高等工程數學 (4)

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

线性方程组

矩阵论（3）

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

1.4 线性方程组

设线性方程组为

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \end{cases} \tag{1.4-1}$$

其中 $a_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 和 $b_i(i=1,2,\cdots,m)$ 都是实数. 记

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

则线性方程组(1.4-1)可写成

$$\boldsymbol{Ax} = \boldsymbol{b}, \tag{1.4-1'}$$

其中 $\boldsymbol{A}$ 称为此线性方程组的**系数矩阵**.

1.4 线性方程组

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.

1.4.1 线性方程组解的存在定理

由于(1.4-1′)式可写成向量形式

$$x_1\boldsymbol{A}_1+x_2\boldsymbol{A}_2+\cdots+x_n\boldsymbol{A}_n=\boldsymbol{b},$$

所以这个线性方程组的解是向量 $\boldsymbol{b}$ 用 $\boldsymbol{A}$ 的列向量 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n$ 线性表出的系数.因此线性方程组(1.4-1)有解的充要条件是,$\boldsymbol{b}$ 可用向量组 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n$ 线性表出,也即向量组 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n$ 与向量组 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n,\boldsymbol{b}$ 等价.于是,下面的定理给出了非齐次线性方程组(1.4-1)有解的充要条件及解的个数.

定理 1.4-1　非齐次线性方程组(1.4-1)有解的充要条件是,系数矩阵 $\boldsymbol{A}$ 的秩等于其增广矩阵 $\boldsymbol{B}$ 的秩.并且在 $r(\boldsymbol{A})=r(\boldsymbol{B})=r$ 时,若 $r=n$,则有唯一解;而若 $r<n$,则有无穷多解,且在其一般解中有 $n-r$ 个自由未知量. ∎

将这个定理应用于齐次线性方程组,便得以下定理.

定理 1.4-2　$\boldsymbol{Ax}=\boldsymbol{0}$有非零解的充要条件是

$$r(\boldsymbol{A})=r<n. \quad ∎$$

特别地,若 $\boldsymbol{A}$ 是 n 阶方阵,则 $\boldsymbol{Ax}=\boldsymbol{0}$有非零解的充要条件是$|\boldsymbol{A}|=0$,即 $\boldsymbol{A}$ 是不可逆的.

推论　设 $\boldsymbol{A}$ 是 $m\times n$ 矩阵,且 $m<n$,则齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{0}$ 必有非零解.

1.4.1 线性方程组解的存在定理

例 1 问 a 和 b 取何值时，线性方程组

$$\begin{cases} x_1+x_2+x_3+x_4=0, \\ x_2+2x_3+2x_4=1, \\ -x_2+(a-3)x_3-2x_4=b, \\ 3x_1+2x_2+x_3+ax_4=-1 \end{cases}$$

有唯一解、无解、无穷多解？并求出无穷多解时的一般解.

解 对方程组的增广矩阵施行行初等变换：

$$\boldsymbol{B}=[\boldsymbol{A} \vdots \boldsymbol{b}]=\left[\begin{array}{cccc:c} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 2 & 1 \\ 0 & -1 & a-3 & -2 & b \\ 3 & 2 & 1 & a & -1 \end{array}\right] \to \left[\begin{array}{cccc:c} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 2 & 1 \\ 0 & -1 & a-3 & -2 & b \\ 0 & -1 & -2 & a-3 & -1 \end{array}\right]$$

$$\to \left[\begin{array}{cccc:c} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 2 & 1 \\ 0 & 0 & a-1 & 0 & b+1 \\ 0 & 0 & 0 & a-1 & 0 \end{array}\right].$$

1.4.1 线性方程组解的存在定理

若令 $x_3=k_1, x_4=k_2$，则得到原方程组的一般解是

$$x_1=-1+k_1+k_2,\quad x_2=1-2k_1-2k_2,\quad x_3=k_1,\quad x_4=k_2.$$

写成向量形式为

$$\begin{bmatrix} x_1\\ x_2\\ x_3\\ x_4 \end{bmatrix}=\begin{bmatrix} -1\\ 1\\ 0\\ 0 \end{bmatrix}+k_1\begin{bmatrix} 1\\ -2\\ 1\\ 0 \end{bmatrix}+k_2\begin{bmatrix} 1\\ -2\\ 0\\ 1 \end{bmatrix},$$

其中 k_1、k_2 为任意实数. $\blacksquare$

1.4.2 线性方程组解的结构

对于齐次线性方程组

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0},$$

容易验证，若 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 均为它的解向量，则 $\boldsymbol{x}_1+\boldsymbol{x}_2$ 也是它的解向量，且 $k\boldsymbol{x}_1$ 亦是它的解向量，其中 k 为任意实数. 因此，若用 $\mathcal{N}(\boldsymbol{A})$ 表示 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的全体解向量所组成的集合，则 $\mathcal{N}(\boldsymbol{A})$ 是 $\mathbf{R}^n$ 的子空间，称为齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的**解空间**. 容易看出，若 $r(\boldsymbol{A})=r<n$，则 $\dim\mathcal{N}(\boldsymbol{A})=n-r$，即解空间 $\mathcal{N}(\boldsymbol{A})$ 是 $n-r$ 维向量空间，亦即它的一个基中含有 $n-r$ 个线性无关的解向量.

设 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_{n-r}$ 是齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 之解空间 $\mathcal{N}(\boldsymbol{A})$ 的一个基，则称这个线性无关的解向量组 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_{n-r}$ 为此齐次线性方程组的一个**基础解系**. 因而，$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的任何一个解都可用这个基础解系线性表出. 于是，齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的一般解为

$$\boldsymbol{x}=k_1\boldsymbol{\alpha}_1+\cdots+k_{n-r}\boldsymbol{\alpha}_{n-r}, \tag{1.4-2}$$

其中 $k_1,\cdots,k_{n-r}$ 为任意实数.

1.4.2 线性方程组解的结构

例 2　设四元齐次线性方程组(Ⅰ)为

$$\begin{cases} x_1+2x_2-x_4=0, \\ 2x_1+x_2+x_4=0, \\ x_1-x_2+2x_4=0, \end{cases}$$

又已知某齐次线性方程组(Ⅱ)的一般解为

$$k_1[0,1,1,0]^{\mathrm{T}}+k_2[-1,2,2,1]^{\mathrm{T}}.$$

(1) 求线性方程组(Ⅰ)的基础解系；

(2) 问线性方程组(Ⅰ)与(Ⅱ)是否有非零公共解？若有，则求出所有的非零公共解. 若没有，则说明理由.

解　(1)对方程组(Ⅰ)的系数矩阵施以行初等变换：

$$\boldsymbol{A}=\begin{bmatrix} 1 & 2 & 0 & -1 \\ 2 & 1 & 0 & 1 \\ 1 & -1 & 0 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & -3 & 0 & 3 \\ 0 & -3 & 0 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

因此方程组(Ⅰ)的基础解系中含有 $4-r(\boldsymbol{A})=4-2=2$ 个线性无关的解向量，取

$$\boldsymbol{\alpha}_1=[0,0,1,0]^{\mathrm{T}},\quad \boldsymbol{\alpha}_2=[-1,1,0,1]^{\mathrm{T}}.$$

(2) 将方程组(Ⅱ)的解

$$[x_1,x_2,x_3,x_4]^{\mathrm{T}}=[-k_2,k_1+2k_2,k_1+2k_2,k_2]^{\mathrm{T}}$$

1.4.2 线性方程组解的结构

代入方程组（Ⅰ），得
$$\begin{cases} -k_2+2(k_1+2k_2)-k_2=0, \\ -2k_2+(k_1+2k_2)+k_2=0, \\ -k_2-(k_1+2k_2)+2k_2=0. \end{cases}$$
解之得 $k_1=-k_2$. 显然，当 $k_1=-k_2\neq 0$ 时，方程组（Ⅱ）的解为
$$[x_1,x_2,x_3,x_4]=k_2[-1,1,1,1]^{\mathrm{T}} \quad (k_2\neq 0).$$
而它即是方程组（Ⅰ）与（Ⅱ）的非零公共解.

于是，方程组（Ⅰ）与（Ⅱ）有非零公共解，且所有的非零公共解是 $k[-1,1,1,1]^{\mathrm{T}}$（k 是不为零的任意实数）. ∎

非齐次线性方程组(1.4-1′)的解具有性质：

(1) 若 $\boldsymbol{x}=\boldsymbol{\eta}_1$ 和 $\boldsymbol{x}=\boldsymbol{\eta}_2$ 都是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解，则 $\boldsymbol{x}=\boldsymbol{\eta}_1-\boldsymbol{\eta}_2$ 是对应齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{0}$ 的解；

(2) 若 $\boldsymbol{x}=\boldsymbol{\eta}$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解，$\boldsymbol{x}=\boldsymbol{\xi}$ 是 $\boldsymbol{Ax}=\boldsymbol{0}$ 的解，则 $\boldsymbol{x}=\boldsymbol{\eta}+\boldsymbol{\xi}$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解.

由此不难推出非齐次线性方程组(1.4-1′)解的结构.

定理 1.4-3 若线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的系数矩阵 $\boldsymbol{A}$ 与增广矩阵 $\boldsymbol{B}=[\boldsymbol{A} \vdots \boldsymbol{b}]$ 的秩相等为 r，且 $r<n$，而 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_{n-r}$ 是对应齐次线性方程组的基础解系，$\boldsymbol{\eta}$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的任意一个特解，那么 $\boldsymbol{Ax}=\boldsymbol{b}$ 的一般解为
$$\boldsymbol{x}=\boldsymbol{\eta}+k_1\boldsymbol{\alpha}_1+\cdots+k_{n-r}\boldsymbol{\alpha}_{n-r}, \tag{1.4-3}$$
其中 $k_1,\cdots,k_{n-r}$ 是任意实数. ∎

1.4.2 线性方程组解的结构

例 3 设 $\boldsymbol{A}$ 是 4 阶方阵，$r(\boldsymbol{A})=2$，$\boldsymbol{b}$ 是 4 维向量. 已知 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3,\boldsymbol{\eta}_4$ 都是线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解，且满足

$$\boldsymbol{\eta}_1+\boldsymbol{\eta}_2=[2,4,0,8]^{\mathrm{T}},\quad 2\boldsymbol{\eta}_2+\boldsymbol{\eta}_3=[3,0,3,3]^{\mathrm{T}},\quad 3\boldsymbol{\eta}_3+\boldsymbol{\eta}_4=[2,1,0,1]^{\mathrm{T}},$$

求这个线性方程组的一般解.

解 由于

$$\boldsymbol{A}\left[\frac{1}{2}(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)\right]=\frac{1}{2}(\boldsymbol{A\eta}_1+\boldsymbol{A\eta}_2)=\frac{1}{2}(\boldsymbol{b}+\boldsymbol{b})=\boldsymbol{b},$$

$$\boldsymbol{A}\left[\frac{1}{3}(2\boldsymbol{\eta}_2+\boldsymbol{\eta}_3)\right]=\frac{1}{3}(2\boldsymbol{A\eta}_2+\boldsymbol{A\eta}_3)=\frac{1}{3}(2\boldsymbol{b}+\boldsymbol{b})=\boldsymbol{b},$$

$$\boldsymbol{A}\left[\frac{1}{4}(3\boldsymbol{\eta}_3+\boldsymbol{\eta}_4)\right]=\frac{1}{4}(3\boldsymbol{A\eta}_3+\boldsymbol{A\eta}_4)=\frac{1}{4}(3\boldsymbol{b}+\boldsymbol{b})=\boldsymbol{b},$$

所以 $\frac{1}{2}(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)$，$\frac{1}{3}(2\boldsymbol{\eta}_2+\boldsymbol{\eta}_3)$，$\frac{1}{4}(3\boldsymbol{\eta}_3+\boldsymbol{\eta}_4)$ 都是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解. 于是

$$\boldsymbol{\alpha}_1=\frac{1}{2}(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)-\frac{1}{3}(2\boldsymbol{\eta}_2+\boldsymbol{\eta}_3)=\frac{1}{2}[2,4,0,8]^{\mathrm{T}}-\frac{1}{3}[3,0,3,3]^{\mathrm{T}}=[0,2,-1,3]^{\mathrm{T}},$$

$$\boldsymbol{\alpha}_2=\frac{1}{3}(2\boldsymbol{\eta}_2+\boldsymbol{\eta}_3)-\frac{1}{4}(3\boldsymbol{\eta}_3+\boldsymbol{\eta}_4)=\frac{1}{3}[3,0,3,3]^{\mathrm{T}}-\frac{1}{4}[2,1,0,1]^{\mathrm{T}}=\left[\frac{1}{2},-\frac{1}{4},1,\frac{3}{4}\right]^{\mathrm{T}}$$

1.4.2 线性方程组解的结构

都是对应齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的解. 因 $r(\boldsymbol{A})=2$，所以 $4-r(\boldsymbol{A})=2$，且 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$ 线性无关，故 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$ 是 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$的基础解系. 于是 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的一般解为

$$\boldsymbol{x}=\frac{1}{2}(\boldsymbol{\eta}_1+\boldsymbol{\eta}_4)+k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2=[1,2,0,4]^{\mathrm{T}}+k_1[0,2,-1,3]^{\mathrm{T}}+k_2\left[\frac{1}{2},\ -\frac{1}{4},\ 1,\ \frac{3}{4}\right]^{\mathrm{T}},$$

其中 k_1，k_2 为任意实数.

作业（第1部分）

习题 1.4

2. λ 取何值时，线性方程组

$$\begin{cases} (2-\lambda)x_1 + 2x_2 - 2x_3 = 1, \\ 2x_1 + (5-\lambda)x_2 - 4x_3 = 2, \\ -2x_1 - 4x_2 + (5-\lambda)x_3 = -\lambda - 1. \end{cases}$$

误差的基本知识

数理统计（1）

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

1.2 数值计算的误差估计及算法稳定性

数值计算中误差传播情况比较复杂，要对每一步计算的误差进行精确估计难以做到，因而通常采用微分误差分析方法估计误差，即误差较小时忽略二阶及二阶以上的误差高阶小量.

设原始数据 $x_1, x_2, \cdots, x_n$ 是相互独立的变量值，把通过算式所得到的结果 y 看成是 $x_1, x_2, \cdots, x_n$ 的函数值，即

$$y = f(x_1, x_2, \cdots, x_n).$$

若 $x_1^*, x_2^*, \cdots, x_n^*$ 依次是 $x_1, x_2, \cdots, x_n$ 的近似值，那么在假设按算式计算本身没有误差的情况下，结果 y 的近似值 y^* 为

$$y^* = f(x_1^*, \cdots, x_n^*).$$

当 $|e_i| = |x_i^* - x_i|\,(i=1,\cdots,n)$ 都很小时，有

$$\begin{aligned} e_y &= y^* - y = f(x_1^*, \cdots, x_n^*) - f(x_1, \cdots, x_n) \\ &= f(x_1^*, \cdots, x_n^*) - \Big[f(x_1^*, \cdots, x_n^*) + \sum_{i=1}^{n} \frac{\partial f(x_1^*, \cdots, x_n^*)}{\partial x_i}(x_i - x_i^*) \\ &\quad + \text{二阶及二阶以上的误差高阶小量}\Big] \\ &\approx \sum_{i=1}^{n} \frac{\partial f(x_1^*, \cdots, x_n^*)}{\partial x_i}(x_i^* - x_i). \end{aligned} \qquad (1.2\text{-}1)$$

1.2 数值计算的误差估计及算法稳定性

由(1.2-1)式得到近似值 y^* 的绝对误差限

$$\varepsilon_y \leqslant \sum_{i=1}^{n} \left| \frac{\partial f(x_1^*, \cdots, x_n^*)}{\partial x_i} \right| \varepsilon_{x_i} \tag{1.2-2}$$

和相对误差限

$$(\varepsilon_y)_r = \frac{\varepsilon_y}{|y^*|} \leqslant \sum_{i=1}^{n} \left| \frac{1}{y^*} \left(\frac{\partial f}{\partial x_i} \right)_* \right| \varepsilon_{x_i} = \sum_{i=1}^{n} \left| \frac{x_i^*}{y^*} \cdot \left(\frac{\partial f}{\partial x_i} \right)_* \right| (\varepsilon_{x_i})_r, \tag{1.2-3}$$

其中 $\left(\frac{\partial f}{\partial x_i} \right)_* = \frac{\partial f(x_1^*, \cdots, x_n^*)}{\partial x_i}$.

值得指出的是，不等式(1.2-2)和(1.2-3)中的等号是可能取到的，因为 $\left(\frac{\partial f}{\partial x_i} \right)_*$ 与 $x_i^* - x_i$ 可能同为正值或负值.

从(1.2-2)式和(1.2-3)式看出，

$$\left| \left(\frac{\partial f}{\partial x_i} \right)_* \right|, \left| \frac{x_i^*}{y^*} \left(\frac{\partial f}{\partial x_i} \right)_* \right| \quad (i = 1, 2, \cdots, n)$$

分别给出了 x_i^* 的绝对误差对 y^* 的绝对误差和 x_i^* 的相对误差对 y^* 的相对误差的影响程度，它们的大小对误差传播起着重要的作用.

1.2 数值计算的误差估计及算法稳定性

我们称

$$\left|\frac{x_i^*}{y^*}\left(\frac{\partial f}{\partial x_i}\right)_*\right| \quad (i=1,2,\cdots,n)$$

为该问题(即计算 $y=f(x_1,\cdots,x_n)$)的**条件数**.

当条件数 $\left|\frac{x_i^*}{y^*}\left(\frac{\partial f}{\partial x_i}\right)_*\right|$ 很大时,即使 x_i^* 的相对误差 $|(e_{x_i})_r|$ 很小,也可能使 y^* 的相对误差很大,这时计算 $y=f(x_1,\cdots,x_n)$ 的问题称为**病态的**,而当这 n 个条件数都不大时,原始数据的相对误差对 y^* 的相对误差影响不大,故称该问题是**良态的**.

一般地,若问题是由输入数据 $x_1,x_2,\cdots,x_n$ 计算所要求的结果 $y_1,y_2,\cdots,y_m$,记

$$\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}, \quad \boldsymbol{y}=[y_1,y_2,\cdots,y_m]^{\mathrm{T}},$$

则所要解决的问题归结为计算向量值函数

$$\boldsymbol{y}=\boldsymbol{f}(\boldsymbol{x})$$

的值,其中 $\boldsymbol{f}(\boldsymbol{x})=[f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x})]^{\mathrm{T}}$,即

$$y_i=f_i(x_1,x_2,\cdots,x_n), \quad i=1,2,\cdots,m.$$

设 $\boldsymbol{x}^*$ 是 $\boldsymbol{x}$ 的近似值,则在假设计算过程不产生误差的条件下,由 $\boldsymbol{x}^*$ 算出的 $\boldsymbol{y}$ 的值为

$$\boldsymbol{y}^*=\boldsymbol{f}(\boldsymbol{x}^*).$$

1.2 数值计算的误差估计及算法稳定性

因此，由微分误差分析方法估计近似值 $\boldsymbol{y}^*$ 的误差向量(因 $\boldsymbol{y}^*-\boldsymbol{y}$ 是 m 维向量，故称为误差向量)为

$$\boldsymbol{y}^*-\boldsymbol{y}=\boldsymbol{f}(\boldsymbol{x}^*)-\boldsymbol{f}(\boldsymbol{x})=D\boldsymbol{f}(\boldsymbol{x}^*)(\boldsymbol{x}^*-\boldsymbol{x}), \tag{1.2-4}$$

其中 $m\times n$ 矩阵 $D\boldsymbol{f}(\boldsymbol{x}^*)$ 为 $\boldsymbol{f}(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处的 Jacobi 矩阵：

$$D\boldsymbol{f}(\boldsymbol{x}^*)=\begin{bmatrix}\dfrac{\partial f_1(\boldsymbol{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial f_1(\boldsymbol{x}^*)}{\partial x_n}\\ \vdots & & \vdots\\ \vdots & & \vdots\\ \dfrac{\partial f_m(\boldsymbol{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial f_m(\boldsymbol{x}^*)}{\partial x_n}\end{bmatrix}. \tag{1.2-5}$$

由(1.2-4)式得到

$$\|\boldsymbol{y}^*-\boldsymbol{y}\|=\|D\boldsymbol{f}(\boldsymbol{x}^*)(\boldsymbol{x}^*-\boldsymbol{x})\|\leqslant\|D\boldsymbol{f}(\boldsymbol{x}^*)\|\cdot\|\boldsymbol{x}^*-\boldsymbol{x}\|, \tag{1.2-6}$$

$$\frac{\|\boldsymbol{y}^*-\boldsymbol{y}\|}{\|\boldsymbol{y}^*\|}=\frac{\|D\boldsymbol{f}(\boldsymbol{x}^*)(\boldsymbol{x}^*-\boldsymbol{x})\|}{\|\boldsymbol{y}^*\|}\leqslant\frac{\|\boldsymbol{x}^*\|}{\|\boldsymbol{y}^*\|}\|D\boldsymbol{f}(\boldsymbol{x}^*)\|\frac{\|\boldsymbol{x}^*-\boldsymbol{x}\|}{\|\boldsymbol{x}^*\|}. \tag{1.2-7}$$

1.2 数值计算的误差估计及算法稳定性

从(1.2-6)式和(1.2-7)式看出，

$$\|Df(x^*)\|, \quad \frac{\|x^*\|}{\|y^*\|}\|Df(x^*)\|$$

分别给出了 x^* 的绝对误差对 y^* 的绝对误差和 x^* 的相对误差对 y^* 的相对误差的影响程度. 自然地,

$$\frac{\|x^*\|}{\|y^*\|}\|Df(x^*)\|$$

就是这个问题的条件数.

例 1 设 $y=-a+\sqrt{a^2+b}$，则

$$\frac{\partial y}{\partial a}=-1+\frac{a}{\sqrt{a^2+b}}=\frac{-y}{\sqrt{a^2+b}}, \quad \frac{\partial y}{\partial b}=\frac{1}{2\sqrt{a^2+b}},$$

从而由(1.2-3)式得

$$(\varepsilon_y)_r \leqslant \left|\frac{a^*}{\sqrt{(a^*)^2+b^*}}\right|(\varepsilon_a)_r+\left|\frac{a^*+\sqrt{(a^*)^2+b^*}}{2\sqrt{(a^*)^2+b^*}}\right|(\varepsilon_b)_r.$$

当 $b^*>0$ 时，由于 $\left|\frac{a^*}{\sqrt{(a^*)^2+b^*}}\right|<1$，$\left|\frac{a^*+\sqrt{(a^*)^2+b^*}}{2\sqrt{(a^*)^2+b^*}}\right|<1$，所以问题是良态的. 但是，若 $b^*\approx-(a^*)^2$，则问题是病态的. ∎

1.2 数值计算的误差估计及算法稳定性

例 2 设 $y=f(x)=Ax$，其中 A 是 n 阶可逆矩阵，则由于 $x^*=A^{-1}y^*$ 得 $\|x^*\|_p=\|A^{-1}y^*\|_p\leqslant\|A^{-1}\|_p\|y^*\|_p$ 和 $Df(x^*)=A$，所以(1.2-7)式给出

$$\frac{\|y^*-y\|_p}{\|y^*\|_p}\leqslant\frac{\|x^*\|_p}{\|y^*\|_p}\|Df(x^*)\|_p\frac{\|x^*-x\|_p}{\|x^*\|_p}$$

$$\leqslant\|A^{-1}\|_p\|A\|_p\cdot\frac{\|x^*-x\|_p}{\|x^*\|_p}.$$

因此该问题的条件数是 $\|A^{-1}\|_p\|A\|_p$，此即方阵 A 的条件数. ∎

对于加减乘除四则运算来说，容易验证

$$\varepsilon_r(x^*\pm y^*)\leqslant\left|\frac{x^*}{x^*\pm y^*}\right|\varepsilon_r(x^*)+\left|\frac{y^*}{x^*\pm y^*}\right|\varepsilon_r(y^*),$$

$$\varepsilon_r(x^*y^*)\leqslant\varepsilon_r(x^*)+\varepsilon_r(y^*),$$

$$\varepsilon_r\left(\frac{x^*}{y^*}\right)\leqslant\varepsilon_r(x^*)+\varepsilon_r(y^*),$$

这里 $\varepsilon_r(z^*)$ 表示近似值 z^* 的相对误差限. 因此，乘除的相对误差传播是不强的，但对于加减运算，当 $x^*\approx-y^*$ ($x^*\approx y^*$) 时，加法(减法)使问题的条件数变得很大，产生灾难性的对消，消去了大量有效数字.

1.2 数值计算的误差估计及算法稳定性

如果计算公式比较复杂，计算步骤很多，用上述方法一步一步地估计误差，不仅困难很大，而且也不实用. 因为在众多步骤的计算中，每步产生的误差有大有小，有正也有负，很可能有一些互相抵消. 如果每一步都按最保险的方式估计，不仅工作量大得几乎无法进行，而且得到的误差限也比实际的大得多，因而失去实用价值.

对于计算步骤较多的一个大的算法，一般不用上述定量估计办法，而采用定性分析方法，即讨论算法的数值稳定性问题.

在介绍算法的数值稳定性概念之前，首先了解一下为什么用不同算法解决同一个问题，一般会产生不同的效果.

一个算法的“好坏”可以从不同的角度来评价，但主要是算法的计算效率和它的可靠性.

所谓算法的计算效率，是指解一个问题所花费的代价，通常是以电子数字计算机上运行该算法所用的时间来衡量的. 由于在浮点机上完成一次乘或除比完成一次加或减所耗的机时要多得多，所以当一个算法中加减运算与乘除运算的次数差不多时，可以用乘除次数来体现算法的效率.

1.2 数值计算的误差估计及算法稳定性

例 3　设 $\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\boldsymbol{A}_4$ 依次是 $10\times20,20\times50,50\times1,1\times100$ 的矩阵，比较一下按不同运算顺序求矩阵乘积 $\boldsymbol{P}=\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4$ 的工作量.

解　我们以三种运算顺序为例说明其计算 $\boldsymbol{P}$ 的工作量：

1) $\boldsymbol{P}=[(\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{A}_3]\boldsymbol{A}_4$；

2) $\boldsymbol{P}=\boldsymbol{A}_1[\boldsymbol{A}_2(\boldsymbol{A}_3\boldsymbol{A}_4)]$；

3) $\boldsymbol{P}=[\boldsymbol{A}_1(\boldsymbol{A}_2\boldsymbol{A}_3)]\boldsymbol{A}_4$.

由于计算 $\boldsymbol{A}_1\boldsymbol{A}_2$ 的每个元素要做 20 次乘法和 19 次加法，所以计算 $\boldsymbol{A}_1\boldsymbol{A}_2$ 总共需要做 $10\times20\times50=10000$ 次乘法和 $10\times19\times50=9500$ 次加法. 因为加法次数与乘法次数差不多，故以下只考虑乘法次数. 由于 $\boldsymbol{A}_1\boldsymbol{A}_2$ 是 10×50 矩阵，所以计算 $(\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{A}_3$ 总共需要做 $10\times20\times50+10\times50\times1=10500$ 次乘法. 又因 $(\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{A}_3$ 是 10×1 矩阵，故计算 $\boldsymbol{P}=[(\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{A}_3]\boldsymbol{A}_4$ 总共需要做 $10\times20\times50+10\times50\times1+10\times1\times100=11500$ 次乘法.

类似地，可算出 $\boldsymbol{A}_1[\boldsymbol{A}_2(\boldsymbol{A}_3\boldsymbol{A}_4)]$ 所需要做的乘法次数为 125000 次，而 $[\boldsymbol{A}_1(\boldsymbol{A}_2\boldsymbol{A}_3)]\boldsymbol{A}_4$ 只需要做 2200 次乘法. 可见，第三种算法的工作量小，亦即效率高. ■

1.2 数值计算的误差估计及算法稳定性

算法的可靠性问题更困难，一般都需要进行误差分析，特别是研究计算中含入误差的传播.

如果用 $fl(x)$表示 x 的浮点近似，那么舍入误差为 $fl(x)-x$，而 $fl(x)$的相对误差是 $\dfrac{fl(x)-x}{x}$，记它为 δ_x，则有

$$\delta_x=\frac{fl(x)-x}{x},$$

即

$$fl(x)=(1+\delta_x)x. \tag{1.2-8}$$

根据(1.2-8)式，加减乘除四则运算产生的舍入误差为

$$\begin{cases} fl(x\pm y)-(x\pm y)=\delta_1(x\pm y), \\ fl(xy)-xy=\delta_2(xy), \\ fl\left(\dfrac{x}{y}\right)-\dfrac{x}{y}=\delta_3\dfrac{x}{y}, \end{cases} \tag{1.2-9}$$

其中 $\delta_i(i=1,2,3)$是由所用电子数字计算机的机器精度决定的相对(舍入)误差. 若用 eps 表示所用数字计算机的精度，那么有$|\delta_i|\leqslant \mathrm{eps}$.

1.2 数值计算的误差估计及算法稳定性

以求和 $y=x_1+x_2+x_3$ 为例，考虑舍入误差的影响. 算法 $y=(x_1+x_2)+x_3$ 可分解为两个步骤，先计算 $\eta=x_1+x_2$，再计算 $y=\eta+x_3$. 根据(1.2-8)式，实际在数字计算机上得到的数值结果是

$$\begin{aligned} fl(\eta) &= (1+\delta_1)(x_1+x_2), \\ fl(y) &= (1+\delta_2)(fl(\eta)+x_3) \\ &= (1+\delta_2)[(1+\delta_1)(x_1+x_2)+x_3] \\ &= (x_1+x_2+x_3)\left[1+\frac{x_1+x_2}{x_1+x_2+x_3}\delta_1(1+\delta_2)+\delta_2\right]. \end{aligned}$$

于是，$fl(y)$的相对误差是

$$(e_y)_r = \frac{fl(y)-y}{y} = \frac{x_1+x_2}{x_1+x_2+x_3}\delta_1(1+\delta_2)+\delta_2.$$

按照微分误差分析方法忽略二阶及二阶以上的误差高阶小量，则有

$$(e_y)_r = \frac{x_1+x_2}{x_1+x_2+x_3}\delta_1+1\cdot\delta_2. \tag{1.2-10}$$

从(1.2-10)式看出，放缩因子 $\dfrac{x_1+x_2}{x_1+x_2+x_3}$ 与 1 分别度量了相对误差 δ_1、δ_2 对 y 的相对误差的影响. 因此，要根据是 $|x_1+x_2|$ 还是 $|x_2+x_3|$ 较小来选用 $y=(x_1+x_2)+x_3$ 还是 $y=x_1+(x_2+x_3)$ 作为计算 $y=x_1+x_2+x_3$ 的算法.

1.2 数值计算的误差估计及算法稳定性

例 4 计算 $y=a^2-b^2$ 有两个算法：

算法Ⅰ为 $\eta_1=a\times a, \eta_2=b\times b, y=\eta_1-\eta_2$；

算法Ⅱ为 $\xi_1=a+b, \xi_2=a-b, y=\xi_1\times\xi_2$.

讨论舍入误差对它们的影响.

解 对于算法Ⅰ，

$$fl(\eta_1)=(1+\delta_1)a^2,\quad fl(\eta_2)=(1+\delta_2)b^2,$$

$$fl(y)=(1+\delta_3)[fl(\eta_1)-fl(\eta_2)]=(1+\delta_3)[(1+\delta_1)a^2-(1+\delta_2)b^2].$$

于是，$fl(y)$的相对误差（忽略误差的二阶项）是

$$(e_y)_r=\frac{1}{a^2-b^2}[a^2\delta_1-b^2\delta_2+(a^2-b^2)\delta_3],$$

从而相对误差限为

$$(\varepsilon_y)_r\leqslant\frac{1}{|a^2-b^2|}(a^2+b^2+|a^2-b^2|)\text{eps}. \tag{1.2-11}$$

对于算法Ⅱ，$fl(\xi_1)=(1+\delta_4)(a+b), fl(\xi_2)=(1+\delta_5)(a-b)$，

$$fl(y)=(1+\delta_6)fl(\xi_1)\cdot fl(\xi_2)=(1+\delta_4)(1+\delta_5)(1+\delta_6)(a^2-b^2).$$

从而相对误差限为

$$(\varepsilon_y)_r\leqslant\frac{1}{|a^2-b^2|}(|\delta_4|+|\delta_5|+|\delta_6|)|a^2-b^2|\leqslant\frac{3}{|a^2-b^2|}|a^2-b^2|\text{eps}. \tag{1.2-12}$$

1.2 数值计算的误差估计及算法稳定性

比较(1.2-11)式与(1.2-12)式可以看出，当$a^2+b^2>2|a^2-b^2|$，即$\frac{1}{3}<\left(\frac{a}{b}\right)^2<3$时，算法Ⅱ比算法Ⅰ数值上更可靠些，因为它的相对误差较小．而当$a^2+b^2<2|a^2-b^2|$时，算法Ⅰ比算法Ⅱ数值上更可靠些． $\blacksquare$

例如，$a=0.3237$，$b=0.3134$，用四位有效数字计算a^2-b^2，可以得到下述结果：

算法Ⅰ　$a\times^* a=0.1048$，$b\times^* b=0.9822\times10^{-1}$，

$$(a\times^* a)-^*(b\times^* b)=0.66\times10^{-2};$$

算法Ⅱ　$a+^* b=0.6371$，$a-^* b=0.1030\times10^{-1}$，

$$(a+^* b)(a-^* b)=0.6562\times10^{-2},$$

其中$+^*$，$-^*$，$\times^*$表示浮点数的加、减、乘运算．a^2-b^2的准确值是0.656213×10^{-2}，可见算法Ⅱ的结果比算法Ⅰ的结果可靠．

由于$\frac{a}{b}=\frac{0.3237}{0.3134}\approx1.032865348$，满足$\frac{1}{3}<(\frac{a}{b})^2<3$，所以由理论分析也知，算法Ⅱ比算法Ⅰ可靠些．

1.2 数值计算的误差估计及算法稳定性

所谓数值稳定的算法是指，在数字计算机执行这个数值算法的过程中，产生的舍入误差能够被控制在一定范围内，并对最终的结果影响不大. 如果计算过程中的舍入误差不断增大，使最终结果与准确值相差较大，这样的算法就是数值不稳定的算法. 具体来说，在一个算法的计算过程中，如果某一步有了舍入误差，其绝对值不大于 ε，而以后各步的计算都假定是准确地进行的，不产生新的舍入误差，那么仅由该步的舍入误差所引起的误差的绝对值始终不超过 ε 的某个不大的倍数，则这个算法是数值稳定的. 这实际上是把算法的数值稳定性的研究，转化为数据误差对算法影响的分析.

对于数值稳定的算法，一般认为其结果是可靠的，必要时再进一步进行定量的误差分析，而数值不稳定的算法则尽量不使用.

1.3 数值计算中应注意的一些原则

为了减小误差的影响，特别是舍入误差的影响，在数值运算中应注意以下一些原则.

1）用数值稳定性好的计算方法，以便控制舍入误差的传播.

例如，要求在四位有效数字的精度下计算定积分：

$$y_n=\int_0^1 \frac{x^n}{x+5}\mathrm{d}x,\quad n=0,1,\cdots,100.$$

由有理函数积分法知，

$$y_n=\int_0^1 \frac{x^n}{x+5}\mathrm{d}x=\int_0^1 \frac{x^n+5x^{n-1}-5x^{n-1}}{x+5}\mathrm{d}x$$

$$=\int_0^1 x^{n-1}\mathrm{d}x-\int_0^1 \frac{5x^{n-1}}{x+5}\mathrm{d}x=\frac{1}{n}-5y_{n-1},$$

因而计算这 101 个定积分的一个算法是

$$\begin{cases} y_0=\int_0^1 \frac{1}{x+5}\mathrm{d}x=\ln 6-\ln 5\approx 0.1823, \\ y_n=\frac{1}{n}-5y_{n-1}, \quad n=1,2,\cdots,100. \end{cases}$$

它是数值稳定性不好的算法.因为 y_0 的舍入误差传播到 y_1 时增大 5 倍.如此进行，传播到 y_{100} 时将增大 5^{100} 倍.

1.3 数值计算中应注意的一些原则

若采用另一个算法，利用估计式

$$\frac{1}{6(n+1)}=\int_0^1 \frac{x^n}{6}\mathrm{d}x < y_n = \int_0^1 \frac{x^n}{x+5}\mathrm{d}x < \int_0^1 \frac{x^n}{5}\mathrm{d}x = \frac{1}{5(n+1)},$$

取 y_{100} 的近似值为 $\frac{1}{2}\left(\frac{1}{606}+\frac{1}{505}\right)=0.1815\times10^{-2}$，再按下式计算这 101 个定积分：

$$\begin{cases} y_{100} = 0.1815\times10^{-2} \\ y_{n-1} = \dfrac{1}{5n} - \dfrac{1}{5}y_n, \quad n = 100,99,\cdots,1. \end{cases}$$

那么该算法具有很好的数值稳定性，因为 y_{100} 的误差在计算过程中每一步都要乘以 $\frac{1}{5}$.

2）两个数量级相差很大的数进行加减运算时，要防止小的那个数加减不到大的数中所引起的严重后果.

例如，在五位十进制计算机上，计算

$$S = 52492 + \sum_{i=1}^{1000} a_i, \quad 0.1 \leqslant a_i \leqslant 0.9.$$

1.3 数值计算中应注意的一些原则

如果该计算机是截位的，即凡超过五位的数字都自动抹掉，那么由于浮点运算时要对阶，在计算机中，52492 表示为 0.52492×10^5. 若取 $a_i=0.9$，对阶时 a_i 表示为 0.000009×10^5，从而它在该计算机中被抹掉数字 9 成为机器零，所以将 52492 与每个 a_i 逐一相加，其结果仍为 52492，出现了大数 52492“吃掉”小数 a_i. 这时，应该先把数量级相同的 1000 个 a_i 相加，然后再与 52492 相加，才能得到较准确的数值结果.

3）避免两个相近的数相减，以免严重损失有效数字.

例如，$x=1.232$，$y=1.231$，在四位十进制的限制下计算 $z=x^3-y^3$ 的值. 如果直接计算，得

$$z=(1.232)^3-(1.231)^3=0.1870\times10^1-0.1865\times10^1=0.5\times10^{-2},$$

所得结果至多有一位有效数字.

若采用 $z=(x-y)(x^2+xy+y^2)$ 进行计算，则得

$$\begin{aligned}z&=x^3-y^3=(x-y)(x^2+xy+y^2)\\&=0.1\times10^{-2}\times(0.1518\times10^1+0.1517\times10^1+0.1515\times10^1)\\&=0.455\times10^{-2}.\end{aligned}$$

如果 x 与 y 都是准确值，则 0.1×10^{-2} 是 $x-y$ 的准确值，所以 0.455×10^{-2} 的绝对误差限为 $\varepsilon_z=0.1\times10^{-2}\times(0.5\times10^{-3}+0.5\times10^{-3}+0.5\times10^{-3})$

$$=0.15\times10^{-5}<0.5\times10^{-5},$$

从而近似值 0.455×10^{-2} 的三位数字 4，5，5 都是有效的.

1.3 数值计算中应注意的一些原则

4）在除法运算中，避免除数的绝对值远小于被除数的绝对值.

从绝对误差来说，$z=\dfrac{x}{y}$的绝对误差限为

$$\varepsilon_z \leqslant \frac{|x|\varepsilon_y+|y|\varepsilon_x}{|y|^2},$$

由此可见，如果$|y|\ll|x|$，其中“≪”表示远小于，则ε_z很大.

5）防止出现机器零和溢出停机.

若计算过程中某些中间结果超出了数字计算机所能表示的数值范围，就出现机器零或溢出，使计算中止. 因此在计算过程中，常常需要用比例因子去乘近似值，使其保持在一定的数值范围内，以便数字计算机正常运行.

6）简化计算步骤，减少运算次数.

减少运算次数不仅可提高效率，并且能够使误差积累减小. 例如计算多项式

$$p_n(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$$

的值. 如果逐项计算后再相加，总共需要做$n+(n-1)+\cdots+1=\dfrac{n(n+1)}{2}$次乘法和$n$次加法. 若按递推算法

$$\begin{cases} u_n=a_n, \\ u_k=xu_{k+1}+a_k, \quad k=n-1,n-2,\cdots,0, \\ p_n(x)=u_0. \end{cases}$$

进行计算，则只需要n次乘法和n次加法. 这个算法称为**秦九韶法**.

作业（第2部分）

习题 1

5. 取$\sqrt{2}\approx 1.4$，欲计算$(\sqrt{2}-1)^6$ 的近似值，有下列四个算式可采用：

(1) $\dfrac{1}{(\sqrt{2}+1)^6}$，　(2) $(3-2\sqrt{2})^3$，

(3) $\dfrac{1}{(3+2\sqrt{2})^3}$，　(4) $99-70\sqrt{2}$.

分析这四个算式哪一个所得结果的误差最小.

6. 如何计算下列函数值才比较准确：

(1) $\dfrac{1}{1+2x}-\dfrac{1-x}{1+x}$，$|x|\ll 1$；　(2) $\sqrt{x+\dfrac{1}{x}}-\sqrt{x-\dfrac{1}{x}}$，$|x|\gg 1$；

(3) $\dfrac{1-\cos x}{x}$，$|x|\ll 1$；　(4) $\arctan(x+1)-\arctan x$，$|x|\gg 1$.

抽样分布

数理统计（1）

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

1.3 数理统计中常见的几个分布

下列三种分布在数理统计中起着重要的作用.

1. χ^2 分布

定义 1.3-1 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立,且都服从标准正态分布 $N(0;1)$,则称随机变量

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \tag{1.3-1}$$

所服从的分布为**自由度 n 的 χ^2 分布**,记为 $\chi^2(n)$.

定理 1.3-1 设随机变量 $\chi^2 \sim \chi^2(n)$,则其概率密度为

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}}, & x > 0, \\ 0, & x \leqslant 0. \end{cases} \tag{1.3-2}$$

证 因为 $X_1, X_2, \cdots, X_n$ 相互独立,且都服从 $N(0;1)$,所以 $\boldsymbol{X} = [X_1, X_2, \cdots, X_n]^{\mathrm{T}}$ 的概率密度为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x_i^2}{2}} = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right).$$

于是,当 $x \geqslant 0, \Delta x > 0$ 时,随机事件 $\{x < \chi^2 \leqslant x + \Delta x\}$ 的概率为

$$P\{x < \chi^2 \leqslant x + \Delta x\} = (2\pi)^{-\frac{n}{2}} \iint_G \cdots \int \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n,$$

1.3 数理统计中常见的几个分布

其中 G 是外半径为 $\sqrt{x+\Delta x}$、内半径为 $\sqrt{x}$ 的 n 维球壳. 记 G 的体积为 ΔV，则有

$$(2\pi)^{-\frac{n}{2}}\mathrm{e}^{-\frac{1}{2}(x+\Delta x)}\Delta V \leqslant P\{x<\chi^2\leqslant x+\Delta x\} \leqslant (2\pi)^{-\frac{n}{2}}\mathrm{e}^{-\frac{1}{2}x}\Delta V. \qquad (1.3\text{-}3)$$

由于半径为 r 的 n 维球体的体积是 Cr^n，这里的 C 是常数，故 $\Delta V = C[(x+\Delta x)^{\frac{n}{2}} - x^{\frac{n}{2}}]$，从而

$$\lim_{\Delta x\to 0}\frac{\Delta V}{\Delta x} = C\lim_{\Delta x\to 0}\frac{1}{\Delta x}[(x+\Delta x)^{\frac{n}{2}} - x^{\frac{n}{2}}] = C\,\frac{n}{2}x^{\frac{n}{2}-1}. \qquad (1.3\text{-}4)$$

因而，由(1.3-3)式和(1.3-4)式，得

$$\lim_{\Delta x\to 0}\frac{1}{\Delta x}P\{x<\chi^2\leqslant x+\Delta x\} = kx^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, \qquad (1.3\text{-}5)$$

其中 $k=(2\pi)^{-\frac{n}{2}}\,\frac{n}{2}C$ 是常数. 这就是说，当 $x>0$ 时，χ^2 的概率密度是 $kx^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}$. 又因 $\chi^2\geqslant 0$，故当 $x\leqslant 0$ 时，χ^2 的概率密度显然为 0. 再由概率密度的性质知，

$$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x = \int_0^{+\infty}kx^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}\mathrm{d}x = k2^{\frac{n}{2}}\int_0^{+\infty}u^{\frac{n}{2}-1}\mathrm{e}^{-u}\mathrm{d}u = k2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right),$$

即 $k=\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$. 于是，(1.3-2) 式成立. ∎

1.3 数理统计中常见的几个分布

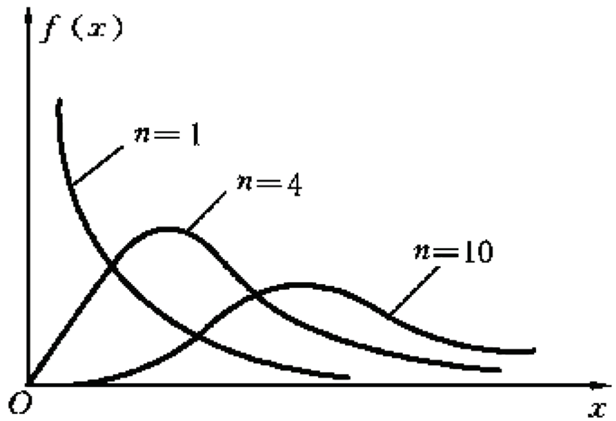

图 1.3-1 $\chi^2(n)$ 的概率密度函数

χ^2 分布的概率密度函数的图形如图1.3-1所示，它随自由度 n 的不同有所变动. 当 $n=1$ 时，$\lim\limits_{x\to 0^+} f(x) = +\infty$；而当 $n \geqslant 2$ 时，$\lim\limits_{x\to 0^+} f(x) = 0$.

定理 1.3-2 设 $X \sim \chi^2(n)$，即 X 是自由度为 n 的 χ^2 随机变量，则

(1) X 的特征函数为

$$\varphi(t) = E(e^{itX}) = (1-2it)^{-\frac{n}{2}}. \tag{1.3-6}$$

(2) X 的数学期望与方差分别为

$$E(X) = n,\ D(X) = 2n. \tag{1.3-7}$$

1.3 数理统计中常见的几个分布

证 (1)由特征函数的定义得

$$\varphi(t)=\int_0^{+\infty}\mathrm{e}^{\mathrm{i}tx}\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\mathrm{e}^{-\frac{x}{2}}x^{\frac{n}{2}-1}\mathrm{d}x=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^{+\infty}x^{\frac{n}{2}-1}\mathrm{e}^{-\left(\frac{1}{2}-\mathrm{i}t\right)x}\mathrm{d}x$$

$$=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}-\mathrm{i}t\right)^{-\frac{n}{2}}\int_0^{+\infty}u^{\frac{n}{2}-1}\mathrm{e}^{-u}\mathrm{d}u=(1-2\mathrm{i}t)^{-\frac{n}{2}}.$$

(2) $$E(X)=\frac{1}{\mathrm{i}}\varphi'(0)=\frac{1}{\mathrm{i}}\left[-\frac{n}{2}(1-2\mathrm{i}t)^{-\frac{n}{2}-1}(-2\mathrm{i})\right]_{t=0}=n;$$

$$E(X^2)=\frac{1}{\mathrm{i}^2}\varphi''(0)=\frac{1}{\mathrm{i}^2}\left[n\mathrm{i}\left(-\frac{n}{2}-1\right)(1-2\mathrm{i}t)^{-\frac{n}{2}-2}(-2\mathrm{i})\right]_{t=0}=n(n+2),$$

从而 $$D(X)=E(X^2)-[E(X)]^2=2n.$$ ∎

1.3 数理统计中常见的几个分布

定理 1.3-3 设 $X_1 \sim \chi^2(n_1)$，$X_2 \sim \chi^2(n_2)$，且 X_1 与 X_2 相互独立，则

$$X_1 + X_2 \sim \chi^2(n_1 + n_2). \tag{1.3-8}$$

证 由于 X_1、X_2 的特征函数分别为

$$\varphi_1(t) = (1 - 2it)^{-\frac{n_1}{2}}, \quad \varphi_2(t) = (1 - 2it)^{-\frac{n_2}{2}},$$

又 X_1 与 X_2 相互独立，故 $X_1 + X_2$ 的特征函数为

$$\varphi(t) = \varphi_1(t)\varphi_2(t) = (1 - 2it)^{-\frac{1}{2}(n_1+n_2)},$$

从而 $X_1 + X_2$ 服从自由度为 $n_1 + n_2$ 的 χ^2 分布. ■

定理 1.3-3 表明，χ^2 分布具有可加性. 更一般地，若 $X_i \sim \chi^2(n_i)(i = 1,2,\cdots,k)$，且 $X_1,X_2,\cdots,X_k$ 相互独立，则有 $\sum\limits_{i=1}^{k} X_i \sim \chi^2\left(\sum\limits_{i=1}^{k} n_i\right)$.

1.3 数理统计中常见的几个分布

2. t 分布

定义 1.3-2 设 $X \sim N(0;1)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，则称随机变量

$$T = \frac{X}{\sqrt{Y/n}} \tag{1.3-9}$$

所服从的分布为**自由度 n 的 t 分布**，记为 $t(n)$。

定理 1.3-4 设随机变量 $T \sim t(n)$，则 T 的概率密度为

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < +\infty. \tag{1.3-10}$$

证 记 $Z = \sqrt{\frac{Y}{n}}$，即 $Y = nZ^2$，则 Z 的概率密度为

$$f_Z(z) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}(nz^2)^{\frac{n}{2}-1}e^{-nz^2/2}2nz, & z > 0, \\ 0, & z \leqslant 0. \end{cases}$$

由于 X 与 Y 相互独立，故 X 与 Z 也相互独立，从而 $[X,Z]^{\mathrm{T}}$ 的概率密度为

$$f(x,z) = \begin{cases} \dfrac{n^{\frac{n}{2}}}{\sqrt{2\pi}2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} z^{n-1}e^{-(x^2+nz^2)/2}, & -\infty < x < +\infty,\ z > 0, \\ 0, & \text{其他}. \end{cases}$$

1.3 数理统计中常见的几个分布

现 $T=\dfrac{X}{Z}$，因而 T 的分布函数

$$F(t)=P\{T\leqslant t\}=P\left\{\frac{X}{Z}\leqslant t\right\}=P\{X\leqslant tZ\}=\int_0^{+\infty}\mathrm{d}z\int_{-\infty}^{tz}f(x,z)\mathrm{d}x.$$

于是，T 的概率密度为

$$f(t)=F'(t)=\int_0^{+\infty}f(tz,z)z\mathrm{d}z=\int_0^{+\infty}\frac{n^{\frac{n}{2}}}{\sqrt{2\pi}2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}z^n\mathrm{e}^{-(n+t^2)z^2/2}\mathrm{d}z$$

$$=\frac{1}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}\int_0^{+\infty}x^{\frac{n+1}{2}-1}\mathrm{e}^{-x}\mathrm{d}x=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}. \quad \blacksquare$$

从(1.3-10)式看出，t 分布的概率密度函数 $f(t)$ 是偶函数，且 $\lim\limits_{t\to\infty}f(t)=0$. $f(t)$ 的图形如图 1.3-2 所示.

我们不加证明地给出，若 $T\sim t(n)$，且 $n\geqslant 2$，则对正整数 $r(r<n)$，$E(T^r)$ 存在，且

$$E(T^r)=\begin{cases}0, & r\text{ 为奇数，}\\ n^{\frac{r}{2}}\dfrac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}, & r\text{ 为偶数.}\end{cases}$$

1.3 数理统计中常见的几个分布

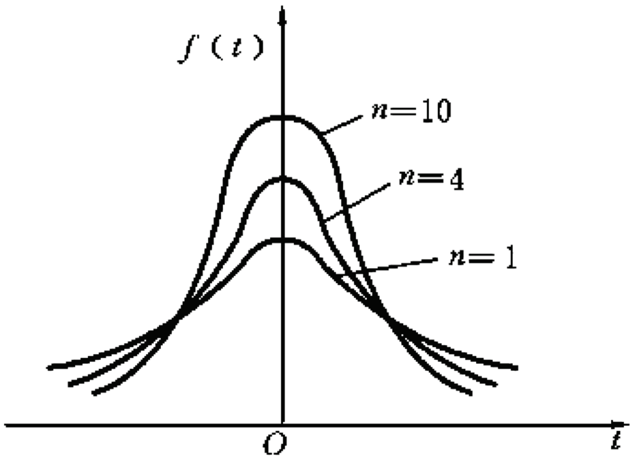

图 1.3-2　$t(n)$ 的概率密度函数

特别地，当 $n \geqslant 2$ 时，有 $E(T)=0$；当 $n \geqslant 3$ 时，有 $D(T)=\dfrac{n}{n-2}$.

当 $n \to +\infty$ 时，利用 Γ 函数的 Stirling 公式，可得

$$\lim_{n\to\infty} f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

即当 n 很大（一般只要 $n>30$）时，t 分布非常近似于 $N(0;1)$.

1.3 数理统计中常见的几个分布

3. F 分布

定义 1.3-3 设随机变量 X 与 Y 相互独立，且 $X \sim \chi^2(m)$，$Y \sim \chi^2(n)$，则称随机变量

$$F = \frac{X}{m}\bigg/\frac{Y}{n} = \frac{n}{m}\frac{X}{Y} \tag{1.3-11}$$

所服从的分布为**自由度为 (m,n) 的 F 分布**，记为 $F(m,n)$，其中 m 称为第一自由度，n 称为第二自由度.

由 F 分布定义可知，当 $F \sim F(m,n)$ 时，则随机变量 $\frac{1}{F}$ 所服从的分布为自由度为 (n,m) 的 F 分布，即 $\frac{1}{F} \sim F(n,m)$.

定理 1.3-5 设随机变量 $F \sim F(m,n)$，则 F 的概率密度为

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}\left(\dfrac{m}{n}\right)\left(\dfrac{m}{n}x\right)^{\frac{m}{2}-1}\left(1+\dfrac{m}{n}x\right)^{-\frac{m+n}{2}}, & x > 0 \\ 0, & x \leqslant 0. \end{cases} \tag{1.3-12}$$

证 令 $Z = \frac{X}{m}$，则由 X 的概率密度得到 Z 的概率密度为

$$f_W(w) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}(mw)^{\frac{n}{2}-1}e^{-\frac{mw}{2}}, & w > 0, \\ 0, & w \leqslant 0. \end{cases}$$

1.3 数理统计中常见的几个分布

由于 X 与 Y 相互独立，故 Z 与 W 相互独立，从而$[Z,W]^{\mathrm{T}}$ 的概率密度为

$$f(z,w)=\begin{cases}\dfrac{m^{\frac{m}{2}}\cdot n^{\frac{n}{2}}}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}z^{\frac{m}{2}-1}w^{\frac{n}{2}-1}\mathrm{e}^{-(mz+nw)/2}, & z>0,\ w>0,\\ 0, & \text{其他}.\end{cases}$$

于是，$F=\dfrac{Z}{W}$ 的分布函数为

$$P\{F\leqslant x\}=P\{Z\leqslant xW\}=\begin{cases}\displaystyle\int_0^{+\infty}\mathrm{d}w\int_0^{xw}f(z,w)\mathrm{d}z, & x>0\\ 0, & x\leqslant 0.\end{cases}$$

因此，当 $x>0$ 时，F 的概率密度为

$$f(x)=\frac{\mathrm{d}}{\mathrm{d}x}P\{F\leqslant x\}=\int_0^{+\infty}wf(xw,w)\mathrm{d}w$$

$$=\frac{m^{\frac{m}{2}}\cdot n^{\frac{n}{2}}}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\int_0^{+\infty}w(xw)^{\frac{m}{2}-1}w^{\frac{n}{2}-1}\mathrm{e}^{-(mx+n)w/2}\mathrm{d}w$$

$$=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)\left(\frac{m}{n}x\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}.$$

而当 $x\leqslant 0$ 时，F 的概率密度显然为 0. 这就证得(1.3-12)式. ▌

1.3 数理统计中常见的几个分布

F 分布的概率密度函数的图形如图 1.3-3 所示.

χ^2 分布、t 分布及 F 分布都是由标准正态分布 $N(0;1)$ 诱导出来的. 因此,它们与正态总体下统计量的分布有着密切的联系.

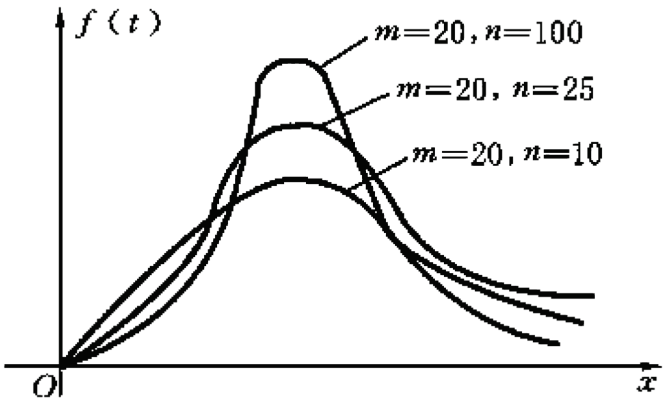

图 1.3-3　$F(m,n)$ 的概率密度函数

1.4 抽样分布

在统计推断中，经常要用到统计量的分布. 在一般情况下，要确定一个统计量的精确分布是很困难的，不过，如果总体服从正态分布，则可以求出一些重要统计量的分布. 正态总体在数理统计中占有特别重要的地位，一方面是因为在某些限制条件下假定总体服从正态分布是合理的，另一方面，数理统计中有许多总体的分布，都可以认为近似于正态分布. 因此研究正态总体下统计量的分布是很重要的.

定理 1.4-1 设 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体 $N(\mu;\sigma^2)$ 的一个样本，$a_1, a_2, \cdots, a_n$ 是已知常数，则统计量 $U = \sum_{i=1}^{n} a_i X_i$ 服从正态分布 $N\left(\mu \sum_{i=1}^{n} a_i; \sigma^2 \sum_{i=1}^{n} a_i^2\right)$.

证 由于 $X_1, X_2, \cdots, X_n$ 相互独立，所以 U 是正态随机变量.

又因为

$$E(U) = \sum_{i=1}^{n} a_i E(X_i) = \mu \sum_{i=1}^{n} a_i,$$

$$D(U) = \sum_{i=1}^{n} D(a_i X_i) = \sum_{i=1}^{n} a_i^2 D(X_i) = \sigma^2 \sum_{i=1}^{n} a_i^2,$$

所以有

$$U \sim N\left(\mu \sum_{i=1}^{n} a_i; \sigma^2 \sum_{i=1}^{n} a_i^2\right). \qquad \blacksquare$$

1.4 抽样分布

定理 1.4-2 设$(X_1, X_2, \cdots, X_n)$是取自正态总体$N(\mu;\sigma^2)$的一个样本，$\mathbf{A}=[a_{ij}]$是已知的$p\times n$常数矩阵，又

$$[Y_1, Y_2, \cdots, Y_p]^{\mathrm{T}} = \mathbf{A}[X_1, X_2, \cdots, X_n]^{\mathrm{T}},$$

则

$$Y_i \sim N\left(\mu\sum_{j=1}^{n} a_{ij}, \sigma^2\sum_{j=1}^{n} a_{ij}^2\right), \quad i=1,2,\cdots,p;$$

$$\mathrm{cov}(Y_i, Y_j) = \sigma^2\sum_{k=1}^{n} a_{ik}a_{jk}, \quad i,j=1,2,\cdots,p.$$

请读者自证.

特别，当$\mu=0$，且$\mathbf{A}$为n阶正交矩阵时，$\mathrm{cov}(Y_i, Y_j) = \begin{cases}\sigma^2, i=j\\ 0, i\neq j\end{cases}$. 因此，统计量$Y_1, Y_2, \cdots, Y_n$相互独立，并有相同的分布$N(0,\sigma^2)$. 这就是说，均值为零且相互独立的同分布正态随机变量$X_1, X_2, \cdots, X_n$，通过正交变换所得的随机变量$Y_1, Y_2, \cdots, Y_n$，也是均值为零且相互独立的同分布正态随机变量.

1.4 抽样分布

定理 1.4-3 设$(X_1,X_2,\cdots,X_n)$是取自正态总体$N(\mu;\sigma^2)$的一个样本，则

(1) $\overline{X}\sim N\left(\mu;\frac{1}{n}\sigma^2\right)$，$\sqrt{n}\frac{\overline{X}-\mu}{\sigma}\sim N(0;1)$，　　(1.4-1)

(2) $\overline{X}$与S^2相互独立，且

$$\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1).\tag{1.4-2}$$

证 (1) 由于$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，所以由定理 1.4-1 令$a_i=\frac{1}{n}(i=1,2,\cdots,n)$，即得$\overline{X}\sim N\left(\mu;\frac{1}{n}\sigma^2\right)$，从而

$$\frac{\overline{X}-\mu}{\sqrt{\frac{1}{n}\sigma^2}}=\sqrt{n}\frac{\overline{X}-\mu}{\sigma}\sim N(0;1).$$

(2) 令$Y_i=\frac{X_i-\mu}{\sigma}(i=1,2,\cdots,n)$，则$Y_1,Y_2,\cdots,Y_n$相互独立，且具有相同的分布$N(0;1)$. 作正交变换

$$\begin{cases}Z_1=\frac{1}{\sqrt{n}}Y_1+\frac{1}{\sqrt{n}}Y_2+\cdots+\frac{1}{\sqrt{n}}Y_n,\\ Z_i=c_{i1}Y_1+c_{i2}Y_2+\cdots+c_{in}Y_n,\quad i=2,3,\cdots,n,\end{cases}$$

1.4 抽样分布

其中常数 $c_{ij}(i=2,3,\cdots,n,\quad j=1,2,\cdots,n)$ 使 n 阶方阵

$$\begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

为正交矩阵，这可以通过求 $n-1$ 维欧氏空间 $V=\left\{\boldsymbol{x}\in\mathbf{R}^n \mid \sum_{i=1}^{n} x_i=0\right\}$ 的标准正交基，得到 c_{ij} 而找到这种正交矩阵. 于是 $Z_1,Z_2,\cdots,Z_n$ 相互独立，且都服从 $N(0;1)$. 记 $\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i=\frac{\bar{X}-\mu}{\sigma}$，则 $Z_1=\sqrt{n}\bar{Y}=\sqrt{n}\frac{\bar{X}-\mu}{\sigma}\sim N(0;1)$. 又因为正交变换不改变向量的长度，故有

$$\sum_{i=1}^{n}Y_i^2=\sum_{i=1}^{n}Z_i^2.$$

1.4 抽样分布

从而

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}-\frac{\bar{X}-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2+\sum_{i=1}^{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)^2$$

$$-2\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)\left(\frac{\bar{X}-\mu}{\sigma}\right)$$

$$=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2+n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2-2n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2-n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2$$

$$=\sum_{i=1}^{n}Z_i^2-Z_1^2=\sum_{i=2}^{n}Z_i^2.$$

因此按 χ^2 分布的定义知

$$\frac{(n-1)S^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=2}^{n}Z_i^2\sim\chi^2(n-1),$$

并且由于 $Z_1,Z_2,\cdots,Z_n$ 相互独立，故 Z_1 与 $\sum_{i=2}^{n}Z_i^2$ 相互独立，即 $\sqrt{n}\frac{\bar{X}-\mu}{\sigma}$ 与 $\frac{(n-1)S^2}{\sigma^2}$ 相互独立，所以 $\bar{X}$ 与 S^2 相互独立. ∎

1.4 抽样分布

定理 1.4-4 设$(X_1,X_2,\cdots,X_n)$是取自正态总体$N(\mu;\sigma^2)$的一个样本，则

$$\sqrt{n}\,\frac{\overline{X}-\mu}{S}\sim t(n-1). \tag{1.4-3}$$

证 因为$\overline{X}$与S^2相互独立，且

$$\sqrt{n}\,\frac{\overline{X}-\mu}{\sigma}\sim N(0,1),\quad \frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1),$$

所以由t分布的定义知 $\sqrt{n}\,\dfrac{\overline{X}-\mu}{S}=\dfrac{\sqrt{n}\,\dfrac{\overline{X}-\mu}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}}\sim t(n-1).$

定理 1.4-5 设$(X_1,X_2,\cdots,X_m)$是取自正态总体$N(\mu_1;\sigma^2)$的样本，$(Y_1,Y_2,\cdots,Y_n)$是取自正态总体$N(\mu_2;\sigma^2)$的样本，且这两个样本相互独立，则

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}\sim t(m+n-2), \tag{1.4-4}$$

其中 $S_w^2=\dfrac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2},\quad S_1^2=\dfrac{1}{m-1}\sum_{i=1}^{m}(X_i-\overline{X})^2,$

$$S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2.$$

1.4 抽样分布

证 由 $\bar{X}$ 与 $\bar{Y}$ 相互独立以及 $\bar{X} \sim N\left(\mu_1; \frac{\sigma^2}{m}\right)$，$\bar{Y} \sim N\left(\mu_2; \frac{\sigma^2}{n}\right)$，得 $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2; \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right)$，即

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0;1).$$

又因
$$\frac{(m-1)S_1^2}{\sigma^2} \sim \chi^2(m-1), \quad \frac{(n-1)S_2^2}{\sigma^2} \sim \chi^2(n-1),$$

且它们相互独立，故由 χ^2 分布的可加性得

$$V = \frac{(m-1)S_1^2}{\sigma^2} + \frac{(n-1)S_2^2}{\sigma^2} \sim \chi^2(m+n-2).$$

从而由 U 与 V 相互独立和 t 分布的定义知

$$\frac{U}{\sqrt{\frac{V}{m+n-2}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2). \quad \blacksquare$$

1.4 抽样分布

定理 1.4-6 设$(X_1,X_2,\cdots,X_m)$是取自正态总体$N(\mu_1;\sigma_1^2)$的样本，$(Y_1,Y_2,\cdots,Y_n)$是取自正态总体$N(\mu_1;\sigma_2^2)$的样本，且这两个样本相互独立，则

$$\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(m-1,n-1). \tag{1.4-5}$$

证 由于$\frac{(m-1)S_1^2}{\sigma_1^2}\sim\chi^2(m-1)$，$\frac{(n-1)S_2^2}{\sigma_2^2}\sim\chi^2(n-1)$，且它们相互独立，故由$F$分布的定义知

$$\frac{(m-1)S_1^2}{\sigma_1^2(m-1)}\bigg/\frac{(n-1)S_2^2}{\sigma_2^2(n-1)}=\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}\sim F(m-1,n-1).$$ ∎

值得注意的是，上述两个定理中，定理 1.4-5 要求两个正态总体的方差相同，称为方差齐次；而定理 1.4-6 并不要求方差是相同的.

对于非正态总体，要求出统计量的分布是比较困难的，即使有时理论上可以求出精确的分布，但其形式复杂而难以应用，这时一般利用中心极限定理推出$n\to+\infty$下的极限分布. 从而当样本容量很大时，总是应用有关统计量的极限分布作为其近似分布.

1.4 抽样分布

定理 1.4-7 设总体 X 的一阶矩、二阶矩存在：$E(X)=\mu, 0<D(X)=\sigma^2<+\infty$，$(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本，则当 n 充分大时，

(1) 样本均值 $\overline{X}_n$ 近似地服从 $N\left(\mu; \dfrac{\sigma^2}{n}\right)$，

(2) 样本方差 S_n^2 依概率收敛于 σ^2，

(3) $\sqrt{n}\dfrac{\overline{X}_n-\mu}{S_n}$ 近似地服从 $N(0;1)$，其中

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i, \quad S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2.$$

（请参阅何迎晖、闵华玲《数理统计》，北京：高等教育出版社，1989） ■

1.5 分位数

数理统计中经常要用到分位数. 分位数是随机变量的一种数字特征，其定义如下.

定义 1.5-1 设随机变量 X 的分布函数为 $F(x)$，p 是一实数，$0<p<1$，若 x_p 使

$$P\{X \leqslant x_p\} = F(x_p) = p, \tag{1.5-1}$$

则称 x_p 为此分布的 p 分位数.

如果 $X \sim N(0;1)$，记它的 p 分位数为 u_p，即 u_p 满足

$$\Phi(u_p) = \int_{-\infty}^{u_p} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = p.$$

对于给定的 p，u_p 的值可由标准正态分布表（见附表一）查得，例如，$u_{0.95} = 1.645$，$u_{0.975} = 1.96$.

由于正态分布的概率密度函数是偶函数，所以有

$$\begin{aligned} P\{X \leqslant -u_{1-p}\} &= P\{X \geqslant u_{1-p}\} = 1 - P\{X < u_{1-p}\} \\ &= 1 - P\{X \leqslant u_{1-p}\} = 1-(1-p) = p. \end{aligned}$$

从而由分位数的定义知 $\quad u_p = -u_{1-p} \quad$ 或 $\quad u_{1-p} = -u_p. \qquad (1.5\text{-}2)$

若 $X \sim \chi^2(n)$，用 $\chi_p^2(n)$ 表示此分布的 p 分位数，即

$$P\{X \leqslant \chi_p^2(n)\} = p.$$

$\chi_p^2(n)$ 的值可由 χ^2 分布表查得. 附表二对于某些 p 和较小的 n 给出了 $\chi_p^2(n)$ 的值. 例如，$\chi_{0.01}^2(8) = 1.646$，$\chi_{0.95}^2(10) = 18.307$. 当 $n > 45$ 时，可以用近似公式

$$\chi_p^2(n) \approx \frac{1}{2}(u_p + \sqrt{2n-1})^2 \tag{1.5-3}$$

得到 χ^2 分布的 p 分位数.

1.5 分位数

$t(n)$ 分布的 p 分位数 $t_p(n)$ 可由 t 分布表查得. 附表三对于某些大于 $\frac{1}{2}$ 的 p 和较小的 n 给出了 $t_p(n)$ 的值. 当 $p < \frac{1}{2}$ 时, 可以利用

$$t_p(n) = -t_{1-p}(n) \quad \text{或} \quad t_{1-p}(n) = -t_p(n) \tag{1.5-4}$$

得到 p 分位数. 例如, $t_{0.975}(14) = 2.1448$, $t_{0.05}(7) = -t_{0.95}(7) = -1.8946$. 当 $n > 45$ 时, 由于 $t(n)$ 近似于 $N(0;1)$, 所以可以用近似公式

$$t_p(n) \approx u_p \tag{1.5-5}$$

得到 t 分布的 p 分位数.

$F(m,n)$ 分布的 p 分位数记为 $F_p(m,n)$. 对于某些大于 $\frac{1}{2}$ 的 p 及某些 m、n, $F_p(m,n)$ 的值可以通过查附表四得到. 当 $p < \frac{1}{2}$ 时, 可以利用公式

$$F_p(m,n) = \frac{1}{F_{1-p}(n,m)} \tag{1.5-6}$$

得到 p 分位数. 这是因为

$$p = P\{F \leqslant F_p(m,n)\} = P\left\{\frac{1}{F} \geqslant \frac{1}{F(m,n)}\right\} = 1 - P\left\{\frac{1}{F} \leqslant \frac{1}{F_p(m,n)}\right\},$$

1.5 分位数

故由 $\frac{1}{F} \sim F(n,m)$ 及分位数定义知，$\frac{1}{F_p(m,n)} = F_{1-p}(n,m)$，亦即(1.5-6)式成立. 例如，

$F_{0.95}(5,9) = 3.48$，$F_{0.05}(5,9) = 1/F_{0.95}(9,5) = 1/4.77 = 0.21$.

对于分位数还需注意下列两种表示方法.

(1) 若 λ 使 $P\{X > \lambda\} = p$，则称 λ 为 X 的上侧 p 分位数. 显然有

$$P\{X > \lambda\} = 1 - P\{X \leqslant \lambda\},$$

故 X 的上侧 p 分位数即为 X 的 $1-p$ 分位数.

(2) 若 λ_1、λ_2 使

$$P\{X \leqslant \lambda_1\} = \frac{p}{2}, P\{X > \lambda_2\} = \frac{p}{2},$$

则称 λ_1、λ_2 为 X 的双侧 p 分位数. 显然，λ_1 为 X 的 $\frac{p}{2}$ 分位数，λ_2 为 X 的 $1-\frac{p}{2}$ 分位数.

作业（第3部分）

习题 1

10. 设总体 X 服从正态分布 $N(12;2^2)$，现随机抽取容量为 5 的样本，问

(1) 此样本的最小值小于 10 的概率是多少？

(2) 此样本的最大值大于 15 的概率是多少？

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

黃 煒
博士·副教授
Dr. Wei Huang