高等工程數學（5）

厦門大學
XIAMEN UNIVERSITY
信息学院
（国家示范性软件学院）
School of Informatics
黄 煒
博士·副教授
Dr. Wei Huang

矩阵论（5）

厦門大學
XIAMEN UNIVERSITY

信息学院
（国家示范性软件学院）
School of Informatics

黄 炜
博士·副教授
Dr. Wei Huang

1.5 特征值与特征向量

特征值与特征向量的概念刻划了方阵的一些本质特征. 这些概念不仅在理论上具有重要意义，而且在众多领域，诸如几何学、力学、控制理论及数量经济学等都有着广泛的应用. 除另有说明外，本节在复数域上讨论问题.

定义 1.5-1 设 $\boldsymbol{A}=[a_{ij}]$是 n 阶方阵，若有数 λ 和非零向量 $\boldsymbol{x}$，使

$$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x} \tag{1.5-1}$$

成立，则称 λ 为 $\boldsymbol{A}$ 的**特征值**，非零向量 $\boldsymbol{x}$ 为 A 的属于(或对应于) 特征值 λ 的**特征向量**.

(1.5-1)式可以写成

$$(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}. \tag{1.5-2}$$

这是关于 $\boldsymbol{x}$ 的齐次线性方程组，它有非零解的充要条件是系数行列式等于零，即

$$\det(\boldsymbol{A}-\lambda\boldsymbol{I})=0, \tag{1.5-3}$$

亦即

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix}=0. \tag{1.5-3'}$$

这是以 λ 为未知数的 n 次代数方程，称为方阵 $\boldsymbol{A}$ 的**特征方程**，其左端 $\det(\boldsymbol{A}-\lambda\boldsymbol{I})$ 是关于 λ 的 n 次多项式，称为方阵 $\boldsymbol{A}$ 的**特征多项式**.

1.5 特征值与特征向量

显然，$\boldsymbol{A}$ 的特征值是特征方程(1.5-3)的解(根)，$\boldsymbol{A}$ 的属于特征值 λ_j 的特征向量就是齐次线性方程组

$$(\boldsymbol{A}-\lambda_j\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0} \tag{1.5-4}$$

的非零解. 我们称方程组(1.5-4)的解空间 $\mathcal{N}(\boldsymbol{A}-\lambda_j\boldsymbol{I})$ 为 $\boldsymbol{A}$ 的关于特征值 λ_j 的**特征子空间**，从而，此特征子空间中除零向量外，其余向量全都是 $\boldsymbol{A}$ 的属于特征值 λ_j 的特征向量.

定义 1.5-2　齐次线性方程组(1.5-4)的解空间的维数称为特征值 λ_j 的**几何重数**.

由于齐次线性方程组 $(\boldsymbol{A}-\lambda_j\boldsymbol{I})\boldsymbol{X}=\boldsymbol{0}$ 的基础解系中解向量的个数等于 n 减去方阵 $\boldsymbol{A}-\lambda_j\boldsymbol{I}$ 的秩，所以 λ_j 的几何重数等于 $n-r(\boldsymbol{A}-\lambda_j\boldsymbol{I})$，即

$$\dim\mathcal{N}(\boldsymbol{A}-\lambda_j\boldsymbol{I})=n-r(\boldsymbol{A}-\lambda_j\boldsymbol{I}). \tag{1.5-5}$$

1.5 特征值与特征向量

例 1 设
$$A=\begin{bmatrix}-1 & 1 & 1\\ 1 & -1 & 1\\ 1 & 1 & -1\end{bmatrix},$$
求 A 的特征值、特征向量和特征子空间.

解 A 的特征多项式为

$$\det(A-\lambda I)=\begin{vmatrix}-1-\lambda & 1 & 1\\ 1 & -1-\lambda & 1\\ 1 & 1 & -1-\lambda\end{vmatrix}=\begin{vmatrix}1-\lambda & 1-\lambda & 1-\lambda\\ 1 & -1-\lambda & 1\\ 1 & 1 & -1-\lambda\end{vmatrix}$$

$$=(1-\lambda)\begin{vmatrix}1 & 1 & 1\\ 1 & -1-\lambda & 1\\ 1 & 1 & -1-\lambda\end{vmatrix}=(1-\lambda)\begin{vmatrix}1 & 1 & 1\\ 0 & -2-\lambda & 0\\ 0 & 0 & -2-\lambda\end{vmatrix}$$

$$=(1-\lambda)(2+\lambda)^2,$$

故 A 的特征值是 $\lambda_1=1,\lambda_2=\lambda_3=-2$.

1.5 特征值与特征向量

对于 $\lambda_1=1$，解齐次线性方程组

$$(\boldsymbol{A}-1\cdot\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0},$$

由
$$\boldsymbol{A}-1\cdot\boldsymbol{I}=\begin{vmatrix}-2 & 1 & 1\\ 1 & -2 & 1\\ 1 & 1 & -2\end{vmatrix}\rightarrow\begin{vmatrix}1 & 1 & -2\\ 0 & 1 & -1\\ 0 & 0 & 0\end{vmatrix}\rightarrow\begin{vmatrix}1 & 0 & -1\\ 0 & 1 & -1\\ 0 & 0 & 0\end{vmatrix}.$$

得其一个基础解系 $\boldsymbol{\alpha}_1=[1,1,1]^{\mathrm{T}}$，从而 $\boldsymbol{A}$ 的属于特征值 $\lambda_1=1$ 的所有特征向量为 $k_1[1,1,1]^{\mathrm{T}}$，其中 k_1 为任意非零实数.

对于 $\lambda_2=\lambda_3=-2$，解齐次线性方程组

$$(\boldsymbol{A}-(-2)\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0},$$

由
$$\boldsymbol{A}-(-2)\boldsymbol{I}=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}\rightarrow\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

得其一个基础解系

$$\boldsymbol{\alpha}_2=[-1,1,0]^{\mathrm{T}},\quad \boldsymbol{\alpha}_3=[-1,0,1]^{\mathrm{T}},$$

1.5 特征值与特征向量

于是 $\boldsymbol{A}$ 的属于特征值 $\lambda_2=\lambda_3=-2$ 的所有特征向量是 $k_2[-1,1,0]^{\mathrm{T}}+k_3[-1,0,1]^{\mathrm{T}}$，其中 k_2、k_3 为不同时为零的任意实数.

$\boldsymbol{A}$ 的关于特征值 1 的特征子空间是

$$\{\boldsymbol{x} \mid \boldsymbol{x}=c[1,1,1]^{\mathrm{T}}, c \in \mathbf{R}\}.$$

$\boldsymbol{A}$ 的关于特征值 -2 的特征子空间是

$$\{\boldsymbol{x} \mid \boldsymbol{x}=c_1[-1,1,0]^{\mathrm{T}}+c_2[-1,0,1]^{\mathrm{T}}, c_1, c_2 \in \mathbf{R}\}. \qquad \blacksquare$$

1.5.1 特征值与特征向量的性质

由代数基本定理知，特征方程(1.5-3)在复数域中必有解，且解的个数等于方程的次数(重根按重数计)，因而 n 阶方阵 $\boldsymbol{A}$ 有 n 个特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$(有重根时重复出现)，于是

$$\det(\boldsymbol{A}-\lambda\boldsymbol{I})=(\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda). \tag{1.5-6}$$

定义 1.5-3 若 λ_i 是方阵 $\boldsymbol{A}$ 的特征方程(1.5-3)的 k 重根，则称特征值 λ_i 的代数重数是 k.

由于 $$(\boldsymbol{A}-\lambda\boldsymbol{I})^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}-\lambda\boldsymbol{I},\ (\boldsymbol{A}-\bar{\lambda}\boldsymbol{I})^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}-\lambda\boldsymbol{I},$$

故由行列式的性质得 $$\det(\boldsymbol{A}^{\mathrm{T}}-\lambda\boldsymbol{I})=\det(\boldsymbol{A}-\lambda\boldsymbol{I}),$$

$$\det(\boldsymbol{A}^{\mathrm{H}}-\lambda\boldsymbol{I})=\overline{\det(\boldsymbol{A}-\bar{\lambda}\boldsymbol{I})}.$$

这就是说，$\boldsymbol{A}^{\mathrm{T}}$ 与 $\boldsymbol{A}$ 有相同的特征多项式，故它们有相同的特征值；$\boldsymbol{A}^{\mathrm{H}}$ 的特征多项式与 $\boldsymbol{A}$ 的特征多项式之差别仅在于对应的各项系数互为共轭复数，故 $\boldsymbol{A}^{\mathrm{H}}$ 的特征值是 $\boldsymbol{A}$ 的特征值的共轭复数.

1.5.1 特征值与特征向量的性质

记
$$\boldsymbol{B}=[\boldsymbol{A}\ \boldsymbol{b}]=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix},$$
称其为此线性方程组的**增广矩阵**.

若 $\boldsymbol{b}=\boldsymbol{0}$,则称此线性方程组为**齐次线性方程组**;否则,称为**非齐次线性方程组**.

解线性方程组的基本方法是通过对增广矩阵(系数矩阵)施行行初等变换来消元,以得到等价的易于求解的线性方程组.

1.5.1 特征值与特征向量的性质

定理 1.5-1 设 $\boldsymbol{A}$ 是 n 阶方阵，则

(1) $\boldsymbol{A}$ 的 n 个特征值之和等于 $\boldsymbol{A}$ 的 n 个主对角线元素之和，即

$$\lambda_1+\lambda_2+\cdots+\lambda_n=a_{11}+a_{22}+\cdots+a_{nn}; \tag{1.5-7}$$

(2) $\boldsymbol{A}$ 的 n 个特征值之积等于 $\boldsymbol{A}$ 的行列式之值，即

$$\lambda_1\cdot\lambda_2\cdots\lambda_n=|\boldsymbol{A}|. \tag{1.5-8}$$

证 利用 n 阶行列式定义将 $\det(\boldsymbol{A}-\lambda\boldsymbol{I})$ 展开，其中有一项是主对角线元素乘积

$$(a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda),$$

而其余各项至多含 $n-2$ 个主对角线元素，这些项中 λ 的次数最高是 $n-2$. 因此在(1.5-6)式中比较两边对应的 λ^{n-1} 项和常数项之系数，即知(1.5-7)式和(1.5-8)式成立. ∎

称方阵 $\boldsymbol{A}$ 的主对角线元素之和 $\sum_{i=1}^{n}a_{ii}$ 为 $\boldsymbol{A}$ 的**迹**，记为 $\mathrm{tr}\boldsymbol{A}$. 据(1.5-7)式，有

$$\mathrm{tr}\boldsymbol{A}=\sum_{i=1}^{n}\lambda_i.$$

1.5.1 特征值与特征向量的性质

例 2 设 λ 是可逆方阵 $\boldsymbol{A}$ 的特征值，证明 $\lambda\neq 0$，且 λ^{-1} 是 $\boldsymbol{A}^{-1}$ 的特征值.

证 因 $\boldsymbol{A}$ 可逆，故 $|\boldsymbol{A}|\neq 0$，所以 $\lambda\neq 0$.

设 $\boldsymbol{x}\neq\boldsymbol{0}$ 是 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量，则有 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$. 于是，由 $\lambda\neq 0$ 得

$$\boldsymbol{x}=\frac{1}{\lambda}\boldsymbol{A}\boldsymbol{x}.$$

用 $\boldsymbol{A}^{-1}$ 左乘上式的两边，得 $$\boldsymbol{A}^{-1}\boldsymbol{x}=\lambda^{-1}\boldsymbol{x}.$$

由 $\boldsymbol{x}\neq\boldsymbol{0}$ 知 λ^{-1} 是 $\boldsymbol{A}^{-1}$ 的特征值. ∎

1.5.1 特征值与特征向量的性质

定理 1.5-2 设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s$ 分别是 $\boldsymbol{A}$ 的属于不同特征值 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 的特征向量，则 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_s$ 线性无关.

证 对 s 用数学归纳法.

$s=1$ 时，由于 $\boldsymbol{x}_1\neq\boldsymbol{0}$，故定理成立. 现设当 $s=k$ 时定理成立. 当 $s=k+1$ 时，设有数 c_1，$c_2,\cdots,c_k,c_{k+1}$ 使

$$c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2+\cdots+c_k\boldsymbol{x}_k+c_{k+1}\boldsymbol{x}_{k+1}=\boldsymbol{0}. \tag{1.5-9}$$

用 $\boldsymbol{A}$ 左乘上式两边得

$$c_1\lambda_1\boldsymbol{x}_1+c_2\lambda_2\boldsymbol{x}_2+\cdots+c_k\lambda_k\boldsymbol{x}_k+c_{k+1}\lambda_{k+1}\boldsymbol{x}_{k+1}=\boldsymbol{0}. \tag{1.5-10}$$

而用 λ_{k+1} 乘(1.5-9)式两边得

$$c_1\lambda_{k+1}\boldsymbol{x}_1+c_2\lambda_{k+1}\boldsymbol{x}_2+\cdots+c_k\lambda_{k+1}\boldsymbol{x}_k+c_{k+1}\lambda_{k+1}\boldsymbol{x}_{k+1}=\boldsymbol{0}. \tag{1.5-11}$$

(1.5-10)式减去(1.5-11)式，得

$$c_1(\lambda_1-\lambda_{k+1})\boldsymbol{x}_1+c_2(\lambda_2-\lambda_{k+1})\boldsymbol{x}_2+\cdots+c_k(\lambda_k-\lambda_{k+1})\boldsymbol{x}_k=\boldsymbol{0}.$$

由归纳假设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_k$ 线性无关，又 $\lambda_1,\lambda_2,\cdots,\lambda_k,\lambda_{k+1}$ 互不相同，故有 $c_1=c_2=\cdots=c_k=0$. 从而由(1.5-9)式和 $\boldsymbol{x}_{k+1}\neq\boldsymbol{0}$ 知 $c_{k+1}=0$. 于是 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{k+1}$ 线性无关. ■

设有关于 λ 的一个 m 次多项式

$$g(\lambda)=a_0\lambda^m+a_1\lambda^{m-1}+\cdots+a_m,\quad a_0\neq 0,$$

$\boldsymbol{A}$ 是 n 阶方阵，规定 $\quad g(\boldsymbol{A})=a_0\boldsymbol{A}^m+a_1\boldsymbol{A}^{m-1}+\cdots+a_m\boldsymbol{I}$，

称它为 $\boldsymbol{A}$ 的多项式. $g(\boldsymbol{A})$ 是用 $\boldsymbol{A}$ 替换多项式 $g(\lambda)$ 中的 λ，并用单位矩阵 $\boldsymbol{I}$ 替换 $\lambda^0=1$ 的结果，显然 $g(\boldsymbol{A})$ 是一个 n 阶方阵.

1.5.1 特征值与特征向量的性质

定理 1.5-3 设 n 阶方阵 $\boldsymbol{A}$ 的 n 个特征值是 $\lambda_1,\lambda_2,\cdots,\lambda_n$，对应于这些特征值的特征向量依次为 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$，而 $g(\lambda)$ 是一多项式，那么 n 阶方阵 $g(\boldsymbol{A})$ 的 n 个特征值是 $g(\lambda_1),g(\lambda_2),\cdots,g(\lambda_n)$，且 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 分别是 $g(\boldsymbol{A})$ 的属于这些特征值的特征向量.

证 不妨设

$$g(\lambda)=a_0\lambda^m+a_1\lambda^{m-1}+\cdots+a_m,$$

由于

$$\boldsymbol{A}\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i,$$

$$\boldsymbol{A}^2\boldsymbol{x}_i=\boldsymbol{A}(\boldsymbol{A}\boldsymbol{x}_i)=\boldsymbol{A}(\lambda_i\boldsymbol{x}_i)=\lambda_i\boldsymbol{A}\boldsymbol{x}_i=\lambda_i^2\boldsymbol{x}_i,$$

$$\cdots\cdots$$

$$\boldsymbol{A}^m\boldsymbol{x}_i=\boldsymbol{A}(\boldsymbol{A}^{m-1}\boldsymbol{x}_i)=\boldsymbol{A}(\lambda_i^{m-1}\boldsymbol{x}_i)=\lambda_i^{m-1}\boldsymbol{A}\boldsymbol{x}_i=\lambda_i^m\boldsymbol{x}_i,$$

所以

$$\begin{aligned}g(\boldsymbol{A})\boldsymbol{x}_i&=(a_0\boldsymbol{A}^m+a_1\boldsymbol{A}^{m-1}+\cdots+a_m\boldsymbol{I})\boldsymbol{x}_i\\&=a_0\boldsymbol{A}^m\boldsymbol{x}_i+a_1\boldsymbol{A}^{m-1}\boldsymbol{x}_i+\cdots+a_m\boldsymbol{x}_i\\&=(a_0\lambda_i^m+a_1\lambda_i^{m-1}+\cdots+a_m)\boldsymbol{x}_i=g(\lambda_i)\boldsymbol{x}_i.\end{aligned}$$

因 $\boldsymbol{x}_i\neq\boldsymbol{0}$，故 $g(\lambda_i)$ 是方阵 $g(\boldsymbol{A})$ 的特征值，$\boldsymbol{x}_i$ 是 $g(\boldsymbol{A})$ 的属于特征值 $g(\lambda_i)$ 的特征向量. ∎

方阵 $\boldsymbol{A}\neq\boldsymbol{O}$，但存在正整数 $k(\geqslant 2)$ 使 $\boldsymbol{A}^k=\boldsymbol{O}$，则称 $\boldsymbol{A}$ 为**幂零矩阵**. 幂零矩阵的特征值必是零. 事实上，若 λ_0 是幂零矩阵 $\boldsymbol{A}$ 的特征值，$\boldsymbol{x}_0\neq\boldsymbol{0}$ 是 $\boldsymbol{A}$ 的属于 λ_0 的特征向量，则由上述定理知

$$\boldsymbol{0}=\boldsymbol{A}^k\boldsymbol{x}_0=\lambda_0^k\boldsymbol{x}_0,$$

1.5.1 特征值与特征向量的性质

从而由 $\boldsymbol{x}_0 \neq \mathbf{0}$ 得 $\lambda_0^k = 0$，即 $\lambda_0 = 0$.

例如，1.2 节例 3 中的方阵

$$\boldsymbol{U} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

是幂零矩阵，其特征多项式为 $(-\lambda)^4$，所有的特征值全为 0. 又如 $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$ 也是幂零矩阵，因为 $\boldsymbol{A}^2 = \boldsymbol{O}$，它的特征多项式是 $(-\lambda)^2$，特征值全都是 0.

1.5.1 特征值与特征向量的性质

定理 1.5-4(Sylverster)　设 $\boldsymbol{A}$ 是 $m\times n$ 矩阵，$\boldsymbol{B}$ 是 $n\times m$ 矩阵，且 $m\geqslant n$，记方阵 $\boldsymbol{AB}$、$\boldsymbol{BA}$ 的特征多项式分别为 $f_{\boldsymbol{AB}}(\lambda)$，$f_{\boldsymbol{BA}}(\lambda)$，则有

$$f_{\boldsymbol{AB}}(\lambda)=(-\lambda)^{m-n}f_{\boldsymbol{BA}}(\lambda). \tag{1.5-12}$$

证　由分块矩阵乘法得

$$\left[\begin{array}{c:c}\boldsymbol{I}_m & \boldsymbol{A}\\ \hdashline \boldsymbol{O} & \lambda\boldsymbol{I}_n\end{array}\right]\left[\begin{array}{c:c}-\lambda\boldsymbol{I}_m & \boldsymbol{A}\\ \hdashline \boldsymbol{B} & -\boldsymbol{I}_n\end{array}\right]=\left[\begin{array}{c:c}\boldsymbol{AB}-\lambda\boldsymbol{I}_m & \boldsymbol{O}\\ \hdashline \lambda\boldsymbol{B} & -\lambda\boldsymbol{I}_n\end{array}\right], \tag{1.5-13}$$

$$\left[\begin{array}{c:c}-\lambda\boldsymbol{I}_m & \boldsymbol{A}\\ \hdashline \boldsymbol{B} & -\boldsymbol{I}_n\end{array}\right]\left[\begin{array}{c:c}\boldsymbol{I}_m & \boldsymbol{A}\\ \hdashline \boldsymbol{O} & \lambda\boldsymbol{I}_n\end{array}\right]=\left[\begin{array}{c:c}-\lambda\boldsymbol{I}_m & \boldsymbol{O}\\ \hdashline \boldsymbol{B} & \boldsymbol{BA}-\lambda\boldsymbol{I}_n\end{array}\right]. \tag{1.5-14}$$

对(1.5-13)式、(1.5-14)式的两边取行列式，则由行列式的乘法公式和行列式的性质，分别得

$$\lambda^n\det\left[\begin{array}{c:c}-\lambda\boldsymbol{I}_m & \boldsymbol{A}\\ \hdashline \boldsymbol{B} & -\boldsymbol{I}_n\end{array}\right]=(-\lambda)^n\det(\boldsymbol{AB}-\lambda\boldsymbol{I}_m),$$

$$\lambda^n\det\left[\begin{array}{c:c}-\lambda\boldsymbol{I}_m & \boldsymbol{A}\\ \hdashline \boldsymbol{B} & -\boldsymbol{I}_n\end{array}\right]=(-\lambda)^m\det(\boldsymbol{BA}-\lambda\boldsymbol{I}_n).$$

因此，对任意的 λ 都有

$$(-\lambda)^n\det(\boldsymbol{AB}-\lambda\boldsymbol{I}_m)=(-\lambda)^m\det(\boldsymbol{BA}-\lambda\boldsymbol{I}_n),$$

从而

$$\det(\boldsymbol{AB}-\lambda\boldsymbol{I}_m)=(-\lambda)^{m-n}\det(\boldsymbol{BA}-\lambda\boldsymbol{I}_n),$$

即(1.5-12)式成立. ∎

(1.5-12)式也称为特征多项式的**降秩计算公式**，它对于计算小秩方阵的特征多项式是很有用的.

1.5.1 特征值与特征向量的性质

例 3 设 n 阶方阵 $\boldsymbol{A}$ 的秩是 1，求 $\boldsymbol{A}$ 的特征多项式.

解 因 $r(\boldsymbol{A})=1$，故由 1.3 定理 1.3-2 知，存在 $\boldsymbol{A}$ 的一个满秩分解，不妨设为

$$\boldsymbol{A}=\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}=[a_1,a_2,\cdots,a_n]^{\mathrm{T}}[b_1\quad b_2\quad \cdots\quad b_n].$$

于是，由(1.5-12)式得

$$\begin{aligned}|\boldsymbol{A}-\lambda\boldsymbol{I}_n| &= \det([a_1,a_2,\cdots,a_n]^{\mathrm{T}}[b_1\quad \cdots\quad b_n]-\lambda\boldsymbol{I}_n)\\ &=\lambda^{n-1}\det(\lambda-[b_1\quad \cdots\quad b_n][a_1,a_2,\cdots,a_n]^{\mathrm{T}})\\ &=\lambda^{n-1}\left(\lambda-\sum_{i=1}^{n}a_ib_i\right).\end{aligned}$$

1.5.2 方阵的相似变换和相似对角化

定义 1.5-4 设 $\boldsymbol{A}$、$\boldsymbol{B}$ 都是 n 阶方阵，若存在可逆矩阵 $\boldsymbol{P}$，使

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{B},$$

则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ **相似**，记为 $\boldsymbol{A}\sim\boldsymbol{B}$，并称从 $\boldsymbol{A}$ 到 $\boldsymbol{B}$ 的这种变换为相似变换，称可逆矩阵 $\boldsymbol{P}$ 为**相似变换矩阵.**

若 $\boldsymbol{A}$ 与一个对角矩阵相似，则称 $\boldsymbol{A}$ 可相似对角化.

定理 1.5-5 若方阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似，则它们有相同的特征多项式，从而有相同的特征值.

证 因 $\boldsymbol{A}\sim\boldsymbol{B}$，故存在可逆矩阵 $\boldsymbol{P}$ 使

$$\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}.$$

于是，由行列式的性质得

$$\begin{aligned}|\boldsymbol{B}-\lambda\boldsymbol{I}|&=|\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}-\lambda\boldsymbol{I}|=|\boldsymbol{P}^{-1}(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{P}|\\&=|\boldsymbol{P}^{-1}|\cdot|\boldsymbol{A}-\lambda\boldsymbol{I}|\cdot|\boldsymbol{P}|=|\boldsymbol{A}-\lambda\boldsymbol{I}|.\end{aligned}$$ ∎

作为这个定理的推论，我们有下述命题：

若 $\boldsymbol{A}\sim\boldsymbol{B}$，则 $\mathrm{tr}\boldsymbol{A}=\mathrm{tr}\boldsymbol{B}$，$|\boldsymbol{A}|=|\boldsymbol{B}|$.

值得指出的是，两个同阶方阵的特征值相同，并不能推出它们是相似的.

方阵之间的相似具有下列性质：

(1) 自反性 $\boldsymbol{A}\sim\boldsymbol{A}$；

(2) 对称性 若 $\boldsymbol{A}\sim\boldsymbol{B}$，则 $\boldsymbol{B}\sim\boldsymbol{A}$；

(3) 传递性 若 $\boldsymbol{A}\sim\boldsymbol{B}$，$\boldsymbol{B}\sim\boldsymbol{C}$，则 $\boldsymbol{A}\sim\boldsymbol{C}$.

1.5.2 方阵的相似变换和相似对角化

例 4 证明：若 n 阶方阵与对角矩阵 $\boldsymbol{D}=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}$ 相似，则 $\lambda_1,\cdots,\lambda_n$ 是 $\boldsymbol{A}$ 的 n 个特征值.

证 因 $\boldsymbol{A}\sim\boldsymbol{D}$，故

$$|\boldsymbol{A}-\lambda\boldsymbol{I}|=|\boldsymbol{D}-\lambda\boldsymbol{I}|=\begin{vmatrix}\lambda_1-\lambda & & \\ & \ddots & \\ & & \lambda_n-\lambda\end{vmatrix}=\prod_{i=1}^{n}(\lambda_i-\lambda).$$

因而 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 $\boldsymbol{A}$ 的 n 个特征值. ∎

在矩阵的运算和应用中，常常要对方阵 $\boldsymbol{A}$ 作相似变换，使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}$ 为尽可能简单的方阵. 显然最简单的方阵是对角矩阵. 因此就要研究方阵可相似对角化的条件及其方法.

定理 1.5-6 n 阶方阵 $\boldsymbol{A}$ 与对角矩阵相似的充要条件是 $\boldsymbol{A}$ 有 n 个线性无关的特征向量.

证 **必要性**：设 $\boldsymbol{A}$ 与对角矩阵 $\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}$ 相似，则存在可逆矩阵 $\boldsymbol{P}$ 使

$$\boldsymbol{P}^{-1}\boldsymbol{AP}=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}. \tag{1.5-15}$$

1.5.2 方阵的相似变换和相似对角化

于是

$$\boldsymbol{AP}=\boldsymbol{P}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}. \tag{1.5-16}$$

对 $\boldsymbol{P}$ 分块，用 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 表示 $\boldsymbol{P}$ 的 n 个列向量，则有 $\boldsymbol{P}=[\boldsymbol{x}_1 \vdots \cdots \vdots \boldsymbol{x}_n]$. 将其代入上式的左端和右端，分别得到

$$\boldsymbol{AP}=\boldsymbol{A}[\boldsymbol{x}_1 \vdots \cdots \vdots \boldsymbol{x}_n]=[\boldsymbol{Ax}_1 \vdots \cdots \vdots \boldsymbol{Ax}_n],$$

$$\boldsymbol{P}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}=[\boldsymbol{x}_1 \vdots \cdots \vdots \boldsymbol{x}_n]\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}=[\lambda_1\boldsymbol{x}_1 \vdots \cdots \vdots \lambda_n\boldsymbol{x}_n].$$

因而有

$$\boldsymbol{Ax}_i=\lambda_i\boldsymbol{x}_i,\quad i=1,2,\cdots,n.$$

由于 $\boldsymbol{P}$ 可逆，故 $\boldsymbol{x}_i\neq\boldsymbol{0}$，且 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 线性无关，所以 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是 $\boldsymbol{A}$ 的 n 个线性无关的特征向量，λ_i 是与 $\boldsymbol{x}_i$ 对应的特征值.

充分性：设 $\boldsymbol{A}$ 有 n 个线性无关的特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$，对应的特征值依次为 $\lambda_1,\lambda_2,\cdots,\lambda_n$，则令 $\boldsymbol{P}=[\boldsymbol{x}_1 \vdots \cdots \vdots \boldsymbol{x}_n]$，由上述必要性的证明过程逆推上去即得(1.5-16)式. 因 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 线性无关，故 $\boldsymbol{P}$ 可逆，从而(1.5-15)式成立. ∎

1.5.2 方阵的相似变换和相似对角化

从定理 1.5-6 的证明可知，当 n 阶方阵 $\boldsymbol{A}$ 可相似对角化时，相似变换矩阵 $\boldsymbol{P}$ 可由 $\boldsymbol{A}$ 的 n 个线性无关特征向量作为其列向量构成.

由定理 1.5-2 和 1.5-6 可得下列推论：

推论 1 若 n 阶方阵 $\boldsymbol{A}$ 有 n 个互不相同的特征值，则 $\boldsymbol{A}$ 必可相似对角化；

推论 2 若 n 阶方阵 $\boldsymbol{A}$ 的每一个 k_i 重特征值 λ_i 都有 k_i 个线性无关的特征向量，即对 $\boldsymbol{A}$ 的每个特征值，其代数重数与几何重数都相等，则 $\boldsymbol{A}$ 必可相似对角化.

例 5 证明 $\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0\\ 1 & 2 & -1\\ -1 & -1 & 2\end{bmatrix}$ 可相似对角化，并求可逆矩阵 $\boldsymbol{P}$ 使 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ 为对角矩阵.

解 由于

$$|\boldsymbol{A}-\lambda\boldsymbol{I}|=\begin{vmatrix}1-\lambda & 0 & 0\\ 1 & 2-\lambda & -1\\ -1 & -1 & 2-\lambda\end{vmatrix}=(1-\lambda)\begin{vmatrix}2-\lambda & -1\\ -1 & 2-\lambda\end{vmatrix}=(3-\lambda)(1-\lambda)^2,$$

所以 $\boldsymbol{A}$ 的特征值是 $\lambda_1=3,\lambda_2=\lambda_3=1$.

1.5.2 方阵的相似变换和相似对角化

从定理 1.5-6 的证明可知，当 n 阶方阵 $\boldsymbol{A}$ 可相似对角化时，相似变换矩阵 $\boldsymbol{P}$ 可由 $\boldsymbol{A}$ 的 n 个线性无关特征向量作为其列向量构成.

由定理 1.5-2 和 1.5-6 可得下列推论：

推论 1 若 n 阶方阵 $\boldsymbol{A}$ 有 n 个互不相同的特征值，则 $\boldsymbol{A}$ 必可相似对角化；

推论 2 若 n 阶方阵 $\boldsymbol{A}$ 的每一个 k_i 重特征值 λ_i 都有 k_i 个线性无关的特征向量，即对 $\boldsymbol{A}$ 的每个特征值，其代数重数与几何重数都相等，则 $\boldsymbol{A}$ 必可相似对角化.

例 5 证明 $\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0\\ 1 & 2 & -1\\ -1 & -1 & 2\end{bmatrix}$ 可相似对角化，并求可逆矩阵 $\boldsymbol{P}$ 使 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ 为对角矩阵.

解 由于

$$|\boldsymbol{A}-\lambda\boldsymbol{I}|=\begin{vmatrix}1-\lambda & 0 & 0\\ 1 & 2-\lambda & -1\\ -1 & -1 & 2-\lambda\end{vmatrix}=(1-\lambda)\begin{vmatrix}2-\lambda & -1\\ -1 & 2-\lambda\end{vmatrix}=(3-\lambda)(1-\lambda)^2,$$

所以 $\boldsymbol{A}$ 的特征值是 $\lambda_1=3,\lambda_2=\lambda_3=1$.

1.5.2 方阵的相似变换和相似对角化

因 $$r(\boldsymbol{A}-1\cdot\boldsymbol{I})=r\left(\begin{bmatrix}0&0&0\\1&1&-1\\-1&-1&1\end{bmatrix}\right)=1,$$

故 $\boldsymbol{A}$ 的特征值 1 的几何重数是 $3-1=2$，等于其代数重数 2. 于是 $\boldsymbol{A}$ 的每个特征值的代数重数都等于其几何重数，故 $\boldsymbol{A}$ 可相似对角化.

对于 $\lambda_1=3$，解 $(\boldsymbol{A}-3\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}$，即解

$$\begin{bmatrix}-2&0&0\\1&-1&-1\\-1&-1&-1\end{bmatrix}\boldsymbol{x}=\boldsymbol{0},$$

得其一个基础解系 $\boldsymbol{x}_1=[0,1,-1]^{\mathrm{T}}$.

对于 $\lambda_2=\lambda_3=1$，解 $(\boldsymbol{A}-1\cdot\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}$，即解

$$\begin{bmatrix}0&0&0\\1&1&-1\\-1&-1&1\end{bmatrix}\boldsymbol{x}=\boldsymbol{0},$$

1.5.2 方阵的相似变换和相似对角化

得到它的一个基础解系

$$x_2 = [-1,1,0]^T,\quad x_3 = [1,0,1]^T.$$

令
$$P = [x_1 \quad x_2 \quad x_3] = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix},$$

则有
$$AP = P\begin{bmatrix} 3 & & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

因 x_1, x_2, x_3 线性无关，故 P 可逆。事实上

$$P^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

于是
$$P^{-1}AP = \begin{bmatrix} 3 & & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

∎

1.5.2 方阵的相似变换和相似对角化

例 6 已知 3 阶方阵 $\boldsymbol{A}$ 的特征值是 1，−1，−1，且 $\boldsymbol{x}_1=[1,0,1]^{\mathrm{T}}$，$\boldsymbol{x}_2=[0,1,-1]^{\mathrm{T}}$，$\boldsymbol{x}_3=[2,2,-1]^{\mathrm{T}}$ 依次为对应的特征向量，求 $\boldsymbol{A}$ 和 $\boldsymbol{A}^8$。

解 由于 $\boldsymbol{A}\boldsymbol{x}_1=1\cdot\boldsymbol{x}_1$，$\boldsymbol{A}\boldsymbol{x}_2=(-1)\cdot\boldsymbol{x}_2$，$\boldsymbol{A}\boldsymbol{x}_3=(-1)\cdot\boldsymbol{x}_3$，

所以若令
$$\boldsymbol{P}=[\boldsymbol{x}_1\quad\boldsymbol{x}_2\quad\boldsymbol{x}_3]=\begin{bmatrix}1&0&2\\0&1&2\\1&-1&-1\end{bmatrix},$$

则有
$$\boldsymbol{AP}=[\boldsymbol{A}\boldsymbol{x}_1\quad\boldsymbol{A}\boldsymbol{x}_2\quad\boldsymbol{A}\boldsymbol{x}_3]=[1\cdot\boldsymbol{x}_1\quad(-1)\cdot\boldsymbol{x}_2\quad(-1)\cdot\boldsymbol{x}_3]$$
$$=[\boldsymbol{x}_1\quad\boldsymbol{x}_2\quad\boldsymbol{x}_3]\begin{bmatrix}1&&\\&-1&\\&&-1\end{bmatrix}=\boldsymbol{P}\begin{bmatrix}1&&\\&-1&\\&&-1\end{bmatrix}.$$

又因 $|\boldsymbol{P}|=-1\neq0$，故 $\boldsymbol{P}$ 可逆，且
$$\boldsymbol{P}^{-1}=\begin{bmatrix}-1&2&2\\-2&3&2\\1&-1&-1\end{bmatrix}.$$

于是
$$\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}1&&\\&-1&\\&&-1\end{bmatrix}\boldsymbol{P}^{-1}=\begin{bmatrix}1&0&2\\0&1&2\\1&-1&-1\end{bmatrix}\begin{bmatrix}1&&\\&-1&\\&&-1\end{bmatrix}\begin{bmatrix}-1&2&2\\-2&3&2\\1&-1&-1\end{bmatrix}$$

1.5.2 方阵的相似变换和相似对角化

$$= \begin{bmatrix} -3 & 4 & 4 \\ 0 & -1 & 0 \\ -2 & 4 & 3 \end{bmatrix};$$

$$\boldsymbol{A}^8 = (\boldsymbol{PDP}^{-1})^8 = \boldsymbol{PDP}^{-1} \cdot \boldsymbol{PDP}^{-1} \cdots \boldsymbol{PDP}^{-1} = \boldsymbol{PD}^8\boldsymbol{P}^{-1}$$

$$= \boldsymbol{P} \begin{bmatrix} 1^8 & & \\ & (-1)^8 & \\ & & (-1)^8 \end{bmatrix} \boldsymbol{P}^{-1} = \boldsymbol{PIP}^{-1} = \boldsymbol{PP}^{-1} = \boldsymbol{I},$$

其中 $\boldsymbol{D} = \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix}$.

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

本段主要讨论 Hermite 矩阵的特征值与特征向量的特性及其相似对角化的问题，而把实对称矩阵看成 Hermite 矩阵在实矩阵下的特例，从而有相应的结论.

定义 1.5-5 设 $\boldsymbol{A}$ 是 n 阶复(实)方阵，如果

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{I}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}), \tag{1.5-17}$$

则称 $\boldsymbol{A}$ 为**酉(正交)矩阵.**

n 阶酉(正交)矩阵 $\boldsymbol{A}$ 具有如下性质：

(1) $\boldsymbol{A}$ 是可逆的，且 $\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{T}})$. 因此

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{I}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{I}); \tag{1.5-17$'$}$$

(2) $\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{T}})$也是酉(正交)矩阵，从而 $\boldsymbol{A}^{-1}$ 也是酉(正交)矩阵；

(3) 对任意 n 维复(实)向量 $\boldsymbol{x}$，$\boldsymbol{Ax}$ 的长度与 $\boldsymbol{x}$ 的长度相等，即

$$|\boldsymbol{Ax}|=|\boldsymbol{x}|;$$

(4) 对任意两个 n 维复(实)向量 $\boldsymbol{x},\boldsymbol{y}$，有

$$\langle\boldsymbol{Ax},\boldsymbol{Ay}\rangle=\langle\boldsymbol{x},\boldsymbol{y}\rangle,$$

即 $\boldsymbol{Ax}$ 与 $\boldsymbol{Ay}$ 的内积和 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积相同.

我们只给出性质(1)和(4)的证明.

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

证 (1) 由(1.5-17)式和行列式乘法公式得

$$|\boldsymbol{A}^{\mathrm{H}}|\cdot|\boldsymbol{A}|=1(|\boldsymbol{A}^{\mathrm{T}}|\cdot|\boldsymbol{A}|=1),$$

故$|\boldsymbol{A}|\neq 0$，从而$\boldsymbol{A}$可逆. 用$\boldsymbol{A}^{-1}$右乘(1.5-17)式两边，得

$$\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{T}}),$$

再用$\boldsymbol{A}$左乘此式两边便得(1.5-17′)式.

(4) $\langle \boldsymbol{Ax},\boldsymbol{Ay}\rangle=(\boldsymbol{Ax})^{\mathrm{T}}\overline{(\boldsymbol{Ay})}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\bar{\boldsymbol{A}}\bar{\boldsymbol{y}}=\boldsymbol{x}^{\mathrm{T}}\overline{(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}\bar{\boldsymbol{y}}=\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{I}}\bar{\boldsymbol{y}}=\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{y}}=\langle \boldsymbol{x},\boldsymbol{y}\rangle.$

例 7 证明酉(正交)矩阵的特征值的模等于1.

证 设$\boldsymbol{x}$是酉(正交)矩阵$\boldsymbol{A}$的属于特征值λ的特征向量，则$\boldsymbol{Ax}=\lambda\boldsymbol{x}$，且$\boldsymbol{x}\neq\boldsymbol{0}$. 于是，由酉(正交)矩阵的性质(4)得

$$\langle \boldsymbol{x},\boldsymbol{x}\rangle=\langle \boldsymbol{Ax},\boldsymbol{Ax}\rangle=\langle \lambda\boldsymbol{x},\lambda\boldsymbol{x}\rangle=\lambda\langle \boldsymbol{x},\lambda\boldsymbol{x}\rangle=\lambda\cdot\bar{\lambda}\langle \boldsymbol{x},\boldsymbol{x}\rangle;$$

因$\langle \boldsymbol{x},\boldsymbol{x}\rangle\neq 0$，故由上式得$\lambda\bar{\lambda}=|\lambda|^2=1$，即

$$|\lambda|=1. \qquad \blacksquare$$

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

定理 1.5-7 n 阶复(实)方阵 $\boldsymbol{A}=[\boldsymbol{A}_1 \vdots \cdots \vdots \boldsymbol{A}_n]$为酉(正交)矩阵的充要条件是

$$\langle \boldsymbol{A}_i, \boldsymbol{A}_j \rangle = \delta_{ij}, \quad i, j = 1, \cdots, n, \tag{1.5-18}$$

即 $\boldsymbol{A}$ 的列向量组 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 是 $\mathbf{C}^n(\mathbf{R}^n)$的一个标准正交基.

证

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1^{\mathrm{H}} \\ \hdashline \vdots \\ \hdashline \boldsymbol{A}_n^{\mathrm{H}} \end{bmatrix} [\boldsymbol{A}_1 \vdots \cdots \vdots \boldsymbol{A}_n] = \begin{bmatrix} \boldsymbol{A}_1^{\mathrm{H}}\boldsymbol{A}_1 & \cdots & \boldsymbol{A}_1^{\mathrm{H}}\boldsymbol{A}_n \\ \vdots & & \vdots \\ \boldsymbol{A}_n^{\mathrm{H}}\boldsymbol{A}_1 & \cdots & \boldsymbol{A}_n^{\mathrm{H}}\boldsymbol{A}_n \end{bmatrix}, \tag{1.5-19}$$

其中 $\boldsymbol{A}_j^{\mathrm{H}}\boldsymbol{A}_i = \overline{\boldsymbol{A}_j^{\mathrm{T}}\overline{\boldsymbol{A}}_i} = \overline{\langle \boldsymbol{A}_j, \boldsymbol{A}_i \rangle} = \langle \boldsymbol{A}_i, \boldsymbol{A}_j \rangle$.

因此，若 $\boldsymbol{A}$ 为酉矩阵，则 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{I}$，故由(1.5-19)式知(1.5-18)式成立. 反之，若(1.5-18)式成立，则由(1.5-19)式知 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{I}$，故 $\boldsymbol{A}$ 为酉矩阵. ■

值得指出的是，由于酉(正交)矩阵的性质(2)，所以 $\boldsymbol{A}$ 的行向量组也是 $\mathbf{C}^n(\mathbf{R}^n)$的标准正交基.

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

定理 1.5-8 Hermite 矩阵的特征值都是实数，且属于不同特征值的特征向量必正交.

证 设 A 是 Hermite 矩阵，λ 是它的任一特征值，$x\neq 0$是 A 的属于特征值λ 的特征向量，则

$$\langle Ax,x\rangle=\langle \lambda x,x\rangle=\lambda\langle x,x\rangle;$$

$$\langle A^{H}x,x\rangle=(A^{H}x)^{T}\bar{x}=x^{T}\bar{A}\bar{x}=x^{T}\overline{(Ax)}=\langle x,Ax\rangle=\langle x,\lambda x\rangle=\bar{\lambda}\langle x,x\rangle.$$

因 $A^{H}=A$，故由上述两式得 $\bar{\lambda}\langle x,x\rangle=\lambda\langle x,x\rangle$. 但$\langle x,x\rangle>0$，从而 $\bar{\lambda}=\lambda$，即 λ 是实数.

再则，设 λ、μ 是 Hermite 矩阵 A 的两个不同的特征值，x_1、x_2 是对应的特征向量，则有 $Ax_1=\lambda x_1$，$Ax_2=\mu x_2$，且 λ 和 μ 都是实数. 于是

$$\begin{aligned}\lambda\langle x_1,x_2\rangle&=\langle \lambda x_1,x_2\rangle=\langle Ax_1,x_2\rangle=\langle A^{H}x_1,x_2\rangle\\&=\langle x_1,Ax_2\rangle=\langle x_1,\mu x_2\rangle=\bar{\mu}\langle x_1,x_2\rangle=\mu\langle x_1,x_2\rangle,\end{aligned}$$

即$(\lambda-\mu)\langle x_1,x_2\rangle=0$.

但 $\lambda\neq\mu$，故$\langle x_1,x_2\rangle=0$，亦即 $x_1\perp x_2$. ∎

推论 实对称矩阵的特征值都是实数，且属于不同特征值的特征向量必正交.

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

定理 1.5-9 设 $\boldsymbol{A}$ 是 n 阶 Hermite 矩阵，则必存在酉矩阵 $\boldsymbol{U}$ 使 $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$ 为对角矩阵.

证 对 $\boldsymbol{A}$ 的阶数 n 用数学归纳法.

$n=1$ 时定理显然成立. 现假设对于 $n-1$ 阶 Hermite 矩阵定理成立，要证阶数为 n 时定理也成立.

令 $\boldsymbol{\varepsilon}_1$ 是 $\boldsymbol{A}$ 的属于特征值 λ_1（实数）的单位特征向量，以 $\boldsymbol{\varepsilon}_1$ 为第一列构造酉矩阵 $\boldsymbol{Q}$（这是做得到的，具体构造方法见下面的例 8），记 $\boldsymbol{Q}$ 的其余各列为 $\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$，因而 $\boldsymbol{Q}=[\boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\cdots\ \boldsymbol{\varepsilon}_n]$. 由于

$$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{\varepsilon}_1=\boldsymbol{Q}^{-1}(\boldsymbol{A}\boldsymbol{\varepsilon}_1)=\lambda_1\boldsymbol{Q}^{-1}\boldsymbol{\varepsilon}_1,$$

而
$$\boldsymbol{Q}^{-1}\boldsymbol{Q}=\boldsymbol{Q}^{-1}[\boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\cdots\ \boldsymbol{\varepsilon}_n]=[\boldsymbol{Q}^{-1}\boldsymbol{\varepsilon}_1\ \ \boldsymbol{Q}^{-1}\boldsymbol{\varepsilon}_2\ \cdots\ \boldsymbol{Q}^{-1}\boldsymbol{\varepsilon}_n]=\boldsymbol{I},$$

故 $\boldsymbol{Q}^{-1}\boldsymbol{\varepsilon}_1=[1,0,\cdots,0]^{\mathrm{T}}$，从而

$$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{\varepsilon}_1=[\lambda_1,0,\cdots,0]^{\mathrm{T}}.$$

因此
$$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}=\left[\begin{array}{c:ccc}\lambda_1 & c_2 & \cdots & c_n\\ \hdashline 0 & & & \\ \vdots & & \boldsymbol{A}_1 & \\ 0 & & & \end{array}\right],$$
其中 $\boldsymbol{A}_1$ 为 $n-1$ 阶方阵.

因为 $\boldsymbol{Q}$ 是酉矩阵，故 $\boldsymbol{Q}^{-1}=\boldsymbol{Q}^{\mathrm{H}}$，所以

$$(\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q})^{\mathrm{H}}=(\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Q})^{\mathrm{H}}=\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Q}=\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}.$$

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

于是 $Q^{-1}AQ$ 是 Hermite 矩阵，故 $c_2=\cdots=c_n=0$，且 A_1 是 $n-1$ 阶 Hermite 矩阵，并且

$$Q^{-1}AQ=\left[\begin{array}{c|ccc}\lambda_1 & 0 & \cdots & 0\\ \hline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right].$$

由归纳假设知，存在 $n-1$ 阶酉矩阵 U_1，使 $U_1^{-1}AU_1$ 为对角矩阵。令 n 阶矩阵

$$V=\left[\begin{array}{c|ccc}1 & 0 & \cdots & 0\\ \hline 0 & & & \\ \vdots & & U_1 & \\ 0 & & & \end{array}\right],$$

因

$$V^{-1}=\left[\begin{array}{c|c}1 & \\ \hline & U_1\end{array}\right]^{-1}=\left[\begin{array}{c|c}1 & \\ \hline & U_1^{-1}\end{array}\right]=\left[\begin{array}{c|c}1 & \\ \hline & U_1^{H}\end{array}\right]=V^{H},$$

故 V 是酉矩阵。取 $U=QV$，则作为两个酉矩阵乘积，它是酉矩阵，且

$$U^{-1}AU=V^{-1}(Q^{-1}AQ)V=V^{-1}\left[\begin{array}{c|c}\lambda_1 & \\ \hline & A_1\end{array}\right]V=\left[\begin{array}{c|c}1 & \\ \hline & U_1^{-1}\end{array}\right]\left[\begin{array}{c|c}\lambda_1 & \\ \hline & A_1\end{array}\right]\left[\begin{array}{c|c}1 & \\ \hline & U_1\end{array}\right]$$

$$=\left[\begin{array}{c|c}\lambda_1 & \\ \hline & U_1^{-1}A_1U_1\end{array}\right]=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}.$$

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

定理 1.5-9 表明，任一 Hermite 矩阵必可酉相似对角化.

推论 任一实对称矩阵必可正交相似对角化.

例 8 求以单位向量 $\boldsymbol{\varepsilon}_1=\frac{1}{2}[1,1,1-\mathrm{i}]^{\mathrm{T}}$ 为第一列的酉矩阵 $\boldsymbol{Q}$.

解 与 $\boldsymbol{\varepsilon}_1$ 正交的向量 $\boldsymbol{x}=[x_1,x_2,x_3]^{\mathrm{T}}$ 必须满足 $\langle \boldsymbol{x},\boldsymbol{\varepsilon}_1\rangle=\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{\varepsilon}}_1=0$，即满足方程

$$1\cdot x_1+1\cdot x_2+(1+\mathrm{i})x_3=0.$$

该方程的一个解是 $x_1=-1,x_2=1,x_3=0$.

将向量 $[-1,1,0]^{\mathrm{T}}$ 单位化后取为 $\boldsymbol{\varepsilon}_2$，即

$$\boldsymbol{\varepsilon}_2=\frac{1}{\sqrt{2}}[-1,1,0]^{\mathrm{T}}.$$

与 $\boldsymbol{\varepsilon}_1$、$\boldsymbol{\varepsilon}_2$ 正交的向量 $\boldsymbol{x}=[x_1,x_2,x_3]^{\mathrm{T}}$ 应同时满足 $\langle \boldsymbol{x},\boldsymbol{\varepsilon}_1\rangle=0$ 和 $\langle \boldsymbol{x},\boldsymbol{\varepsilon}_2\rangle=0$，即有

$$\begin{cases} 1\cdot x_1+1\cdot x_2+(1+\mathrm{i})x_3=0, \\ -1\cdot x_1+1\cdot x_2+0\cdot x_3=0. \end{cases}$$

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

解此方程组得其一个解是

$$x_1 = x_2 = -\frac{1}{2}(1+\mathrm{i}), \quad x_3 = 1.$$

将向量 $\left[-\frac{1}{2}(1+\mathrm{i}), -\frac{1}{2}(1+\mathrm{i}), 1\right]^{\mathrm{T}}$ 单位化取为 $\boldsymbol{\varepsilon}_3$，则

$$\boldsymbol{\varepsilon}_3 = \left[-\frac{1}{2\sqrt{2}}(1+\mathrm{i}), -\frac{1}{2\sqrt{2}}(1+\mathrm{i}), \frac{1}{\sqrt{2}}\right]^{\mathrm{T}}.$$

于是，以 $\boldsymbol{\varepsilon}_1$ 为第一列的一个酉矩阵是

$$\boldsymbol{Q} = [\boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\ \boldsymbol{\varepsilon}_3] = \begin{bmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1+\mathrm{i}}{2\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & -\frac{1+\mathrm{i}}{2\sqrt{2}} \\ \frac{1-\mathrm{i}}{2} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

例 9 求正交矩阵 Q 使 $Q^{-1}AQ$ 为对角矩阵，其中

$$A=\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}.$$

解

$$|A-\lambda I|=\begin{vmatrix}1-\lambda&2&2\\2&1-\lambda&2\\2&2&1-\lambda\end{vmatrix}=\begin{vmatrix}5-\lambda&5-\lambda&5-\lambda\\2&1-\lambda&2\\2&2&1-\lambda\end{vmatrix}=(5-\lambda)\begin{vmatrix}1&1&1\\2&1-\lambda&2\\2&2&1-\lambda\end{vmatrix}$$

$$=(5-\lambda)\begin{vmatrix}1&1&1\\0&-1-\lambda&0\\0&0&-1-\lambda\end{vmatrix}=(5-\lambda)(1+\lambda)^2,$$

故 A 的特征值是 $\lambda_1=5,\lambda_2=\lambda_3=-1$.

对 $\lambda_1=5$，由 $(A-5I)x=0$，即

$$\begin{bmatrix}-4&2&2\\2&-4&2\\2&2&-4\end{bmatrix}x=\mathbf{0}$$

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

得其一个解 $\boldsymbol{x}_1=[1,1,1]^{\mathrm{T}}$. 将它单位化得

$$\boldsymbol{\varepsilon}_1=\frac{1}{\sqrt{3}}[1,1,1]^{\mathrm{T}}.$$

对 $\lambda_2=\lambda_3=-1$，由 $(\boldsymbol{A}-(-1)\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}$，即

$$\begin{bmatrix}2 & 2 & 2\\ 2 & 2 & 2\\ 2 & 2 & 2\end{bmatrix}\boldsymbol{x}=\boldsymbol{0}$$

亦即 $1\cdot x_1+1\cdot x_2+1\cdot x_3=0$，得它的一个解 $[-1,1,0]^{\mathrm{T}}$，单位化后得

$$\boldsymbol{\varepsilon}_2=\frac{1}{\sqrt{2}}[-1,1,0]^{\mathrm{T}}.$$

再解线性方程组
$$\begin{cases}1\cdot x_1+1\cdot x_2+1\cdot x_3=0,\\ -1\cdot x_1+1\cdot x_2+0\cdot x_3=0,\end{cases}$$

得其一个解 $\left[-\frac{1}{2},\ -\frac{1}{2},\ 1\right]^{\mathrm{T}}$. 显然，这个向量是 $\boldsymbol{A}$ 的属于特征值 -1 的特征向量，且与 $\boldsymbol{\varepsilon}_2$ 正交. 将此向量单位化得

$$\boldsymbol{\varepsilon}_3=\frac{1}{\sqrt{6}}[-1,-1,2]^{\mathrm{T}}.$$

1.5.3 Hermit矩阵和实矩阵的特征值与特征向量

令 $\boldsymbol{Q}=[\boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\ \boldsymbol{\varepsilon}_3]$，得正交矩阵

$$\boldsymbol{Q}=\begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix},$$

且有
$$\boldsymbol{AQ}=\boldsymbol{A}[\boldsymbol{\varepsilon}_1 \mid \boldsymbol{\varepsilon}_2 \mid \boldsymbol{\varepsilon}_3]=[5\cdot\boldsymbol{\varepsilon}_1 \mid -1\cdot\boldsymbol{\varepsilon}_2 \mid -1\cdot\boldsymbol{\varepsilon}_3]$$

$$=[\boldsymbol{\varepsilon}_1 \mid \boldsymbol{\varepsilon}_2 \mid \boldsymbol{\varepsilon}_3]\begin{bmatrix} 5 & & \\ & -1 & \\ & & -1 \end{bmatrix}=\boldsymbol{Q}\begin{bmatrix} 5 & & \\ & -1 & \\ & & -1 \end{bmatrix}.$$

因此
$$\boldsymbol{Q}^{-1}\boldsymbol{AQ}=\begin{bmatrix} 5 & & \\ & -1 & \\ & & -1 \end{bmatrix}.$$ ∎

作业（第1部分）

习题 1.5

2. 已知 3 阶方阵 $\boldsymbol{A}$ 的属于特征值 1,0,−1 的特征向量依次为$\boldsymbol{x}_1=[1,2,2]^{\mathrm{T}}$，$\boldsymbol{x}_2=[2,-2,1]^{\mathrm{T}}$，$\boldsymbol{x}_3=[-2,-1,2]^{\mathrm{T}}$，求 $\boldsymbol{A}$ 和 $\boldsymbol{A}^8$.

4. 求正交矩阵 $\boldsymbol{Q}$ 使 $\boldsymbol{Q}^{-1}\boldsymbol{AQ}$ 为对角矩阵，其中

$$\boldsymbol{A}=\begin{bmatrix} 5 & -4 & 2 \\ -4 & 5 & -2 \\ 2 & -2 & 2 \end{bmatrix}.$$

5. 设 $\boldsymbol{A}=\begin{bmatrix} -1 & 2 & 4 \\ 2 & x & 2 \\ 4 & 2 & -1 \end{bmatrix}$ 与 $\boldsymbol{B}=\begin{bmatrix} 5 & & \\ & y & \\ & & -5 \end{bmatrix}$ 相似，求 x,y.

数理统计（5）

厦門大學 XIAMEN UNIVERSITY
信息学院（国家示范性软件学院）School of Informatics
黄 炜 博士·副教授 Dr. Wei Huang

2 参数估计

数理统计学的基本任务之一，是根据样本所提供的信息对未知的总体分布的某些方面进行推断。在数理统计中，总体分布的某些未知的数字特征称为**参数**，所谓**参数估计**就是由样本对总体的未知参数作出估计。参数估计又分为**点估计**和**区间估计**两种形式。本章讨论估计量的求法，评价估计量好坏的标准，以及总体均值与方差的区间估计等问题。

2.1 点估计

在实际问题中，常常遇到这样一类统计问题：总体分布虽然未知，但凭借历史资料和有关知识可以认为总体分布的类型是确定的，总体分布的未知因素集中表现在一个或几个参数上．这时，只要对这些未知参数作出估计，就可以认为总体分布确定了，从而可在此基础上解决某些概率论方面的问题．这种参数称为**总体参数**，其可能取值的范围称为**参数空间**．

用 θ 表示未知的总体参数，Θ 表示参数空间．为了强调总体分布与总体参数之间的密切联系，通常把总体分布函数记作 $F(x;\theta)$，把总体概率密度（或概率函数）$f(x)$ 记作 $f(x;\theta)$．这时 F 与 f 的形式都是已知的，且 θ 的取值范围 Θ 也是已知的．

如果表征总体分布的总体参数不止一个，例如有 k 个总体参数 $\theta_1,\theta_2,\cdots,\theta_k$，那么可以用 k 维向量 $\boldsymbol{\theta}=[\theta_1,\theta_2,\cdots,\theta_k]^{\mathrm{T}}$ 表示总体参数，并仍用 Θ 表示参数空间，这时 Θ 是 k 维欧氏空间 $\mathbf{R}^k$ 中的某个已知集合．例如，正态总体 $N(\mu;\sigma^2)$，其中 μ 与 σ^2 均未知，那么总体参数 $\boldsymbol{\theta}=[\mu,\sigma^2]^{\mathrm{T}}$，且参数空间 Θ 为

$$\Theta=\{[\mu;\sigma^2]^{\mathrm{T}}\mid\mu\in(-\infty,+\infty),\sigma^2>0\}.$$

2.1 点估计

相应地，总体分布函数 $F(x;\theta_1,\theta_2,\cdots,\theta_k)$，总体概率密度（或概率函数）$f(x;\theta_1,\theta_2,\cdots,\theta_k)$可以分别表示为 $F(x;\boldsymbol{\theta})$，$f(x;\boldsymbol{\theta})$.

参数估计问题的实质是，给出了一族分布函数$\{F(x;\theta)\mid\theta\in\Theta\}$，只知道其中有一个是总体分布函数，但不知道究竟具体是哪一个，需要根据样本来估计这个实际上的总体分布函数. 由于参数 θ 表征了分布函数，所以问题归结为根据样本估计这个实际上的 θ 值. 一旦根据样本所提供的信息对 θ 作出了估计，也就相当于在分布函数族$\{F(x;\theta)\mid\theta\in\Theta\}$中选定了一个具体的分布函数作为总体分布函数的估计.

在估计问题中，除了要对 θ 作估计外，有时还要对 θ 的某个函数 $g(\theta)$作估计，这里的 g 是一个已知函数. 例如，当总体 $X\sim P(\lambda)$，其中 λ 是总体参数时，我们要估计总体 X 的二阶原点矩$E(X^2)=\lambda+\lambda^2$. 显然，它是总体参数 λ 的一个函数. 总体参数 θ 的这种函数 $g(\theta)$称为**待估函数**.

2.1 点估计

对于点估计问题，所谓根据样本$(X_1,X_2,\cdots,X_n)$来估计θ，就是要构造一个统计量$h(X_1,X_2,\cdots,X_n)$，并在取得样本观测值$(x_1,x_2,\cdots,x_n)$后，用统计量h的观测值$h(x_1,x_2,\cdots,x_n)$作为θ的估计值. $h(X_1,X_2,\cdots,X_n)$称为未知参数θ的估计量，记为$\hat{\theta}(X_1,X_2,\cdots,X_n)$；而$h(x_1,x_2,\cdots,x_n)$称为$\theta$的估计值，记为$\hat{\theta}(x_1,x_2,\cdots,x_n)$. 为了方便起见，用$\hat{\theta}$表示$\theta$的估计量或估计值，具体的含义视上下文而定. 类似地，用$\hat{g}$表示待估函数$g(\theta)$的估计量或估计值.

求点估计的方法很多，这里介绍其中的两种方法：矩法和极大似然法. 另一种求点估计的方法是最小二乘法，我们将在第四章中讨论.

2.1.1 矩估计

1.2 节中的Гливенко定理表明，经验分布函数是总体分布函数的一个很好的近似. 因此自然会想到，能否用样本矩来估计总体分布的相应矩呢？实践证明，这个方法是可取的. 这种用样本矩“替代”相应的总体矩的估计方法称为**矩估计法**. 例如，用样本 k 阶原点矩来估计总体的 k 阶原点矩；用样本 k 阶中心矩来估计总体的 k 阶中心矩. 特别是用样本均值估计总体均值，用样本方差估计总体方差.

例 1 设总体 X 服从指数分布 $E(\lambda)$，其中 $\lambda>0$ 但未知，$(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本，且 $(x_1, x_2, \cdots, x_n)$ 是样本观测值，现要估计未知参数 λ. 由于总体均值 $E(X)=\frac{1}{\lambda}$，故 $\lambda=\frac{1}{E(X)}$，所以用矩估计法得 λ 的矩估计量为

$$\hat{\lambda}=\frac{1}{\bar{X}}=\frac{n}{\sum_{i=1}^{n} X_i},$$

λ 的矩估计值为

$$\hat{\lambda}=\frac{1}{\bar{x}}=\frac{n}{\sum_{i=1}^{n} x_i}.$$ ∎

值得指出的是，估计量是统计量，从而是随机变量，而估计值是该统计量的观测值.

2.1.1 矩估计

例 2 设总体 $X \sim B(1,p)$，其中 $0 \leqslant p \leqslant 1$ 但未知，$(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本，$(x_1, x_2, \cdots, x_n)$ 是样本观测值，则由于 $E(X)=p$，所以 p 的矩估计量与估计值分别为

$$\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{p} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

一般地，设总体 X 的分布函数族为 $\{F(x;\theta_1,\theta_2,\cdots,\theta_k) \mid (\theta_1,\theta_2,\cdots,\theta_k) \in \Theta\}$，其中 $\theta_1, \theta_2, \cdots, \theta_k$ 是 k 个未知的总体参数，$(X_1, X_2, \cdots, X_n)$ 是取自这个总体的一个样本. 假定 X 的前 k 阶原点矩 $\alpha_j \triangleq E(X^j)$ 存在，$j=1,2,\cdots,k$，它们通常是 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数，即有

$$\begin{cases} \alpha_1 = f_1(\theta_1,\theta_2,\cdots,\theta_k), \\ \alpha_2 = f_2(\theta_1,\theta_2,\cdots,\theta_k), \\ \quad \vdots \\ \alpha_k = f_k(\theta_1,\theta_2,\cdots,\theta_k). \end{cases} \tag{2.1-1}$$

如果方程组(2.1-1)有解，则 $\theta_1,\theta_2,\cdots,\theta_k$ 可以解出而表示为 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 的函数：

$$\begin{cases} \theta_1 = h_1(\alpha_1,\alpha_2,\cdots,\alpha_k), \\ \theta_2 = h_2(\alpha_1,\alpha_2,\cdots,\alpha_k), \\ \quad \vdots \\ \theta_k = h_k(\alpha_1,\alpha_2,\cdots,\alpha_k). \end{cases} \tag{2.1-2}$$

2.1.1 矩估计

由于矩估计法是用样本矩来估计总体矩的，所以 $\alpha_j(j=1,2,\cdots,k)$ 可以用样本的 j 阶原点矩 $M_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j$ 作为其估计量，即 $\hat{\alpha}_j=M_j$. 从而 θ_j 的矩估计量为

$$\hat{\theta}_j = h_j(M_1,M_2,\cdots,M_k) \triangleq \hat{\theta}_j(X_1,X_2,\cdots,X_n), \quad j=1,2,\cdots k. \tag{2.1-3}$$

上述求矩估计量的方法也可以如下进行：在方程组(2.1-1)的左端用 $M_1,M_2,\cdots,M_k$ 分别代替 $\alpha_1,\alpha_2,\cdots,\alpha_k$，然后求解此方程组，所得的一组解就是 $\theta_1,\theta_2,\cdots,\theta_k$ 的矩估计量.

2.1.1 矩估计

例 3 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\theta(\theta+1)x^{\theta-1}(1-x), & 0<x<1,\\ 0, & \text{其他},\end{cases}$$

其中 $\theta>0$ 但未知，$(x_1,x_2,\cdots,x_n)$ 是取自这个总体的样本观测值，求 θ 的矩估计值.

解 由于

$$\alpha_1=E(X)=\int_0^1 x\theta(\theta+1)x^{\theta-1}(1-x)\mathrm{d}x$$

$$=\theta(\theta+1)\int_0^1 x^{\theta}(1-x)\mathrm{d}x=\theta(\theta+1)\left(\frac{x^{\theta+1}}{\theta+1}-\frac{x^{\theta+2}}{\theta+2}\right)\Bigg|_0^1$$

$$=\theta(\theta+1)\left(\frac{1}{\theta+1}-\frac{1}{\theta+2}\right)=\frac{\theta}{\theta+2},$$

故 $\theta=\dfrac{2\alpha_1}{1-\alpha}$，所以 θ 的矩估计值为

$$\hat{\theta}=\frac{2\bar{x}}{1-\bar{x}}=\frac{2(x_1+x_2+\cdots+x_n)}{n-(x_1+x_2+\cdots+x_n)}.$$

2.1.1 矩估计

例 4 设总体 X 在区间(θ_1,θ_2)上服从均匀分布，其中 θ_1、θ_2 均未知，求 θ_1 与 θ_2 的矩估计量.

解 由于

$$\alpha_1 = E(X) = \frac{1}{2}(\theta_1 + \theta_2), \quad \alpha_2 = E(X^2) = \frac{1}{3}(\theta_1^2 + \theta_1\theta_2 + \theta_2^2),$$

解之得

$$\theta_1 = \alpha_1 - \sqrt{3(\alpha_2 - \alpha_1^2)}, \quad \theta_2 = \alpha_1 + \sqrt{3(\alpha_2 - \alpha_1^2)},$$

所以 θ_1 与 θ_2 的矩估计量分别为

$$\hat{\theta}_1 = M_1 - \sqrt{3(M_2 - M_1^2)} = M_1 - \sqrt{3M'_2} = \bar{X} - \sqrt{3\tilde{S}^2},$$

$$\hat{\theta}_2 = M_1 + \sqrt{3(M_2 - M_1^2)} = M_1 + \sqrt{3M'_2} = \bar{X} + \sqrt{3\tilde{S}^2}. \quad \blacksquare$$

2.1.1 矩估计

例 5 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1}, & 0<x<1,\\ 0, & \text{其他},\end{cases}$$

其中 $\theta>0$ 未知，$(0.65,0.80,0.75,0.70,0.85)$ 是来自这个总体 X 的一个容量为 5 的样本观测值，用矩估计法估计概率 $P\{X\leqslant 0.80\}$。

解 由于

$$E(X)=\int_0^1 x\theta x^{\theta-1}\mathrm{d}x=\frac{\theta}{\theta+1},$$

所以 θ 的矩估计量为

$$\hat{\theta}=\frac{\bar{X}}{1-\bar{X}}.$$

由样本观测值算得样本均值的观测值为

$$\bar{x}=\frac{1}{5}(0.65+0.80+0.75+0.70+0.85)=0.75,$$

从而 θ 的估计值是

$$\hat{\theta}=\frac{\bar{x}}{1-\bar{x}}=\frac{0.75}{1-0.75}=3.$$

这就是说，用矩估计法可以认为总体 X 的概率密度是

$$f(x)=\begin{cases}3x^2, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$$

于是

$$P\{X\leqslant 0.80\}=\int_0^{0.80}3x^2\mathrm{d}x=x^3\Big|_0^{0.80}=0.512.$$

2.1.1 矩估计

由上述讨论可以看出，矩估计法是一种简便易行的估计方法，并且只要样本容量 n 充分大，其估计的精度也比较高. 此外矩估计法不依赖总体分布的具体形式，因而适用性较广. 其不足之处在于方法太一般化，从而对某些特殊的分布，它可能不如某些特殊的估计量好. 另外矩估计量可能不唯一，例如 Poisson 分布的数学期望和方差都是 λ，因而样本均值 $\overline{X}$ 和修正的样本方差 $\tilde{S}^2$ 都可以作为 λ 的矩估计量. 再则，矩估计法的前提是总体矩要存在，但有些分布不满足这个要求，例如概率密度为 $f(x)=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\theta)^2}(\lambda>0)$ 的 Cauchy 分布，若要估计未知参数 θ，就不能用样本均值，因为它的一阶原点矩不存在.

2.1.2 极大似然估计法

在参数估计问题中,极大似然估计法是一种广泛使用的求点估计的方法. 极大似然估计法是建立在最大似然原理(即概率最大的事件最可能出现)基础上的一种统计方法. 例如一个随机试验,如果有若干个可能的结果 $A,B,C,\cdots$,现在一次试验中,结果 A 出现了,那么认为 A 出现的概率最大.

设总体 X 的概率密度(或概率函数)为 $f(x;\theta_1,\theta_2,\cdots,\theta_k)$,$(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta$,$(X_1,X_2,\cdots,X_n)$是取自这个总体的一个样本,则这个样本的联合概率密度(或联合概率函数)是

$$\prod_{i=1}^{n} f(x_i;\theta_1,\theta_2,\cdots,\theta_k).$$

对于样本的一个实现$(x_1,x_2,\cdots,x_n)$,它是 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数,称为**似然函数**,记为

$$L(\theta_1,\theta_2,\cdots,\theta_k)=\prod_{i=1}^{n} f(x_i;\theta_1,\theta_2,\cdots,\theta_k). \tag{2.1-4}$$

当 X 是离散型随机变量时,$f(x;\theta_1,\theta_2,\cdots,\theta_k)$是概率函数,似然函数 $L(\theta_1,\theta_2,\cdots,\theta_k)$ 可以看成为:当总体参数取 $\theta_1,\theta_2,\cdots,\theta_k$ 时,样本$(X_1,X_2,\cdots,X_n)$取得观测值$(x_1,x_2,\cdots,x_n)$的概率,即

$$L(\theta_1,\theta_2,\cdots,\theta_k)=P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}.$$

2.1.2 极大似然估计法

当 X 是连续型随机变量时，

$$L(\theta_1,\theta_2,\cdots,\theta_k)\Delta x_1\Delta x_2\cdots\Delta x_n=\prod_{i=1}^{n}f(x_i;\theta_1,\cdots,\theta_k)\Delta x_i$$

近似地表示样本$(X_1,X_2,\cdots,X_n)$落在$(x_1,x_2,\cdots,x_n)$的邻域内的概率，这个概率是 θ_1，θ_2，$\cdots$，θ_k 的函数. 由于 $\Delta x_i(1\leqslant i\leqslant n)$与 $\theta=[\theta_1,\theta_2,\cdots,\theta_k]^{\mathrm{T}}$ 无关，所以 $L(\theta_1,\theta_2,\cdots,\theta_k)$可以看成当总体参数为 $\theta_1,\theta_2,\cdots,\theta_k$ 时，样本$(X_1,X_2,\cdots,X_n)$在$(x_1,x_2,\cdots,x_n)$的邻域内取值的可能性大小. 在观测到$(x_1,x_2,\cdots,x_n)$的前提下，自然应当选取使 $L(\theta_1,\theta_2,\cdots,\theta_k)$取得最大值的 $\theta_1,\theta_2,\cdots,\theta_k$ 值作为未知总体参数的估计值，因为这时出现所获得的观测值的可能性最大. 这种求未知参数的估计的方法称为**极大似然估计法**.

2.1.2 极大似然估计法

定义 2.1-1 如果 $\hat{\theta}_1=\hat{\theta}_1(x_1,x_2,\cdots,x_n),\hat{\theta}_2=\hat{\theta}_2(x_1,x_2,\cdots,x_n),\cdots,\hat{\theta}_k=\hat{\theta}_k(x_1,x_2,\cdots,x_n)$满足

$$L(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)=\max_{(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta}L(\theta_1,\theta_2,\cdots,\theta_k),\tag{2.1-5}$$

则 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 分别称为 $\theta_1,\theta_2,\cdots,\theta_k$ 的**极大似然估计值**. 相应地,$\hat{\theta}_1(X_1,X_2,\cdots,X_n)$,$\hat{\theta}_2(X_1,X_2,\cdots,X_n),\cdots,\hat{\theta}_k(X_1,X_2\cdots,X_n)$分别称为 $\theta_1,\theta_2,\cdots,\theta_k$ 的**极大似然估计量**.

如果待估函数为 $g(\theta_1,\theta_2,\cdots,\theta_k)$,则 g 的极大似然估计量为

$$\hat{g}(X_1,X_2,\cdots,X_n)\triangleq g(\hat{\theta}_1(X_1,X_2,\cdots,X_n),$$
$$\hat{\theta}_2(X_1,X_2,\cdots,X_n),\cdots,\hat{\theta}_k(X_1,X_2,\cdots,X_n)),$$

其中 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 分别是 $\theta_1,\theta_2,\cdots,\theta_k$ 的极大似然估计量.

由于似然函数是一些函数的乘积,取对数后运算会简化,又因对数函数是单调增函数,故(2.1-5)式等价于

$$\begin{aligned}\ln L(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)&=\max_{(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta}\ln L(\theta_1,\theta_2,\cdots,\theta_k)\\&=\max_{(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta}\sum_{i=1}^{n}\ln f(x_i;\theta_1,\theta_2,\cdots,\theta_k).\end{aligned}\tag{2.1-6}$$

2.1.2 极大似然估计法

当 $f(x;\theta_1,\theta_2,\cdots,\theta_k)$ 关于 $\theta_1,\theta_2,\cdots,\theta_k$ 可微时，则 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 必须满足方程组

$$\frac{\partial}{\partial\theta_j}\ln L(\theta_1,\theta_2,\cdots,\theta_k)=\sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\ln f(x_i;\theta_1,\theta_2,\cdots,\theta_k)=0,\quad j=1,2,\cdots,k. \tag{2.1-7}$$

(2.1-7)式也称为**似然方程组**. 一般来说，由解似然方程组(2.1-7)所得的解只是 L 的驻点，是否为 L 的极大值点还应加以验证，不过，本书不需要再验证.

需要指出的是，当由似然方程组(2.1-7)求出的解没有意义，或者 $f(x;\theta_1,\theta_2,\cdots,\theta_k)$ 关于 $\theta_1,\theta_2,\cdots,\theta_k$ 的偏导数不存在时，只能针对具体问题采用别的方法寻求使 L 达到极大的估计 $\hat{\theta}_j$，$j=1,2,\cdots,k$.

2.1.2 极大似然估计法

例 6 设总体 X 服从 Poisson 分布 $P(\lambda)$，其中 $\lambda>0$ 未知，求 λ 的极大似然估计量.

解 Poisson 分布 $P(\lambda)$ 的概率函数是

$$f(x;\lambda)=P\{X=x\}=\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda},\quad x=0,1,2,\cdots,$$

故似然函数(即 $P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$)

$$L(\lambda)=\prod_{i=1}^{n}f(x_i;\lambda)=\prod_{i=1}\frac{\lambda^{x_i}}{x_i!}\mathrm{e}^{-\lambda}=\mathrm{e}^{-n\lambda}\frac{\lambda^{\sum\limits_{i=1}^{n}x_i}}{\prod\limits_{i=1}^{n}x_i!},$$

$$\ln L(\lambda)=-n\lambda+\left(\sum_{i=1}^{n}x_i\right)\ln\lambda-\sum_{i=1}^{n}\ln(x_i!).$$

因此，似然方程是 $$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ln L(\lambda)=-n+\frac{1}{\lambda}\sum_{i=1}^{n}x_i=0,$$

解之得 $$\lambda=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

于是，λ 的极大似然估计量为 $$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$$

其中$(X_1,X_2,\cdots,X_n)$是取自总体 X 的样本. ∎

2.1.2 极大似然估计法

例 7 设$(X_1,X_2,\cdots,X_n)$是取自正态总体$N(\mu;\sigma^2)$的样本，其中μ、σ^2 均未知，求μ与σ^2 的极大似然估计量，以及σ的极大似然估计量.

解 似然函数为

$$L(\mu;\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right]=(2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right],$$

$$\ln L(\mu;\sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

似然方程组是
$$\begin{cases}\dfrac{\partial}{\partial\mu}\ln L(\mu;\sigma^2)=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0,\\[2ex]\dfrac{\partial}{\partial\sigma^2}\ln L(\mu;\sigma^2)=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2=0.\end{cases}$$

解此方程组得
$$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},\quad \sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\tilde{s}^2,$$

故μ与σ^2 的极大似然估计量分别为

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X},\quad \widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\tilde{S}^2,$$

由于$\sigma=\sqrt{\sigma^2}$，所以σ的极大似然估计量为

$$\hat{\sigma}=\sqrt{\widehat{\sigma^2}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}=\tilde{S}.$$ ■

2.1.2 极大似然估计法

例 8 设总体 X 服从 Maxwell 分布，其概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{4x^2}{\theta^3\sqrt{\pi}}e^{-\frac{x^2}{\theta^2}}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

其中 $\theta>0$ 但未知，求 θ 的极大似然估计量.

解 设$(X_1,X_2,\cdots,X_n)$是取自总体 X 的样本，则似然函数为

$$L(\theta)=\left(\frac{4}{\sqrt{\pi}}\right)^n\theta^{-3n}\left(\prod_{i=1}^n x_i^2\right)\exp\left(-\frac{1}{\theta^2}\sum_{i=1}^n x_i^2\right),$$

$$\ln L(\theta)=n\ln\frac{4}{\sqrt{\pi}}-3n\ln\theta+\sum_{i=1}^n\ln(x_i^2)-\frac{1}{\theta^2}\sum_{i=1}^n x_i^2,$$

故似然方程是 $$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta)=-\frac{3n}{\theta}+\frac{2}{\theta^3}\sum_{i=1}^n x_i^2=0.$$

解之得 $$\theta=\sqrt{\frac{2}{3n}\sum_{i=1}^n x_i^2}.$$

于是，θ 的极大似然估计量为 $$\hat{\theta}=\sqrt{\frac{2}{3n}\sum_{i=1}^n X_i^2}.$$ ∎

2.1.2 极大似然估计法

例 9 设总体 X 服从 Cauchy 分布，其概率密度为

$$f(x;\theta)=\frac{1}{\pi[1+(x-\theta)^2]},\quad -\infty<x<+\infty,$$

其中 θ 未知，$-\infty<\theta<+\infty$，$(x_1,x_2,\cdots,x_n)$ 是取自 X 的样本观测值，求 θ 的极大似然估计值.

解 似然函数为 $$L(\theta)=\pi^{-n}\prod_{i=1}^{n}[1+(x_i-\theta)^2]^{-1},$$

$$\ln L(\theta)=-n\ln\pi-\sum_{i=1}^{n}\ln[1+(x_i-\theta)^2],$$

故似然方程是 $$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta)=2\sum_{i=1}^{n}\frac{x_i-\theta}{1+(x_i-\theta)^2}=0.$$

由此可得 θ 的极大似然估计值 $\hat{\theta}$ 所满足的方程为

$$\hat{\theta}=\frac{\sum_{i=1}^{n}\frac{x_i}{1+(x_i-\hat{\theta})^2}}{\sum_{i=1}^{n}\frac{1}{1+(x_i-\hat{\theta})^2}}. \tag{2.1-8}$$

(2.1-8)式是 $\hat{\theta}$ 的非线性方程，一般是用迭代法求其近似值. 如果用样本中位数的观测值 $\mathrm{med}(x_1,x_2,\cdots,x_n)$ 作为 $\hat{\theta}$ 的初始值 $\hat{\theta}^{(0)}$，则可用迭代公式

$$\hat{\theta}^{(k+1)}=\frac{\sum_{i=1}^{n}\frac{x_i}{1+(x_i-\hat{\theta}^{(k)})^2}}{\sum_{i=1}^{n}\frac{1}{1+(x_i-\hat{\theta}^{(k)})^2}}$$

求得 $\hat{\theta}$ 的近似值. ∎

2.1.2 极大似然估计法

例 10 设总体 X 服从(θ_1,θ_2)上的均匀分布，其中$-\infty<\theta_1<\theta_2<+\infty$，但 θ_1 与 θ_2 均未知，$(X_1,X_2,\cdots,X_n)$是取自 X 的样本，求 θ_1 与 θ_2 的极大似然估计量.

解 由于总体 X 的概率密度为

$$f(x;\theta_1,\theta_2)=\begin{cases}\dfrac{1}{\theta_2-\theta_1}, & \theta_1<x<\theta_2,\\ 0, & \text{其他},\end{cases}$$

所以似然函数为 $L(\theta_1,\theta_2)=\begin{cases}\dfrac{1}{(\theta_2-\theta_1)^n}, & \theta_1<x_1,x_2,\cdots,x_n<\theta_2,\\ 0, & \text{其他}.\end{cases}$

这时不能用微分法求得 θ_1、θ_2 的估计值，但根据极大似然估计法的思想知，θ_1 与 θ_2 的选取应使 $L(\theta_1,\theta_2)$取得最大值. 显然，θ_2 与 θ_1 之差越小，则 $L(\theta_1,\theta_2)$的值越大；不过，θ_1 与 θ_2 又必须满足下式

$$\theta_1<x_1,x_2,\cdots,x_n<\theta_2,$$

所以有 $$\theta_1<\min_{1\leqslant i\leqslant n}x_i=x_{(1)},\quad \theta_2>\max_{1\leqslant i\leqslant n}x_i=x_{(n)}.$$

因此，θ_1 与 θ_2 的极大似然估计值分别为

$$\hat{\theta}_1=x_{(1)},\quad \hat{\theta}_2=x_{(n)}.$$

从而 θ_1 与 θ_2 的极大似然估计量分别为

$$\hat{\theta}_1=X_{(1)},\quad \hat{\theta}_2=X_{(n)}.$$ ∎

2.1.2 极大似然估计法

从这个例子看出，次序统计量可以作为未知参数的估计量. 事实上，次序统计量及其函数在点估计问题中有相当广泛的应用. 例如，当总体 X 的分布为对称分布时，可以用样本中位数来估计其对称中心位置，如例 9 所述. 对次序统计量及其应用有兴趣的读者可参阅 David, H. A. Order statistics, Wiley, New York, 1970.

极大似然估计充分利用了总体分布的类型的信息，从而极大似然估计量具有许多重要的性质. 如果总体分布的具体形式已知，通常总是对其未知参数应用极大似然估计法求出未知参数的极大似然估计. 当然，如果总体分布的具体形式无法确定，这时多选用矩估计法.

2.2 估计量的评选标准

对于同一个未知参数，可以构造不同的估计量，例如，设总体 X 为$(0,\theta)$上的均匀分布，其中 $\theta>0$ 未知，则容易验证，θ 的矩估计量为 $2\overline{X}$，而极大似然估计量为$\max\limits_{1\leqslant i\leqslant n} X_i = X_{(n)}$，那么究竟采用哪一个估计量好呢？这就涉及用什么标准来评价估计量的问题.

由于估计量是样本的函数，它是随机变量，由不同的样本观测值所得到的估计值也就不同，所以当我们采用 θ 的某个估计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$时，必须考察 $\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta$ 的性态. 直观上说，$\hat{\theta}-\theta$ 反映了估计的误差，一个好的估计自然应当使 $\hat{\theta}-\theta$ 尽可能接近于 0，但要使它等于 0 是不现实的，从而只能在概率论意义上提出如下一些要求：

(1) 无系统偏差，即 $\hat{\theta}$ 平均起来应与 θ 值相同，也就是说 $E[\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta]=0$. 这就是无偏性的要求.

(2) 波动性小，即 $|\hat{\theta}-\theta|$ 平均起来应尽可能小，为数学上便于处理，可用 $\hat{\theta}(X_1,X_2,\cdots,X_n)$的均方误差 $E[(\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta)^2]$表示. 这就是有效性的要求.

(3) 随着样本容量 n 的增大，$\hat{\theta}$ 应越来越接近于 θ，亦即$|\hat{\theta}-\theta|$按某种概率意义收敛于 0. 这就是相合性的要求.

下面分别讨论这些要求.

2.2.1 无偏估计

定义 2.2-1 设 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是未知参数 θ 的估计量，如果对任意 $\theta\in\Theta$ 均有

$$E[\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta]=0, \tag{2.2-1}$$

则称 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 θ 的**无偏估计量**，简称 $\hat{\theta}$ 为 θ 的**无偏估计**. 如果(2.2-1)式不成立，则称 $\hat{\theta}$ 为 θ 的**有偏估计**.

(2.2-1)式可以等价地表示为

$$E[\hat{\theta}(X_1,X_2,\cdots,X_n)]=\theta. \tag{2.2-1'}$$

如果 $g(\theta)$ 是待估函数，那么 $g(\theta)$ 的估计量 $\hat{g}(X_1,X_2,\cdots,X_n)$ 满足

$$E[\hat{g}(X_1,X_2,\cdots,X_n)]=g(\theta) \tag{2.2-2}$$

时，则称 $\hat{g}$ 为 g 的无偏估计.

例 1 设总体 X 的数学期望 $E(X)=\mu$ 未知，$-\infty<\mu<+\infty$，$(X_1,X_2,\cdots,X_n)$ 是取自总体 X 的样本，问下列统计量是否为 μ 的无偏估计：

(1) $\hat{\mu}_1=\overline{X}$；(2) $\hat{\mu}_2=\sum_{i=1}^{n}a_iX_i$，其中 $\sum_{i=1}^{n}a_i=1$.

解 (1) $E(\hat{\mu}_1)=E(\overline{X})=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu$，所以 $\hat{\mu}_1$ 是 μ 的无偏估计.

(2) $E(\hat{\mu}_2)=E\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_iE(X_i)=\mu\sum_{i=1}^{n}a_i=\mu$，故 $\hat{\mu}_2$ 也是 μ 的无偏估计. ∎

2.2.1 无偏估计

例 2 设$(X_1,X_2,\cdots,X_n)$是取自总体 X 的样本，$E(X)=\mu$ 和 $D(X)=\sigma^2$ 均未知，证明$(\overline{X})^2$，$\tilde{S}^2$ 分别是 μ^2、σ^2 的有偏估计.

证 由 $E[(\overline{X})^2]=D(\overline{X})+[E(\overline{X})]^2$，$D(\overline{X})=\frac{1}{n}\sigma^2$，$E(\overline{X})=\mu$ 知，

$$E[(\overline{X})^2]=\frac{\sigma^2}{n}+\mu^2\neq\mu^2,$$

故$(\overline{X})^2$ 是 μ^2 的有偏估计.

由于
$$\tilde{S}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2$$
$$=\frac{1}{n}[\sum_{i=1}^{n}(X_i-\mu)^2-2\sum_{i=1}^{n}(X_i-\mu)(\overline{X}-\mu)+\sum_{i=1}^{n}(\overline{X}-\mu)^2]$$
$$=\frac{1}{n}[\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2],$$

所以 $E(\tilde{S}^2)=\frac{1}{n}E[\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2]$

$$=\frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu)^2]-E[(\overline{X}-\mu)^2]=\sigma^2-D(\overline{X})=(1-\frac{1}{n})\sigma^2\neq\sigma^2,$$

故 $\tilde{S}^2$ 是 σ^2 的有偏估计. ∎

2.2.1 无偏估计

例 3 设总体 X 服从 $(0,\theta)$ 上的均匀分布，其中 $\theta>0$ 未知，问 θ 的矩估计量 $\hat{\theta}_M=2\overline{X}$，$\theta$ 的极大似然估计量 $\hat{\theta}_L=X_{(n)}$ 是否为 θ 的无偏估计量？若不是，则由其构造无偏估计量.

解 由于 $E(X)=\int_0^\theta x\cdot\frac{1}{\theta}\mathrm{d}x=\frac{\theta}{2}$，所以

$$E(2\overline{X})=\frac{2}{n}\sum_{i=1}^{n}E(X_i)=\frac{2}{n}\cdot\frac{n}{2}\theta=\theta,$$

故 $\hat{\theta}_M=2\overline{X}$ 是 θ 的无偏估计量.

由于 $X_{(n)}$ 的分布函数为 $F(x)=\begin{cases}0, & x\leqslant 0,\\ \left(\dfrac{x}{\theta}\right)^n, & 0<x\leqslant\theta,\\ 1, & x>\theta,\end{cases}$

所以 $X_{(n)}$ 的概率密度为 $f(x)=\begin{cases}\dfrac{nx^{n-1}}{\theta^n}, & 0<x<\theta,\\ 0, & \text{其他}.\end{cases}$

因此 $E(X_{(n)})=\int_0^\theta x\frac{nx^{n-1}}{\theta^n}\mathrm{d}x=\frac{n}{n+1}\theta\neq\theta,$

故 $\hat{\theta}_L=X_{(n)}$ 不是 θ 的无偏估计量.

显然，$\hat{\theta}_L=X_{(n)}$ 是 θ 的渐近无偏估计，并且可由纠偏公式(2.2-5)得到 θ 的一个无偏估计量 $\frac{n+1}{n}X_{(n)}$. ∎

2.2.1 无偏估计

值得指出的是，即使 $\hat{\theta}$ 是未知参数 θ 的无偏估计，但 $\hat{\theta}$ 的函数 $g(\hat{\theta})$ 一般不是 $g(\theta)$ 的无偏估计. 例如，例 2 中的样本均值 $\overline{X}$ 是总体均值 μ 的一个无偏估计，但 $(\overline{X})^2$ 并不是 μ^2 的无偏估计. 又如，样本方差 S^2 是 σ^2 的无偏估计，但 S 不是 σ 的无偏估计. 事实上当 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体 $N(\mu;\sigma^2)$ 的样本时，$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，从而

$$E\left(\frac{\sqrt{n-1}S}{\sigma}\right) = \int_0^{+\infty} \frac{\sqrt{x}}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} e^{-\frac{x}{2}} x^{\frac{n-1}{2}-1} dx = \frac{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} = \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},$$

故有

$$E(S) = \sqrt{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \sigma \neq \sigma.$$

2.2.2 有效估计和最小方差估计

无偏性是评选估计量的一个重要且有用的概念，但不是唯一的标准. 事实上，较为合理的一个标准应该是，选取估计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 使 $E[(\hat{\theta}-\theta)^2]$ 尽可能小.

定义 2.2-3 设 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是未知参数 θ 的估计量，称 $E[(\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta)^2]$ 为估计量 $\hat{\theta}$ 的**均方误差**，记为 $r(\hat{\theta})$，即

$$r(\hat{\theta})=E[(\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta)^2]. \tag{2.2-6}$$

定理 2.2-1 未知参数 θ 的估计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 的均方误差 $r(\hat{\theta})$ 与 $\hat{\theta}$ 的方差 $D(\hat{\theta})$ 和偏差 $b(\hat{\theta})$ 有关系式：

$$r(\hat{\theta})=D(\hat{\theta})+[b(\hat{\theta})]^2. \tag{2.2-7}$$

证 由 $r(\hat{\theta})$ 和 $b(\hat{\theta})$ 的定义，并注意到

$$E[\hat{\theta}-E(\hat{\theta})]=E(\hat{\theta})-E(\hat{\theta})=0,$$

便得

$$\begin{aligned} r(\hat{\theta}) &= E[(\hat{\theta}-\theta)^2]=E\{[(\hat{\theta}-E(\hat{\theta}))-(\theta-E(\hat{\theta}))]^2\} \\ &= E[(\hat{\theta}-E(\hat{\theta}))^2]+E[(\theta-E(\hat{\theta}))^2]-2E(\hat{\theta}-E(\hat{\theta}))(\theta-E(\hat{\theta})) \\ &= D(\hat{\theta})+(\theta-E(\hat{\theta}))^2=D(\hat{\theta})+[b(\hat{\theta})]^2. \end{aligned}$$ ■

若以均方误差作为评选估计量的标准，那么均方误差较小的估计量较优.

厦門大學 XIAMEN UNIVERSITY 信息学院（国家示范性软件学院）School of Informatics 黄 炜 博士·副教授 Dr. Wei Huang

2.2.2 有效估计和最小方差估计

例 4 设$(X_1,X_2,\cdots,X_n)$是取自正态总体$N(0;\sigma^2)$的样本，其中$\sigma^2>0$未知，问正常数c取何值时，估计量$\widehat{\sigma}^2=c\sum_{i=1}^{n}X_i^2$的均方误差最小？

解 由χ^2分布的定义知，$\frac{1}{\sigma^2}\sum_{i=1}^{n}X_i^2\sim\chi^2(n)$，而(1.3-7)式给出

$$E\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}X_i^2\right)=n,\quad D\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}X_i^2\right)=2n,$$

从而

$$E\left(\sum_{i=1}^{n}X_i^2\right)=n\sigma^2,\quad D\left(\sum_{i=1}^{n}X_i^2\right)=2n\sigma^4.$$

于是，

$$b(\widehat{\sigma}^2)=E(\widehat{\sigma}^2)-\sigma^2=cE\left(\sum_{i=1}^{n}X_i^2\right)-\sigma^2=(nc-1)\sigma^2;$$

$$D(\widehat{\sigma}^2)=c^2D\left(\sum_{i=1}^{n}X_i^2\right)=2nc^2\sigma^4;$$

$$\begin{aligned}r(\widehat{\sigma}^2)&=D(\widehat{\sigma}^2)+[b(\widehat{\sigma}^2)]^2=2nc^2\sigma^4+(nc-1)^2\sigma^4\\&=[(n^2+2n)c^2-2nc+1]\sigma^4.\end{aligned}\tag{2.2-8}$$

(2.2-8)式右端作为c的二次函数，由高等数学知，当$c=\frac{1}{n+2}$时，均方误差$r(\widehat{\sigma}^2)$达到最小，且最小值为$\frac{2\sigma^4}{n+2}$. ∎

易知，$\frac{1}{n+2}\sum_{i=1}^{n}X_i^2$既不是$\sigma^2$的无偏估计，也不是$\sigma^2$的极大似然估计，但它是$\sigma^2$的一个较优的估计量.

2.2.2 有效估计和最小方差估计

例 5 设总体 X 服从 $[0,\theta]$ 上的均匀分布，其中 $\theta>0$ 未知，$(X_1,X_2,\cdots,X_n)$ 是取自这个总体的样本. 现以均方误差作为估计优劣的标准，比较 θ 的估计 $2\overline{X}$、$X_{(n)}$ 和 $\frac{n+1}{n}X_{(n)}$ 中哪一个较优？

解 由于 θ 的矩估计 $2\overline{X}$ 是无偏估计，又因 $D(X)=\frac{\theta^2}{12}$，所以有

$$r(2\overline{X}) = D(2\overline{X}) = \frac{4}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{4}{n^2}\cdot\frac{n\theta^2}{12} = \frac{\theta^2}{3n}. \qquad (2.2\text{-}9)$$

θ 的极大似然估计 $X_{(n)}$ 是有偏估计，在例 3 中已经得到 $E(X_{(n)})=\frac{n}{n+1}\theta$，因而有

$$b(X_{(n)}) = E(X_{(n)}) - \theta = -\frac{\theta}{n+1}.$$

又因为

$$E(X_{(n)}^2) = \int_0^\theta x^2\,\frac{nx^{n-1}}{\theta^n}\mathrm{d}x = \frac{n}{n+2}\theta^2,$$

所以得到

$$r(X_{(n)}) = D(X_{(n)}) + [b(X_{(n)})]^2 = E(X_{(n)}^2) - [E(X_{(n)})]^2 + [b(X_{(n)})]^2$$

$$= \frac{n}{n+2}\theta^2 - \left(\frac{n\theta}{n+1}\right)^2 + \frac{\theta^2}{(n+1)^2} = \frac{2\theta^2}{(n+1)(n+2)}. \qquad (2.2\text{-}10)$$

2.2.2 有效估计和最小方差估计

再则，由例 3 知，估计量 $\frac{n+1}{n}X_{(n)}$ 是 θ 的无偏估计量，且 $D(X_{(n)}) = E(X_{(n)}^2) - [E(X_{(n)})]^2 = \frac{n\theta^2}{(n+1)^2(n+2)}$，故有

$$r\left(\frac{n+1}{n}X_{(n)}\right) = D\left(\frac{n+1}{n}X_{(n)}\right) = \frac{(n+1)^2}{n^2}D(X_{(n)}) = \frac{\theta^2}{n(n+2)}. \qquad (2.2\text{-}11)$$

将(2.2-11)式与(2.2-9)和(2.2-10)式比较，不难验证，当 $n \geqslant 2$ 时，有

$$r\left(\frac{n+1}{n}X_{(n)}\right) < r(2\overline{X}), \quad r\left(\frac{n+1}{n}X_{(n)}\right) < r(X_{(n)}).$$

因此，以均方误差为标准，则这三个估计量中最优的是 $\frac{n+1}{n}X_{(n)}$. ■

定义 2.2-4 如果存在未知参数 θ 的一个估计量 $\hat{\theta}^*$，使得对 θ 的任意一个估计量 $\hat{\theta}$，都有

$$r(\hat{\theta}^*) \leqslant r(\hat{\theta}), \qquad (2.2\text{-}12)$$

则称 $\hat{\theta}^*$ 为 θ 的**最小均方误差估计**.

从(2.2-7)式看出，如果限于讨论 θ 的无偏估计，那么由于 $b(\hat{\theta})=0$，故有 $r(\hat{\theta}) = D(\hat{\theta})$，从而均方误差的大小等于估计量方差的值.

2.2.2 有效估计和最小方差估计

定义 2.2-5 设 $\hat{\theta}_1$、$\hat{\theta}_2$ 是未知参数 θ 的两个无偏估计量，若

$$D(\hat{\theta}_1) < D(\hat{\theta}_2), \tag{2.2-12'}$$

则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

例 6 设总体 X 在 $(0,\theta)$ 上服从均匀分布，其中 $\theta>0$ 未知，$(X_1,X_2,\cdots,X_n)$ 是取自这个总体的样本，证明 $(n+1)X_{(1)}$ 是 θ 的无偏估计量，并问 θ 的两个无偏估计量 $(n+1)X_{(1)}$ 与 $\dfrac{n+1}{n}X_{(n)}$ 中哪一个有效？

解 由于 $X_{(1)}$ 的分布函数是

$$F^*(x)=\begin{cases}0, & x\leqslant 0,\\ 1-\left(\dfrac{\theta-x}{\theta}\right)^n, & 0<x\leqslant\theta,\\ 1, & x>\theta,\end{cases}$$

所以 $X_{(1)}$ 的概率密度为 $f^*(x)=\begin{cases}\dfrac{n(\theta-x)^{n-1}}{\theta^n}, & 0<x<\theta,\\ 0, & \text{其他}.\end{cases}$

因此，$E(X_{(1)})=\displaystyle\int_0^\theta x\frac{n(\theta-x)^{n-1}}{\theta^n}\mathrm{d}x=\frac{n}{\theta^n}\int_0^\theta[\theta-(\theta-x)](\theta-x)^{n-1}\mathrm{d}x=\frac{\theta}{n+1}.$

2.2.2 有效估计和最小方差估计

于是，由纠偏公式(2.2-5)知，$(n+1)X_{(1)}$ 是 θ 的一个无偏估计. 又因为

$$E(X_{(1)}^2)=\int_0^\theta x^2\frac{n(\theta-x)^{n-1}}{\theta^n}\mathrm{d}x=\int_0^\theta[(\theta-x)^2+2\theta x-\theta^2]\frac{n(\theta-x)^{n-1}}{\theta^n}\mathrm{d}x$$

$$=\frac{2}{(n+2)(n+1)}\theta^2,$$

所以有 $$D(X_{(1)})=E(X_{(1)}^2)-[E(X_{(1)})]^2=\frac{2}{(n+2)(n+1)}\theta^2-\frac{1}{(n+1)^2}\theta^2$$

$$=\frac{n}{(n+1)^2(n+2)}\theta^2,$$

$$D((n+1)X_{(1)})=(n+1)^2D(X_{(1)})=\frac{n}{n+2}\theta^2. \tag{2.2-13}$$

将(2.2-13)式与(2.2-11)式比较便知，

$$D\left(\frac{n+1}{n}X_{(n)}\right)\leqslant D((n+1)X_{(1)}),$$

故 $\frac{n+1}{n}X_{(n)}$ 比 $(n+1)X_{(1)}$ 有效. ∎

定义 2.2-6 设 $\hat{\theta}^*$ 是未知参数 θ 的无偏估计量，若对于 θ 的任意一个无偏估计量 $\hat{\theta}$，都有

$$D(\hat{\theta}^*)\leqslant D(\hat{\theta}), \tag{2.2-14}$$

则称 $\hat{\theta}^*$ 为 θ 的**最小方差无偏估计**.

2.2.2 有效估计和最小方差估计

于是，由纠偏公式(2.2-5)知，$(n+1)X_{(1)}$是θ的一个无偏估计. 又因为

$$E(X_{(1)}^2)=\int_0^\theta x^2\frac{n(\theta-x)^{n-1}}{\theta^n}\mathrm{d}x=\int_0^\theta[(\theta-x)^2+2\theta x-\theta^2]\frac{n(\theta-x)^{n-1}}{\theta^n}\mathrm{d}x$$

$$=\frac{2}{(n+2)(n+1)}\theta^2,$$

所以有 $$D(X_{(1)})=E(X_{(1)}^2)-[E(X_{(1)})]^2=\frac{2}{(n+2)(n+1)}\theta^2-\frac{1}{(n+1)^2}\theta^2$$

$$=\frac{n}{(n+1)^2(n+2)}\theta^2,$$

$$D((n+1)X_{(1)})=(n+1)^2D(X_{(1)})=\frac{n}{n+2}\theta^2. \tag{2.2-13}$$

将(2.2-13)式与(2.2-11)式比较便知，

$$D\left(\frac{n+1}{n}X_{(n)}\right)\leqslant D((n+1)X_{(1)}),$$

故$\frac{n+1}{n}X_{(n)}$比$(n+1)X_{(1)}$有效. ∎

定义 2.2-6 设$\hat{\theta}^*$是未知参数θ的无偏估计量，若对于θ的任意一个无偏估计量$\hat{\theta}$，都有

$$D(\hat{\theta}^*)\leqslant D(\hat{\theta}), \tag{2.2-14}$$

则称$\hat{\theta}^*$为θ的**最小方差无偏估计**.

2.2.2 有效估计和最小方差估计

例 7 设总体 X 的均值 $E(X)=\mu$ 未知，$-\infty<\mu<+\infty$，$(X_1,X_2,\cdots,X_n)$ 是取自总体 X 的样本，在形如 $\hat{\mu}=\sum_{i=1}^{n}a_iX_i$ 且 $\sum_{i=1}^{n}a_i=1$ 的无偏估计量中求出最小方差无偏估计量 $\hat{\mu}^*$.

解 记 $D(X)=\sigma^2$，则

$$D(\hat{\mu})=D(\sum_{i=1}a_iX_i)=\sum_{i=1}a_i^2D(X_i)=(\sum_{i=1}a_i^2)\sigma^2.$$

于是，问题归结为在条件 $\sum_{i=1}^{n}a_i=1$ 下，求 $a_1,a_2,\cdots,a_n$ 使 $\sum_{i=1}^{n}a_i^2$ 取得最小值.

由高等数学知识可得下述结论：当 $a_1=a_2=\cdots=a_n=\frac{1}{n}$，即 $\hat{\mu}=\hat{\mu}^*=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$ 时，$D(\hat{\mu})$ 取得最小值 $D(\hat{\mu}^*)=\frac{1}{n}\sigma^2$. 也就是说，样本均值 $\overline{X}$ 是形如 $\sum_{i=1}^{n}a_iX_i$ 且 $\sum_{i=1}^{n}a_i=1$ 的无偏估计量中方差最小的. ∎

如何求未知参数的最小方差无偏估计呢?在样本容量一定的条件下，待估参数的无偏估计量的方差是否可以任意小呢?Rao-Cramér 定理回答了这些问题.

2.2.2 有效估计和最小方差估计

定理 2.2-2 设总体 X 的概率密度(或概率函数)为 $f(x;\theta)$，$\theta\in\Theta$，这里 Θ 是 $(-\infty,+\infty)$ 上的一个开区间，$(X_1,X_2,\cdots,X_n)$ 是取自总体 X 的样本，待估函数是 $g(\theta)$，且 $\hat{g}(X_1,X_2,\cdots,X_n)$ 是 $g(\theta)$ 的任意一个无偏估计量，如果

(1) $E\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right]>0$，

(2) $\frac{\partial}{\partial\theta}f(x;\theta)$ 存在，且 $\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}f(x;\theta)\mathrm{d}x=\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}f(x;\theta)\mathrm{d}x$，

(3) $\frac{\partial}{\partial\theta}g(\theta)$ 存在，且

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\hat{g}(x_1,\cdots,x_n)\prod_{i=1}^{n}f(x_i;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n$$
$$=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\hat{g}(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\left(\prod_{i=1}^{n}f(x_i;\theta)\right)\mathrm{d}x_1\cdots\mathrm{d}x_n,$$

则有

$$D[\hat{g}(X_1,X_2,\cdots,X_n)]\geqslant\frac{[g'(\theta)]^2}{nI(\theta)},\qquad(2.2\text{-}15)$$

其中

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

特别地，当 $g(\theta)=\theta$ 时，(2.2-15)式为

$$D[\hat{\theta}(X_1,X_2,\cdots,X_n)]\geqslant\frac{1}{nI(\theta)}.\qquad(2.2\text{-}16)$$

定理 2.2-2 的证明，请参阅参考书[2]. ▌

2.2.2 有效估计和最小方差估计

上述的 $I(\theta)$ 称为 Fisher 信息量，由于

$$\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)=\frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]=\frac{\partial}{\partial\theta}\left[\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right]$$

$$=-\left[\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right]^2+\frac{1}{f(X;\theta)}\frac{\partial^2}{\partial\theta^2}f(X;\theta)$$

$$=-\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2+\frac{1}{f(X;\theta)}\frac{\partial^2}{\partial\theta^2}f(X;\theta),$$

而

$$E\left[\frac{1}{f(X;\theta)}\frac{\partial^2}{\partial\theta^2}f(X;\theta)\right]=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial^2}{\partial\theta^2}f(x_1,\cdots,x_n;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n$$

$$=\frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(x_1,\cdots,x_n;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n=0,$$

所以

$$I(\theta)=E\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right]=-E\left(\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right). \tag{2.2-17}$$

(2.2-15)式右端的 $\frac{[g'(\theta)]^2}{nI(\theta)}$ 称为 **Rao-Cramér** 下界.

(2.2-16) 式表明，如果未知参数 θ 的某个无偏估计量 $\hat{\theta}$，其方差 $D(\hat{\theta})$ 等于 Rao-Cramér 下界 $\frac{1}{nI(\theta)}$，那么这个无偏估计量 $\hat{\theta}$ 必是最小方差无偏估计量.

2.2.2 有效估计和最小方差估计

例 8 设$(X_1, X_2, \cdots, X_n)$是取自总体 $B(1; p)$的一个样本，其中 $p \in [0,1]$但未知，求 p 的无偏估计量方差的下界.

解 这时 X 的概率函数为

$$f(x; p) = p^x (1-p)^{1-x}, x = 0, 1.$$

由于 $$\frac{\partial}{\partial p}\ln(x; p) = \frac{\partial}{\partial p}\{x \ln p + (1-x)\ln(1-p)\} = \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)},$$

所以 $$I(p) = E\left[\left(\frac{X-p}{p(1-p)}\right)^2\right] = \frac{1}{p^2(1-p)^2} E[(X-p)^2]$$

$$= \frac{1}{p^2(1-p)^2} D(X) = \frac{1}{p^2(1-p)^2} p(1-p) = \frac{1}{p(1-p)}.$$

因此，p 的无偏估计量的方差下界是$\dfrac{p(1-p)}{n}$. ■

如果取样本均值 $\overline{X}$ 为 p 的估计量，它显然是无偏估计量，并且由 $D(\overline{X}) = \dfrac{1}{n}D(X) = p(1-p)/n$ 知，$\overline{X}$ 的方差等于 Rao-Cramér 下界 $p(1-p)/n$，故 $\overline{X}$ 是 p 的最小方差无偏估计量.

2.2.2 有效估计和最小方差估计

例 9 设$(X_1,X_2,\cdots,X_n)$是取自正态总体$N(\mu;\sigma^2)$的样本，讨论μ、σ^2的无偏估计量的方差下界.

解

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\},$$

$$\ln f(x;\mu,\sigma^2)=-\frac{1}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}(x-\mu)^2.$$

由于

$$\frac{\partial}{\partial\mu}\ln f(x;\mu,\sigma^2)=\frac{x-\mu}{\sigma^2},\quad \frac{\partial^2}{\partial\mu^2}\ln f(x;\mu,\sigma^2)=-\frac{1}{\sigma^2},$$

所以由(2.2-17)式得

$$I(\mu)=\frac{1}{\sigma^2}.$$

于是，μ的无偏估计的方差下界是σ^2/n. 因为样本均值$\bar{X}$的方差$D(\overline{X})=\sigma^2/n$达到这个下界，从而$\overline{X}$是$\mu$的最小方差无偏估计量.

现在讨论σ^2的无偏估计的方差下界. 由于

$$\frac{\partial}{\partial\theta^2}\ln f(x;\mu,\sigma^2)=-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}(x-\mu)^2,$$

2.2.2 有效估计和最小方差估计

$$\frac{\partial^2}{\partial(\sigma^2)^2}\ln f(x;\mu,\sigma^2)=\frac{1}{2\sigma^4}-\frac{(x-\mu)^2}{\sigma^6},$$

故由(2.2-17)式得 $$I(\sigma^2)=-\left(\frac{1}{2\sigma^4}-\frac{1}{\sigma^4}\right)=\frac{1}{2\sigma^4}.$$

因此，σ^2 的无偏估计的方差下界是 $\frac{2}{n}\sigma^4$.

当 μ 已知时，σ^2 的极大似然估计量 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ 是无偏的. 由于 $\sum_{i=1}^{n}\left(\frac{X^i-\mu}{\sigma}\right)^2\sim\chi^2(n)$，所以由(1.3-7)式知，

$$D\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right]=2n,\quad D\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]=2n\sigma^4.$$

于是，$$D(\hat{\sigma}^2)=\frac{1}{n^2}D\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]=\frac{2}{n}\sigma^4,$$

即 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ 的方差达到 Rao-Cramér 下界，因此它是 σ^2 的最小方差无偏估计量.

2.2.2 有效估计和最小方差估计

若 μ 未知，则样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是 σ^2 的无偏估计．由于$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2\sim\chi^2(n-1)$，所以由(1.3-28)式知，

$$D\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]=2(n-1),\quad D\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]=2(n-1)\sigma^4.$$

于是，
$$D(S^2)=D\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)=\frac{1}{(n-1)^2}D\left(\sum_{i=1}^{n}(X_i-\overline{X})^2\right)$$
$$=\frac{1}{(n-1)^2}\cdot 2(n-1)\sigma^4=\frac{2}{n-1}\sigma^4>\frac{2}{n}\sigma^4,$$

即 S^2 的方差没有达到 Rao-Cramér 下界． ∎

定义 2.2-7 设 $\hat{g}(X_1,X_2,\cdots,X_n)$是待估函数 $g(\theta)$的一个无偏估计量，称

$$e(\hat{g})=\frac{[g'(\theta)]^2}{nI(\theta)}\Big/D(\hat{g}) \tag{2.2-18}$$

为无偏估计量 $\hat{g}$ 的**效率**．显然 $e(\hat{g})\leqslant 1$，如果 $e(\hat{g})=1$，则称 $\hat{g}$ 为 g 的**优效估计**；如果 $\hat{g}$ 满足 $\lim\limits_{n\to+\infty}e(\hat{g})=1$，则称 $\hat{g}$ 为 g 的**渐近优效估计**．

由例 9 的结果可知，对于正态总体 $N(\mu;\sigma^2)$，样本均值 $\overline{X}$ 是 μ 的优效估计，样本方差 S^2 是 σ^2 的渐近优效估计．

作业（第2部分）

习题 2

1. 设样本(1.3,0.6,1.7,2.2,0.3,1.1)是取自具有概率密度 $f(x;\beta)=\begin{cases}\dfrac{1}{\beta}, & 0<x<\beta \\ 0, & \text{其余}\end{cases}$ 的总体，用矩估计法估计总体均值、总体方差及参数 β.

4. 设总体 X 服从对数正态分布 $LN(\mu;\sigma^2)$，其概率函数为

$$f(x;\mu,\sigma^2)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\dfrac{1}{2\sigma^2}(\ln x-\mu)^2\right\}, & x>0, \\ 0, & x\leqslant 0,\end{cases}$$

其中 μ,σ^2 均未知，$(X_1,X_2,\cdots,X_n)$是取自这个总体的样本，求 μ 与 σ^2 的极大似然估计量.

8. 设(X_1,X_2,X_3)是取自总体 X 的样本，证明下列统计量都是总体均值$E(X)$的无偏估计量：

$$t_1(X_1,X_2,X_3)=\frac{2}{5}X_1+\frac{1}{5}X_2+\frac{2}{5}X_3;$$

$$t_2(X_1,X_2,X_3)=\frac{1}{6}X_1+\frac{1}{3}X_2+\frac{1}{2}X_3,$$

$$t_3(X_1,X_2,X_3)=\frac{1}{7}X_1+\frac{3}{14}X_2+\frac{9}{14}X_3.$$

并问哪一个无偏估计量的方差最小？

謝謝觀看！

厦門大學
XIAMEN UNIVERSITY

信息学院
（国家示范性软件学院）
School of Informatics

黄 煒
博士，副教授
Dr. Wei Huang