Índice general

Capítulo 1

Complementos de álgebra matricial

1.1. Introducción

El contenido de este capítulo, amén de justificarse por sí solo como herramientas dentro del Álgebra Matricial, encuentra plena justificación dentro del contexto del Análisis Multivariante. En efecto, la manipulación algebraica de vectores aleatorios puede ser abordada sin ningún problema, pero la situación se vuelve distinta cuando consideramos matrices aleatorias ya que el tratamiento de las mismas debe ser tal que generalice el realizado sobre vectores. El empleo de matrices aleatorias en Análisis Multivariante no es una mera cuestión teórica sino que aparecen de forma natural. Por ejemplo, al considerar la distribución conjunta de una muestra aleatoria simple procedente de una población vectorial o al estimar la matriz de varianzas-covarianzas de dicha población.

1.2. Operación Vec

Como comentábamos con anterioridad, el tratamiento sobre matrices aleatorias debe ser visto como una extensión del que se realiza para vectores. Por ello lo habitual es *vectorizar* dicha matriz, o sea, tratarla como un vector sin más que tener en cuenta que los espacios $\mathbb{M}_{n\times q}$ (espacio vectorial de las matrices de dimensión $n \times q$) y $\mathbb{R}^{nq}$ son isomorfos. Evidentemente esta es una solución cómoda que será útil si somos capaces de conocer bien los mecanismos que ligan las expresiones matriciales y *vectorizadas*.

Definición 1.2.1. *Sea* $\mathbf{X}$ *una matriz de orden* $n \times q$*. Se define* $\mathrm{Vec}(\mathbf{X})$ *como el vector de dimensión* $nq \times 1$ *formado al apilar las columnas de* X *una tras otra, o sea, si notamos por* *columnas* $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_q]$, *entonces*

$$\mathrm{Vec}(\mathbf{X}) = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_q \end{pmatrix}.$$

El siguiente teorema proporciona el resultado comentado acerca de que los espacios vectoriales $\mathbb{M}_{n\times q}$ y $\mathbb{R}^{nq}$ son isomorfos.

Teorema 1.2.1. *La aplicación* $\mathrm{Vec} : \mathbb{M}_{n\times q} \longrightarrow \mathbb{R}^{nq}$ *es un isomorfismo de espacios vectoriales.*

Ejemplo 1.2.1. *Sean* $\mathbf{x}_1, \ldots, \mathbf{x}_N$ *vectores aleatorios* p*-dimensionales con igual media* μ*. Sea la matriz aleatoria* $\mathbf{X}_{N\times p} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^t$ *y consideremos el vector*

$$\mathrm{Vec}(\mathbf{X}^t) = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{pmatrix}.$$

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

$$\mathrm{E}\left[\mathrm{Vec}(\mathbf{X}^t)\right] = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} = \mathrm{Vec}\left([\mu, \mu, \ldots, \mu]\right) = \mathrm{Vec}(\mu\mathbf{1}_{\mathbf{N}}^{\mathbf{t}})\ .$$

Nota 1.2.1. *En ocasiones estaremos interesados en calcular la esperanza matemática de una cierta matriz aleatoria, que es una matriz cuyas componentes serán las esperanzas de cada componente. Sin embargo habrá situaciones en las que dicho cálculo será más fácil realizarlo si calculamos la esperanza de su* vectorización *y después, en virtud del isomorfismo anterior, deshacemos dicho proceso.*

1.3. Producto Kronecker

Como se ha comentado al final de la sección anterior (ver nota 1.2.1), la *vectorización* puede ser muy útil en determinadas ocasiones cuando ello puede, por ejemplo, facilitar el cálculo de esperanzas de matrices (o funciones suyas). En otras situaciones es obligado su empleo, como muestra el siguiente ejemplo:

Ejemplo 1.3.1. *Continuando con el ejemplo 1.2.1, supongamos además que* $\mathbf{x}_1, \ldots, \mathbf{x}_N$ *son independientes y con igual matriz de covarianzas* Σ. *Entonces se define la matriz de covarianzas de* $\mathbf{X}^t$ *como* $\mathrm{Cov}\left[\mathrm{Vec}(\mathbf{X}^t)\right]$ *ya que dicha matriz contendrá 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 (ver apéndice B), se verifica*

$$\begin{aligned}
\mathrm{Cov}\left[\mathrm{Vec}(\mathbf{X}^t)\right] &= \mathrm{E}\left[\left[\mathrm{Vec}(\mathbf{X}^t) - \mathrm{E}\left[\mathrm{Vec}(\mathbf{X}^t)\right]\right]\left[\mathrm{Vec}(\mathbf{X}^t) - \mathrm{E}\left[\mathrm{Vec}(\mathbf{X}^t)\right]\right]^t\right] \\
&= \mathrm{E}\left[\begin{pmatrix} \mathbf{X}_1 - \mu \\ \mathbf{X}_2 - \mu \\ \vdots \\ \mathbf{X}_N - \mu \end{pmatrix} \left((\mathbf{X}_1 - \mu)^t, \ldots, (\mathbf{X}_N - \mu)^t\right)\right] = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma \end{pmatrix}.
\end{aligned}$$

La expresión anterior que adopta la matriz $\mathrm{Cov}\left[\mathrm{Vec}(\mathbf{X}^t)\right]$ nos sugiere la introducción del producto Kronecker de matrices que veremos a continuación.

Definición 1.3.1. *Sean* $\mathbf{A}_{m\times n}$ *y* $\mathbf{B}_{p\times q}$ *dos matrices. Se define el producto Kronecker de ellas como la matriz de dimensiones* $mp \times nq$ *siguiente*

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

Nota 1.3.1. *A la vista de la definición anterior es inmediato que en el ejemplo 1.3.1 se tiene* $\mathrm{Cov}\left[\mathrm{Vec}(\mathbf{X}^t)\right] = \mathbf{I}_N \otimes \Sigma$.

Nota 1.3.2. *Evidentemente el ejemplo anterior no es suficiente justificación para la introducción de esta operación ya que se puede comentar que no deja de ser una forma de abreviar la notación. Otra, de las múltiples razones que se pueden argumentar, es la siguiente. Sea el producto* $\mathbf{A}\otimes\mathbf{B}$*, con* $\mathbf{A}_{n\times n}$ *y* $\mathbf{B}_{3\times 3}$ *y consideremos el sistema de ecuaciones* $\mathbf{x} = (\mathbf{A}\otimes\mathbf{B})\mathbf{y}$*. Si* $\mathbf{A}\otimes\mathbf{B}$ *es no singular, para resolver dicho sistema habrá que invertir una matriz* $3n \times 3n$*. Como veremos posteriormente, se verifica* $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$*, por lo que será suficiente con invertir dos matrices de orden inferior a la anterior, con el consiguiente ahorro de cálculo y, seguramente, con una ganancia en lo que se refiere a la precisión de la solución.*

Nota 1.3.3. *Como sabemos, el producto usual de matrices está relacionado con la composición de aplicaciones lineales. El producto Kronecker definido ahora lo está con el producto tensorial.*

Algunas de las propiedades básicas de esta operación se resumen en el siguiente resultado.

Teorema 1.3.1. *Se verifican las siguientes propiedades*

1. *Dados* $\alpha, \beta \in \mathbb{R}$*,* $\mathbf{A}_{m\times n}$ *y* $\mathbf{B}_{p\times q}$*, entonces*

$$(\alpha\mathbf{A})\otimes(\beta\mathbf{B}) = \alpha\beta(\mathbf{A}\otimes\mathbf{B}) = \alpha\beta\mathbf{A}\otimes\mathbf{B} = \mathbf{A}\otimes(\alpha\beta)\mathbf{B}\ .$$

2. *Dadas* $\mathbf{A}_{m\times n}$*,* $\mathbf{B}_{m\times n}$*,* $\mathbf{C}_{p\times q}$ *y* $\mathbf{D}_{p\times q}$*, entonces*
   a) $(\mathbf{A}\otimes\mathbf{C}) + (\mathbf{B}\otimes\mathbf{C}) = (\mathbf{A}+\mathbf{B})\otimes\mathbf{C}$.
   b) $(\mathbf{A}\otimes\mathbf{C}) + (\mathbf{A}\otimes\mathbf{D}) = \mathbf{A}\otimes(\mathbf{C}+\mathbf{D})$.
   c) $(\mathbf{A}+\mathbf{B})\otimes(\mathbf{C}+\mathbf{D}) = (\mathbf{A}\otimes\mathbf{C}) + (\mathbf{A}\otimes\mathbf{D}) + (\mathbf{B}\otimes\mathbf{C}) + (\mathbf{B}\otimes\mathbf{D})$.
3. *Dadas* $\mathbf{A}_{m\times n}$*,* $\mathbf{B}_{p\times q}$*,* $\mathbf{C}_{r\times s}$ *, entonces* $(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C})$.
4. *Dadas* $\mathbf{A}_{m\times n}$*,* $\mathbf{B}_{n\times p}$*,* $\mathbf{C}_{q\times r}$ *y* $\mathbf{D}_{r\times s}$*, entonces* $(\mathbf{A}\otimes\mathbf{C})(\mathbf{B}\otimes\mathbf{D}) = \mathbf{AB}\otimes\mathbf{CD}$.
5. *Si* $\mathbf{A}_{m\times m}$ *y* $\mathbf{B}_{n\times n}$ *son no singulares entonces* $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$.
6. *Dadas* $\mathbf{A}_{m\times n}$ *y* $\mathbf{B}_{p\times q}$*, entonces* $(\mathbf{A}\otimes\mathbf{B})^t = \mathbf{A}^t\otimes\mathbf{B}^t$.
7. *Si* $\mathbf{A}_{m\times m}$ *y* $\mathbf{B}_{n\times n}$ *son ortogonales, entonces* $\mathbf{A}\otimes\mathbf{B}$ *es ortogonal.*
8. *Si* $\mathbf{A}_{m\times m}$ *y* $\mathbf{B}_{n\times n}$ *son matrices triangulares superiores (inferiores), entonces* $\mathbf{A}\otimes\mathbf{B}$ *es triangular superior (inferior).*

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

10. *Dadas* $\mathbf{A}_{m\times n} = [\mathbf{A}_1, \ldots, \mathbf{A}_k]$ *y* $\mathbf{B}_{p\times q}$*, entonces* $\mathbf{A}\otimes\mathbf{B} = [\mathbf{A}_1\otimes\mathbf{B}, \ldots, \mathbf{A}_k\otimes\mathbf{B}]$*. En particular, si* $\mathbf{a}_{m\times 1}$ *y* $\mathbf{b}_{p\times 1}$ *son dos vectores se tiene* $\mathbf{a}\otimes\mathbf{b}^t = \mathbf{a}\mathbf{b}^t = \mathbf{b}^t\otimes\mathbf{a}$.

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

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

13. *Sean* $\mathbf{A}_{m\times m}$ *y* $\mathbf{B}_{n\times n}$ *matrices reales con autovalores reales respectivos* $\lambda_1, \ldots, \lambda_m$ *y* $\mu_1, \ldots, \mu_n$*. Entonces* $\mathbf{A}\otimes\mathbf{B}$ *tiene como autovalores* $\lambda_i\mu_j$*,* $i = 1, \ldots, m$*;* $j = 1, \ldots, n$*. Como consecuencia* $\operatorname{rg}(\mathbf{A}\otimes\mathbf{B}) = \operatorname{rg}(\mathbf{A})\operatorname{rg}(\mathbf{B})$.

14. *Dadas* $\mathbf{A}_{m\times m}$ *y* $\mathbf{B}_{n\times n}$*, entonces* $|\mathbf{A}\otimes\mathbf{B}| = |\mathbf{A}|^n|\mathbf{B}|^m$.

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

En el ejemplo 1.2.1 habíamos obtenido $\mathrm{E}\left[\operatorname{Vec}(\mathbf{X}^t)\right] = \operatorname{Vec}\left([\mu, \mu, \ldots, \mu]\right) = \operatorname{Vec}(\mu\mathbf{1}_{\mathbf{N}}^{\mathbf{t}})$ y, a la vista de la definición de producto Kronecker, es inmediato comprobar que esa expresión no es más que $\mathbf{1}_N\otimes\mu$, lo cual es una primera (y evidente) muestra de que ambas operaciones pueden conducir a resultados relacionados entre ellas. A continuación vamos a exponer algunas propiedades que ponen en relación las dos operaciones introducidas.

Teorema 1.3.2. *Se verifican las siguientes afirmaciones:*

1. *Si* $\mathbf{a}_{n\times 1}$ *y* $\mathbf{b}_{q\times 1}$ *son dos vectores, entonces* $\operatorname{Vec}(\mathbf{a}\mathbf{b}^t) = \mathbf{b}\otimes\mathbf{a}$.

2. *Sean* $\mathbf{A}_{n\times q}$*,* $\mathbf{B}_{q\times p}$ *y* $\mathbf{C}_{p\times r}$ *tres matrices cualesquiera. Entonces* $\operatorname{Vec}(\mathbf{ABC}) = (\mathbf{C}^t\otimes\mathbf{A})\operatorname{Vec}(\mathbf{B})$.

3. *Sean* $\mathbf{A}_{n\times q}$ *y* $\mathbf{B}_{q\times n}$*. Entonces* $\operatorname{tr}[\mathbf{AB}] = \operatorname{Vec}'(\mathbf{A}^t)\operatorname{Vec}(\mathbf{B}) = \operatorname{Vec}'(\mathbf{B}^t)\operatorname{Vec}(\mathbf{A})$.

4. *Sea* $\{\mathbf{e}_i : i = 1, \ldots, n\}$ *la base canónica de* $\mathbb{R}^n$*. Entonces* $\operatorname{Vec}(\mathbf{I}_n) = \sum_{i=1}^{n}(\mathbf{e}_i\otimes\mathbf{e}_i)$.

5. *Sea* $\{\mathbf{J}_{ij} : i = 1, \ldots, n; j = 1, \ldots, q\}$ *la base canónica del espacio de matrices* $M_{n\times q}$*. Entonces* $\sum_{i=1}^{n}\sum_{j=1}^{q}(\mathbf{J}_{ij}\otimes\mathbf{J}_{ij}) = \operatorname{Vec}(\mathbf{I}_n)\operatorname{Vec}'(\mathbf{I}_q)$.

1.3.2. Ejercicios

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

2. Verificar las siguientes igualdades

   a) Sean $\mathbf{A}_{m\times n}$ y $\mathbf{B}_{n\times p}$. Entonces

   $$\mathrm{Vec}(\mathbf{AB}) = (\mathbf{B}^t \otimes \mathbf{I}_m)\,\mathrm{Vec}(\mathbf{A}) = (\mathbf{I}_p \otimes \mathbf{A})\,\mathrm{Vec}(\mathbf{B}) = (\mathbf{B}^t \otimes \mathbf{A})\,\mathrm{Vec}(\mathbf{I}_n)\ .$$

   b) Sea $\mathbf{A}_{m\times n}$. Entonces

   $$\mathrm{Vec}(\mathbf{A}) = (\mathbf{A}^t \otimes \mathbf{I}_m)\,\mathrm{Vec}(\mathbf{I}_m) = (\mathbf{I}_n \otimes \mathbf{I}_m)\,\mathrm{Vec}(\mathbf{A}) = (\mathbf{I}_n \otimes \mathbf{A})\,\mathrm{Vec}(\mathbf{I}_n)\ .$$

   c) Sean $\mathbf{A}_{m\times n}$, $\mathbf{B}_{n\times p}$ y $\mathbf{C}_{p\times q}$. Entonces

   $$\begin{aligned}\mathrm{Vec}(\mathbf{ABC}) &= (\mathbf{C}^t \otimes \mathbf{A})\,\mathrm{Vec}(\mathbf{B}) = (\mathbf{C}^t\mathbf{B}^t \otimes \mathbf{I}_m)\,\mathrm{Vec}(\mathbf{A}) = (\mathbf{C}^t \otimes \mathbf{I}_m)\,\mathrm{Vec}(\mathbf{AB})\\ &= (\mathbf{I}_q \otimes \mathbf{AB})\,\mathrm{Vec}(\mathbf{C})\ .\end{aligned}$$

   d) Sean $\mathbf{B}_{m\times n}$, $\mathbf{C}_{n\times p}$ y $\mathbf{D}_{p\times m}$. Entonces

   $$\mathrm{tr}[\mathbf{BCD}] = \mathrm{Vec}'(\mathbf{B}^t)(\mathbf{I}_n \otimes \mathbf{C})\,\mathrm{Vec}(\mathbf{D}) = \mathrm{Vec}'(\mathbf{D}^t)(\mathbf{C}^t \otimes \mathbf{I}_m)\,\mathrm{Vec}(\mathbf{B})\ .$$

   e) Sea $\mathbf{A}_{n\times n}$. Entonces $\mathrm{tr}[\mathbf{A}] = \mathrm{Vec}'(\mathbf{A}^t)\,\mathrm{Vec}(\mathbf{I}_n) = \mathrm{Vec}'(\mathbf{I}_n)\,\mathrm{Vec}(\mathbf{A})$.

   f) Sean $\mathbf{A}_{m\times n}$, $\mathbf{B}_{m\times n}$, $\mathbf{C}_{n\times p}$ y $\mathbf{D}_{n\times p}$. Entonces

   $$\begin{aligned}\mathrm{Vec}((\mathbf{A}+\mathbf{B})(\mathbf{C}+\mathbf{D})) &= [(\mathbf{I}_p \otimes \mathbf{A}) + (\mathbf{I}_p \otimes \mathbf{B})]\,[\mathrm{Vec}(\mathbf{C}) + \mathrm{Vec}(\mathbf{D})]\\ &= \left[(\mathbf{C}^t \otimes \mathbf{I}_m) + (\mathbf{D}^t \otimes \mathbf{I}_m)\right]\,[\mathrm{Vec}(\mathbf{A}) + \mathrm{Vec}(\mathbf{B})]\ .\end{aligned}$$

3. Sea la matriz aleatoria $\mathbf{X}_{N\times p}$ de los ejemplos 1.2.1 y 1.3.1. Sea $\mathbf{Y}_{r\times s} = \mathbf{B}_{r\times p}\mathbf{X}^t\mathbf{C}_{N\times s}$. Probar que

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

   b) $\mathrm{Cov}\,[\mathrm{Vec}(\mathbf{Y})] = \mathbf{C}^t\mathbf{C} \otimes \mathbf{B}\Sigma\mathbf{B}^t$.

1.4. La matriz conmutación

Los primeros antecedentes sobre esta matriz hay que buscarlos en 1938 con Murhaghan en un texto sobre representaciones de grupos, o Vartak 1955 en las primeras aplicaciones del producto Kronecker en Estadística. Posteriormente han sido diversos los autores que la han considerado, incluso otorgándole diversos nombres. Así, por ejemplo, se le conoce como *matriz permutación* (Tracy y Dwyer, 1969), *matriz identidad permutada* (MacRae, 1974), *matriz conmutación* (Magnus y Neudecker, 1979), *matriz vec-permutación* (Searle, 1979) o *matriz conmutador tensorial* (Pollock, 1979). La diversidad de denominaciones viene dada bien por la forma de introducirla o bien por alguna propiedad concreta. Existen diversas formas de introducirla si bien todas inciden en la propiedad que esta matriz tiene para reordenar los elementos de una matriz (de ahí el nombre de *matriz permutación*). Una propiedad importante es la de transformar $\mathrm{Vec}(\mathbf{A})$ en $\mathrm{Vec}(\mathbf{A}^t)$, donde $\mathbf{A}_{n\times q}$ es una matriz arbitraria (de ahí el nombre de *matriz vec-permutación*). MacRae demostró que puede ser usado para invertir un producto Kronecker (que como sabemos no es conmutativo y de ahí el nombre de *matriz conmutador tensorial*), siendo esta una propiedad muy útil para el cálculo de derivadas matriciales así como para obtener esperanzas de vectores y matrices aleatorias (ver por ejemplo, Magnus y Neudecker, 1979).

En primer lugar notemos que, dada $\mathbf{A}_{n\times q}$, los vectores $\mathrm{Vec}(\mathbf{A})$ y $\mathrm{Vec}(\mathbf{A}^t)$ contienen los mismos elementos pero en un orden diferente. Por ello es lógico que exista una única matriz de permutación de orden $nq \times nq$ que transforme el vector $\mathrm{Vec}(\mathbf{A})$ en $\mathrm{Vec}(\mathbf{A}^t)$.

Definición 1.4.1. *Sea $\mathbf{A}_{n\times q}$. Se define la matriz conmutación como la matriz de permutación, $\mathbf{K}_{nq}$, de dimensiones $nq \times nq$ tal que $\mathbf{K}_{nq}\,\mathrm{Vec}(\mathbf{A}) = \mathrm{Vec}(\mathbf{A}^t)$.*

Nota 1.4.1. *Debido a que $\mathbf{K}_{nq}$ es una matriz de permutación, es ortogonal, por lo que $\mathbf{K}_{nq}^t = \mathbf{K}_{nq}^{-1}$. Además $\mathbf{K}_{qn}\mathbf{K}_{nq}\,\mathrm{Vec}(\mathbf{A}) = \mathbf{K}_{qn}\,\mathrm{Vec}(\mathbf{A}^t) = \mathrm{Vec}(\mathbf{A})$, de donde se puede concluir que $\mathbf{K}_{qn}\mathbf{K}_{nq} = \mathbf{I}_{nq}$ y con ello $\mathbf{K}_{nq}^{-1} = \mathbf{K}_{qn} = \mathbf{K}_{nq}^t$. Además es inmediato comprobar que $\mathbf{K}_{n1} = \mathbf{I}_n$ y $\mathbf{K}_{1q} = \mathbf{I}_q$.*

El siguiente resultado nos proporciona la forma explícita que adopta la matriz de conmutación.

Teorema 1.4.1. *Sea $\{\mathbf{J}_{ij} : i = 1, \ldots, n; j = 1 \ldots, q\}$ la base canónica del espacio de matrices*

$\mathbb{M}_{n\times q}$. *Entonces* $\mathbf{K}_{nq} = \sum_{i=1}^{n}\sum_{j=1}^{q} \mathbf{J}_{ij} \otimes \mathbf{J}_{ij}^t$.

Nota 1.4.2. *A partir del teorema 1.4.1 se podría haber introducido la matriz conmutación de forma constructiva sin más que comentar que es aquella que verifica* $\mathrm{Vec}(\mathbf{A}^t) = \sum_{i=1}^{n}\sum_{j=1}^{q}(\mathbf{J}_{ij} \otimes \mathbf{J}_{ij}^t)\,\mathrm{Vec}(\mathbf{A})$.

Ejemplo 1.4.1. *Sea* $\mathbf{A}_{3\times 2}$. *A partir del teorema 1.4.1 es inmediato comprobar que*

$$\mathbf{K}_{32} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

A partir del ejemplo anterior se puede observar que la matriz $\mathbf{K}_{32}$ no es más que una permutación de la matriz identidad $\mathbf{I}_6$. Ello motiva otra posible introducción de la matriz conmutación. Ahora se parte de un planteamiento más general en el sentido de definir las permutaciones de una matriz por filas y por columnas. Esta forma de considerar esta matriz fue introducida por Tracy y Dwyer (1969).

Sea $\mathbf{A}_{n\times q}$ y consideremos dos números naturales, r y s, que sean factores propios de n y q respectivamente. Definimos ahora las matrices $\mathbf{A}_{(r)}$ y $\mathbf{A}^{(s)}$ como sigue:

1. $\mathbf{A}_{(r)}$ es la matriz formada a partir de $\mathbf{A}$ colocando cada r-ésima fila de $\mathbf{A}$, empezando por la primera hasta acabar con el número de filas. Después se continúa con la segunda fila y así sucesivamente hasta completar la matriz. Por ejemplo, $\mathbf{A}^{(2)}$ se forma poniendo la primera fila de $\mathbf{A}$, luego la tercera, la quinta y así sucesivamente hasta concluir con las filas de $\mathbf{A}$. Después se tomaría la segunda fila, la cuarta y sucesivamente. A esta matriz la llamaremos la r-permutación por filas de $\mathbf{A}$.

2. $\mathbf{A}^{(s)}$ es la matriz formada siguiendo la misma idea anterior pero considerando cada s-ésima columna de $\mathbf{A}$. A esa matriz se la conoce como s-permutación por columnas de $\mathbf{A}$.

Ejemplo 1.4.2. *Sea* $\mathbf{A}$ *una matriz de orden* 6×4. *Las posibles permutaciones por filas son* $\mathbf{A}_{(2)}$ *y* $\mathbf{A}_{(3)}$, *mientras que por columnas sólo se puede calcular* $\mathbf{A}^{(2)}$. *En concreto se tendrá*

$$\mathbf{A}_{(2)} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{51} & a_{52} & a_{53} & a_{54} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{61} & a_{62} & a_{63} & a_{64} \end{pmatrix},\ \mathbf{A}_{(3)} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{51} & a_{52} & a_{53} & a_{54} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{61} & a_{62} & a_{63} & a_{64} \end{pmatrix} \text{ y } \mathbf{A}^{(2)} = \begin{pmatrix} a_{11} & a_{13} & a_{12} & a_{14} \\ a_{21} & a_{23} & a_{22} & a_{24} \\ a_{31} & a_{33} & a_{32} & a_{34} \\ a_{41} & a_{43} & a_{42} & a_{44} \\ a_{51} & a_{53} & a_{52} & a_{54} \\ a_{61} & a_{63} & a_{62} & a_{64} \end{pmatrix}$$

A continuación consideremos las matrices identidades de orden n y q y tomemos $\mathbf{I}_n\otimes\mathbf{I}_q$, producto que tiene por resultado $\mathbf{I}_{nq}$. Teniendo en cuenta los comentarios anteriores tenemos la siguiente definición alternativa para la matriz conmutación.

Definición 1.4.2. *Se define la matriz conmutación* $\mathbf{K}_{nq}$ *como la matriz resultante de la* n*-permutación por filas de la matriz* $\mathbf{I}_n \otimes \mathbf{I}_q$. *Esta definición asegura que* $\mathrm{Vec}(\mathbf{A}^t) = \mathbf{K}_{nq}\,\mathrm{Vec}(\mathbf{A})$.

Nota 1.4.3. *En virtud de la definición anterior, en ocasiones se usa la notación* $\mathbf{I}_{(q,n)}$ *para esta matriz, notación que fue introducida por MacRae en 1974). Nosotros emplearemos la más habitual de* $\mathbf{K}_{nq}$.

Nota 1.4.4. *Retomando el ejemplo 1.4.1, se tiene* $\mathbf{K}_{32} = \mathbf{I}_{6_{(3)}} = \mathbf{I}_{(2,3)}$.

Nota 1.4.5. *Evidentemente* $\mathbf{K}_{qn}$ *es la matriz resultante de la* q*-permutación por filas de la matriz* $\mathbf{I}_n \otimes \mathbf{I}_q$.

Otras expresiones alternativas de la matriz conmutación están recogidas en el siguiente resultado.

Teorema 1.4.2. *Sean* $\{\mathbf{e}_i : i = 1,\ldots,n\}$ *y* $\{\mathbf{u}_j : j = 1,\ldots,q\}$ *las bases canónicas de* $\mathbb{R}^n$ *y* $\mathbb{R}^q$, *respectivamente. Entonces*

$$\mathbf{K}_{nq} = \sum_{j=1}^{q}(\mathbf{u}_j^t \otimes \mathbf{I}_n \otimes \mathbf{u}_j) = \sum_{i=1}^{n}(\mathbf{e}_i \otimes \mathbf{I}_q \otimes \mathbf{e}_i^t)\ .$$

Enunciemos a continuación algunas propiedades de esta matriz.

Teorema 1.4.3. *Consideremos* $\mathbf{A}_{m\times n}$, $\mathbf{B}_{p\times q}$, $\mathbf{C}_{q\times s}$, $\mathbf{D}_{n\times t}$, $\mathbf{E}_{m\times n}$, $\mathbf{a}_{m\times 1}$, $\mathbf{b}_{p\times 1}$ *y* $\mathbf{z}$ *un vector cualquiera. Entonces*

1. $\mathbf{K}_{pm}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{qn}$.

2. $\mathbf{z}^t \otimes \mathbf{A} \otimes \mathbf{b} = \mathbf{K}_{mp}(\mathbf{b}\mathbf{z}^t \otimes \mathbf{A})$.

3. $\mathrm{tr}[\mathbf{K}_{mn}(\mathbf{A}^t \otimes \mathbf{E})] = \mathrm{tr}[\mathbf{A}^t\mathbf{E}] = \mathrm{Vec}'(\mathbf{A}^t)\mathbf{K}_{mn}\,\mathrm{Vec}(\mathbf{E})$.

4. $\mathrm{tr}[\mathbf{K}_{mn}] = 1 + f(m-1, n-1)$, *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* $\mathbf{K}_{nn}$ *son* 1 *con multiplicidad* $\frac{n(n+1)}{2}$ *y* -1 *con multiplicidad* $\frac{n(n-1)}{2}$.

6. $|\mathbf{K}_{nn}| = (-1)^{\frac{n(n-1)}{2}}$ *y* $|\mathbf{K}_{mn}| = (-1)^{\frac{m(m-1)n(n-1)}{4}}$.

7. *Supongamos que* $\mathrm{rg}(\mathbf{A}) = r$ *y sean* $\lambda_1, \ldots, \lambda_r$ *los autovalores de* $\mathbf{A}^t\mathbf{A}$. *Sea la matriz* $\mathbf{P} = \mathbf{K}_{mn}(\mathbf{A}^t \otimes \mathbf{A})$. *Entonces* $\mathbf{P}$ *es una matriz simétrica con rango* r^2, *con* $\mathrm{tr}[\mathbf{P}] = \mathrm{tr}[\mathbf{A}^t\mathbf{A}]$. *Además se verifica* $\mathbf{P}^2 = (\mathbf{A}\mathbf{A}^t) \otimes (\mathbf{A}^t\mathbf{A})$ *y los autovalores de* $\mathbf{P}$ *son los anteriores más* $\pm(\lambda_i\lambda_j)^{\frac{1}{2}}$, *con* $i \neq j$.

El siguiente resultado nos proporciona la expresión de la vectorización de un producto Kronecker.

Teorema 1.4.4. *Sean* $\mathbf{A}_{m\times n}$ *y* $\mathbf{B}_{p\times q}$. *Entonces* $\mathrm{Vec}(\mathbf{A}\otimes\mathbf{B}) = [\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p]\,[\mathrm{Vec}(\mathbf{A}) \otimes \mathrm{Vec}(\mathbf{B})]$.

A continuación vamos a generalizar la matriz conmutación a más de dos índices en el siguiente sentido. Sean las matrices $\mathbf{A}_{m\times n}$, $\mathbf{B}_{s\times t}$ y $\mathbf{C}_{p\times q}$. Definimos la matriz $\mathbf{K}_{st,n}$ de igual manera a $\mathbf{K}_{mn}$ donde se toma $m = st$ y actúa en la forma

$$\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C} = \mathbf{K}_{ms,p}(\mathbf{C} \otimes \mathbf{A} \otimes \mathbf{B})\mathbf{K}_{q,nt} = \mathbf{K}_{m,sp}(\mathbf{B} \otimes \mathbf{C} \otimes \mathbf{A})\mathbf{K}_{tq,n} \; .$$

Teorema 1.4.5. *Se verifican las siguientes afirmaciones*

1. $\mathbf{K}_{mn,p} = \mathbf{K}_{nm,p}$ *y* $\mathbf{K}_{m,np} = \mathbf{K}_{m,pn}$.

2. $\mathbf{K}_{mn,p} = (\mathbf{I}_m \otimes \mathbf{K}_{np})(\mathbf{K}_{mp} \otimes \mathbf{I}_n) = (\mathbf{I}_n \otimes \mathbf{K}_{mp})(\mathbf{K}_{np} \otimes \mathbf{I}_m)$.

3. $\mathbf{K}_{m,np} = (\mathbf{K}_{mn} \otimes \mathbf{I}_p)(\mathbf{I}_n \otimes \mathbf{K}_{mp}) = (\mathbf{K}_{mp} \otimes \mathbf{I}_n)(\mathbf{I}_p \otimes \mathbf{K}_{mn})$.

4. $\mathbf{K}_{mn,p} = \mathbf{K}_{m,np}\mathbf{K}_{n,pm} = \mathbf{K}_{n,mp}\mathbf{K}_{m,pn}$ *y* $\mathbf{K}_{m,np} = \mathbf{K}_{mn,p}\mathbf{K}_{pm,n} = \mathbf{K}_{mp,n}\mathbf{K}_{nm,p}$.

5. $\mathbf{I}_{nmp} = \mathbf{K}_{mn,p}\mathbf{K}_{pm,n}\mathbf{K}_{np,m} = \mathbf{K}_{mn,p}\mathbf{K}_{pn,m}\mathbf{K}_{mp,n} = \mathbf{K}_{m,np}\mathbf{K}_{n,pm}\mathbf{K}_{p,mn} = \mathbf{K}_{m,np}\mathbf{K}_{p,nm}\mathbf{K}_{n,mp}$.

6. Cualquier producto de dos matrices de este tipo con el mismo conjunto de índices conmuta.

1.4.1. Ejercicios

1. A partir de la expresión explícita de la matriz $\mathbf{K}_{nn}$, demostrar que $\mathrm{tr}[\mathbf{K}_{nn}] = n$.

2. Sean $\mathbf{A}_{m\times n}$, $\mathbf{B}_{p\times q}$, $\mathbf{C}_{q\times s}$, $\mathbf{D}_{n\times t}$ y $\mathbf{b}_{p\times 1}$. Demostrar las siguientes igualdades:

   a) $\mathbf{K}_{pm}(\mathbf{A}\otimes\mathbf{B})\mathbf{K}_{nq} = \mathbf{B}\otimes\mathbf{A}$.

   b) $\mathbf{K}_{pm}(\mathbf{A}\otimes\mathbf{b}) = \mathbf{b}\otimes\mathbf{A}$.

   c) $\mathbf{K}_{mp}(\mathbf{b}\otimes\mathbf{A}) = \mathbf{A}\otimes\mathbf{b}$.

3. Sea $\mathbf{N}_n = \frac{1}{2}\left[\mathbf{I}_{n^2} + \mathbf{K}_{nn}\right]$. Probar las siguientes afirmaciones:

   a) $\mathbf{N}_n$ es simétrica e idempotente.

   b) $\mathrm{rg}(\mathbf{N}_n) = \mathrm{tr}[\mathbf{N}_n] = \frac{1}{2}n(n+1)$.

   c) $\mathbf{N}_n\mathbf{K}_{nn} = \mathbf{K}_{nn}\mathbf{N}_n = \mathbf{N}_n$.

1.5. Producto $*$ de dos matrices

A continuación vamos a introducir otro producto entre matrices cuya utilidad principal va a ser la de simplificar cálculos a la hora de obtener derivadas matriciales. La siguiente definición es debida a MacRae en 1974.

Definición 1.5.1. *Sean* $\mathbf{X}_{m\times n}$ *e* $\mathbf{Y}_{mp\times nq}$*. Particionemos* $\mathbf{Y}$ *en la forma*

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

Se define $\mathbf{X}*\mathbf{Y}$ *como la matriz* $p\times q$ $\mathbf{X}*\mathbf{Y} = \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}\mathbf{Y}_{ij}$.

El siguiente resultado recoge algunas propiedades de esta operación matricial.

Teorema 1.5.1. *Sean* $\mathbf{X}_{m\times n}$, $\mathbf{Y}_{mp\times nq}$, $\mathbf{W}_{mp\times nq}$ *y* $\mathbf{Z}_{s\times t}$. *Entonces:*

1. $(\mathbf{X}*\mathbf{Y})^t = \mathbf{X}*\mathbf{Y}^t$.

2. *Si* $p=q=1$, *entonces* $\mathbf{X}*\mathbf{Y} = \operatorname{tr}[\mathbf{X}\mathbf{Y}^t] = \operatorname{tr}[\mathbf{Y}\mathbf{X}^t] = \operatorname{tr}[\mathbf{X}^t\mathbf{Y}] = \operatorname{tr}[\mathbf{Y}^t\mathbf{X}]$.

3. $(\mathbf{X}*\mathbf{Y})\otimes\mathbf{Z} = \mathbf{X}*(\mathbf{Y}\otimes\mathbf{Z})$.

4. $\mathbf{X}*(\mathbf{Y}+\mathbf{W}) = \mathbf{X}*\mathbf{Y}+\mathbf{X}*\mathbf{W}$.

Por último he aquí otras propiedades importantes acerca del producto $*$.

Teorema 1.5.2. *Sean* $\mathbf{X}_{m\times n}$, $\mathbf{Y}_{n\times p}$, *y* $\mathbf{Z}_{p\times q}$. *Entonces:*

1. $\mathbf{XYZ} = \mathbf{Y}*\operatorname{Vec}(\mathbf{X})\operatorname{Vec}'(\mathbf{Z}^t)$.

2. $\mathbf{XYZ} = \mathbf{Y}^t*(\mathbf{Z}\otimes\mathbf{I}_m)\mathbf{K}_{qm}(\mathbf{X}\otimes\mathbf{I}_q)$.

1.5.1. Ejercicios

1. Sean $\mathbf{X}_{m\times n}$, $\mathbf{Y}_{n\times p}$. Demostrar que $\mathbf{Y}*\mathbf{K}_{np} = \mathbf{Y}^t$.

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

Consideremos $\mathbf{A}_{n\times n}$ y notemos por $\operatorname{Vech}(\mathbf{A})$ al vector $\frac{1}{2}n(n+1)$ dimensional obtenido a partir del vector $\operatorname{Vec}(\mathbf{A})$ eliminando los elementos de la diagonal superior de $\mathbf{A}$. Por ejemplo, si $n=3$ tenemos $\operatorname{Vech}(\mathbf{A}) = (a_{11},a_{21},a_{31},a_{22},a_{32},a_{33})^t$.

En particular, si $\mathbf{A}$ es simétrica entonces $\operatorname{Vech}(\mathbf{A})$ contiene sus elementos distintos. Puesto que los elementos de $\operatorname{Vec}(\mathbf{A})$ son los de $\operatorname{Vech}(\mathbf{A})$ con algunos elementos repetidos, debe existir una matriz que sea de orden $n^2\times\frac{1}{2}n(n+1)$ y que transforme $\operatorname{Vec}(\mathbf{A})$ en $\operatorname{Vech}(\mathbf{A})$. Esta matriz recibe el nombre de *matriz duplicación* y la notaremos $\mathbf{D}_n$.

Este planteamiento no es más que un caso particular de matrices asociadas a un nuevo operador, Vech. En efecto, consideremos una matriz $\mathbf{A}_{m\times n}$ cuyos elementos no sean todos matemáticamente independientes. Por ejemplo, dicha matriz contiene s elementos repetidos y/o v

elementos constantes, de tal forma que el número de elementos matemáticamente independientes son $r = mn - s - v$. Algunos posibles ejemplos son los siguientes:

1. $\mathbf{A}$ es simétrica. En este caso $v = 0$ y $s = m(m-1)/2$, por lo que $r = m(m+1)/2$.
2. $\mathbf{A}$ es antisimétrica ($\mathbf{A} = -\mathbf{A}^t$). En este caso $v = m$ y $s = m(m-1)/2$, por lo que $r = m(m-1)/2$.
3. $\mathbf{A}$ es triangular inferior o superior. En este caso $s = 0$ y $s = m(m-1)/2$, por lo que $r = m(m+1)/2$.
4. $\mathbf{A}$ es diagonal. En este caso $v = m(m-1)/2$ y $s = 0$, por lo que $r = m$.
5. $\mathbf{A}$ es una matriz de correlación. En este caso $v = m$ y $s = m(m-1)/2$, por lo que $r = m(m-1)/2$.

Para todas estas matrices podemos considerar el vector formado a partir de su vectorización excluyendo los elementos repetidos y/o constantes. Este operador es conocido, en general, como $\text{Vech}(\mathbf{A})$ si bien algunos autores eligen la notación $\text{Vecp}(\mathbf{A})$ o $v(\mathbf{A})$. Los operadores $\text{Vec}(\mathbf{A})$ y $\text{Vech}(\mathbf{A})$ deben estar relacionados puesto que ordenan los elementos de la matriz $\mathbf{A}$ salvo los excluidos. La relación entre ellos se consigue mediante una matriz que transforme un vector en el otro, matriz que se conoce genéricamente como *matriz de transición*. En el caso de matrices simétricas esta matriz es conocida como *matriz duplicación* (Magnus y Neudecker, 1978).

Definición 1.6.1. *Sea $\mathbf{A}_{m\times n}$ y r el número de elementos de $\mathbf{A}$ 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 $\mathbf{TR}\,\text{Vech}(\mathbf{A}) = \text{Vec}(\mathbf{A})$.*

Definición 1.6.2. *Si $\mathbf{A}_{n\times n}$ es simétrica, se define la matriz duplicación como aquella matriz de orden $n^2 \times \frac{1}{2}n(n+1)$ que verifique la relación $\mathbf{D}_n\,\text{Vech}(\mathbf{A}) = \text{Vec}(\mathbf{A})$.*

La forma explícita que adopta esta matriz queda expuesta en el siguiente resultado.

Teorema 1.6.1. *Sea $\mathbf{A}$ una matriz simétrica de orden n. Entonces las columnas de $\mathbf{D}_n$ vienen dadas por $\mathbf{e}_{(i-1)n+j} + \mathbf{e}_{(j-1)n+i}(1-\delta_{ij})$, $j = 1,\ldots,n$, $i = j,\ldots,n$, donde los vectores anteriores son vectores básicos de $\mathbb{R}^{n^2}$ y δ_{ij} representa la función delta de Kronecker.*

Nota 1.6.1. *A la vista del resultado anterior es inmediato comprobar que* $\mathbf{D}_n$ *tiene rango por columnas completo. En efecto, si* $\mathbf{D}_n \operatorname{Vech}(\mathbf{A}) = 0$, *entonces* $\operatorname{Vec}(\mathbf{A}) = 0$ *y con ello* $\operatorname{Vech}(\mathbf{A}) = 0$. *Por lo tanto su g-inversa viene dada por* $(\mathbf{D}_n)_g = (\mathbf{D}_n^t \mathbf{D}_n)^{-1}\mathbf{D}_n^t$.

El siguiente teorema resume algunos resultados asociados a la matriz duplicación

Teorema 1.6.2. *Sean* $\mathbf{b}_{n\times 1}$ *y* $\mathbf{A}_{n\times n}$. *Entonces se verifica*

1. $\mathbf{K}_{nn}\mathbf{D}_n = \mathbf{D}_n$.

2. $\mathbf{D}_n(\mathbf{D}_n)_g = \mathbf{N}_n$.

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

4. $\mathbf{D}_n(\mathbf{D}_n)_g(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n = (\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$.

5. $\mathbf{D}_n(\mathbf{D}_n)_g(\mathbf{A} \otimes \mathbf{A})(\mathbf{D}_n)_g^t = (\mathbf{A} \otimes \mathbf{A})(\mathbf{D}_n)_g^t$.

donde $\mathbf{N}_n = \frac{1}{2}[\mathbf{I}_{n^2} + \mathbf{K}_{nn}]$.

En el caso de matrices diagonales o triangulares disponemos de resultados análogos a los anteriores. En concreto el siguiente teorema, cuya demostración se deja como ejercicio, proporciona las expresiones que adoptan las correspondientes matrices de transición.

Teorema 1.6.3. *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* $\mathbf{e}_{(i-1)n+i}$, $i = 1, \ldots, n$.

2. *Las columnas de* $\mathbf{TSup}_n$ *vienen dadas por* $\mathbf{e}_{(j-1)n+i}$, $j = 1, \ldots, n$, $i = 1, \ldots, j$.

3. *Las columnas de* $\mathbf{TInf}_n$ *vienen dadas por* $\mathbf{e}_{(j-1)n+i}$, $j = 1, \ldots, n$, $i = j, \ldots, n$.

4. *Las columnas de* $\mathbf{TInf}_n$ *vienen dadas por* $\mathbf{e}_{(j-1)n+i}$, $j = 1, \ldots, n-1$, $i = j+1, \ldots, n$.

donde los vectores $\mathbf{e}_k$ *son vectores básicos de* $\mathbb{R}^{n^2}$.

Nota 1.6.2. *Es obvio que las matrices* $\mathbf{Di}_n$, $\mathbf{TSup}_n$ *y* $\mathbf{TInf}_n$ *tienen rango por columnas completo ya que dichas columnas son vectores básicos de* $\mathbb{R}^{n^2}$. *Además, el producto de ellas por su transpuesta es la matriz identidad y, como consecuencia, sus g-inversas coinciden con sus transpuestas.*

Estos dos últimos resultados no serán demostrados y serán usados en un futuro:

1. $\mathbf{D}_n^t \operatorname{Vec}(\mathbf{A}) = \operatorname{Vech}(\mathbf{A} + \mathbf{A}^t - \operatorname{diag}(\mathbf{A}))$.

2. $|\mathbf{D}_n^-(\mathbf{A}^{-1} \otimes \mathbf{A}^{-1})(\mathbf{D}_n^-)^t| = 2^{-\frac{n(n-1)}{2}} |\mathbf{A}|^{n+1}$.

1.6.1. Ejercicios

1. Si $\mathbf{A}$ es no singular, $[(\mathbf{D}_n)_g(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n]^{-1} = (\mathbf{D}_n)_g(\mathbf{A} \otimes \mathbf{A})^{-1}\mathbf{D}_n$.

2. Si $\mathbf{A}$ es no singular, $\left[\mathbf{D}_n'(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n\right]^{-1} = (\mathbf{D}_n)_g(\mathbf{A} \otimes \mathbf{A})^{-1}(\mathbf{D}_n)_g^t$.

Capítulo 2

Derivación matricial

2.1. Introducción

La notación vectorial y matricial es una herramienta muy importante en Estadística, sobre todo en Análisis Multivariante, donde gran cantidad de modelos y funciones asociadas a ellos deben ser escritos de forma resumida, dada su complejidad, en términos de vectores y matrices. Este hecho conlleva una mayor claridad en las expresiones pero puede presentar algún tipo de inconveniente desde el punto de vista operativo. En efecto, la complejidad de las expresiones y el tipo de resultados que se pretenden obtener a partir de ellas ha hecho necesario un esfuerzo adicional en el campo del Álgebra Matricial, como se puede deducir tras el capítulo anterior.

En este capítulo pretendemos dar una introducción general al problema de la derivación respecto de vectores y, principalmente, matrices. Este es un problema muy frecuente en ambiente multivariante sobre todo relacionado con la optimización de determinadas funciones. Pensemos en la obtención de estimadores máximo-verosímiles o mínimo-cuadráticos, o en la obtención de matrices de información de Fisher o cotas tipo Crámer-Rao o cualquier otro problema donde se deba optimizar alguna función respecto de parámetros vectoriales o matriciales.

Una cuestión adicional que hay que tener en cuenta es que las funciones de las que estamos hablando no tienen por qué ser solamente escalares o vectoriales sino que en un gran número de ocasiones son matriciales. Ello complica aún más la situación ya que si bien el análisis real nos dice que la matriz derivada de una función $f : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ es una matriz $m \times n$ de componentes $\frac{\partial f_i(\mathbf{x})}{\partial x_j}$, $i = 1, \ldots m$, $j = 1, \ldots, n$ (matriz jacobiana), en un ambiente más genérico con funciones matriciales de argumento matricial, $F : S \subseteq \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$, dispondremos en principio de *mpnq* derivadas parciales $\frac{\partial F_{ij}(\mathbf{X})}{\partial x_{uv}}$, de tal forma que nos encontramos ante el problema de *ordenar* dichas parciales en una matriz adecuada. Es precisamente este punto el que plantea más controversia puesto que han sido varias las aproximaciones que se han dado como solución a la pregunta planteada y que desarrollaremos en este capítulo.

Los orígenes de la derivación matricial, como la conocemos hoy día, podemos encontrarlos en autores como Dwyer y MacPhail (1948). Posteriormente, Trawinski y Bargmann (1964) desarrollaron un cálculo matricial muy próximo al de Dwyer pero cuya notación era mucho más simple, si bien seguía incidiendo en funciones escalares de matrices y funciones matriciales de escalares. En 1974, MacRae extiende el cálculo de Bargmann al caso de funciones matriciales de matrices con la ayuda de operaciones como el producto Kronecker y el producto $*$. Por último, otro grupo de autores entre los que destacan Neudecker (1969), Tracy y Dwyer (1969), McDonald y Swaminathan (1973) desarrollan un nuevo procedimiento mediante la *vectorización* de las matrices implicadas en el proceso. De esta forma el problema se trasladaba a la situación vectorial con algunos beneficios: por una parte la operatividad y por otra parte se extendía el cálculo en varias variables al matricial, lo que permite hablar de diferenciabilidad de una función matricial de argumento matricial, comprobar que la matriz de derivadas es la matriz jacobiana y hablar de matrices hessianas. Pero, por otro lado, esta nueva situación planteó tener que seguir desarrollando nuevos aspectos de álgebra matricial (matrices de transición, matrices de duplicación,...) amén de no resolver del todo el problema, sobre todo en lo que se refiere a derivadas de orden superior.

Empezaremos tratando precisamente la última aproximación comentando. Así, en la línea de Magnus y Neudecker(1999) definiremos la diferenciabilidad de una función matricial (previo resumen del caso vectorial) y se introducirá la matriz jacobiana junto con resultados que permitan identificarla a partir de la diferencial de una función matricial y muestren diversas propiedades. Posteriormente trataremos algunas definiciones alternativas en el caso de argumento matricial (MacRae y Dwyer) con sus respectivas propiedades. Como es lógico, relacionaremos las tres principales definiciones y terminaremos con algunos resultados para determinadas funciones de amplio uso en Estadística.

2.2. Diferencial primera y jacobianos

2.2.1. Diferencial de una función vectorial

Definición 2.2.1. *Consideremos una función vectorial* $\mathbf{f} : S \to \mathbb{R}^m$ *con* $S \subseteq \mathbb{R}^n$. *Sea* $\mathbf{c}$ *un punto interior de* S *y consideremos una bola centrada en* $\mathbf{c}$ *y de radio* r, $\mathbb{B}(\mathbf{c}; r)$. *Sea* $\mathbf{u}$ *un punto de* $\mathbb{R}^n$ *con* $||\mathbf{u}|| < r$, *o sea,* $\mathbf{c} + \mathbf{u} \in \mathbb{B}(\mathbf{c}; r)$.

Diremos que $\mathbf{f}$ *es diferenciable en* $\mathbf{c}$ *si existe una matriz real de orden* $m \times n$, *que dependede* $\mathbf{c}$ *y no de* $\mathbf{u}$, *tal que se verifique* $\mathbf{f}(\mathbf{c} + \mathbf{u}) - \mathbf{f}(\mathbf{c}) = \mathbf{A}(\mathbf{c})\mathbf{u} + \mathbf{r_c}(\mathbf{u})$ *con* $\lim\limits_{\mathbf{u}\to\mathbf{0}} \dfrac{\mathbf{r_c}(\mathbf{u})}{||\mathbf{u}||} = 0$. *Además, se define la primera diferencial de* $\mathbf{f}$ *en el punto* $\mathbf{c}$ *con incremento* $\mathbf{u}$ *como* $\mathbf{df}(\mathbf{c}; \mathbf{u}) = \mathbf{A}(\mathbf{c})\mathbf{u}$.

Nota 2.2.1. *Si* $\mathbf{f}$ *es diferenciable en* $\mathbf{c}$ *entonces* $\mathbf{f}$ *es continua en* $\mathbf{c}$.

Nota 2.2.2. *La expresión* $\mathbf{f}(\mathbf{c} + \mathbf{u}) - \mathbf{f}(\mathbf{c}) = \mathbf{A}(\mathbf{c})\mathbf{u} + \mathbf{r_c}(\mathbf{u})$ *es una expresión vectorial con componentes* $f_i(\mathbf{c}+\mathbf{u}) - f_i(\mathbf{c}) = \sum_{j=1}^{n} a_{ij}(\mathbf{c})u_j + r_c^i(\mathbf{u})$ *con* $\lim\limits_{\mathbf{u}\to\mathbf{0}} \dfrac{r_c^i(\mathbf{u})}{||\mathbf{u}||} = 0$, *de donde se deduce que* $\mathbf{f}$ *es diferenciable en* $\mathbf{c}$ *sí y solamente sí* f_i *lo es en* $\mathbf{c}$ $(i = 1, \ldots, m)$.

Definición 2.2.2. *Sea* $\mathbf{f} : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *y sea* $f_i : S \to \mathbb{R}$ *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* $\mathbf{f}$ *respecto a la* j*-ésima coordenada como*

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

Veamos a continuación un teorema muy útil ya que va a ser el que nos ayude en su momento a identificar los jacobianos asociados a este tipo de funciones puesto que nos indica la existencia de una correspondencia biunívoca entre la diferencial de una función vectorial $\mathbf{f}$ y su matriz jacobiana. Asimismo nos asegura la unicidad de la diferencial.

Teorema 2.2.1. *(Primer teorema de identificación para funciones vectoriales)*

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

Nota 2.2.3. *La existencia de las derivadas parciales es necesaria pero no suficiente para la diferenciabilidad de una función (ver ejercicio 2 del apartado 2.2.3). Sin embargo, hay que hacer notar que si las parciales existieran y fueran continuas entonces sí es cierto que ello implica que la función es diferenciable.*

A continuación enunciamos el teorema que nos proporciona la regla de la cadena para funciones vectoriales.

Teorema 2.2.2. *Sea* $\mathbf{f} : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *diferenciable en un punto* $\mathbf{c}$ *interior de* S. *Sea* T *un subconjunto de* $\mathbb{R}^m$ *tal que* $\mathbf{f}(\mathbf{x}) \in T$, $\forall \mathbf{x} \in S$ *y supongamos que* $\mathbf{g} : T \to \mathbb{R}^p$ *es diferenciable en un punto* $\mathbf{b}$ *(*$\mathbf{b} = \mathbf{f}(\mathbf{c})$*) de* T. *Entonces la función compuesta* $\mathbf{h} : S \to \mathbb{R}^p$ *definida por* $\mathbf{h}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$ *es diferenciable en* $\mathbf{c}$ *y* $\mathbf{Dh}(\mathbf{c}) = \mathbf{Dg}(\mathbf{b})\mathbf{Df}(\mathbf{c})$.

La regla de la cadena relaciona las derivadas parciales de la función compuesta $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ con las de $\mathbf{f}$ y $\mathbf{g}$, pero ¿cómo lo hacen las diferenciales? El siguiente resultado, conocido como regla de invarianza de Cauchy, resuelve esta pregunta.

Teorema 2.2.3. *(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 diferencial de* $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ *es* $\mathbf{dh}(\mathbf{c}; \mathbf{u}) = \mathbf{dg}(\mathbf{b}; \mathbf{df}(\mathbf{c}; \mathbf{u}))$.

2.2.2. Diferencial de una función matricial

El caso matricial se puede tratar de forma inmediata a partir del vectorial. Para ello basta con extrapolar lo que ocurre en el caso vectorial haciendo uso de la operación Vec.

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

Definición 2.2.3. *En las condiciones anteriores, se dice que* $\mathbf{F}$ *es diferenciable en* $\mathbf{C}$ *si existe una matriz* $\mathbf{A}$ *de dimensiones* $mp \times nq$, *que dependa de* $\mathbf{C}$ *y no de* $\mathbf{U}$ *y tal que*

$$\mathrm{Vec}\,(\mathbf{F}(\mathbf{C}+\mathbf{U})) - \mathrm{Vec}\,(\mathbf{F}(\mathbf{C})) = \mathbf{A}(\mathbf{C})\,\mathrm{Vec}(\mathbf{U}) + \mathrm{Vec}\,(\mathbf{R_C}(\mathbf{U})) \quad \textit{donde} \quad \lim_{\mathbf{U}\to\mathbf{0}} \frac{\mathbf{R_C}(\mathbf{U})}{||\mathbf{U}||} = 0\ .$$

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

Nota 2.2.4. *A la vista de la anterior definición, todas las propiedades de cálculo para las funciones matriciales se siguen de las correspondientes propiedades de las funciones vectoriales ya que en lugar de la función matricial* $\mathbf{F}$ *se considerará la función vectorial* $\mathbf{f} : \mathrm{Vec}(S) \to \mathbb{R}^{mp}$ *dada por* $\mathbf{f}\left(\mathrm{Vec}(\mathbf{X})\right) = \mathrm{Vec}\left(\mathbf{F}(\mathbf{X})\right)$. *De esta forma, notemos que*

$$\mathrm{Vec}(\mathbf{F}(\mathbf{C}+\mathbf{U})) - \mathrm{Vec}(\mathbf{F}(\mathbf{C})) = \mathbf{f}(\mathrm{Vec}(\mathbf{C}+\mathbf{U})) - \mathbf{f}(\mathrm{Vec}(\mathbf{C})) = \mathbf{df}(\mathrm{Vec}(\mathbf{C});\mathrm{Vec}(\mathbf{U})) + \mathrm{Vec}(\mathbf{R}_{\mathbf{C}}(\mathbf{U}))$$

por lo que $\mathrm{Vec}\left(\mathbf{dF}(\mathbf{C};\mathbf{U})\right) = \mathbf{df}\left(\mathrm{Vec}(\mathbf{C});\mathrm{Vec}(\mathbf{U})\right) = \mathbf{A}(\mathbf{C})\,\mathrm{Vec}(\mathbf{U})$. *Con ello se define la matriz jacobiana de* $\mathbf{F}$ *en* $\mathbf{C}$ *como* $\mathbf{DF}(\mathbf{C}) = \mathbf{Df}\left(\mathrm{Vec}(\mathbf{C})\right)$, *que es una matriz* $mp \times nq$ *cuyo* ij*-ésimo elemento es la parcial del* i*-ésimo elemento de* $\mathrm{Vec}\left(\mathbf{F}(\mathbf{X})\right)$ *respecto al* j*-ésimo elemento de* $\mathrm{Vec}(\mathbf{X})$, *evaluada en* $\mathbf{X} = \mathbf{C}$.

Al igual que ocurría en el caso vectorial, también hace falta ahora un teorema que ayude a identificar los jacobianos de funciones matriciales que se consideren con posterioridad.

Teorema 2.2.4. *(Primer teorema de identificación para funciones matriciales)*

Sea $\mathbf{F} : S \subseteq \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$ *diferenciable en un punto interior* $\mathbf{C}$ *de* S. *Entonces se verifica* $\mathrm{Vec}\left(\mathbf{dF}(\mathbf{C};\mathbf{U})\right) = \mathbf{A}(\mathbf{C})\,\mathrm{Vec}(\mathbf{U}) \Leftrightarrow \mathbf{DF}(\mathbf{C}) = \mathbf{A}(\mathbf{C})$.

Ejemplo 2.2.1. *Sea* $\mathbf{F}(\mathbf{X}) = \mathbf{AXB}$, *donde* $\mathbf{A}_{m\times n}$ *y* $\mathbf{B}_{q\times p}$ *son matrices constantes y* $\mathbf{X}$ *es de dimensiones* $n \times q$. *Se tiene*

$$\mathrm{Vec}(\mathbf{F}(\mathbf{C}+\mathbf{U})) - \mathrm{Vec}(\mathbf{F}(\mathbf{C})) = \mathrm{Vec}(\mathbf{A}(\mathbf{C}+\mathbf{U})\mathbf{B}) - \mathrm{Vec}(\mathbf{ACB}) = \mathrm{Vec}(\mathbf{AUB}) = (\mathbf{B}^t \otimes \mathbf{A})\,\mathrm{Vec}(\mathbf{U})$$

de donde $\mathrm{Vec}(\mathbf{dF}(\mathbf{C};\mathbf{U})) = (\mathbf{B}^t \otimes \mathbf{A})\,\mathrm{Vec}(\mathbf{U})$ *y con ello* $\mathbf{dF}(\mathbf{C};\mathbf{U}) = \mathbf{AUB}$ *y* $\mathbf{DF}(\mathbf{C}) = (\mathbf{B}^t \otimes \mathbf{A})$.

De igual manera se tienen los siguientes resultados, análogos a los del caso vectorial:

Teorema 2.2.5. *(Regla de la cadena para funciones matriciales)*

Sea $\mathbf{F} : S \subseteq \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$ *diferenciable en un punto* $\mathbf{C}$ *interior de* S. *Sea* T *un subconjunto de* $\mathbb{M}_{m\times p}$ *tal que* $\mathbf{F}(\mathbf{X}) \in T$, $\forall \mathbf{X} \in S$ *y supongamos que* $\mathbf{G} : T \to \mathbb{M}_{r\times s}$ *es diferenciable en un punto* $\mathbf{B}$ *(*$\mathbf{B} = \mathbf{F}(\mathbf{C})$*) de* T. *Entonces la función compuesta* $\mathbf{H} : S \to \mathbb{M}_{r\times s}$ *definida por* $\mathbf{H}(\mathbf{X}) = \mathbf{G}(\mathbf{F}(\mathbf{X}))$ *es diferenciable en* $\mathbf{C}$ *y* $\mathbf{DH}(\mathbf{C}) = \mathbf{DG}(\mathbf{B})\mathbf{DF}(\mathbf{C})$.

Teorema 2.2.6. *(Regla de invarianza de Cauchy para funciones matriciales)*

En las condiciones del teorema anterior, $\mathbf{dH}(\mathbf{C};\mathbf{U}) = \mathbf{dG}(\mathbf{B};\mathbf{dF}(\mathbf{C};\mathbf{U}))$, $\forall \mathbf{U} \in \mathbb{M}_{n\times q}$.

El siguiente resultado nos proporciona algunas propiedades de las diferenciales matriciales así como de las matrices jacobianas:

Teorema 2.2.7. *Sean* $\mathbf{F}$ *y* $\mathbf{G}$ *dos funciones matriciales,* $\mathbf{A}$ *una matriz constante y* $\alpha \in \mathbb{R}$.

1. *Si* $\mathbf{F}(\mathbf{X}) = \mathbf{A}_{m\times p}$, *entonces* $\mathbf{F}$ *es diferenciable. Además* $\mathbf{dF}(\mathbf{X}) = \mathbf{0}$ *y* $\mathbf{DF}(\mathbf{X}) = \mathbf{0}$.

2. *Si* $\mathbf{G}_{m\times p}$ *es diferenciable y* α *un escalar distinto de cero y* $\mathbf{F}(\mathbf{X}) = \alpha\mathbf{G}(\mathbf{X})$, *entonces* $\mathbf{F}$ *es diferenciable. Además* $\mathbf{dF}(\mathbf{X}) = \alpha\mathbf{dG}(\mathbf{X})$ *y* $\mathbf{DF}(\mathbf{X}) = \alpha\mathbf{DG}(\mathbf{X})$.

3. *Si* $\mathbf{G}_{m\times p}$ *es diferenciable y* $\mathbf{F}(\mathbf{X}) = \mathrm{Vec}(\mathbf{G}(\mathbf{X}))$, *entonces* $\mathbf{F}$ *es diferenciable. Además se verifica* $\mathbf{dF}(\mathbf{X}) = \mathrm{Vec}(\mathbf{dG}(\mathbf{X}))$ *y* $\mathbf{DF}(\mathbf{X}) = \mathbf{DG}(\mathbf{X})$.

4. *Si* $\mathbf{G}_{m\times p}$ *es diferenciable y* $\mathbf{F}(\mathbf{X}) = \mathbf{G}^t(\mathbf{X})$, *entonces* $\mathbf{F}$ *es diferenciable. Además se verifica* $\mathbf{dF}(\mathbf{X}) = (\mathbf{dG}(\mathbf{X}))^t$ *y* $\mathbf{DF}(\mathbf{X}) = \mathbf{K}_{mp}\mathbf{DG}(\mathbf{X})$.

5. *Si* $\mathbf{G}_{p\times p}$ *es diferenciable y* $\mathbf{F}(\mathbf{X}) = \mathrm{tr}[\mathbf{G}(\mathbf{X})]$, *entonces* $\mathbf{F}$ *es diferenciable. Además se verifica* $\mathbf{dF}(\mathbf{X}) = \mathrm{tr}[\mathbf{dG}(\mathbf{X})]$ *y* $\mathbf{DF}(\mathbf{X}) = \mathrm{Vec}'(\mathbf{I}_p)\mathbf{DG}(\mathbf{X})$.

6. *Si* $\mathbf{G}_{m\times p}$ *y* $\mathbf{H}_{m\times p}$ *son diferenciables y* $\mathbf{F}(\mathbf{X}) = (\mathbf{G} \pm \mathbf{H})(\mathbf{X})$, *entonces* $\mathbf{F}$ *es diferenciable. Además* $\mathbf{dF}(\mathbf{X}) = \mathbf{dG}(\mathbf{X}) \pm \mathbf{dH}(\mathbf{X})$ *y* $\mathbf{DF}(\mathbf{X}) = \mathbf{DG}(\mathbf{X}) \pm \mathbf{DH}(\mathbf{X})$.

7. *Si* $\mathbf{G}_{m\times r}$ *y* $\mathbf{H}_{r\times p}$ *son diferenciables y* $\mathbf{F}(\mathbf{X}) = (\mathbf{GH})(\mathbf{X})$, $\mathbf{F}$ *es diferenciable. Además* $\mathbf{dF}(\mathbf{X}) = \mathbf{dG}(\mathbf{X})\mathbf{H}(\mathbf{X}) + \mathbf{G}(\mathbf{X})\mathbf{dH}(\mathbf{X})$ *y* $\mathbf{DF}(\mathbf{X}) = (\mathbf{H}^t(\mathbf{X}) \otimes \mathbf{I}_m)\mathbf{DG}(\mathbf{X}) + (\mathbf{I}_p \otimes \mathbf{G}(\mathbf{X}))\mathbf{DH}(\mathbf{X})$.

8. *Si* $\mathbf{G}_{m\times m}$ *es diferenciable y* $\mathbf{F}(\mathbf{X}) = (\mathbf{G}(\mathbf{X}))^{-1}$, *entonces para los puntos en los que* $\mathbf{F}$ *exista se verifica que dicha función es diferenciable. Además* $\mathbf{dF}(\mathbf{X}) = -\mathbf{F}(\mathbf{X})\mathbf{dG}(\mathbf{X})\mathbf{F}(\mathbf{X})$ *y* $\mathbf{DF}(\mathbf{X}) = -(\mathbf{F}^t(\mathbf{X}) \otimes \mathbf{F}(\mathbf{X}))\mathbf{DG}(\mathbf{X})$.

9. *Dadas* $\mathbf{G}_{m\times p}$ *y* Φ *una función real, ambas diferenciables, y* $\mathbf{F}(\mathbf{X}) = \Phi(\mathbf{X})\mathbf{G}(\mathbf{X})$, *entonces*

$$\mathbf{dF}(\mathbf{X}) = \Phi(\mathbf{X})\mathbf{dG}(\mathbf{X}) + \mathbf{G}(\mathbf{X})d\Phi(\mathbf{X}) \quad y \quad \mathbf{DF}(\mathbf{X}) = \mathrm{Vec}(\mathbf{G}(\mathbf{X}))D\Phi(\mathbf{X}) + \Phi(\mathbf{X})\mathbf{DG}(\mathbf{X}).$$

10. *Dadas* $\mathbf{G}_{m\times p}$ *y* $\mathbf{H}_{r\times s}$ *diferenciables, si* $\mathbf{F}(\mathbf{X}) = (\mathbf{G}\otimes\mathbf{H})(\mathbf{X})$*, entonces* $\mathbf{F}$ *es diferenciable. Además* $\mathbf{dF}(\mathbf{X}) = [\mathbf{dG}(\mathbf{X})\otimes\mathbf{H}(\mathbf{X})] + [\mathbf{G}(\mathbf{X})\otimes\mathbf{dH}(\mathbf{X})]$ *y*

$$\mathbf{DF}(\mathbf{X}) = [\mathbf{I}_p\otimes(\mathbf{K}_{sm}\otimes\mathbf{I}_r)(\mathbf{I}_m\otimes\mathrm{Vec}(\mathbf{H}(\mathbf{X}))]\mathbf{DG}(\mathbf{X})+[(\mathbf{I}_p\otimes\mathbf{K}_{sm})(\mathrm{Vec}(\mathbf{G}(\mathbf{X}))\otimes\mathbf{I}_s)\otimes\mathbf{I}_r]\mathbf{DH}(\mathbf{X}).$$

Nota 2.2.5. *A la hora de encontrar la matriz jacobiana lo usual no será encontrar todas las derivadas sino su diferencial. En el caso de una función vectorial* $\mathbf{f}$*, el primer teorema de identificación nos proporciona una correspondencia biunívoca entre diferenciales y matrices jacobianas. Más concretamente, la expresión* $\mathbf{df}(\mathbf{x}) = \mathbf{A}(\mathbf{x})\mathbf{dx}$ *conduce a* $\mathbf{Df}(\mathbf{x}) = \mathbf{A}(\mathbf{x})$*.*

En cuanto a las funciones matriciales, el teorema de identificación nos indica que si se verifica $\mathbf{d}\,\mathrm{Vec}(\mathbf{F}(\mathbf{X})) = \mathbf{A}(\mathbf{X})\mathbf{d}\,\mathrm{Vec}(\mathbf{X})$ [1] *entonces* $\mathbf{DF}(\mathbf{X}) = \mathbf{A}(\mathbf{X})$*. Así para calcular las matrices jacobianas para este tipo de funciones se procederá de la siguiente forma:*

1. *Calcular* $\mathbf{dF}(\mathbf{X})$*.*
2. *Vectorizar la expresión resultante y obtener* $\mathbf{d}\,\mathrm{Vec}(\mathbf{F}(\mathbf{X})) = \mathbf{A}(\mathbf{X})\mathbf{d}\,\mathrm{Vec}(\mathbf{X})$*.*
3. *Concluir que* $\mathbf{DF}(\mathbf{X}) = \mathbf{A}(\mathbf{X})$*.*

Ejemplo 2.2.2. *Sea* $\mathbf{F}_{m\times p}$ *una función diferenciable y consideremos* $\mathbf{G}(\mathbf{X}) = \mathbf{AF}(\mathbf{X})\mathbf{B}$ *con* $\mathbf{A}_{n\times m}$ *y* $\mathbf{B}_{p\times q}$ *matrices constantes. Entonces, aplicando el teorema anterior, se tiene* $\mathbf{dG}(\mathbf{X}) = \mathbf{AdF}(\mathbf{X})\mathbf{B}$*, de donde* $\mathbf{d}\,\mathrm{Vec}(\mathbf{G}(\mathbf{X})) = (\mathbf{B}^t\otimes\mathbf{A})\mathbf{d}\,\mathrm{Vec}(\mathbf{F}(\mathbf{X})) = (\mathbf{B}^t\otimes\mathbf{A})\mathbf{DF}(\mathbf{X})\mathbf{d}\,\mathrm{Vec}(\mathbf{X})$ *y por lo tanto* $\mathbf{DG}(\mathbf{X}) = (\mathbf{B}^t\otimes\mathbf{A})\mathbf{DF}(\mathbf{X})$*.*

Más generalmente, como consecuencia de las propiedades anteriores, se tiene el siguiente corolario

Corolario 2.2.1. *Sean* $\mathbf{A}_{t\times m}$*,* $\mathbf{B}_{p\times r}$*,* $\mathbf{C}_{s\times u}$ *matrices constantes y sean* $\mathbf{G}_{m\times p}$ *y* $\mathbf{H}_{r\times s}$ *dos funciones matriciales diferenciables. Si* $\mathbf{F}(\mathbf{X}) = \mathbf{AG}(\mathbf{X})\mathbf{BH}(\mathbf{X})\mathbf{C}$*, entonces*

$$\mathbf{dF}(\mathbf{X}) = \mathbf{AdG}(\mathbf{X})\mathbf{BH}(\mathbf{X})\mathbf{C} + \mathbf{AG}(\mathbf{X})\mathbf{BdH}(\mathbf{X})\mathbf{C}$$

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

[1] Aquí hemos usado que $\mathrm{Vec}(\mathbf{dF}) = \mathbf{d}\,\mathrm{Vec}(\mathbf{F})$

A continuación mostramos las matrices jacobianas asociadas a ciertas funciones matriciales que aparecen frecuentemente en la práctica y que se deduce del corolario anterior:

Corolario 2.2.2. *Sea* $\mathbf{X}_{n\times q}$.

- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}$, $\mathbf{DF}(\mathbf{X}) = \mathbf{I}_{nq}$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t$, $\mathbf{DF}(\mathbf{X}) = \mathbf{K}_{nq}$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}\mathbf{X}^t$, $\mathbf{DF}(\mathbf{X}) = 2\mathbf{N}_n(\mathbf{X}\otimes\mathbf{I}_n)$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t\mathbf{X}$, $\mathbf{DF}(\mathbf{X}) = 2\mathbf{N}_n(\mathbf{I}_q\otimes\mathbf{X}^t)$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^p$, $p\in\mathbb{N}$, $\mathbf{DF}(\mathbf{X}) = \sum_{i=1}^{p}\left[(\mathbf{X}^t)^{p-i}\otimes\mathbf{X}^{i-1}\right]$.
- *Si* $\mathbf{A}_{q\times n}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{XAX}$, $\mathbf{DF}(\mathbf{X}) = (\mathbf{X}^t\mathbf{A}^t\otimes\mathbf{I}_n) + (\mathbf{I}_q\otimes\mathbf{XA})$..
- *Si* $\mathbf{A}_{q\times n}$ *y* $\mathbf{B}_{q\times r}$ *son constantes y* $\mathbf{F}(\mathbf{X}) = \mathbf{AXB}$, $\mathbf{DF}(\mathbf{X}) = (\mathbf{B}^t\otimes\mathbf{A})$.
- *Si* $\mathbf{A}_{n\times n}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t\mathbf{AX}$, $\mathbf{DF}(\mathbf{X}) = \mathbf{K}_{qq}\left[\mathbf{I}_q\otimes\mathbf{X}^t\mathbf{A}^t\right] + \left[\mathbf{I}_q\otimes\mathbf{X}^t\mathbf{A}\right]$.
- *Si* $\mathbf{A}_{q\times q}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{XAX}^t$, $\mathbf{DF}(\mathbf{X}) = \left[\mathbf{XA}^t\otimes\mathbf{I}_n\right] + \mathbf{K}_{nn}\left[\mathbf{XA}\otimes\mathbf{I}_n\right]$.
- *Si* $\mathbf{A}_{n\times q}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t\mathbf{AX}^t$, $\mathbf{DF}(\mathbf{X}) = \left[\left[\mathbf{XA}^t\otimes\mathbf{I}_q\right] + \left[\mathbf{I}_n\otimes\mathbf{X}^t\mathbf{A}\right]\right]\mathbf{K}_{nq}$.
- *Si* $q = n$, $\mathbf{X}$ *es no singular y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^{-1}$, $\mathbf{DF}(\mathbf{X}) = -\left[(\mathbf{X}^t)^{-1}\otimes\mathbf{X}^{-1}\right]$.
- *Si* $q = n$, $\mathbf{X}$ *es no singular,* $\mathbf{A}_{p\times n}$, $\mathbf{B}_{n\times q}$ *son constantes y* $\mathbf{F}(\mathbf{X}) = \mathbf{AX}^{-1}\mathbf{B}$, $\mathbf{DF}(\mathbf{X}) = -(\mathbf{B}^t\otimes\mathbf{A})(\mathbf{X}^t\otimes\mathbf{X})^{-1}$.

2.2.3. Ejercicios

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

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

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

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

3. Si $\mathbf{X}_{n\times n}$ es una matriz simétrica y $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$ es diferenciable, demostrar que $d\,\mathrm{Vec}(\mathbf{F}(\mathbf{X})) = \mathbf{D}_n\mathbf{DF}(\mathbf{X})d\,\mathrm{Vech}(\mathbf{X})$, mientras que $d\,\mathrm{Vec}(\mathbf{F}(\mathbf{X})) = \mathbf{N}_n\mathbf{DF}(\mathbf{X})d\,\mathrm{Vec}(\mathbf{X})$. donde $\mathbf{N}_n = \frac{1}{2}[\mathbf{I}_{n^2} + \mathbf{K}_{nn}]$.

2.3. Matrices jacobianas y derivadas matriciales

Una de los principales inconvenientes a la hora de usar el cálculo diferencial matricial es que no existe una sola definición para la derivada de una función de argumento matricial. En primer lugar planteamos las definiciones clásicas para funciones reales de argumento vectorial y vectoriales, tanto de argumento real como vectorial

Definición 2.3.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$. Se define la derivada de f respecto de $\mathbf{x} \in \mathbb{R}^n$ como el vector $1 \times n$ dado por* $\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}^t} = (\dfrac{\partial f(\mathbf{x})}{\partial x_1}, \ldots, \dfrac{\partial f(\mathbf{x})}{\partial x_n})$.

Nota 2.3.1. $\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}^t} = \displaystyle\sum_{i=1}^{n} \dfrac{\partial f(\mathbf{x})}{\partial x_i}\mathbf{e}_i^t$*, siendo $\mathbf{e}_i$ el i-ésimo vector de la base canónica de $\mathbb{R}^n$.*

Nota 2.3.2. *La notación $\dfrac{f(\mathbf{x})}{\partial \mathbf{x}^t}$ hace referencia a que la derivada es un vector fila. En ocasiones puede encontrarse la definición anterior pero considerando la derivada como un vector columna. En tal caso se define $\dfrac{f(\mathbf{x})}{\partial \mathbf{x}} = (\dfrac{\partial f(\mathbf{x})}{\partial x_1}, \ldots, \dfrac{\partial f(\mathbf{x})}{\partial x_n})^t$. Es obvio que se verifica $\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}^t} = \left(\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right)^t$.*

Definición 2.3.2. *Sea $\mathbf{f} : \mathbb{R} \to \mathbb{R}^m$. Se define la derivada de $\mathbf{f}$ respecto de $x \in \mathbb{R}$ como el vector $m \times 1$ dado por* $\dfrac{\partial \mathbf{f}(x)}{\partial x} = (\dfrac{\partial f_1(x)}{\partial x}, \ldots, \dfrac{\partial f_m(x)}{\partial x})^t$.

Nota 2.3.3. $\dfrac{\partial \mathbf{f}(x)}{\partial x} = \displaystyle\sum_{j=1}^{m} \dfrac{\partial f_j(x)}{\partial x}u_j$*, siendo u_j el j-ésimo vector de la base canónica de $\mathbb{R}^m$.*

Definición 2.3.3. *Sea $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. Se define la derivada de $\mathbf{f}$ respecto de $\mathbf{x} \in \mathbb{R}^n$ como la matriz $m \times n$*

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^t} = \begin{pmatrix} \dfrac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}^t} \\ \vdots \\ \dfrac{\partial f_m(\mathbf{x})}{\partial \mathbf{x}^t} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \cdots & \dfrac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \cdots & \dfrac{\partial f_m(\mathbf{x})}{\partial x_n} \end{pmatrix}.$$

Nota 2.3.4. $\dfrac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\partial f_i(\mathbf{x})}{\partial x_j}\mathbf{J}_{ij}$, *con* $\mathbf{J}_{ij}$ *el ij-ésimo elemento de la base canónica de* $\mathbb{M}_{m\times n}$.

Definición 2.3.4. *Sea* $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{R}$. *Se define la derivada de* $\mathbf{F}$ *respecto de* $\mathbf{X} \in \mathbb{M}_{n\times q}$ *como la matriz* $n \times q$

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

Nota 2.3.5. $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{q}\frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{ij}}\mathbf{J}_{ij}$, *con* $\mathbf{J}_{ij}$ *el ij-ésimo elemento de la base canónica de* $\mathbb{M}_{n\times q}$.

Definición 2.3.5. *Sea* $\mathbf{F} : \mathbb{R} \to \mathbb{M}_{m\times p}$. *Se define la derivada de* $\mathbf{F}$ *respecto de* $x \in \mathbb{R}$ *como la matriz* $m \times p$

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

Nota 2.3.6. $\dfrac{\partial \mathbf{F}(x)}{\partial x} = \displaystyle\sum_{s=1}^{m}\sum_{t=1}^{p}\frac{\partial F_{st}(x)}{\partial x}\mathbf{E}_{st}$, *con* $\mathbf{E}_{st}$ *el st-ésimo elemento de la base canónica de* $\mathbb{M}_{m\times p}$.

Definición 2.3.6. *Sea* $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{R}^m$. *Se define la derivada de* $\mathbf{F}$ *respecto de* $\mathbf{X} \in \mathbb{M}_{n\times q}$ *como la matriz* $nm \times q$

$$\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \begin{pmatrix} \dfrac{\partial F_1(\mathbf{X})}{\partial \mathbf{X}} \\ \vdots \\ \dfrac{\partial F_m(\mathbf{X})}{\partial \mathbf{X}} \end{pmatrix} \quad \text{donde} \quad \frac{\partial F_i(\mathbf{X})}{\partial \mathbf{X}} = \begin{pmatrix} \dfrac{\partial F_i(\mathbf{X})}{\partial x_{11}} & \cdots & \cdots & \dfrac{\partial F_i(\mathbf{X})}{\partial x_{1q}} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial F_i(\mathbf{X})}{\partial x_{n1}} & \cdots & \cdots & \dfrac{\partial F_i(\mathbf{X})}{\partial x_{nq}} \end{pmatrix}.$$

Nota 2.3.7. *Observemos que*

$$\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \sum_{l=1}^{m}\left[\mathbf{u}_l \otimes \frac{\partial F_l(\mathbf{X})}{\partial \mathbf{X}}\right] = \sum_{l=1}^{m}\left[\mathbf{u}_l \otimes \sum_{i=1}^{n}\sum_{j=1}^{q}\frac{\partial F_l(\mathbf{X})}{\partial x_{ij}}\mathbf{J}_{ij}\right] = \sum_{l=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{q}\frac{\partial F_l(\mathbf{X})}{\partial x_{ij}}(\mathbf{u}_l \otimes \mathbf{J}_{ij})$$

con $\mathbf{J}_{ij}$ *el* ij*-ésimo elemento de la base canónica de* $\mathbb{M}_{n\times q}$ *y* $\mathbf{u}_l$ *el* l*-ésimo elemento de la base canónica de* $\mathbb{R}^m$*.*

Definición 2.3.7. *Sea* $\mathbf{F} : \mathbb{R}^n \to \mathbb{M}_{m\times p}$*. Se define la derivada de* $\mathbf{F}$ *respecto de* $\mathbf{x} \in \mathbb{R}^n$ *como la matriz* $m \times np$

$$\frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}^t} = \begin{pmatrix} \frac{\partial F_{11}(\mathbf{x})}{\partial \mathbf{x}^t} & \cdots & \cdots & \frac{\partial F_{1p}(\mathbf{x})}{\partial \mathbf{x}^t} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{m1}(\mathbf{x})}{\partial \mathbf{x}^t} & \cdots & \cdots & \frac{\partial F_{mp}(\mathbf{x})}{\partial \mathbf{x}^t} \end{pmatrix} \quad \textit{donde} \quad \frac{\partial F_{ij}(\mathbf{x})}{\partial \mathbf{x}^t} = \left(\frac{\partial F_{ij}(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial F_{ij}(\mathbf{x})}{\partial x_n} \right).$$

Nota 2.3.8. *Observemos que en este caso*

$$\frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}^t} = \sum_{s=1}^{m}\sum_{t=1}^{p}\left[\mathbf{E}_{st} \otimes \frac{\partial F_{st}(\mathbf{x})}{\partial \mathbf{x}^t}\right] = \sum_{s=1}^{m}\sum_{t=1}^{p}\left[\mathbf{E}_{st} \otimes \sum_{i=1}^{n} \frac{\partial F_{st}(\mathbf{x})}{\partial x_i}\mathbf{e}_i^t\right] = \sum_{s=1}^{m}\sum_{t=1}^{p}\sum_{i=1}^{n} \frac{\partial F_{st}(\mathbf{x})}{\partial x_i}(\mathbf{E}_{st}\otimes \mathbf{e}_i^t)$$

con $\mathbf{E}_{st}$ *el* st*-ésimo elemento de la base canónica de* $\mathbb{M}_{m\times p}$ *y* $\mathbf{e}_i$ *el* i*-ésimo elemento de la base canónica de* $\mathbb{R}^n$*.*

Por último tenemos la expresión general para cualquier tipo de función matricial según dos autores. La existencia de ambas definiciones se deduce del hecho de que si $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$ podemos considerar cada una de las $m \times p$ componentes y derivarlas respecto de la matriz $\mathbf{X}$ (definición de Mac Rae) o bien considerar la totalidad de la función e ir derivando respecto de cada componente de la matriz $\mathbf{X}$ (Dwyer).

Definición 2.3.8. *(Definición de derivada matricial según Mac Rae). Sea* $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$*. Se define la derivada de* $\mathbf{F}$ *respecto de* $\mathbf{X} \in \mathbb{M}_{n\times q}$ *como la matriz* $nm \times pq$

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

Nota 2.3.9. *Observemos que*

$$\begin{aligned}\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} &= \sum_{s=1}^{m}\sum_{t=1}^{p}\left[\mathbf{E}_{st}\otimes\frac{\partial \mathbf{F}_{st}(\mathbf{X})}{\partial \mathbf{X}}\right] = \sum_{s=1}^{m}\sum_{t=1}^{p}\left[\mathbf{E}_{st}\otimes\sum_{i=1}^{n}\sum_{j=1}^{q}\frac{\partial \mathbf{F}_{st}(\mathbf{X})}{\partial x_{ij}}\mathbf{J}_{ij}\right]\\ &= \sum_{s=1}^{m}\sum_{t=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{q}\frac{\partial \mathbf{F}_{st}(\mathbf{X})}{\partial x_{ij}}(\mathbf{E}_{st}\otimes\mathbf{J}_{ij})\end{aligned}$$

con $\mathbf{J}_{ij}$ *y* $\mathbf{E}_{st}$ *los elementos* ij*-ésimo y* st*-ésimo de las bases canónicas de* $\mathbb{M}_{n\times q}$ *y* $\mathbb{M}_{m\times p}$*, respectivamente.*

Definición 2.3.9. *(Definición de derivada matricial según Dwyer). Sea* $\mathbf{F}:\mathbb{M}_{n\times q}\to\mathbb{M}_{m\times p}$*. Se define la derivada de* $\mathbf{F}$ *respecto de* $\mathbf{X}\in\mathbb{M}_{n\times q}$ *como la matriz* $nm\times pq$

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

Nota 2.3.10. *Observemos que*

$$\begin{aligned}\frac{\partial \mathbf{F}(\mathbf{X})}{\underline{\underline{\partial \mathbf{X}}}} &= \sum_{i=1}^{n}\sum_{j=1}^{q}\left[\mathbf{J}_{ij}\otimes\frac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{ij}}\right] = \sum_{i=1}^{n}\sum_{j=1}^{q}\left[\mathbf{J}_{ij}\otimes\sum_{s=1}^{m}\sum_{t=1}^{p}\frac{\partial F_{st}(\mathbf{X})}{\partial x_{ij}}\mathbf{E}_{st}\right]\\ &= \sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{s=1}^{m}\sum_{t=1}^{p}\frac{\partial F_{st}(\mathbf{X})}{\partial x_{ij}}(\mathbf{J}_{ij}\otimes\mathbf{E}_{st})\end{aligned}$$

con $\mathbf{J}_{ij}$ *y* $\mathbf{E}_{st}$ *los elementos* ij*-ésimo y* st*-ésimo de las bases canónicas de* $\mathbb{M}_{n\times q}$ *y* $\mathbb{M}_{m\times p}$*, respectivamente.*

Aunque las dos definiciones anteriores conducen a dos distintas derivadas de la función matricial de argumento matricial $\mathbf{F}$, es esperable que ambas estén relacionadas en algún sentido. En efecto, se tiene el siguiente resultado

Teorema 2.3.1. *Sea* $\mathbf{F}:\mathbb{M}_{n\times q}\to\mathbb{M}_{m\times p}$*. Entonces se verifica* $\mathbf{K}_{nm}\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}}\mathbf{K}_{pq} = \frac{\partial \mathbf{F}(\mathbf{X})}{\underline{\underline{\partial \mathbf{X}}}}$.

Nota 2.3.11. *Es inmediato comprobar que las definiciones de Mac Rae y Dwyer generalizan a las definiciones anteriormente planteadas (definiciones 2.3.1 a 2.3.7). En concreto, identificando* $\mathbb{R}^s$ *con* $\mathbb{M}_{1\times s}$ *en el espacio de partida y con* $\mathbb{M}_{s\times 1}$ *en el espacio de llegada, se verifica que las* *definiciones de Mac Rae y Dwyer coinciden con las dadas en las definiciones 2.3.1 a 2.3.5 mientras que las definiciones 2.3.6 y 2.3.7 se corresponden con la de Mac Rae (y por lo tanto relacionadas con la de Dwyer vía el teorema 2.3.1).*

A partir de las definiciones anteriores hay que hacer notar algunas salvedades. En primer lugar las tres primeras no plantean ningún tipo de inconveniente pues coinciden con la definición conocida de matriz jacobiana. Sin embargo las relacionadas con las funciones matriciales pueden no conducir necesariamente a la matriz jacobiana definida con anterioridad.

Ejemplo 2.3.1. *Sea* $F : \mathbb{M}_{n\times n} \to \mathbb{R}$ *dada por* $F(\mathbf{X}) = \mathrm{tr}[\mathbf{X}]$. *A partir de la definición dada*

$$\frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial F(\mathbf{X})}{\partial x_{ij}}\mathbf{J}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} \frac{\partial x_{ll}}{\partial x_{ij}}\mathbf{J}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} \delta_{li}\delta_{lj}\mathbf{J}_{ij} = \sum_{l=1}^{n}\mathbf{J}_{ll} = \mathbf{I}_n$$

mientras que $dF(\mathbf{X}) = d\,\mathrm{tr}[\mathbf{X}] = \mathrm{tr}[\mathbf{dX}] = \mathrm{Vec}^t(\mathbf{I}_n)\,\mathrm{Vec}(\mathbf{dX}) = \mathrm{Vec}^t(\mathbf{I}_n)\mathbf{d}\,\mathrm{Vec}(\mathbf{X})$, *con lo cual* $DF(\mathbf{X}) = \mathrm{Vec}^t(\mathbf{I}_n)$ *que es un vector* $n^2 \times 1$ *(lo cual corresponde con la definición de matriz jacobiana dada con anterioridad).*

Ejemplo 2.3.2. *Sea* $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{M}_{n\times q}$ *dada por* $\mathbf{F}(\mathbf{X}) = \mathbf{X}$. *Es obvio que* $\mathbf{DF}(\mathbf{X}) = \mathbf{I}_{nq}$ *ya que* $\mathrm{Vec}(\mathbf{C}+\mathbf{U}) - \mathrm{Vec}(\mathbf{C}) = \mathrm{Vec}(\mathbf{U}) = \mathbf{I}_{nq}\,\mathrm{Vec}(\mathbf{U})$, *pero, sin embargo,*

$$\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{s=1}^{n}\sum_{t=1}^{q} \delta_{is}\delta_{jt}(\mathbf{J}_{ij}\otimes \mathbf{J}_{st}) = \sum_{i=1}^{n}\sum_{j=1}^{q}(\mathbf{J}_{ij}\otimes \mathbf{J}_{ij}) = \mathrm{Vec}(\mathbf{I}_n)\,\mathrm{Vec}^t(\mathbf{I}_q)$$

Notemos que en el primer caso el jacobiano asociado es uno mientras que en el segundo es cero puesto que $\mathrm{rg}(\mathrm{Vec}(\mathbf{I}_n)\,\mathrm{Vec}^t(\mathbf{I}_q)) = 1$.

Este hecho llevó a Magnus y Neudecker a definir la derivada matricial a partir de cómo se introdujo la diferencial de una función matricial.

Definición 2.3.10. *(Definición de derivada matricial según Magnus y Neudecker)*

Sea $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$. *Se define la derivada de* $\mathbf{F}$ *respecto de* $\mathbf{X} \in \mathbb{M}_{n\times q}$ *como la matriz* $mp \times nq$ *dada por* $\mathbf{DF}(\mathbf{X}) = \dfrac{\partial\,\mathrm{Vec}(\mathbf{F}(\mathbf{X}))}{\partial\,\mathrm{Vec}^t(\mathbf{X})}$.

Nota 2.3.12. *La notación empleada no es casual ya que esa matriz, por definición, sí coincide con la matriz jacobiana asociada a la función matricial dada. Por ejemplo dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}$ *es inmediato que* $\mathbf{DF}(\mathbf{X}) = \mathbf{I}_{nq}$ *a partir de la expresión anterior.*

Veamos a continuación algunas relaciones importantes útiles a la hora de calcular matrices jacobianas a partir de las derivadas matriciales y viceversa. Claramente las definiciones coinciden para funciones vectoriales, mientras que para el resto tenemos el siguiente resultado

Teorema 2.3.2. *Para las siguientes funciones matriciales y de variable matricial se tiene:*

1. *Dada* $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$, $\mathbf{DF}(\mathbf{X}) = \dfrac{\partial \operatorname{Vec}(\mathbf{F}(\mathbf{X}))}{\partial \operatorname{Vec}^t(\mathbf{X})} = \sum_{i=1}^{n}\sum_{j=1}^{q} \operatorname{Vec}\left(\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial x_{ij}}\right) \operatorname{Vec}^t(\mathbf{J}_{ij})$.
2. *Dada* $\mathbf{F} : \mathbb{R} \to \mathbb{M}_{m\times p}$, $\mathbf{DF}(x) = \dfrac{\partial \operatorname{Vec}(\mathbf{F}(x))}{\partial x} = \operatorname{Vec}\left(\dfrac{\partial \mathbf{F}(x)}{\partial x}\right)$.
3. *Dada* $F : \mathbb{M}_{n\times q} \to \mathbb{R}$, $DF(\mathbf{X}) = \dfrac{\partial F(\mathbf{X})}{\partial \operatorname{Vec}^t(\mathbf{X})} = \operatorname{Vec}^t\left(\dfrac{\partial F(\mathbf{X})}{\partial \mathbf{X}}\right)$.
4. *Dada* $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{R}^m$, $\mathbf{DF}(\mathbf{X}) = \dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \operatorname{Vec}^t(\mathbf{X})} = \sum_{l=1}^{m} \mathbf{u}_l \operatorname{Vec}^t\left(\dfrac{\partial F_l(\mathbf{X})}{\partial \mathbf{X}}\right) = \sum_{l=1}^{m} \mathbf{u}_l DF_l(\mathbf{X})$, *con* $\{\mathbf{u}_j : j = 1, \ldots, m\}$ *la base canónica de* $\mathbb{R}^m$.
5. *Sea* $\mathbf{F} : \mathbb{R}^n \to \mathbb{M}_{m\times p}$, *entonces*

$$\mathbf{DF}(\mathbf{x}) = \frac{\partial \operatorname{Vec}(\mathbf{F}(\mathbf{x}))}{\partial \mathbf{x}^t} = \sum_{s=1}^{m}\sum_{t=1}^{p} \operatorname{Vec}(\mathbf{E}_{st}) \frac{\partial F_{st}(\mathbf{x})}{\partial \mathbf{x}^t} = \sum_{s=1}^{m}\sum_{t=1}^{p} \operatorname{Vec}(\mathbf{E}_{st}) DF_{st}(\mathbf{x}).$$

Nota 2.3.13. *Puesto que la definición de derivada matricial de Magnus y Neudecker coincide con la matriz jacobiana introducida a partir de la diferencial matricial, las reglas de derivación no son más que las correspondientes a las matrices jacobianas. No obstante, se pueden obtener a partir de las expresiones que relacionan la matriz jacobiana con la derivada matricial (ver ejercicio 8, apartado 2.3.1).*

A continuación exponemos las principales reglas de derivación según Mac Rae, que será la que consideraremos desde ahora en adelante).

Teorema 2.3.3. *(Propiedades de la derivada matricial según Mac Rae)*

1. *Sean* $\mathbf{F}, \mathbf{G} : M_{n\times q} \to M_{m\times p}$. *Entonces* $\dfrac{\partial (\mathbf{F}+\mathbf{G})(\mathbf{X})}{\partial \mathbf{X}} = \dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} + \dfrac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}}$.
2. *Sean* $\mathbf{F} : M_{n\times q} \to M_{m\times p}$ *y* $\mathbf{G} : M_{n\times q} \to M_{p\times r}$. *Entonces*

$$\frac{\partial (\mathbf{FG})(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} (\mathbf{G}(\mathbf{X}) \otimes \mathbf{I}_q) + (\mathbf{F}(\mathbf{X}) \otimes \mathbf{I}_n) \frac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}}.$$

3. *Sea* $\mathbf{F} : M_{n\times q} \to M_{m\times m}$ *y supongamos que* $\mathbf{F}(\mathbf{X})$ *es no singular* $\forall \mathbf{X} \in M_{n\times q}$. *Entonces*

$$\frac{\partial(\mathbf{F}(\mathbf{X}))^{-1}}{\partial \mathbf{X}} = -\left[(\mathbf{F}(\mathbf{X}))^{-1} \otimes \mathbf{I}_n\right] \frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} \left[(\mathbf{F}(\mathbf{X}))^{-1} \otimes \mathbf{I}_q\right].$$

4. *Sea* $\mathbf{F} : M_{n\times q} \to M_{m\times p}$ *y* $\Phi : M_{n\times q} \to \mathbb{R}$. *Entonces*

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

5. *Sea* $\mathbf{A}_{m\times p}$ *constante y* $\mathbf{G} : M_{n\times q} \to M_{r\times s}$. *Entonces* $\dfrac{\partial(\mathbf{A} \otimes \mathbf{G}(\mathbf{X}))}{\partial \mathbf{X}} = \mathbf{A} \otimes \dfrac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}}$.

6. *Sean* $\mathbf{F} : M_{n\times q} \to M_{m\times p}$ *y* $\mathbf{G} : M_{n\times q} \to M_{r\times s}$. *Entonces*

$$\frac{\partial(\mathbf{F}(\mathbf{X}) \otimes \mathbf{G}(\mathbf{X}))}{\partial \mathbf{X}} = (\mathbf{K}_{mr} \otimes \mathbf{I}_n)\left(\mathbf{G}(\mathbf{X}) \otimes \frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}}\right)(\mathbf{K}_{sp} \otimes \mathbf{I}_q) + \mathbf{F}(\mathbf{X}) \otimes \frac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}}.$$

7. *Sean* $\mathbf{F} : M_{n\times q} \to M_{m\times p}$ *y* $\Phi : M_{m\times p} \to \mathbb{R}$. *Entonces, notando* $\mathbf{Y} = \mathbf{F}(\mathbf{X})$, *se verifica*

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

8. *Sean* $\mathbf{F} : M_{n\times q} \to M_{m\times p}$ *y* $\mathbf{G} : M_{n\times q} \to M_{ml\times pr}$. *Entonces*

$$\frac{\partial(\mathbf{F}(\mathbf{X}) * \mathbf{G}(\mathbf{X}))}{\partial \mathbf{X}} = \mathbf{F}(\mathbf{X}) * \frac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}} + \mathbf{G}(\mathbf{X}) * (\mathbf{I}_m \otimes \mathrm{Vec}(\mathbf{I}_l) \otimes \mathbf{I}_n)\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}}(\mathbf{I}_p \otimes \mathrm{Vec}^t(\mathbf{I}_r) \otimes \mathbf{I}_q).$$

9. *Sean* $\mathbf{F}, \mathbf{G} : M_{n\times q} \to M_{m\times p}$. *Entonces* $\dfrac{\partial(\mathbf{F}(\mathbf{X}) * \mathbf{G}(\mathbf{X}))}{\partial \mathbf{X}} = \mathbf{F}(\mathbf{X}) * \dfrac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}} + \mathbf{G}(\mathbf{X}) * \dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}}$.

El siguiente resultado es bastante útil para el cálculo de algunas derivadas:

Teorema 2.3.4. *Sea* $\mathbf{F} : M_{n\times q} \to M_{m\times p}$. *Entonces*

1. $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}^t} = \left(\dfrac{\partial(\mathbf{F}(\mathbf{X}))^t}{\partial \mathbf{X}}\right)^t$.

2. $\dfrac{\partial(\mathbf{F}(\mathbf{X}))^t}{\partial \mathbf{X}^t} = \left(\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}}\right)^t$.

El siguiente resultado engloba la derivada matricial, según Mac Rae, de múltiples funciones matriciales y es consecuencia inmediata del teorema anterior

Corolario 2.3.1. *Sean* $\mathbf{A}_{t\times m}$, $\mathbf{B}_{p\times r}$, $\mathbf{C}_{s\times u}$ *tres matrices constantes y* $\mathbf{G}_{m\times p}$, $\mathbf{H}_{r\times s}$ *dos funciones matriciales de argumento matricial* $\mathbf{X}_{n\times q}$. *Si* $\mathbf{F}(\mathbf{X}) = \mathbf{A}\mathbf{G}(\mathbf{X})\mathbf{B}\mathbf{H}(\mathbf{X})\mathbf{C}$, *entonces*

$$\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = (\mathbf{A}\otimes \mathbf{I}_n)\frac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}}(\mathbf{B}\mathbf{H}(\mathbf{X})\mathbf{C}\otimes \mathbf{I}_q) + (\mathbf{A}\mathbf{G}(\mathbf{X})\mathbf{B}\otimes \mathbf{I}_n)\frac{\partial \mathbf{H}(\mathbf{X})}{\partial \mathbf{X}}(\mathbf{C}\otimes \mathbf{I}_q).$$

y más particularmente se verifica

Corolario 2.3.2. *Sea* $\mathbf{X}_{n\times q}$.

- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \operatorname{Vec}(\mathbf{I}_n)\operatorname{Vec}^t(\mathbf{I}_q)$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{K}_{qn}$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}\mathbf{X}^t$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \operatorname{Vec}(\mathbf{I}_n)\operatorname{Vec}^t(\mathbf{X}^t) + \mathbf{K}_{nn}[\mathbf{I}_n\otimes \mathbf{X}]$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t\mathbf{X}$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = [\mathbf{I}_q\otimes \mathbf{X}]\mathbf{K}_{qq} + \operatorname{Vec}(\mathbf{X})\operatorname{Vec}^t(\mathbf{I}_q)$.
- *Dada* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^p$, $p\in\mathbb{N}$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \displaystyle\sum_{i=1}^{p}\operatorname{Vec}((\mathbf{X}^t)^{i-1})\operatorname{Vec}^t(\mathbf{X}^{p-i})$.
- *Si* $\mathbf{A}_{q\times n}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}\mathbf{A}\mathbf{X}$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \operatorname{Vec}(\mathbf{I}_n)\operatorname{Vec}^t(\mathbf{A}\mathbf{X}) + \operatorname{Vec}(\mathbf{A}^t\mathbf{X}^t)\operatorname{Vec}^t(\mathbf{I}_q)$.
- *Si* $\mathbf{A}_{q\times n}$ *y* $\mathbf{B}_{q\times r}$ *son constantes y* $\mathbf{F}(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{B}$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \operatorname{Vec}(\mathbf{A}^t)\operatorname{Vec}^t(\mathbf{B})$.
- *Si* $\mathbf{A}_{n\times n}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t\mathbf{A}\mathbf{X}$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = [\mathbf{I}_q\otimes \mathbf{A}\mathbf{X}]\mathbf{K}_{qq} + \operatorname{Vec}(\mathbf{A}^t\mathbf{X})\operatorname{Vec}^t(\mathbf{I}_q)$.
- *Si* $\mathbf{A}_{q\times q}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}\mathbf{A}\mathbf{X}^t$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \operatorname{Vec}(\mathbf{I}_n)\operatorname{Vec}^t(\mathbf{A}\mathbf{X}^t) + \mathbf{K}_{nn}[\mathbf{I}_n\otimes \mathbf{X}\mathbf{A}]$.
- *Si* $\mathbf{A}_{n\times q}$ *es constante y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^t\mathbf{A}\mathbf{X}^t$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = \big[[\mathbf{I}_q\otimes \mathbf{A}\mathbf{X}^t] + [\mathbf{X}^t\mathbf{A}\otimes \mathbf{I}_n]\big]\mathbf{K}_{qn}$.
- *Si* $q=n$, $\mathbf{X}$ *es no singular y* $\mathbf{F}(\mathbf{X}) = \mathbf{X}^{-1}$, $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = -\operatorname{Vec}((\mathbf{X}^{-1})^t)\operatorname{Vec}^t(\mathbf{X}^{-1})$.
- *Si* $q=n$, X *es no singular,* $\mathbf{A}_{p\times n}$ *y* $\mathbf{B}_{n\times q}$ *son constantes y* $\mathbf{F}(\mathbf{X}) = \mathbf{A}\mathbf{X}^{-1}\mathbf{B}$, *entonces* $\dfrac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} = -\operatorname{Vec}((\mathbf{A}\mathbf{X}^{-1})^t)\operatorname{Vec}^t(\mathbf{X}^{-1}\mathbf{B})$.

2.3.1. Ejercicios

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

   a) Sea $\mathbf{X}_{n\times n}$ y $\mathbf{F}(\mathbf{X}) = \mathrm{tr}[\mathbf{X}]$. Entonces $\mathbf{DF}(\mathbf{X}) = \mathrm{Vec}^t(\mathbf{I}_n)$.

   b) Sea ahora $\mathbf{X}_{n\times q}$ y $\mathbf{F}(\mathbf{X}) = \mathbf{X}$. Entonces $\mathbf{DF}(\mathbf{X}) = \mathbf{I}_q \otimes \mathbf{I}_n = \mathbf{I}_{nq}$

2. Sea $\mathbf{X}_{n\times q}$. Demostrar las siguientes igualdades

   a) $\dfrac{\partial \mathbf{X}^t}{\partial \mathbf{X}} = \mathbf{K}_{qn}$

   b) $\dfrac{\partial \mathbf{X}}{\partial \mathbf{X}^t} = \mathbf{K}_{nq}$.

   c) $\dfrac{\partial \mathbf{X}^t}{\partial \mathbf{X}^t} = \mathrm{Vec}(\mathbf{I}_q)\,\mathrm{Vec}^t(\mathbf{I}_n)$.

3. Demostrar que si $\mathbf{X}_{n\times n}$ es no singular entonces $\dfrac{\partial \mathbf{X}^{-1}}{\partial \mathbf{X}} = -\,\mathrm{Vec}((\mathbf{X}^{-1})^t)\,\mathrm{Vec}^t(\mathbf{X}^{-1})$.

2.3.2. Derivadas matriciales de funciones escalares de un vector

Las principales funciones de este tipo que suelen aparecer en las aplicaciones prácticas son:

- Formas lineales: $\phi(\mathbf{x}) = \mathbf{a}^t\mathbf{x}$.

- Formas cuadráticas: $\phi(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x}$.

para las cuales es inmediato comprobar que

- $d\phi(\mathbf{x}) = \mathbf{a}^t d\mathbf{x} \Rightarrow D\phi(\mathbf{x}) = \dfrac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}^t} = \mathbf{a}^t$.

- $d\phi(\mathbf{x}) = \mathbf{x}^t(\mathbf{A}+\mathbf{A}^t)d\mathbf{x} \Rightarrow D\phi(\mathbf{x}) = \dfrac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}^t} = \mathbf{x}^t(\mathbf{A}+\mathbf{A}^t)$.

Además podemos considerar algunos otros casos que aparecen recopilados en el siguiente resultado:

Teorema 2.3.5. *Sean* $\mathbf{a}$ *y* $\mathbf{A}$*, respectivamente, un vector y una matriz constantes y sean* $\mathbf{f}$ *y* $\mathbf{g}$ *dos funciones vectoriales del vector* $\mathbf{x}$*. Entonces*

- *Si* $\phi(\mathbf{x}) = \mathbf{a}^t\mathbf{f}(\mathbf{x})$*, entonces* $D\phi(\mathbf{x}) = \dfrac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}^t} = \mathbf{a}^t\mathbf{Df}(\mathbf{x})$.

- *Si* $\phi(\mathbf{x}) = \mathbf{f}(\mathbf{x})^t\mathbf{g}(\mathbf{x})$, *entonces* $D\phi(\mathbf{x}) = \dfrac{\partial\phi(\mathbf{x})}{\partial\mathbf{x}^t} = \mathbf{g}(\mathbf{x})^t\mathbf{D}\mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{x})^t\mathbf{D}\mathbf{g}(\mathbf{x})$.
- *Si* $\phi(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{f}(\mathbf{x})$, *entonces* $D\phi(\mathbf{x}) = \dfrac{\partial\phi(\mathbf{x})}{\partial\mathbf{x}^t} = \mathbf{f}(\mathbf{x})^t\mathbf{A}^t + \mathbf{x}^t\mathbf{A}\mathbf{D}\mathbf{f}(\mathbf{x})$.
- *Si* $\phi(\mathbf{x}) = \mathbf{f}(\mathbf{x})^t\mathbf{A}\mathbf{f}(\mathbf{x})$, *entonces* $D\phi(\mathbf{x}) = \dfrac{\partial\phi(\mathbf{x})}{\partial\mathbf{x}^t} = \mathbf{f}(\mathbf{x})^t(\mathbf{A}^t + \mathbf{A})\mathbf{D}\mathbf{f}(\mathbf{x})$.
- *Si* $\phi(\mathbf{x}) = \mathbf{f}(\mathbf{x})^t\mathbf{A}\mathbf{g}(\mathbf{x})$, *entonces* $D\phi(\mathbf{x}) = \dfrac{\partial\phi(\mathbf{x})}{\partial\mathbf{x}^t} = \mathbf{g}(\mathbf{x})^t\mathbf{A}^t\mathbf{D}\mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{x})^t\mathbf{A}\mathbf{D}\mathbf{g}(\mathbf{x})$.
- *Si* $\mathbf{x} = (\mathbf{x}_1^t, \mathbf{x}_2^t)^t$ *y* $\phi(\mathbf{x}) = \mathbf{x}_1^t\mathbf{A}\mathbf{x}_2$, *entonces* $D\phi(\mathbf{x}) = \dfrac{\partial\phi(x)}{\partial\mathbf{x}^t} = \mathbf{x}^t\begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^t & \mathbf{0} \end{pmatrix}$.

2.3.3. Derivadas matriciales de funciones escalares de matrices

En este apartado trataremos con dos tipos de funciones importantes como son la traza y el determinante.

Derivadas matriciales asociadas a trazas

El resultado básico sobre el que desarrollaremos los posteriores está recogido en el siguiente teorema.

Teorema 2.3.6. *Sea* $\mathbf{F} : \mathbb{M}_{n\times q} \to \mathbb{M}_{p\times p}$ *y* $\Phi(\mathbf{X}) = \mathrm{tr}[\mathbf{F}(\mathbf{X})]$. *Entonces* $d\Phi(\mathbf{X}) = \mathrm{tr}[\mathbf{dF}(\mathbf{X})]$, $D\Phi(\mathbf{X}) = \mathrm{Vec}^t(\mathbf{I}_p)\mathbf{D}\mathbf{F}(\mathbf{X})$ *y*

$$\frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \mathbf{I}_p * \frac{\partial\mathbf{F}(\mathbf{X})}{\partial\mathbf{X}} = \sum_{l=1}^{p}\frac{\partial F_{ll}(\mathbf{X})}{\partial\mathbf{X}}$$

Teorema 2.3.7. *Sean* $\mathbf{F}_{p\times m}$ *y* $\mathbf{G}_{m\times p}$ *dos funciones matriciales de* $\mathbf{X}_{n\times q}$. *Entonces se verifica*

$$\begin{aligned}
&d\,\mathrm{tr}[(\mathbf{FG})(\mathbf{X})] = \mathrm{tr}[\mathbf{F}(\mathbf{X})\mathbf{dG}(\mathbf{X}) + \mathbf{F}(\mathbf{X})\mathbf{dG}(\mathbf{X})] \\
&D\,\mathrm{tr}[(\mathbf{FG})(\mathbf{X})] = \mathrm{Vec}^t(\mathbf{G}^t(\mathbf{X}))\mathbf{DF}(\mathbf{X}) + \mathrm{Vec}^t(\mathbf{F}^t(\mathbf{X}))\mathbf{DG}(\mathbf{X}) \\
&\frac{\partial\,\mathrm{tr}[(\mathbf{FG})(\mathbf{X})]}{\partial\mathbf{X}} = \mathbf{G}^t(\mathbf{X}) * \frac{\partial\mathbf{F}(\mathbf{X})}{\partial\mathbf{X}} + \mathbf{F}^t(\mathbf{X}) * \frac{\partial\mathbf{G}(\mathbf{X})}{\partial\mathbf{X}}
\end{aligned}$$

El resultado siguiente es una consecuencia inmediata de los anteriores.

Corolario 2.3.3. *Para cada función escalar de variable matricial,* Φ, *se verifica*

- *Dada* $\mathbf{X}_{n\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}]$, $D\Phi(\mathbf{X}) = \operatorname{Vec}^t(\mathbf{I}_n)$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \mathbf{I}_n$.

- *Dada* $\mathbf{X}_{n\times q}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}^t\mathbf{X}] = \operatorname{tr}[\mathbf{X}\mathbf{X}^t]$, $D\Phi(\mathbf{X}) = 2\operatorname{Vec}^t(\mathbf{X})$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = 2\mathbf{X}$.

- *Dada* $\mathbf{X}_{n\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}^p]$, $(p\in\mathbb{N})$, $D\Phi(\mathbf{X}) = p\operatorname{Vec}^t(\mathbf{X}^{p-1'})$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = p(\mathbf{X}^t)^{p-1}$.

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{q\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{A}\mathbf{X}] = \operatorname{tr}[\mathbf{X}\mathbf{A}]$, $D\Phi(\mathbf{X}) = \operatorname{Vec}^t(\mathbf{A}^t)$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \mathbf{A}^t$.

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{q\times q}$, $\mathbf{B}_{n\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}\mathbf{A}\mathbf{X}^t\mathbf{B}]$, $D\Phi(\mathbf{X}) = \operatorname{Vec}^t(\mathbf{B}^t\mathbf{X}\mathbf{A}^t + \mathbf{B}\mathbf{X}\mathbf{A})$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \mathbf{B}^t\mathbf{X}\mathbf{A}^t + \mathbf{B}\mathbf{X}\mathbf{A}$.

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{q\times n}$, $\mathbf{B}_{q\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}\mathbf{A}\mathbf{X}\mathbf{B}]$, $D\Phi(\mathbf{X}) = \operatorname{Vec}^t(\mathbf{B}^t\mathbf{X}^t\mathbf{A}^t + \mathbf{A}^t\mathbf{X}^t\mathbf{B}^t)$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \mathbf{B}^t\mathbf{X}^t\mathbf{A}^t + \mathbf{A}^t\mathbf{X}^t\mathbf{B}^t$.

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{n\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}^t\mathbf{A}\mathbf{X}]$, $D\Phi(\mathbf{X}) = \operatorname{Vec}^t((\mathbf{A}+\mathbf{A}^t)\mathbf{X})$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = (\mathbf{A} + \mathbf{A}^t)\mathbf{X}$.

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{q\times q}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}\mathbf{A}\mathbf{X}^t]$, $D\Phi(\mathbf{X}) = \operatorname{Vec}^t(\mathbf{X}(\mathbf{A}^t+\mathbf{A}))$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \mathbf{X}(\mathbf{A}^t + \mathbf{A})$.

- *Dadas* $\mathbf{X}_{n\times n}$, $\mathbf{A}_{n\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{X}\mathbf{A}\mathbf{X}]$, $D\Phi(\mathbf{X}) = \operatorname{Vec}^t(\mathbf{X}^t\mathbf{A}^t + \mathbf{A}^t\mathbf{X}^t)$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \mathbf{X}^t\mathbf{A}^t + \mathbf{A}^t\mathbf{X}^t$.

- *Dada* $\mathbf{X}_{n\times q}$, $\mathbf{a}_{n\times 1}$ *y* $\phi(\mathbf{X}) = \mathbf{a}^t\mathbf{X}\mathbf{X}^t\mathbf{a}$, $D\Phi(\mathbf{X}) = 2\operatorname{Vec}^t(\mathbf{a}\mathbf{a}^t\mathbf{X})$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = 2\mathbf{a}\mathbf{a}^t\mathbf{X}$.

- *Dada* $\mathbf{X}_{n\times q}$, $\mathbf{a}_{q\times 1}$ *y* $\phi(\mathbf{X}) = \mathbf{a}^t\mathbf{X}^t\mathbf{X}\mathbf{a}$, $D\Phi(\mathbf{X}) = 2\operatorname{Vec}^t(\mathbf{X}\mathbf{a}\mathbf{a}^t)$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = 2\mathbf{X}\mathbf{a}\mathbf{a}^t$.

- *Dada* $\mathbf{X}_{n\times n}$ *tal que exista* $\mathbf{X}^{-1}$, $\mathbf{A}_{n\times n}$ *y* $\Phi(\mathbf{X}) = \operatorname{tr}[\mathbf{A}\mathbf{X}^{-1}]$, $D\Phi(\mathbf{X}) = -\operatorname{Vec}^t((\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1})^t)$ *y* $\dfrac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1})^t$.

Derivadas matriciales asociadas a determinantes

El determinante de una matriz es una función muy empleada en Estadística pues forma parte de la verosimilitud asociada a múltiples matrices aleatorias. Al igual que hicimos en el apartado anterior sobre la función traza, los resultados se basarán en los siguientes teoremas:

Teorema 2.3.8. *Sea una función matricial* $\mathbf{F} : S \to \mathbb{R}^{p\times p}$*,* $p \geq 2$*, donde* S *es un abierto de* $\mathbb{R}^{n\times q}$*. Si la función* $\mathbf{F}$ *es diferenciable en* S*, entonces lo es también la función* $|\mathbf{F}| : S \to \mathbb{R}$ *definida por* $|\mathbf{F}|(\mathbf{X}) = |\mathbf{F}(\mathbf{X})|$*, verificándose las relaciones* $d|\mathbf{F}|(\mathbf{X}) = \operatorname{tr}[\mathbf{F}^*(\mathbf{X})\mathbf{dF}(\mathbf{X})]$*,* $D|\mathbf{F}|(\mathbf{X}) = \operatorname{Vec}^t(\mathbf{F}^{*'}(\mathbf{X}))\mathbf{DF}(\mathbf{X})$ *y*

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

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

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

A la vista del teorema anterior se tiene

Teorema 2.3.9. *Sean* $\mathbf{F}_1 : \mathbb{M}_{n\times q} \to \mathbb{M}_{p\times r}$ *y* $\mathbf{F}_2 : \mathbb{M}_{n\times q} \to \mathbb{M}_{r\times p}$*. Entonces en los puntos* $\mathbf{X}$ *en los que la función producto* $\mathbf{F}_1\mathbf{F}_2$ *tenga determinante distinto de cero se verifica*

$$\begin{aligned} d|\mathbf{F}_1\mathbf{F}_2| &= |\mathbf{F}_1\mathbf{F}_2|\operatorname{tr}\left[\mathbf{F}_2(\mathbf{F}_1\mathbf{F}_2)^{-1}\mathbf{dF}_1 + (\mathbf{F}_1\mathbf{F}_2)^{-1}\mathbf{F}_1\mathbf{dF}_2\right]\\ D|\mathbf{F}_1\mathbf{F}_2| &= |\mathbf{F}_1\mathbf{F}_2|\left[\operatorname{Vec}^t\left((\mathbf{F}_1\mathbf{F}_2)^{-1'}\mathbf{F}_2^t\right)\mathbf{DF}_1 + \operatorname{Vec}^t\left(\mathbf{F}_1^t(\mathbf{F}_1\mathbf{F}_2)^{-1'}\right)\mathbf{DF}_2\right]\end{aligned}$$

y

$$\begin{aligned}\frac{\partial|\mathbf{F}_1\mathbf{F}_2|}{\partial\mathbf{X}} &= |\mathbf{F}_1\mathbf{F}_2|(\mathbf{F}_1\mathbf{F}_2)^{-1'} * \left(\frac{\partial\mathbf{F}_1}{\partial\mathbf{X}}[\mathbf{F}_2\otimes\mathbf{I}_q] + [\mathbf{F}_1\otimes\mathbf{I}_n]\frac{\partial\mathbf{F}_2}{\partial\mathbf{X}}\right)\\ &= |\mathbf{F}_1\mathbf{F}_2|\sum_{i=1}^{n}\sum_{j=1}^{q}\operatorname{tr}\left[\mathbf{F}_2(\mathbf{F}_1\mathbf{F}_2)^{-1}\frac{\partial\mathbf{F}_1}{\partial x_{ij}} + (\mathbf{F}_1\mathbf{F}_2)^{-1}\mathbf{F}_1\frac{\partial\mathbf{F}_2}{\partial x_{ij}}\right]\mathbf{J}_{ij}\end{aligned}$$

Siguiendo un desarrollo paralelo al de la función traza tenemos

Corolario 2.3.4. *Sean* $\mathbf{F}_{p\times m}$ *y* $\mathbf{G}_{r\times s}$ *dos funciones matriciales de* $\mathbf{X}_{n\times q}$ *y sean* $\mathbf{A}_{m\times r}$ *y* $\mathbf{B}_{s\times p}$ *dos funciones constantes. Entonces, si llamamos* $\mathbf{H}(\mathbf{X}) = \mathbf{F}(\mathbf{X})\mathbf{A}\mathbf{G}(\mathbf{X})\mathbf{B}$ *y consideramos* $\Phi(\mathbf{X}) =$

$|\mathbf{H}(\mathbf{X})|$, *se tiene, para los puntos en los que existe la inversa de* $\mathbf{H}(\mathbf{X})$*:*

$$\begin{aligned}D\Phi(\mathbf{X}) = |\mathbf{H}(\mathbf{X})| \big[&\mathrm{Vec}^t(\mathbf{G}^t(\mathbf{X}))(\mathbf{A}^t \otimes \mathbf{B}(\mathbf{H}(\mathbf{X}))^{-1})\mathbf{DF}(\mathbf{X}) \\ &+ \mathrm{Vec}^t(\mathbf{F}^t(\mathbf{X}))((\mathbf{H}(\mathbf{X}))^{-1'}\mathbf{B}^t \otimes \mathbf{A})\mathbf{DG}(\mathbf{X})\big]\end{aligned}$$

y

$$\begin{aligned}\frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} &= |\mathbf{H}(\mathbf{X})|(\mathbf{H}(\mathbf{X}))^{-1'} * \left[\frac{\partial\mathbf{F}(\mathbf{X})}{\partial\mathbf{X}}(\mathbf{AG}(\mathbf{X})\mathbf{B} \otimes \mathbf{I}_q) + (\mathbf{F}(\mathbf{X})\mathbf{A} \otimes \mathbf{I}_n)\frac{\partial\mathbf{G}(\mathbf{X})}{\partial\mathbf{X}}(\mathbf{B} \otimes \mathbf{I}_q)\right] \\ &= |\mathbf{H}(\mathbf{X})| \sum_{i=1}^{n}\sum_{j=1}^{q} \mathrm{tr}\left[\mathbf{BG}(\mathbf{X})(\mathbf{H}(\mathbf{X}))^{-1}\mathbf{A}\frac{\partial\mathbf{F}(\mathbf{X})}{\partial x_{ij}} + (\mathbf{H}(\mathbf{X}))^{-1}\mathbf{AF}(\mathbf{X})\mathbf{B}\frac{\partial\mathbf{G}(\mathbf{X})}{\partial x_{ij}}\right]\mathbf{J}_{ij}\end{aligned}$$

Otros resultados interesantes en las aplicaciones son los siguientes

Corolario 2.3.5. *Sean* $\mathbf{A}$ *y* $\mathbf{B}$ *dos matrices constantes y tales que los determinantes que a continuación se consideran sean distintos de cero.*

- *Dada* $\mathbf{X}_{n\times q}$ *y* $\Phi(\mathbf{X}) = |\mathbf{X}\mathbf{X}^t|$*,*

$$D\Phi(\mathbf{X}) = 2\Phi(\mathbf{X})\,\mathrm{Vec}^t((\mathbf{X}\mathbf{X}^t)^{-1}\mathbf{X}) \;\; y \;\; \frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = 2\Phi(\mathbf{X})(\mathbf{X}\mathbf{X}^t)^{-1}\mathbf{X}$$

- *Dada* $\mathbf{X}_{n\times q}$ *y* $\Phi(\mathbf{X}) = |\mathbf{X}^t\mathbf{X}|$*, entonces*

$$D\Phi(\mathbf{X}) = 2\Phi(\mathbf{X})\,\mathrm{Vec}^t(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}) \;\; y \;\; \frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = 2\Phi(\mathbf{X})\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}$$

- *Dada* $\mathbf{X}_{n\times n}$ *y* $\Phi(\mathbf{X}) = |\mathbf{X}^p|$ *(*$p \in \mathbb{Z}$*), entonces*

$$D\Phi(\mathbf{X}) = p\Phi(\mathbf{X})\,\mathrm{Vec}^t((\mathbf{X}^t)^{-1}) \;\; y \;\; \frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = p\Phi(\mathbf{X})(\mathbf{X}^t)^{-1}$$

- *Dadas* $\mathbf{X}_{n\times q}$*,* $\mathbf{A}_{q\times q}$*,* $\mathbf{B}_{n\times n}$ *y* $\Phi(\mathbf{X}) = |\mathbf{XAX}^t\mathbf{B}|$*, entonces*

$$D\Phi(\mathbf{X}) = \Phi(\mathbf{X})\,\mathrm{Vec}^t(\left[\mathbf{XAX}^t\mathbf{B}\right]^{-1'}\mathbf{B}^t\mathbf{XA}^t + \mathbf{B}\left[\mathbf{XAX}^t\mathbf{B}\right]^{-1}\mathbf{XA})$$

  y

$$\frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \Phi(\mathbf{X})\left(\left[\mathbf{XAX}^t\mathbf{B}\right]^{-1'}\mathbf{B}^t\mathbf{XA}^t + \mathbf{B}\left[\mathbf{XAX}^t\mathbf{B}\right]^{-1}\mathbf{XA}\right)$$

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{n\times n}$, $\mathbf{B}_{q\times q}$ *y* $\Phi(\mathbf{X}) = |\mathbf{X}^t\mathbf{A}\mathbf{X}\mathbf{B}|$, *entonces*

$$D\Phi(\mathbf{X}) = \Phi(\mathbf{X})\,\mathrm{Vec}^t(\mathbf{AXB}\left[\mathbf{X}^t\mathbf{AXB}\right]^{-1} + \mathbf{A}^t\mathbf{X}\left[\mathbf{X}^t\mathbf{AXB}\right]^{-1'}\mathbf{B}^t)$$

y

$$\frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \Phi(\mathbf{X})\left(\mathbf{AXB}\left[\mathbf{X}^t\mathbf{AXB}\right]^{-1} + \mathbf{A}^t\mathbf{X}\left[\mathbf{X}^t\mathbf{AXB}\right]^{-1'}\mathbf{B}^t\right)$$

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{q\times n}$, $\mathbf{B}_{q\times n}$ *y* $\Phi(\mathbf{X}) = |\mathbf{XAXB}|$, *entonces*

$$D\Phi(\mathbf{X}) = \Phi(\mathbf{X})\,\mathrm{Vec}^t((\mathbf{AXB}(\mathbf{XAXB})^{-1} + \mathbf{B}(\mathbf{XAXB})^{-1}\mathbf{XA})^t)$$

y

$$\frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \Phi(\mathbf{X})\left(\mathbf{AXB}(\mathbf{XAXB})^{-1} + \mathbf{B}(\mathbf{XAXB})^{-1}\mathbf{XA}\right)^t$$

- *Dadas* $\mathbf{X}_{n\times q}$, $\mathbf{A}_{q\times n}$, $\mathbf{B}_{q\times n}$ *y* $\Phi(\mathbf{X}) = |\mathbf{AXB}|$, *entonces*

$$D\Phi(\mathbf{X}) = \Phi(\mathbf{X})\,\mathrm{Vec}^t((\mathbf{B}\left(\mathbf{AXB}\right)^{-1}\mathbf{A})^t) \ \ y\ \ \frac{\partial\Phi(\mathbf{X})}{\partial\mathbf{X}} = \Phi(\mathbf{X})(\mathbf{B}\left(\mathbf{AXB}\right)^{-1}\mathbf{A})^t$$

Por último tenemos la expresión que adopta la diferencial y derivada matricial del logaritmo de un determinante, cuestión ésta que aparece en repetidas ocasiones en el ambiente estadístico, por ejemplo en algunas funciones de verosimilitud multidimensionales (normal, logarítmico-normal...).

Corolario 2.3.6. *Sea* S *un abierto de* $\mathbb{M}_{n\times q}$ *y* $\mathbf{F} : S \to \mathbb{M}_{p\times p}$ *tal que* $\mathbf{F}$ *es diferenciable sobre* S. *Entonces, para los puntos* $\mathbf{X}$ *en los que* $|\mathbf{F}(\mathbf{X})| > 0$, *la función* $\log|\mathbf{F}| : S \to \mathbb{R}$ *dada por* $(\log|\mathbf{F}|)(\mathbf{X}) = \log|\mathbf{F}(\mathbf{X})|$ *es diferenciable. Además* $d\log(|\mathbf{F}|(\mathbf{X})) = \mathrm{tr}\left[\mathbf{F}^{-1}(\mathbf{X})\mathbf{dF}(\mathbf{X})\right]$, $D\log(|\mathbf{F}|(\mathbf{X})) = \mathrm{Vec}^t(\mathbf{F}^{-1'}(\mathbf{X}))\mathbf{DF}(\mathbf{X})$ *y*

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

2.3.4. Jacobianos de funciones vectoriales

Con frecuencia uno se encuentra en Estadística con que un vector aleatorio $\mathbf{Y} = (y_1, \ldots, y_m)^t$ es una función lineal de otra colección de variables e interesa poder expresar la derivada de ese vector $\mathbf{Y}$ respecto a las otras variables. Vamos a presentar dos posibles casos que pueden ocurrir como son que el vector $\mathbf{Y}$ sea una función lineal, concretamente cada componente sea una combinación lineal de los elementos de otro vector aleatorio, siendo el otro posible caso el que las variables independientes sean los elementos de una determinada matriz aleatoria $\mathbf{X}$.

- Consideremos un conjunto de variables $y_1, \ldots, y_m$ y supongamos que estas variables son combinaciones lineales desconocidas de otro conjunto de variables $x_1, \ldots, x_n$, por lo que podemos expresar $y_i = \sum_{j=1}^{n} a_{ij} x_j \quad i = 1, \ldots, m$ o bien en forma compacta $\mathbf{y} = \mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Evidentemente, como $\mathbf{df}(\mathbf{x}) = \mathbf{A}\mathbf{dx}$ se verifica $\mathbf{Df}(\mathbf{x}) = \mathbf{A}$.

- Supongamos ahora que las variables y_i, $i = 1, \ldots, m$ están relacionadas con una matriz de variables x_{ij} en la forma $\mathbf{y} = \mathbf{f}(\mathbf{X}) = \mathbf{X}\mathbf{a}$. En este caso se tiene

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

  por lo que se desprende en este caso que $\mathbf{Df}(\mathbf{X}) = \mathbf{a}^t \otimes \mathbf{I}_n$.

2.4. Diferencial segunda y hessianos

2.4.1. Matriz hessiana para una función escalar de un vector

Definición 2.4.1. *Sea $\Phi : S \subseteq \mathbb{R}^n \to \mathbb{R}$. Consideremos c un punto interior de S donde las n^2 parciales de segundo orden $D^2_{kj}\Phi(c)$ existan. Entonces se define la matriz hessiana $H\Phi(c)$ como la matriz cuadrada de orden n siguiente*

$$H\Phi(c) = \begin{pmatrix} D^2_{11}\Phi(c) & \ldots & D^2_{1n}\Phi(c) \\ \vdots & \vdots & \vdots \\ D^2_{n1}\Phi(c) & \ldots & D^2_{nn}\Phi(c) \end{pmatrix}$$

2.4.2. Matriz hessiana para una función vectorial de un vector

Definición 2.4.2. *Sea $f : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Consideremos c un punto interior de S donde las mn^2 parciales de segundo orden $D^2_{kj} f_i(c)$ 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 2.4.3. *Consideremos una función real $\Phi : S \subseteq \mathbb{R}^n \to \mathbb{R}$ 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*

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

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

Profundizando un poco más en esta cuestión, dada una función Φ que sea dos veces diferenciable en un punto c interior de S, si definimos la función $\Psi(x) = d\Phi(x; u)$, entonces la diferencial segunda de Φ no es más que la diferencial de la función Ψ. Así pues

$$\Psi(x) = \sum_{j=1}^{n} u_j D_j \Phi(x) \quad \text{con parciales} \quad D_i\Psi(x) = \sum_{j=1}^{n} u_j D^2_{ij}\Phi(x)\ ,\ i = 1, \ldots, n$$

Con ello la primera diferencial en u de Ψ será

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

de donde se deduce

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

que no es más que una forma cuadrática. Además la matriz hessiana es única.

A continuación lo que conviene es tener alguna expresión que nos permita identificar, al hacer la diferencial segunda de una función, quien es la matriz hessiana asociada a tal función.

Teorema 2.4.1. *(Segundo teorema de identificación)*

Sea $\Phi : S \subseteq \mathbb{R}^n \to \mathbb{R}$ *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*

$$d^2\Phi(c; u) = u' B(c) u \qquad \forall u \in \mathbb{R}^n$$

entonces

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

2.4.4. Diferencial segunda de una función vectorial

Definición 2.4.4. *Consideremos una función* $f : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$. *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* $D_j f_i$ *es diferenciable en* c, *diremos que* f *es dos veces diferenciable en* c.

La diferencial segunda no es más que la diferencial de la primera diferencial, o sea $d^2 f = d(df)$. Como df es por definición una función de dos conjuntos de variables, llamémosles x y u, la expresión $d(df)$ requiere alguna explicación adicional. Consideraremos que df es una función sólo de x mientras que u se considerará constante. Por lo tanto el mismo valor de u vale tanto para la primera como para la segunda diferencial.

Definición 2.4.5. *Sea* $f : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *y dos veces diferenciable en un punto interior* c *de* S. *Sea* $\mathbb{B}(c)$ *una bola abierta en* 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 incrememto* u *y la notaremos* $d^2 f(c; u)$.

Vamos a desarrollar la expresión anterior para ver la forma exacta que adopta la diferencial segunda de la función f

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

y con ello

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

Al igual que ocurría en el caso escalar, conviene tener un teorema que nos permita identificar quien es la matriz hessiana asociada a una determinada función vectorial. Necesitamos una definición previa

Definición 2.4.6. *Sean $A_1, \ldots, A_m$ matrices cuadradas de orden n y sea la matriz $A = [A_1, \ldots, A_m]$. Definamos la matriz de orden $mn \times n$ siguiente*

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

En particular, si $B_1, \ldots, B_m$ son matrices cuadradas y tenemos la matriz

$$B = \begin{pmatrix} B_1 \\ \vdots \\ \vdots \\ B_m \end{pmatrix}$$

entonces $B' = [B_1^t, \ldots, B_m^t]$ y con ello se tiene

$$B'_v = \begin{pmatrix} B'_1 \\ \vdots \\ \vdots \\ B'_m \end{pmatrix}$$

Teorema 2.4.2. *(Segundo teorema de identificación para funciones vectoriales)*

Sea $f : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ y dos veces diferenciable en un punto c interior de S. Sea $u \in \mathbb{R}^n$. Entonces se tiene

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

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

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

entonces

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

2.4.5. Diferencial segunda de una función matricial

Para abordar esta situación lo que se va a hacer es trasladar la situación matricial a la vectorial mediante la operación Vec.

Así, dada la función $F : S \subseteq \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$ se tiene asociada una función vectorial f : $\text{Vec}(S) \to \mathbb{R}^{mp}$ definida como $f(\text{Vec}(X)) = \text{Vec}(F(X))$.

De esta manera se tiene que $DF(\mathbf{C}) = Df(\text{Vec}(C))$ lo que lleva a definir la matriz hessiana asociada a F como

$$HF(\mathbf{C}) = Hf(\text{Vec}(C))$$

Esta matriz es de orden $mnpq \times nq$ y está formada a partir de los hessianos de las mp funciones de las que está formada F, F_{st}, $s = 1, \ldots, m$; $t = 1, \ldots, p$

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

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

Si notamos por $G(X) = dF(X; U)$, entonces

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

Como la diferencial de F y f están relacionadas por la expresión

$$\mathrm{Vec}\left[dF(C; U)\right] = df\left[\mathrm{Vec}(C); \mathrm{Vec}(\mathbf{U})\right]$$

es obvio que las segundas diferenciales lo estén también, teniéndose en este caso

$$\mathrm{Vec}\left[d^2F(C; U)\right] = d^2f\left[\mathrm{Vec}(C); \mathrm{Vec}(\mathbf{U})\right]$$

Para finalizar vamos a ver el homólogo del teorema de identificación en el caso matricial

Teorema 2.4.3. *(Segundo teorema de identificación para funciones matriciales)*

Sea $F : S \subseteq \mathbb{R}^{n\times q} \to \mathbb{R}^{m\times p}$. Supongamos que la función F es dos veces diferenciable en un punto C interior de S. Entonces se verifica

$$\mathrm{Vec}\left[d^2F(C; U)\right] = \left[I_{mp} \otimes \mathrm{Vec}^t(U)\right] B(C)\,\mathrm{Vec}(\mathbf{U}) \quad \forall U \in \mathbb{R}^{n\times q} \Leftrightarrow HF(\mathbf{C}) = \frac{1}{2}\left[B(C) + B(C)_v^t\right]$$

Capítulo 3

Clase esférica y elíptica de distribuciones

3.1. Clases esférica y elíptica de distribuciones en $\mathbb{R}^p$.

Definición 3.1.1. *Dado un vector aleatorio* $\mathbf{X} = (X_1, \ldots, X_p)^t$*, se dice que se distribuye en la clase esférica de distribuciones en* $\mathbb{R}^p$ *si* $\mathbf{X}$ *y* $\mathbf{HX}$ *tienen la misma distribución,* $\forall \mathbf{H} \in O(p)$*, siendo* $O(p)$ *el grupo de matrices ortogonales de orden* p*.*

Comentario 3.1.1. *La definición anterior nos dice que* $\mathbf{X}$ *pertenece a la clase esférica si su distribución es invariante frente a transformaciones ortogonales, lo cual significará que su densidad depende de* $\mathbf{x}$ *sólo a través de* $\mathbf{x^t x}$*.*

Algunos ejemplos de densidades esféricas son las siguientes:

- Distribución uniforme en la hiperesfera de radio r:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\Gamma\left(\frac{p+2}{2}\right)}{\pi^{p/2}\, r^p}\, I_{[\mathbf{x}^t\mathbf{x} \leq r^2]}$$

  En este caso, g es la función indicadora en el círculo $\mathbf{x}^t\mathbf{x} \leq r^2$ y $K_p = \dfrac{\Gamma\left(\frac{p+2}{2}\right)}{\pi^{\frac{p}{2}}\, r^p}$.

- Normal p-dimensional esférica $\mathrm{N}[0; \mathbf{I}_p]$:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\left(-\frac{\mathbf{x}^t\mathbf{x}}{2}\right), \quad \mathbf{x} \in \mathbb{R}^p.$$

  En este caso la función g es la exponencial y $K_p = \dfrac{1}{(2\pi)^{\frac{p}{2}}}$.

- T-Student esférica con n grados de libertad

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma\left(\frac{n+p}{2}\right)}{(n\pi)^{\frac{p}{2}}\,\Gamma\left(\frac{n}{2}\right)\left(1+\dfrac{\mathbf{x}^t\mathbf{x}}{n}\right)^{\frac{n+p}{2}}}, \quad \mathbf{x} \in \mathbb{R}^p.$$

Ahora $g(x) = (1 + x/n)^{(n+p)/2}$ y $K_p = \dfrac{\Gamma\left(\frac{n+p}{2}\right)}{(n\pi)^{\frac{p}{2}}\,\Gamma\left(\frac{n}{2}\right)}$.

Nota: Si $n = 1$ se tiene la distribución de Cauchy p-dimensional, mientras que si $p = 1$ se tiene la distribución t de Student con n grados de libertad.

El siguiente resultado proporciona la forma que adopta la función característica de cualquier variable que se distribuya en la clase esférica. Para ello es necesario utilizar algunos aspectos de la teoría de invarianza (ver apéndice A).

Teorema 3.1.1. *Sea* $\mathbf{X}$ *un vector aleatorio con distribución en la clase esférica. Entonces su función característica es de la forma* $\phi(\mathbf{t}) = \psi(\mathbf{t}^t\mathbf{t})$*, con* ψ *una cierta función. Además,* $\mathrm{E}[\mathbf{X}] = \mathbf{0}$ *y* $\mathrm{Cov}[\mathbf{X}] = -2\psi'(0)\mathbf{I}_p$.

Ejercicio: Demostrar el teorema anterior.

Definición 3.1.2. *Dado un vector aleatorio* $\mathbf{X} = (X_1, \ldots, X_p)^t$*, se dice que se distribuye en la clase elíptica de parámetros* $\mu \in \mathbb{R}^p$ *y* $\mathbf{V}_{p\times p}$ *(*$\mathbf{V} > 0$*) si su densidad es de la forma*

$$f(\mathbf{x}) = C_p|\mathbf{V}|^{-\frac{1}{2}} h\left((\mathbf{x}-\mu)^t\mathbf{V}^{-1}(\mathbf{x}-\mu)\right)$$

donde C_p *es una constante y* h *una función suficientemente regular. A la clase elíptica la notaremos por* $E_p(\mu;\mathbf{V})$.

Algunos ejemplos de densidades elípticas son los siguientes:

- Distribución uniforme en el elipsoide de dimensión p:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma\left(\frac{p+2}{2}\right)|\mathbf{A}|^{\frac{1}{2}}}{\pi^{\frac{p}{2}}\, r^p}\, I_{[(\mathbf{x}-\mu)^t\mathbf{A}(\mathbf{x}-\mu)\le r^2]}.$$

En este caso, $\mathbf{V} = \mathbf{A}^{-1}$, h es la función indicadora en el elipsoide $(\mathbf{x}-\mu)^t\mathbf{A}(\mathbf{x}-\mu) \leq r^2$ y

$$C_p = \frac{\Gamma\left(\frac{p+2}{2}\right)}{\pi^{\frac{p}{2}}\, r^p}.$$

- Normal p-dimensional $\mathrm{N}[\mu;\Sigma]$:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^t\Sigma^{-1}(\mathbf{x}-\mu)\right), \quad \mathbf{x} \in \mathbb{R}^p.$$

En este caso, $\mathbf{V} = \mathbf{\Sigma}$, la función h es la exponencial y

$$C_p = \frac{1}{(2\pi)^{\frac{p}{2}}}.$$

- T-Student elíptica con n grados de libertad

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma\left(\frac{n+p}{2}\right)}{(n\pi)^{\frac{p}{2}}\,\Gamma\left(\frac{n}{2}\right)|\Sigma|^{\frac{1}{2}}\left(1+\frac{(\mathbf{x}-\mu)^t\Sigma^{-1}(\mathbf{x}-\mu)}{n}\right)^{\frac{n+p}{2}}}, \quad \mathbf{x} \in \mathbb{R}^p.$$

Ahora $\mathbf{V} = \mathbf{\Sigma}$, $h(x) = (1+x/n)^{(n+p)/2}$ y

$$C_p = \frac{\Gamma\left(\frac{n+p}{2}\right)}{(n\pi)^{\frac{p}{2}}\,\Gamma\left(\frac{n}{2}\right)}.$$

Nota: Si $n = 1$ se tiene la distribución de Cauchy p-dimensional.

Ejercicio:

- Comprobar que la clase esférica coincide con la elíptica $E_p(\mathbf{0};\mathbf{I}_p)$.

- Verificar la siguiente caracterización, que generaliza la vista en el caso normal:

$$\mathbf{X} \in E_p(\mu;\mathbf{V}) \Longleftrightarrow \mathbf{X} = \mu + \mathbf{CU}$$

donde $\mu \in \mathbb{R}^p$, $\mathbf{C}$ es la matriz dada por la descomposición de Cholesky de $\mathbf{V}$, o sea, $\mathbf{V} = \mathbf{CC}^t$ y $\mathbf{U}$ es un vector de la clase esférica p-dimensional.

- Verificar que si $\mathbf{X} \in E_p(\mu; \mathbf{V})$, entonces $\mathrm{E}[\mathbf{X}] = \mu + \mathbf{C}\,\mathrm{E}[\mathbf{U}]$ y $\mathrm{Cov}[\mathbf{X}] = \mathbf{C}\,\mathrm{Cov}[\mathbf{U}]\mathbf{C}^t$, siendo $\mathbf{U}$ perteneciente a la clase esférica.

Con respecto a la función característica, se plantea el siguiente ejercicio en el cual se pide calcular su forma genérica así como ciertas propiedades inmediatas obtenidas gracias a ella, y que generalizan las obtenidas para la distribución normal.

Ejercicio:

- Sea $\mathbf{X}$ un vector aleatorio con distribución en la clase elíptica $E_p(\mu; \mathbf{V})$. Entonces su función característica es de la forma $\phi(\mathbf{u}) = e^{i\mathbf{u}^t\mu}\psi(\mathbf{u}^t\mathbf{V}\mathbf{u})$, con ψ una cierta función.

- Sea $\mathbf{X}$ un vector aleatorio con distribución en la clase elíptica $E_p(\mu; \mathbf{V})$. Comprobar que si $\mathbf{A}_{q\times p}$ es una matriz de constantes con $\mathrm{rg}(\mathbf{A}) = q \leq p$, y $\mathbf{c} \in \mathbb{R}^p$, entonces $\mathbf{Y} = \mathbf{c} + \mathbf{A}\mathbf{X} \in E_q(\mathbf{c} + \mathbf{A}\mu; \mathbf{A}\mathbf{V}\mathbf{A}^{\mathbf{t}})$.

- Sea $\mathbf{X} \in E_p(\mu; \mathbf{V})$. Entonces $\mathrm{E}[\mathbf{X}] = \mu$ y $\mathrm{Cov}[\mathbf{X}] = -2\psi'(0)\mathbf{V}$.

- Comprobar que todas las distribuciones de la clase elíptica $E_p(\mu; \mathbf{V})$ tienen igual matriz de correlaciones.

Ejercicio: Sea $\mathbf{X} \in E_p(\mu; \mathbf{V})$. Particionemos el vector $\mathbf{X}$ en la forma $\mathbf{X} = (\mathbf{X}^t_{(1)}|\mathbf{X}^t_{(2)})^t$ donde $\mathbf{X}_{(1)}$ es de dimensión $q \times 1$ y $\mathbf{X}_{(2)}$ lo es $(p-q) \times 1$. Consideremos en μ y $\mathbf{V}$ las particiones inducidas

$$\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} \quad ; \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$$

Entonces se verifica

- $\mathbf{X}_{(1)} \in E_q(\mu_{(1)}; \mathbf{V}_{11})$

- $\mathbf{X}_{(2)} \in E_{(p-q)}(\mu_{(2)}; \mathbf{V}_{22})$

Nota: Hacerlo usando la función característica y también mediante la caracterización obtenida en el primer ejercicio.

3.2. Algunas distribuciones de las clases esférica y elíptica de distribuciones en $\mathbb{R}^p$.

Aquí incluimos algunas distribuciones de la clase esférica de distribuciones en $\mathbb{R}^p$. Todos los desarrollos se dejan como ejercicio.

3.2.1. Distribución uniforme en el círculo de radio r

Sea $S^1 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq r^2\}$ el círculo centrado en el origen y de radio $r > 0$. Dado un vector aleatorio $\mathbf{X} = (X_1, X_2)^t$ se dice que sigue la distribución uniforme en S^1 si su densidad es

$$f(x_1, x_2) = K\, I_{S^1},$$

donde $K > 0$ y I_{S^1} indica la función indicadora en S^1.

Nota: Observemos que la densidad anterior depende de $\mathbf{x} = (x_1, x_2)^t$ a través del producto escalar $\mathbf{x}^t\mathbf{x}$.

Ejercicio:

- Verificar que $K = \dfrac{1}{\pi\, r^2}$. **Indicación:** Emplear el cambio a coordenadas polares:

$$\begin{array}{l} x_1 = \rho\cos(\theta) \\ x_2 = \rho\,\mathrm{sen}(\theta) \end{array}, \qquad \rho > 0, \quad 0 < \theta \leq 2\pi.$$

- Comprobar que $\mathrm{E}[\mathbf{X}] = \mathbf{0}$ y $\mathrm{Cov}[\mathbf{X}] = \dfrac{r^2}{4}\mathbf{I}_2$. **Indicación:** Emplear el cambio a coordenadas polares.

- Calcular las distribuciones marginales así como las condicionadas. Por simetría basta calcular la distribución de X_1 y la de $X_1|X_2 = x_2$.

Solución:

$$f_{X_1}(x_1) = \frac{2\sqrt{r^2 - x_1^2}}{\pi\, r^2}, \quad -r \le x_1 \le r$$

$$f_{X_1|X_2=x_2}(x_1) = \frac{1}{2\sqrt{r^2 - x_2^2}}, \quad x_1^2 \le r^2 - x_2^2; \quad\quad (-r \le x_2 \le r)\,.$$

Nota: Comprobar que las distribuciones condicionadas son uniformes en un determinado intervalo. Asimismo, tanto las densidades marginales comos la condicionadas verifican que dependen de la norma euclídea al cuadrado del argumento, o sea pertenecen a la clase esférica.

3.2.2. Distribución uniforme en la esfera de radio r

Sea $S^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 \le r^2\}$ el interior y el borde de la esfera centrada en el origen y de radio $r > 0$. Dado un vector aleatorio $\mathbf{X} = (X_1, X_2, X_3)^t$ se dice que sigue la distribución uniforme en S^2 si su densidad es

$$f(x_1, x_2, x_3) = K\, I_{S^2},$$

donde $K > 0$ y I_{S^2} indica la función indicadora en S^2.

Nota: Observemos que la densidad anterior depende de $\mathbf{x} = (x_1, x_2, x_3)^t$ a través del producto escalar $\mathbf{x}^t\mathbf{x}$.

Ejercicio:

- Verificar que $K = \dfrac{3}{4\pi\, r^3}\cdot$ **Indicación:** Emplear el cambio a coordenadas polares:

$$\begin{aligned} x_1 &= \rho\cos(\theta) \\ x_2 &= \rho\,\mathrm{sen}(\theta)\cos(\phi)\ , \qquad\qquad \rho > 0, \quad 0 < \theta \le \pi, \quad 0 < \phi \le 2\pi \\ x_3 &= \rho\,\mathrm{sen}(\theta)\,\mathrm{sen}(\phi) \end{aligned}$$

- Comprobar que $\mathrm{E}[\mathbf{X}] = \mathbf{0}$ y $\mathrm{Cov}[\mathbf{X}] = \dfrac{r^2}{5}\mathbf{I}_3$. **Indicación:** Emplear el cambio a coordenadas polares.

- Calcular la distribuciones marginales unidimensionales y bidimensionales. Dada la simetría, basta con calcular la distribución de X_1 y la de $(X_1, X_2)^t$.

  Solución:

$$f_{X_1}(x_1) = \frac{3(r^2 - x_1^2)}{4\,r^3}, \quad -r \leq x_1 \leq r$$

$$f_{X_1,X_2}(x_1, x_2) = \frac{3\sqrt{r^2 - x_1^2 - x_2^2}}{2\pi\,r^3}, \quad x_1^2 + x_2^2 \leq r^2$$

- Calcular las distribuciones $(X_2, X_3)^t | X_1 = x_1$ y $X_3 | X_1 = x_1, X_2 = x_2$. Deducir que son distribuciones uniformes en un círculo y en un intervalo, respectivamente y, en consecuencia, calcular los dos primeros momentos.

  Solución:

$$f_{X_2,X_3|X_1=x_1}(x_2, x_3) = \frac{1}{\pi(r^2 - x_1^2)}, \quad x_2^2 + x_3^2 \leq r^2 - x_1^2; \qquad (-r \leq x_1 \leq r)$$

$$f_{X_3|X_1=x_1,X_2=x_2}(x_3) = \frac{1}{2\sqrt{r^2 - x_1^2 - x_2^2}}, \quad x_3^2 \leq r^2 - x_1^2 - x_2^2; \qquad \left(x_1^2 + x_2^2 \leq r^2\right).$$

  Nota: Obsérvese que tanto las densidades marginales comos la condicionadas verifican que dependen de la norma euclídea al cuadrado del argumento, o sea, pertenecen a la clase esférica.

3.2.3. Distribución uniforme en la hiperesfera de radio r

Los anteriores resultados se pueden generalizar a mayores dimensiones. Así tenemos la distribución uniforme en la hiperesfera de radio r.

Sea $S^{p-1} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{x}^t\mathbf{x} \leq r^2\}$ el interior y el borde de la hiperesfera centrada en el origen y de radio $r > 0$. Dado un vector aleatorio $\mathbf{X} = (X_1, \ldots, X_p)^t$ se dice que sigue la distribución uniforme en S^{p-1} si su densidad es

$$f(\mathbf{x}) = K\,I_{S^{p-1}},$$

donde $K > 0$ y $I_{S^{p-1}}$ indica la función indicadora en S^{p-1}.

NOTA: Observemos que la densidad anterior depende de $\mathbf{x}$ a través del producto escalar $\mathbf{x}^t\mathbf{x}$.

Ejercicio:

- Verificar que $K = \dfrac{\Gamma\left(\frac{p+2}{2}\right)}{\pi^{p/2}\, r^p}$. **Indicación:** Emplear el cambio a coordenadas polares:

$$\begin{aligned}
x_1 &= \rho\cos(\theta_1)\\
x_2 &= \rho\,\mathrm{sen}(\theta_1)\cos(\theta_2)\\
&\vdots\\
x_{p-1} &= \rho\mathrm{sen}(\theta_1)\mathrm{sen}(\theta_2)\cdots\mathrm{sen}(\theta_{p-2})\cos(\theta_{p-1})\\
x_p &= \rho\mathrm{sen}(\theta_1)\mathrm{sen}(\theta_2)\cdots\mathrm{sen}(\theta_{p-2})\mathrm{sen}(\theta_{p-1})
\end{aligned}$$

  para $\rho > 0$, $0 < \theta_j \leq \pi\,(j = 1,\ldots,p-2)$ y $0 < \theta_{p-1} \leq 2\pi$, y cuyo jacobiano viene dado por $|J| = \rho^{p-1}\displaystyle\prod_{j=1}^{p-2}(\mathrm{sen}(\theta_j))^{p-j-1}$ (**No hay que calcularlo**.)

- Comprobar que $\mathrm{E}[\mathbf{X}] = \mathbf{0}$ y $\mathrm{Cov}[\mathbf{X}] = \dfrac{r^2}{p+2}\mathbf{I}_p$. **Indicación:** usar el cambio a coordenadas polares (prestando atención en cada caso a las variables que están involucradas en las integrales que se deban hacer). Además, es aconsejable recordar que
  - $\displaystyle\int_{\pi/2}^{\pi}\mathrm{sen}^k(\theta)\,d\theta = \int_0^{\pi/2}\cos^k(\theta)\,d\theta.$
  - $\displaystyle\int_0^{\pi/2}\mathrm{sen}^k(\theta)\,d\theta = \int_0^{\pi/2}\cos^k(\theta)\,d\theta = \frac{1}{2}Beta\left(\frac{k+1}{2},\frac{1}{2}\right) = \frac{1}{2}\frac{\Gamma\left(\frac{k+1}{2}\right)\sqrt{\pi}}{\Gamma\left(\frac{k+2}{2}\right)}.$
  - $\displaystyle\int_0^{\pi}\mathrm{sen}^k(\theta)\,d\theta = \frac{\Gamma\left(\frac{k+1}{2}\right)\sqrt{\pi}}{\Gamma\left(\frac{k+2}{2}\right)}.$
  - $\displaystyle\int_0^{\pi}\mathrm{sen}^k(\theta)\cos(\theta)\,d\theta = 0.$
  - $\displaystyle\int_0^{\pi}\mathrm{sen}^k(\theta)\cos^2(\theta)\,d\theta = \frac{1}{k+1}\frac{\Gamma\left(\frac{k+3}{2}\right)\sqrt{\pi}}{\Gamma\left(\frac{k+4}{2}\right)}.$

  (**No hay que calcular estas integrales**.)

- Calcular la distribuciones marginales unidimensionales y bidimensionales. Dada la simetría basta con calcular, por ejemplo, las de X_1 y $(X_1, X_2)^t$.

Solución:

$$f_{X_1}(x_1) = \frac{\Gamma\left(\frac{p+2}{2}\right)}{\sqrt{\pi}\, r^p\, \Gamma\left(\frac{p+1}{2}\right)}(r^2 - x_1^2)^{\frac{p-1}{2}}, \quad -r \leq x_1 \leq r$$

$$f_{X_1,X_2}(x_1,x_2) = \frac{\Gamma\left(\frac{p+2}{2}\right)}{\pi\, r^p\, \Gamma\left(\frac{p}{2}\right)}(r^2 - x_1^2 - x_2^2)^{\frac{p-2}{2}}, \quad x_1^2 + x_2^2 \leq r^2$$

- Si consideramos $\mathbf{X}$ partido en la forma $\mathbf{X} = \left(\mathbf{X}_{(1)}^t | \mathbf{X}_{(2)}^t\right)^t$, donde $\mathbf{X}_{(1)}$ es de dimensión q y $\mathbf{X}_{(1)}$ lo es $p-q$, calcular la distribución de $\mathbf{X}_{(1)}$.

Solución:

$$f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)}) = \frac{\Gamma\left(\frac{p+2}{2}\right)}{\pi^{\frac{q}{2}}\, r^p\, \Gamma\left(\frac{p-q+2}{2}\right)}\left(r^2 - \mathbf{x}_{(1)}^t \mathbf{x}_{(1)}\right)^{\frac{p-q}{2}}, \qquad \mathbf{x}_{(2)}^t \mathbf{x}_{(2)} \leq r^2$$

Indicación: Para el cálculo observa la regla de generación de las anteriores distribuciones.

- Calcular la distribución condicionada $\mathbf{X}_{(2)} | \mathbf{X}_{(1)} = \mathbf{x}_{(1)}$. Deducir que es una distribución uniforme en S^{p-q-1}, o sea, la esfera de dimensión $p-q$. En consecuencia, calcular los dos primeros momentos.

Solución:

$$f_{\mathbf{X}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1)}}(\mathbf{x}_{(\mathbf{2})}) = \frac{\Gamma\left(\frac{p-q+2}{2}\right)}{\pi^{\frac{p-q}{2}}\left(r^2 - \mathbf{x}_{(1)}^t \mathbf{x}_{(1)}\right)^{\frac{p-q}{2}}}, \quad \mathbf{x}_{(2)}^t \mathbf{x}_{(2)} \leq r^2 - \mathbf{x}_{(1)}^t \mathbf{x}_{(1)}; \quad \left(\mathbf{x}_{(1)}^t \mathbf{x}_{(1)} \leq r^2\right).$$

Nota: Obsérvese que tanto las densidades marginales comos la condicionadas verifican que dependen de la norma euclídea al cuadrado del argumento, o sea, pertenecen a la clase esférica.

3.2.4. Distribución T-Student esférica

Sea $\mathbf{X} = (X_1, \ldots, X_p)'$ un vector aleatorio. Diremos que se distribuye según la distribución t de student esférica p-dimensional con n grados de libertad si su función de densidad es

$$f(\mathbf{X}) = \frac{\Gamma(\frac{n+p}{2})}{\Gamma(\frac{n}{2})(n\pi)^{\frac{p}{2}}} \frac{1}{[1 + \frac{1}{n}\mathbf{x^t x}]^{\frac{n+p}{2}}}, \quad \mathbf{x} \in \mathbb{R}^p.$$

Demostrar que los momentos de esta distribución son:

$$E[\mathbf{X}] = 0 \qquad \text{y} \qquad Cov[\mathbf{X}] = \frac{n}{n-2}\mathbf{I}_p, \qquad n > 2.$$

Indicación: Usar argumentos similares al caso de la hiperesfera unidad, y tener en cuenta además que

- $\displaystyle\int_0^\pi \cos^2\theta_j [\operatorname{sen}\theta_j]^{p-j-1} d\theta_j = \frac{1}{p-j} \; \frac{\sqrt{\pi}\Gamma(\frac{p-j+2}{2})}{\Gamma(\frac{p-j+3}{2})}$
- $\displaystyle\int_0^\pi [\operatorname{sen}\theta_e]^{p-e+1} d\theta_e = \frac{\sqrt{\pi}\Gamma(\frac{p-e+2}{2})}{\Gamma(\frac{p-e+3}{2})}$
- $\int_0^\pi [\operatorname{sen}\theta_e]^{p-e-1} d\theta_e = \frac{\sqrt{\pi}\Gamma(\frac{p-e}{2})}{\Gamma(\frac{p-e+1}{2})}$

(**No hay que calcular estas integrales**.)

3.2.5. Versiones elípticas de las densidades uniformes en la hiperesfera y t-student

- Sea $\mathbf{U}$ un vector p-dimensional distribuido de forma uniforme en el interior y borde de la hiperesfera de dimensión p y radio $r > 0$. Consideremos μ un vector de $\in \mathbb{R}^p$ y $\mathbf{V}_{p\times p}$ una matriz definida positiva descompuesta en la forma $\mathbf{V} = \mathbf{CC}^t$. Calcular la densidad del vector aleatorio $\mathbf{X} = \mu + \mathbf{CU}$.

  Solución:

  $$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma\left(\frac{p+2}{2}\right)|\mathbf{A}|^{\frac{1}{2}}}{\pi^{\frac{p}{2}}\, r^p} I_{E_r^p}$$

  donde E_r^p es el elipsoide dado por $E_r^p = \left\{\mathbf{x} \in \mathbb{R}^p : \, (\mathbf{x}-\mu)^t\mathbf{V}^{-1}(\mathbf{x}-\mu) \leq r^2\right\}$.

  Nota: La distribución obtenida es la denominada distribución uniforme en el elipsoide E_r^p.

- Calcular $E[\mathbf{X}]$ y Cov$[\mathbf{X}]$ para la distribución uniforme en el elipsoide E_r^p.

- Sea $\mathbf{U}$ un vector p-dimensional distribuido según una t de Student multivariante esférica. Consideremos μ un vector de $\in \mathbb{R}^p$ y $\mathbf{V}_{p\times p}$ una matriz definida positiva descompuesta en la forma $\mathbf{V} = \mathbf{CC}^t$. Calcular la densidad del vector aleatorio $\mathbf{X} = \mu + \mathbf{CU}$.

  Solución: Debe resultar la densidad t de Student elíptica dada en los ejemplos de la introducción.

- Calcular $E[\mathbf{X}]$ y Cov$[\mathbf{X}]$ para la distribución anterior.

Capítulo A

Notas sobre invarianza

Sea G un grupo de transformaciones de un espacio $\mathcal{X}$ en sí mismo. Sean g_1 y g_2 dos elementos de G y definamos la operación $(g_1g_2)x = g_1(g_2)(x)$. Evidentemente, al ser G un grupo, se verificará

- Si g_1 y $g_2 \in G$ entonces $g_1g_2 \in G$.
- Si $g \in G$ entonces $g^{-1} \in G$, donde g^{-1} verifica $gg^{-1} = g^{-1}g = e$ con e la transformación identidad en $\mathcal{X}$.

Definición A.0.1. *Dos puntos x_1 y x_2 de $\mathcal{X}$ se dicen equivalentes bajo G, y escribiremos $x_1 \sim x_2$ (mod G), si existe $g \in G$ tal que $x_2 = gx_1$. Como tal relación de equivalencia verificará*

- *$x \sim x$ (mod G)*
- *$x \sim y$ (mod G) $\Rightarrow y \sim x$ (mod G).*
- *$x \sim y$ (mod G), $y \sim z$ (mod G) $\Rightarrow x \sim z$ (mod G).*

Las clases de equivalencia formadas a partir de la relación anterior se llaman las órbitas de $\mathcal{X}$ bajo G y serán de la forma $\mathrm{Orb}(x) = \{gx : \ g \in G\}$. Dos órbitas o bien son la misma o bien son disjuntas, por lo que el conjunto de las órbitas constituyen una partición del espacio $\mathcal{X}$.

Definición A.0.2. *Se dice que una función ϕ sobre el espacio $\mathcal{X}$ es invariante bajo G si verifica $\phi(gx) = \phi(x)$, $\forall x \in \mathcal{X}$, $\forall g \in G$.*

Notemos que esta definición lo que está diciendo es que ϕ es invariante bajo G sí y sólo sí es constante en cada órbita formada por la acción de G.

Definición A.0.3. *Una función ϕ sobre $\mathcal{X}$ se dice invariante maximal bajo G si es invariante bajo G y además verifica $\phi(x_1) = \phi(x_2) \Rightarrow x_1 \sim x_2$(mod G).*

O sea, ϕ es invariante maximal sí y sólo sí es constante sobre cada órbita y además da un valor distinto a cada órbita.

El siguiente resultado es muy importante puesto que establece que toda función invariante es función de un invariante maximal.

Teorema A.0.1. *Sea ϕ un invariante maximal bajo G para $\mathcal{X}$. Entonces una función ψ sobre $\mathcal{X}$ es invariante bajo G sí y sólo sí es función de ϕ.*

Demostración

Supongamos en primer lugar que ψ es función de ϕ. En tal caso debe existir una función f tal que $\psi(x) = f(\phi(x))$, $\forall x \in \mathcal{X}$. Entonces, $\forall g \in G$ y $\forall x \in \mathcal{X}$ se tiene $\psi(gx) = f(\phi(gx)) = f(\phi(x)) = \psi(x)$ y por lo tanto ψ es invariante. Notemos que no hace falta para esta implicación que ϕ sea maximal, lo que hemos probado realmente es que toda función de un invariante es también invariante.

Recíprocamente, supongamos que ψ es invariante. Veremos que los valores que ψ da a los puntos de x dependen de los valores que ϕ otorga. En efecto, supongamos que $\phi(x_1) = \phi(x_2)$, lo cual significa que $x_1 \sim x_2 \pmod G$ ya que ϕ es invariante maximal. Entonces debe existir $g \in G$ tal que $x_2 = gx_1$. Así se tiene $\psi(x_2) = \psi(gx_1) = \psi(x_1)$ ya que ψ es invariante. Por lo tanto los valores que ψ da a los puntos de $\mathcal{X}$ depende de los que otorga ϕ, con lo cual $\psi(x)$ depende de x sólo a través de $\phi(x)$.

Veamos a continuación algunos ejemplos interesantes en el contexto del Análisis Multivariante.

Ejemplo A.0.1. *Sea $\mathcal{X} = \mathbb{R}^p$ y $G = O(p)$ el grupo de matrices ortogonales de orden $p \times p$. Definamos la acción de $\mathbf{H} \in G$ sobre $\mathbf{x} \in \mathcal{X}$ como $\mathbf{x} \to \mathbf{Hx}$ y consideremos el producto de matrices en G. Vamos a probar que $\phi(\mathbf{x}) = \mathbf{x}^{\mathbf{t}}\mathbf{x} = ||\mathbf{x}||_2^2$ es un invariante maximal bajo G.*

Notemos en primer lugar que la órbita de un elemento $\mathbf{x}$ de $\mathcal{X}$ bajo G será el conjunto $Orb(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{y} = \mathbf{Hx}, \mathbf{H} \in G\}$ y para comprobar lo que pretendemos bastará con probar que esa órbita coincide con el conjunto de puntos de $\mathbb{R}^p$ que distan del origen igual que $\mathbf{x}$. En efecto, si $\mathbf{y} \in Orb(\mathbf{x})$ entonces $\mathbf{y} = \mathbf{Hx}$ con $\mathbf{H} \in G$. Por lo tanto $||\mathbf{y}||_2^2 = ||\mathbf{Hx}||_2^2 = ||\mathbf{x}||_2^2$, lo cual ya prueba que ϕ es invariante. Ahora bien si $\mathbf{x}$ e $\mathbf{y}$ son tales que $\mathbf{y}^{\mathbf{t}}\mathbf{y} = \mathbf{x}^{\mathbf{t}}\mathbf{x}$, consideremos $\mathbf{H_x}$ y $\mathbf{H_y}$ en G tales que verifiquen $\mathbf{H_x x} = (||\mathbf{x}||_2, 0, \ldots, 0)^t$ y $\mathbf{H_y y} = (||\mathbf{y}||_2, 0, \ldots, 0)^t$ (lo cual se puede conseguir de forma fácil, por ejemplo, tomando $\mathbf{H_x}$ como la matriz con primera fila

$\mathbf{x}^t/||\mathbf{x}||_2$ *y las demás filas ortogonales a ella). Evidentemente se tiene que* $\mathbf{H_x x} = \mathbf{H_y y}$*, por lo que* $\mathbf{y} = \mathbf{Hx}$ *con* $\mathbf{H} = \mathbf{H_y}^t\mathbf{H_x}$ *que pertenece a* G*. Así pues* $\mathbf{x} \sim \mathbf{y}$ *(mod* G*) con lo cual se prueba que* ϕ *es invariante maximal. De esta manera, cualquier función sobre* $\mathbb{R}^p$ *que sea invariante bajo el grupo ortogonal dependerá de* $\mathbf{x}$ *sólo a través de su norma euclídea.*

Ejemplo A.0.2. *Sea* $\mathcal{X} = \mathcal{M}_{n\times p}$ *el espacio de matrices de orden* $n \times p$ *y* $G = O(n)$ *el grupo de matrices ortogonales de orden* $n \times n$*. Definamos la acción de* $\mathbf{H} \in G$ *sobre* $X \in \mathcal{X}$ *como* $\mathbf{X} \to \mathbf{HX}$ *y consideremos el producto de matrices en* G*. Vamos a probar que* $\phi() = \mathbf{X^t X}$ *es un invariante maximal bajo* G*.*

En efecto, que ϕ *es invariante es inmediato ya que* $\phi(\mathbf{HX}) = \mathbf{X}^t\mathbf{H}^t\mathbf{HX} = \mathbf{X}^t$*. Por otro lado sean* $\mathbf{X}$ *e* $\mathbf{Y}$ *pertenecientes a* $\mathcal{X}$ *dadas por columnas como* $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_p]$ *e* $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_p]$ *y supongamos que* $\mathbf{X}^t\mathbf{X} = \mathbf{Y}^t\mathbf{Y}$*. Ello significa que* $\mathbf{x}_i^t\mathbf{x}_i = \mathbf{y}_i^t\mathbf{y}_i$*,* $i = 1, \ldots, p$ *y que* $\mathbf{y}_i^t\mathbf{y}_j = \mathbf{x}_i^t\mathbf{x}_j$*,* $i, j = 1, \ldots, p$*, (*$i \neq j$*). Por el ejemplo anterior deben existir matrices* $\mathbf{H}_i$*,* $i = 1, \ldots, p$ *tales que* $\mathbf{y}_i = \mathbf{H}_i\mathbf{x}_i$*, pero puesto que* $\mathbf{y}_i^t\mathbf{y}_j = \mathbf{x}_i^t\mathbf{x}_j$*, entonces se tiene que verificar que* $\mathbf{H}_i^t\mathbf{H}_j = \mathbf{I}_n$ *lo cual, junto a que tanto* $\mathbf{H}_i$ *como* $\mathbf{H}_j$ *son ortogonales conducen a que* $\mathbf{H}_i = \mathbf{H}_j = \mathbf{H}$ *(*$i \neq j$*). Por lo tanto se tiene* $\mathbf{Y} = [\mathbf{Hx}_1, \ldots, \mathbf{Hx}_p] = HX$*, por lo que* $\mathbf{X} \sim \mathbf{Y}$ *y de esta forma* ϕ *es invariante maximal.*

Ejemplo A.0.3. *Sea* $\mathcal{X} = \mathbb{R}^2 \times \mathcal{S}_2$*, siendo* $\mathcal{S}_2$ *el espacio de matrices simétricas definidas positivas de orden 2 y sea* G *el siguiente grupo de transformaciones*

$$G = \left\{ (\mathbf{B}, \mathbf{c}) \;:\; \mathbf{B} = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},\; b_1 > 0, b_2 > 0 \right\}$$

Definamos en G *la operación* $(\mathbf{B}_1, \mathbf{c}_1)(\mathbf{B}_2, \mathbf{c}_2) = (\mathbf{B}_1\mathbf{B}_2, \mathbf{B}_1\mathbf{c}_2 + \mathbf{c}_1)$
y la acción de G *sobre* $\mathcal{X}$ *como* $(\mu, \Sigma) \to (\mathbf{B}, \mathbf{c})(\mu, \Sigma) = (\mathbf{B}\mu + \mathbf{c}, \mathbf{B}\Sigma\mathbf{B}^t)$.

Vamos a probar que un invariante maximal sobre G *es*

$$\phi(\mu, \Sigma) = \frac{\sigma_{12}}{(\sigma_{11}\sigma_{22})^{\frac{1}{2}}}$$

En primer lugar notemos que si $(\mathbf{B}, \mathbf{c})$ *pertenece a* G *entonces*

$$\mathbf{B}\Sigma\mathbf{B}^t = \begin{pmatrix} b_1^2\sigma_{11} & b_1b_2\sigma_{12} \\ b_1b_2\sigma_{12} & b_2^2\sigma_{22} \end{pmatrix}$$

por lo que

$$\phi(\mathbf{B}\mu+\mathbf{c},\mathbf{B}\Sigma\mathbf{B}^t)=\frac{b_1b_2\sigma_{12}}{\left(b_1^2\sigma_{11}b_2^2\sigma_{22}\right)^{\frac{1}{2}}}=\frac{\sigma_{12}}{\left(\sigma_{11}\sigma_{22}\right)^{\frac{1}{2}}}=\phi(\mu,\Sigma)$$

y así ϕ es invariante. Por otro lado, si $\phi(\mu,\Sigma)=\phi(\tau,\Gamma)$, entonces

$$\frac{\sigma_{12}}{\left(\sigma_{11}\sigma_{22}\right)^{\frac{1}{2}}}=\frac{\gamma_{12}}{\left(\gamma_{11}\gamma_{22}\right)^{\frac{1}{2}}}$$

Consideremos $b_1=\left(\frac{\gamma_{11}}{\sigma_{11}}\right)^{\frac{1}{2}}$, $b_2=\left(\frac{\gamma_{22}}{\sigma_{22}}\right)^{\frac{1}{2}}$ y $c=-B\mu+\tau$. Así

$$\mathbf{B}\Sigma\mathbf{B}^t=\begin{pmatrix}\gamma_{11} & \dfrac{\sigma_{12}\gamma_{11}^{\frac{1}{2}}\gamma_{22}^{\frac{1}{2}}}{\sigma_{11}^{\frac{1}{2}}\sigma_{22}^{\frac{1}{2}}}\\ \dfrac{\sigma_{12}\gamma_{11}^{\frac{1}{2}}\gamma_{22}^{\frac{1}{2}}}{\sigma_{11}^{\frac{1}{2}}\sigma_{22}^{\frac{1}{2}}} & \gamma_{22}\end{pmatrix}=\Gamma$$

por lo que $(\mathbf{B},\mathbf{c})(\mu,\Sigma)=(\tau,\Gamma)$ y con ello $(\mu,\Sigma)\sim(\tau,\Gamma)$ y ϕ es invariante maximal.

Si ahora contemplamos el espacio $\mathcal{X}$ como el espacio de todos los posibles vectores media y matrices de varianzas-covarianzas de un vector aleatorio bidimensional, $\mathbf{X}=(X_1,X_2)^t$, la acción de grupo anterior es inducida por el cambio de variable $\mathbf{Y}=\mathbf{B}\mathbf{X}+\mathbf{c}$ en el sentido de que la media de $\mathbf{Y}$ es $\mathbf{B}\mu+\mathbf{c}$ y la matriz de varianzas-covarianzas de $\mathbf{Y}$ es $\mathbf{B}\Sigma\mathbf{B}^t$. Por lo tanto lo que hemos probado, aplicado a este ambiente, es que el coeficiente de correlación entre las variables X_1 y X_2 es un invariante maximal bajo G y por lo tanto cualquier función invariante para $\mathcal{X}$, bajo G, debe ser función del coeficiente de correlación.

Ejemplo A.0.4. *Sea ahora $\mathcal{X}=\{(\mu,\Sigma):\ \mu\in\mathbb{R}^p,\Sigma_{p\times p}>0\}$ y G el grupo de transformaciones $G=\{(\mathbf{H},\mathbf{c}):\ \mathbf{H}\in O(p),\mathbf{c}\in\mathbb{R}^p\}$ (con la misma operación definida anteriormente) actuando sobre $\mathcal{X}$ de la forma $(\mu,\Sigma)\to(\mathbf{H},\mathbf{c})(\mu,\Sigma)=(\mathbf{H}\mu+\mathbf{c},\mathbf{H}\Sigma\mathbf{H}^t)$.*

Entonces un invariante maximal bajo G es $\phi(\mu,\Sigma)=\lambda(\Sigma)=\mathrm{diag}(\lambda_1,\ldots,\lambda_p)$, siendo $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_p>0$ los autovalores de Σ.

En efecto, ϕ es invariante puesto que Σ y $\mathbf{H}\Sigma\mathbf{H}^t$, con $\mathbf{H}\in O(p)$ tienen los mismos autovalores. Para probar que es maximal supongamos que $\phi(\mu,\Sigma)=\phi(\tau,\Gamma)$, es decir, Σ y Γ tienen los mismos autovalores. Si ello es así existirán matrices ortogonales $\mathbf{H}_1$ y $\mathbf{H}_2$ que factorizan de forma conjunta a Σ y Γ, esto es, $\mathbf{H}_1\Sigma\mathbf{H}_1^t=\mathbf{H}_2\Gamma\mathbf{H}_2^t=\mathrm{diag}(\lambda_1,\ldots,\lambda_p)$. Por lo tanto podemos *escribir* $\Gamma = \mathbf{H}_2^t\mathbf{H}_1\Sigma\mathbf{H}_1^t\mathbf{H}_2 = \mathbf{H}\Sigma\mathbf{H}^t$, *con* $\mathbf{H} = \mathbf{H}_2^t\mathbf{H}_1 \in O(p)$. *Tomando* $\mathbf{c} = -\mathbf{H}\mu + \tau$ *se tendrá* $(\mathbf{H},\mathbf{c})(\mu,\Sigma) = (\mathbf{H}\mu + \mathbf{c}, \mathbf{H}\Sigma\mathbf{H}^t) = (\tau,\Gamma)$, *por lo que* $(\mu,\Sigma) \sim (\tau,\Gamma)$ *y así* ϕ *es maximal invariante.*

Ejemplo A.0.5. *Sea* $\mathcal{X} = \mathcal{M}_{n\times p}$ *el espacio de matrices de dimensión* $n \times p$ *y* G *el grupo dado por* $G = \{(\mathbf{H},\mathbf{P}) : \ \mathbf{H} \in O(n), \mathbf{P} \in O(p)\}$. *Consideremos en* G *el producto de matrices por pares y la acción de* G *sobre* $\mathcal{X}$ *dada por* $(\mathbf{H},\mathbf{P})\mathbf{X} = \mathbf{HXP}$. *Entonces un invariante maximal bajo* G *es* $\phi(\mathbf{X}) = \lambda(\mathbf{X}^t\mathbf{X})$, *las raices características de* $\mathbf{X}^t\mathbf{X}$.

En primer lugar tenemos que $\phi(\mathbf{HXP}) = \lambda((\mathbf{HXP})^t(\mathbf{HXP})) = \lambda(\mathbf{P}^t\mathbf{X}^t\mathbf{XP}) = \lambda(\mathbf{X}^t\mathbf{X}) = \phi(\mathbf{X})$, *por lo que* ϕ *es invariante. Supongamos ahora que* $\phi(\mathbf{X}) = \phi(\mathbf{Y})$, *o sea,* $\mathbf{Y}^t\mathbf{Y}$ *y* $\mathbf{X}^t\mathbf{X}$ *tienen los mismos autovalores. Razonando como en el ejemplo A.0.4 debe existir una matriz* $\mathbf{P} \in O(p)$ *tal que* $\mathbf{Y}^t\mathbf{Y} = \mathbf{PX}^t\mathbf{XP}^t = (\mathbf{XP}^t)^t(\mathbf{XP}^t)$ *y, aplicando el ejemplo A.0.2, debe existir* $\mathbf{H} \in O(n)$ *tal que* $\mathbf{Y} = \mathbf{HXP}^t$, *por lo que* $\mathbf{X} \sim \mathbf{Y}$ *y de esta forma* ϕ *es invariante maximal.*