



Estadística Multivariante

Complementos del álgebra matricial. Derivación matricial.

Trabajo B

Antonio R. Moya
Martín-Castaño
Elena Romero Contreras
Nuria Rodríguez
Barroso
Universidad de Granada
anmomar85@correo.ugr.es
elenaromeroc@correo.ugr.es
rbnuria6@gmail.com

Índice

1. Complementos del álgebra matricial.	2
1.1. Introducción	2
1.2. Operación Vec.	2
1.3. Producto Kronecker	3
1.3.1. Producto Kronecker y Vec. Relaciones de interés entre ambas operaciones.	5
1.3.2. Ejercicios	5
1.4. La matriz conmutación	9
1.4.1. Ejercicios	10
1.5. Producto * de dos matrices	13
1.5.1. Ejercicios	13
1.6. Operación Vech. Matrices de transición. La matriz duplicación	13
1.6.1. Ejercicios	16
2. Derivación matricial.	17
2.1. Introducción	17
2.2. Diferencial primera y jacobianos	17
2.2.1. Diferencial de una función vectorial	17
2.2.2. Diferencial de una función matricial	18
2.2.3. Ejercicios	20
2.3. Matrices jacobianas y derivadas matriciales	21
2.3.1. Ejercicios	23
2.3.2. Derivadas matriciales de funciones escalares de un vector	25
2.3.3. Derivadas matriciales de funciones escalares de matrices	25
2.3.4. Jacobianos de funciones vectoriales	27
2.4. Diferencial segunda y hessianos	27
2.4.1. Matriz hessiana para una función escalar de un vector	27
2.4.2. Matriz hessiana para una función vectorial de un vector	27
2.4.3. Diferencial segunda de una función escalar	27
2.4.4. Diferencial segunda de una función vectorial	28
2.4.5. Diferencial segunda de una función matricial	30
3. Aplicaciones	31
3.1. Estimación máximo verosímil de los parámetros de la distribución normal multivariante.	31
3.2. Distribución normal matricial	32

1. Complementos del álgebra matricial.

1.1. Introducción

A pesar de que la manipulación algebraica de vectores aleatorios puede ser abordada sin ninguna complicación, las matrices aleatorias suponen un esfuerzo extra ya que el resultado obtenido sobre las mismas debe generalizar al realizado sobre vectores. Además, el uso de matrices aleatorias en el Análisis Multivariante es muy importante dado que aparecen de forma natural, por ejemplo, al estimar la matriz de covarianzas de una población vectorial.

1.2. Operación Vec.

Para que el trato de las matrices aleatorias generalice a la manipulación de vectores aleatorios, debemos *vectorizar* las matrices, esto es, tratarlas como si se tratasen de vectores. Esta consideración se puede hacer teniendo en cuenta que los espacios $\mathbb{M}_{n \times q}$ y \mathbb{R}^{nq} son isomorfos.

Vemos que el objetivo puede ser muy cómodo a la hora de manipular matrices aleatorias, sin embargo, debemos entender primero bien las propiedades que unen las expresiones matriciales y las *vectorizadas*. Introducimos formalmente la definición de *Vec*

Definición 1.1. Sea \mathbf{X} una matriz de orden $n \times q$. Se define $\text{Vec}(\mathbf{X})$ como el vector de dimensión $nq \times 1$ formado al apilar las columnas de \mathbf{X} una tras otra, o sea, si notamos por columnas $\mathbf{X} = [x_1, x_2, \dots, x_q]$, entonces

$$\text{Vec}(\mathbf{X}) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}$$

Como ya habíamos comentado, la identificación entre matrices y las *vectorizadas* se basa en un resultado que formalizamos a continuación:

Teorema 1.1. La aplicación $\text{Vec}: \mathbb{M}_{n \times q} \rightarrow \mathbb{R}^{nq}$ es un isomorfismo de espacios vectoriales.

Este isomorfismo presentado en este teorema nos puede servir para calcular la esperanza matemática. Así, cuando sea más fácil calcular la esperanza de su vectorización, utilizando la propiedad que obtenemos el en ejemplo siguiente, bastaría con deshacer el cambio con el isomorfismo anterior.

EJEMPLO 1.1: Sean x_1, \dots, x_N vectores aleatorios p -dimensionales con igual media μ . Sea la matriz aleatoria $\mathbf{X}_{N \times X} = [x_1, \dots, x_N]^t$ y consideremos el vector

$$Vec(X^t) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

Entonces, si notamos $\mathbf{1}_N$ al vector N dimensional cuyas componentes son todas iguales a uno, se verifica

$$E[Vec(X^t)] = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} = Vec([\mu, \mu, \dots, \mu]) = Vec(\mu \mathbf{1}_N^t)$$

1.3. Producto Kronecker

Como ya hemos visto en el ejercicio anterior, utilizar la vectorización puede facilitar ciertos cálculos, como el de la esperanza de una matriz aleatoria. Sin embargo, hay ocasiones en las que es obligatorio usar la vectorización de matrices, como en el ejemplo siguiente:

EJEMPLO 1.2: En las hipótesis del ejercicio anterior, suponemos además que x_1, \dots, x_N son independientes y con igual matriz de covarianzas Σ . Entonces se define la matriz de covarianzas de \mathbf{X}^t como $Cov[Vec(X^t)]$ ya que dicha matriz contiene todas las matrices de covarianzas entre las columnas de \mathbf{X}^t . A partir de la definición de la matriz de covarianzas de un vector aleatorio, se verifica

$$Cov[Vec(X^t)] = E[[Vec(X^t) - E[Vec(X^t)]] [Vec(X^t) - E[Vec(X^t)]]^t] =$$

$$= E \begin{bmatrix} X_1 - \mu \\ X_2 - \mu \\ \vdots \\ (X_1 - \mu)^t \quad \dots \quad (X_N - \mu)^t \\ \vdots \\ X_N - \mu \\ \mu \end{bmatrix} = \begin{pmatrix} \Sigma & 0 & \dots & 0 \\ 0 & \Sigma & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & \dots & 0 & \Sigma \mu \end{pmatrix}$$

A raíz de la expresión obtenida en el ejemplo, vamos a definir el producto de Kronecker de matrices:

Definición 1.2. Sean $A_{m \times n}$ y $B_{p \times q}$ dos matrices. Se define el **producto Kronecker** de ellas como la matriz de dimensiones $mp \times nq$ siguiente

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & \dots & \dots & \dots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix} = (a_{ij}B)_{ij}; i = 1, \dots, m, j = 1, \dots, n$$

Observamos la utilidad de la definición anterior obteniendo que en el ejemplo anterior se tiene $Cov[Vec(X^t)] = I_N \otimes \Sigma$.

Aunque hemos utilizado el resultado anterior para justificar la introducción del producto Kronecker, no es una justificación formal. Otra de las justificaciones que podríamos haber dado para introducirlo podría ser el resolver el sistema de ecuaciones $x = (A \otimes B)y$, que es no singular.

Ahora bien, una vez presentada esta nueva operación nos proponemos estudiar sus propiedades. Al igual que el producto usual matricial estaba relacionado con la composición de aplicaciones lineales, este nuevo producto lo estará con el producto tensorial. Vemos algunas propiedades en el siguiente teorema:

Teorema 1.2. *Se verifican las siguientes propiedades:*

1. *Dados $\alpha, \beta \in \mathbb{R}$, $A_{m \times n}$ y $B_{p \times q}$, entonces*

$$(\alpha A) \otimes (\beta B) = \alpha\beta(A \otimes B) = \alpha\beta A \otimes B = A \otimes (\alpha\beta)B$$

2. *Dadas $A_{m \times n}$, $B_{m \times n}$, $C_{p \times q}$ y $D_{p \times q}$, entonces*

- a) $(A \otimes C) + (B \otimes C) = (A + B) \otimes C$

- b) $(A \otimes C) + (A \otimes D) = A \otimes (C + D)$

- c) $(A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D)$

3. *Dadas $A_{m \times n}$, $B_{p \times q}$, $C_{r \times s}$, entonces $(A \otimes B) \otimes C = A \otimes (B \otimes C)$*

4. *Dadas $A_{m \times n}$, $B_{n \times p}$, $C_{q \times r}$ y $D_{r \times s}$, entonces $(A \otimes C)(B \otimes D) = AB \otimes CD$*

5. *Dadas $A_{m \times m}$ y $B_{n \times n}$ son no singulares entonces $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$*

6. *Si $A_{m \times n}$ y $B_{p \times q}$, entonces $(A \otimes B)^t = A^t \otimes B^t$*

7. *Si $A_{m \times m}$ y $B_{n \times n}$ son ortogonales, entonces $A \otimes B$ es ortogonal*

8. *Si $A_{m \times m}$ y $B_{n \times n}$ son matrices triangulares (inferiores), entonces $A \otimes B$ es triangular superior (inferior).*

9. *Si $A_{m \times m}$ y $B_{n \times n}$ son definidas positivas, entonces $A \otimes B$ es definida positiva*

10. *Dadas $A_{m \times n} = [A_1, \dots, A_k]$ y $B_{p \times q}$, entonces $A \otimes B = [A_1 \otimes B, \dots, A_k \otimes B]$. En particular, si $a_{m \times 1}$ y $b_{p \times 1}$ son dos vectores se tiene que $a \otimes b^t = ab^t = b^t \otimes a$*

11. *Dadas $A_{m \times n} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ y $B_{p \times q}$, entonces $A \otimes B = \begin{pmatrix} A_{11} \otimes B & A_{12} \otimes B \\ A_{21} \otimes B & A_{22} \otimes B \end{pmatrix}$*

12. *Dadas $A_{m \times m}$ y $B_{n \times n}$, entonces $tr[A \otimes B] = tr[A]tr[B]$*

13. Sean $A_{m \times m}$ y $B_{n \times n}$ matrices reales con autovalores reales respectivos $\lambda_1, \dots, \lambda_m$ y μ_1, \dots, μ_n . Entonces $A \otimes B$ tiene como autovalores $\lambda_i \mu_j, i = 1, \dots, m; j = 1, \dots, n$. Como consecuencia $rg(A \otimes B) = rg(A)rg(B)$

14. Dadas $A_{m \times m}$ y $B_{n \times n}$, entonces $|A \otimes B| = |A|^n |B|^m$

1.3.1. Producto Kronecker y Vec. Relaciones de interés entre ambas operaciones.

En este apartado vamos a exponer varias propiedades que relacionan las dos operaciones recientemente introducidas.

Teorema 1.3. Se verifican las siguientes afirmaciones:

- Si $a_{n \times 1}$ y $b_{q \times 1}$ son dos vectores, entonces $Vec(ab^t) = b \otimes a$
- Sean $A_{n \times q}$, $B_{q \times p}$ y $C_{p \times r}$ tres matrices cualesquiera. Entonces $Vec(ABC) = (C^t \otimes A)Vec(B)$
- Sean $A_{n \times q}$ y $B_{q \times n}$. Entonces $tr[AB] = Vec'(A^t)Vec(B) = Vec'(B^t)Vec(A)$
- Sea $\{e_i : i = 1, \dots, n\}$ la base canónica de \mathbb{R}^n . Entonces

$$Vec(I_n) = \sum_{i=1}^n (e_i \otimes e_i)$$

- Sea $\{J_{ij} : i = 1, \dots, n : j = 1, \dots, q\}$ la base canónica del espacio de matrices $\mathbb{M}_{n \times q}$. Entonces

$$\sum_{i=1}^n \sum_{j=1}^q (J_{ij} \otimes J_{ij}) = Vec(I_n)Vec'(I_q)$$

1.3.2. Ejercicios

Ejercicio 1.1 Demostrar que la operación Vec define un isomorfismo de espacios vectoriales.

Solución:

Definimos la operación Vec como $Vec : \mathbb{M}_{n \times q} \rightarrow \mathbb{R}^{nq}$ con $Vec(X) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}$ donde $X = [x_1, \dots, x_q]$

Para ver que es isomorfismo tenemos que comprobar que:

1. Existe la inversa de Vec .
2. $Vec(aX + bY) = aVec(X) + bVec(Y)$ con $a, b \in \mathbb{R}$

En primer lugar, dado que $Vec : \mathbb{M}_{n \times q} \rightarrow \mathbb{R}^{nq}$ es $Vec : \mathbb{R}^{nq} \rightarrow \mathbb{M}_{n \times q}$ verificando:

$$Vec(Vec^{-1}(x)) = x \forall x \in \mathbb{R}^{nq}$$

$$Vec(Vec^{-1}(X)) = X \forall X \in \mathbb{M}_{n \times q}$$

Problema: saber que es nq y no qn , podemos solucionarlo con notación $Vec_{n,q}$

Comprobemos ahora la segunda condición:

Tenemos $X = [x_1, \dots, x_n], Y = [y_1, \dots, y_m] \Rightarrow aX + bY = [ax_1 + by_1, \dots, ax_n + by_n]$

$$\text{Entonces } Vec(aX + bY) = \begin{pmatrix} ax_1 + by_1 \\ ax_2 + by_2 \\ \vdots \\ ax_n + by_n \end{pmatrix} = a \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + b \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = aVec(X) + bVec(Y)$$

Ejercicio 1.2 Verificar las siguientes igualdades:

a) Sean $A_{m \times n}$ y $B_{n \times p}$. Entonces:

$$Vec(AB) = (B^t \otimes I_m)Vec(A) = (I_p \otimes A)Vec(B) = (B^t \otimes A)Vec(I_n).$$

Solución:

En primer lugar, demostremos de forma genérica la segunda propiedad del Teorema 1.3:

Sean $X_{m \times n}$, $Y_{n \times p}$ y $W_{p \times q}$ tres matrices cualesquiera. Entonces $Vec(XYW) = (W^t \otimes X)Vec(Y)$

Demostración:

Sea $W = [w_1, \dots, w_p]$, entonces la k -ésima columna de XYW es

$$\begin{aligned} (XYW)_{i,k} &= XYW_k = X \sum_{i=1}^p y_i w_{i,k} = [w_{1,k}X, w_{2,k}, \dots, w_{p,k}] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = [w_{1,k}X, w_{2,k}X, \dots, w_{p,k}X]Vec(Y) \\ &= ([w_{1,k}, w_{2,k}, \dots, w_{p,k}] \otimes X)Vec(Y) = (w_k^t \otimes X)Vec(Y) \end{aligned}$$

Ahora bien, considerando todas las columnas de W tenemos:

$$Vec(XYW) = \begin{bmatrix} (XYW)_{i,1} \\ (XYW)_{i,2} \\ \vdots \\ (XYW)_{i,q} \end{bmatrix} = \begin{bmatrix} W_1^t \otimes X \\ W_2^t \otimes X \\ \vdots \\ W_q^t \otimes X \end{bmatrix} Vec(Y) = (W^t \otimes X)Vec(Y)$$

Luego para nuestro ejercicio basta con considerar:

- $AB = I_m AB \Rightarrow Vec(AB) = Vec(I_m AB) = (B^t \otimes I_m)Vec(A)$
- $AB = AI_n B \Rightarrow Vec(AB) = Vec(AI_n B) = (B^t \otimes A)Vec(I_n)$
- $AB = ABI_p \Rightarrow Vec(AB) = Vec(ABI_p) = (I_p \otimes A)Vec(B)$

b) Sea $A_{m \times n}$. Entonces:

$$Vec(A) = (A^t \otimes I_m)Vec(I_m) = (I_n \otimes I_m)Vec(A) = (I_n \otimes A)Vec(I_n).$$

Solución:

Utilizando el resultado genérico ya demostrado basta considerar:

- $A = I_m I_m A \Rightarrow \text{Vec}(A) = \text{Vec}(I_m I_m A) = (A^t \otimes I_m) \text{Vec}(I_m)$
- $A = I_m A I_n \Rightarrow \text{Vec}(A) = \text{Vec}(I_m A I_n) = (I_n \otimes I_m) \text{Vec}(A)$
- $A = A I_n I_n \Rightarrow \text{Vec}(A) = \text{Vec}(A I_n I_n) = (I_n \otimes A) \text{Vec}(I_n)$

c) Sean $A_{m \times n}$, $B_{n \times p}$ y $C_{p \times q}$. Entonces

$$\text{Vec}(ABC) = (C^t \otimes A) \text{Vec}(B) = (C^t B^t \otimes I_m) \text{Vec}(A) = (C^t \otimes I_m) \text{Vec}(AB) = (I_q \otimes AB) \text{Vec}(C).$$

Solución:

De nuevo utilizamos el resultado demostrado en el apartado (a).

- La primera igualdad es aplicar directamente el resultado.
- $ABC = I_m A(BC) \Rightarrow \text{Vec}(ABC) = \text{Vec}(I_m A(BC)) = (C^t B^t \otimes I_m) \text{Vec}(A)$
- $ABC = I_m (AB)C \Rightarrow \text{Vec}(ABC) = \text{Vec}(I_m (AB)C) = (C^t \otimes I_m) \text{Vec}(AB)$
- $ABC = (AB)C I_q \Rightarrow \text{Vec}(ABC) = \text{Vec}((AB)C I_q) = (I_q \otimes AB) \text{Vec}(C)$

d) Sean $B_{m \times n}$, $C_{n \times p}$ y $D_{p \times m}$. Entonces

$$\text{tr}[BCD] = \text{Vec}'(B^t)(I_n \otimes C) \text{Vec}(D) = \text{Vec}'(D^t)(C^t \otimes I_m) \text{Vec}(B).$$

Solución:

En primer lugar, demostraremos de forma genérica la tercera propiedad del teorema 1.3:

$$\text{tr}[X^t Y] = \text{Vec}'(X^t) \text{Vec}(Y)$$

Demostración:

$$\begin{aligned} X_{m \times n} Y_{m \times n} \Rightarrow X^t Y &= \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{1n} & & & x_{mn} \end{pmatrix}_{n \times m} \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ y_{m1} & & & y_{nm} \end{pmatrix}_{m \times n} \\ &= \begin{pmatrix} \sum_{i=1}^m x_{i1}^t y_{i1} & & & \\ & \sum_{i=1}^m x_{i2}^t y_{i2} & & \\ & & \ddots & \\ & & & \sum_{i=1}^m x_{in}^t y_{in} \end{pmatrix} = \sum_{j=1}^n \sum_{i=1}^m x_{ij}^t y_{ij} \end{aligned}$$

Por otro lado,

$$\text{Vec}(X^t)^t \text{Vec}(Y) = \sum_{i=1}^n x_{:,i} y_{i,:} = \sum_{j=1}^m \sum_{i=1}^n x_{ji} y_{ij} = \sum_{j=1}^m \sum_{i=1}^n x_{ij}^t y_{ij}$$

que es la igualdad que buscábamos.

En nuestro ejercicio $X^t = B$, $Y = CD$ y usando $CD = CD I_m$ tenemos que:

$$\text{tr}[BCD] = \text{Vec}'(B^t) \text{Vec}(CD) = \text{Vec}'(B^t)(I_m \otimes C) \text{Vec}(D)$$

Para la segunda igualdad sabemos que $\text{tr}(A) = \text{tr}(A^t) \forall A \in \mathbb{M}_{p \times q}$ y que

$$\text{tr}[X^t Y] = \text{tr}[XY^t] = \text{tr}[Y^t X] = \text{tr}[YX^t]$$

Usamos $X^t = BC$ y $Y = D$ y tenemos que:

$$\text{tr}[X^t Y] = \text{tr}[DBC] = \text{Vec}'(D^t)(C^t \otimes I_n)\text{Vec}(B)$$

e) Sea $A_{n \times n}$. Entonces $\text{tr}[A] = \text{Vec}'(A^t)\text{Vec}(I_n) = \text{Vec}'(I_n)\text{Vec}(A)$.

Solución:

Aplicamos la propiedad demostrada en el apartado anterior y usamos que $A = I_n A = A I_n$. Así, tenemos que

$$\text{tr}[A] = \text{tr}[A I_n] = \text{Vec}'(A^t)\text{Vec}(I_n) = \text{Vec}'(I_n)\text{Vec}(A)$$

f) Sean $A_{m \times n}$, $B_{m \times n}$, $C_{n \times p}$ y $D_{n \times p}$. Entonces

$$\text{Vec}((A+B)(C+D)) = [(I_p \otimes A) + (I_p \otimes B)][\text{Vec}(C) + \text{Vec}(D)] = [(C^t \otimes I_m) + (D^t \otimes I_m)][\text{Vec}(A) + \text{Vec}(B)].$$

Solución:

Desarrollamos la primera expresión y usamos la propiedad demostrada en el apartado (a) se tiene:

$$\begin{aligned} \text{Vec}((A+B)(C+D)) &= \text{Vec}(AC + AD + BC + BD) = \text{Vec}(AC) + \text{Vec}(AD) + \text{Vec}(BC) + \text{Vec}(BD) \\ &= \text{Vec}(AC I_p) + \text{Vec}(AD I_p) + \text{Vec}(BC I_p) + \text{Vec}(BD I_p) \\ &= (I_p \otimes A)\text{Vec}(C) + (I_p \otimes A)\text{Vec}(D) + (I_p \otimes B)\text{Vec}(C) + (I_p \otimes B)\text{Vec}(D) \\ &= [(I_p \otimes A) + (I_p \otimes B)][\text{Vec}(C) + \text{Vec}(D)] \end{aligned}$$

De forma análoga, si usamos que $AC = I_n AC$ en lugar de $AC I_p$ e para los demás pares de matrices, obtenemos la segunda igualdad aplicando la propiedad del apartado (a).

Ejercicio 1.3 Sea la matriz aleatoria $X_{N \times p}$ de los ejemplos 1.1 y 1.2. Sea $Y_{r \times s} = B_{r \times p} X^t C_{N \times s}$. Probar que

a) $E[\text{Vec}(Y)] = C^t 1_N \otimes B\mu$

Solución:

$$\text{Como } Y = B X^t C \Rightarrow \text{Vec}(Y) = \text{Vec}(B X^t C) = (C^t \otimes B)\text{Vec}(X^t).$$

$$\text{Así, tenemos que: } E[\text{Vec}(Y)] = E[(C^t \otimes B)\text{Vec}(X^t)] = (C^t \otimes B)E[\text{Vec}(X^t)] = (C^t \otimes B)\text{Vec}(\mu 1_N^t)$$

Usamos ahora la primera propiedad del Teorema 1.3 y la cuarta del Teorema 1.2:

$$(C^t \otimes B)\text{Vec}(\mu 1_N^t) = (C^t \otimes B)(1_N \otimes \mu) = C^t 1_N \otimes B\mu$$

b) $\text{Cov}[\text{Vec}(Y)] = C^t C \otimes B \Sigma B^t$

$$\text{De la misma forma que en el apartado anterior, se tiene } \text{Cov}[\text{Vec}(Y)] = \text{Cov}[(C^t \otimes B)\text{Vec}(X^t)]$$

Usando ahora cierta propiedad de la covarianza y sustituyendo por la expresión deducida del ejemplo 1.2:

$$\begin{aligned} \text{Cov}[(C^t \otimes B)\text{Vec}(X^t)] &= (C^t \otimes B)\text{Cov}[\text{Vec}(X^t)](C^t \otimes B)^t = (C^t \otimes B)(I_N \otimes \Sigma)(C^t \otimes B)^t \\ &= (C^t I_N \otimes B \Sigma)(C \otimes B^t) = C^t C \otimes B \Sigma B^t \end{aligned}$$

1.4. La matriz conmutación

En esta sección vamos a introducir un tipo de matriz que toma diversos nombres dado a que ha aparecido a lo largo de la historia de diferentes maneras. Nosotros tomaremos la notación de *matriz conmutación*.

Una característica común de todas las introducciones de esta matriz es la propiedad de reordenar los elementos de una matriz. Una propiedad importante es la de transformar $Vec(A)$ en $Vec(A^t)$ siendo A una matriz arbitraria. También sirve para invertir un producto Kronecker, siendo una propiedad muy utilizada en la derivación matricial.

Definamos formalmente lo que venimos introduciendo:

Definición 1.3. Sea $A_{n \times q}$. Se define la **matriz de conmutación** como la matriz de permutación, K_{nq} , de dimensión $nq \times nq$ tal que

$$K_{nq}Vec(A) = Vec(A^t)$$

Notar que dado que K_{nq} es una matriz de permutación, es ortogonal, esto es $K_{nq}^t = K_{nq}^{-1}$. Además, claramente se tiene que $K_{qn}K_{nq}Vec(A) = K_{qn}Vec(A^t) = Vec(A)$, por lo que se obtiene que $K_{qn}K_{nq} = I_{nq}$, por lo que $K_{nq}^{-1} = K_{qn} = K_{nq}^t$. Además, es fácilmente comprobable que $K_{n1} = I_n$ y $K_{1q} = I_q$.

Para saber cómo se expresa de forma explícita la matriz de conmutación introducimos el siguiente teorema.

Teorema 1.4. Sea $\{J_{ij} : i = 1, \dots, n; j = 1, \dots, q\}$ la base canónica del espacio de matrices $\mathbb{M}_{n \times q}$. Entonces

$$K_{nq} = \sum_{i=1}^n \sum_{j=1}^q J_{ij} \otimes J_{ij}^t$$

Otra introducción de la matriz de conmutación es de forma constructiva, es decir, introduciéndola como aquella matriz $A_{n \times q}$ que verifica que

$$Vec(A^t) = \sum_{i=1}^n \sum_{j=1}^q (J_{ij} \otimes J_{ij}^t)Vec(A)$$

Tengamos en cuenta ahora que el resultado de $I_n \otimes I_q$ es I_{nq} . Así, podemos dar una nueva definición de la matriz de conmutación como la n -permutación por filas de la matriz $I_n \otimes I_q$, con lo que se asegura que $Vec(A^t) = K_{nq}Vec(A)$.

Notemos que con vistas a esta definición se puede usar la notación $I_{(q,n)}$ para hacer referencia a K_{nq} . Además se guarda cierta relación con K_{qn} que no deja de ser la q -permutación por filas de $I_n \otimes I_q$. Veamos ahora el siguiente teorema:

Teorema 1.5. Sean $\{e_i : i = 1 \dots n\}$ y $\{u_j : j = 1 \dots q\}$ las bases canónicas de \mathbb{R}^n y \mathbb{R}^q respectivamente. Entonces:

$$K_{nq} = \sum_{j=1}^q (u_j^t \otimes I_n \otimes u_j) = \sum_{i=1}^n e_i \otimes I_q \otimes e_i^t$$

Veamos ahora algunas propiedades de esta matriz:

Teorema 1.6. Consideramos las matrices $A_{m \times n}$, $B_{p \times q}$, $C_{q \times s}$, $D_{n \times t}$, $E_{m \times n}$ y los vectores $a_{m \times 1}$, $b_{p \times 1}$ y z un vector cualquiera. Entonces:

1. $K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$
2. $z^t \otimes A \otimes B = K_{mp}(bz^t \otimes A)$.
3. $\text{tr}[K_{mn}(A^t \otimes E)] = \text{tr}[A^t E] = \text{Vec}'(A^t)K_{mn}\text{Vec}(E)$
4. $\text{tr}[K_{mn}] = 1 + f(m-1, n-1)$, donde:

$$f(m-1, n-1) = \begin{cases} m.c.d(m-1, n-1) \\ f(0, n) = f(n, 0) = n \end{cases}$$

5. Los autovalores de K_{nn} son 1 con multiplicidad $\frac{n(n+1)}{2}$ y -1 con multiplicidad $\frac{n(n-1)}{2}$.
6. $|K_{nn}| = (-1)^{\frac{n(n-1)}{2}}$ y $|K_{mn}| = (-1)^{\frac{m(m-1)n(n-1)}{4}}$
7. Supongo que $\text{rg}(A) = r$ y $\lambda_1 \dots \lambda_r$ los autovalores de $A^t A$. Sea $P = K_{mn}(A^t \otimes A)$. Entonces P es una matriz simétrica con rango r^2 y que cumple: $\text{tr}[P] = \text{tr}[A^t A]$. Además se verifica que $P^2 = (AA^t) \otimes (A^t A)$ y sus autovalores son los anteriores más $\pm(\lambda_i \lambda_j)^{\frac{1}{2}}$ con $i \neq j$.

Ahora podemos plantearnos como realizar la vectorización de un producto de Kronecker. Para ello mostramos el siguiente resultado:

Teorema 1.7. Sean $A_{m \times n}$, $B_{p \times q}$. Entonces $\text{Vec}(A \otimes B) = [I_n \otimes K_{qm}I_p][\text{Vec}(A) \otimes \text{Vec}(B)]$.

Vamos a llevar la matriz de conmutación un paso más adelante. Para ello vamos a definirla para más de dos índices $K_{st,n}$ como se hacía anteriormente pero tomando $st = m$. Su actuación se ve como:

$$A \otimes B \otimes C = K_{ms,p}(C \otimes A \otimes B)K_{q,nt} = K_{m,sp}(B \otimes C \otimes A)K_{tq,n}$$

Así, podemos enunciar el siguiente teorema en el que se muestran algunas propiedades:

Teorema 1.8. Se cumple:

1. $K_{mn,p} = K_{nm,p}$ y $K_{m,np} = K_{m,pn}$
2. $K_{mn,p} = (I_m \otimes K_{np})(K_{mp} \otimes I_n) = (I_n \otimes K_{mp})(K_{np} \otimes I_m)$
3. $K_{m,np} = (K_{mn} \otimes I_p)(I_n \otimes K_{mp}) = (K_{mp} \otimes I_n)(I_p \otimes K_{mn})$
4. $K_{mn,p} = K_{m,np}K_{n,pm} = K_{n,mp}K_{m,pn}$ y $K_{m,np} = K_{mn,p}K_{pm,n} = K_{mp,n}K_{nm,p}$
5. $I_{nmp} = K_{mn,p}K_{pm,n}K_{np,m} = K_{mn,p}K_{pn,m}K_{mp,n} = K_{m,np}K_{n,pm}K_{p,mn} = K_{m,np}K_{p,nm}K_{n,mp}$
6. Cualquier producto de dos matrices de este tipo con el mismo conjunto de índices conmuta.

1.4.1. Ejercicios

Ejercicio 1.4 A partir de la expresión explícita de la matriz K_{nn} , demostrar que $\text{tr}[K_{nn}] = n$.

Solución:

Tengamos en cuenta que $K_n n = \sum_{i=1}^n \sum_{j=1}^n n(J_{ij} \otimes J_{ij}^t)$. Por tanto:

$$\text{tr}[K_n n] = \text{tr}\left[\sum_{i=1}^n \sum_{j=1}^n n(J_{ij} \otimes J_{ij}^t)\right] = \sum_{i=1}^n \sum_{j=1}^n \text{tr}[J_{ij} \otimes J_{ij}^t] = \sum_{i=1}^n \sum_{j=1}^n \text{tr}[J_{ij}] \text{tr}[J_{ij}^t]$$

Puesto que $\text{tr}[J_{ij}] = \text{tr}[J_{ij}^t]$ y

$$\text{tr}[J_{ij}] = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}$$

Entonces se cumple que

$$\text{tr}[K_n n] = \sum_{i=1}^n 1 = n$$

Pues solo se suma cuando coincidan i y j.

Ejercicio 1.5 Sean $A_{m \times n}$, $B_{p \times q}$, $C_{q \times s}$, $D_{n \times t}$ y $b_{p \times 1}$. Demostrar las siguientes igualdades:

a) $K_{pm}(A \otimes B)K_{nq} = B \otimes A$.

Solución:

Sabemos que $K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$. Por tanto:

$$K_{pm}(A \otimes B)K_{nq} = (B \otimes A)K_{qn}K_{nq} = (B \otimes A)K_{qn}K_{qn}^{-1}$$

Por tanto $K_{pm}(A \otimes B)K_{nq} = B \otimes A$.

b) $K_{pm}(A \otimes b) = b \otimes A$.

Solución:

Si consideramos que ahora $q = 1$, se saca que en particular: $K_{pm}(A \otimes b) = b \otimes A$

c) $K_{mp}(b \otimes A) = A \otimes b$.

Solución

Partiendo del apartado anterior: $K_{pm}(A \otimes b) = b \otimes A$.

Por tanto:

$$K_{mp}K_{pm}(A \otimes b) = K_{mp}(b \otimes A)$$

Y entonces: $A \otimes b = K_{mp}(b \otimes A)$.

Ejercicio 1.6 Sea $N_n = \frac{1}{2}[I_{n^2} + K_{nn}]$. Probar las siguientes afirmaciones:

a) N_n es simétrica e idempotente.

Solución: Primero veamos que es simétrica. Consideramos N_n^t :

$$N_n^t = \frac{1}{2}(I_{n^2} + K_{nn})^t = \frac{1}{2}(K_{nn}^t + I_{n^2}^t) = \frac{1}{2}(K_{nn} + I_{n^2}) = N_n$$

Por tanto es simétrica. Veamos ahora que es idempotente, esto es: $N_n^2 = N_n$:

$$N_n^2 = \frac{1}{4}(I_{n^2} + K_{nn})(I_{n^2} + K_{nn}) = \frac{1}{4}(I_{n^2}^2 + 2K_{nn} + K_{nn}^2) = \frac{1}{4}(2I_{n^2} + 2K_{nn}) = \frac{1}{2}(I_{n^2} + K_{nn}) = N_n$$

Por lo tanto es idempotente.

b) $rg(N_n) = tr[N_n] = \frac{1}{2}n(n+1)$

Solución:

Primero veamos la parte de la traza:

$$tr[N_n] = \frac{1}{2}tr[(I_{n^2} + K_{nn})] = \frac{1}{2}(tr[I_{n^2}] + tr[K_{nn}]) = \frac{1}{2}(n^2 + n) = \frac{1}{2}n(n+1)$$

Para el rango nos vale con observar la forma de N_n :

$$\begin{pmatrix} 2 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 1 & \cdots & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 1 & 0 & \cdots & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 1 & \cdots & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & 2 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 1 & 0 & \cdots & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 2 \end{pmatrix}$$

De las n primeras filas, podemos observar como la segunda coincide con la $n+1$, la tercera con la $2n+1$, etc, luego podemos eliminar de las n primeras (para el rango), todas menos 1. Cuando nos vamos al segundo bloque de n filas, las dos primeras no tienen coincidencia pero el resto sí por lo que podemos eliminar $n-2$. Así para los n bloques. Por tanto, $rg(N_n) = n^2 - \sum_{i=1}^n (n-i) = n^2 - \frac{n^2}{2} + \frac{n}{2} = \frac{n^2}{2} + \frac{n}{2} = \frac{1}{2}n(n+1)$.

Por tanto se deduce el resultado que se pedía demostrar.

c) $N_n K_{nn} = K_{nn} N_n = N_n$

Solución:

Veamos que $N_n K_{nn} = \frac{1}{2}(I_{n^2} + K_{nn})K_{nn} = \frac{1}{2}(K_{nn} + K_{nn}^2) = \frac{1}{2}(K_{nn} + I_{n^2}) = N_n$

Por otra parte: $K_{nn} N_n = K_{nn} \frac{1}{2}(I_{n^2} + K_{nn}) = \frac{1}{2}(K_{nn} + K_{nn}^2) = \frac{1}{2}(K_{nn} + I_{n^2}) = N_n$

Por tanto: $N_n K_{nn} = K_{nn} N_n = N_n$

1.5. Producto * de dos matrices

Para simplificar cálculos en las derivadas matriciales, se va a introducir un nuevo producto:

Definición 1.4. Sean $X_{m \times n}$ e $Y_{mp \times nq}$. Particionamos Y como sigue:

$$Y = \begin{pmatrix} Y_{11} & \cdots & \cdots & Y_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{m1} & \cdots & \cdots & Y_{mn} \end{pmatrix} \text{ donde } Y_{ij} \text{ es de dimensión } p \times q$$

Así, se define $X * Y = \sum_{i=1}^m \sum_{j=1}^n x_{ij} Y_{ij}$

En los dos siguientes teoremas vemos propiedades de esta operación:

Teorema 1.9. Sean $X_{m \times n}$, $Y_{mp \times nq}$, $W_{mp \times nq}$, $Z_{s \times t}$. Se cumple:

1. $(X * Y)^t = X * Y^t$
2. Si $p = q = 1$, entonces $X * Y = \text{tr}[XY^t] = \text{tr}[YX^t] = \text{tr}[X^tY] = \text{tr}[Y^tX]$
3. $(X * Y) \otimes Z = X * (Y \otimes Z)$
4. $X * (Y + W) = X * Y + X * W$

Teorema 1.10. Sean $X_{m \times n}$, $Y_{n \times p}$, $Z_{p \times q}$:

1. $XYZ = Y * \text{Vec}(X) * \text{Vec}'(Z^t)$
2. $XYZ = Y^t * (Z \otimes I_m) K_{qm}(X \otimes I_q)$

1.5.1. Ejercicios

Ejercicio 1.7 Sean $X_{m \times n}$, $Y_{n \times p}$. Demostrar que $Y * K_{np} = Y^t$

Solución:

Sabemos que $I_p Y^t I_n = Y^t$. Ahora:

$$\begin{aligned} I_p Y^t I_n &= Y * (I_n \otimes I_p) K_{np} (I_p \otimes I_n) = Y * I_{np} K_{np} I_{np} = \\ &= Y * K_{np} \end{aligned}$$

Por tanto $Y^t = Y * K_{np}$.

1.6. Operación Vech. Matrices de transición. La matriz duplicación

Vamos a realizar un planteamiento que nos lleve a un nuevo operador (**Vech**). Para ello, si consideramos $A_{n \times n}$, se puede ver $\text{Vech}(A)$ como el vector obtenido a partir de $\text{Vec}(A)$ eliminando los elementos de la diagonal superior de A. Por ejemplo si A es simétrica, entonces $\text{Vech}(A)$ contiene sus elementos distintos.

Es claro que se puede realizar una transformación que lleve de $Vec(A)$ a $Vech(A)$ mediante el producto por una matriz D_n . A esta matriz la llamaremos la **matriz de duplicación**.

Así se puede ver esto como un nuevo operador. Sea $A_{m \times n}$ una matriz con s elementos repetidos y/o v elementos constantes, el número de elementos matemáticamente independientes se puede calcular como sigue: $r = mn - s - v$.

Algunos ejemplos son los siguientes:

EjemPLO 1.3: ■ Si A es simétrica, $v = 0$, $s = \frac{m(m-1)}{2}$ y por tanto $r = \frac{m(m+1)}{2}$.

■ Si A es antisimétrica, $v = m$, $s = \frac{m(m-1)}{2}$ y $r = \frac{m(m-1)}{2}$

■ Si A es diagonal, $v = \frac{m(m-1)}{2}$, $s = 0$ y $r = m$

■ Si A es una matriz de correlación, entonces $v = m$, $s = \frac{m(m-1)}{2}$ y $r = \frac{m(m-1)}{2}$

Así en estas matrices $Vech(A)$ se podría ver como el vector eliminando los elementos repetidos y/o constantes. De este modo, podemos establecer una relación entre $Vec(A)$ y $Vech(A)$, pues ordenan los elementos de la matriz A salvo excluidos. Esta relación se consigue a través de una matriz que se denomina **matriz de transición**. En el caso de matrices simétricas la conocemos como **matriz de duplicación**.

Definición 1.5. Sea $A_{m \times n}$ y r el número de elementos de A matemáticamente independientes en el sentido comentado anteriormente. Se define la matriz de transición como aquella matriz \mathbf{TR} de orden $mn \times r$ que verifique la relación $\mathbf{TR}Vec(A) = Vec(A)$.

Veamos cuál es su forma explícita en el siguiente teorema.

Teorema 1.11. Sea A una matriz simétrica de orden n . Entonces las columnas de D_n vienen dadas por $e_{(j-1)n+i} + e_{(i-1)n+j}(1 - \delta_{ij})$, $i=1, \dots, n$, $j=i, \dots, n$, donde los vectores anteriores son vectores básicos de \mathbb{R}^{n^2} y δ_{ij} representa la función delta de Kronecker.

Demostración:

Sea X una matriz simétrica de orden $n \times n$, entonces $X = \sum_{i=1}^n \sum_{j=1}^n x_{ij} [J_{ij} + J_{ji}(1 - \delta_{ij})]$.

Por tanto, $Vec(X) = \sum_{i=1}^n \sum_{j=1}^n x_{ij} [e_{(j-1)n+i} + e_{(i-1)n+j}(1 - \delta_{ij})]$ y las columnas de la matriz de transición

son $e_{(j-1)n+i} + e_{(i-1)n+j}(1 - \delta_{ij})$ para $i = 1, \dots, n$, $j = i, \dots, n$. Es de dimensión $n^2 \times \frac{n(n+1)}{2}$.

□

Se comprueba fácilmente que D_n tiene rango por columnas completo y su g-inversa viene dada por $(D_n)_g = (D_n^t D_n)^{-1} D_n^t$.

Teorema 1.12. Sean $b_{n \times 1}$ y $A_{n \times n}$. Entonces se verifica

1. $K_{nn} D_n = D_n$.

2. $D_n (D_n)_g = N_n$.

$$3. D_n(D_n)_g(b \otimes A) = \frac{1}{2}(b \otimes A + A \otimes b).$$

$$4. D_n(D_n)_g(A \otimes A)D_n = (A \otimes A)D_n.$$

$$5. D_n(D_n)_g(A \otimes A)(D_n)_g^t = (A \otimes A)(D_n)_g^t.$$

donde $N_n = \frac{1}{2}[I_{n^2} + K_{nn}]$.

El siguiente teorema nos proporciona las expresiones de las matrices de transición en el caso de matrices diagonales, triangulares y antisimétricas.

Teorema 1.13. Si denotamos por \mathbf{Di}_n , \mathbf{TSup}_n , \mathbf{TInf}_n y \mathbf{Ant}_n a las matrices de transición correspondientes a una matriz diagonal, triangular superior, triangular inferior y antisimétrica, respectivamente, entonces se verifica:

1. Las columnas de \mathbf{Di}_n vienen dadas por $e_{(i-1)n+i}$, $i=1, \dots, n$.
2. Las columnas de \mathbf{TSup}_n vienen dadas por $e_{(j-1)n+i}$, $i=1, \dots, n$, $j=i, \dots, n$.
3. Las columnas de \mathbf{TInf}_n vienen dadas por $e_{(j-1)n+i}$, $i=1, \dots, n$, $j=1, \dots, i$.
4. Las columnas de \mathbf{Ant}_n vienen dadas por $e_{(j-1)n+i} - e_{(i-1)n+j}$, $i=1, \dots, n-1$, $j=i+1, \dots, n$.

donde los vectores e_k son vectores básicos de \mathbb{R}^{n^2} .

Demostración:

1. Sea $X_{n \times n}$ diagonal, entonces $X = \sum_{i=1}^n x_{ii} J_{ii}$. Por tanto, $Vec(X) = \sum_{i=1}^n x_{ii} e_{(i-1)n+i}$ y las columnas de la matriz de transición son $e_{(i-1)n+i}$, $i = 1, \dots, n$. Es de dimensión $n^2 \times n$.
2. Sea $X_{n \times n}$ triangular superior, entonces $X = \sum_{i=1}^n \sum_{j=i}^n x_{ij} J_{ij}$. Por tanto, $Vec(X) = \sum_{i=1}^n \sum_{j=i}^n x_{ij} e_{(j-1)n+i}$ y las columnas de la matriz de transición son $e_{(j-1)n+i}$, $i = 1, \dots, n$, $j = i, \dots, n$. Es de dimensión $n^2 \times \frac{n(n+1)}{2}$.
3. Sea $X_{n \times n}$ triangular inferior, entonces $X = \sum_{i=1}^n \sum_{j=1}^i x_{ij} J_{ij}$. Por tanto, $Vec(X) = \sum_{i=1}^n \sum_{j=1}^i x_{ij} e_{(j-1)n+i}$ y las columnas de la matriz de transición son $e_{(j-1)n+i}$, $i = 1, \dots, n$, $j = 1, \dots, i$. Es de dimensión $n^2 \times \frac{n(n+1)}{2}$.
4. Sea $X_{n \times n}$ antisimétrica (en \mathbb{R} los valores de la diagonal son 0), entonces $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_{ij} [J_{ij} - J_{ji}]$. Por tanto, $Vec(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_{ij} [e_{(j-1)n+i} - e_{(i-1)n+j}]$ y las columnas de la matriz de transición son $e_{(j-1)n+i} - e_{(i-1)n+j}$, $i = 1, \dots, n-1$, $j = i+1, \dots, n$. Es de dimensión $n^2 \times \frac{n(n-1)}{2}$.

□

Las matrices \mathbf{Di}_n , \mathbf{TSup}_n y \mathbf{TInf}_n también tienen rango por columnas completo. Además, como el producto de ellas por su traspuesta es la matriz identidad, sus g-inversas coinciden con sus traspuestas.

Los siguientes resultados serán útiles más adelante:

1. $D_n^t Vec(A) = Vech(A + A^t - diag(A))$.
2. $|D_n^-(A^{-1} \otimes A^{-1})(D_n^-)^t| = 2^{-\frac{n(n-1)}{2}} |A|^{n+1}$.

1.6.1. Ejercicios

Ejercicio 1.8 Si A es no singular, $[(D_n)_g(A \otimes A)D_n]^{-1} = (D_n)_g(A \otimes A)^{-1}D_n$.

Solución:

Veamos que $[(D_n)_g(A \otimes A)D_n][(D_n)_g(A \otimes A)^{-1}D_n] = I_{\frac{1}{2}n(n+1)}$:

$$\begin{aligned} & [(D_n)_g(A \otimes A)D_n][(D_n)_g(A \otimes A)^{-1}D_n] = \\ & = [(D_n)_g(A \otimes A)][D_n(D_n)_g(A^{-1} \otimes A^{-1})D_n] = \\ & = (D_n)_g(A \otimes A)(A \otimes A)^{-1}D_n = \\ & = (D_n)_gI_{n^2}D_n = (D_n)_gD_n = \\ & = ((D_n^t D_n)^{-1}D_n^t)D_n = (D_n^t D_n)^{-1}(D_n^t D_n) = I_{\frac{1}{2}n(n+1)} \end{aligned}$$

Ejercicio 1.9 Si A es no singular, $[D_n'(A \otimes A)D_n]^{-1} = (D_n)_g(A \otimes A)^{-1}(D_n)_g^t$.

Solución:

Veamos ahora que $[D_n^t(A \otimes A)D_n][(D_n)_g(A \otimes A)^{-1}(D_n)_g^t] = I_{\frac{1}{2}n(n+1)}$. Para ello:

$$\begin{aligned} & [D_n^t(A \otimes A)D_n][(D_n)_g(A \otimes A)^{-1}(D_n)_g^t] = \\ & = [D_n^t(A \otimes A)][D_n(D_n)_g(A^{-1} \otimes A^{-1})(D_n)_g^t] = \\ & = D_n^t(A \otimes A)(A \otimes A)^{-1}(D_n)_g^t = \\ & = D_n^tI_{n^2}(D_n)_g^t = D_n^t(D_n)_g^t = \\ & = ((D_n)_gD_n)^t = (I_{\frac{1}{2}n(n+1)})^t = I_{\frac{1}{2}n(n+1)} \end{aligned}$$

2. Derivación matricial.

2.1. Introducción

En este apartado se va a tratar la derivación respecto a vectores y matrices, que es muy necesaria en estadística multivariante sobre todo desde el punto de vista de la optimización. Así, permite calcular datos tales como estimador máximo verosímil, matrices de información de Fisher, o cotas tipo Crámer-Rao. Más importancia tiene todavía este tema si tenemos en cuenta que, si ya la derivación vectorial puede dar lugar a cálculos costosos, en el caso de la matricial se pueden generar un enorme número de derivadas que pueden resultar difícil de ordenar con sentido en una matriz.

Muchas son las aproximaciones que se han dado y en este caso lo que se hará es seguir una línea en la que las matrices se conviertan en vectores, mucho más fáciles de manejar.

2.2. Diferencial primera y jacobianos

2.2.1. Diferencial de una función vectorial

Definición 2.1. Consideramos una función vectorial $f : S \rightarrow \mathbb{R}^m$ con $S \subset \mathbb{R}^n$. Sea \mathbf{c} un punto interior de S y consideremos una bola cerrada con centro en \mathbf{c} y radio r , $B(\mathbf{c}, r)$. Sea \mathbf{u} un punto de \mathbb{R}^n tal que $\|\mathbf{u}\| \leq r$ es decir, $\mathbf{c} + \mathbf{u} \in B(\mathbf{c}, r)$.

Diremos que f es **diferenciable** en \mathbf{c} si existe una matriz real de orden $m \times n$ que depende de \mathbf{c} y no de \mathbf{u} y que cumple que $f(\mathbf{c} + \mathbf{u}) - f(\mathbf{c}) = A(\mathbf{c})\mathbf{u} + r_c(\mathbf{u})$ con $\lim_{\mathbf{u} \rightarrow 0} \frac{r_c(\mathbf{u})}{\|\mathbf{u}\|} = 0$. Además, se define la **primera diferencial** de f en el punto \mathbf{c} con incremento \mathbf{u} como: $df(\mathbf{c}; \mathbf{u}) = A(\mathbf{c})\mathbf{u}$.

Notemos que si f es diferenciable en \mathbf{c} , entonces f es continua en \mathbf{c} .

Definición 2.2. Sea $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ y sea $f_i : S \rightarrow \mathbb{R}$ su i -ésima componente. Sea \mathbf{c} un punto interior de S y \mathbf{e}_j el j -ésimo vector de la base canónica de \mathbb{R}^n . Se define la **derivada parcial** de f respecto a la j -ésima coordenada como:

$$D_j f_i(\mathbf{c}) = \lim_{t \rightarrow 0} \frac{f_i(\mathbf{c} + t\mathbf{e}_j) - f_i(\mathbf{c})}{t}, t \in \mathbb{R}$$

El siguiente teorema nos permitirá identificar la matriz jacobiana a partir de la diferencial de una función vectorial:

Teorema 2.1. (Primer teorema de identificación para funciones vectoriales)

Sea $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un punto \mathbf{c} interior de S y $\mathbf{u} \in \mathbb{R}^n$. Entonces $df(\mathbf{c}; \mathbf{u}) = (Df(\mathbf{c}))\mathbf{u}$ donde $Df(\mathbf{c})$ es una matriz $m \times n$ cuyos elementos $D_j f_i(\mathbf{c})$ son las derivadas parciales de f evaluadas en \mathbf{c} y que recibe el nombre de matriz jacobiana. Recíprocamente, si $A(\mathbf{c})$ es una matriz que verifica que $df(\mathbf{c}; \mathbf{u}) = A(\mathbf{c})\mathbf{u} \forall \mathbf{u} \in \mathbb{R}^n$, entonces $A(\mathbf{c}) = Df(\mathbf{c})$.

La existencia de derivadas parciales no implica la diferenciabilidad de una función en general, pero si estas parciales existen y son continuas sí podríamos afirmar que la función es diferenciable.

En los dos siguientes teoremas vemos el equivalente a la regla de la cadena para funciones vectoriales: en el primero para derivadas parciales y en el segundo para las diferenciales.

Teorema 2.2. Sea $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un punto c interior de S . Sea T un subconjunto de \mathbb{R}^m tal que $f(x) \in T \forall x \in S$ y supongamos que $g : T \rightarrow \mathbb{R}^p$ es diferenciable en un punto \mathbf{b} ($\mathbf{b} = f(c)$) de T . Entonces la función compuesta $h : S \rightarrow \mathbb{R}^p$ definida por $h(x) = g(f(x))$, es diferenciable en c y $Dh(c) = Dg(\mathbf{b})Df(c)$.

Teorema 2.3. (Regla de invarianza de Cauchy)

En el ambiente del teorema anterior, si \mathbf{f} es diferenciable en \mathbf{c} y \mathbf{g} lo es en $\mathbf{b}=\mathbf{f}(\mathbf{c})$, entonces la diferenciable $h = g \circ f$ es $dh(\mathbf{c}; u) = dg(\mathbf{b}; d\mathbf{f}(\mathbf{c}; u))$.

2.2.2. Diferencial de una función matricial

Basta con extrapolar lo dicho en el caso vectorial usando la operación Vec .

Consideramos una función matricial $F : S \rightarrow \mathbb{M}_{m \times p}$ donde $S \subseteq \mathbb{M}_{n \times q}$. Sea C un punto interior de S , $\mathbb{B}(C; r) \subseteq S$ una bola abierta y U un punto de $\mathbb{M}_{n \times q}$ con $\|U\| < r$, por lo que $C + U$ pertenece a $\mathbb{B}(C; r)$, donde hemos considerado la norma matricial $\|U\| = (\text{tr}[U^t U])^{\frac{1}{2}}$.

Definición 2.3. En las condiciones anteriores, se dice que F es diferenciable en C si existe una matriz A de dimensiones $mp \times nq$, que dependa de C y no de U y tal que

$$\text{Vec}(F(C + U)) - \text{Vec}(F(C)) = A(C)\text{Vec}(U) + \text{Vec}(R_C(U)) \quad \text{donde} \quad \lim_{U \rightarrow 0} \frac{R_C(U)}{\|U\|} = 0$$

Se define la **matriz diferencial** de F en C con incremento U como la matriz $dF(C; U)$ de dimensiones $m \times p$ que verifique $\text{Vec}(dF(C; U)) = A(C)\text{Vec}(U)$ y a la matriz $A(C)$ se le llama la **primera derivada** de F en C .

Todas las propiedades de cálculo para las funciones matriciales se deducen de las correspondientes propiedades de las funciones vectoriales.

Al igual que en el caso anterior tenemos el siguiente teorema que nos ayuda a identificar los jacobianos de funciones matriciales.

Teorema 2.4. (Primer teorema de identificación para funciones matriciales) Sea $F : S \subseteq \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ diferencialbe en un punto interior C de S . Entonces se verifica $\text{Vec}(dF(C; U)) = A(C)\text{Vec}(U) \Leftrightarrow DF(C) = A(C)$.

Tenemos los siguientes resultados análogos a los del caso vectorial:

Teorema 2.5. (Regla de la cadena para funciones matriciales) Sea $F : S \subseteq \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ diferencialbe en un punto interior C de S . Sea T un subconjunto de $\mathbb{R}^{r \times s}$ tal que $F(X) \in T, \forall X \in S$ y supongamos que $G : T \rightarrow \mathbb{M}_{r \times s}$ es diferenciable en un punto B ($B = F(C)$) de T . Entonces la función compuesta $H : S \rightarrow \mathbb{M}_{r \times s}$ definida por $H(X) = G(F(X))$ es diferenciable en C y $DH(C) = DG(B)DF(C)$.

Teorema 2.6. (Regla de invarianza de Cauchy para funciones matriciales) En las condiciones del teorema anterior, $dH(C; U) = dG(B; dF(C; U)), \forall U \in \mathbb{M}_{n \times q}$.

En el siguiente teorema se recogen algunas propiedades útiles de las diferenciales matriciales y de las matrices jacobianas:

Teorema 2.7. Sean F y G dos funciones matriciales, A una matriz constante y $\alpha \in \mathbb{R}$.

1. Si $F(X) = A_{m \times p}$, entonces F es diferenciable. Además $dF(X)=0$ y $DF(X)=0$.
2. Si $G_{m \times p}$ es diferenciable, α un escalar distinto de cero y $F(X) = \alpha G(X)$, entonces F es diferenciable. Además, $dF(X) = \alpha dG(X)$ y $DF(X) = \alpha DG(X)$.
3. Si $G_{m \times p}$ es diferenciable y $F(x) = \text{Vec}(G(X))$, entonces F es diferenciable. Además se verifica $dF(X) = \text{Vec}(dG(X))$ y $DF(X)=DG(X)$.
4. Si $G_{m \times p}$ es diferenciable y $F(X)=G^t(X)$, entonces F es diferenciable. Además se verifica $dF(X)=(dG(X))^t$ y $DF(X)=K_{mp}DG(X)$.
5. Si $G_{p \times p}$ es diferenciable y $F(X)=\text{tr}[G(X)]$, entonces F es diferenciable. Además se verifica $dF(X)=\text{tr}[dG(X)]$ y $DF(X)=\text{Vec}'(I_p)DG(X)$.
6. Si $G_{m \times p}$ y $H_{m \times p}$ son diferenciables y $F(X) = (G \pm H)(X)$, entonces F es diferenciable. Además $dF(X) = dG(X) \pm dH(X)$ y $DF(X)=DG(X) \pm DH(X)$.
7. Si $G_{m \times r}$ y $H_{r \times p}$ son diferenciables y $F(X) = (GH(X))$, entonces F es diferenciable. Además $dF(X) = dG(X)H(X) + G(X)dH(X)$ y $DF(X) = (H^t(X) \otimes I_m)DG(X) + (I_p \otimes G(X))DH(X)$.
8. Si $G_{m \times m}$ es diferenciable y $F(X) = (G(X))^{-1}$, entonces para los puntos en los que F exista se verifica que dicha función es diferenciable. Además $dF(X) = -F(X)dG(X)F(X)$ y $DF(X) = -(F^t(X) \otimes F(X))DG(X)$.
9. Dados $G_{m \times p}$ y Φ una función real, ambas diferenciables y $F(X) = \Phi(X)G(X)$, entonces $dF(X) = \Phi(X)dG(X) + G(X)d\Phi(X)$ y $DF(X) = \text{Vec}(G(X))D\Phi(X) + \Phi(X)DG(X)$.
10. Dados $G_{m \times p}$ y $H_{r \times s}$, diferenciables, si $F(X) = (G \otimes H)(X)$ entonces F es diferenciable. Además $dF(X) = [dG(X) \otimes H(X)] + [G(X) \otimes dH(X)]$ y

$$DF(X) = [I_p \otimes (K_{sm} \otimes I_r)(I_m \otimes \text{Vec}(H(X)))]DG(X) + [(I_r \otimes K_{sm})(\text{Vec}(G(X)) \otimes I_s) \otimes I_r]DH(X)$$

Así, en el caso de funciones vectoriales, se cumple que $df(X) = A(X)dX$ y por tanto $Df(X) = A(X)$, mientras que en el caso de funciones matriciales, $d\text{Vec}(F(X)) = A(X)d\text{Vec}(X)$, entonces $DF(X) = A(X)$. Así, para calcular matrices jacobianas para este tipo de funciones:

- Calculamos $dF(X)$.
- Vectorizar la expresión resultante y obtener $d\text{Vec}(F(X)) = A(X)d\text{Vec}(X)$.
- Concluir con el cálculo de $DF(X)$ que es igual a $A(X)$.

Veamos un ejemplo de cálculo de matrices jacobianas:

EJEMPLO 2.1: Sea $F_{m \times p}$ una función diferenciable y consideremos $G(X) = AF(X)B$ con $A_{n \times m}$ y $B_{p \times q}$ matrices constantes. Entonces, siguiendo el procedimiento que acabamos de explicar y aplicando el teorema anterior, se tiene $dG(X) = AdF(X)B$, de donde $d\text{Vec}(G(X)) = (B^t \otimes A)d\text{Vec}(F(X)) = (B^t \otimes A)DF(X)d\text{Vec}(X)$ y, por lo tanto, concluimos que $DG(X) = (B^t \otimes A)DF(X)$.

Corolario 2.8. Sean A_{txm} , B_{pxr} y C_{sxu} matrices constantes y sean $G_{m \times p}$ y $H_{r \times s}$ dos funciones matriciales diferenciables. Si $F(X) = AG(X)BH(X)C$, entonces

$$dF(X) = AdG(X)BH(X)C + AG(X)BdH(X)C$$

$$\text{y } DF(X) = (C^t H^t(X) B^t \otimes A) DG(X) + (C^t \otimes AG(X) B) DH(X)$$

Del corolario anterior se pueden deducir algunas propiedades que se utilizan frecuentemente en la práctica.

2.2.3. Ejercicios

Ejercicio 2.1 Sea $h : \mathbb{R}^k \rightarrow \mathbb{R}$ definida por $h(\beta) = (y - X\beta)^t(y - X\beta)$ donde $y \in \mathbb{R}^n$ y $X \in \mathbb{M}_{n \times k}$. Haciendo uso de la regla de invarianza de Cauchy demostrar que

$$dh(c; u) = dg(y - Xc; df(c; u)) = dg(y - Xc; -Xu) = -2(y - Xc)^t Xu$$

y con ello $Dh(c) = -2(y - Xc)^t X$.

Solución:

Para comenzar, tengamos en cuenta que h es composición de $f : \mathbb{R}^k \rightarrow \mathbb{R}^n$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}$, con $f(c) = (y - Xc)$ con X una matriz $n \times k$ y con $y \in \mathbb{R}^n$; y $g(x) = x^t x$, de modo que $h(x) = g(f(x)) \forall x \in \mathbb{R}^k$. Así, es claro que $dh(c; u) = dg(y - Xc; df(c; u))$.

Utilizando el primer teorema de identificación para funciones vectoriales y que $Df(c) = -X$ se tiene que

$$df(c; u) = (Df(c))u = -Xu$$

Entonces,

$$dh(c; u) = dg(y - Xc; df(c; u)) = dg(y - Xc; -Xu)$$

Utilizando el teorema de identificación para g y utilizando que $Dg(x) = 2(y - Xc)^t$ tenemos:

$$dg(y - Xc; -Xu) = (Dg(y - Xc))(-Xu) = -2(y - Xc)^t Xu$$

Por tanto,

$$dh(c; u) = dg(y - Xc; df(c; u)) = -2(y - Xc)^t Xu$$

.

Para finalizar, tengamos en cuenta que $dh(c; u) = (Dh(c))u$, por tanto, $Dh(c) = -2(y - Xc)^t X$

Ejercicio 2.2 Sea $F(X) = AG(X)B$, donde $A_{m \times r}$ y $B_{s \times p}$ son matrices constantes y $G(X)_{r \times s}$ es una función diferenciable. Calcular $DF(C)$ a partir de la definición de diferencial matricial.

Solución:

$F(X) = AG(X)B$. Por tanto:

$$\begin{aligned} \text{Vec}(F(C+U)) - \text{Vec}(F(C)) &= \text{Vec}(AG(C+U)B) - \text{Vec}(AG(C)B) = \\ &= (B^t \otimes A)\text{Vec}(G(C+U)) - (B^t \otimes A)\text{Vec}(G(C)) = (B^t \otimes A)[\text{Vec}(G(C+U)) - \text{Vec}(G(C))] \end{aligned}$$

.

Por tanto, como G es diferenciable:

$$\text{Vec}(F(C+U)) - \text{Vec}(F(C)) = (B^t \otimes A)K(C)\text{Vec}(U) + \text{Vec}(R_c(U))$$

con $K(C)$ la primera derivada de F en C . Si llamamos $Q(C) = (B^t \otimes A)K(C)$, entonces F es diferenciable, $Q(C)$ (es decir, $(B^t \otimes A)K(C)$) es su primera derivada y $DF(C) = Q(C) = (B^t \otimes A)$

Ejercicio 2.3 Si $X_{n \times n}$ es una matriz simétrica y $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ es diferenciable, demostrar que $d\text{Vec}(F(X)) = D_n DF(X) d\text{Vec}(X)$. Por tanto, la matriz jacobiana de F respecto de $\text{Vec}(X)$ es $DF(X)D_n$, siendo $DF(X)$ la correspondiente a $\text{Vec}(X)$.

Solución:

Teniendo en cuenta que $d\text{Vec}(F(X)) = DF(X)d\text{Vec}(X)$ y que $D_n \text{Vec}(X) = \text{Vec}(X)$ dado que X es simétrica. Por tanto,

$$d\text{Vec}(X) = d(D_n \text{Vec}(X)) = D_n d(\text{Vec}(X))$$

Entonces,

$$d\text{Vec}(F(X)) = DF(X)d\text{Vec}(X) = DF(X)D_n d(\text{Vec}(X))$$

Por tanto, por el primer teorema de identificación se tiene que la matriz jacobiana de F respecto de $\text{Vec}(X)$ es $DF(X)D_n$, habiendo comprobado que $DF(X)$ es la correspondiente a $\text{Vec}(X)$.

2.3. Matrices jacobianas y derivadas matriciales

No existe una única definición para la derivada de una función de argumento matricial y esto supone un problema a la hora de usar el cálculo diferencial matricial. Son muchas las definiciones clásicas que existen, pero nosotros nos quedaremos con las de Mac Rae y Dwyer que generalizan a las demás y que están relacionadas como veremos también.

Definición 2.4. (*Definición de derivada matricial según Mac Rae*)

Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n \times q}$ como la matriz $nm \times pq$

$$\frac{\partial F(x)}{\partial x} = \begin{pmatrix} \frac{\partial F_{11}(x)}{\partial x} & \cdots & \cdots & \frac{\partial F_{1p}(x)}{\partial x} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{m1}(x)}{\partial x} & \cdots & \cdots & \frac{\partial F_{mp}(x)}{\partial x} \end{pmatrix}$$

Definición 2.5. (Definición de derivada matricial según Dwyer)

Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n \times q}$ como la matriz $nm \times pq$

$$\frac{\partial F(x)}{\partial x} = \begin{pmatrix} \frac{\partial F(x)}{\partial x_{11}} & \cdots & \cdots & \frac{\partial F(x)}{\partial x_{1q}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F(x)}{\partial x_{n1}} & \cdots & \cdots & \frac{\partial F(x)}{\partial x_{nq}} \end{pmatrix}$$

Teorema 2.9. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Entonces se verifica que $K_{nm} \frac{\partial F(x)}{\partial x} K_{pq} = \frac{\partial F(x)}{\partial x}$

Puede ocurrir que estas definiciones no conduzcan a la matriz jacobiana definida previamente. Así surge la siguiente definición en la que precisamente se emplea la misma notación que la matriz jacobiana debido a su coincidencia.

Definición 2.6. (Definición de derivada matricial según Magnus y Neudecker)

Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n \times q}$ como la matriz $mp \times nq$ dada por $DF(X) = \frac{\partial \text{Vec}(F(x))}{\partial \text{Vec}^t(x)}$

Podemos ahora mostrar algunas relaciones en el cálculo de matrices jacobianas a partir de derivadas matriciales y viceversa:

Teorema 2.10. Para las siguientes funciones matriciales y de variable matricial se tiene:

1. Dada $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$, $DF(X) = \frac{\partial \text{Vec}(F(x))}{\partial \text{Vec}^t(x)} = \sum_{i=1}^n \sum_{j=1}^q \text{Vec}\left(\frac{\partial F(x)}{\partial x_{ij}}\right) \text{Vec}^t(J_{ij})$.
2. Dada $F : \mathbb{R} \rightarrow \mathbb{M}_{m \times p}$, $DF(x) = \frac{\partial \text{Vec}(F(x))}{\partial x} = \text{Vec}\left(\frac{\partial F(x)}{\partial x}\right)$
3. Dada $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{R}$, $DF(X) = \frac{\partial F(X)}{\partial \text{Vec}^t(X)} = \text{Vec}\left(\frac{\partial F(X)}{\partial X}\right)$
4. Dada $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{R}^m$, $DF(X) = \frac{\partial F(X)}{\partial \text{Vec}^t(X)} = \sum_{l=1}^m u_l \text{Vec}^t\left(\frac{\partial F_l(X)}{\partial X}\right) = \sum_{l=1}^m u_l DF_l(X)$ con $\{u_j : j = 1, \dots, m\}$ base canónica de \mathbb{R}^m
5. Sea $F : \mathbb{R}^n \rightarrow \mathbb{M}_{m \times p}$, entonces

$$DF(x) = \frac{\partial \text{Vec}(F(X))}{\partial x^t} = \sum_{s=1}^m \sum_{t=1}^p \text{Vec}(E_{st}) \frac{\partial F_{st}(x)}{\partial x^t} = \sum_{s=1}^m \sum_{t=1}^p \text{Vec}(E_{st}) DF_{st}(x)$$

Veamos ahora las principales reglas de derivación según Mac Rae.

Teorema 2.11. (Propiedades de la derivada matricial según Mac Rae):

1. Sean $F, G : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Entonces $\frac{\partial (F+G)(X)}{\partial X} = \frac{\partial F(X)}{\partial X} + \frac{\partial G(X)}{\partial X}$.
2. Sean $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$, $G : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{p \times r}$. Entonces:

$$\frac{\partial (FG)(X)}{\partial X} = \frac{\partial F(X)}{\partial X} (G(X) \otimes I_q) + (F(X) \otimes I_n) \frac{\partial G(X)}{\partial X}$$

3. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times m}$, y supongamos que $F(X)$ no es singular $\forall X \in \mathbb{M}_{n \times q}$. Entonces:

$$\frac{\partial (F(X))^{-1}}{\partial X} = -[(F(X))^{-1} \otimes I_n] \frac{\partial F(X)}{\partial X} [(F(X))^{-1} \otimes I_q]$$

4. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ y $\Phi : \mathbb{M}_{n \times q} \rightarrow \mathbb{R}$. Entonces:

$$\frac{\partial(\Phi F)(X)}{\partial X} = F(X) \otimes \frac{\partial \Phi(X)}{\partial X} + \Phi(X) \frac{\partial F(X)}{\partial X}$$

5. Sea $A_{m \times p}$, constante y $G : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{r \times s}$. Entonces $\frac{\partial(A \otimes G(X))}{\partial X} = A \otimes \frac{\partial G(X)}{\partial X}$

6. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ y $G : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{r \times s}$. Entonces:

$$\frac{\partial(F(X) \otimes G(X))}{\partial X} = (K_{mr} \otimes I_n)(G(X) \otimes \frac{\partial F(X)}{\partial X})(K_{sp} \otimes I_q) + F(X) \frac{\partial G(X)}{\partial X}$$

7. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ y $\Phi : \mathbb{M}_{m \times p} \rightarrow \mathbb{R}$. Entonces notando $Y = F(X)$, se verifica:

$$\frac{\partial \Phi(F(X))}{\partial X} = \frac{\partial \Phi(X)}{\partial Y} \frac{\partial Y}{\partial X}$$

8. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ y $G : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{ml \times pr}$. Entonces:

$$\frac{\partial(F(X) * G(X))}{\partial X} = F(X) * \frac{\partial G(X)}{\partial X} + G(X)(I_m \otimes \text{Vec}(I_l) \otimes I_n) \frac{\partial F(X)}{\partial X} (I_p \otimes \text{Vec}^t(I_r) \otimes I_q)$$

9. Sean $F, G : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Entonces $\frac{\partial(F(X) * G(X))}{\partial X} = F(X) * \frac{\partial G(X)}{\partial X} + G(X) * \frac{\partial F(X)}{\partial X}$

Teorema 2.12. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Entonces

1. $\frac{\partial F(x)}{\partial x^t} = \left(\frac{\partial(F(x))^t}{\partial x} \right)^t$
2. $\frac{\partial(F(x))^t}{\partial x^t} = \left(\frac{\partial F(x)}{\partial x} \right)^t$

Como consecuencia obtenemos el siguiente corolario que engloba la derivada matricial de múltiples funciones matriciales.

Corolario 2.13. Sean A_{txm} , B_{pxr} y C_{sxu} matrices constantes y sean G_{mxp} y H_{rxs} dos funciones de argumento matricial $X_{n \times q}$. Si $F(X) = AG(X)BH(X)C$, entonces

$$\frac{\partial F(\partial x)}{\partial x} = (A \otimes I_n) \frac{\partial G(x)}{\partial x} (BH(\partial x)C \otimes I_q) + (AG(x)B \otimes I_n) \frac{\partial H(x)}{\partial x} (C \otimes I_q)$$

2.3.1. Ejercicios

Ejercicio 2.4 A partir de las relaciones existentes entre la derivada matricial y la matriz jacobiana, verificar las siguientes expresiones:

a) Sea $X_{n \times n}$ y $F(X) = \text{tr}[X]$. Entonces $DF(X) = \text{Vec}^t(I_n)$.

Solución:

Tengamos en cuenta que $DF(X) = \text{Vec}^t \frac{\partial F(X)}{\partial x}$. Además,

$$\frac{\partial F(X)}{\partial x} = \begin{pmatrix} \frac{\partial F(X)}{\partial x_{11}} & \cdots & \cdots & \frac{\partial F(X)}{\partial x_{1n}} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial F(X)}{\partial x_{n1}} & \cdots & \cdots & \frac{\partial F(X)}{\partial x_{nn}} \end{pmatrix}$$

Puesto que $\frac{\partial F(X)}{\partial x_{ij}} = \frac{\partial \sum_{i=1}^n x_{ii}}{\partial x_{ij}} = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}$

Finalmente, $\frac{\partial F(X)}{\partial x} = I_n \Rightarrow DF(X) = Vec^t(I_n)$

b) Sea ahora $X_{n \times q}$ y $F(X) = X$. Entonces $DF(X) = I_q \otimes I_n = I_{nq}$.

Ejercicio 2.5 Sea $X_{n \times q}$. Demostrar las siguientes igualdades:

a) $\frac{\partial X^t}{\partial X} = K_{qn}$.

Solución:

$$\frac{\partial X^t}{\partial X} = \sum_{i=1}^q \sum_{j=1}^n [J_{ij} \otimes \frac{\partial X^t}{\partial x_{ij}}] = \sum_{i=1}^q \sum_{j=1}^n [J_{ij} \otimes J_{ij}^t] = K_{qn}$$

b) $\frac{\partial X}{\partial X^t} = K_{nq}$.

Solución:

$$\frac{\partial X}{\partial X^t} = \frac{\partial X^t}{\partial X} = K_{qn}^t = K_{nq}$$

c) $\frac{\partial X^t}{\partial X^t} = Vec(I_q)Vec^t(I_n)$.

Solución:

Basta con tomar $Y = X^t$. De este modo:

$$\frac{\partial X^t}{X^t} = \frac{\partial Y}{\partial Y} = Vec(I_q)Vec^t(I_n)$$

Ejercicio 2.6 Demostrar que si $X_{n \times n}$ es no singular entonces $\frac{\partial X^{-1}}{\partial X} = -Vec((X^{-1})^t)Vec^t(X^{-1})$.

Solución:

Para comenzar tengamos en cuenta que

$$\begin{aligned} \frac{\partial X^{-1}}{\partial X} &= -[X^{-1} \otimes I_n] \frac{\partial X}{\partial X} [X^{-1} \otimes I_q] = -[X^{-1} \otimes I_n] Vec(I_n) Vec^t(I_q) [X^{-1} \otimes I_q] = -Vec[(X^{-1})^t] [((X^{-1})^t \otimes I_q^t) Vec(I_q)]^t = \\ &= Vec[(X^{-1})^t] [Vec((X^{-1})^t)]^t = Vec(X^{-1})^t Vec^t(X^{-1}) \end{aligned}$$

2.3.2. Derivadas matriciales de funciones escalares de un vector

Las principales funciones que suelen aparecer en la práctica son:

- Formas lineales: $\phi(x) = a^t x$
- Formas cuadráticas: $\phi(x) = x^t A x$

De las que se pueden sacar una serie de **propiedades**:

- $d\phi(x) = a^t dx \rightarrow D\phi(x) = \frac{\partial \phi(x)}{\partial x^t} = a^t$
- $d\phi(x) = x^t (A + A^t) dx \rightarrow D\phi(x) = \frac{\partial \phi(x)}{\partial x^t} = x^t (A + A^t)$

Y otras tomando f y g como dos funciones vectoriales del vector x :

- Si $\phi(x) = a^t f(x) \rightarrow D\phi(x) = a^t Df(x)$
- Si $\phi(x) = f(x)^t g(x) \rightarrow D\phi(x) = g(x)^t Df(x) + f(x)^t Dg(x)$
- Si $\phi(x) = x^t A f(x) \rightarrow D\phi(x) = f(x)^t A^t + x^t A Df(x)$
- Si $\phi(x) = f(x)^t A f(x) \rightarrow D\phi(x) = f(x)^t (A + A^t) Df(x)$
- Si $\phi(x) = f(x)^t A g(x) \rightarrow D\phi(x) = g(x)^t A^t Df(x) + f(x)^t A Dg(x)$
- Si $x = (x_1^t, x_2^t)^t$ y Si $\phi(x) = x_1^t A x_2 \rightarrow D\phi(x) = x^t \begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}$

2.3.3. Derivadas matriciales de funciones escalares de matrices

Cabe preguntarse también qué ocurre con las derivadas de dos funciones importantes como son la traza y el determinante.

Derivadas matriciales asociadas a trazas

Teorema 2.14. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{p \times p}$ y $\phi(x) = \text{tr}[F(x)]$. Entonces $d\phi(x) = \text{tr}[dF(x)]$, $D\phi(x) = \text{Vec}^t(I_p) DF(x)$ y

$$\frac{\partial \phi(x)}{\partial x} = I_p * \frac{\partial F(x)}{\partial x} = \sum_{l=1}^p \frac{\partial F_{ll}(x)}{\partial x}$$

Teorema 2.15. Sean $F_{p \times m}$, $G_{m \times p}$ dos funciones matriciales de $X_{n \times q}$. Entonces se verifica:

- $d\text{tr}[FG(x)] = \text{tr}[F(x)dG(x)] + \text{tr}[dF(x)G(x)]$
- $D\text{tr}[(FG)(x)] = \text{Vec}^t(G^t(x))DF(x) + \text{Vec}^t(F^t(x))DG(x)$
- $\frac{\partial \text{tr}[(FG)(x)]}{\partial x} = G^t(x) * \frac{\partial F(x)}{\partial x} + F^t(x) * \frac{\partial G(x)}{\partial x}$

De estos teoremas se pueden deducir numerosas propiedades.

Derivadas matriciales asociadas a determinantes

El determinante es una función muy empleada en Estadística ya que forma parte de la verosimilitud asociada a matrices aleatorias, así que veamos qué ocurre con su derivada.

Teorema 2.16. *Sea una función matricial $F : S \rightarrow \mathbb{R}^{p \times p}$, $p \geq 2$, donde S es un abierto de $\mathbb{R}^{n \times p}$. Si la función F es diferenciable en S , entonces lo es también la función $|F| : S \rightarrow \mathbb{R}$ definida por $|F|(X) = |F(X)|$, verificándose las relaciones $d|F|(X) = \text{tr}[F^*(X)dF(X)]$, $D|F|(X) = \text{Vec}^t(F^{*'}(X))DF(X)$ y*

$$\frac{\partial |F|(X)}{\partial X} = F^{*'}(X) * \frac{\partial F(X)}{\partial X} = \sum_{i=1}^n \sum_{j=1}^q \left(F^{*'}(X) * \frac{\partial F(X)}{\partial x_{ij}} \right) J_{ij} = \sum_{i=1}^n \sum_{j=1}^q \text{tr} \left[F^*(X) \frac{\partial F(X)}{\partial x_{ij}} \right] J_{ij}$$

donde $F^*(X)$ es la matriz adjunta de $F(X)$. En particular, en los puntos donde $\text{rg}(F(X))=p$, se verifica $d|F|(X) = |F(X)|\text{tr}[F^{-1}(X)dF(X)]$, $D|F|(X) = |F(X)|\text{Vec}^t(F^{-1'}(X))DF(X)$ y

$$\begin{aligned} \frac{\partial |F|(X)}{\partial X} &= |F(X)|F^{-1'}(X) * \frac{\partial F(X)}{\partial X} = |F(X)| \sum_{i=1}^n \sum_{j=1}^q \left(F^{-1'}(X) * \frac{\partial F(X)}{\partial x_{ij}} \right) J_{ij} \\ &= |F(X)| \sum_{i=1}^n \sum_{j=1}^q \text{tr} \left[F^{-1}(X) \frac{\partial F(X)}{\partial x_{ij}} \right] J_{ij} \end{aligned}$$

Por el teorema anterior se tiene:

Teorema 2.17. *Sean $F_1 : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{p \times r}$ y $F_2 : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{r \times p}$. Entonces en los puntos X en los que la función producto $F_1 F_2$ tenga determinante distinto de cero se verifica*

$$d|F_1 F_2| = |F_1 F_2| \text{tr}[F_2(F_1 F_2)^{-1}dF_1 + (F_1 F_2)^{-1}F_1 dF_2]$$

$$D|F_1 F_2| = |F_1 F_2| \left[\text{Vec}^t \left((F_1 F_2)^{-1'} F_2^t \right) DF_1 + \text{Vec}^t \left(F_1^t (F_1 F_2)^{-1'} \right) DF_2 \right]$$

y

$$\begin{aligned} \frac{\partial |F_1 F_2|}{\partial X} &= |F_1 F_2| (F_1 F_2)^{-1'} * \left(\frac{\partial F_1}{\partial X} [F_2 \otimes I_q] + [F_1 \otimes I_n] \frac{\partial F_2}{\partial X} \right) \\ &= |F_1 F_2| \sum_{i=1}^n \sum_{j=1}^q \text{tr} \left[F_2(F_1 F_2)^{-1} \frac{\partial F_1}{\partial x_{ij}} + (F_1 F_2)^{-1} F_1 \frac{\partial F_2}{\partial x_{ij}} \right] J_{ij} \end{aligned}$$

Veamos también la expresión de la diferencial y la derivada matricial del logaritmo de un determinante.

Corolario 2.18. *Sea S un abierto de $\mathbb{M}_{n \times q}$ y $F : S \rightarrow \mathbb{M}_{p \times p}$ tal que F es diferenciable sobre S . Entonces, para los puntos X en los que $\det F(X) \neq 0$, la función $\log|F| : S \rightarrow \mathbb{R}$ dada por $(\log|F|)(X) = \log|F(X)|$ es diferenciable. Además $d\log(|F|(X)) = \text{tr}[F^{-1}(X)dF(X)]$, $D\log(|F|(X)) = \text{Vec}^t(F^{-1'}(X))DF(X)$ y*

$$\begin{aligned} \frac{\partial \log(|F|(X))}{\partial X} &= F^{-1'}(X) * \frac{\partial F(X)}{\partial X} = \sum_{i=1}^n \sum_{j=1}^q \left(F^{-1'}(X) * \frac{\partial F(X)}{\partial x_{ij}} \right) J_{ij} \\ &= \sum_{i=1}^n \sum_{j=1}^q \text{tr} \left[F^{-1}(X) \frac{\partial F(X)}{\partial x_{ij}} \right] J_{ij} \end{aligned}$$

2.3.4. Jacobianos de funciones vectoriales

Es habitual encontrarnos en Estadística con que un vector aleatorio $Y = (y_1, \dots, y_m)^t$ es una función lineal de otra colección de variables y queremos poder expresar la derivada de este vector respecto a las otras variables. Pueden darse dos casos:

- Consideramos un conjunto de variables y_1, \dots, y_m y supongamos que estas variables son combinaciones lineales desconocidas de otro conjunto de variables x_1, \dots, x_n , por lo que podemos expresar $y_i = \sum_{j=1}^n a_{ij}x_j$ $i = 1, \dots, m$ o bien $y = f(x) = Ax$. Evidentemente, como $df(x) = Adx$ se verifica $Df(x) = A$.
- Supongamos ahora que las variables independientes sean los elementos de una matriz de variables x_{ij} en la forma $y = f(X) = Xa$. En este caso se tiene

$$df(X) = (dX)a = \text{Vec}((dX)a) = [a^t \otimes I_n]d\text{Vec}(X)$$

por lo que se desprende en este caso que $Df(X) = a^t \otimes I_n$.

2.4. Diferencial segunda y hessianos

2.4.1. Matriz hessiana para una función escalar de un vector

Definición 2.7. Sea $\Phi : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Consideremos c un punto interior de S donde las n^2 parciales de segundo orden $D_{kj}^2 \Phi(c)$ existan. Entonces se define la **matriz hessiana** $H\Phi(c)$ como la matriz cuadrada de orden n siguiente

$$H\Phi(c) = \begin{pmatrix} D_{11}^2 \Phi(c) & \cdots & D_{1n}^2 \Phi(c) \\ \vdots & \ddots & \vdots \\ D_{n1}^2 \Phi(c) & \cdots & D_{nn}^2 \Phi(c) \end{pmatrix}$$

2.4.2. Matriz hessiana para una función vectorial de un vector

Definición 2.8. Sea $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Consideremos c un punto interior de S donde las mn^2 parciales de segundo orden $D_{kj}^2 f_i(c)$ existan. Entonces se define la **matriz hessiana** $Hf(c)$ como la matriz de orden $mn \times n$ siguiente

$$Hf(c) = \begin{pmatrix} Hf_1(c) \\ \vdots \\ Hf_m(c) \end{pmatrix}$$

2.4.3. Diferencial segunda de una función escalar

Definición 2.9. Consideremos una función real $\Phi : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ que sea diferenciable en un punto interior c de S . Supongamos además que existe una matriz B que dependa de c y no de u y tal que verifique la relación

$$\Phi(c+u) = \Phi(c) + [D\Phi(c)]u + \frac{1}{2}u'Bu + r_c(u), \lim_{u \rightarrow 0} \frac{r_c(u)}{\|u\|^2} = 0$$

En tal caso se dice que la función Φ es dos veces diferenciable en c .

Ahora, siendo Φ una función dos veces diferenciable en un punto c interior del conjunto S , si notamos $d\Phi(x; u) = \Psi(x)$, entonces la diferencial segunda de Φ sería la diferencial de Ψ . Entonces

$$\Psi(x) = \sum_{j=1}^n u_j D_j \Phi(x)$$

cuyas derivadas parciales son

$$D_i \Psi(x) = \sum_{j=1}^n u_j D_{ij}^2 \Phi(x)$$

Por tanto, la primera diferencial de Φ en u sería

$$d\Psi(x; u) = \sum_{i=1}^n u_i D_i \Psi(x) = \sum_{i=1}^n \sum_{j=1}^n u_i u_j D_{ij}^2 \Phi(x)$$

por tanto, la diferencial segunda de Φ no es más que

$$d^2\Phi(x; u) = u' H\Phi(x) u$$

que es una forma cuadrática.

Por tanto, hemos deducido un resultado interesante sobre la identificación de diferenciales segundas que recogemos en el siguiente teorema.

Teorema 2.19 (Segundo teorema de identificación). *Sea $\Phi : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ que sea dos veces diferenciable en un punto interior c de S . Sea u un vector cualquiera de \mathbb{R}^n . Entonces*

$$d^2\Phi(c; u) = u' H\Phi(c) u$$

Además, si existe una matriz B que dependa de c y no de u y tal que

$$d^2\Phi(c; u) = u' B(c) u$$

entonces,

$$H\Phi(c) = \frac{1}{2}[B(c) + B(c)^t]$$

2.4.4. Diferencial segunda de una función vectorial

Definición 2.10. *Consideremos la función $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$. Sea c un punto interior de S . Si f es diferenciable en alguna bola abierta $\mathbb{B}(c)$ y cada una de las derivadas parciales $D_i f_i$ es diferenciable en c , diremos que f es dos veces diferenciable en c .*

Como es natural, la diferencial segunda no puede ser otra que la diferencial de la diferencial primera.

Definición 2.11. Sea $F : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ y dos veces diferenciable en un punto interior c de S . Sea $\mathbb{B}(c)$ una bola abierta de S tal que f es diferenciable en cada punto de $\mathbb{B}(c)$ y sea la función $g(x) = df(x; u)$.

Entonces la diferencial de g en c con incremento u se llama la diferencial segunda de f en c con incremento u y la notaremos $d^2 f(c; u)$.

Desarrollando la expresión, tenemos

$$d^2 f(c; u) = \begin{pmatrix} d^2 f_1(c; u) \\ \vdots \\ d^2 f_m(c; u) \end{pmatrix} = \begin{pmatrix} u' H f_1(c) u \\ \vdots \\ u' H f_m(c) u \end{pmatrix} = [I_m \otimes u'] \begin{pmatrix} H f_1(c) \\ \vdots \\ H f_m(c) \end{pmatrix} u$$

por tanto,

$$d^2 f(c; u) = [I_m \otimes u'] H f(c) u$$

De forma análoga al caso escalar, buscamos un teorema que nos de una identificación de la matriz hessiana con una función vectorial. Para introducirlo, necesitamos la siguiente definición

Definición 2.12. Sean A_1, \dots, A_m matrices cuadradas de orden n y sea la matriz $A = [A_1, \dots, A_m]$. Definimos la matriz de orden $mn \times n$ siguiente

$$A_v = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{pmatrix}$$

En particular, si B_1, \dots, B_m son matrices cuadradas y tenemos la matriz $B = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{pmatrix}$ entonces

$$B' = [B_1^t, \dots, B_m^t] \text{ y con ello tenemos que } B'_v = \begin{pmatrix} B'_1 \\ B'_2 \\ \vdots \\ B'_m \end{pmatrix}$$

Estamos ahora en condición de enunciar el teorema buscado.

Teorema 2.20 (Segundo teorema de identificación para funciones vectoriales). Sea $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ y dos veces diferenciable en un punto c interior de S . Sea $u \in \mathbb{R}^n$. Entonces se tiene

$$d^2 f(c; u) = [I_m \otimes u'] Hf(c) u$$

Además, si existe una matriz $B(c)$ tal que verifique

$$d^2 f(c; u) = [I_m \otimes u'] B(c) u, \forall u \in \mathbb{R}^n$$

entonces

$$Hf(c) = \frac{1}{2} [B(c) + B(c)^t]$$

2.4.5. Diferencial segunda de una función matricial

Definiremos la diferencial segunda de una función matricial trasladando la situación matricial a la vectorial haciendo uso de la ya estudiada operación Vec .

Dada una función $F : S \subset \mathbb{R}^{n \times q} \rightarrow \mathbb{R}^{m \times p}$ se tiene asociada una función vectorial $f : \text{Vec}(S) \rightarrow \mathbb{R}^{mp}$ definida como $f(\text{Vec}(X)) = \text{Vec}(F(X))$. Por tanto, se tendría que $DF(C) = Df(\text{Vec}(C))$, por tanto, definimos la matriz hessiana asociada a F como $HF(C) = Hf(\text{Vec}(C))$.

Esta matriz se forma con los hessianos de las mp funciones de las que está formada F , así

$$HF(C) = \begin{pmatrix} HF_{11}(C) \\ \vdots \\ HF_{m1}(C) \\ \vdots \\ HF_{1p}(C) \\ \vdots \\ HF_{mp}(C) \end{pmatrix}$$

Claramente, las matrices utilizadas $HF_{st}(C)$ son de orden $nq \times nq$ y el ij -ésimo elemento de estas es la segunda derivada parcial de $F_{st}(C)$ respecto de los i -ésimo y j -ésimo elementos de $\text{Vec}(X)$ evaluados en $X = C$.

Al igual que en el resto de secciones, notando $G(X) = df(X; U)$, entonces

$$d^2 F(C; U) = dG(C; U)$$

.

Utilizando la relación de la diferencial de F y f , obtenemos la relación de las diferenciales segundas

$$\text{Vec}[d^2 F(C; U)] = d^2 f[\text{Vec}(C); \text{Vec}(U)]$$

Por último, al igual que hemos realizado en las secciones anteriores, vemos un teorema homólogo del teorema de identificación en el caso matricial

Teorema 2.21 (Segundo teorema de identificación para funciones matriciales). *Sea $F : S \subset \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{m \times p}$. Supongamos que la función F es dos veces diferenciable en un punto C interior de S . Entonces se verifica*

$$Vec[d^2 F(C; U)] = [I_{mp} \otimes Vec^t(U)] B(C) Vec(U), \forall U \in \mathbb{R}^{n \times 1} \iff HF(C) = \frac{1}{2}[B(C) + B(C)^t]$$

3. Aplicaciones

3.1. Estimación máximo verosímil de los parámetros de la distribución normal multivariante.

Sea $X \rightsquigarrow Np[\mu, \Sigma]$ con $\Sigma > 0$ y sea $\mathbb{X}_{N \times p} = [X_1, \dots, X_N]^t$ una muestra aleatoria simple extraída de dicha población. A partir de la función de verosimilitud, para un valor muestral observado χ , calcular los estimadores máximo verosímiles de μ y Σ

Solución:

Tengamos en cuenta que tenemos que maximizar la función logaritmo de la función de verosimilitud, es decir:

$$\log(\mathcal{L}_X(\mu, \Sigma)) = -\frac{NP}{2} \log(2\pi) - \frac{N}{2} \log(|\Sigma|) - \frac{1}{2} tr[\Sigma^{-1} A] - \frac{N}{2} (\bar{X} - \mu)^t \Sigma^{-1} (\bar{X} - \mu)$$

Para el caso del estimador de μ es claro que para minimizar $(\bar{X} - \mu)^t \Sigma^{-1} (\bar{X} - \mu)$ puesto que $\Sigma > 0$, se minimiza cuando $\mu = \bar{X}$, por tanto $\hat{\mu} = \bar{X}$.

Sin embargo, para el estimador de Σ , tenemos que maximizar $F(\Sigma) = -\frac{N}{2} \log(|\Sigma|) - \frac{1}{2} tr[\Sigma^{-1} A]$

Para ello, derivamos. Para comenzar tengamos en cuenta que:

$$-\frac{1}{2} \frac{\partial tr[\Sigma^{-1} A]}{\partial \Sigma} = -\frac{1}{2} \frac{\partial tr[A \Sigma^{-1}]}{\partial \Sigma} = \frac{1}{2} (\Sigma^{-1} A \Sigma^{-1})^t.$$

Ahora, atendiendo a la derivada del logaritmo de un determinante, tenemos que:

$$\frac{-N}{2} \frac{\partial \log(|\Sigma|)}{\partial \Sigma} = \frac{-N}{2} (\Sigma^{-1})^t * \frac{\partial \Sigma}{\partial \Sigma} = \frac{-N}{2} (\Sigma^{-1})^t * Vec(I_n) Vec(I_n^t) = \frac{-N}{2} (\Sigma^{-1})^t$$

Uniendo tenemos que:

$$\frac{\partial F(\Sigma)}{\partial \Sigma} = -\frac{N}{2} (\Sigma^{-1})^t + \frac{1}{2} (\Sigma^{-1} A \Sigma^{-1})^t = -\frac{N}{2} (\Sigma^{-1}) + \frac{1}{2} (\Sigma^{-1} A^t \Sigma^{-1}) = -\frac{N}{2} (\Sigma^{-1}) + \frac{1}{2} (\Sigma^{-1} A \Sigma^{-1})$$

$$-\frac{N}{2} (\Sigma^{-1}) + \frac{1}{2} (\Sigma^{-1} A \Sigma^{-1}) = 0 \Leftrightarrow N(\Sigma^{-1}) = (\Sigma^{-1} A \Sigma^{-1}) \Leftrightarrow N = \Sigma^{-1} A \Leftrightarrow \Sigma = \frac{1}{N} A$$

Por tanto $\hat{\Sigma} = \frac{1}{N} A$.

3.2. Distribución normal matricial

Obtención de la expresión de su densidad. Como aplicación, obtención de la distribución de la matriz $\mathbb{X}_{N \times p} = [X_1, \dots, X_N]^t$ que contiene una muestra aleatoria simple extraída de una población normal multivariante.

Solución:

Si $\mathbb{Y}_{r \times s}$ es una matriz aleatoria. En general, la densidad de esa matriz es la correspondiente a $Vec(\mathbb{Y}^t)$:

Sea $\mathbb{Y}_{r \times s}$ una matriz aleatoria. Sean $M_{r \times s}$, $C_{r \times r}$, $D_{s \times s}$. Con C y D definidas positivas. Se dice que $Y \rightsquigarrow N_{r \times s}[M; C \otimes D]$ si $y = Vec(\mathbb{Y}^t) \rightsquigarrow N_{rs \times 1}[m, C \otimes D]$ con $m = Vec(M^t)$.

Teniendo en cuenta esa caracterización, se obtiene el siguiente resultado:

Sea $\mathbb{Y}_{r \times s}$ una matriz aleatoria normal $N_{r \times s}[M; C \otimes D]$. Entonces su densidad es:

$$f(\mathbb{Y}) = (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \exp\left(-\frac{1}{2} \text{tr}[C^{-1}(\mathbb{Y} - M)D^{-1}(\mathbb{Y} - M)^t]\right)$$

Esto se puede demostrar, basta con trabajar con la densidad de $y = Vec(\mathbb{Y}^t)$:

$$\begin{aligned} & (2\pi)^{-\frac{rs}{2}} |C \otimes D|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y - m)^t(C \otimes D)^{-1}(y - m)\right) = \\ & (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \exp\left(-\frac{1}{2}(Vec[(\mathbb{Y} - M)^t])^t(C^{-1} \otimes D^{-1})(Vec[(\mathbb{Y} - M)])^t\right) = \\ & = (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \exp\left(-\frac{1}{2}(Vec[(\mathbb{Y} - M)^t])^t(C^{-1} \otimes I_s)(I_r \otimes D^{-1})(Vec[\mathbb{Y} - M])^t\right) = \\ & = (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \exp\left(-\frac{1}{2}(C^{-1} \otimes I_s)(Vec[(\mathbb{Y} - M)^t])^t(I_r \otimes D^{-1})(Vec[(\mathbb{Y} - M)])^t\right) = \\ & = (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \exp\left(-\frac{1}{2}Vec^t[(\mathbb{Y} - M)^t]C^{-1}Vec[D^{-1}(\mathbb{Y} - M)^t]\right) = \\ & = (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \exp\left(-\frac{1}{2}Vec^t[(C^{-1}(\mathbb{Y} - M))^t]Vec[D^{-1}(\mathbb{Y} - M)^t]\right) = \\ & = (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \exp\left(-\frac{1}{2}\text{tr}[C^{-1}(\mathbb{Y} - M)D^{-1}(\mathbb{Y} - M)^t]\right) = \\ & = (2\pi)^{-\frac{rs}{2}} |C|^{-\frac{s}{2}} |D|^{-\frac{r}{2}} \text{etr}\left(-\frac{1}{2}C^{-1}(\mathbb{Y} - M)D^{-1}(\mathbb{Y} - M)^t\right). \end{aligned}$$

Si consideramos ahora $\mathbb{X}_{N \times p} = [X_1 \dots X_N]^t$ la matriz que contiene la muestra aleatoria simple X_1, \dots, X_N procedente de una población $N_p[\mu; \Sigma]$, con $\Sigma > 0$, entonces $\mathbb{X} \rightsquigarrow N_{N \times p}[1_N \mu^t; I_N \otimes \Sigma]$