Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática

Apuntes para el Curso:

Métodos Estadísticos Predictivos

Prof. Nancy Lacourly
Félix Carrasco - Claudio Pareja
2009

PREFACIO

El curso de métodos estadísticos predictivos, obligatorio para los alumnos de ingeniería matemá-tica, profundiza y complementa los temas de análisis multivariados vistos en el curso de estadística. Se trata de dar justificaciones matemáticas de los métodos así como aspectos aplicados.

Los modelos pretenden representar estructuras de un fenómeno descrito mediante datos. Todo modelo estadístico se basa en supuestos y simplifica la realidad. Es entonces importante verificar la validez del modelo, tanto los supuestos en los cuales se basa el modelo así como la calidad de la aproximación que el modelo hace del fenómeno.

Índice general

1. FUNDAMENTOS MATEMÁTICOS **7**

1.1. Derivación matricial. 9
1.2. Inversa generalizada 10
1.3. Elementos relacionados con formas cuadráticas 11
1.3.1. M-simetría. 11
1.3.2. Proyectores. 11
1.3.3. Matriz de varianza-covarianza de un vector aleatorio 12
1.3.4. Esperanza de una forma cuadrática 13
1.4. Distribución normal multivariada. 13
1.4.1. Definiciones y propiedades. 13
1.4.2. Varianza de una forma cuadrática. 16
1.5. Distribuciones derivadas de la normal. 17
1.5.1. La distribución χ^2_n. 17
1.5.2. La distribución $F_{m,n}$. 20
1.6. Inferencia Estadística 20
1.6.1. Cantidad de información de Fisher y la Desigualdad de Cramer-Rao 20
1.6.2. Test de hipótesis. 23
1.6.3. Caso de dos hipótesis simples 24
1.6.4. Test uniformemente más potente (UMP). 25
1.6.5. Test de razón de verosimilitudes. 26
1.6.6. Ejemplos 26
1.7. Ejercicios. 28

Capítulo 1

FUNDAMENTOS MATEMÁTICOS

1.1. Derivación matricial.

Definición 1.1 *1. Sea f una variable dependiente de un vector $X \in \mathbb{R}^p$ de componentes x_i. Se define el vector* **gradiente** ∇f *como el vector de las derivadas de f con respecto a los elementos de X:*

$$\nabla f = \frac{\partial f}{\partial X} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{pmatrix}.$$

Si f depende de una matriz $A = (a_{ij})_{i,j}$, entonces $\dfrac{\partial f}{\partial A}$ es una matriz de término general $\dfrac{\partial f}{\partial a_{ij}}$.

2. Se define el **Hessiano** *H como la matriz simétrica de las segundas derivadas:*

$$H_f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j} = \frac{\partial \nabla f}{\partial X} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_p} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_p \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_p^2} \end{bmatrix}.$$

Ejercicio 1.1 *Sean u y $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^r$, $A \in \mathcal{M}_{r,p}$ y $B \in \mathcal{M}_{p,r}$, demostrar las siguientes propiedades básicas:*

$$\begin{array}{ll} f = u^t X = X^t u & \partial f/\partial X = u \\ f = AX & \partial f/\partial X = A \\ f = X^t BY & \partial f/\partial X = BY \\ f = X^t AX \quad (p = r) & \partial f/\partial X = (A + A^t)X \\ f = X^t AX \quad (p = r) & \partial f/\partial A = XX^t \\ f = Traza(BA) & \partial f/\partial A = B^t \\ f = Traza(BAA^t) \quad (p = r) & \partial f/\partial A = (B + B^t)A \end{array}$$

Proposición 1.1 *Si X es una matriz cuadrada invertible se tiene la siguiente regla de derivación:*

$$\frac{\partial \mid X \mid}{\partial X} = \mid X \mid (X^T)^{-1}$$

Dem: Denotemos X^{ij} a la matriz formada por las columnas de X pero sin la columna i y la fila j. Fijemos x_{kl} y calculemos $\frac{\partial |X|}{\partial x_{kl}}$.

$$\begin{aligned} \mid X \mid &= \sum_i (-1)^{(i+k)} x_{ki} \mid X^{ki} \mid \\ \Rightarrow \frac{\partial \mid X \mid}{\partial x_{kl}} &= (-1)^{(l+k)} \mid X^{kl} \mid \end{aligned}$$

Por otro lado tenemos que, si denotamos $X^{-1} = (b_{ij})_{i,j}$,

$$\begin{aligned} b_{ij} &= (-1)^{i+j} \frac{\mid X^{ji} \mid}{\mid X \mid} \\ \Rightarrow (-1)^{l+k} \mid X^{lk} \mid &= \mid X \mid b_{kl} \end{aligned}$$

Reemplazando esto en el paso anterior, se concluye.

■

1.2. Inversa generalizada

En un sistema de ecuaciones: $\Gamma x = y$, si la matriz Γ es invertible (Γ^{-1}), entonces $x = \Gamma^{-1}y$. Al Estudiar la definición de **inversa generalizada o g-inversa** se quiere ampliar el concepto de solución para un sistema lineal, donde la matriz Γ resulta ser no invertible.

Sea Γ de orden p y de rango r. Si $r = p$, entonces Γ es invertible, si no, existen matrices Γ^{-}, g-inversa de Γ tales que $\Gamma\Gamma^{-}\Gamma = \Gamma$. Es decir: $x = \Gamma^{-}y$ es una solución de $\Gamma x = y$. Se observará que $\Gamma^{-}\Gamma$ no es igual a la identidad salvo si Γ es invertible, pero se tiene:

$$(\Gamma\Gamma^{-})^2 = \Gamma\Gamma^{-} \quad \text{y} \quad (\Gamma^{-}\Gamma)^2 = \Gamma^{-}\Gamma.$$

Notar que la matriz g-inversa no es invertible ni única.

Definición 1.2 *Sea Γ una matriz y Γ^{-} una g-inversa. Si además la g-inversa de Γ^{-} es Γ, con $\Gamma\Gamma^{-}$ y $\Gamma^{-}\Gamma$ simétricas, entonces tal inversa generalizada es la de* Penrose. *La inversa generalizada de Penrose se denotará Γ^{+}.*

Proposición 1.2 (Propiedades básicas) . *Sea $X(nxp)$ de rango incompleto $r < p$ y G una g-inversa de X^tX. Se tiene que*

1. *Si G es de la Penrose, entonces es única.*
2. *G^t es una g-inversa de X^tX.*
3. *GX^t es una g-inversa de X.*
4. *XGX^t es invariante para cualquier g-inversa G de X^tX.*
5. *XGX^t es simétrica aún si G no lo es.*

<u>Dem:</u>

1. Ver página 28, ejercicio 5.
2. Directa.
3. Mostramos que $XGX^tX = X$. Como G es una g-inversa de X^tX, se tiene: $X^tXGX^tX = X^tX$. Luego $X^tXGX^tX - X^tX = 0 \Rightarrow (GX^tX - I)(X^tXGX^tX - X^tX) = 0 \Rightarrow (XGX^tX - X)^t(XGX^tX - X) = 0 \Rightarrow XGX^tX - X = 0$.
4. Del resultado anterior se deduce que si G_1 y G_2 son g-inversas de X^tX, $XG_1X^tX = XG_2X^tX$, o sea $XG_1X^t = XG_2X^t$.
5. Si G_1 es una g-inversa simétrica de X^tX entonces XG_1X^t es simétrica. De la propiedad de invarianza se deduce el resultado.

■

Se nota que las tercera y cuarta propiedades anteriores se basan en el siguiente resultado:

Proposición 1.3 *Sea $A(mxn)$. Para toda matriz $B(nxp)$ y $C(nxp)$, se tiene*

$$AB = AC \Leftrightarrow A^tAB = A^tAC$$

<u>Dem:</u> Si $AB = AC \Rightarrow A^tAB = A^tAC$.
Recíprocamente, si $A^tAB = A^tAC \Rightarrow A^tAB - A^tAC = 0$
Luego $(B^t - C^t)(A^tAB - A^tAC) = 0 \Rightarrow (AB - AC)(AB - AC) = 0 \Rightarrow AB - AC = 0$.

■

1.3. Elementos relacionados con formas cuadráticas

1.3.1. M-simetría.

Sea F un espacio vectorial real (e.v.r) de dimensión n ($F = \mathbb{R}^n$) y M una matriz simétrica definida positiva sobre F. A todo x e y de F se le asocia el producto escalar M, que define una metrica euclideana en F:

$$< x,y >_M = M(x,y) = x^t My = y^t Mx$$

Se obtiene entonces la distancia euclidiana entre x e y:

$$d(x,y) = \sqrt{< x-y, x-y >_M} = \|x-y\|_M$$

Se define también la ***M*-ortogonalidad**: $x \perp_M y \iff < x,y >_M = 0$.

Definición 1.3 *Se dice que una aplicación lineal A de F en F es M-simétrica si y solo si M es simétrica, definida positiva y $\forall x,y \in F :< x,A(y) >_M =< A(x),y >_M$.*

Se deduce que si A es M-simétrica, entonces $(M \circ A)^t = M \circ A$. En particular, si $M = I_n$ se obtiene la métrica usual y entonces A es simétrica. En ese caso los valores propios de la matriz A, que es simétrica, son reales y existe una base ortonormal de F formada de vectores propios de A. Si A es M-simétrica, se extienden estos resultados:

Proposición 1.4 *Si A es M-simétrica, sus valores propios son reales y existe una base M-ortonormal de F formada de vectores propios de A. Si $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los valores propios de A, entonces existe una matriz U tal que si $U_1, U_2, \ldots, U_n$ son las columnas de U entonces $AU_j = \lambda_j U_j$ con $U^t MU = I_n$.*

1.3.2. Proyectores.

Se definen los proyectores en un e.v.r. F a partir de la descomposición en suma directa de F:

$$F = H \oplus G \quad \text{ssi} \quad \forall x \in F \quad \exists!(u,v) \in H \times G \quad tq \quad x = u + v$$

El vector u es la proyección de x sobre H paralelamente a G y v es la proyección de x sobre G paralelamente a H.

Proposición 1.5 *Sean P el proyector sobre H paralelamente a G: $u = P(x)$ y Q el proyector sobre G paralelamente a H: $v = Q(x)$. Se cumplen*

1. *Los operadores P y Q son lineales e idempotentes de orden 2 ($P \circ P = P$ y $Q \circ Q = Q$).*
2. *La imagen de P: $Im(P) = H = Ker(Q)$ y $Im(Q) = G = Ker(P)$.*
3. *El rango de P es la dimensión de H y el rango de Q es la dimensión de G.*
4. *Los operadores P y Q tienen dos valores propios distintos: el valor propio 1 de espacio propio igual a H para P (y G para Q) y el valor propio 0 de espacio propio igual a G para P (y H para Q). Además $P + Q = I_n$ y $P \circ Q = 0_n$.*

Consideremos ahora el caso particular en que H y G son M-ortogonales: $< u,v >_M = 0$. En este caso se habla de proyecciones M-ortogonales y los proyectores P y Q tienen además la propiedad de ser M-simétricos. En efecto, si $x = P(x) + Q(x)$ e $y = P(y) + Q(y)$, entonces:

$$< x,P(y) >_M =< P(x) + Q(x), P(y) >_M =$$

$$< P(x),P(y) >_M=< P(x),P(y)+Q(y) >_M=< P(x),y >_M .$$

Luego, si P es un proyector M-ortogonal, existe una base M-ortonormal de vectores propios de $M \circ P$.

Sea $y \in \mathbb{R}^n$ y H un s.e.v. de $\mathbb{R}^n$. La proyección M-ortogonal Py de y sobre H es el punto de H lo más cercano de y en el sentido de la métrica M.

Buscamos la expresión de un proyector M-ortogonal a partir de un conjunto generador $\{x_1,\ldots,x_r\}$ de H. La proyección Py es tal que $(y-Py) \perp_M H$, por lo tanto $\forall j : (y-Py) \perp_M x_j$. Luego si X es la matriz formada de los vectores x_j en columnas, $X^tM(y-Py) = 0$, o sea

$$X^tMy = X^tMPy$$

Ahora bien $Py \in H$, luego existe un vector $b \in \mathbb{R}^r$ tal que $Py = Xb$. De aquí se obtienen las **ecuaciones normales**:

$$X^tMXb = X^tMy$$

Si X es de rango r, entonces se obtiene el vector b y la expresión del proyector:

$$b = (X^tMX)^{-1}X^tMy$$

$$P = X(X^tMX)^{-1}X^tM$$

Si X no es de rango r (los x_j no forman una base de H) se pueden suprimir columnas de X hasta extraer una base de H.

1.3.3. Matriz de varianza-covarianza de un vector aleatorio

Sea Y un vector real aleatorio de dimensión p de media μ y matriz de varianza-covarianza Γ:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}, \qquad \mathbb{E}(Y) = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \mu,$$

$$\Gamma = \mathbb{V}(Y) = (\gamma_{ij})_{ij} = \mathbb{E}\{(Y-\mu)(Y-\mu)^t\} = \mathbb{E}(YY^t) - \mu\mu^t$$

con $\gamma_{i,j} = Cov(y_i,y_j)$ si $i \neq j$ y $\gamma_{i,i} = \mathbb{V}(y_i)$. Notar que la matriz Γ es semi-definida positiva.

Si $Z = AY$ en que A es una transformación lineal, se tiene:

$$\mathbb{E}(Z) = A\mathbb{E}(Y) = A\mu$$

$$\mathbb{V}(Z) = \mathbb{E}(ZZ^t) - (A\mu)(A\mu)^t = A\Gamma A^t$$

Sea $K = Ker(\Gamma)$ el núcleo de Γ. Si $u \in K$, entonces $\mathbb{V}(u^tY) = 0$ y $u^t(Y-\mu) \overset{c.s.}{=} 0$. Si Γ es de rango r, K es de dimensión $p-r$ y $Y-\mu$ pertenece (c.s.) a un s.e.v. S de dimensión r.

Sea Γ invertible. Se asocia al vector Y el **elipsoide de concentración**:

$$\{Z \in \mathbb{R}^p | (Z-\mu)^t\Gamma^{-1}(Z-\mu) = c\}$$

en que c es una constante dada.

Sea la diagonalización de Γ: $\Gamma = UDU^t$, en las columnas de U se tienen los vectores propios de Γ y D es la matriz diagonal con los valores propios asociados.Como D es semidefinida positiva, se puede escribir entonces:

$$\Gamma = TT^t,\ T := UD^{1/2}$$

Ahora bien, existen una infinidad de descomposiciones de Γ, pero si se pide que T sea triangular inferior, se obtiene la descomposición de **Choleski**. Por otro lado, si Γ es de rango r ($r < p$) y la matriz Λ restricción de T a los r vectores asociados a los valores propios no nulos, entonces: $\Gamma = \Lambda\Lambda^T$.

Si el vector aleatorio Z es un vector de media nula y matriz de varianza-covarianza igual a I_p, entonces $\mu + \Lambda Z$ es un vector de media μ y matriz de varianza-covarianza $\Gamma = \Lambda\Lambda^t$. Si Γ es de rango p, $\Lambda = T$ es invertible, entonces $Z = \Lambda^{-1}(Y-\mu)$ es un vector aleatorio de dimensión p de media nula y matriz de varianza-covarianza I_p. El cambio de variables $X = \Lambda^{-1}Y$ permite pasar de la métrica Γ^{-1} a la métrica I_p:

$$\|X\|^2_{I_p} = X^tX = Y^t(\Lambda^t)^{-1}\Lambda^{-1}Y = Y^t\Gamma^{-1}Y = \|Y\|^2_{\Gamma^{-1}}$$

Luego

$$\|Z\|^2_{I_p} = \|Y-\mu\|^2_{\Gamma^{-1}}$$

Calculemos la media y la varianza de $\|Y-\mu\|^2_{\Gamma^{-1}}$.

$$\mathbb{E}(\|Y-\mu\|^2_{\Gamma^{-1}}) = \mathbb{E}(\|Z\|^2_{I_p}) = \sum_{i=1}^{p}\mathbb{E}(z_i^2) = \sum_{i=1}^{p}\mathbb{V}(z_i) = p$$

$$\mathbb{V}(\|Y-\mu\|^2_{\Gamma^{-1}}) = \mathbb{V}(\|Z\|^2_{I_p}) = \mathbb{V}\left(\sum_{i=1}^{p} z_i^2\right) = \sum_{i=1}^{p}\mathbb{V}\left(z_i^2\right) = p$$

Pues las p v.a. z_i son no correlacionadas entre si y todas de varianza igual a 1.

1.3.4. Esperanza de una forma cuadrática

Sea la matriz simétrica $A \in \mathcal{M}_{n,n}$ y $X \in \mathbb{R}^n$. Se define la forma cuadrática:

$$Q = X^TAX = Traza(X^TAX) = Traza(AXX^T).$$

Entonces: $\mathbb{E}(Q) = Traza(A\mathbb{V}(X)) + \mathbb{E}(X)^TA\mathbb{E}(X)$. En efecto:

$$\begin{aligned}
\mathbb{E}(X^TAX) &= \mathbb{E}((X-\mathbb{E}(X))^TA(X-\mathbb{E}(X))) + \mathbb{E}\{\mathbb{E}(X)^TAX + X^TA\mathbb{E}(X) - \mathbb{E}(X)^TA\mathbb{E}(X)\} \\
&= \mathbb{E}((X-\mathbb{E}(X))^TA(X-\mathbb{E}(X))) + \mathbb{E}(X)^TA\mathbb{E}(X) \\
&= \mathbb{E}(tr(A(X-\mathbb{E}(X))(X-\mathbb{E}(X))^T)) + \mathbb{E}(X)^TA\mathbb{E}(X) \\
&= tr(A\mathbb{V}(X)) + \mathbb{E}(X)^TA\mathbb{E}(X)
\end{aligned}$$

el resultado buscado.

1.4. Distribución normal multivariada.

1.4.1. Definiciones y propiedades.

Definición 1.4 *Se dice que Y es un vector normal multivariado de orden p de vector de media μ y de matriz de varianza-covarianza Γ (se denota $Y \sim N_p(\mu,\Gamma)$), si y sólo si:*

$$\forall u \in \mathbb{R}^p \to \mathbb{R} : u^TY \sim N(u^T\mu, u^T\Gamma u).$$

Es decir, Y es un vector normal ssi toda combinación lineal de Y es una variable aleatoria normal.

Definición 1.5 *Se dice que $Y \sim N_p(\mu,\Gamma)$ si y sólo si, su función característica es:*

$$\forall v \in \mathbb{R}^p : \Psi_Y(v) = \exp\left\{iv^T\mu - \frac{1}{2}v^T\Gamma v\right\}.$$

Proposición 1.6 *Algunas propiedades interesantes:*

- *Tomando como vector u a los distintos vectores canónicos, se obtiene que las leyes marginales de Y son normales, pero la recíproca es falsa: un vector formado de variables normales no es necesariamente un vector normal.*
- *Sea $A \in \mathcal{M}_{q,p} : Y \sim N_p(\mu,\Gamma) \Rightarrow X = AY \sim N_q(A\mu, A\Gamma A^T)$.*
- *Las v.a. y_i son independientes $\iff$ Γ es diagonal.*
- *Si Γ es de rango r, existe una matriz $\Lambda \in \mathcal{M}_{p,r}$ tal que $\Gamma = \Lambda\Lambda^T$ entonces*

  $$Y \sim N_p(\mu,\Gamma) \iff Y = \mu + \Lambda X \quad \text{con} \quad X \sim N_r(0, I_r)$$

  es decir que las componentes del vector X son centradas, normalizadas e independientes entre si.
- *Si Γ es invertible, Λ es invertible también y $X = \Lambda^{-1}(Y-\mu)$. Se puede escribir también la transformación de* Mahalanobis $X = \Gamma^{-1/2}(Y-\mu)$.

Esta última propiedad permite calcular la densidad del vector X en el caso que Γ es invertible. En efecto, se puede calcular la densidad del vector $X \sim N_p(0, I_p)$:

$$f(X) = \prod_{i=1}^{p} f_i(x_i) = \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{p} x_i^2\right) = \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} e^{-\frac{1}{2}X^TX}.$$

Como $X^TX = (\Lambda^{-1}(Y-\mu))^T\Lambda^{-1}(Y-\mu) = (Y-\mu)^T\Gamma^{-1}(Y-\mu)$, el jacobiano de la transformación es:

$$|J(X \longrightarrow Y)| = \frac{1}{|\Lambda|} = \frac{1}{\sqrt{|\Gamma|}}.$$

Luego la densidad de Y es:

$$h(Y) = \frac{1}{(2\pi)^{p/2}\sqrt{|\Gamma|}} \exp\left\{-\frac{1}{2}(Y-\mu)^T\Gamma^{-1}(Y-\mu)\right\}.$$

Entonces se observará que la densidad de la distribución $N_p(0, I_p)$ es constante sobre los elipsoides de la forma: $(Y-\mu)^T\Gamma^{-1}(Y-\mu) = d^2$.

Proposición 1.7 *Sean dos vectores normales $Y_1 \sim N_{p_1}(\mu_1, \Gamma_{11})$ e $Y_2 \sim N_{p_2}(\mu_2, \Gamma_{22})$, con Γ_{12} como la matriz de covarianza entre Y_1 e Y_2. Entonces la distribución condicional de Y_1 dado Y_2 es una normal:*

$$Y_1|Y_2 \sim N_{p_1}\left(\mu_1 - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^T(Y_2-\mu_2), \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^T\right)$$

Dem: Considerando las particiones siguientes:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}$$

Con $Y_1 \in \mathbb{R}^{p_1}$ y $Y_2 \in \mathbb{R}^{p_2}$ $(p = p_1 + p_2)$.

Determinemos la ley condicional de Y_1 dado Y_2 cuando la matriz $\Gamma_{2,2}$ es invertible. Sea el cambio de variables $T = \begin{pmatrix} T_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Y_1 - \Gamma_{12}\Gamma_{22}^{-1}Y_2 \\ Y_2 \end{pmatrix}$, su esperanza y su matriz de varianza-covarianza queda dada por:

$$\mathbb{E}(T) = \begin{pmatrix} \mu_1 - \Gamma_{12}\Gamma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}$$

$$\mathbb{V}(T) = \begin{bmatrix} \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^T & \Gamma_{12} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{22} \\ \Gamma_{12}^T - \Gamma_{12}^T\Gamma_{22}^{-1}\Gamma_{22} & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^T & 0 \\ 0 & \Gamma_{22} \end{bmatrix}$$

Se deduce que los vectores T_1 y Y_2 son normales e independientes entre sí (ver proposición 1.6).
La densidad conjunta de Y_1 e Y_2 puede escribirse a partir de la densidad de T_1 y Y_2. Sea $g(\cdot,y)$ la función densidad de su primer argumento en $y = (y_1, y_2)$ entonces:

$$g(Y,y) = g(T(Y),y)\left|J(T \to Y)\big|_{Y=y}\right|$$

en donde $|J(T \to Y)|$ es el determinante del jacobiano de la transformación de T a Y que es igual a:

$$|J(T \to Y)| = \left|\frac{DT}{DY}\right| = \begin{vmatrix} I_{p_1} & -\Gamma_{12}\Gamma_{22}^{-1} \\ 0 & I_{p_2} \end{vmatrix} = 1$$

Sigue que

$$\begin{aligned} g(T(Y),y) &= g(T_1,y_1)g(Y_2,y_2) \\ \Rightarrow g(Y_1|Y_2,y) &= \frac{g(Y,y)}{g(Y_2,y_2)} = \frac{g(T(Y),y)}{g(Y_2,y_2)} \\ &= g(T_1(Y_1,Y_2)) \end{aligned}$$

Sabemos ahora que $Y_1|Y_2$ posee una distribución normal. Calculemos su esperanza y su matriz de varianza-covarianza, usando lo hecho para T:

$$\begin{aligned} \mathbb{E}(Y_1|Y_2) &= \mathbb{E}\left(T_1 + \Gamma_{12}\Gamma_{22}^{-1}T_2|T_2\right) = \mu_1 + \Gamma_{12}\Gamma_{22}^{-1}(Y_2 - \mu_2) \\ \mathbb{V}(Y_1|Y_2) &= \mathbb{V}(T_1) = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^t \end{aligned}$$

Obs:

- La esperanza condicional $\mathbb{E}(Y_1|Y_2) = \mathbb{E}(T_1 + \Gamma_{12}\Gamma_{22}^{-1}T_2|T_2) = \mu_1 + \Gamma_{12}\Gamma_{22}^{-1}(Y_2 - \mu_2)$ es una función lineal de Y_2.
- La matriz de varianza condicional $\mathbb{V}(Y_1|Y_2) = \Gamma_{11\bullet 2} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^T$ es independiente de Y_2.

La siguiente propiedad es importante ya que será usada más adelante:

Proposición 1.8 *Sea $Y \sim N_p(\mu,\Gamma)$ y Γ regular, entonces si u y $v \in \mathbb{R}^p$, A y B son dos matrices simétricas de orden p, se tienen las propiedades de independencia siguientes:*

1. *u^TY y v^TY son independientes $\iff u^T\Gamma v = 0$.*
2. *u^TY y Y^TAY son independientes $\iff u^T\Gamma A = 0$.*
3. *Y^TAY y Y^TBY son independientes $\iff A\Gamma B = 0$.*

Dem: Ejercicio. (**Hint**: Calcular covarianzas).

1.4.2. Varianza de una forma cuadrática.

Sea la matriz simétrica $A \in \mathcal{M}_{n,n}$ y $X \in \mathbb{R}^n$. Como antes, definimos $Q = X^T AX = Traza(X^T AX) = Traza(AXX^T)$, calcularemos la varianza de la forma cuadrática cuando el vector X sea una normal multivariada $(x_1, x_2, \ldots, x_n) \sim N_n(\theta, \sigma^2 I_n)$. Estos cálculos nos darán:

$$\mathbb{V}(Q) = \mathbb{V}(X^T AX) = 2\sigma^4 Traza(A^2) + 4\sigma^2 \theta^T A^2 \theta.$$

En efecto,

$$\begin{array}{rcl} Q & = & (X-\theta)^T A(X-\theta) + 2\theta^T A(X-\theta) + \theta^T A\theta \\ \Rightarrow Q^2 & = & \{(X-\theta)^T A(X-\theta)\}^2 + 4\{\theta^T A(X-\theta)\}^2 + (\theta^T A\theta)^2 \\ & & \ldots 2\theta^T A\theta\{(X-\theta)^T A(X-\theta) + 2\theta^T A(X-\theta)\} + 4\theta^T A(X-\theta)(X-\theta)^T A(X-\theta) \end{array}$$

Consideremos el cambio de variable $Y = X - \theta$. Calculemos la esperanza de cada sumando de Q^2 por separado:

$$\mathbb{E}\{((X-\theta)^T A(X-\theta))^2\} = \mathbb{E}\{(Y^T AY)^2\} = \mathbb{E}\left\{\sum_{i,j,k,l} a_{i,j} a_{k,l} Y_i Y_j Y_k Y_l\right\}.$$

En el caso de la distribución normal se tiene:

$$\begin{array}{rcl} \mathbb{E}(Y_i Y_j Y_k Y_l) & = & \left\{ \begin{array}{ccc} 3\sigma^4 & \text{si} & i=j=k=l \\ \sigma^4 & \text{si} & (i=j\neq k=l) \text{ o } (i=l\neq j=k) \text{ o } (i=k\neq j=l) \\ 0 & \text{si} & \text{no} \end{array} \right. \\ \Rightarrow \mathbb{E}\{(Y^T AY)^2\} & = & 3\sigma^4 \sum_{i=1}^n a_{ii}^2 + \sigma^4 \left[\sum_{i\neq k} a_{ii}a_{kk} + \sum_{i\neq j} a_{ij}a_{ji} + \sum_{i\neq j} a_{ij}^2\right] \\ & = & \sigma^4 \left[\left(\sum_{i=1}^n a_{ii}^2 + \sum_{i\neq k} a_{ii}a_{kk}\right) + 2\left(\sum_{i=1}^n a_{ii}^2 + \sum_{i\neq j} a_{ij}^2\right)\right] = \sigma^4[(Traza(A))^2 + 2Traza(A^2)] \end{array}$$

Dado que

$$\begin{array}{rcl} A^T & = & A \\ \Rightarrow Traza(A^2) & = & \sum_{i,j} a_{i,j}^2 \\ \Rightarrow \mathbb{E}[(\theta^T AY)^2] & = & \mathbb{E}\left(\sum_{i,j} (A\theta)_i (A\theta)_j Y_i Y_j\right) \\ \Rightarrow \mathbb{E}[(\theta^T AY)^2] & = & \sigma^2 \sum_{i=1}^n (A\theta)_i^2 \\ & = & \sigma^2 (A\theta)^T (A\theta) = \sigma^2 \theta^T A^2 \theta \end{array}$$

Como[1]

$$\begin{array}{rcl} \mathbb{E}[(\theta^T AY)(Y^T AY)] & = & 0 \; y \\ \mathbb{E}(Q) = \mathbb{E}(X^T AX) & = & \sigma^2 Traza(A) + \theta^T A\theta \end{array}$$

entonces:

$$\mathbb{V}(Q) = \mathbb{E}(Q^2) - \mathbb{E}(Q)^2 = 2\sigma^4 Traza(A^2) + 4\sigma^2 \theta^T A^2 \theta.$$

En particular si $\theta = 0$, entonces $\mathbb{V}(Q) = 2\sigma^4 Traza(A^2)$. Para el caso general[2] $Y \sim N_p(\mu, \Gamma)$ se tiene $\mathbb{V}(Y^T AY) = 2Traza((A\Gamma)^2) + 4\mu^T A\Gamma A\mu$.

[1] Por qué?

[2] Ver página 28, ejercicio 7.

1.5. Distribuciones derivadas de la normal.

Se repasan aquí las distribuciones univariadas clásicas derivadas de la normal y algunas de sus aplicaciones.

1.5.1. La distribución χ^2_n.

Definición 1.6 *Si $X \sim N_n(0, I_n)$, entonces:*

$$\|X\|^2 = \sum_{i=1}^{n} x_i^2 \sim \chi^2_n.$$

Decimos que sigue una distribución chi cuadrado a n grados de libertad.
Cuando $X \sim N_n(\mu, I_n)$, se define la distribución χ^2_n decentrada:

$$\|X\|^2 = \sum_{i=1}^{n} x_i^2 \sim \chi^2_n(\delta^2)$$

con el decentramiento $\delta^2 = \sum_{i=1}^{n} \mu_i^2 = \|\mu\|^2$, aunque algunos autores definen el decentramiento como $\frac{1}{2}\|\mu\|^2$.

Se observa que $\|X\|^2 \sim \chi^2_n(\|\mu\|^2)$ si sólo si $\|X-\mu\|^2 \sim \chi^2_n$. Se deduce que:

$$E(\|X\|^2) = n + \|\mu\|^2 \quad \text{y} \quad Var(\|X\|^2) = 2n + 4\|\mu\|^2.$$

La función de densidad de la variable $U \sim \chi^2_n$ es:

$$f(u) = \frac{u^{\frac{n}{2}-1} e^{-\frac{u}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \mathbb{1}_{\{u>0\}}.$$

Luego, la función de distribución de U: $F(U) = \mathbb{P}(U \leq u)$ define el interior de la esfera centrada en el origen de $\mathbb{R}^n$ y de radio $\sqrt{u}$. En el caso de decentramiento $U \sim \chi^2_n(\|\mu\|^2)$ la esfera es de centro μ.

Proposición 1.9 *Se tienen las siguientes propiedades básicas.*

1. *La suma de variables $\chi^2_{r_i}$ independientes tiene una distribución χ^2_n con $\sum_i r_i = n$.*
2. *Existe una relación con la distribución Gamma: $\chi^2_n = Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.*
3. *Su distribución asintótica es normal: se tiene como aproximación para n suficientemente grande $\chi^2_n \approx N(n, 2n)$. Una aproximación un poco más exacta es $\sqrt{2\chi^2_n} - \sqrt{2n-1} \approx N(0,1)$.*

Proposición 1.10 *Si $Y \sim N_p(\mu, \Gamma)$ con Γ de rango r, entonces $\|Y-\mu\|^2_{\Gamma^+} \sim \chi^2_r$, en donde Γ^+ es la inversa generalizada de Penrose de Γ.*

Dem: Como $\Gamma = \Lambda\Lambda^T$, con Λ de mismo rango r que Γ, existe X tal que $Y = \mu + \Lambda X$, con $X \sim N_r(0, I_r)$. Pero

$$\|X\|^2 = \sum_{i=1}^{r} x_i^2 \sim \chi^2_r.$$

Como se puede escribir $X = (\Lambda^T\Lambda)^{-1}\Lambda^T(Y-\mu)$, luego:

$$\|X\|^2 = \|Y-\mu\|^2_{\Gamma^+} \sim \chi^2_r$$

en que $\Gamma^+ = \Lambda(\Lambda^T\Lambda)^{-2}\Lambda^T$ es la inversa de Penrose de Γ. Si Γ es invertible, $\Gamma^+ = \Gamma^{-1}$.

$\blacksquare$

El siguiente teorema tiene muchas aplicaciones en el estudio de los modelos lineales.

Teorema 1.1 *Si $Y \sim N_n(0,\Gamma)$ con Γ regular, A una matriz simétrica de rango r, entonces*

$$Q = Y^T AY \sim \chi^2_r \iff A\Gamma \text{ es idempotente de orden 2.}$$

Dem: Vemos la implicancia hacia la izquierda usando la función característica. Sea $A\Gamma$ idempotente, tiene rango r, entonces $A\Gamma$ tiene sus valores propios λ_i iguales a 1 con multiplicidad r y 0 con multiplicidad $n-r$. La función característica de Q está dada por:

$$\begin{aligned} \mathbb{E}\left(e^{tQ}\right) &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\Gamma|^{-\frac{1}{2}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} e^{tQ-\frac{1}{2}Y^T\Gamma^{-1}Y} dy_1 \ldots dy_n \quad \text{con}\, t \in]-\infty, 1/2] \\ &= |I_n - 2tA\Gamma|^{-\frac{1}{2}} = \prod_{i=1}^{n} (1-2t\lambda_i)^{-\frac{1}{2}} = (1-2t)^{-\frac{r}{2}} \end{aligned}$$

Que es la f.g.m. de χ^2_r

Ahora la implicancia hacia la derecha. Sea $Q \sim \chi^2_r$, entonces $E\left(e^{tQ}\right) = (1-2t)^{-\frac{r}{2}} = |I_n - 2tA\Gamma|^{-\frac{1}{2}}$. Sea $u = 2t$, entonces

$$(1-u)^r = |I_n - uA\Gamma| = \prod_{i=1}^{n} (1-u\lambda_i) \quad \text{con}\, u \in]-\infty, 1].$$

Tal ecuación no puede tener más de r términos no nulos. Luego tomamos:

$$\begin{aligned} (1-u)^r &= \prod_{i=1}^{r} (1-u\lambda_i) \\ \Rightarrow r\ln(1-u) &= \sum_{i=1}^{r} \ln(1-u\lambda_i) \end{aligned}$$

Y se obtiene que $\forall i = 1, \ldots, r : \lambda_i = 1$. Como los otros valores propios son nulos, entonces $A\Gamma$ es idempotente de rango r.

$\blacksquare$

Se deduce un teorema muy importante en el estudio de las formas cuadráticas.

Teorema 1.2 *Dado el vector aleatorio $Y \sim N_n(\mu,\Gamma)$ con Γ regular, se considera la forma cuadrática $Q = Y^T AY$ con A simétrica de rango r y las p formas cuadráticas $Q_h = Y^T A_h Y$ con A_h simétrica de rango r_h ($1 \leq h \leq p$) tales que*

$$\left(A = \sum_{h=1}^{p} A_h\right), \quad Q = \sum_{h=1}^{p} Q_h$$

Sean las cuatro proposiciones siguientes:

1. $Q \sim \chi^2_r(\mu^T A\mu)$
2. $\forall h,\ Q_h \sim \chi^2_{r_h}(\mu^T A_h \mu)$
3. $\forall h \neq k : Q_h$ *y* Q_k *son independientes*
4. $\sum_{h=1}^{p} r_h = r$.

Entonces dos de las proposiciones, salvo las dos últimas que son equivalentes, implican las dos otras.

Dem: Del teorema anterior y la prop (1.8) se deduce que:

- La proposición (1) es equivalente a decir que $A\Gamma$ es un operador idempotente de rango r y $A\Gamma A = A$.
- (2) es equivalente a decir que los $A_h\Gamma$ son operadores idempotentes de rango r_h y $A_h\Gamma A_h = A_h$.
- (3) es equivalente a decir que $\forall h \neq k : A_h\Gamma A_k = 0$

Utilizaremos estas equivalencias para la demostración:

Veamos que las proposiciones (3) y (4) son equivalentes

$$A = \sum_{h=1}^{p} A_h \Rightarrow A\Gamma = \sum_{h=1}^{p} A_h\Gamma \Rightarrow Im(A\Gamma) = \sum_{h=1}^{p} Im(A\Gamma).$$

Entonces se tiene las equivalencias siguientes:

$$\begin{array}{rl} & \forall h \neq k : A_h\Gamma A_k = 0 \\ \Longleftrightarrow & \forall h \neq k : A_h\Gamma A_k\Gamma = 0 \\ \Longleftrightarrow & \forall h \neq k, \forall x, y \in \mathbb{R}^n < A_h\Gamma x, A_k\Gamma y >= 0 \\ \Longleftrightarrow & Im(A_h\Gamma) \perp Im(A_k\Gamma) \\ \Longleftrightarrow & Im(A\Gamma) = \bigoplus_{h=1}^{p} Im(A_h\Gamma) \\ \Longleftrightarrow & \sum_{h=1}^{p} r_h = r \end{array}$$

Ahora, pdq (1)+(2) $\Rightarrow$ (3). Primero notemos dos propiedades acerca de A y A_h. Si (1) y (2) se cumplen, luego sigue (el argumento es el mismo para A_h):

$$(A\Gamma)^2 = A\Gamma \Rightarrow A\Gamma A = A$$

Ya que Γ es invertible. Además, A es semi definida positiva, en efecto, sea $x \in \mathbb{R}^n$:

$$x^T A x = x^T A\Gamma A x = \tilde{x}^T \Gamma \tilde{x} > 0$$

Pues Γ es s.d.p. Con esto en mente, volvemos a Q.

$$Q = Y^T A\Gamma A Y = Y^T A Y = Y^T A\Gamma\Gamma^{-1}\Gamma A Y = \|\Gamma A Y\|^2_{\Gamma^{-1}}$$

Similarmente

$$\|\Gamma A Y\|^2_{\Gamma^{-1}} = Q = \sum_{h=1}^{p} Y^T A_h Y = \sum_{h=1}^{p} Q_h = \sum_{h=1}^{p} Y^T A_h\Gamma A_h Y = \sum_{h=1}^{p} \|\Gamma A_h Y\|^2_{\Gamma^{-1}}.$$

Luego

$$\sum_{k,h=1;\ k\neq h}^{p} < \Gamma A_h Y, \Gamma A_k Y >_{\Gamma^{-1}} = 0$$

Como A_h y Γ son s.d.f. sigue que

$$\begin{array}{rccl} \forall h \neq k : & < \Gamma A_h Y, \Gamma A_k Y >_{\Gamma^{-1}} & = & 0 \\ \Rightarrow \forall h \neq k : & Y^T A_h \Gamma\Gamma^{-1}\Gamma A_k Y & = & Y^T A_h\Gamma A_k Y = 0 \\ \Rightarrow \forall h \neq k : & A_h\Gamma A_k & = & 0 \end{array}$$

$\Rightarrow \forall h \neq k : Q_h$ y Q_k son independientes.

Finalmente, (1)+(3) $\Rightarrow$ (2):
Como se tiene

$$Im(A\Gamma) = \sum_{h=1}^{p} Im(A_h\Gamma) \Rightarrow Im(A_h\Gamma) \subseteq Im(A\Gamma).$$

Luego si $Z = A_h\Gamma Y \Rightarrow Z \in Im(A\Gamma)$ y

$$A_h\Gamma Y = A\Gamma Z = A\Gamma A_h\Gamma Y = \sum_{k=1}^{p} A_k\Gamma A_h\Gamma Y = (A_h\Gamma)^2 Y.$$

■

Corolario 1.1 (Teorema de Cochran) *Dado el vector aleatorio $Y \sim N_n(0,\Gamma)$ con Γ regular, se considera la forma cuadrática $Q = Y^T AY$ con A simétrica de rango r y las p formas cuadráticas $Q_h = Y^T A_h Y$ con A_h simétrica de rango r_h ($1 \leq h \leq p$) tales que*

$$\left(A = \sum_{h=1}^{p} A_h\right), \quad Q = \sum_{h=1}^{p} Q_h$$

Entonces:

$$A\Gamma \text{ idempotente de orden } 2 \text{ y } \sum_{h=1}^{p} r_h = r \iff Q_h \sim \chi^2_{r_h} \wedge \forall h \neq k : Q_h \text{ y } Q_k \text{ independientes}.$$

1.5.2. La distribución $F_{m,n}$.

Se estudia el cociente de dos formas cuadráticas independientes.

Definición 1.7 *Si $U \sim \chi^2_m$ y $V \sim \chi^2_n$ con U y V independientes, se dice que $F = \dfrac{U/m}{V/n}$ sigue una forma distribución F de Fisher a m y n grados de libertad (se denota $F_{m,n}$).*
Se define $F_{m,n}$ no centrada cuando el numerador es no centrado.

Proposición 1.11 *Propiedades básicas:*

- *Su función densidad es* $f(x) = \frac{\Gamma((r+s)/2)r^{r/2}s^{s/2}x^{-1+r/2}}{\Gamma(r/2)\Gamma(s/2)(rx+s)^{(r+s)/2}}$.
- $\mathbb{E}(F_{m,n}) = \frac{n}{n-2}$, *si* $n > 2$.
- $\mathbb{V}(F_{m,n}) = \frac{2n^2(m+n-2)}{m(n-4)(n-2)^2}$, *si* $n > 4$.
- $F_{1,n} = t_n^2$.

1.6. Inferencia Estadística

1.6.1. Cantidad de información de Fisher y la Desigualdad de Cramer-Rao

La desigualdad de Cramer-Rao, que vamos a establecer, permite dar una cota inferior de la varianza de un estimador, lo cual es de bastante utilidad cuando no se tiene mucha información con respecto a esta cantidad. Esta cota se basa en la cantidad de la información de Fisher.

Para las definiciones en esta sección consideraremos una v.a. X con función de densidad o función de probabilidad $f(x|\theta)$ en donde θ es un parámetro desconocido en el conjunto Ω.

Definición 1.8 *Se llama* **cantidad de información de Fisher** *dada por X sobre el parámetro* θ *a la cantidad*

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial \ln f}{\partial \theta}\right)^2\right].$$

Además bajo ciertas condiciones es posible deducir otras dos formas equivalentes a la cantidad de información de Fisher, las cuales enunciamos:

Teorema 1.3 *Si el dominio S de X no depende de* θ*, entonces*

$$I(\theta) = \mathbb{V}\left(\frac{\partial \ln f}{\partial \theta}\right).$$

<u>Dem:</u> Sea S el dominio de X, entonces como

$$\forall\theta \in \Omega : \int_S f(x|\theta)dx = 1 \Rightarrow \forall\theta \in \Omega : \int_S f'(x|\theta)dx = 0.$$

Además $\dfrac{\partial \ln f}{\partial \theta} = \dfrac{f'}{f}$, luego $\mathbb{E}\left(\dfrac{\partial \ln f}{\partial \theta}\right) = 0$ y $\forall\theta \in \Omega : I(\theta) = \mathbb{V}\left(\dfrac{\partial \ln f}{\partial \theta}\right)$.

■

El teorema siguiente nos da otra expresión para $I(\theta)$, que a menudo es más fácil de determinar.

Teorema 1.4 *Si el dominio S de X no depende de* θ*, entonces:*

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ln f}{\partial \theta^2}\right]$$

si esta cantidad existe.

<u>Dem:</u> Si $\forall\theta \in \Omega : \dfrac{\partial^2 f}{\partial \theta^2}$ existe, al igual que antes, $\mathbb{E}\left(\dfrac{\partial^2 f}{\partial \theta^2}\right) = 0$. Además

$$\frac{\partial^2 \ln f}{\partial \theta^2} = \frac{f f'' - (f')^2}{f^2} = \frac{f''}{f} - \left(\frac{\partial \ln f}{\partial \theta}\right)^2.$$

Como

$$\mathbb{E}\left(\frac{\partial^2 \ln f}{\partial \theta^2}\right) = \int_S f''(x|\theta)dx - I(\theta),$$

se deduce que

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ln f}{\partial \theta^2}\right).$$

■

Las definición que dimos para la información de Fisher solo considera el caso de una variable aleatoria, enunciaremos ahora para el caso de una muestra aleatoria simple. Sea una m.a.s. $\{x_1, x_2, \ldots, x_n\}$, x_i con función de densidad o de probabilidad $f(x_i|\theta)$, donde θ es un parámetro desconocido en el conjunto Ω, esta descripción será utilizada a lo largo de todo este capítulo. Además sea L la función de verosimilitud de la muestra.

Definición 1.9 *Se llama* **cantidad de información de Fisher** *de una muestra aleatoria de tamaño n sobre el parámetro* θ *a la cantidad*

$$I_n(\theta) = \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right].$$

Donde

$$L = f(x_1, \ldots, x_n|\theta)$$

Además se tienen las mismas 2 expresiones equivalentes a $I_n(\theta)$ como en el caso de una v.a. X, de las cuales dejamos la demostración como ejercicio:

$$I_n(\theta) = \mathbb{V}\left(\frac{\partial \ln L}{\partial \theta}\right) = -\mathbb{E}\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right).$$

Con estas expresiones es fácil deducir que:

Teorema 1.5 *Si* $I(\theta)$ *es la cantidad de Fisher dada por cada* x_i *sobre el parámetro* θ, *entonces*

$$I_n(\theta) = nI(\theta).$$

Enunciamos a continuación el resultado más importante de esta sección. Sea una m.a.s, para esta tenemos la siguiente desigualdad

Teorema 1.6 (Desigualdad de Cramer-Rao) *Si el dominio* S *de* X *no depende de* θ, *para todo estimador* T *insesgado de* θ *se tiene:*

$$\mathbb{V}(T) \geq \frac{1}{I_n(\theta)}.$$

Además si T *es un estimador insesgado de* $h(\theta)$, *entonces*

$$\mathbb{V}(T) \geq \frac{(h'(\theta))^2}{I_n(\theta)}.$$

Dem: Utilizando el mismo argumento que se utilizo en la demostración del Teorema 1.3, se tiene que:

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial \ln L}{\partial \theta}\right) &= 0 \\
\Rightarrow \mathbb{C}ov\left(T, \frac{\partial \ln L}{\partial \theta}\right) &= \mathbb{E}\left(T\frac{\partial \ln L}{\partial \theta}\right) = \int_S T\frac{\partial \ln L}{\partial \theta} L dx = \int_S T \frac{\partial L}{\partial \theta} dx \\
&= \frac{\partial \mathbb{E}(T)}{\partial \theta} = h'(\theta)
\end{aligned}$$

Por otro lado, de la desigualdad de Cauchy-Schwartz se obtiene:

$$\left(\mathbb{C}ov\left(T, \frac{\partial \ln L}{\partial \theta}\right)\right)^2 \leq \mathbb{V}(T)\mathbb{V}\left(\frac{\partial \ln L}{\partial \theta}\right).$$

Es decir que

$$(h'(\theta))^2 \leq \mathbb{V}(T) I_n(\theta).$$

■

1.6.2. Test de hipótesis.

En general cuando se realizan estudios estadísticos se asumen ciertas condiciones sobre los datos, para poder aplicar según conste el caso, algún test sobre estos, de esta manera es importante saber si dichas hipótesis son satisfechas. Se formalizan en este capitulo las nociones de hipótesis estadistica y se dan a conocer los test más importantes para su comparación.

Considere un vector aleatorio $x = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^{p\times n}$ con distribución conjunta $F_n(x_1, x_2, \ldots, x_n; \theta)$ en donde θ es un parámetro vectorial de dimensión r que toma valores en Ω, una región abierta de $\mathbb{R}^r$. Sea también θ_0 el valor verdadero de θ en la población Ω_0, un subconjunto de Ω. Bajo este contexto pasamos a definir de manera general, las 2 hipótesis estadísticas, entre las cuales buscamos rechazar o no una de ellas.

Definición 1.10 *Consideremos la* **Hipótesis de Trabajo o Nula** $H_0 : \theta_0 \in \Omega_0$ *contra la* **Hipótesis alternativa** $H_1 : \theta_0 \in \Omega \setminus \Omega_0$. *Cuando* Ω_0 *está reducido a un solo punto se habla de hipótesis simple, sino de hipótesis compuesta..*

Notar que la unión de la region de trabajo con la de la hipótesis alternativa corresponde a todo el espacio al cual pertenece el parámetro, por lo que es razonable preguntarse **¿Con qué grado de desacuerdo uno tiene que abandonar la hipótesis nula para adoptar la hipótesis alternativa?**. Ante esto es necesario definir lo que en estadística se denomina una regla de decisión:

Definición 1.11 (Regla de decisión) *Sean una hipótesis nula y alternativa, dada una muestra de tamaño n en la cual los valores muestrales se encuentran en* $\mathbb{R}^p$*, una regla de decisión* δ *consiste en dividir el dominio* $\mathbb{R}^{p\times n}$ *del conjunto de todas las muestras de tamaño n en dos partes disjuntas. Una en la cual se rechaza la hipótesis nula y otra, su complemento, en la cual se rechaza la hipótesis alternativa.*

Dada una **regla de decisión**, δ, definimos $\alpha(\delta)$ como la probabilidad de equivocarse cuando la hipótesis nula es cierta y $\beta(\delta)$ la probabilidad de equivocarse cuando la hipótesis alternativa es cierta, con estas definiciones es razonable buscar entonces aquellas reglas de decisión que minimizan alguna función de estas probabilidades (a modo de ejemplo, puede ser la combinación convexa de estas). La manera de representar la idea de “equivocarse", nos la dan los conceptos de errores de tipo I y II, que enunciamos a continuación.

Definición 1.12 *Definimos el* **Error de tipo I** α*, si dada una hipótesis nula* H_0*,* $\alpha(\delta)$ *es la probabilidad condicional de rechazar la hipótesis* H_0 *con la regla* δ *cuando* H_0 *es cierta. De la misma manera definimos el* **Error de tipo II** β*, si dada una hipótesis alternativa* H_1*,* $\beta(\delta)$ *es la probabilidad de no rechazar* H_0 *(o bien rechazar* H_1*) con la regla de decisón* δ *cuando* H_0 *es cierta.*

Ahora bien la regla δ se basa en los valores muestrales, lo cual nos da pie para definir la **Regíon crítica del test**, entendiendo por esta la región en la cual se rechaza H_0, mas formalmente:

Definición 1.13 *Dada una regla de decisión* δ *y una muestra de tamaño n donde sus valores se encuentran en* $\mathbb{R}^p$*, la región donde se rechaza la hipótesis Nula* H_0*, se denomina* **Región crítica***, que denotamos por* W_n*, además denotamos por* $\overline{W_n}$ *la región donde se rechaza la hipótesis alternativa.*

Como la región crítica del test es aquella en donde se rechaza H_0, y eso sugiere aceptar la hipótesis alternativa, una regla de decisión consiste entonces en determinar la región crítica del test en función de las dos hipótesis.

Hasta ahora no hemos tomado en cuenta el parámetro θ, en términos de como cuantificamos el error que cometemos en función de este. La siguiente definición introduce la manera en la cual se hace en estadística:

Definición 1.14 *La función*

$$\pi(\theta) = \int_{W_n} dF_n = \mathbb{P}(\text{rechazar}\, H_0|\theta)$$

se llama **función de potencia** *del test.*

Obs:

La región crítica ideal es aquella que produce una función de potencia tal que:

$$\pi(\theta) = \begin{cases} 0 & \text{si}\, \theta \in \Omega_0 \\ 1 & \text{si}\, \theta \in \Omega \setminus \Omega_0 \end{cases} .$$

En efecto, para todo $\theta \in \Omega_0$ la decisión de rechazar H_0 es una decisión equivocada, entonces $\pi(\theta)$ es **una probabilidad de error de tipo I** (o riesgo de primer especie). Por otro lado, para todo $\theta \in \Omega \setminus \Omega_0$, la decisión de rechazar H_0 es una decisión correcta, entonces $1-\pi(\theta)$ es una **probabilidad de error de tipo II** (o riesgo de segundo especie).

El problema es que tal región crítica ideal no existe. Es por esta razón que se buscan propiedades más débiles, por ejemplo se fija un nivel de error aceptable y se les pide a los test que sean insesgados o consistentes, características que definimos a continuación:

Definición 1.15 *Se llama* **Nivel de significación** *del test al valor que uno se fija como cota máxima del error de tipo I.*

Definición 1.16 *Se dice que un test es* **Insesgado** *si dado un nivel de significación* α *se tiene a la vez* $\mathbb{P}(x \in W_n|\theta \in \Omega_0) \leq \alpha$ *y* $\mathbb{P}(x \in W_n|\theta \in \Omega \setminus \Omega_0) > \alpha$ *. Es decir que el error de tipo I está controlado y no está sobrepasado por el error de tipo II.*

Definición 1.17 *Se llama* **tamaño del test** *a* $\alpha_0 = \sup\{\pi(\theta)|\theta \in \Omega_0\}$.

Definición 1.18 *Si*

$$\forall \theta \in \Omega \setminus \Omega_0 : \lim_{n \to +\infty} \mathbb{P}(x \in W_n|\theta) = 1$$

entonces se dice que el test es **consistente** *de tamaño* α *para la hipótesis* H_0 *contra la hipótesis* H_1.

Definición 1.19 *Si* W_n *y* W_n^* *son dos regiones críticas para la hipótesis* H_0 *contra la hipótesis* H_1 *con un tamaño del test igual a* α*, se dice que* W_n^* *es* **Uniformemente Más Potente (U.M.P.)** *que* W_n *para* H_0 *contra* H_1 *si y sólo si:*

$$\forall \theta \in \Omega \setminus \Omega_0 : \mathbb{P}(x \in W_n^*|\theta) > \mathbb{P}(x \in W_n|\theta)$$

1.6.3. Caso de dos hipótesis simples

Buscamos en esta sección encontrar herramientas que nos permitan decidir sobre las hipótesis planteadas pero en un caso más simple. Consideramos aquí el caso de Ω_0 reducido a un solo punto: $\Omega_0 = \{\theta_0\}$ y $\Omega \setminus \Omega_0 = \{\theta_1\}$ reducido a un punto también: es decir $\Omega = \{\theta_0, \theta_1\}$. Bajo ciertas condiciones, existe un test insesgado y más potente para la hipótesis H_0 contra la hipótesis H_1. Este es el resultado importante en este capítulo, y que denominamos como el Lema de Neyman-Pearson.

Lema 1.1 (Lema de Neyman-Pearson) *Sea* $\{x_1, x_2, \ldots, x_n\}$ *una m.a.s. con función de verosimilitud* $f_n(x_1, x_2, \ldots, x_n|\theta)$ *y el espacio muestral* Ω *de* θ *con dos puntos* θ_0 *y* θ_1*. Sea* $W_n \subset \mathbb{R}^n$ *la región definida por:*

- $\mathbb{P}(x \in W_n|\theta_0) = \alpha$ *(Nivel de significación)*
- $f_n(x_1, x_2, \ldots, x_n|\theta_1) \geq c_\alpha f_n(x_1, x_2, \ldots, x_n|\theta_0)$ [3]

Si W_n^* *es un subconjunto de* $\mathbb{R}^n$ *tal que* $\mathbb{P}(x \in W_n^*|\theta_0) \leq \alpha$ *es decir con menor nivel de significación que* W_n*, entonces*

$$\mathbb{P}(x \in W_n|\theta_1) \geq \mathbb{P}(x \in W_n^*|\theta_1).$$

Lo cual se interpreta que la región crítica W_n *es insesgada y más potente que cualquier* W_n^* *para* $H_0 : \theta = \theta_0$ *contra* $H_1 : \theta = \theta_1$ *al nivel* α*.*

Obs: Gracias a este Lema es posible encontrar de manera explícita la región más potente para el test con hipótesis simples, tal como lo mostrarán los ejemplos más adelante para casos específicos de distribuciones.

Proposición 1.12 *Algunas propiedades del test son:*

- *El test es insesgado.*
- *El test es consistente.*
- *Cuando existe un estadístico suficiente* T *para* θ*, es decir cuando podemos escribir* $f_n(x_1, x_2, \ldots, x_n|\theta) = g(T, \theta)h(x_1, x_2, \ldots, x_n)$*, entonces el test se reduce a*

$$\frac{g(T, \theta_1)}{g(T, \theta_0)} > c_\alpha.$$

1.6.4. Test uniformemente más potente (UMP).

En el capítulo anterior encontramos un test con región crítica más potente en el caso de hipótesis simples. Nos enfocamos ahora en el caso con hipótesis compuestas.

Definición 1.20 *Se dice que un test es UMP (uniformemente más potente) cuando existe una región crítica óptima común para todo valor de la hipótesis alternativa* H_1*.*

Sea la hipótesis nula $H_0 : \theta = \theta_0$ y la hipótesis alternativa $H_1 : \theta > \theta_0$ (o $H_1 : \theta \neq \theta_0$). La región crítica óptima de nivel de significación α no cambia para todo $\theta > \theta_0$ pero si cambia para $\theta \neq \theta_0$.

La existencia de un test UMP está dada por el teorema de Lehmann, que enunciamos sin demostración:

Teorema 1.7 *Existe un test UMP si para un estadístico* T *el cociente*

$$\frac{f_n(x_1, x_2, \ldots, x_n|\theta_1)}{f_n(x_1, x_2, \ldots, x_n|\theta_2)}$$

es una función monótona creciente cuando $\theta_1 > \theta_2$*.*

Esta condición está asegurada con estadísticos suficientes T que tienen una distribución de tipo exponencial.

[3] La constante se denota c_α pues depende del nivel de confianza que coloquemos al test.

1.6.5. Test de razón de verosimilitudes.

Este test permite extender el caso anterior cuando no existe un test UMP. Sea la hipótesis nula $H_0 : \theta \in \Omega_0$ contra la hipótesis alternativa $H_1 : \theta \in \Omega_1$ con $\Omega = \Omega_0 \cup \Omega_1$. Se define la razón de verosimilitudes:

$$\Lambda = \frac{L(x,\Omega_0)}{L(x,\Omega)}$$

en donde

$$L(x,\Omega_0) = \sup_{\theta\in\Omega_0} f_n(x_1,x_2,\ldots,x_n|\theta)$$
$$L(x,\Omega) = \sup_{\theta\in\Omega} f_n(x_1,x_2,\ldots,x_n|\theta).$$

Obs:

- $\Lambda \in [0,1]$. Mientras más cerca Λ de 1, más verosímil es la hipótesis H_o.
- Fijado un nivel de significación α, la región crítica W_n satisface que $\mathbb{P}(x \in W_n|\theta_0 \in \Omega_0) = \alpha$ y satisface que $\Lambda \leq c_\alpha$[4], dado que H_0 es más aceptable cuando Λ se acerca a 1.

1.6.6. Ejemplos

Se presentan en esta sección algunos ejemplos ilustrativos de como utilizar los test y encontrar las regiones críticas en cada caso:

Ejemplo 1.1 Sea una m.a.s. $\{x_1,x_2,\ldots,x_n\}$ con $\forall i : x_i \sim N(\mu,\sigma^2)$. Se considera las hipótesis $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$. Aquí $\Omega = \mathbb{R}$ y $\Omega_0 = \{\mu_0\}$.

$$L(x|\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\Lambda = \frac{\max\limits_{\Omega_0} L(x|\mu,\sigma^2)}{\max\limits_{\Omega} L(x|\mu,\sigma^2)} = \frac{L(x|\mu_0,s_0^2)}{L(x|\bar{x},s_n^2)}$$

con

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$
$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2$$

Se sabe que son los máximos en esas cantidades pues son los estimadores de máxima verosimilitud, además esto implica que:

$$\Lambda = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\mu_0)^2}\right)^{\frac{n}{2}} = \left(\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^{-\frac{n}{2}} = \left(\frac{1+n(\bar{x}-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^{-\frac{n}{2}}$$
$$= \left(1+\frac{(\bar{x}-\mu_0)^2}{s_n^2}\right)^{-\frac{n}{2}} = \left(1+\frac{a^2}{n-1}\right)^{-\frac{n}{2}}$$

[4]Obs: la constante se denota c_α pues depende del nivel de confianza que coloquemos al test.

En donde a sigue una distribución t de Student a $n-1$ grados de libertad. El test de razón de verosimilitudes equivale en este caso al test t de Student.

Ejemplo 1.2 Sea el vector Y formado de una m.a.s. $\{y_1, y_2, \ldots, y_n\}$ con $\forall i : y_i \sim N(\mu, \sigma^2)$. Sea la matriz formada de vectores constantes $X \in M_{n,p}$ con $X = (X_0|X_1)$, $X_0 \in M_{n,p_0}$, $X_1 \in M_{n,p_1}$, $p = p_1 + p_2$ y $\beta = (\beta_0, \beta_1)^T$.

Sean las hipótesis $H_0 : E(Y) = X_0\beta_0$ contra $H_1 : E(Y) = X_1\beta_1$. Los conjuntos Ω y Ω_0 son subespacios vectoriales de $\mathbb{R}^p$: $\Omega = \text{Imagen}(X)$ y $\Omega_0 = \text{Imagen}(X_0)$.

$$L(Y|\mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\|Y-\mu\|^2}$$

Para el denominador:

$$\max_{E(Y)=X\beta} L(Y|\mu, \sigma^2) = \left(\frac{n}{2\pi\|Y - X\beta\|^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}$$

dado que en este caso el estimador de σ^2 es
$|Y - X\beta\|^2 \frac{1}{n}$.

Para el numerador:

$$\max_{E(Y)=X_0\beta_0} L(Y|\mu, \sigma^2) = \left(\frac{n}{2\pi\|Y - X_0\beta_0\|^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}$$

dado que en este caso el estimador de σ^2 es $\dfrac{\|Y - X_0\beta_0\|^2}{n}$.

Si $S_\Omega = \|Y - X\beta\|^2$ y $S_{\Omega_0} = \|Y - X_0\beta_0\|^2$, entonces

$$\Lambda = \left(\frac{S_\Omega}{S_{\Omega_0}}\right)^{\frac{n}{2}} = \left(\frac{S_{\Omega_0}}{S_\Omega}\right)^{-\frac{n}{2}}.$$

Del teorema de Cochran se obtiene que $\dfrac{S_\Omega}{\sigma^2} \sim \chi^2_{n-p}$ y $\dfrac{S_{\Omega_0} - S_\Omega}{\sigma^2} \sim \chi^2_{p_1}$ son independientes entre si. De aquí obtenemos el estadístico

$$F = \left(\frac{n-p}{p_1}\right) \frac{S_{\Omega_0} - S_\Omega}{S_\Omega} \sim F_{p_1, n-p}$$

bajo la hipótesis H_0 y

$$\Lambda = \left(1 + \frac{p_1}{n-p} F\right)^{-\frac{n}{2}}.$$

Es decir que

$$\frac{n-p}{p_1}(\Lambda^{-2/n} - 1) \sim F_{p_1, n-p}$$

bajo la hipótesis nula H_0.

En estos dos casos se puede fácilmente deducir una región crítica, pero en casos más generales, para encontrar el valor c_α y calcular la potencia del test se requiere conocer la distribución de Λ. Presentamos a continuación un resultado asintótico para la distribución de una función de la razón de verosimilitud.

Teorema 1.8 (Wilks) *Si $H_0 : \theta \in \Omega_0 \wedge H_1 : \theta \in \Omega_1$ con $\Omega_0 \subset \Omega_1 \subset \mathbb{R}^q$ espacios vectoriales de dimensión r y q, respectivamente, entonces bajo ciertas condiciones de regularidad (que en general se cumplen):*

$$-2\ln(\Lambda) \to \chi^2_{q-r},\ n \to \infty$$

1.7. Ejercicios.

1. Sea $X \in M_{p,q}$. Si Y es un vector de $\mathbb{R}^p$, encuentre el vector $Z \in \mathbb{R}^p$ de la forma Xb lo más cercano a Y con respecto a la métrica N.

2. Sea un vector $Y \in \mathbb{R}^p$ normal de media $E(Y) = \mu$ y de matriz de varianza-covarianza $Var(Y) = \Gamma$ invertible ($A = \Gamma^{-1}$). Sea $\{Y_1, Y_2, \ldots, Y_n\}$ una muestra aleatoria de realizaciones independientes del vector Y. Encuentre el estimador de máxima verosimilitud para μ y Γ.

3. Sean A y B dos matrices simétricas del mismo orden, B invertible. Muestre que el cociente $\dfrac{u^T Au}{u^T Bu}$ es máximo para el vector propio u de $B^{-1}A$, asociado al mayor valor propio.

4. Sea Γ una matriz cuadrada no invertible de rango r. Muestre que existe una inversa generalizada de Γ que puede escribirse como:
$$\Gamma^- = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$
en donde A es una matriz cuadrada invertible de orden r.

5. Demuestre que existe una única inversa generalizada de Penrose para una matriz dada.

6. Muestre el teorema siguiente:

   Teorema 1.9: Sea Y un vector aleatorio en $\mathbb{R}^p$ de matriz de varianza-covarianza Γ invertible. Consideramos la descomposición en suma directa: $\mathbb{R}^p = E_1 \oplus E_2$, $Y = Y_1 + Y_2$, con $Y_1 \in E_1$ e $Y_2 \in E_2$. Se llaman Γ_1 y Γ_2 las respectivas matrices de varianza-covarianza de Y_1 e Y_2. Entonces las dos propiedades siguientes son equivalentes:

   a) $\Gamma = \Gamma_1 + \Gamma_2$ (Y_1 e Y_2 son no correlacionados)
   b) E_1 y E_2 son Γ^{-1}-ortogonales.

7. Muestre que si $Y \sim N_p(\mu, \Gamma)$, A es una matriz simétrica de orden p, entonces
$$Var(Y^T AY) = 2Traza((A\Gamma)^2) + 4\mu^T A\Gamma A\mu.$$

8. Sea $Y \sim N_p(\mu, \Gamma)$, Γ regular, entonces si u y $v \in \mathbb{R}^p$, A y B son dos matrices simétricas de orden p y $L \in M_{m,p}$ se tienen las siguientes propiedades de independencia:

   a) $u^T Y$ y $v^T Y$ son independientes $\iff u^T \Gamma v = 0$
   b) LY e $Y^T AY$ son independientes $\iff L\Gamma A = 0$
   c) $Y^T AY$ e $Y^T BY$ son independientes $\iff A\Gamma B = 0$.

   Muestre la 3 proposiciones anteriores.

9. Aplique el teorema de Cochran para demostrar la independencia entre la media empírica y la varianza empírica de una variable normal.

10. Sea $X \in M_{n,p}$ la matriz que tiene en fila las realizaciones independientes $X_i \sim N_p(\mu, \Gamma)$. Muestre que
$$\frac{1}{n}D = \frac{1}{n}(X - 1_n\mu^T)^T(X - 1_n\mu^T)$$
es una estimación insesgada de Γ cuando el vector de medias μ es conocido y que
$$E(D^{-1}) = \frac{1}{n-p-1}\Gamma^{-1}$$
si $n - p - 1 > 0$.

11. Muestre que la cantidad de información de Fisher dada por una v.a. de Bernoulli sobre su parámetro p es $I(p) = \dfrac{1}{p(1-p)}$.

12. Muestre que la cantidad de información de Fisher dada por una v.a. $X \sim N(\mu, \sigma^2)$ sobre el parámetro μ desconocido y la varianza σ^2 conocida es $I(\mu) = \dfrac{1}{\sigma^2}$.

Bibliografía

[1] NDERSON T.W., An Introduction to Multivariate Statistical Analysis, Wiley.

[2] REIMAN L. et al (1993), Classification and Regression Trees, Chapman and Hall.

[3] HRISTENSEN R. (1990), Linear Models for Multivariate, Time Series and Spatial Data, Springer.

[4] OX D.R., SNELL E.G. (1992), Analysis of Binary Data, Chapman and Hall.

[5] RAPER N., SMITH H. (1998), Applied Regression Analysis, Wiley.

[6] UKUNAGA K. (1972), Introduction to Statistical Pattern Recognition, Academic Press.

[7] OLDSTEIN M., DILLON W. (1978), Discrete Discriminante Analysis, Wiley.

[8] OURIEROUX C. (1984), Econométrie des Variables Qualitatives, Economica.

[9] RAYBILL F.A. (1961), An Introduction Linear Statistical Models, McGraw-Hill.

[10] ASTIE T., TIBSHIRANI R., FRIEDMAN J. (2001), The Elements of Statistical Learning. Data Mining, Inference and Prediction, Springer.

[11] OCKING R. (1996), Methods and Applications of Linear Models, Wiley.

[12] OSMER D. W., LEMESHOW S. (2000), Applied Logistic Regression, Wiley.

[13] EBART L. (1979), Traitement des Donnés Statistiques, Dunod.

[14] ARDIA K. (1979) Multivariate Analysis, Academic Press.

[15] ILLER R.G. (1986), Beyond ANOVA, Basics of Applied Statistics, Wiley.

[16] ORRISON D.F. (1976), Multivariate Statistical Methods, McGraw-Hill.

[17] AO C.R. (1973), Linear Statistical Inference and its Applications, Wiley.

[18] AO C.R., TOUTENBURG H. (1995), Linear Models, Least squares and Alternatives, Springer

[19] AVISHANDER N, DIPAK K. D. (2002), A First Course in Linear Model Theory, Chapman and Hall.

[20] APORTA G. (1990), Probabilités, Analyse des Donnï¿½es et Statistique, Editions Technip.

[21] CHEFFE H. (1959), The Analysis of Variance, Wiley.

[22] EARLE S.R. (1971), Linear Models, Wiley.

[23] EBER G.A.F. (1977), Linear Regression Analysis, Wiley.

[24] OMASSONE R. et al. (1988), Discrimination et Classement, Masson.