矩阵代数

张伟平

zwp@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~zwp/

论坛 http://fisher.stat.ustc.edu.cn

目录

1.1 矩阵和向量基础

多元数据常用矩阵和向量来表示, 本讲我们回顾一下基础的矩阵代数. 本节小写黑体字母表示向量, 如 $\mathbf{x}$; 大写黑体字母表示矩阵, 如 $\mathbf{X}$.

一. 向量

- n 维列向量: $\mathbf{x}' = [x_1, \ldots, x_n]$
- 长度: $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$
- $\mathbf{x}$ 和 $\mathbf{y}$ 的夹角: $cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{<\mathbf{x},\mathbf{y}>}{\|\mathbf{x}\|\|\mathbf{y}\|}$
- 向量 $\mathbf{x}$ 在 $\mathbf{y}$ 上的投影 $= \frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}\mathbf{y} = \frac{<\mathbf{x},\mathbf{y}>}{\|\mathbf{y}\|}\frac{\mathbf{y}}{\|\mathbf{y}\|} = \|\mathbf{x}\|cos(\theta)\frac{\mathbf{y}}{\|\mathbf{y}\|} = \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}'\mathbf{x} = P_y\mathbf{x}$, 其中 θ 为 $\mathbf{x}$ 和 $\mathbf{y}$ 的夹角. $P_y = \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}'$ 称为投影阵.

- 向量集 $\mathbf{x}_1,\ldots,\mathbf{x}_p$**线性相关**(linearly dependent): 存在不全为零的常数 $c_1,\ldots,c_p$ 使得

$$c_1\mathbf{x}_1+\cdots+c_p\mathbf{x}_p=0$$

如果不存在满足上式的常数 $c_1,\ldots,c_p$, 则称向量集 $\mathbf{x}_1,\ldots,\mathbf{x}_p$**线性无关**(linearly independent).

  - 任何含有零向量的向量集总是线性相关的
  - 向量集 $\mathbf{x}_1,\ldots,\mathbf{x}_p$ 线性无关的充要条件是: 如果一组常数 $c_1,\ldots,c_p$ 使得 $c_1\mathbf{x}_1+\cdots+c_p\mathbf{x}_p=0$, 则必有 $c_1=\cdots=c_p=0$.
  - 设 $\mathbf{x}_1,\ldots,\mathbf{x}_p$ 是非零向量, 它们线性相关的充要条件是: 存在 i, 使得

$$\mathbf{x}_i=b_1\mathbf{x}_1+\cdots+b_{i-1}\mathbf{x}_{i-1}+b_{i+1}\mathbf{x}_{i+1}+\cdots+b_p\mathbf{x}_p$$

其中 $b_1,\cdots,b_p$ 为常数.

称 R^n 的子集 H 为**线性空间**，如果
(1) 对任意的 $\mathbf{x} \in H, \mathbf{y} \in H$ 必有 $\mathbf{x}+\mathbf{y} \in H$，以及
(2) $c\mathbf{x} \in H$ 对一切实数 c 成立.

Definition

- $\{\mathbf{x} = [x_1, \ldots, x_n]' \in R^n | x_1 + \cdots + x_n = 0\}$ 为 R^n 的线性子空间
- 对给定的 $\mathbf{x} \in R^n$, 所有正交于 $\mathbf{x}$ 的向量构成线性空间
- 给定 R^n 中的一些向量 $\mathbf{x}_1, \ldots, \mathbf{x}_k$, 令

$$\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \{\sum_{i=1}^{k} c_i \mathbf{x}_i | (c_1, \ldots, c_k)' \in R^n\}$$

则 $\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ 为线性空间, 称为是 $\mathbf{x}_1, \ldots, \mathbf{x}_k$ 所张成的线性空间.

- $n\times p$ 维矩阵 $A=[\mathbf{x}_1,\ldots,\mathbf{x}_p]$ 的列向量生成的线性空间记为 $\mathcal{L}(A)$. 可以证明 $Rank(A)=dim(\mathcal{L}(A))$.

- 设 H 和 G 是两个线性子空间, 若对任意的 $\mathbf{a}\in H,\mathbf{b}\in G$, 有 $\mathbf{a}'\mathbf{b}=0$, 则称 H 和 G<u>正交</u>, 记作 $H\perp G$. 进一步, 若 $H+G=R^n$, 则称 G 为 H 的<u>正交补空间</u>, 记作 $G=H^{\perp}$.
  - 若 A 为投影阵, 则 $\mathcal{L}(A)$ 与 $\mathcal{L}(I-A)$ 互为正交补空间.
  - A 为任意矩阵, 则 $\mathcal{L}(A)=\mathcal{L}(AA'),\mathcal{L}(A')=\mathcal{L}(A'A)$.

- (零空间) 称线性子空间 $\mathcal{N}(A)=\{\mathbf{x}|A\mathbf{x}=0\}$ 为矩阵 A 的零空间. $\mathcal{N}(A)$ 是 A 零特征根对应的特征向量所张成的线性空间. 从而

$$dim(\mathcal{N}(A))+rank(A)=q(A\text{的列数})$$

易知 $\mathcal{N}(A'A)=\mathcal{N}(A),\mathcal{N}(AA')=\mathcal{N}(A')$.

二. 矩阵, 行列式, 逆和秩

- np 个实数排出的 n 行 p 列阵称为 $n \times p$ 维 (实) 矩阵. $n = p$ 时称为方阵.

- 若 A 为方阵, 且 $A' = A$, 则称 A 为对称阵; 若 $A' = -A$, 则称 A 为斜对称阵.

- 方阵 A 称为是正交矩阵, 若 $AA' = A'A = I$; 方阵 A 称为是幂等的, 若 $A^2 = A$; 对称的幂等阵称为是投影阵.

若 $A = (a_{ij})$ 为 p 阶方阵, 记 $|A| = \sum_\pi \epsilon_\pi a_{1j_1} a_{2j_2} \cdots a_{pj_p}$, 其中 $(j_1, \ldots, j_p)$ 为 $(1, 2, \cdots, p)$ 的任一置换. $\sum_\pi$ 是对全部可能的 $p!$ 个置换求和, $\epsilon_\pi = 1$ 或 -1 相应地取决于偶置换或奇置换. $|A|$ 称为 A 的行列式.

Definition

- $|A| = \sum_{j=1}^{p} a_{ij} A_{ij}$, 其中 A_{ij} 为元素 a_{ij} 的代数余子式.

- 设 A 为 p 阶方阵, 则 $|aA| = a^p |A|$.

- 设 A 和 B 分别为 $p \times q$ 和 $q \times p$ 的矩阵, 则 $|I_p + AB| = |I_q + BA|$

- 若 A 为正交阵, 则 $|A| = \pm 1$

设 A 为 p 阶方阵，若 $|A| \neq 0$，则称 A 是非退化阵；若 $|A| = 0$，则称 A 是退化阵．设 A 是 p 阶非退化阵，若存在一个唯一的矩阵 B，使得 $AB = BA = I_p$，则称 B 为矩阵 A 的逆，记为 $B = A^{-1}$．

Definition

- $(A')^{-1} = (A^{-1})'$, $(AC)^{-1} = C^{-1} A^{-1}$, $|A^{-1}| = |A|^{-1}$

- 上 (下) 三角矩阵的逆仍为上 (下) 三角矩阵.

- 若 A 和 B 分别为 p 和 q 阶可逆方阵, C 和 D 分别为 $p\times q$ 阶和 $q\times p$ 矩阵. 令 $T=A+CBD$, 则

$$T^{-1}=A^{-1}-A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1}$$

设 A 为 $p\times q$ 矩阵，若存在 A 的一个 r 阶子方阵的行列式不为零，而 A 的所有 $r+1$ 阶子方阵的行列式均为零，则称 A 的秩为 r，记为 $Rank(A)=r$. Definition

- $Rank(A)=Rank(A')=Rank(A'A)=Rank(AA')$
- $Rank(A)=0$, 当且仅当 $A=0$.
- $0\le Rank(A)\le min(p,q)$.
- $Rank(AB)\le min(Rank(A),Rank(B))$.

- $Rank(A+B) \leq Rank(A) + Rank(B)$.
- 若 A 和 C 为非退化方阵, 则 $Rank(ABC) = Rank(B)$.

1.2 正定阵, 非负定阵和投影阵

称 p 阶对称阵 A 为正定矩阵, 如果对一切 $\mathbf{x} \neq 0, \mathbf{x} \in R^p$, 有 $\mathbf{x}' A \mathbf{x} > 0$. 记作 $A > 0$;
称 p 阶对称阵 A 为非负定矩阵, 如果对一切 $\mathbf{x} \in R^p$, 有 $\mathbf{x}' A \mathbf{x} \geq 0$. 记作 $A \geq 0$.

Definition

正定阵和非负定阵的一些性质:

- 一个对称阵是正 (非负) 定的, 当且仅当它的特征根为正 (非负).

- 若 $A > 0$, 则 $A^{-1} > 0$.
- 若 $A > 0, B > 0, A - B > 0$, 则 $B^{-1} - A^{-1} > 0$, 且 $|A| > |B|$.
- 若 $A > 0$, 将 A 分块为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

其中 A_{11} 为方阵, 则 $A_{11} > 0$, $A_{22} > 0$, $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21} > 0$, $A_{22\cdot 1} = A_{22} - A_{21}A_{11}^{-1}A_{12} > 0$.

- 若 $A \geq 0$, 则必存在正交矩阵 Γ 使得

$$\Gamma' A \Gamma = diag\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$$

其中 $\lambda_i \geq 0, i = 1, \ldots, p$ 为 A 的特征根.

在非负定矩阵中, 有一类重要的矩阵叫**投影阵**(对称幂等阵). 它有以下性质:

- 若 A 为投影阵, 则 $tr(A) = Rank(A)$.
- 若 A 为投影阵, 则 $I - A$ 也为投影阵.
- 若 A 和 B 为投影阵, 且 $A + B = I$, 则 $AB = BA = 0$.
- 若 X 为 $n \times p$ 矩阵, $n \geq p$, $Rank(X) = p$, 则 $P_X = X(X'X)^{-1}X'$ 为投影阵, 且 $Rank(P_X) = p$.

定理 1. n 阶方阵 P 为正交投影阵, 当且仅当对任给向量 $\mathbf{x} \in R^n$ 有

$$\|\mathbf{x} - P\mathbf{x}\| = inf\|\mathbf{x} - \mathbf{u}\|, \mathbf{u} \in \mathcal{L}(P)$$

这个定理刻画了正交投影阵的距离最短性, 即在线性子空间 $\mathcal{L}(P)$ 的所有向量中, 只有和 $\mathbf{x}$ 的正交投影 $P\mathbf{x}$ 到 $\mathbf{x}$ 的距离最短. 这个结果在最小二乘估计理论中有重要应用.

设 P_1 和 P_2 为两个正交投影阵, 则

- $P_1 + P_2$ 为正交投影阵 $\Longleftrightarrow P_1P_2 = P_2P_1 = 0.$
- P_1P_2 为正交投影阵 $\Longleftrightarrow P_1P_2 = P_2P_1.$
- $P_1 - P_2$ 为正交投影阵 $\Longleftrightarrow P_1P_2 = P_2P_1 = P_2.$

对对称阵 A 的特征根 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, 有以下结果

- $\sup\limits_{\mathbf{x}\neq 0} \frac{\mathbf{x}'A\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \sup\limits_{\|\mathbf{x}\|=1} \mathbf{x}'A\mathbf{x} = \lambda_1$
- $\inf\limits_{\mathbf{x}\neq 0} \frac{\mathbf{x}'A\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \inf\limits_{\|\mathbf{x}\|=1} \mathbf{x}'A\mathbf{x} = \lambda_p$

1.3 矩阵的因子分解

- 若 A 为 p 阶对称方阵, 则存在正交阵 Γ 及对角阵 $\Lambda = diag\{\lambda_1, \ldots, \lambda_p\}(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p)$, 使得

$$A = \Gamma' \Lambda \Gamma$$

若记 $\Gamma' = [l_1, l_2, \cdots, l_p]$, 则

$$A = \sum_{i=1}^{p} \lambda_i l_i l_i'$$

称为矩阵 A 的**谱分解**(spectral decomposition).

- 若 $A > 0(\geq 0)$, 则存在 $A^{1/2} > 0(\geq 0)$, 使得 $A = A^{1/2}A^{1/2}$.
- 设 $A \geq 0$ 是 p 阶秩为 r 的矩阵, 则
  - 存在一个秩为 r 的 $p \times r$ 阵 B, 使得 $A = BB'$.

– 存在一个 p 阶非退化阵 C, 使得 $A = C \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} C'$.

– 存在一个对角元素非负的上三角矩阵 T, 使得 $A = T'T$. (Cholesky 分解), 若 $A > 0$, 则此分解是唯一的.

- (**奇异值分解**) 设 A 是 $n \times p$ 矩阵, $Rank(A) = r$, 则存在 $n \times n$ 阵正交阵 U 和 $p \times p$ 阶正交阵 V 使得

$$A = UDV'$$

其中 $n \times p$ 阶矩阵

$$D = \begin{pmatrix} D_{11} & 0_{r\times(p-r)} \\ 0 & 0 \end{pmatrix},$$

第一块 $D_{11} = diag\{\lambda_1, \ldots, \lambda_r\}$, $\lambda_1^2 \geq \ldots \geq \lambda_r^2 > 0$ 是矩阵 $A'A$ 非零特征根 (λ_i 称为矩阵 A 的奇异值.) V 的列向量为矩阵 $A'A$ 的特征向量, U 的列向量为 AA' 的特征向量.

- (QR 分解) 任何实矩阵 A 可以分解为 $A = QR$, 其中 Q 为列正交矩阵, R 为上三角矩阵.

- 设 A 为任一 $n \times p$ 矩阵, 其秩为 r, 则存在 $n \times r$ 阵 P 和 $p \times r$ 矩阵 Q 使得 $A = PQ'$.

- (同时对角化) 设 $A_1, \ldots, A_k$ 为 p 阶对称阵, 且 $A_i A_j = 0, i \neq j, i, j = 1, 2, \ldots, k$, 则存在一个正交阵 H, 使得 $H' A_i H = \Lambda_i$, Λ_i 为对角阵, $i = 1, \ldots, k$.

- 设 A 和 B 为 p 阶对称方阵, 且 $A > 0$, 则存在 p 阶非退化阵 H, 使得
$$A = H'H, \quad B = H'\Lambda H$$
其中 $\Lambda = diag\{\lambda_1, \ldots, \lambda_p\}$ 是 $A^{-1}B$ 的特征根.

1.4 分块矩阵

设 A 为 $n \times p$ 矩阵, 将其分为四块:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

其中 A_{11} 为 $k \times l$ 阶. 称此为矩阵 A 的分块表示.

性质 1 若 A 为方阵, A_{11} 也为方阵, 则

- 当 $|A_{11}| \neq 0$ 时, $|A| = |A_{11}||A_{22\cdot 1}|$, 其中 $A_{22\cdot 1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$.
- 当 $|A_{22}| \neq 0$ 时, $|A| = |A_{11\cdot 2}||A_{22}|$, 其中 $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$.

性质 2 若 A 为可逆方阵, A_{11} 和 A_{22} 也为方阵, 则

- 当 $|A_{11}| \neq 0$ 时,

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1} \\ -A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} & A_{22\cdot 1}^{-1} \end{pmatrix}$$

- 当 $|A_{22}| \neq 0$ 时,

$$A^{-1} = \begin{pmatrix} A_{11\cdot 2}^{-1} & -A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}$$

- 当 $|A_{11}| \neq 0$, $|A_{22}| \neq 0$ 时,

$$A^{-1} = \begin{pmatrix} A_{11\cdot 2}^{-1} & -A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} & A_{22\cdot 1}^{-1} \end{pmatrix}$$

1.5 矩阵的广义逆

减号逆

设 A 为 $n \times p$ 阶矩阵，若存在一个 $p \times n$ 矩阵 X，使得 $AXA = A$，则称 X 为 A 的广义逆，记作 $X = A^-$. Definition

- 对任意矩阵 A, 一定存在 (不一定唯一) 广义逆.
- 若 A 非退化, 则 A^- 唯一, 且 $A^- = A^{-1}$.
- $Rank(A^-) \geq Rank(A)$, $Rank(A) = Rank(AA^-) = tr(AA^-)$.
- 对任意矩阵 A, 有 $A'A(A'A)^-A' = A'$, $A(A'A)^-A'A = A$.
- $A(A'A)^-A'$ 为投影阵, 且与 $(A'A)^-$ 的取法无关

加号逆

设 A 为 $n\times p$ 阶矩阵，若存在一个 $p\times n$ 矩阵 X，使得

$$AXA=A, XAX=X, (AX)'=AX, (XA)'=XA$$

Definition

则称 X 为 A 的 Moore-Penrose 广义逆，记作 $X=A^+$.

- 对任意矩阵 A, 加号逆存在且唯一.
- $(A')^+=(A^+)'$.
- 若 $Rank(A)=n$, 则 $A^+=A'(A'A)^{-1}$; 若 $Rank(A)=p$, 则 $A^+=(A'A)^{-1}A'$; 若 $Rank(A)=n=p$, 则 $A^+=A^{-1}$.
- $(A^+)^+=A$

- $(A'A)^+ = A^+(A^+)'$
- $A^+ = (A'A)^+A' = A'(AA')^+$.
- 若 A 是投影阵, 则 $A^+ = A$.
- AA^+ 与 A^+A 是投影阵.
- 若 $A = PQ'$, $Rank(A) = Rank(P) = Rank(Q) = r$, 则 $A^+ = (Q')^+P^+$.
- 若 $A' = A$, 则 $A = H'\Lambda H$, H 为正交阵. $\Lambda = diag\{\lambda_1, \ldots, \lambda_p\}$, 令 $\lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0 \\ 0, & \lambda = 0 \end{cases}$ 则 $A^+ = H'diag\{\lambda_1^+, \ldots, \lambda_p^+\}H$.

1.6 矩阵的拉直运算

所谓拉直运算, 就是将矩阵拉成一个长向量. 通过它建立矩阵和向量之间的联系.

设 $A = (a_{ij}) = (\mathbf{a}_1, \ldots, \mathbf{a}_p)$ 为 $n \times p$ 矩阵, $\mathbf{a}_j$ 表示其第 j 列, $(j = 1, \ldots, p)$. 则 $vec(A) = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_p \end{pmatrix}$ 称为矩阵 A 的按列拉直运算, 记为 $vec(A)$ 或 $\overrightarrow{A}$.

Definition

- 若 c, d 为实数, A 和 B 是两个大小相同的矩阵, 则 $vec(c \cdot A + d \cdot B) = c \cdot vec(A) + d \cdot vec(B)$.
- $tr(AB) = \sum_{ij} a_{ij} b_{ji} = (vec(A'))'(vec(B))$.

- $A=\sum_{ij}a_{ij}E_{ij}=\sum_{ij}a_{ij}\mathbf{e}_i\mathbf{e}_j'=\sum_j\mathbf{a}_j\mathbf{e}_j'=\sum_i\mathbf{e}_i\mathbf{a}_{(i)}'$, 其中 $\mathbf{e}_i$ 表示第 i 维单位向量, $E_{ij}=\mathbf{e}_i\mathbf{e}_j'$, $\mathbf{a}_{(i)}$ 表示 A 的第 i 行.

- $\mathbf{a}_j=A\mathbf{e}_j$, $\mathbf{a}_{(i)}=A'\mathbf{e}_i$.

- $a_{ij}=\mathbf{e}_i'A\mathbf{e}_j$, $tr(E_{ij}'A)=a_{ij}$

设 $A=(a_{ij})$ 和 B 分别为 $n\times p$ 和 $m\times q$ 阵, A 和 B 的 `Kronecker` 积 $A\otimes B$ 定义为

$$A\otimes B=(a_{ij}B)=\begin{pmatrix} a_{11}B & \cdots & a_{1p}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{np}B \end{pmatrix}$$

Definition

- 若 a 为任一实数, 正则 $(aA)\otimes B=A\otimes(aB)=a(A\otimes B)$

- $(A \otimes B)' = A' \otimes B'$, $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

- 若 A, B 为方阵, 则 $tr(A \otimes B) = (tr(A))(tr(B))$

- 若 $\mathbf{x}$ 和 $\mathbf{y}$ 为向量, 则 $\mathbf{x}'\mathbf{y} = \mathbf{x} \otimes \mathbf{y}' = \mathbf{y}' \otimes \mathbf{x}$.

- 若 A 为 $n \times n$ 阵, 其特征根为 $\{\lambda_i, i = 1, \ldots, n\}$, 相应的特征向量为 $\{\mathbf{x}_i, i = 1, \ldots, n\}$. B 为 $m \times m$ 阵, 其特征根为 $\{\mu_i, i = 1, \ldots, m\}$, 相应的特征向量为 $\{\mathbf{y}_i, i = 1, \ldots, m\}$. 则 $A \otimes B$ 的特征根为 $\{\lambda_i \mu_j, i = 1, \ldots, n; j = 1, \ldots, m\}$, 相应的特征向量为 $\{\mathbf{x}_i \otimes \mathbf{y}_j, i = 1, \ldots, n; j = 1, \ldots, m\}$.

- 设 A 为 n 阶方阵, B 为 m 阶方阵, 则 $|A \otimes B| = |A|^m |B|^n$.

- $Rank(A \otimes B) = Rank(A)Rank(B) = Rank(B \otimes A)$

矩阵的拉直运算性质:

- $vec(AXB) = (B' \otimes A)vec(X)$
- $vec(AB) = (I \otimes A)vec(B)$
- $tr(ABC) = vec(A')'(I \otimes B)vec(C)$
- $vec(\mathbf{a}\mathbf{a}') = \mathbf{a} \otimes \mathbf{a}$
- $tr(A'BCD') = vec(A)'(D \otimes B)vec(C)$

1.7 矩阵的微商和变换的雅可比

矩阵微商是通常微商的推广，是求极大似然估计和最小二乘估计的工具，利用它可以方便地求多变量积分变换的雅可比行列式.

矩阵对标量的微商

设 $Y = (y_{ij}(t))$ 是 $p \times q$ 矩阵，它的元素是 t 的函数，则

$$\frac{\partial Y}{\partial t} = \left(\frac{\partial y_{ij}(t)}{\partial t}\right)$$

Definition

称为 Y 对 t 的微商.

- $\frac{\partial X+Y}{\partial t} = \frac{\partial X}{\partial t} + \frac{\partial Y}{\partial t}$
- $\frac{\partial XY}{\partial t} = \frac{\partial X}{\partial t} Y + X \frac{\partial Y}{\partial t}$

- $\frac{\partial X \otimes Y}{\partial t} = \frac{\partial X}{\partial t} \otimes Y + X \otimes \frac{\partial Y}{\partial t}$
- $\left(\frac{\partial X}{\partial t}\right)' = \frac{\partial X'}{\partial t}$
- $\frac{\partial X}{\partial x_{ij}} = E_{ij}$
- $\frac{\partial AXB}{\partial x_{ij}} = AE_{ij}B$
- $\frac{\partial X^{-1}}{\partial t} = -X^{-1}\frac{\partial X}{\partial t}X^{-1}$

矩阵的标量函数对矩阵的微商

设 $y = f(X)$ 是 $p \times q$ 阶矩阵 X 的标量函数，y 对 X 的微商定义为

$$\frac{\partial y}{\partial X} = \left(\frac{\partial y}{\partial x_{ij}}\right)$$

Definition

- $\left(\frac{\partial y}{\partial X}\right)' = \frac{\partial y}{\partial X'}$
- 若 X 为 p 阶方阵, 则 $\frac{\partial tr(X)}{\partial X} = I_p$
- $\frac{\partial tr(AXB)}{\partial X} = A'B'$
- $\frac{\partial tr(AX)}{\partial X} = \begin{cases} A', & X \neq X' \\ A + A' - diag(a_{11}, \ldots, a_{pp}), & X = X' \end{cases}$
- $\frac{\partial x'Ax}{\partial x} = (A + A')x$, x 为列向量; $\frac{\partial tr(X'AX)}{\partial X} = (A + A')X$, X 为矩阵.
- $\frac{\partial \mathbf{a}'x}{\partial x} = \mathbf{a}$, 其中 $\mathbf{a}, x$ 为列向量.
- $\frac{\partial det(X)}{\partial X} = (X')^{-1} det(X)$
- 设 X 为一方阵, 则 $\frac{\partial tr(X^{-1}A)}{\partial X} = -(X^{-1}AX^{-1})'$

向量对向量的微商

Definition

设 $\mathbf{x}$ 为 n 维列向量，m 维列向量 $\mathbf{y} = (y_1, \ldots, y_m)' = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))'$，则 $\mathbf{y}$ 对 $\mathbf{x}$ 的微商定义为

$$\frac{\partial \mathbf{y}'}{\partial \mathbf{x}} = \left(\frac{\partial y_j}{\partial x_i} \right)$$

- 若 $\mathbf{y} = A\mathbf{x}$, 则 $\partial \mathbf{y}'/\partial \mathbf{x} = A'$.
- 若 $Y = AXB$, 则 $\frac{\partial (vec(Y))'}{\partial vec(X)} = B \otimes A'$.

矩阵对矩阵的微商

设 X 是 $m \times n$ 阶矩阵，$F(X)$ 是 $p \times q$ 阶矩阵，记 $F(X) = (f_{ij}(X)) = (f_{ij})$，则

$$\frac{\partial F(X)}{\partial X} = \frac{\partial [vec(F(X))]'}{\partial vec(X)}$$

Definition

称为 $F(X)$ 对 X 的微商.

- 若 X 为 $m \times n$ 阶矩阵, 则 $\frac{\partial X}{\partial X} = I_{mn}$
- 若 $F(X)$ 为 $p \times q$ 阶矩阵, $G(X)$ 是 $q \times r$ 阶矩阵, X 是 $m \times n$ 阶矩阵, 则

$$\frac{\partial F(X)G(X)}{\partial X} = \frac{\partial F(X)}{\partial X}(G(X) \otimes I_p) + \frac{\partial G(X)}{\partial X}(I_r \otimes F'(X))$$

- $\frac{\partial AXB}{\partial X} = B \otimes A'$
- $\frac{\partial X^{-1}}{\partial X} = -[X^{-1} \otimes (X^{-1})']$

定理 2 (Chain rule). 设 $\psi(X)$ 是矩阵变量 X 的数值函数, X 的每个元素 x_{ij} 均为变量 t 的函数, 则

$$\frac{\partial \psi(X)}{\partial t} = tr\left[\frac{\partial X}{\partial t}\left(\frac{\partial \psi}{\partial X}\right)'\right]$$

定理 3 (Chain rule). 设 $F(G(X))$ 是两个矩阵函数的复合函数, 则

$$\frac{\partial F(G(X))}{\partial X} = \frac{\partial G(X)}{\partial X}\frac{\partial F(G)}{\partial G}$$

- $\frac{\partial det(X(t))}{\partial t} = det(X)tr\left(X^{-1}\frac{\partial X}{\partial t}\right)$
- $\frac{\partial lndet(X)}{\partial t} = tr\left(X^{-1}\frac{\partial X}{\partial t}\right)$

雅可比行列式的计算

$$\int_D g(x_1, \ldots, x_n) dx_1 \cdots dx_n = \int_T g(f^{-1}(\mathbf{y}))|J(\mathbf{x} \to \mathbf{y})| dy$$

其中 $D \subset R^n$, $T = \{y | y = f(\mathbf{x}), \mathbf{x} \in D\}$, f 是一一变换; $|J(\mathbf{x} \to \mathbf{y})| = |\frac{\partial \mathbf{x}'}{\partial \mathbf{y}}|_+$, $|A|_+$ 表示 A 行列式值的绝对值.

Definition

设 $X \in R^{m\times n}$ 是矩阵变量，$Y = F(X) \in R^{m\times n}$ 是一一变换，$F(X)$ 可微，则变换 $Y = F(X)$ 的雅可比行列式定义为

$$J(Y \to X) = J(Y : X) = |\frac{\partial Y}{\partial X}|_+$$

(1) 当矩阵变量为一般矩阵的情况

- 若 $Y = AXB$, 且 A 为 m 阶非退化方阵, B 为 n 阶非退化方

阵, 则有

$$J(Y \to X) = |(det(A))^n (det(B))^m| = |A|_+^n |B|_+^m$$

- 若 X 为 n 阶非退化方阵, $Y = X^{-1}$, 则 $J(Y \to X) = |X|^{-2n}$.

(2) 当矩阵变量为三角矩阵的情况

- 设 G 为给定的 n 阶非退化下三角矩阵, X 为 n 阶下三角矩阵, 令 $Y = GX$, 则有

$$J(Y \to X) = |\prod_{i=1}^{n} g_{ii}^{i}|$$

其中 $g_{11}, g_{22}, \ldots, g_{nn}$ 为 G 的主对角元.

- 设 G 为给定的 n 阶非退化下三角矩阵, X 为 n 阶下三角矩阵, 令 $Y = XG$, 则有

$$J(Y \to X) = |\prod_{i=1}^{n} g_{ii}^{n-i+1}|$$

- 设 X 为上三角矩阵变量, 且其对角元素为正, 变换 $Y = XX'$, 则有

$$J(Y \to X) = 2^n |\prod_{i=1}^{n} x_{ii}^{n-i+1}|$$

注意: 若要求 X 的对角线元素非负, 则 $Y = XX'$ 是一一变换.

(3) 当矩阵变量为对称矩阵的情况

- 设 $X' = X$, G 是非退化的上 (下) 三角矩阵, 令 $Y = G'XG$, 则有

$$J(Y \to X) = |G|_+^{n+1}$$

- 设 $X' = X$, P 是非退化矩阵, 令 $Y = P'XP$, 则有

$$J(Y \to X) = |P|_+^{n+1}$$

补充作业:

- 设 A 和 B 为两个 $n \times p$ 矩阵, 且 $\mathcal{L}(A) \subset \mathcal{L}(B)$. $P_A = A(A'A)^- A'$ 和 $P_B = B(B'B)^{-1}B'$ 为 R^n 到 $\mathcal{L}(A) \subset \mathcal{L}(B)$ 上的投影阵, 证明 $P_A P_B = P_B P_A = P_A$.