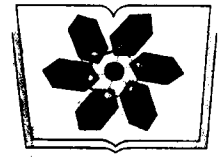
成分数据统计分析引论

张尧庭 著

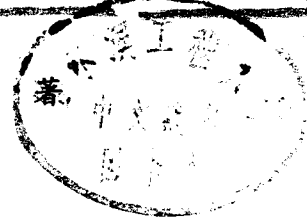
科学出版社

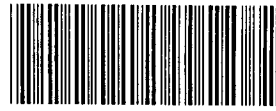


中国科学院科学出版基金资助出版

成分数据统计分析引论

张尧庭 著





科学出版社

2000

内 容 简 介

成分数据统计分析的主要内容是以成分数据为目标的统计理论与方法，其基础与多种分布，如逻辑正态分布族，狄氏分布族等有关．本书在此基础上介绍了成分数据统计分析的理论与分析方法，以及这一方向在国内外的最新成果．本书每章末附有习题，以便读者更好地理解本书内容．

本书可供应用统计工作者、科研人员及大学有关专业高年级学生、研究生、教师阅读．

图书在版编目(CIP)数据

成分数据统计分析引论/张尧庭著．-北京:科学出版社,2000
ISBN 7-03-008264-8

Ⅰ．成… Ⅱ．张… Ⅲ．数理统计 Ⅳ．O212.1

中国版本图书馆 CIP 数据核字（2000）第 01520 号

科学出版社 出版
北京东黄城根北街 16 号
邮政编码:100717

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2000 年 8 月第 一 版 开本：850×1168 1/32
2000 年 8 月第一次印刷 印张：5 3/8
印数：1—2 000 字数：138 000

定价：14.00 元

（如有印装质量问题，我社负责调换〈新欣〉）

前　言

成分数据的统计问题早就提出了，早在1897年，K·皮尔逊就指出解决这些问题是很困难的．直到1986年，才有第一本书专门论述这一类问题，作者J．Aitchison也因在这一方向研究的特殊贡献，荣获英国皇家统计学会1988年的研究奖章．此书的中译本（《成分数据的统计分析》）在1990年由中国地质大学出版社出版．艾奇逊在书的前言中阐明了他写这本书的目的是："……在于清晰而完整地介绍近年来专为成分数据而设计的一套新的统计方法．……强调提供有效而可行的方法，而不是进行详细的证明和推导．"这样的写法对于想进一步了解理论的人，是会感到不满足的．新的方法主要是指基于加性逻辑正态分布的统计分析方法．

成分数据的统计分析与单形上的分布有密切的联系，逻辑正态、狄氏分布等等都是单形上的分布族．只有把单形上的分布类型作一些合理的分类，才可能把成分数据的统计分析放在一个坚实的基础上，才有可能提供各种恰当的分析方法．从理论上看，这样也比较系统和自然．

从成分数据的实际背景来看，加性、乘性逻辑正态分布仅仅反映了一种类型的分布．被艾奇逊认为不够丰富的狄氏分布及其推广，事实上也是重要的．本书发展了这一方面的统计分析方法，对理论和实际都提供了新的工具．

有了单形上的分布，混料试验的设计就可以有一个比较恰当的理论模型．一方面既可以说明为什么混料试验中添加对数项往往会有效，另一方面也推动了设计的改进和分析方法的多样化，这些内容在艾奇逊书中没有展开，而在本书中有一些反映．本书尽量不和艾奇逊书的内容重复，因为我的观点和他有些不同．我

认为，直到现在，真正的成分数据的分析仍然是非常困难的．本书一个重要的目的是显示它的困难，但我没有办法解决，而是希望有更多的人来从事这一方面的研究．

本书断断续续写了几乎 7 年，因为这几年有不少其他的事情，迫使我中断下来，所以本书的一、二章和三、四章有明显的差别，我自己也感到不满意，由于能力有限，只能这样了．书中有一些内容是没有发表过的成果，但我感到没有做完、做好，只能重印时再说了．

本书的前两章，邹国华同志看了之后，提出了不少好的意见，并改正了不少错误，在这里我表示衷心的感谢；另外书中也汇集了章栋恩同志近年来的工作，有些即将发表，我得到他的允许，也反映到本书中．

本书的出版得到了中国科学院出版基金的支持，我深表感谢．希望这本书既能填补我国这一方面的空白，也希望为将来的研究提供条件．我还要感谢多年来一直合作很好的毕颖同志，她为本书的出版和编辑付出了大量的精力．

希望得到读者的批评和帮助，以便改正由于我水平有限所出现的不足．

张尧庭

1999 年 11 月

目　录

第一章　准备知识

§1. n 维欧氏空间与单形

我们在这一节引入本书常用的一些记号，并列举一些重要的概念，读者可以从一般的线性代数的教材中查阅有关的说明. 我们假定读者已具备线性代数、数理统计的基本知识，并对多元统计分析的方法有一些了解.

用 R_n 表示实数域上的 n 维线性空间，R_n 中的向量用小写字母表示，如 $a,b,c,\cdots,x,y,z$ 等. 向量的分量以相应的带脚标的字母表示，如

$$a=\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix},\qquad x=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}.$$

通常矩阵用大写字母表示，如 $A,B,C,\cdots,X,Y,Z$ 等. 它的大小列举在字母下面，如 $\underset{n\times m}{A}$ 表示 A 是 n 行 m 列的矩阵，从上下文可以确定其大小时，就不再列举，A 中的元素用双脚标表示，如

$$A=(a_{ij}),\qquad X=(x_{ij}).$$

矩阵的转置用“′”表示，A' 表示将 A 转置后的矩阵. 我们总是把向量也看成矩阵，当

$$a=\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix}$$

时，a 的转置 $a'=(a_1,a_2,\cdots,a_n)$. 于是向量 a 与 b 的内积可以用 $a'b$ 写出，即

$$a'b = (a_1,\cdots,a_n)\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i,$$

在 R_n 中引入内积后，R_n 就是一个 n 维的欧氏空间. 对于矩阵，有时也用它的行向量或列向量表示，通常都用一个脚标，带括号表示行指标，不带括号表示列指标，即

$$\underset{n\times m}{A} = (a_{ij}) = \begin{pmatrix} a'_{(1)} \\ \vdots \\ a'_{(n)} \end{pmatrix} = (a_1 \quad a_2 \quad \cdots \quad a_m),$$

易见

$$a'_{(i)} = (a_{i1},a_{i2},\cdots,a_{im}), \quad a'_i = (a_{1i},a_{2i},\cdots,a_{ni}).$$

我们用 $\mathbb{1}_n$ 表示元素全为 1 的 n 维向量，当维数由上下文确定时，用 $\mathbb{1}$ 表示，于是有

$$\mathbb{1}'a = a'\mathbb{1} = (a_1,\cdots,a_n)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{n} a_i,$$

在统计中常见的一组数据 $a_1,\cdots,a_n$ 的均值和方差均可用向量、矩阵的运算表示出来，因为

$$\frac{1}{n}\sum_{i=1}^{n} a_i = \frac{1}{n}\mathbb{1}'a = \frac{1}{n}a'\mathbb{1} = \bar{a},$$

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{a})^2 &= \frac{1}{n}\left(\sum_{i=1}^{n} a_i^2 - n\bar{a}^2\right) \\ &= \frac{1}{n}\left(a'a - n\left(\frac{1}{n}\mathbb{1}'a\right)^2\right) \\ &= \frac{1}{n}a'\left(I_n - \frac{1}{n}\mathbb{1}\mathbb{1}'\right)a, \end{aligned}$$

其中 $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$，$I_n$ 就是 $n\times n$ 的单位阵.

当向量 a 的每个分量均大于零时，我们用 $a>0$ 表示，类似的 $a\geqslant 0$ 就表示 a 的分量均为非负的实数. 然而矩阵 $A\geqslant 0$ 则表示 A 是非负定的方阵，$A>0$ 表示 A 是正定阵.

向量 $\underset{n\times 1}{x}$ 的一个线性函数 $\sum_{i=1}^{n} a_i x_i$ 可以写成 $a'x$，其中常数向量 a 是由系数 $a_1,\cdots,a_n$ 形成的. 于是线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_m. \end{cases}$$

可以写成矩阵或向量的形式：

$$\underset{n\times m}{A}\ \underset{m\times 1}{x} = \underset{n\times 1}{b}$$

或

$$a_1x_1 + a_2x_2 + \cdots + a_mx_m = b,$$

而

$$A = (a_{ij}),\quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix},\quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

A 也可写成 $(a_1\ \ a_2\ \ \cdots\ \ a_m)$.

给定矩阵 $\underset{n\times m}{A} = (a_1\ \ a_2\ \ \cdots\ \ a_m)$ 后，A 的列向量 $a_1,\cdots,a_m$ 所张成的线性子空间用 $R(A)$ 表示，即

$$R(A) = \{\underset{n\times m}{A}\ \underset{m\times 1}{x} : x \in R_m\}.$$

很明显，$R(A)$ 是 R_n 中的一个子空间. 当 $a'b=0$ 时，我们称 a 与 b 正交，记为 $a\perp b$. 若 a 与集合 S 中每一个向量都正交，就记为 $a\perp S$. 如果 b 与矩阵 $A=(a_1\ \ a_2\ \ \cdots\ \ a_m)$ 中的每一个列向量都正交，即 $b\perp a_i, i=1,2,\cdots,m$，则 $b\perp R(A)$. 与 $R(A)$ 正交的向量形成一个子空间，称为 $R(A)$ 的正交补空间，记为 $R(A)^{\perp}$，即

$$R(A)^{\perp} = \{x : x \perp R(A)\}.$$

很明显，$R(A)^{\perp}$ 也可以用另一个形式表示，即

$$R(A)^{\perp} = \{x : x \perp a_i, i = 1,2,\cdots,m\}$$

$$= \{x : a_i'x = 0, i = 1,2,\cdots,m\}$$
$$= \{x : A'x = 0\}.$$

考虑一个特殊的正交补空间，它在今后的统计分析中起重要的作用.若 $a'\mathbb{1}=0$，则称 $a'x$ 是向量 x 的一个对比，a 称为这个对比的系数向量.如

$$x_1 - x_2, \quad x_1 - 2x_2 + x_n,$$
$$x_1 - \frac{1}{n-1}(x_2 + \cdots + x_n), \cdots$$

都是 x 的对比.很明显，$R(\mathbb{1})^{\perp}$ 就是对比系数向量组成的子空间，它就是方程

$$\mathbb{1}'x = 0$$

的全部的解.

为了方便，今后用一些记号代表 R_n 中的一些特定的集合.这些集合是：

$$R_n^+ = \{x : x \in R_n, x_i > 0, i = 1,2,\cdots,n\},$$
$$\overline{R}_n^+ = \{x : x \in R_n, x_i \geqslant 0, i = 1,2,\cdots,n\},$$
$$S_n = \{x : x \in R_n^+, \mathbb{1}'x = 1\},$$
$$\overline{S}_n = \{x : x \in \overline{R}_n^+, \mathbb{1}'x = 1\},$$
$$D_n = \{x : x \in R_n^+, \mathbb{1}'x < 1\},$$
$$\overline{D}_n = \{x : x \in \overline{R}_n^+, \mathbb{1}'x \leqslant 1\},$$

其中集合 $\overline{S}_n$ 通常称为 n 维空间中的单形，或简称为单形(有的书上称为单纯形)，S_n 是单形 $\overline{S}_n$ 的内部.

成分数据，也就是由百分比组成的数据，因此成分数据的取值范围就是单形 $\overline{S}_n$，由于讨论时在数学上更便于处理，我们常常先讨论 S_n 内取值的情况，所以 S_n，$\overline{S}_n$ 是我们今后经常会遇到的集合.

如果 $a_i \geqslant 0, i = 1,2,\cdots,k$，且 $\sum_{i=1}^{k} a_i = 1$，也即 $a = \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} \in \overline{S}_k$，

$b_1, b_2, \cdots, b_k$ 均为 R_n 中的向量，则称

$$a_1b_1 + a_2b_2 + \cdots + a_kb_k = (b_1\ b_2 \cdots\ b_k)a$$

是 $b_1, \cdots, b_k$ 的凸线性组合. 用矩阵 $\underset{n\times k}{B} = (b_1 \cdots\ b_k)$ 来表示，就是：只要 $a \in \overline{S}_k$，Ba 就是 B 的列向量的一个凸线性组合，集合

$$\{Ba : a \in \overline{S}_k\}$$

就是向量 $b_1, b_2, \cdots, b_k$ 所生成的凸包. R_n 中一个集合对凸线性组合是封闭的，也即其中任意有限个向量的凸线性组合仍属于这个集合，则称这个集合是凸集. 很明显，R_n^+，$\overline{R}_n^+$，S_n，$\overline{S}_n$，D_n，$\overline{D}_n$ 都是凸集，$b_1, \cdots, b_k$ 生成的凸包 $\{Ba : a \in \overline{S}_k\}$ 是包含 $b_1, \cdots, b_k$ 的最小的凸集.

§2. 基和成分

给定一个向量 $\omega \in \overline{R}_{n+1}$，$\omega \neq 0$，$\omega' = (\omega_0, \omega_1, \cdots, \omega_n)$，我们就称 ω 是一个基向量，令

$$\begin{cases} t = \sum_{i=0}^{n} \omega_i, \\ x_i = \dfrac{\omega_i}{t}, \end{cases} \qquad i = 0,1,2,\cdots,n. \tag{2.1}$$

则称 t 是 ω 相应的总量，$x' = (x_0, x_1, \cdots, x_n)$ 称为 ω 相应的成分，当 ω 明确时，简称为总量、成分. 易见 ω, t, x 有下列关系：

$$\begin{cases} \omega = tx, x = \omega t^{-1} (\omega \neq 0), \\ t \geqslant 0, x \geqslant 0, 1'x = 1 (\omega \neq 0). \end{cases} \tag{2.2}$$

为了今后讨论方便，以下无特殊声明时，我们总假定 $\omega \in R_{n+1}^+$，于是相应有

$$\begin{cases} \omega = tx, x = \omega t^{-1}, \\ t > 0, x > 0, 1'x = 1. \end{cases} \tag{2.3}$$

此时，我们对向量 ω 或 x 的各个分量取对数是有意义的，取对数后形成的向量分别用 $\ln\omega$ 和 $\ln x$ 表示，即

$$\ln\omega = \begin{pmatrix} \ln\omega_0 \\ \ln\omega_1 \\ \vdots \\ \ln\omega_n \end{pmatrix}, \quad \ln x = \begin{pmatrix} \ln x_0 \\ \ln x_1 \\ \vdots \\ \ln x_n \end{pmatrix}.$$

我们称 $\ln\omega$ 的一个对比 $a'\ln\omega$ 是 ω 的一个对数对比. 对成分 x 也是一样,当 $\mathbb{1}'a=0$ 时,$a'\ln x$ 称为 x 的一个对数对比.

如果把 $x\in S_{n+1}$ 看成一个基向量,则 x 相应的成分还是它自己. 因此基向量 ω 和相应的成分 x 可以看成是 ω 在 S_{n+1} 上的"投影",$x\in S_{n+1}$ 时它的"投影"还是自己,不同的 ω 可以有相同的"投影". 把这一概念更一般化,就引出形状和大小的抽象概念.

定义 2.1 假定 $\omega\in R_{n+1}^{+}$,即 ω 是一个基向量,如果 $G(\omega)$ 是一个 ω 的正值函数,且有

$$G(c\omega) = cG(\omega), \tag{2.4}$$

对一切 $c>0$ 成立,则称 $G(\omega)$ 是 ω 的大小(Size),令

$$z_G(\omega) = \omega/G(\omega), \tag{2.5}$$

称 $z_G(\omega)$ 是 $G(\omega)$ 这个大小相应的形状(shape).

从定义 2.1 可以看出,总量 t 是一个 ω 的大小,它相应的形状就是成分 x. 很明显,ω 的大小可以列举很多,如

$$\Big(\sum_{i=0}^{n}\omega_i^2\Big)^{\frac{1}{2}}, \quad \omega_0, \quad \max_{0\leqslant i\leqslant n}\omega_i$$

等都是,它们各自相应的形状是不同的. 然而形状向量之间却有如下的重要关系.

定理 2.1 假定 $\omega\in R_{n+1}^{+}$,G_1, G_2 是 ω 的两个大小,z_1, z_2 是 G_1, G_2 分别相应的形状,则有

$$z_1(\omega) = z_2(\omega)/G_1(z_2(\omega)). \tag{2.6}$$

证明 $z_1(\omega)=\omega/G_1(\omega)$

$$\begin{aligned} &= \Big[\frac{\omega}{G_2(\omega)}\Big]\Big/\Big[\frac{G_1(\omega)}{G_2(\omega)}\Big] \\ &= z_2(\omega)/G_1(\omega/G_2(\omega)) \\ &= z_2(\omega)/G_1(z_2(\omega)), \end{aligned}$$

这就告诉我们,形状向量是可以相互表示的.

现在引入子成分的概念.成分的一部分并不再是一个成分向量,例如将成分向量 x 分为二段,即

$$x = \begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix},$$

$x_{(i)}$是 $n_i \times 1$ 的向量,$i=1,2$.由于 $x_{(i)}>0$,必然有 $0<\mathbb{1}'x_{(i)}<1$,$i=1,2$,它们各自的分量之和一定小于 1,不能是一个成分向量.子成分是把成分分解为若干段,把每一段看成一个基向量,这些基向量相应的成分称为子成分.用数学的术语描述就是如下的定义.

定义 2.2 把基向量 ω 和相应的成分向量 x 同样分成 k 段,即有

$$x = \begin{pmatrix} x_{(1)} \\ \vdots \\ x_{(k)} \end{pmatrix} \begin{matrix} n_1 \\ \vdots \\ n_k \end{matrix}, \quad \omega = \begin{pmatrix} \omega_{(1)} \\ \vdots \\ \omega_{(k)} \end{pmatrix} \begin{matrix} n_1 \\ \vdots \\ n_k \end{matrix}.$$

$$n_1 + n_2 + \cdots + n_k = n + 1.$$

令 $S_{(i)} = x_{(i)}(\mathbb{1}'x_{(i)})^{-1}$,$i=1,2,\cdots,k$,则称 $S_{(i)}$是 x 的子成分.

当 $n_i>1$ 时,$S_{(i)}$是一个向量;当 $n_i=1$ 时,$S_{(i)}=1$.这表明成分向量 x 的子成分在 $n_i=1$ 时需要特殊考虑,这一点在今后的讨论中会经常遇到.另一方面,从子成分的定义还可以看出子成分 $S_{(i)}$有以下的性质:

(1)$S_{(i)} = \omega_{(i)}(\mathbb{1}'\omega_{(i)})^{-1}$.

这是因为 $S_{(i)} = x_{(i)}(\mathbb{1}'x_{(i)})^{-1} = tx_{(i)}(t\,\mathbb{1}'x_{(i)})^{-1}$,而 $tx_{(i)} = w_{(i)}$,于是就有 $S_{(i)} = \omega_{(i)}(\mathbb{1}'\omega_{(i)})^{-1}$.这告诉我们子成分 $S_{(i)}$既是 $x_{(i)}$的成分也是 $\omega_{(i)}$的成分,$\omega_{(i)}$的有些性质可以在 $S_{(i)}$上反映出来.

(2)$S_{(i)}$的对数对比一定是 x 的对数对比,也是 $\omega_{(i)}$的对数对比.

这是由于对数对比的两个性质推出的,这两个性质是:成分 x 的对数对比一定是基 ω 的同一个对数对比(同一个指系数向量相

同);ω 的一部分 $\omega_{(i)}$ 的对数对比也是 ω 的一个对数对比.用数学的式子表示就是:

因为当 $a'\mathbb{1}=0$ 时,

$$a'\ln x = a'\ln(\omega/t) = a'\ln\omega - a'\,\mathbb{1}\ln t$$
$$= a'\ln\omega;$$

当 $a'_{(i)}\mathbb{1}_{n_i}=0$ 时,$a'_{(i)}\ln\omega_{(i)}=(0,0,\cdots,a'_{(i)},0,\cdots,0)\ln\omega$,取 $a=(0,0,\cdots,a'_{(i)},0,\cdots,0)$,则有

$$a'\,\mathbb{1}_{n+1} = a'_{(i)}\,\mathbb{1}_{n_i} = 0, a'_{(i)}\ln\omega_{(i)} = a'\ln\omega.$$

一个对比的系数是一个 $n+1$ 维的向量,它满足齐次方程 $a'\mathbb{1}=0$,也即 $\mathbb{1}'a=0$.这一齐次方程的全部解是 $R(\mathbb{1})$ 的正交补空间 $R^{\perp}(\mathbb{1})$,是一个 n 维的子空间,它的一组基是 $\begin{pmatrix} I_n \\ -\mathbb{1}' \end{pmatrix}$ 这一矩阵的 n 个列向量,它相应的投影矩阵是

$$I_{n+1} - P_{\mathbb{1}} = I_{n+1} - \frac{1}{n+1}\mathbb{1}\mathbb{1}'.$$

容易验证:

$$\left(I_{n+1} - \frac{1}{n+1}\mathbb{1}\mathbb{1}'\right)\mathbb{1} = 0,$$

因此任一 $a=\left(I_{n+1}-\frac{1}{n+1}\mathbb{1}\mathbb{1}'\right)x, x\in R_{n+1}$,就一定有性质 $a'\mathbb{1}=0$,也即 a 是一个对比系数;反之,任一对比系数 a,它一定可以表成 $\left(I_{n+1}-\frac{1}{n+1}\mathbb{1}\mathbb{1}'\right)b$ 或 $\begin{pmatrix} I_n \\ -\mathbb{1}' \end{pmatrix}g$.这些在今后的讨论中都会用到.

§3. 多元正态分布

这一节概要地叙述多元正态分布的一些基本的性质,一些较长的证明都不列出,读者可以参阅一般多元统计分析的教材.

n 维多元标准正态分布用 $N(0,I_n)$ 表示,它的密度函数写成

$\varphi(x)$,$\varphi(\dot{x})$的表达式是

$$\varphi(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \mathrm{e}^{-\frac{1}{2}x'x}. \tag{3.1}$$

容易看出,当随机向量 $\underset{n\times 1}{x}$ 遵从 $N(0,I_n)$分布时,它的分量 $x_1,\cdots,x_n$ 是独立、同分布 $N(0,1)$的随机变量,期望值为 0,方差是 1. $N(0,1)$是一元的标准正态分布.

我们用 Ex 表示随机向量 x 的期望值组成的向量,$\mathrm{Var}(x)$表示 x 相应协方差矩阵,即

$$Ex = \begin{pmatrix} Ex_1 \\ \vdots \\ Ex_n \end{pmatrix},$$

$$\mathrm{Var}(x) = E(x - Ex)(x - Ex)'.$$

当 $x\sim N(0,I_n)$时,若 $y=\underset{m\times n}{A}x+\mu$,则有

$$\begin{cases} Ey = AEx + \mu = \mu, \\ \mathrm{Var}(y) = AE(x - Ex)(x - Ex)'A' = AA'. \end{cases} \tag{3.2}$$

只要 $AA'>0$,y 就有密度,记$\sum=AA'$后,y 的分布密度是

$$\left(\frac{1}{\sqrt{2\pi}}\right)^m \left|\sum\right|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(y-\mu)'\sum^{-1}(y-\mu)}. \tag{3.3}$$

这是一般 m 维多元正态分布,记为 $N(\mu,\sum)$,μ 的维数或$\sum$的阶数就是分布的维数.对于 $N(\mu,\sum)$我们用引理的形式列举它的性质,并给以扼要的证明.

引理 3.1 设 $\underset{m\times 1}{y}\sim N(\mu,\sum)$,$\sum>0$,则有

(1)$\sum^{-\frac{1}{2}}(y-\mu)\sim N(0,I_m)$;

(2)$(y-\mu)'\sum^{-1}(y-\mu)\sim\chi^2(m)$.

证明 当$\sum>0$时,它的特征根 $\lambda_1,\cdots,\lambda_m$ 均为正数,λ_i^α 对任一实数α 均有意义.由对角化的定理知道存在正交阵 Γ 使

$$\sum = \Gamma \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix} \Gamma'.$$

我们规定

$$\sum^{\alpha} = \Gamma \begin{pmatrix} \lambda_1^{\alpha} & & 0 \\ & \ddots & \\ 0 & & \lambda_m^{\alpha} \end{pmatrix} \Gamma',$$

于是$\sum^{\frac{1}{2}}, \sum^{-\frac{1}{2}}$均有意义，且$\sum^{\frac{1}{2}} \sum^{-\frac{1}{2}} = \sum^{-\frac{1}{2}} \sum^{\frac{1}{2}} = I_m$.

由于 y 的线性函数依然是标准正态分布随机变量 x 的线性函数，因此它还是正态分布，只需求出它的期望值与协方差矩阵，就可以完全确定它的分布，而

$$E\sum^{-\frac{1}{2}}(y-\mu) = \sum^{-\frac{1}{2}}(E(y)-\mu) = 0,$$

$$\mathrm{Var}\left(\sum^{-\frac{1}{2}}(y-\mu)\right) = \sum^{-\frac{1}{2}} \mathrm{Var}(y) \sum^{-\frac{1}{2}} = I_m,$$

这就证明了(1). 由 $x = \sum^{-\frac{1}{2}}(y-\mu) \sim N(0, I_m)$，知道

$$(y-\mu)' \sum^{-1} (y-\mu) = x'x = \sum_{i=1}^{m} x_i^2,$$

因此从 $x_1, \cdots, x_m$ 独立同分布 $N(0,1)$ 得到

$$x'x \sim \chi^2(m).$$

这就证明了(2).

引理 3.2 若 $y \sim N(\mu, \sum)$，将 y, μ 与 $\sum$ 作相应的分块，即

$$y = \begin{pmatrix} y_{(1)} \\ y_{(2)} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \quad \sum = \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix},$$

则有

(1) $y_{(i)} \sim N(\mu_{(i)}, \sum_{ii})$, $i = 1, 2$;

(2) $y_{(1)}$ 对 $y_{(2)}$ 的条件分布还是正态，且

$$\begin{cases} E\{y_{(1)} \mid y_{(2)}\} = \mu_{(1)} + \sum_{12} \sum_{22}^{-1} (y_{(2)} - \mu_{(2)}), \\ \mathrm{Var}(y_{(1)} \mid y_{(2)}) = \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21}. \end{cases} \tag{3.4}$$

证明 这只需将(3.3)的分布密度用边缘密度和条件分布密度乘积的形式写出就得. 关键是指数上二次型的分解，用到了一个