Estadística Multivariante
Derivación matricial

Trabajo B

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

Índice

1. Introducción

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

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

2. Diferencial primera y jacobianos

2.1. Diferencial de una función vectorial

Definición 2.2.1: Consideramos una función vectorial $f : S \to \mathbb{R}^m$ con $S \subset \mathbb{R}^n$. Sea **c** un punto interior de S y consideremos una bola cerrada con centro en **c** y radio **r**, $B(c, r)$. Sea u un punto de $\mathbb{R}^n$ tal que $||u|| \leq r$ es decir, $c + r \in B(c, r)$.
Diremos que f es **diferenciable** en **c** si existe una matriz real de orden m x n que depende de c y no de u y que cumple que $f(c + u) - f(c) = A(c)u + r_c(u)$ con $\lim_{u\to 0} \frac{r_c(u)}{||u||} = 0$. Además, se define la **primera diferencial** de f en el punto c con incremento u como: $df(c; u) = A(c)u$.

Definición 2.2.2: Sea $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ y sea $f_i : S \to \mathbb{R}$ su i-ésima componente. Sea c un punto interior de S y e_j el j-ésimo vector de la base canónica de $\mathbb{R}^n$. Se define la **derivada parcial** de f respecto a la j-ésima coordenada como:

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

Teorema 2.1. *(Primer teorema de identificación para funciones vectoriales)*
*Sea $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ diferenciable en un punto c interior de S y $u \in \mathbb{R}^n$. Entonces $df(c; u) = (Df(c))u$ donde $Df(c)$ es una matriz mxn cuyos elementos $D_j f_i(c)$ son las derivadas parciales de **f** evaluadas en c y que recibe el nombre de matriz jacobiana. Recíprocamente, si $A(c)$ es una matriz que verifica que $df(c; u) = A(c)u \forall u \in \mathbb{R}^n$, entonces $A(c) = Df(c)$.*

El siguiente teorema nos proporciona la regla de la cadena para funciones vectoriales.

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

Teorema 2.3. *(Regla de invarianza de Cauchy)*
*En el ambiente del teorema anterior, si **f** es diferenciable en **c** y **g** lo es en **b=f(c)**, entonces la diferenciable $h = gof$ es $dh(c; u) = dg(b; df(c; u))$.*

2.2. Diferencial de una función matricial

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

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

Definición 2.1. En las condiciones anteriores, se dice que F es diferenciable en C si existe una matriz A de dimensiones $mp \times nq$, que dependa de C y no de U y tal que

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

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

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

Tenemos los siguientes resultados análogos a los del caso vectorial:

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

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

Teorema 2.6. *(Regla de invarianza de Cauchy para funciones matriciales) En las condiciones del teorema anterior, $dH(C;U) = dG(B;dF(C;U))$, $\forall U \in \ltimes \times \shortparallel$.*

Corolario 2.7. *Sean A_{txm}, B_{pxr} **y** C_{sxu} matrices constantes y sean G_{mxp} **y** H_{rxs} dos funciones matriciales diferenciables. Si $F(X) = AG(X)BH(X)C$, entonces*

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

y $DF(X) = (C^tH^t(X)B^t \bigotimes A)DG(X) + (C^t \bigotimes AG(X)B)DH(X)$

*****************COPIAMOS LAS MILDOSCIENTAS34 PROPIEDADES????

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

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

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

Solución: Para comenzar, tengamos en cuenta que h es composición de $f : \mathbb{R}^k \to \mathbb{R}^n$ y $g : \mathbb{R}^n \to \mathbb{R}$, con $f(c) = (y - Xc)$ con X una matriz nxk y con $y \in \mathbb{R}^n$; y $g(X) = X^tX$, de modo que $h(x) = g(f(x))\forall x \in \mathbb{R}^k$. Así, es claro que $dh(c;u) = dg(y - Xc; df(c;u))$. *********************** Parece evidente pero no se que escribir xd***************

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

Solución:

EJERCICIO 2.3: Si $X_{n\times n}$ es una matriz simétrica y $F : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$ es diferenciable, demostrar que $d\text{Vec}(F(X)) = D_n DF(X) d\text{Vech}(X)$, mientras que $d\text{Vec}(F(X)) = N_n DF(X) d\text{Vec}(X)$ donde $N_n = \frac{1}{2}[I_{n^2} + K_{nn}]$.

Solución:

3. Matrices jacobianas y derivadas matriciales

No existe una única definición para la derivada de una función de argumento matricial y esto supone un problema a la hora de usar el cálculo diferencial matricial.

Veamos las definiciones clásicas para funciones reales de argumento vectorial y vectoriales, tanto de argumento real como vectorial.

Definición 3.1. Sea $f : \mathbb{R}^n \to \mathbb{R}$. Se define la derivada de f respecto de $x \in \mathbb{R}^n$ como el vector $1 \times n$ dado por $\dfrac{\partial f(x)}{\partial x^t} = \left(\dfrac{\partial f(x)}{\partial x_1}, ..., \dfrac{\partial f(x)}{\partial x_n}\right)$.

Definición 3.2. Sea $f : \mathbb{R} \to \mathbb{R}^m$. Se define la derivada de f respecto de $x \in \mathbb{R}$ como el vector $m \times 1$ dado por $\dfrac{\partial f(x)}{\partial x} = \left(\dfrac{\partial f_1(x)}{\partial x}, ..., \dfrac{\partial f_m(x)}{\partial x}\right)^t$.

Definición 3.3. Sea $f : \mathbb{R}^n \to \mathbb{R}^m$. Se define la derivada de f respecto de $x \in \mathbb{R}^n$ como la matriz $m \times n$

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

Definición 3.4. Sea $F : \mathbb{M}_{n\times q} \to \mathbb{R}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n\times q}$ como la matriz $n \times q$

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

Definición 3.5. Sea $F : \mathbb{R} \to \mathbb{M}_{m\times p}$. Se define la derivada de F respecto de $x \in \mathbb{R}$ como la matriz $m \times p$

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

**********CREO que las importantes son solo las dos ultimas, quizas haya que borrar algunas de estas aunque eso se decidirá en un futuro xd ******

Existen así muchas formas de definir la derivada de una función de argumento matricial, aunque ahora vamos a mostrar dos que generalizan a las demás:

Definición 3.6. (Definición de derivada matricial según Mac Rae)
Sea $F : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n\times q}$ como la matriz $nm \times pq$

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

Definición 3.7. (Definición de derivada matricial según Dwyer)
Sea $F : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n\times q}$ como la matriz $nm \times pq$

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

Teorema 3.1. *Sea $F : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$. Entonces se verifica que $K_{nm}\frac{\partial F(x)}{\partial x}K_{pq} = \frac{\partial F(x)}{\partial x}$*

Definición 3.8. (Definición de derivada matricial según Magnus y Neudecker)
Sea $F : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n\times q}$ como la matriz $mp \times nq$ dada por $DF(X) = \frac{\partial Vec(F(x))}{\partial Vec^t(x)}$

********************************MILLONES Y MILLONES DE PROPIEDADES XDXDXDXDXD-MUERO *************

Teorema 3.2. *Sea $F : \mathbb{M}_{n\times q} \to \mathbb{M}_{m\times p}$. Entonces*

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

Corolario 3.3. *Sean A_{txm}, B_{pxr} y C_{sxu} matrices constantes y sean G_{mxp} y H_{rxs} dos funciones de argumento matricial $X_{n\times q}$. Si $F(X) = AG(X)BH(X)C$, entonces*

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

Ejercicio 3.1: A partir de las relaciones existentes entre la derivada matricial y la matriz jacobiana, verificar las siguientes expresiones:

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

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

Ejercicio 3.2: Sea $X_{n\times q}$. Demostrar las siguientes igualdades:

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

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

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

Ejercicio 3.3: Demostrar que si $X_{n\times n}$ es no singular entonces $\dfrac{\partial X^{-1}}{\partial X} = -\mathrm{Vec}((X^{-1})^t)\mathrm{Vec}^t(X^{-1})$.

Solución:

3.1. Derivadas matriciales de funciones escalares de un vector

Las principales funciones que suelen aparecer en la práctica son:

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

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

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

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

- Si $\phi(x) = a^t f(x) \to D\phi(x) = a^t Df(x)$
- Si $\phi(x) = f(x)^t g(x) \to D\phi(x) = g(x)^t Df(x) + f(x)^t Dg(x)$
- Si $\phi(x) = x^t Af(x) \to D\phi(x) = f(x)^t A^t + x^t ADf(x)$
- Si $\phi(x) = f(x)^t Af(x) \to D\phi(x) = f(x)^t(A + A^t)Df(x)$
- Si $\phi(x) = f(x)^t Ag(x) \to D\phi(x) = g(x)^t A^t Df(x) + f(x)^t ADg(x)$
- Si $x = (x_1^t, x_2^t)^t$ y Si $\phi(x) = x_1^t Ax_2 \to D\phi(x) = x^t(\begin{matrix} 0 & A \\ A^t & 0 \end{matrix})$

3.2. Derivadas matriciales de funciones escalares de matrices

************PUEDE QUE FALTE LITERATURA EN MUCHAS PARTES XD*******

Derivadas matriciales asociadas a trazas

Teorema 3.4. *Sea* $F : \mathbb{M}_{n\times q} \to \mathbb{M}_{p\times p}$ *y* $\phi(x) = tr[]F(x)]$. *Entonces* $d\phi(x) = tr[dF(x)]$, $D\phi(x) = Vec^t(I_p)DF(x)$ *y*

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

Teorema 3.5. *Sean* $F_{p\times m}$, $G_{m\times p}$ *dos funciones matriciales de* $X_{n\times q}$. *Entonces se verifica:*

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

4. Diferencial segunda y hessianos