidad, solo la existencia de la esperanza)

- (3) $(X \sim N_p(\mu, \Sigma) \Rightarrow E[X'AX] = tr(A\Sigma) + \mu' A\mu$.

$$E[X'AX] = E[tr(X'AX)] = E[tr(AXX')] = tr(E[AXX']) = tr(AE[XX'])$$

  donde hemos usado que $X'AX$ es de dim 1 y que $tr(CB) = tr(BC)$.

  Usaremos ahora que:

$$E[XX'] = \Sigma + \mu\mu'$$

  ya que:

$$\Sigma = E[(X-\mu)(X-\mu)'] = E[XX'] - E[X\mu'] - E[\mu X'] + E[\mu\mu'] =$$

$$= E[XX'] - \mu' E[X] - \mu E[X'] + \mu\mu' = E[XX'] - \mu'\mu - \mu\mu' + \mu\mu' = E[XX'] - \mu\mu'$$

Entonces:

$$E[X'AX] = tr(AE[XX']) = tr(A(\Sigma + \mu\mu')) = tr(A\Sigma + A\mu\mu') =$$
$$= tr(A\Sigma) + tr(A\mu\mu') = tr(A\Sigma) + tr(\mu' A\mu) = tr(A\Sigma) + \mu' A\mu$$

- Resultado 1:
  Sea $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$. Entonces:
  1. $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_p$ (4).
  2. $X'\Sigma^{-1}X \sim \chi^2_p(\delta)$, con $\delta = \mu'\Sigma^{-1}\mu$ (5).
- (4) $X \sim N_p(\mu, \Sigma), \Sigma > 0, \Sigma = CC'$.
  Hacemos el cambio $Y = C^{-1}(X-\mu) \sim N_p(0, I_p)$.

  $$(X-\mu)'\Sigma^{-1}(X-\mu) = (X-\mu)(CC')^{-1}(X-\mu) = (X-\mu)(C^{-1})'C^{-1}(X-\mu) = Y'Y$$

  Como $Y \sim N_p(0, I_p) \Rightarrow YY' \sim \chi^2_p$.
- (5) $V = C^{-1}X \sim N_p(C^{-1}\mu, C^{-1}\Sigma(C^{-1})') \equiv N_p(C^{-1}\mu, C^{-1}CC'(C^{-1})') \equiv N_p(C^{-1}\mu, I_p)$.

  $$X'\Sigma^{-1}X = X'(CC')^{-1}X = [...] = V'V$$

  Como $V \sim N_p(C^{-1}\mu, I_p) \Rightarrow V'V \sim \chi^2_p((C^{-1}\mu)'(C^{-1}\mu)) \equiv \chi^2_p(\mu\Sigma^{-1}\mu)$, donde hemos usado:

  $$(C^{-1}\mu)'(C^{-1}\mu) = \mu'(C^{-1})'C^{-1}\mu = \mu'(CC')^{-1}\mu = \mu'\Sigma^{-1}\mu$$

- Resultado 2:
  Sea $X \sim N_p(\mu, I_p)$ y sea B una matriz (cte.) $(p \times p)$-dimensional simétrica. Entonces, $X'BX$ tiene una distribución χ^2 no centrada si y sólo si B es idempotente (i.e. $B^2 = B$), en cuyo caso los grados de libertad y el parámetro de no centralidad son, respectivamente, $k = rango(B) = tr(B)$ y $\delta = \mu' B\mu$ (6).
- (6)

  $\Rightarrow$ Supongamos $X'BX \sim \chi^2_k(\delta)$. Sea $r = rango(B)$.
  Por ser B simétrica podemos escribir la diagonalización

  $$H'BH = D \qquad (\iff B = HDH')$$

  siendo H matriz $p \times p$ ortogonal y D una matriz $p \times p$ diagonal.

  $$D = \begin{pmatrix} \lambda_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \lambda_r & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}$$

  siendo $\lambda_1, \dots, \lambda_r \neq 0$.
  Transformamos X mediante H, $V = H'X$, siendo entonces

  $$V \sim N_p(H'\mu, H'\Sigma H) \equiv N_p(H'\mu, I_p) \equiv N_p(\mathcal{V}, I_p)$$

  Podemos escribir:

  $$X'BX = X'HDH'X = V'DV = \sum_{j=1}^{r} \lambda_j V_j^2$$

Cada V_j^2 tiene distribución $V_j^2 \sim \chi_1^2\ (\mathcal{V}_j^2)$, siendo $V_1^2, \dots, V_r^2$ independientes.

La función característica de $X'BX$ se obtienen como:

$$\Phi_{X'BX}(t) = \Phi_{\sum_{j=1}^r \lambda_j V_j^2}(t) = \prod_{j=1}^r \Phi_{\lambda_j V_j^2}(t) = \prod_{j=1}^r \Phi_{V_j^2}(\lambda_j t)$$

Cada función característica individual $\Phi_{V_j^2}(\cdot)$ es de la forma:

$$\Phi_{\chi_1^2(\delta)}(s) = (1-2is)^{-\frac{1}{2}} exp\{\frac{is\delta}{1-2is}\}$$

con $\delta = \mathcal{V}_j^2$.
Luego, conjuntamente, se tiene:

$$\Phi_{X'BX}(t) = \prod_{j=1}^r (1-2i\lambda_j t)^{-\frac{1}{2}} exp\{\frac{i\lambda_j t \mathcal{V}_j^2}{1-2i\lambda_j t}\} =$$

$$= (\prod_{j=1}^r (1-2i\lambda_j t)^{-\frac{1}{2}}) exp\{it \sum_{j=1}^r \frac{\lambda_j \mathcal{V}_j^2}{1-2i\lambda_j t}\}$$

Por otra parte, al ser $X'BX \sim \chi_k^2(\delta)$, se tiene que:

$$\Phi_{X'BX}(t) = (1-2it)^{-\frac{k}{2}} exp\{\frac{it\delta}{1-2it}\}$$

Igualando las dos expresiones (para todo $t \in \mathbb{R}$), han de ser:

- $\lambda_j = 1 \qquad j = 1, \dots, r.$
- $\delta = \sum_{j=1}^r \lambda_j \mathcal{V}_j^2 = \sum_{j=1}^r \mathcal{V}_j.$
- $r = k.$

En consecuencia

$$H'BH = D = \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$$

que es una matriz idempotente, es decir, $(H'BH)(H'BH) = H'BH$. Entonces:

$$(H'BH)(H'BH) = H'B^2H = H'BH \Rightarrow B^2 = B$$

por lo que B es idempotente. Además:

$$\delta = \sum_{j=1}^r \mathcal{V}_j^2 = \mathcal{V}'D\mathcal{V} = (H'\mu)'D(H'\mu) = \mu'HDH'\mu = \mu'B\mu$$

$\Leftarrow$ Supongamos B (simétrica) es idempotente y de rango k. Podemos escribir:

$$H'BH = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$$

para alguna matriz ortogonal H.

Definimos el vector:

$$V = H'X \sim N_p(H'\mu, H'I_pH) \equiv N_p(H'\mu, I_p)$$

Por tanto, dado que podemos escribir:

$$X'BX = X'H \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} H'X = V' \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} V = \sum_{i=1}^{k} V_j^2$$

y, puesto que las variables $V_1, \ldots, V_k$ son independientes, se tiene que:

$$X'BX \sim \chi_k^2(\delta)$$

con parámetro de no centralidad:

$$\delta = E[V'] \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} E[V]$$

donde $E[V] = H'\mu$, así que tenemos:

$$\delta = (H'\mu)' \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} (H'\mu) = \mu' H \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} H'\mu = \mu' B \mu$$

- Resultados auxiliares:
  - Una matriz simétrica es idempotente si y sólo si sus autovalores son 0 o 1.

    A parir de esto, si B simétrica idempotente, entonces:

    $$rango(B) = tr(B)$$

  - (Lema 10.1.2 Harville, p. 134) Sea A una matriz cuadrada. Entonces:
    1. A' es idempotente $\iff$ A es idempotente.
    2. $I - A$ es idempotente $\iff$ A es idempotente.
- Resultado 3:
  Sea $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$. Supongamos el particionamiento

  $$X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}$$

  con $X_{(1)}$ y $\mu_{(1)}$ subvectores q-dimensionales y $\Sigma_{(11)}$ submatriz $(q \times q)$-dimensional. Entonces,

  $$Q := (X - \mu)'\Sigma^{-1}(X - \mu) - (X_{(1)} - \mu_{(1)})'\Sigma_{(11)}^{-1}(X_{(1)} - \mu_{(1)}) \sim \chi_{p-1}^2 \qquad (7)$$

- (7) Como $X \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$, se puede descomponer $\Sigma = CC'$. Ponemos $C = \begin{pmatrix} C_{(1)} \\ C_{(2)} \end{pmatrix}$, tal que $\Sigma = CC' = \begin{pmatrix} C_{(1)}C'_{(1)} & C_{(1)}C'_{(2)} \\ C_{(2)}C'_{(1)} & C_{(2)}C'_{(2)} \end{pmatrix}$.

  Normalizamos, definiendo $Y = C^{-1}(X - \mu) \Rightarrow X = CY + \mu$, $Y \sim N_p(0, I_p)$, luego $X_{(1)} = C_{(1)}Y + \mu_{(1)}$.

  $$Q = (CY')\Sigma^{-1}(CY) - Y'C'_{(1)}(C_{(1)}C'_{(1)})^{-1}C_{(1)}Y \Rightarrow$$

  $$\Rightarrow Q = Y'C'(CC')^{-1}CY - [...] = Y'[I_p - C'_{(1)}(C_{(1)}C'_{(1)})^{-1}C_{(1)}]Y$$

$$[C'_{(1)}(C_{(1)}C'_{(1)})^{-1}C_{(1)}]^2 = C'_{(1)}(C_{(1)}C'_{(1)})^{-1}C_{(1)}C'_{(1)}(C_{(1)}C'_{(1)})^{-1}C_{(1)} = C'_{(1)}I_q(C_{(1)}C'_{(1)})^{-1}C_{(1)}$$

$\exists B$ tal que $B^2 = B$ y $Q = YBY \sim \chi^2_k(\delta)$, $\delta = \mu' B\mu = 0'B0 = 0$, $k = rango(B) = tr(B) = tr(I_p) - tr(C'_{(1)}(C_{(1)}C'_{(1)})^{-1}C_{(1)}) = p - tr(\underbrace{C_{(1)}C'_{(1)}(C_{(1)}C_{(1)})^{-1}}_{=I_q}) = p - q$.

- Resultado 4:
  Sea $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$ y sea B una matriz (cte.) $(p \times p)$-dimensional simétrica. Entonces, $X'BX$ tiene una distribución $\chi^2_k(\delta)$, con $k = rango(B)$ y $\delta = \mu' B\mu$, si y sólo si $B\Sigma$ es idempotente (i.e. $(B\Sigma)^2 = B\Sigma$; equivalentemente, en este caso, $B\Sigma B = B$) (8).
- (8) Sea $\Sigma = CC'$, con $C_{p\times p}$ no singular. Sea $Y = C^{-1}X$.
  Entonces $Y \sim N_p(C^{-1}\mu, C^{-1}\Sigma(C^{-1})') \equiv N_p(C^{-1}\mu, I_p)$.

  Entonces $X'BX = (CY)'B(CY) = Y'C'BCY$. Por el Resultado 2 tenemos:

  $$X'BX \sim \chi^2_k(\delta) \iff C'B \text{ idempotente}$$

  Hay que probar entonces que $C'BC$ es idempotente $\iff$ $B\Sigma$ es idempotente.

  $$B\Sigma \text{ idempotente} \iff (B\Sigma)(B\Sigma) = B\Sigma \iff (BCC')(BCC') = BCC' \iff$$

  $$\iff BCC'BCC'(C')^{-1} = BCC'(C')^{-1} \iff BCC'BC = BC \iff$$

  $$\iff (C')BCC'BC = (C')BC \iff (C'BC)'(C'BC) = C'BC \iff C'BC \text{ idempotente}$$

  Además, por el resultado 2:

  $$k = rango(C'BC) = tr(C'BC) = tr(BCC') = tr(B\Sigma) = rango(B\Sigma) = rango(B)$$

  $$\delta = (C^{-1}\mu)'(C'BC)(C^{-1}\mu) = \mu'(C')^{-1}C'BCC^{-1}\mu = \mu' B\mu$$

1.8.4. APÉNDICE: Funciones hipergeométricas generalizadas

- Definición: Se denomina función hipergeométrica generalizada de órdenes p y a a

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

  donde $(a)_k = a(a+1)\ldots(a+k-1)$, siendo $a_1, \ldots, a_p, b_1, \ldots, b_q$ parámetros (posiblemente complejos) y $z \in \mathbb{C}$ el argumento de la función.
- Algunas observaciones:
  - Ningún parámetro b_j puede ser 0 o un entero negativo (en este caso, uno de los denominadores de la serie sería 0 a partir de un cierto k).
  - Si algún parámetro en el numerador es 0 o un entero negativo, los términos de la serie se anulan a partir de un cierto k y queda un polinomio en z.
  - La serie
    - converge para todo z finito si $p \le q$.
    - converge para $|z| < 1$ y diverge para $|z| > 1$ si $p = q + 1$.
    - diverge para todo $z \ne 0$ si $p > q + 1$.
  - El término 'generalizada' se refiere a que ${}_pF_q$ es una generalización de la función hipergeométrica cñásica (o gaussiana), ${}_2F_1$.

- $_1F_1$ se denomina función hipergeométrica confluente.
- $_0F_1$, definida como
$$_0F_1(;b;z) := \lim_{a\to\infty} {}_1F_1(a;b;\frac{z}{a})$$
se denomina función hipergeométrica confluente límite.

1.9. Temario examen

- Fundamentación probabilística $\to$ No entra.
- Aspectos generales sobre vectores aleatorios:
  - Propiedad 3 de Esperanza y covarianzas si entra.
  - Normalización: $Z \sim (0_p, I_{p\times p})$ si entra.
  - Propiedades [...] $= k$ si entra.
  - Semidefinida positiva: Observación 2 si entra.
  - Teorema $\alpha' X$ de función característica si entra.
  - Caso lineal del teorema de cambio de variable si entra.
- Caso $\Sigma > 0$ entra casi entero:
  - Algunas propiedades: Caso (3) sobre el condicionamiento no entra.
  - Caracterización en términos de la densidad esférica, el resultado sobre U y Z independientes no entra.
- Caso $\Sigma \geq 0$ entra casi entero:
  - Extensión de la caracterización C-I es importante que sepamos la diferencia con el caso no general.
  - Discusión en relación con la posible no existencia de densidad en todo $\mathbb{R}^p$ si entra.
  - Resultado sobre normalidad de combinaciones lineales de vectores aleatorios con DNM no entra.
  - El Complemento del problema de Lévy-Crámer, la generalización asumiendo que es cierto para 1 si entra.
- Distribuciones esféricas y elípticas.
  - Los 3 ejemplos entran.
  - El resultado de independencia no entra.
  - Distribuciones elípticas entran sin demostraciones.
  - El ejemplo de demostraciones elípticas de $V = \Sigma$ si entra.
- Formas cuadráticas basadas en vectores aleatorios con DNM.
  - No hay que aprenderse las funciones de densidad ni características.
  - Saber calcular los momentos a partir de la característica si entra.
  - Resultado sobre χ^2 si entra.
  - Si hay que saber como se definen las distribuciones.
  - Formas cuadráticas $X'AX$ entra el resultado a).
  - La prueba del resultado 2 no entra.

- El apéndice no entra.

- Ejercicios: no va a haber ejercicios de tipo numérico, como los 7, 8 y 9 de la relación.
- Los ejercicios extra son interesantes.

  2 preguntas de teoría y 2 problemas. Entre las preguntas de teoría podemos tener varias opciones entre las cuales elegimos 2.

2. Inferencia en la Distribución Normal Multivariante

2.1. Introducción

2.1.1. “Muestra aleatoria simple” vectorial

- Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio con DNM no singular,

$$X \sim N_p(\mu, \Sigma) \qquad (\Sigma > 0)$$

(Por ejemplo, consideramos que las variables representan, en una situación práctica, p magnitudes relativas a “características” de los “individuos” de una “población”, con esa distribución de probabilidad conjunta).

- Se trata de realizar la inferencia (bajo distintas formas) sobre la distribución, supuestos desconocidos total o parcialmente a sus parámetros, a partir de una *muestra aleatoria simple* obtenida de ésta (i.e., un conjunto de “observaciones” dadas en términos de *variables aleatorias independientes e indénticamente distribuidas* con igual distribución de X), que denotaremos

$$\{X_\alpha : \alpha = 1, \dots, N\} \qquad (N = \text{ “tamaño muestral”})$$

siendo $X_{\alpha i}$ la componente $i-$ésima (correspondiente a la variable X_i) de la observación (vectorial) $\alpha-$ésima, X_α.

- Se utilizará la misma notación pero en minúsculas cuando las variables de una muestra aleatoria simple tomen valores numéricos concretos, es decir, se tenga una realización de ésta:

$$\{x_\alpha : \alpha = 1, \dots, N\} \qquad (N = \text{ “tamaño muestral”})$$

- A efectos algebraicos y computacionales, es frecuente disponer las variables componentes de la muestra en la forma de “matriz de datos” (aleatoria),

$$X = \begin{pmatrix} X_1' \\ \vdots \\ X_\alpha' \\ \vdots \\ X_p' \end{pmatrix} = \begin{pmatrix} X_{11} & \dots & X_{1i} & \dots & X_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{\alpha 1} & \dots & X_{\alpha i} & \dots & X_{\alpha p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{N1} & \dots & X_{Ni} & \dots & X_{Np} \end{pmatrix}$$

(y análogamente, usando minúsculas, en el caso de las correspondientes realizaciones de la muestra).

2.1.2. Medidas muestrales de primer y segundo orden

- En primer lugar, como paso previo a otros aspectos de la inferencia, se abordará en el apartado siguiente la *estimación puntual* de los parámetros μ y Σ. Se adoptará el criterio basado en la *maximización de la función de verosimilitud.*

  Como se verá, los *estimadores máximo-verosímiles* de μ y Σ se expresarán en términos de medidas descriptivas de la muestra dadas por el *vector de medias muestral* y la *matriz de dispersiones muestral* (respecto del vector de medias muestral).

- VECTOR DE MEDIAS MUESTRALES (COMPLETAR).
- MATRIZ DE DISPERSIONES MUESTRAL ($p \times p$)

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

$$= \begin{pmatrix} \sum_{\alpha=1}^{N}(X_{\alpha 1}-X_1)^2 & \sum_{\alpha=1}^{N}(X_{\alpha 1}-\bar{X})(X_{\alpha 2}-\bar{X}_2) & \dots & \sum_{\alpha=1}^{N}(X_{\alpha 1}-\bar{X}_1)(X_{\alpha p}-\bar{X}_p) \\ \sum_{\alpha=1}^{N}(X_{\alpha 2}-\bar{X}_2)(X_{\alpha 1}-\bar{X}_1) & \sum_{\alpha=1}^{N}(X_{\alpha 2}-\bar{X}_2)^2 & \dots & \sum_{\alpha=1}^{N}(X_{\alpha 2}-\bar{X}_2)(X_{\alpha p}-\bar{X}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{\alpha=1}^{N}(X_{\alpha p}-\bar{X}_p)(X_{\alpha 1}-\bar{X}_1) & \sum_{\alpha=1}^{N}(X_{\alpha p}-\bar{X}_p)(X_{\alpha 2}-\bar{X}_2) & \dots & \sum_{\alpha=1}^{N}(X_{\alpha p}-\bar{X}_p)^2 \end{pmatrix}$$

A partir de la matriz de dispersiones muestral, se definen:

- MATRIZ DE COVARIANZAS MUESTRAL: $S_N = \frac{A}{N} = (s_{ij})_{i,j=1,\dots,p}$.
- MATRIZ DE CUASI-COVARIANZAS MUESTRAL: $S_{N-1} = \frac{A}{N-1}$.
- MATRIZ DE CORRELACIONES MUESTRAL: $R = (\frac{s_{ij}}{s_{ii}^{\frac{1}{2}} s_{jj}^{\frac{1}{2}}})_{i,j=1,\dots,p}$

  (puede escribirse como $R = D^{-\frac{1}{2}} S_N D^{-\frac{1}{2}}$ con $D = \text{dia}(s_{11}, \dots, s_{pp})$).

2.1.3. Fórmula de momentos multivariante muestral

- LEMA: Sea $\{X_\alpha : \alpha = 1, \dots, N\}$ una muestra aleatoria de una distribución p-dimensional, y sea $\bar{X}$ el vector de medias muestral. Entonces, para cualquier vector $b \in \mathbb{R}^p$, se verifica que

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

$$= A + N(\bar{X} - b)(\bar{X} - b)' \qquad (1)$$

En particular, se tiene:

- Con $b = \mu$ (vector de medias poblacional, si existe)

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

- Con $b = 0$

$$\sum_{\alpha=1}^{N} X_\alpha X_\alpha' = A + N\bar{X}\bar{X}'$$

  ('fórmula de momentos multivariante muestral')

- (1)

$$\sum_{\alpha=1}^{N}(X_\alpha - b)(X_\alpha - b)' = \sum_{\alpha=1}^{N}[(X_\alpha - \bar{X}) + (\bar{X} - b)][(X_\alpha - \bar{X}) + (\bar{X} - b)]' =$$

$$= \sum_{\alpha=1}^{n}[(X_\alpha - \bar{X}) + (\bar{X} - b)][(X_\alpha - \bar{X})' + (\bar{X} - b)'] =$$

$$= \underbrace{\sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(X_\alpha - \bar{X})'}_{A} + \underbrace{\sum_{\alpha=1}^{N}(\bar{X} - b)(\bar{X} - b)'}_{B} + \underbrace{\sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(\bar{X} - b)'}_{C} + \underbrace{\sum_{\alpha=1}^{N}(\bar{X} - b)(X_\alpha - \bar{X})'}_{D} = \star$$

$$B \Rightarrow \sum_{\alpha=1}^{N}(\bar{X} - b)(\bar{X} - b)' = N(\bar{X} - b)(\bar{X} - b)'$$

$$C \Rightarrow \sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(\bar{X} - b)' = \sum_{\alpha=1}^{N} X_\alpha \bar{X}' - \sum_{\alpha=1}^{N} X_\alpha b' - \sum_{\alpha=1}^{N} \bar{X}\bar{X}' + \sum_{\alpha=1}^{N} \bar{X} b' =$$

$$= N\bar{X}\bar{X}^{-1} - N\bar{X}b' - N\bar{X}\bar{X}^{-1} + N\bar{X}b^{-1} = 0$$

$$D \Rightarrow \sum_{\alpha=1}^{N}(\bar{X}-b)(X_\alpha - \bar{X})' = \sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(\bar{X}-b)' = 0$$

$$\star = A + N(\bar{X}-b)(\bar{X}-b)'$$

2.2. Inferencia en la DNM

2.2.1. Función de verosimilitud

- Sea $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$, y sea $\{X_\alpha : \alpha = 1, \ldots, N\}$ una muestra aleatoria simple de dicha distribución.

  Para una realización dada de la muestra $\{x_\alpha : \alpha = 1, \ldots, N\}$, la *función de verosimilitud* se expresa como la función (de argumentos μ y Σ).

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

  $$= \prod_{\alpha=1}^{N} \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x_\alpha - \mu)'\Sigma^{-1}(x_\alpha - \mu)\} =$$

  $$= \frac{1}{(2\pi)^{\frac{pN}{2}}|\Sigma|^{\frac{N}{2}}} \exp\{-\frac{1}{2}\sum_{\alpha=1}^{N}(x_\alpha - \mu)'\Sigma^{-1}(x_\alpha - \mu)\}$$

  De forma abreviada, se denotará simplemente $L(\mu, \Sigma)$, sobreentendiéndose implícita la realización muestral de referencia.

- Mediante operaciones algebraicas, y usando la fórmula de momentos multivariante generalizada (caso $b = \mu$), se comprueba que la función de verosimilitud puede expresarse de la forma

  $$L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{pN}{2}}|\Sigma|^{\frac{N}{2}}} \exp\{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}\bar{A}) - \frac{N}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)\}$$

  con $\bar{A} = \sum_{\alpha=1}^{N}(x_\alpha - \bar{x})(x_\alpha - \bar{x})'$ (es decir, la matriz de dispersiones muestral evaluada en la realización dada de la muestra, $\{x_\alpha : \alpha = 1, \ldots, N\}$) (1).

  Más convenientemente, a efectos de maximización, se suele usar la transformación

  $$\ln(L(\mu, \Sigma)) = -\frac{pN}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\operatorname{tr}(\Sigma^{-1}\bar{A}) - \frac{N}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

  (Es habitual proceder, equivalentemente, a la minimización de la función $-\ln(L(\mu, \Sigma))$, no negativa).

- (1)

  $$L(\mu, \Sigma; x_1, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{pN}{2}}|\Sigma|^{\frac{N}{2}}} \exp\{-\frac{1}{2}\sum_{\alpha=1}^{N}(x_\alpha - \mu)'\Sigma^{-1}(x_\alpha - \mu)\}$$

  $$\sum_{\alpha=1}^{N}(x_\alpha - \mu)'\Sigma^{-1}(x_\alpha - \mu) = \sum_{\alpha=1}^{N}[(x_\alpha - \bar{x}) + (\bar{X} - \mu)]'\Sigma^{-1}[(x_\alpha - \bar{x}) + (\bar{x} - \mu)] =$$

  $$= \sum_{\alpha=1}^{N}[(x_\alpha - \bar{x})' + (\bar{X} - \mu)']\Sigma^{-1}[(x_\alpha - \bar{x}) + (\bar{x} - \mu)] =$$

$$= \sum_{\alpha=1}^{N}(x_\alpha - \bar{X})'\Sigma^{-1}(x_\alpha - \mu) + \sum_{\alpha=1}^{N}(x_\alpha - \bar{x})'\Sigma^{-1}(\bar{x} - \mu) +$$

$$+ \sum_{\alpha=1}^{N}(\bar{x} - \mu)\Sigma^{-1}(x_\alpha - \bar{x}) + \sum_{\alpha=1}^{N}(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) =$$

$$= \sum_{\alpha=1}^{N} \text{traza}(x_\alpha - \bar{x})'\Sigma^{-1}(x_\alpha - \mu) + N(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) =$$

$$= \sum_{\alpha=1}^{N} \text{traza}[\Sigma^{-1}(x_\alpha - \mu)(x_\alpha - \bar{x})'] + N(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) =$$

$$= \text{traza}[\sum_{\alpha=1}^{N} \Sigma^{-1}(x_\alpha - \mu)(x_\alpha - \bar{x}'] + N(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) =$$

$$= \text{traza}[\Sigma^{-1} \sum_{\alpha=1}^{N}(x_\alpha - \mu)(x_\alpha - \bar{x})'] + N(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) =$$

$$= \text{traza}(\Sigma^{-1}\bar{A}) + N(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)$$

$$L(\mu, \Sigma; x_1, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{pN}{2}}|\Sigma|^{\frac{N}{2}}} \exp\{-\frac{1}{2}\,\text{traza}(\Sigma^{-1}\bar{A}) - \frac{N}{2}(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)\}$$

2.2.2. Estimadores máximo-verosímiles de μ y Σ

- RESULTADO: Sea $X \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$, y sea $\{X_\alpha : \alpha = 1, \ldots, N\}$ una muestra aleatoria simple de dicha distribución. Entonces, los *estimadores máximo-verosímiles* (EMV) de μ y Σ son, respectivamente,

  $$\bar{X} \qquad \text{y} \qquad \frac{A}{N} = S_N$$

  este último bajo la condición de ser A definida positiva $(*)$.

  [$(*)$ Este aspecto se discutirá más adelante, tras la demostración, en relación con el 'Teorema de Dykstra']

- El problema de optimización se podría abordar de forma directa mediante derivación matricial. No obstante, es interesante probar el resultado por el procedimiento que se expondrá a continuación.

- Para la demostración, se utilizará el siguiente resultado auxiliar en la parte relativa al estimador del parámetro Σ:
  LEMA (Watson): Sea

  $$f(G) = -N\ln(|G|) - \text{tr}(G^{-1}D)$$

  con argumento de G una matriz simétrica definida positiva, y siendo D una matriz simétrica definida positiva dada, ambas $p \times p$. Entonces, existe (y es único) el máximo de f respecto a G, y se alcanza en $G = \frac{1}{N}D$, siendo el valor máximo alcanzado

  $$f(\frac{1}{N}D) = pN\ln(N) - N\ln(|D|) - pN \qquad (2)$$

- (2) (No examen) Por ser D simétrica y definida positiva, podemos representar $D = EE'$ con E matriz no singular. Sea $H = E'G^{-1}E \Rightarrow G^{-1} = (E')^{-1}HE^{-1} \Rightarrow G = EH^{-1}E'$. Se tiene que

  $$|G| = |EH^{-1}E'| = |E||H^1||E'| = |H^{-1}||EE'| = \frac{|D|}{|H|}$$

$$\text{traza}(G^{-1}D) = \text{traza}(G^{-1}EE') = \text{traza}(E'G^{-1}E) = \text{traza}(H)$$

Podemos escribir:

$$f(G) = g(H) = -N\ln(|D|) + N\ln(|H|) - \text{traza}(H)$$

Con H en el espacio de matrices simétricas y definidas positivas $(p \times p)$.

Consideramos la descomposición de Cholesky de H:

$$H = T'T$$

con T matriz triangular superior con elementos diagonales estrictamente positivos.
Escribimos

$$f(G) = g(H) = h(T) = -N\ln(|D|) + N\ln(|T|^2) - \text{traza}(T'T)$$

Con T en el espacio de matrices triangulares superiores con elementos diagonales estrictamente positivos.

$$|T|^2 = \sum_{i=1}^{p} t_{ii}^2 \Rightarrow \ln(|T|^2) = \sum_{i=1}^{p} \ln(t_{ii}^2) \Rightarrow N\ln(|T|^2) = N\sum_{i=1}^{p} \ln(t_{ii}^2) = \sum_{i=1}^{p} N\ln(t_{ii}^2)$$

$$\text{traza}(T'T) = \sum_{i=1}^{p} t_{ii}^2 + \sum_{\substack{i,j=1 \\ i<j}}^{p} t_{ij}^2$$

$$T'T = \begin{pmatrix} t_{11} & 0 & \dots & 0 \\ t_{12} & t_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ t_{1p} & t_{2p} & \dots & t_{pp} \end{pmatrix} \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1p} \\ 0 & t_{22} & \dots & t_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t_{pp} \end{pmatrix} = \begin{pmatrix} t_{11}^2 & * & \dots & * \\ * & t_{12}^2 + t_{22}^2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & t_{1p}^2 + t_{2p}^2 + \dots + t_{pp}^2 \end{pmatrix}$$

Entonces:

$$h(T) = -N\ln(|D|) + \sum_{i=1}^{p} [N\ln(t_{ii}^2) - t_{ii}^2] - \sum_{\substack{i,j=1 \\ i<j}}^{p} t_{ij}^2$$

Alcanza el máximo para $t_{ij} = 0, \forall i < j \Rightarrow T$ diagonal.

$$y = N\ln(x) - x \Rightarrow y' = \frac{N}{x} - 1 \Rightarrow \frac{N}{x} - 1 = 0 \Rightarrow x = N \Rightarrow y'' = \frac{-N}{x^2} < 0 \Rightarrow \text{máximo}$$

Para que $h(T)$ alcance el máximo, elegimos $t_{ii}^2 = N, \forall i \Rightarrow t_{ii} = \sqrt{N}$.

Entonces tenemos:

$$T = \sqrt{N} I_p \Rightarrow H = T'T = NI_p$$

El máximo de $f(G)$ se obtiene para:

$$G = E'H^1E' = E(NI_p)^{-1}E' = \frac{1}{N}EI_pE' = \frac{1}{N}EE' = \frac{1}{N}D$$

- Demostración:
  - Maximización de $\ln(L(\mu, \Sigma))$ en μ: (Se puede hacer, en este caso, independientemente de Σ).
    De la forma obtenida para $\ln(L(\mu, \Sigma))$, se desprende que, independientemente del valor de Σ, el máximo se alcanzará donde se minimice la forma cuadrática

    $$(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)$$

    Ahora bien, dado que Σ^{-1} es definida positiva (por serlo Σ), la forma cuadrática alcanza el valor mínimo 0 para (y sólo para) $\bar{x} - \mu = 0$, es decir, $\mu = \bar{x}$. Por tanto, el estadístico definido por:

    $$\hat{\mu} := \bar{X} \qquad \text{(el vector de medias muestral)}$$

    es el EMV (único, c.s.) de μ.

- Maximización de $\ln(L(\bar{x},\Sigma))$ en Σ:
  Dado que tratamos de maximizar la función

$$\ln(L(\bar{x},\Sigma)) = -\frac{pN}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\operatorname{tr}(\Sigma^{-1}\bar{A})$$

  (donde ya ha desaparecido el término correspondiente a la forma cuadrática), identificamos, en la notación del 'lema de Watson',

$$G := \Sigma \qquad D := \bar{A}$$

$$f(\Sigma) := 2[\ln L(\bar{x},\Sigma) + \frac{pN}{2}\ln(2\pi)] = pN\ln(2\pi) + 2\ln(L(\bar{x},\Sigma)) = -N\ln(|\Sigma|) - \operatorname{tr}(\Sigma^{-1}\bar{A})$$

  por lo que el máximo de f en Σ (igualmente, entonces, el máximo de $L(\bar{x},\Sigma)$ en Σ) se alcanza para (y sólo para) $\Sigma = \frac{\bar{A}}{N} =: \bar{S}_N$. Por tanto, el estadístico definido por

$$\hat{\Sigma} := \frac{A}{N} = S_N \qquad \text{(la matriz de covarianzas muestral)}$$

  es el EMV (único, c.s.) de Σ.

- Se comprueba que el valor máximo alcanzado por la función de verosimilitud en el punto $(\mu,\Sigma) = (\bar{x},\bar{S}_N)$ del espacio paramétrico es:

$$L(\mu,\Sigma) = \frac{1}{(2\pi e)^{\frac{pN}{2}}|\bar{S}_N|^{\frac{N}{2}}} = [\frac{N}{(2\pi e)|\bar{A}|^{\frac{1}{p}}}]^{\frac{pN}{2}} \qquad (3)$$

  Como puede verse [...]

2.2.3. Teorema de Dykstra

- En relación con el resultado anterior sobre el EMV del par paramétrico (μ,Σ), se plantea la cuestión siguiente:

  Formalmente, puede ocurrir que para una realización concreta de la muestra aleatoria la matriz $\bar{A}$ resulte ser singular (es decir, semidefinada positiva). Ahora bien, ¿con qué probabilidad puede darse este caso?

- La respuesta viene dada por el 'teorema de Dykstra' (1970):
  TEOREMA (Dykstra): Sea $X \sim N_p(\mu,\Sigma)$ $(\Sigma > 0)$. Sea $\{X_\alpha : \alpha = 1,\dots,N\}$ una muestra aleatoria simple de dicha distribución y sea $A = \sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ la correspondiente matriz de dispersiones muestral. Entonces, A es definida positiva, con probabilidad 1, si y sólo si $N > p$.

- Para la demostración del teorema, se establecen algunos resultados previos:
  LEMA 1: Sean $\{X_\alpha : \alpha = 1,\dots,N\}$ vectores (aleatorios o no) de dimensión p. Sea $C = (c_{\alpha\beta})_{\alpha,\beta=1,\dots,N}$ una matriz ortogonal de dimensión $N \times N$. Entonces, definiendo

$$Y_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}X_\beta, \qquad \alpha = 1,\dots,N$$

  se tiene que

$$\sum_{\alpha=1}^{n} X_\alpha X_\alpha' = \sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' \qquad (4)$$

- (4)

$$\sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' = \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} c_{\alpha\beta}X_\beta)(\sum_{\gamma=1}^{N} c_{\alpha\gamma}X_\gamma)' = \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\sum_{\gamma=1}^{N} c_{\alpha\beta}c_{\alpha\gamma}X_\beta X_\gamma' =$$

$$= \sum_{\beta=1}^{N}\sum_{\gamma=1}^{N} X_\beta X'_\gamma (\sum_{\alpha=1}^{N} c_{\alpha\beta}c_{\alpha\gamma}) =$$

Como tenemos $(\sum_{\alpha=1}^{N} c_{\alpha\beta}c_{\alpha\gamma}) =< c_\beta, c_\gamma >= \delta_{\beta\gamma} = \begin{cases} 0 & \text{si } \beta \neq \gamma \\ 1 & \text{si } \beta = \gamma \end{cases}$, entonces:

$$= \sum_{\beta=1}^{N} X_\beta X'_\beta$$

- LEMA 2: Sean $\{X_\alpha : \alpha = 1, \dots, N\}$ vectores aleatorios de dimensión p, con $X_\alpha \sim N_p(\mu_\alpha, \Sigma), \alpha = 1, \dots, N$ (es decir, tienen la misma matriz de covarianzas, aunque posiblemente distinto vector de medias), independientes. Sea $C = (c_{\alpha\beta})_{\alpha,\beta=1,\dots,N}$ una matriz ortogonal de dimensión $N \times N$. Entonces, definiendo
$$Y_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}X_\beta, \qquad \alpha = 1, \dots, N$$
se tiene
  - $Y_\alpha \sim N_p(\mathcal{V}_\alpha, \Sigma)$, con $\mathcal{V}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}\mu_\beta, \alpha = 1, \dots, N$ (5).
  - Los vectores $\{Y_\alpha : \alpha = 1, \dots, N\}$ son independientes (6).
- (5)
$$c_{\alpha\beta}X_\beta = C_{\alpha\beta}I_pX_\beta = \begin{pmatrix} c_{\alpha\beta} & 0 & \dots & 0 \\ 0 & c_{\alpha\beta} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & c_{\alpha\beta} \end{pmatrix} X_\beta = A_{\alpha\beta}X_\beta$$
$$Y_\alpha = \sum_{\beta=1}^{N} A_{\alpha\beta}X_\beta \sim N_p(\sum_{\beta=1}^{N}(A_{\alpha\beta}\mu_\beta), \sum_{\beta=1}^{N} A_{\alpha\beta}\Sigma A'_{\alpha\beta}) \equiv N_p(\sum_{\alpha=1}^{N} c_{\alpha\beta}\mu_\beta, \sum_{\beta=1}^{N} c_{\alpha\beta}^2\Sigma) \equiv N_p(\mathcal{V}_\alpha, \Sigma)$$
donde hemos usado que $\sum_{\beta=1}^{N} c_{\alpha\beta}c_{\alpha\beta}^2 = 1$.

  Entonces
$$\Phi_{Y_\alpha}(t) = e^{it'\mathcal{V}_\alpha - \frac{1}{2}t'\Sigma t}$$
- (6) Consideramos el vector $(pN) \times 1$:
$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}$$
Para cada $t \in \mathbb{R}^{pN}$, tenemos:
$$t = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix}$$
con $t_\alpha \in \mathbb{R}, \alpha = 1, \dots, N$.

  La función característica de Y se calcula como:
$$\Phi_{Y_1,\dots,Y_N}(t) = E[e^{it'Y}] = E[e^{i\sum_{\alpha=1}^{N} t'_\alpha Y_\alpha}] = E[e^{i\sum_{i=1}^{N} t'_\alpha(\sum_{\beta=1}^{N} c_{\alpha\beta}X_\beta)}] =$$

$$= E[e^{i\sum_{\beta=1}^N(\sum_{\alpha=1}^N t'_\alpha c_{\alpha\beta})X_\beta)}] = E[\prod_{\beta=1}^N e^{i(\sum_{\alpha=1}^N t'_\alpha c_{\alpha\beta}X_\beta}] =$$

Como $X_1, \ldots, X_N$ son independientes, tenemos que la esperanza del producto es el producto de las esperanzas:

$$= \prod_{\beta=1}^N E[e^{i(\sum_{\alpha=1}^N t'_\alpha c_{\alpha\beta})X_\beta}] = \prod_{\beta=1}^N \Phi_{X_\beta}(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta}) =$$

Como $X_\beta \sim N(\mu_\beta, \Sigma), \beta = 1, \ldots, N$, tenemos:

$$= \prod_{\beta=1}^N e^{i(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta})'\mu_\beta - \frac{1}{2}(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta})'\Sigma(\sum_{\gamma=1}^N t_\gamma c_{\gamma\beta})} =$$

$$= e^{i\sum_{\beta=1}^N(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta})'\mu_\beta - \frac{1}{2}\sum_{\beta=1}^N[(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta})'\Sigma(\sum_{\gamma=1}^N t_\gamma c_{\gamma\beta})]} =$$
$$= e^{i\sum_{\alpha=1}^N t'_\alpha(\sum_{\beta=1}^N c_{\alpha\beta}\mu_\beta) - \frac{1}{2}\sum_{\alpha=1}^N\sum_{\gamma=1}^N[t'_\alpha\Sigma t_\gamma(\sum_{\beta=1}^N c_{\alpha\beta}c_{\gamma\beta})]} =$$

donde tenemos que $\sum_{\beta=1}^N c_{\alpha\beta}c_{\gamma\beta} = \delta_{\alpha\gamma}$.

$$= e^{i\sum_{\alpha=1}^N t'_\alpha \mathcal{V}_\alpha - \frac{1}{2}\sum_{\alpha=1}^N t'_\alpha\Sigma t_\alpha} = \prod_{\alpha=1}^N e^{it'_\alpha\mathcal{V}_\alpha - \frac{1}{2}t'_\alpha\Sigma t_\alpha} = \prod_{\alpha=1}^N \Phi_{Y_\alpha}(t_\alpha)$$

$\Rightarrow Y_1, \ldots, Y_N$ independientes.

- Demostración (esquema): Consideremos, en nuestro caso, una matriz $B = (b_{\alpha\beta})_{\alpha,\beta=1,\ldots,N}$ de dimensión $N \times N$, ortogonal, cuya última fila sea $(\frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}})$. Es decir, el resto de filas han de ser vectores ortogonales con éste y entre sí, con norma igual a 1. Definimos el vector:

$$Z_\alpha = \sum_{\beta=1}^N b_{\alpha\beta}X_\beta, \qquad \alpha = 1, \ldots, N$$

Por los resultados anteriores (lemas 1 y 2), tenemos que:

- $Z_\alpha \sim N_p(\mathcal{V}_\alpha, \Sigma)$, con:

$$\mathcal{V}_\alpha = \sum_{\beta=1}^N b_{\alpha\beta}\mu = (\sum_{\beta=1}^N b_{\alpha\beta})\mu = \begin{cases} 0, & \text{si } \alpha = 1, \ldots, N-1, \\ \frac{N}{\sqrt{N}}\mu = \sqrt{N}\mu, & \text{si } \alpha = N \end{cases}$$

- Los vectores $\{Z_\alpha : \alpha = 1, \ldots, N\}$ son independientes.
- $Z_m = \sum_{\beta=1}^N \frac{1}{\sqrt{N}}X_\beta = \sqrt{N}(\frac{1}{N}\sum_{\beta=1}^N X_\beta) = \sqrt{N}\bar{X}$.
(es decir, $\bar{X} = \frac{1}{\sqrt{N}}Z_N$).
- $A = \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})' = \sum_{\alpha=1}^N X_\alpha X'_\alpha - N\bar{X}\bar{X}' =$
$= \sum_{\alpha=1}^N Z_\alpha Z'_\alpha - Z_N Z'_N = \sum_{\alpha=1}^{N-1} Z_\alpha Z'_\alpha$.
- Como consecuencia, $\bar{X}$ y A son independientes.

Por último, observamos que podemos escribir

$$A = Z'Z$$

con $Z = \begin{pmatrix} Z'_1 \\ \vdots \\ Z'_{N-1} \end{pmatrix}$, matriz $(N-1) \times p$, $Z' = \begin{pmatrix} Z_1 & \ldots & Z_{N-1} \end{pmatrix}$, matriz $p \times (N-1)$.

Esto equivale a que la matriz A es definida no negativa.

Por otra parte, se cumple que

$$\text{rango}(Z) = \text{rango}(Z') = \text{rango}(Z'Z) = \text{rango}(A)$$

Por tanto, la demostración quedará concluida probando que rango$(Z) = p$ con probabilidad 1 si y sólo si $N > p$ (6).

- (6)

  $\Rightarrow$ Si fuese $N \leq p$, sería $N - 1 < p$. Es decir, el rango de Z sería como máximo igual a $N - 1 < p$.

  $\Leftarrow$ Supongamos que $N > p$. Vamos a probar que, entonces, rango$(Z) =$ rango$(A) = p$ con probabilidad 1.

  Probamos primero para $N = p+1$, y después se argumentará por inducción para cualquier $N > p$.

  $N = p + 1 \Rightarrow N - 1 = p \Rightarrow Z' = (Z_1, \ldots, Z_p)_{p\times p}$ y $Z_\alpha \sim N_p(0, \Sigma), \alpha = 1, \ldots, N - 1$ independientes.

  - NOTA: Para cada $\alpha = 1, \ldots, N - 1 = p$, se tiene que

    $$P[Z_\alpha \in S] = 0, \forall S \in \mathcal{B}^p \text{ tal que } \underbrace{\lambda_p}_{\text{medida de Lebesgue en } (\mathbb{R}^p, \mathcal{B}^p)} (S) = 0$$

    $$P_{Z_\alpha} \underbrace{<<}_{\text{“absolutamente continua”}} \lambda_p$$

    En particular, es cierto para cualquier hiperplano $(p - 1)$-dimensional

    $$S(a_1, \ldots, a_{p-1})$$

    engendrado por los vectores $a_1, \ldots, a_{p-1} \in \mathbb{R}^p$.

  Vamos a probar que la matriz $Z' = \big(Z_1, \ldots, Z_{p=N-1}\big)$ tiene rango p.

  $$P[\text{rango}(Z) < p] = P[Z_1, \ldots, Z_p \text{ son linealmente dependientes}] =$$

  $$= P[\{\omega \in \Omega : Z_1(\omega), \ldots, Z_p(\omega) \text{ son linealmente dependientes}\}] \leq$$

  $$\leq \sum_{i=1}^{p} P[Z_i \in S(Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_p)] = pP[Z_1 \in S(Z_2, \ldots, Z_p)] =$$

  $$= p \cdot 0 = 0$$

  $S(Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_p)$ hiperplano generado por $Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_p$.
  Falta probar que $P[Z_1 \in S(Z_2, \ldots, Z_p)] = 0$.

  - NOTA: En general,

    $$P[A] = E_p[I_A(\cdot)] = \int_\Omega I_A(\omega)P(dw) = \int_A P(dw)$$

  Para ello:

  $$P[Z_1 \in S(Z_2, \ldots, Z_p)] = E_p[I_{[z_1 \in S(z_2,\ldots,z_p)]}(\cdot)] =$$

  $$= \int_{\mathbb{R}^p \times \cdots \times \mathbb{R}^p} I_{[z_1 \in S(z_2,\ldots,z_p)]}(z_1, \ldots, z_p) P_{Z_1 \ldots Z_p}(dz_1, \ldots, dz_p) =$$

  - Como son independientes

    $$P_{Z_1,\ldots,Z_p}(dz_1 \ldots dz_p) = P_{Z_1}(dz_1)P_{Z_2,\ldots,Z_p}(dz_2 \ldots dz_p)$$

$$= \int_{\mathbb{R}^p\times\cdots\times\mathbb{R}^p} \left(\int_{\mathbb{R}^p} I_{[z_1\in S(z_2,\ldots,z_p)}(z_1,\ldots,z_p) P_{Z_1}(dz_1) \right) P_{z_2,\ldots,z_p}(dz_2,\ldots,dz_p) =$$

$$= \int_{\mathbb{R}^p\times\cdots\times\mathbb{R}^p} (P_{Z_1}[Z_1 \in S(z_2,\ldots,z_p)]) P_{z_2,\ldots,z_p}(dz_2\ldots dz_p) =$$

$$= \int_{\mathbb{R}^p\times\cdots\times\mathbb{R}^p} 0 P_{z_2,\ldots,z_p}(dz_2\ldots dz_p) = 0$$

Finalmente, veamos la inducción. Supongamos que la implicación se cumple para cierto $N > p$ y vamos a ver que entonces se cumple para $N' = N+1$.

Definimos $Z^{[N]}$ como la matriz Z definida previamente asociada al conjunto $X_1,\ldots,X_N$, y $Z^{[N']}$ como la matriz asociada al conjunto $X_1,\ldots,X_N,X_{N'}$.

$$B^{[N]} = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N-1\ 1} & b_{N-1\ 2} & \ldots & b_{N-1\ N} \\ \frac{1}{\sqrt{N}} & \frac{1}{\sqrt{N}} & \ldots & \frac{1}{\sqrt{N}} \end{pmatrix}$$

$$B^{[N']} = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1N} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{N-1\ 1} & b_{N-1\ 2} & \ldots & b_{N-1\ N} & 0 \\ b_{N1} & b_{N2} & \ldots & b_{NN} & b_{NN+1} \\ \frac{1}{\sqrt{N}} & \frac{1}{\sqrt{N}} & \ldots & \frac{1}{\sqrt{N}} & \frac{1}{\sqrt{N}} \end{pmatrix}$$

$$Z^{[N]} = \begin{pmatrix} Z_1^{[N]'} \\ \vdots \\ Z_{N-1}^{[N]'} \end{pmatrix}$$

$$Z^{[N']} = \begin{pmatrix} Z_1^{[N']'} = Z_1^{[N]'} \\ \vdots \\ Z_{N-1}^{[N']'} = Z_{N-1}^{[N]'} \\ Z_N^{[N']'} \end{pmatrix}$$

2.2.4. Teorema de Fisher multivariante

De la demostración del teorema de Dykstra, se tiene también la prueba del siguiente resultado, conocido como 'teorema de Fisher multivariante':

TEOREMA (Fihser): Dada una muestra aleatoria simple $X_1,\ldots,X_N$ de distribución $N_p(\mu,\Sigma)$, el vector de medias muestral $\bar{X}$ se distribuye según una $N_p(\mu,\frac{\Sigma}{N})$, y la matriz de dispersiones muestral de igual modo que $\sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha'$, con $Z_1,\ldots,Z_{N-1}$ vectores aleatorios independientes e idénticamente distribuidos según una $N_p(0,\Sigma)$, siendo el vector $\bar{X}$ y la matriz A independientes.

2.2.5. Propiedades de los EMV $\hat{\mu}$ y $\hat{\Sigma}$

En este apartado se trata, a continuación, sobre el posible cumplimiento de algunas propiedades importantes por los EMV $\hat{\mu} = \hat{X}$ y $\hat{\Sigma} = S_N$. Concretamente:

- Insesgadez.
- Consistencia (fuerte, débil).
- Eficiencia.

2.2.6. Propiedades de los EMV $\hat{\mu}$ y $\hat{\Sigma}$ Sobre INSESGADEZ

DEFINICIÓN: Sea $\{X_\alpha : \alpha = 1, \ldots, N\}$ una muestra aleatoria simple de una distribución dependiente de un parámetro (en general, multidimensional), $\theta = (\theta_1, \ldots, \theta_k)' \in \Theta \subseteq \mathbb{R}^k$. Un estimador $\hat{\theta}$ (i. e., función medible de la muestra, $\hat{\theta} = T(X)$) del parámetro θ se dice insesgado si

$$E_\theta[\hat{\theta} - \theta] = 0, \quad \forall \theta \in \Theta \quad \text{(es decir, } E_\theta[\hat{\theta}] = \theta, \quad \forall \theta \in \Theta)$$

Sea $\mathbf{X} \sim N_p(\mu, \Sigma) \quad (\Sigma > 0)$

- VECTOR DE MEDIAS MUESTRAL:
  $\hat{\mu} := \bar{X}$ es un estimador insesgado de μ

  (En realidad, en este caso no hace falta suponer normalidad, solo la existencia de μ)

- MATRIZ DE COVARIANZAS MUESTRAL: $\hat{\Sigma} := S_N$ no es un estimador insesgado de Σ

  (Tener en cuenta, por el teorema de Fisher, que $A \stackrel{d}{=} \sum_{\alpha=1}^{N-1} Z_\alpha Z'_\alpha$, con $Z_\alpha \sim N_p(0, \Sigma), \alpha = 1, \ldots, N-1$, independientes)

  Se tiene, en este caso, que:
  $S_{N-1} = \frac{N}{N-1} S_N$ sí es un esimador insesgado de Σ

2.2.7. Propiedades de los EMV $\hat{\mu}$ y $\hat{\Sigma}$: Sobre CONSISTENCIA

DEFINICIÓN: Sea $\{X_\alpha : \alpha = 1, \ldots, N\}$ una muestra aleatoria simple (de tamaño N, considerado variable) de una distribución dependiente de un parámetro (en general, multidimensional), $\theta(\theta_1, \ldots, \theta_k)' \in \Theta \subseteq \mathbb{R}^k$. Un estimador $\hat{\theta}_N$ i. e., función medible de la muestra, $\hat{\theta}_N = T_N(X)$) del parámetro θ se dice

(a) débilmente consistente si $\hat{\theta}_N$ converge en probabilidad a θ, es decir,

$$\forall \epsilon > 0, \quad \lim_{N \to \infty} P_\theta[\|\hat{\theta}_N - \theta\| < \epsilon] = 1$$

(b) fuertemente consistente si $\hat{\theta}_N$ converge casi seguramente a θ, es decir,

$$P_\theta[\lim_{N \to \infty} \hat{\theta}_N = \theta] = 1$$

Sea $X \sim N_p(\mu, \Sigma) \quad (\Sigma > 0)$. Se demuestra que:
$\hat{\mu} := \bar{X}$ $\hat{\Sigma} := S_N$ son estimadores fuertemente consistentes de $\boldsymbol{\mu}$ y Σ, respectivamente.

(En el caso de $\hat{\boldsymbol{\mu}} := \overline{X}$, no se requiere normalidad. S_{N-1} también es un estimador fuertemente consistente de Σ) [Se omite la demostración]

2.2.8. Propiedades de los EMV $\hat{\mu}$ y $\hat{\Sigma}$: Sobre EFICIENCIA

(Para estimadores insesgados)

DEFINICIÓN: Sea $T = (T_1, \ldots, T_k)'$ un estimador insesgado de un parámetro (en general, multidimensional) $\theta = (\theta_1, \ldots, \theta_k)' \in \bar{\Theta}$, siendo $\Theta = \mathbb{R}^k$ o un 'rectángulo' en $\mathbb{R}^k$. Se dice que T es eficiente para $\theta = (\theta_1, \ldots, \theta_k)'$ si, para cualquier otro estimador insesgado U de θ, se verifica que la diferencia de matrices de covarianzas $\overline{\mathrm{Cov}_\theta(U)} - \mathrm{Cov}_\theta(T)$, para todo $\theta \in \Theta$, es una matriz definida no negativa [se suele denotar $\mathrm{Cov}_\theta(U) \geq \mathrm{Cov}_\theta(T)$]

Sea $X \sim N_p(\mu, \Sigma) \quad (\Sigma > 0)$. Se demuestra que:

- $\bar{X}$ es un estimador eficiente de μ en el espacio $\mathbb{R}^p$
- S_{N-1} es un estimador eficiente de Σ en el espacio de matrices simétricas definidas positivas de dimensión $p \times p$.

2.2.9. Teorema de Zehna: INVARIANCIA de los estimadores MV

TEOREMA (Zehna, 1966): Sea $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ una familia de distribuciones de probabilidad sobre el espacio $(\mathbb{R}^p, \mathcal{B}^p)$. Sea $g : \Theta \to \Lambda$ una función arbitraria dada. Si $\hat{\theta}$ es un estimador máximo-verosímil de θ, entonces $g(\hat{\theta})$ es un estimador máximo-verosímil de $g(\theta)$

[OBSERVACIÓN: Este teorema hace uso del concepto de 'función de verosimilitud inducida por g : Para cada $\lambda \in g(\Theta) \subseteq \Lambda$, se define

$$M(\lambda) = \sup_{\theta \in \Theta_\lambda} L(\theta), \qquad \text{con } \Theta_\lambda = \{\theta \in \Theta : g(\theta) = \lambda\}$$

Se entiende entonces que '$g(\hat{\theta})$ es un EMV de $g(\theta)$' en el sentido de que $g(\hat{\theta})$ es un máximo para M

En el contexto de la inferencia sobre la DNM, se tiene que para distintos coeficientes que dependen funcionalmente de μ y Σ (por ejemplo, coeficientes de correlación, regresión, etc.) los correspondientes estimadores máximo-verosímiles se obtienen mediante sustitución de μ y Σ por $\overline{X}$ y S_N, respectivamente, en las expresiones que definen dichos coeficientes

2.2.10. [Complemento] Distribuciones asintóticas de $\hat{\mu}$ y $\hat{\Sigma}$, con distribución de referencia no necesariamente normal

- MOTIVACIÓN (referida al vector de medias muestral):

  Recordemos que, al enunciar y probar el teorema de Fisher, se ha obtenido que para una muestra aleatoria simple $X_1, \ldots, X_N$ de una distribución $N_p(\mu, \Sigma)$ se tiene que

  $$\overline{X} \sim N_p(\mu, \frac{\Sigma}{N});$$

  es decir, normalizando en media y tamaño muestral,

  $$N^{\frac{1}{2}}(\overline{X} - \mu) = N^{-\frac{1}{2}} \sum_{\alpha=1}^{N} (X_\alpha - \mu) \sim N_p(0, \Sigma)$$

  (la distribución límite ya no depende de N).

- ¿Qué puede afirmarse, al menos como aproximación límite (para $N \to \infty$), cuando la distribución de origen no es necesariamente una DNM?

- Distribución asintótica del vector de medias muestral, $\overline{X}$.

  RESULTADO: Sea $X_1, \ldots, X_N, \ldots$ una sucesión de vectores aleatorios p-dimensionales independientes e idénticamente distribuidos, con vector de medias μ y matriz de covarianzas Σ (i. e., $X_\alpha \sim (\mu, \Sigma), \alpha = 1, \ldots, N, \ldots$). Sea $\overline{X}_N = \frac{1}{N} \sum_{\alpha=1}^{N} X_\alpha$, $\forall N \geq 1$. Entonces, cuando $N \to \infty$, se tiene la distribución asintótica

  $$N^{\frac{1}{2}}(\overline{X}_N - \mu) = N^{-\frac{1}{2}} \sum_{\alpha=1}^{N} (X_\alpha - \mu) \underset{N \to \infty}{\sim} N_p(0, \Sigma)$$

  (Se trata de una versión del Teorema Central del Límite, para vectores i.i.d. $\sim (\mu, \Sigma)$)

- Distribución asintótica de la matriz de dispersiones muestral, A.
  RESULTADO: Sea $X_1, \ldots, X_N, \ldots$ una sucesión de vectores aleatorios p-dimensionales indpendientes e idénticamente distribuidos, con vector de medias μ y matriz de covarianzas Σ (i.e., $X_\alpha \sim (\mu, \Sigma), \alpha = 1, \ldots, N, \ldots, N)$, u con momentos de orden cuatro finitos. Sean $\bar{X}_N = \frac{1}{N} \sum_{\alpha=1}^{N} X_\alpha$ y $A_n = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X}_N)(X_\alpha - \bar{X}_N)', \forall N \geq 1$. Entonces, cuando $N \to \infty$, se tiene la distribución asintótica

  $$N^{-\frac{1}{2}}(A_N - N\Sigma) \underset{N \to \infty}{\sim} N_{p^2}(0, V)$$

(en el sentido de que $N^{-\frac{1}{2}}(\mathrm{Vec}(A_N - N\,\mathrm{Vec}(\Sigma)) \underset{N\to\infty}{\sim} N_{p^2}(0, V))$, con $V = \mathrm{Cov}(\mathrm{Vec}((X_\alpha - \mu)(X_\alpha - \mu)'))$. (OBSERVACIÓN: La matriz V será necesariamente singular, por lo que la distribución límite se refiere al caso general de la DNM)