多元统计分析与应用

余锦华　杨维权　编著

SPSS

SAS

2

中山大学出版社

多元统计分析与应用

余锦华 杨维权 编著

SPSS

中山大学出版社
·广州·

图书在版编目（CIP）数据

多元统计分析与应用/余锦华，杨维权编著．—广州：中山大学出版社，2005.2
ISBN 7-306-02276-8

Ⅰ.多…　Ⅱ.①余…②杨…　Ⅲ.多元分析：统计分析　Ⅳ.O212.4

中国版本图书馆 CIP 数据核字（2005）第 006590 号

责任编辑：李　慈
封面设计：尔　晗
责任校对：李　引
责任技编：黄少伟
出版发行：中山大学出版社
编辑部电话（020）84111996，84113349
发行部电话（020）84111998，84111160
地　　址：广州市新港西路 135 号
邮　　编：510275　　传真：（020）84036565
印 刷 者：番禺市新华印刷有限公司
经 销 者：广东省新华发行集团
规　　格：787mm×960mm　1/16　23.5 印张　470 千字
版次印次：2005 年 2 月第 1 版　2005 年 2 月第 1 次印刷
定　　价：38.00 元

本书如有印装质量问题影响阅读，请与承印厂联系

前　　言

多元统计分析是近几十年来迅速发展起来的一门学科，随着电子计算机的广泛应用，及统计软件的普及，多元统计方法已被广泛应用于自然科学乃至社会科学各个领域。

早在1983年，我们就为我校数学系本科生开设了“多元统计分析”课程并编写了讲义。20多年来，我们在多元统计分析的教学与科研、理论与应用研究中不断探索，今天积20多年之心得体会，编写出这本教材。

本教材旨在介绍多元统计分析的基础知识、基本理论及其应用。全书由两大部分组成：

第一部分包括第一章至第四章，用尽可能少的篇幅简明扼要地阐述多元统计的基础知识，包括多元正态分布及其性质、多元的参数估计、假设检验、分布理论，其中所用到的矩阵知识收集在附录一，若干繁难的定理证明则放在附录二中。

第二部分包括第五章至第十二章，详细介绍当前卓有成效的各种多元统计分析方法，包括多元回归分析、方差分析、判别分析、聚类分析、主成分分析、因子分析、对应分析、多维标度法及典型相关分析。

本书可根据读者的不同需求在三个层次上使用：

第一层次，着重于应用的读者可略去第一部分及第二部分的理论推导，着重理解、掌握各种多元统计方法,通过学习书中应用实例,能够运用这些方法解决有关实际问题。

第二层次，偏重于应用而又希望对多元统计理论有所了解的读者，可略去全书的理论推导，将精力集中在基本概念、基本知识和基本结果及各种多元统计方法的背景、功能、作用、计算方法及实际应用上。

第三层次，既强调基础理论又注重实际应用的读者，通过学习全书，既能深刻系统地掌握多元统计的基本理论，又能掌握好各种多元统计方法，并能够应用这些方法解决实际问题。这是对数学、应用数学、统计科学专业本科生（或研究生）的要求，也是我们编写这本教材的本意。

由于我们学识有限，书中难免有错漏不当之处，敬请读者和专家们指正。

目　　录

第一章　多元正态分布
(Multivariate Normal Distribution)

多元统计分析是概率论与数理统计的后继课程。数理统计学指的是一元统计，主要研究客观事物中单个随机变量的统计规律性，其主要基础是一元正态分布。多元统计分析则研究客观事物中多个随机变量（或多个指标）的统计规律性，包括变量之间相互联系，它的主要基础是多元正态分布。多元统计分析是一元统计的延伸和拓广。

为了研究客观事物中多个随机变量（或多个指标）的统计规律性，我们首先要将一元分布拓广到多元分布，将一元正态分布拓广到多元正态分布。

多元正态分布是最常用的一种多元概率分布。除此之外，还有多元对数正态分布、多项分布、多元 β 分布、多元 χ^2 分布、多元指数分布等。

第一节　多元分布的基本概念

多元分布的基本概念，可由二元概率分布的自然推广而得到，如联合分布、边沿分布、条件分布、独立性、特征函数、数字特征等。对此，只作扼要的综述。

定义 1.1.1 随机向量、随机矩阵　由定义在同一概率空间（Ω，F，P）上的 p 个随机变量 x_1，x_2，…，x_p 构成的 p 维向量

$$X=(x_1,\ x_2,\ \cdots,\ x_p)'=\begin{pmatrix}x_1\\x_2\\\vdots\\x_p\end{pmatrix}$$

称为 p 维随机向量，它的概率分布称为 p 元分布。由定义在同一概率空间（Ω，F，P）上的 $p\times n$ 个随机变量组成的矩阵

$$X=(x_{ij})_{p\times n}$$

称为随机矩阵，矩阵中每个元素 x_{ij}都为一随机变量，常记为

$$X=(x_{ij})_{p\times n}=\begin{pmatrix}x_{11}&x_{12}&\cdots&x_{1n}\\x_{21}&x_{22}&\cdots&x_{2n}\\\vdots&\vdots&&\vdots\\x_{p1}&x_{p2}&\cdots&x_{pn}\end{pmatrix}=(X^{(1)},X^{(2)},\cdots,X^{(n)})_{p\times n}=\begin{pmatrix}X'_{(1)}\\X'_{(2)}\\\vdots\\X'_{(p)}\end{pmatrix}_{p\times n}\tag{1.1.1}$$

在此，以 $X^{(1)}$，$\cdots X^{(n)}$表示 X 的列向量，$X_{(1)}$，$\cdots$，$X_{(p)}$表示 X 的行向量的转置向量。X 是 $p\times n$ 阶随机矩阵，它可看成 n 个 p 维（列）随机向量所构成的，也可看成 p 个 n 维（行）向量构成的。X 的概率分布是指按列拉直的全体元素（随机变量）x_{11}，$\cdots$，x_{p1}，x_{12}，$\cdots$，x_{p2}，$\cdots$，x_{1n}，$\cdots$，x_{pn}组成的 pn 元分布（注：以后所称向量都是指列向量）。

定义 1.1.2 数字特征 依次称

$$E(X)=\begin{pmatrix} Ex_1\\ Ex_2\\ \vdots\\ Ex_p\end{pmatrix},\quad E(Y)=\begin{pmatrix} Ey_1\\ Ey_2\\ \vdots\\ Ey_q\end{pmatrix}$$

为随机向量 $X=(x_1, x_2, \cdots, x_p)'$及 $Y=(y_1, y_2, \cdots, y_q)'$的均值向量，称

$$E(X)=(Ex_{ij})_{p\times n}$$

为随机矩阵 X 的均值矩阵。称

$$\begin{aligned}\operatorname{cov}(X,X)&=E[X-E(X)][X-E(X)]'\\ &=\begin{pmatrix}\operatorname{cov}(x_1,x_1) & \cdots & \operatorname{cov}(x_1,x_p)\\ \operatorname{cov}(x_2,x_1) & \cdots & \operatorname{cov}(x_2,x_p)\\ \vdots & & \vdots\\ \operatorname{cov}(x_p,x_1) & \cdots & \operatorname{cov}(x_p,x_p)\end{pmatrix}_{p\times p}\end{aligned} \tag{1.1.2}$$

为 p 维随机向量 X 的自协方差阵，简称为 X 的自协差阵。称$|\operatorname{cov}(X, X)|$为 X 的广义方差，它是自协差阵的行列式值。注意，有时亦记 $\operatorname{cov}(X, X)$ 为 $D(X)$。称

$$\begin{aligned}\operatorname{cov}(X,Y)&=E[X-E(X)][Y-E(Y)]'\\ &=\begin{pmatrix}\operatorname{cov}(x_1,y_1) & \cdots & \operatorname{cov}(x_1,y_q)\\ \operatorname{cov}(x_2,y_1) & \cdots & \operatorname{cov}(x_2,y_q)\\ \vdots & & \vdots\\ \operatorname{cov}(x_p,y_1) & \cdots & \operatorname{cov}(x_p,y_q)\end{pmatrix}_{p\times q}\end{aligned} \tag{1.1.3}$$

为 X 与 Y 的互协方差阵，简称为 X 与 Y 的互协差阵。

显而易见

$$\begin{aligned}\operatorname{cov}(X, Y)&=E[X-E(X)][Y-E(Y)]'\\ &=\{E[Y-E(Y)][X-E(X)]'\}'\\ &=\{\operatorname{cov}(Y, X)\}'\end{aligned}$$

对于 p 维随机向量 $X=(x_1, \cdots, x_p)$，称

$$\rho_{ij}=\frac{\operatorname{cov}(x_i,x_j)}{\sqrt{D(x_i)}\sqrt{D(x_j)}},\qquad i,j=1,2,\cdots,p$$

为分量 x_i 与 x_j 之间的相关系数，称 $R=(\rho_{ij})_{p\times p}$ 为 X 的相关矩阵。对于两个不同的随机向量 X 及 Y，它们之间的相关问题将在典型相关分析中进一步讨论。

定义 1.1.3 特征函数 随机向量 X 的特征函数定义为

$$\varphi(t)=E(e^{it'X}) \tag{1.1.4}$$

其中 $i=\sqrt{-1}$，t 是实变向量，其维数与 X 的维数相同，若记 $t=(t_1, \cdots, t_p)'$，则

$$\varphi(t)=E[\exp(i\sum_{\alpha=1}^{p}t_\alpha x_\alpha)]$$

随机矩阵 $X=(x_{ij})_{p\times n}$ 的特征函数是用 pn 维随机向量 $(x_{11}, \cdots, x_{p1}, \cdots x_{1n}, \cdots x_{pn})'$ 的特征函数来定义的，即为：

$$E\{\exp\{i\sum_{\alpha=1}^{p}\sum_{\beta=1}^{n}t_{\alpha\beta}x_{\alpha\beta}\}\}=E\{\exp\{itr(T'X)\}\}\triangleq\varphi(T) \tag{1.1.5}$$

其中 T 是与 X 的阶数相同的实变矩阵，tr（$T'X$）是矩阵 $T'X$ 之迹，即 $T'X$ 的对角线上元素之和。

由上述定义及向量、矩阵有关运算的基本知识，可以证明下述等式成立。设 A，B，C 为常数矩阵，$\mathbf{X}$，$\mathbf{Y}$ 为随机矩阵，X，Y，U，V 为随机向量，则有：

$$E(A\mathbf{X}B+C)=AE(\mathbf{X})B+C \tag{1.1.6}$$

$$E(A\mathbf{X}+B\mathbf{Y})=AE(\mathbf{X})+BE(\mathbf{Y}) \tag{1.1.7}$$

$$\mathrm{cov}(X,Y)=E[XY']-[E(X)][E(Y)]' \tag{1.1.8}$$

$$\mathrm{cov}(AX,BY)=A\,\mathrm{cov}(X,Y)B' \tag{1.1.9}$$

$$\begin{aligned}\mathrm{cov}(aX+bY,dU+eV)=&ad\,\mathrm{cov}(X,U)+\\&ae\,\mathrm{cov}(X,V)+bd\,\mathrm{cov}(Y,U)+be\,\mathrm{cov}(Y,V)\end{aligned} \tag{1.1.10}$$

其中 a，b，d，e 皆为实常数。

$$D(BX+K)=BD(X)B' \tag{1.1.11}$$

其中 K 为实常数向量。

这里假定上述各式的运算总可进行（如协方差阵的存在及阶数、维数协调一致等)。详细证明留作练习。

p 维随机向量 $X=(x_1, \cdots, x_p)$ 的协方差阵及相关矩阵都是非负定矩阵。事实上

$$\mathrm{cov}(X,X)=(\mathrm{cov}(x_i,x_j))_{p\times p}$$

记

$$\mathrm{cov}(x_i,x_j)=\sigma_{ij} \qquad i,j=1,\cdots,p$$

则

$$\sigma_{ij}=\mathrm{cov}(x_i,x_j)=\mathrm{cov}(x_j,x_i)=\sigma_{ji}$$

可见自协差阵是对称的。又对任意 p 维实向量 $C=(c_1,\cdots,c_p)'$，总有

$$C'\operatorname{cov}(X,X)C=\sum_{i=1}^{p}\sum_{j=1}^{p}c_ic_jE(x_i-Ex_i)(x_j-Ex_j)$$
$$=E\Big[\sum_{i=1}^{p}c_i(x_i-Ex_i)\Big]^2\geqslant 0 \qquad (1.1.12)$$

可见，协方差阵是非负定阵（常用 $\operatorname{cov}(X,X)\geqslant\square$ 表示非负定阵，$\square$ 表示元素为零的矩阵）。由此可知，相关矩阵 R 也是非负定阵。

定义 1.1.4 条件分布、独立性 将 p 维随机向量 $X=(x_1,\cdots,x_p)'$ 分成两个子向量：$1\leqslant q<p$，

$$X=\begin{pmatrix}X_{(1)}\\X_{(2)}\end{pmatrix},X_{(1)}=\begin{pmatrix}x_1\\\vdots\\x_q\end{pmatrix},X_{(2)}=\begin{pmatrix}x_{q+1}\\\vdots\\x_p\end{pmatrix} \qquad (1.1.13)$$

在此，$X_{(1)}$，$X_{(2)}$ 分别表示 X 的两个子向量，也为随机向量。记 $F(x_1,\cdots,x_p)$ 为 X 的联合分布函数，$F_1(x_1,\cdots,x_q)$，$F_2(x_{q+1},\cdots,x_p)$ 分别为 $X_{(1)}$，$X_{(2)}$ 的边沿分布函数。随机向量 $X_{(1)}$ 与 $X_{(2)}$ 相互独立是指

$$F(x_1,\cdots,x_p)=F_1(x_1,\cdots,x_q)\cdot F_2(x_{q+1},\cdots,x_p) \qquad (1.1.14)$$

若记 $\varphi(t)$ 为 X 的特征函数，$\varphi_1(t_{(1)})$ 及 $\varphi_2(t_{(2)})$ 分别表示 $X_{(1)}$ 与 $X_{(2)}$ 的特征函数，其中

$$t=\begin{pmatrix}t_{(1)}\\t_{(2)}\end{pmatrix},t_{(1)}=\begin{pmatrix}t_1\\\vdots\\t_q\end{pmatrix},t_{(2)}=\begin{pmatrix}t_{q+1}\\\vdots\\t_p\end{pmatrix}$$

则 $X_{(1)}$ 与 $X_{(2)}$ 相互独立等价于

$$\varphi(t)=\varphi_1(t_{(1)})\varphi_2(t_{(2)}) \qquad (1.1.15)$$

条件分布通常是分别就离散型，连续型给出定义。现在就连续型情形讨论条件分布问题，对于离散型情形可类似地给出有关结论。设 X 具有分布密度函数 $f(x_1,\cdots,x_p)$，则 $X_{(1)}$ 及 $X_{(2)}$ 具有各自分布的密度函数，分别记作 $f_1(x_1,\cdots,x_q)$ 及 $f_2(x_{q+1},\cdots,x_p)$。由（1.1.14）式知，$X_{(1)}$ 与 $X_{(2)}$ 相互独立又等价于

$$f(x_1,\cdots,x_p)=f_1(x_1,\cdots,x_p)\cdot f_2(x_{q+1},\cdots,x_p) \qquad (1.1.16)$$

若（1.1.16）式不成立，则 $X_{(1)}$ 及 $X_{(2)}$ 中一方的概率分布必与另一方有关。当给定 $X_{(2)}=x_{(2)}$ 的数值向量时，若 $f_2(x_{q+1},\cdots,x_p)\neq 0$，则 $X_{(1)}$ 具有条件分布密度函数 $f_1(x_1,\cdots,x_q\mid x_{q+1},\cdots,x_p)$，可以证明

$$f_1(x_1,\cdots,x_q \mid x_{q+1},\cdots,x_p) = \frac{f(x_1,\cdots,x_p)}{f_2(x_{q+1},\cdots,x_p)} \tag{1.1.17}$$

同理，在给定 $X_{(1)} = x_{(1)}$的数值向量时，若 f_1（x_1，…，x_q）$\neq 0$，则 $X_{(2)}$具有条件分布密度函数 f_2（x_{q+1}，…，$x_p | x_1$，…，x_q），可以证明

$$f_2(x_{q+1},\cdots,x_p \mid x_1,\cdots,x_q) = \frac{f(x_1,\cdots,x_p)}{f_1(x_1,\cdots,x_q)} \tag{1.1.18}$$

对照（1.1.16）、（1.1.17）、（1.1.18）三式可知，$X_{(1)}$与 $X_{(2)}$相互独立等价于

$$f_1(x_1,\cdots,x_q \mid x_{q+1},\cdots,x_p) = f_1(x_1,\cdots,x_q)$$

或

$$f_2(x_{q+1},\cdots,x_p \mid x_1,\cdots,x_q) = f_2(x_{q+1},\cdots,x_p)$$

第二节　多元正态分布

多元正态分布是一元正态分布及二元正态分布的自然推广，当然内容更为丰富。我们知道，若 y 服从 N（0，1），记 $x = \mu + \sigma y$，其中 μ 及$\sigma > 0$ 为实数，则 x 服从 N（μ，σ^2）。就是说，可用标准正态分布来定义一般正态分布。在上节中，是用 x，X 和 X分别表示随机变量、随机向量和随机矩阵；在本节中，常用 X，Y 分别表示随机向量，而 x，y 分别表示 X，Y 这些随机向量的观察值向量。

一、多元正态分布的定义

定义 1.2.1　设 y_1，…，y_q 是相互独立同分布 N（0，1）的随机变量，则称 $Y =$（y_1，…，y_q）$'$服从 q 元标准正态分布，其中

$$E(Y) = (0,\cdots,0)' = □$$

$$\operatorname{cov}(Y,Y) = E(YY') = \begin{pmatrix} D(y_1) & 0 & \cdots & 0 \\ 0 & D(y_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D(y_q) \end{pmatrix}_{q\times q}$$

$$= \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}_{q\times q} = I_q$$

简记 $Y \sim N_q$（□，I_q）。其中□为零向量，I_q 为 q 阶单位阵。由（1.1.4）式知 Y 的特征函数

$$\varphi_Y(t) = E(e^{it'Y}) = E(e^{i\sum_{j=1}^{q} t_j y_j})$$
$$= E(\prod_{j=1}^{q} e^{it_j y_j}) = \prod_{j=1}^{q} E(e^{it_j y_j})$$
$$= \prod_{j=1}^{q} e^{-\frac{1}{2}t_j^2} = e^{-\frac{1}{2}t't}$$

定义 1.2.2 设 Y 服从 N_q（□，I_q）。B 为 $p\times q$ 实数矩阵，μ 为 p 维实向量，作变换 $X=\mu+BY$，则称 X 服从 N_p（μ，Σ），其中$\Sigma=BB'$为 p 阶非负定阵。

定理 1.2.1 设 X 服从 N_p（μ，Σ），则有

$$E(X)=\mu,\ \mathrm{cov}(X,X)=\Sigma \tag{1.2.1}$$

$$\varphi_X(t)=\exp\left(it'\mu-\frac{1}{2}t'\Sigma t\right) \tag{1.2.2}$$

证明 由定义 1.2.2 知，$X=\mu+BY$，Y 服从 N_q（□，I_q），且$\Sigma=BB'$。于是

$$E(X)=\mu+BE(Y)=\mu$$
$$\mathrm{cov}(X,X)=B\mathrm{cov}(Y,Y)B'=BI_qB'$$
$$=\Sigma$$
$$\varphi_X(t)=E(e^{it'X})=E(e^{it'(\mu+BY)})$$
$$=E(e^{it'\mu}e^{i(B't)'Y})$$
$$=\exp\{it'\mu\}\exp\left\{-\frac{1}{2}(B't)'(B't)\right\}$$
$$=\exp\left\{it'\mu-\frac{1}{2}t'\Sigma t\right\}$$

类似于一元概率分布由其特征函数惟一确定的情形，对于随机向量而言，多元概率分布亦可由其特征函数惟一地确定。由定理 1.2.1 可见，若 X 服从 N_p（μ，Σ），则 X 的特征函数φ_X（t）由（1.2.2）式所确定，反之也可用特征函数来定义多元正态分布。就是说，若随机向量 X 的特征函数为（1.2.2）式所确定，则称 X 服从 N_p（μ，Σ）。

多元正态分布的随机向量的线性变换是否仍为正态分布？

定理 1.2.2 （1）设 X 服从 N_p（μ，Σ），A 为 $r\times p$ 阶实数矩阵，b 为 r 维实向量，作变换 $Z=AX+b$，则 Z 服从 N_r（$A\mu+b$，$A\Sigma A$）。这个性质称为正态分布在线性变换下具有不变性。

（2）设 X 服从 N_p（μ，Σ），c 为一实数，则 cX 服从 N_p（$c\mu$，$c^2\Sigma$）。

（3）设 X 服从 N_p（μ，Σ），L 为 P 维实向量，则 $L'X$ 服从 N（$L'\mu$，$L'\Sigma L$）。

证明 （1）令 $X=\mu+BY$，其中 Y 服从 N_p（□，I_p），$\Sigma=BB'$，于是

$$Z = AX + b = (A\mu + b) + (AB)Y$$

并有

$$E(Z) = A\mu + b$$

$$\mathrm{cov}(Z, Z) = (AB)\mathrm{cov}(Y, Y)(AB)' = ABI_qB'A' = A\Sigma A'$$

则由定义 1.2.2 知，Z 服从 $N_r(A\mu + b, A\Sigma A')$。

(2) 在 (1) 中取 $b = \square$，$A = cI_p$，则 cX 服从 $N_p(c\mu, c^2\Sigma)$，其中

$$cI_p\Sigma(cI_p)' = c^2\Sigma$$

(3) 在 (1) 中取 $b = \square$，$A = L'$，即有

$$AX + b = L'X \sim N(L'\mu, L'\Sigma L)$$

推论 1 多元正态分布的边沿分布仍为正态分布。(证明留作习题)

推论 2 对 $\forall \mu \in R^p$ 及 $\Sigma_{p\times p} \geqslant \square$，必存在 p 维随机向量 X，使 $X \sim N_p(\mu, \Sigma)$。

证明 这里的关键在于能否找出一个矩阵 B，使 $BB' = \Sigma$，这实质上是求非负定阵的方根的问题。根据矩阵知识补充（见附录一）可知，对于 $\forall \Sigma \geqslant \square$，必存在 B，使 $BB' = \Sigma$。特别地，必存在 $C \geqslant \square$，使 $CC' = \Sigma$，特记 C' 为 $\Sigma^{1/2}$。

今令 $X = \mu + \Sigma^{1/2}Y$，其中 Y 服从 $N_p(\square, I_p)$。则 X 为随机向量，且由定理 1.2.2 可知 $X \sim N_p(\mu, \Sigma)$。

多元正态分布与一元正态分布有下述密切联系。

定理 1.2.3 设 X 为 p 维随机向量，则 X 服从多正态分布的充分必要条件为对于任一 p 维实数向量 L，$y = L'X$ 为一元正态分布随机变量。

证明 必要性：由定理 1.2.2 之 (3) 即知结论成立。

充分性：用特征函数来证。设 $E(X) = \mu$，$\mathrm{cov}(X, X) = \Sigma$，则

$$E(y) = L'\mu,\ \mathrm{cov}(y, y) = L'\Sigma L$$

且 y 服从 $N_1(L'\mu, L'\Sigma L)$，于是 y 的特征函数为：

$$\varphi_y(t) = E(e^{ity}) = E(e^{itL'X}) = \exp\left\{it(L'\mu) - \frac{1}{2}t^2L'\Sigma L\right\}$$

今取 $t = 1$，便得

$$E(e^{iL'X}) = \exp\left\{iL'\mu - \frac{1}{2}L'\Sigma L\right\}$$

其中 L 为任一 p 维实数向量。将上式与定理 1.2.1 的 (1.2.2) 式相比，可知 X 服从 $N_p(\mu, \Sigma)$。

定理 1.2.4 设 $X \sim N_p(\mu, \Sigma)$，且 Σ 为正定阵，则随机向量 X 具有分布密度函数

$$f(x;\mu,\Sigma)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}\tag{1.2.3}$$

证明　作变换 $Y=\Sigma^{-\frac{1}{2}}(X-\mu)$，则 Y 服从 N_p（□，I_p），由定义 1.2.1 知，Y 的各个分量，即 y_1，…，y_p 相互独立且服从一元标准正态分布，所以 Y 的分布密度函数为

$$f_Y(y)=\prod_{j=1}^{p}\left[\frac{1}{\sqrt{2\pi}}e^{-y_j^2/2}\right]=\frac{1}{(2\pi)^{p/2}}\exp\left\{-\frac{1}{2}y'y\right\}$$

逆变换为 $X=\Sigma^{1/2}Y+\mu$，其中 $\Sigma^{1/2}(\Sigma^{1/2})=\Sigma$。通过 Y 的分布密度函数 $f_Y(y)$ 可导出 X 的分布密度函数。

这里，X 到 Y 的变换为 1－1 变换，此变换的雅可比行列式的绝对值为

$$\left|\frac{\partial(Y)}{\partial(X)}\right|=\left|\frac{\partial(y_1,\cdots,y_p)}{\partial(x_1,\cdots,x_p)}\right|=|\Sigma^{-1/2}|=|\Sigma|^{-1/2}$$

因此，有

$$\begin{aligned}f_X(x;\mu,\Sigma)&=f_Y\left[\Sigma^{-1/2}(X-\mu)\right]|\Sigma|^{-1/2}\\&=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\right\}\end{aligned}$$

反之，不难证明其逆定理亦成立。

定理 1.2.4 是在 Σ 为正定阵的情况下，导出 X 的分布密度函数（1.2.3）式，称 X 具有非退化的 p 元正态分布。当 Σ 不是正定阵时，则称 X 为退化的 p 元正态分布随机向量，此时 X 没有分布密度函数。如定理 1.2.2 中 $p<r$ 的情形，则线性变换后得到的 Z 具有退化的 r 元正态分布。

至此，关于随机向量 X 具有多元正态分布的定义提出了四种等价的方式：定义 1.2.2，用特征函数（1.2.2）式定义，用分布函数（1.2.3）式定义，定理 1.2.3。

定理 1.2.5　设 X 服从 N_p（□，Σ），Σ 为正定阵，则 $\eta=X'\Sigma^{-1}X$ 服从 χ_p^2，即服从自由度为 p 的卡方分布。

证明　作变换 $Y=\Sigma^{-1/2}X$，则 Y 服从 N_p（□，I_p），因为

$$\mathrm{cov}(Y,Y)=\Sigma^{-1/2}\mathrm{cov}(X,X)(\Sigma^{-1/2})'=\Sigma^{-1/2}\Sigma(\Sigma^{-1/2})=I_p$$

其中 $\Sigma=\Sigma^{1/2}\Sigma^{1/2}$，$(\Sigma^{-1/2})'=\Sigma^{-1/2}$。若记 $Y=(y_1,\cdots,y_p)'$，则 y_1，…，y_p 是 p 个独立的标准正态分布变量，根据卡方分布的定义知，$Y'Y=\sum_{j=1}^{p}y_j^2$ 服从 χ_p^2 分布。另一方面，$Y'Y=X'\Sigma^{-1/2}\Sigma^{-1/2}X=X'\Sigma^{-1/2}X=\eta$，所以 η 服从 χ_p^2 分布。

二、边沿分布与条件分布

由定理 1.2.2 之（1）可知，多元正态分布的边沿分布仍为正态分布。下面将证明多元正态分布的条件分布仍为正态分布。

定理 1.2.6　设 X 服从 $N_p(\mu, \Sigma)$，Σ 为正定阵，作相应的剖分：

$$X=\begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}\begin{matrix} q\times 1 \\ (p-q)\times 1 \end{matrix} \qquad \mu=\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}\begin{matrix} q\times 1 \\ (p-q)\times 1 \end{matrix} \qquad 1\leqslant q<p$$

$$\Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中 Σ_{11} 为 q 阶正定阵，Σ_{22} 为（$p-q$）阶正定阵，$\Sigma'_{21}=\Sigma_{12}$。记

$$\Sigma_{11\cdot 2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\Sigma_{22\cdot 1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{1.2.4}$$

则有：

（1）$X_{(1)}$ 服从 $N_q(\mu_{(1)}, \Sigma_{11})$，$X_{(2)}$ 服从 $N_{p-q}(\mu_{(2)}, \Sigma_{22})$；

（2）$X_{(1)}$ 与 $X_{(2)}$ 相互独立的充要条件为 $\Sigma_{12}=\square$；

（3）$X_{(2)}$ 与 $X_{(1)}-\Sigma_{12}\Sigma_{22}^{-1}X_{(2)}$ 相互独立，且 $X_{(1)}-\Sigma_{12}\Sigma_{22}^{-1}X_{(2)}$ 服从

$$N_q(\mu_{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\mu_{(2)}, \Sigma_{11\cdot 2})$$

（4）给定 $X_{(2)}=x_{(2)}$ 时，$X_{(1)}$ 有条件分布

$$N_q(\mu_{(1)}+\Sigma_{12}\Sigma_{22}^{-1}(x_{(2)}-\mu_{(2)}), \Sigma_{11\cdot 2})$$

（5）同理，$X_{(1)}$ 与 $X_{(2)}-\Sigma_{21}\Sigma_{11}^{-1}X_{(1)}$ 相互独立，且 $X_{(2)}-\Sigma_{21}\Sigma_{11}^{-1}X_{(1)}$ 服从 $N_{p-q}(\mu_{(2)}-\Sigma_{21}\Sigma_{11}^{-1}\mu_{(1)}, \Sigma_{22\cdot 1})$；给定 $X_{(1)}=x_{(1)}$ 时，$X_{(2)}$ 有条件分布

$$N_{p-q}(\mu_{(2)}+\Sigma_{21}\Sigma_{11}^{-1}(x_{(1)}-\mu_{(1)})), \Sigma_{22\cdot 1})$$

证明（1）由定理 1.2.2 之（1）即知，亦可由下面（2）的证明中利用特征函数一并证明。

（2）利用特征函数。记 $t=\begin{pmatrix} t_{(1)} \\ t_{(2)} \end{pmatrix}$，其中 $t_{(1)}=(t_1, \cdots, t_q)'$，$t_{(2)}=(t_{q+1}, \cdots, t_p)'$。既然 $X\sim N_p(\mu, \Sigma)$ 则 X 的特征函数

$$\varphi_X(t)=E\{e^{it'X}\}=E\left\{\exp\left\{i\begin{pmatrix} t_{(1)} \\ t_{(2)} \end{pmatrix}'\begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}\right\}\right\}$$

$$= \exp\left\{ i \begin{pmatrix} t_{(1)} \\ t_{(2)} \end{pmatrix}' \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} t_{(1)} \\ t_{(2)} \end{pmatrix}' \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} t_{(1)} \\ t_{(2)} \end{pmatrix} \right\}$$

$$= \exp \{ i [t'_{(1)}\mu_{(1)} + t'_{(2)}\mu_{(2)}] - \frac{1}{2} [t'_{(1)}\Sigma_{11} t_{(1)} + t'_{(1)}\Sigma_{12} t_{(2)} + t'_{(2)}\Sigma_{21} t_{(1)} + t'_{(2)}\Sigma_{22} t_{(2)}] \}$$

$$= \exp \{ it'_{(1)}\mu_{(1)} - \frac{1}{2} t'_{(1)}\Sigma_{11} t_{(1)} \} \cdot \exp \{ it'_{(2)}\mu_{(2)} - \frac{1}{2} t'_{(2)}\Sigma_{22} t_{(2)} \} \cdot \exp \{ -\frac{1}{2} [t'_{(1)}\Sigma_{12} t_{(2)} + t'_{(2)}\Sigma_{21} t_{(1)}] \} \quad (1.2.5)$$

由（1）知（亦可由上式分别令 $t_{(1)} = \square$及 $t_{(2)} = \square$直接推出）

$$\varphi_{X_{(1)}}(t_{(1)}) = \exp \{ it'_{(1)}\mu_{(1)} - \frac{1}{2} t'_{(1)}\Sigma_{11} t_{(1)} \}$$

$$\varphi_{X_{(2)}}(t_{(2)}) = \exp \{ it'_{(2)}\mu_{(2)} - \frac{1}{2} t'_{(2)}\Sigma_{22} t_{(2)} \}$$

又根据（1.1.15）式知，$X_{(1)}$与 $X_{(2)}$相互独立的充要条件为：

$$\varphi_X(t) = \varphi_{X_{(1)}}(t_{(1)}) \cdot \varphi_{X_{(2)}}(t_{(2)})$$

即等价于

$$\exp \{ -\frac{1}{2} [t'_{(1)}\Sigma_{12} t_{(2)} + t'_{(2)}\Sigma_{21} t_{(1)}] \} = 1$$

即

$$t'_{(1)}\Sigma_{12} t_{(2)} + t'_{(2)}\Sigma_{21} t_{(1)}] = 0$$

由于$\Sigma'_{21} = \Sigma_{12}$，且 $t'_{(1)}\Sigma_{12} t_{(2)}$与 $t'_{(2)}\Sigma_{21} t_{(1)}$是实数，因此有

$$(t'_{(1)}\Sigma_{12} t_{(2)})' = (t'_{(2)}\Sigma_{21} t_{(1)}) = t'_{(1)}\Sigma_{12} t_{(2)}$$

所以 $t'_{(1)}\Sigma_{12} t_{(2)} + t'_{(2)}\Sigma_{21} t_{(1)} = 2t'_{(1)}\Sigma_{12} t_{(2)} = 0$，故$\Sigma_{12} = \square$

（3）与（4）的证明如下：记

$$Y = X_{(1)} - \Sigma_{12}\Sigma_{22}^{-1} X_{(2)}$$

$$Z = \begin{pmatrix} Y \\ X_{(2)} \end{pmatrix} = \begin{pmatrix} I_q & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-q} \end{pmatrix} \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix} \triangleq AX$$

由定理 1.2.2 之（1）知，Z 服从 $N_p(A\mu,\ A\Sigma A')$，其中

$$A\mu = \begin{pmatrix} \mu_{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu_{(2)} \\ \mu_{(2)} \end{pmatrix} \begin{matrix} {}^{q\times 1} \\ {}_{(p-q)\times 1} \end{matrix}$$

$$A\Sigma A' = A \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} A' = \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

由（2）知，Y 与 $X_{(2)}$独立，而且 Y 服从 $N_q(\mu_{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu_{(2)},\ \Sigma_{11\cdot 2})$。这个结论不论 $X_{(2)}$取怎么样的观察值向量都是成立的，也就是说，给定 $X_{(2)} = x_{(2)}$时，Y 的条