# 第1章 多元正态分布

## 学习目标

1. 掌握多元分布的有关概念；
2. 掌握统计距离的概念；
3. 理解多元正态分布的定义及其有关性质；
4. 了解常用多元分布及其抽样分布的定义和基本性质。

在基础统计学中，随机变量的正态分布在理论和实际应用中都有着重要的地位。同样，在多元统计学中，多元正态分布也占有相当重要的位置。原因是许多实际问题研究中的随机向量确实遵从或近似遵从多元正态分布；对于多元正态分布，已有一整套统计推断方法，并且可以得到许多完整的结果。

多元正态分布是最常用的一种多元概率分布。此外，还有多元对数正态分布、多项式分布、多元超几何分布、多元 $\beta$ 分布、多元 $\chi^2$ 分布、多元指数分布等。本章从多维变量及多元分布的基本概念开始，着重介绍多元正态分布的定义及一些重要性质，以及常用多元分布及其抽样分布的定义和基本性质。

## 1.1 多元分布的基本概念

在研究社会、经济现象和许多实际问题时，经常遇到的是多指标的问题。例如研究职工薪酬构成情况时，计时工资、基础工资与职务工资、各种奖金、各种津贴等都是同时需要考察的指标；又如研究公司的运营情况时，要涉及公司的资金周转能力、偿债能力、获

利能力及竞争能力等财务指标，这些都是多指标研究的问题。显然，由于这些指标之间往往不独立，仅研究某个指标或者将这些指标割裂开来分别研究，都不能从整体上把握所研究问题的实质。一般，假设所研究的问题涉及 $p$ 个指标，进行了 $n$ 次独立观测，将得到 $np$ 个数据，我们的目的就是对观测对象进行分组、分类，或分析这 $p$ 个变量之间的相互关联程度，或找出内在规律等。下面简要介绍多元分析中涉及的一些基本概念。

### 1.1.1 随机向量

假定所讨论的是多个变量的总体，所研究的数据是同时观测 $p$ 个指标（即变量），进行了 $n$ 次观测得到的，我们把这 $p$ 个指标表示为 $X_1$，$X_2$，…，$X_p$，常用向量

$$\boldsymbol{X}=(X_1,X_2,\cdots,X_p)'$$

表示对同一个体观测的 $p$ 个变量。若观测了 $n$ 个个体，则可得到如表 1-1 所示的数据，称每一个体的 $p$ 个变量为一个样品，而全体 $n$ 个样品形成一个样本。

**表 1-1**

| 序号 | 变量 | | | |
|---|---|---|---|---|
| | $X_1$ | $X_2$ | $\cdots$ | $X_p$ |
| 1 | $x_{11}$ | $x_{12}$ | $\cdots$ | $x_{1p}$ |
| 2 | $x_{21}$ | $x_{22}$ | $\cdots$ | $x_{2p}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ |
| $n$ | $x_{n1}$ | $x_{n2}$ | $\cdots$ | $x_{np}$ |

横看表 1-1，记

$$\boldsymbol{X}_{(\alpha)}=(x_{\alpha1},x_{\alpha2},\cdots,x_{\alpha p})',\quad \alpha=1,2,\cdots,n$$

它表示第 $\alpha$ 个样品的观测值。竖看表 1-1，第 $j$ 列的元素：

$$\boldsymbol{X}_j=(x_{1j},x_{2j},\cdots,x_{nj})',\quad j=1,2,\cdots,p$$

表示对第 $j$ 个变量 $X_j$ 的 $n$ 次观测数值。

因此，样本资料矩阵可用矩阵语言表示为：

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_p)=\begin{bmatrix} \boldsymbol{X}'_{(1)} \\ \boldsymbol{X}'_{(2)} \\ \vdots \\ \boldsymbol{X}'_{(n)} \end{bmatrix}$$

若无特别说明，本书所称向量均指列向量。

**定义 1.1** 设 $X_1$，$X_2$，…，$X_p$ 为 $p$ 个随机变量，由它们组成的向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$称为随机向量。

### 1.1.2 分布函数与密度函数

描述随机变量的最基本工具是分布函数。类似地，描述随机向量的最基本工具还是分

布函数。

**定义 1.2** 设 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$ 是一随机向量，它的多元分布函数是

$$F(\boldsymbol{x})=F(x_1,x_2,\cdots,x_p)=P(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_p\leqslant x_p) \tag{1.1}$$

式中，$\boldsymbol{x}=(x_1, x_2, \cdots, x_p)\in R^p$，并记成 $\boldsymbol{X}\sim F$。

多元分布函数的有关性质此处从略。

**定义 1.3** 设 $\boldsymbol{X}\sim F(\boldsymbol{x})=F(x_1, x_2, \cdots, x_p)$，若存在一个非负的函数 $f(\cdot)$，使得

$$F(\boldsymbol{x})=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p}f(t_1,t_2,\cdots,t_p)\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_p \tag{1.2}$$

对一切 $\boldsymbol{x}\in R^p$ 成立，则称 $\boldsymbol{X}$（或 $F(\boldsymbol{x})$）有分布密度 $f(\cdot)$，并称 $\boldsymbol{X}$ 为连续型随机向量。

一个 $p$ 维变量的函数 $f(\cdot)$ 能作为 $R^p$ 中某个随机向量的分布密度，当且仅当

(i) $f(\boldsymbol{x})\geqslant 0,\quad \forall \boldsymbol{x}\in R^p$；

(ii) $\int_{R^p}f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=1$。

**例 1-1**

若随机向量 $(X_1, X_2, X_3)$ 有密度函数

$$f(x_1,x_2,x_3)=x_1^2+6x_3^2+\frac{1}{3}x_1x_2$$

$$0<x_1<1,\quad 0<x_2<2,\quad 0<x_3<\frac{1}{2}$$

容易验证它符合分布密度函数的两个条件 (i) 和 (ii)。

最重要的连续型多元分布——多元正态分布将留在 1.3 节讨论。

### 1.1.3 多元变量的独立性

**定义 1.4** 两个随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 称为相互独立的，若

$$P(\boldsymbol{X}\leqslant\boldsymbol{x},\boldsymbol{Y}\leqslant\boldsymbol{y})=P(\boldsymbol{X}\leqslant\boldsymbol{x})P(\boldsymbol{Y}\leqslant\boldsymbol{y}) \tag{1.3}$$

对一切 $\boldsymbol{x}$，$\boldsymbol{y}$ 成立。若 $F(\boldsymbol{x}, \boldsymbol{y})$ 为 $(\boldsymbol{X}, \boldsymbol{Y})$ 的联合分布函数，$G(\boldsymbol{x})$ 和 $H(\boldsymbol{y})$ 分别为 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的分布函数，则 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 独立当且仅当

$$F(\boldsymbol{x},\boldsymbol{y})=G(\boldsymbol{x})H(\boldsymbol{y}) \tag{1.4}$$

若 $(\boldsymbol{X}, \boldsymbol{Y})$ 有密度 $f(\boldsymbol{x}, \boldsymbol{y})$，用 $g(\boldsymbol{x})$ 和 $h(\boldsymbol{y})$ 分别表示 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的分布密度，则 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 独立当且仅当

$$f(\boldsymbol{x},\boldsymbol{y})=g(\boldsymbol{x})h(\boldsymbol{y}) \tag{1.5}$$

注意在上述定义中，$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的维数一般是不同的。

类似地，若它们的联合分布等于各自分布的乘积，则称 $p$ 个随机向量 $\boldsymbol{X}_1$，$\boldsymbol{X}_2$，…，$\boldsymbol{X}_p$ 相互独立。由 $\boldsymbol{X}_1$，$\boldsymbol{X}_2$，…，$\boldsymbol{X}_p$ 相互独立可以推知任何 $\boldsymbol{X}_i$ 与 $\boldsymbol{X}_j$（$i\neq j$）独立，但是，若已知任何 $\boldsymbol{X}_i$ 与 $\boldsymbol{X}_j$（$i\neq j$）独立，并不能推出 $\boldsymbol{X}_1$，$\boldsymbol{X}_2$，…，$\boldsymbol{X}_p$ 相互独立。

### 1.1.4 随机向量的数字特征

1. 随机向量 $\boldsymbol{X}$ 的均值

设 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$ 有 $p$ 个分量。若 $E(X_i)=\mu_i$（$i=1, 2, \cdots, p$）存在，定义随机向量 $\boldsymbol{X}$ 的均值为：

$$E(\boldsymbol{X})=\begin{bmatrix}E(X_1)\\E(X_2)\\\vdots\\E(X_p)\end{bmatrix}=\begin{bmatrix}\mu_1\\\mu_2\\\vdots\\\mu_p\end{bmatrix}=\boldsymbol{\mu} \tag{1.6}$$

$\boldsymbol{\mu}$ 是一个 $p$ 维向量，称为均值向量。

当 $\boldsymbol{A}$，$\boldsymbol{B}$ 为常数矩阵时，由定义可立即推出如下性质：

(1) $E(\boldsymbol{AX})=\boldsymbol{A}E(\boldsymbol{X})$ (1.7)

(2) $E(\boldsymbol{AXB})=\boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}$ (1.8)

2. 随机向量 $\boldsymbol{X}$ 的协方差阵

$$\boldsymbol{\Sigma}=\mathrm{cov}(\boldsymbol{X},\boldsymbol{X})=E(\boldsymbol{X}-E\boldsymbol{X})(\boldsymbol{X}-E\boldsymbol{X})'=D(\boldsymbol{X})$$

$$=\begin{bmatrix}D(X_1)&\mathrm{cov}(X_1,X_2)&\cdots&\mathrm{cov}(X_1,X_p)\\\mathrm{cov}(X_2,X_1)&D(X_2)&\cdots&\mathrm{cov}(X_2,X_p)\\\vdots&\vdots&&\vdots\\\mathrm{cov}(X_p,X_1)&\mathrm{cov}(X_p,X_2)&\cdots&D(X_p)\end{bmatrix}$$

$$=(\sigma_{ij})_{p\times p} \tag{1.9}$$

称它为 $p$ 维随机向量 $\boldsymbol{X}$ 的协方差阵，简称为 $\boldsymbol{X}$ 的协方差阵。

称 $|\mathrm{cov}(\boldsymbol{X}, \boldsymbol{X})|$ 为 $\boldsymbol{X}$ 的广义方差，它是协方差阵的行列式之值。

3. 随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的协方差阵

设 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$ 和 $\boldsymbol{Y}=(Y_1, Y_2, \cdots, Y_q)'$ 分别为 $p$ 维和 $q$ 维随机向量，它们之间的协方差阵定义为一个 $p\times q$ 矩阵，其元素是 $\mathrm{cov}(X_i, Y_j)$，即

$$\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})=(\mathrm{cov}(X_i,Y_j)),\quad i=1,2,\cdots,p;j=1,2,\cdots,q \tag{1.10}$$

若 $\mathrm{cov}(\boldsymbol{X}, \boldsymbol{Y})=\boldsymbol{0}$，称 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是不相关的。

当 $\boldsymbol{A}$，$\boldsymbol{B}$ 为常数矩阵时，由定义可推出协方差阵有如下性质：

(1) $D(\boldsymbol{AX})=\boldsymbol{A}D(\boldsymbol{X})\boldsymbol{A}'=\boldsymbol{A\Sigma A}'$

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

(3) 设 $\boldsymbol{X}$ 为 $p$ 维随机向量，期望和协方差存在，记 $\boldsymbol{\mu}=E(\boldsymbol{X})$，$\boldsymbol{\Sigma}=D(\boldsymbol{X})$，$\boldsymbol{A}$ 为 $p\times p$

常数阵，则

$$E(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})+\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$$

对于任何随机向量 $\boldsymbol{X}=(X_1,\ X_2,\ \cdots,\ X_p)'$来说，其协方差阵 $\boldsymbol{\Sigma}$ 都是对称阵，同时总是非负定（也称半正定）的。大多数情形下是正定的。

4. 随机向量 $\boldsymbol{X}$ 的相关阵

若随机向量 $\boldsymbol{X}=(X_1,\ X_2,\ \cdots,\ X_p)'$的协方差阵存在，且每个分量的方差大于零，则 $\boldsymbol{X}$ 的相关阵定义为：

$$\boldsymbol{R}=(\mathrm{corr}(X_i,X_j))=(r_{ij})_{p\times p}$$

$$r_{ij}=\frac{\mathrm{cov}(X_i,X_j)}{\sqrt{D(X_i)}\sqrt{D(X_j)}},\quad i,j=1,2,\cdots,p \tag{1.11}$$

$r_{ij}$也称为分量 $X_i$与 $X_j$之间的（线性）相关系数。

对于两组不同的随机向量 $\boldsymbol{X}$ 及 $\boldsymbol{Y}$，它们之间的相关问题将在典型相关分析的章节中详细讨论。

在数据处理时，为了克服由于指标的量纲不同对统计分析结果的影响，往往在使用某种统计分析方法之前，将每个指标“标准化”，即做如下变换：

$$X_j^*=\frac{X_j-E(X_j)}{[D(X_j)]^{1/2}},\quad j=1,2,\cdots,p \tag{1.12}$$

$$\boldsymbol{X}^*=(X_1^*,X_2^*,\cdots,X_p^*)'$$

于是

$$E(\boldsymbol{X}^*)=\boldsymbol{0}$$

$$D(\boldsymbol{X}^*)=\mathrm{corr}(\boldsymbol{X})=\boldsymbol{R}$$

即标准化数据的协方差阵正好是原指标的相关阵：

$$\boldsymbol{R}=D(\boldsymbol{X}^*)=E(\boldsymbol{X}^*\boldsymbol{X}^{*\prime}) \tag{1.13}$$

## 1.2 统计距离

在多指标统计分析中，距离的概念十分重要，样品间的不少特征都可用距离来描述。大部分多元方法是建立在简单的距离概念基础上的，即平时人们熟悉的欧氏距离，或称直线距离。如几何平面上的点 $P=(x_1,\ x_2)$ 到原点 $O=(0,\ 0)$ 的欧氏距离，依勾股定理有

$$d(O,P)=(x_1^2+x_2^2)^{1/2} \tag{1.14}$$

一般，若点 $P$ 的坐标 $P=(x_1,\ x_2,\ \cdots,\ x_p)$，则它到原点 $O=(0,\ 0,\ \cdots,\ 0)$ 的欧氏距离，依勾股定理有

$$d(O,P)=\sqrt{x_1^2+x_2^2+\cdots+x_p^2} \tag{1.15}$$

所有与原点距离为 $C$ 的点满足方程

$$d^2(O,P)=x_1^2+x_2^2+\cdots+x_p^2=C^2 \tag{1.16}$$

因为这是一个球面方程（$p=2$ 时是圆），所以，与原点等距离的点构成一个球面，任意两个点 $P=(x_1, x_2, \cdots, x_p)$ 与 $Q=(y_1, y_2, \cdots, y_p)$ 之间的欧氏距离为：

$$d(P,Q)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_p-y_p)^2} \tag{1.17}$$

但就大部分统计问题而言，欧氏距离是不能令人满意的。这是因为每个坐标对欧氏距离的贡献是同等的。当坐标轴表示测量值时，它们往往带有大小不等的随机波动，在这种情况下，合理的办法是对坐标加权，使变化大的坐标比变化小的坐标有较小的权系数，这就产生了各种距离。

欧氏距离还有一个缺点，那就是当各个分量为不同性质的量时，“距离”的大小竟然与指标的单位有关。例如，横轴 $x_1$ 代表重量（以 kg 为单位），纵轴 $x_2$ 代表长度（以 cm 为单位）。有四个点 $A$，$B$，$C$，$D$，它们的坐标如图 1－1 所示。

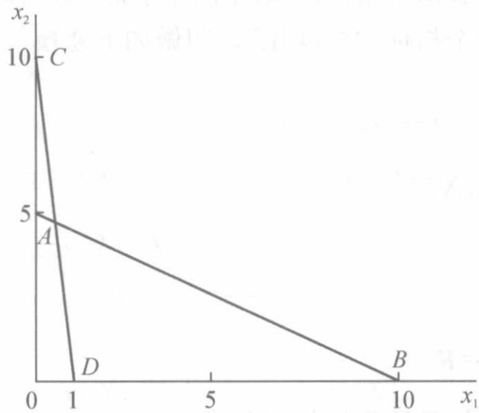


**图 1－1**

这时

$$AB=\sqrt{5^2+10^2}=\sqrt{125}$$

$$CD=\sqrt{10^2+1^2}=\sqrt{101}$$

显然，$AB$ 要比 $CD$ 长。

现在，如果 $x_2$ 用 mm 作单位，$x_1$ 单位保持不变，此时点 $A$ 坐标为（0，50），点 $C$ 坐标为（0，100），则

$$AB=\sqrt{50^2+10^2}=\sqrt{2\,600}$$

$$CD=\sqrt{100^2+1^2}=\sqrt{10\,001}$$

结果 $CD$ 反而比 $AB$ 长！这显然是不够合理的。因此，有必要建立一种距离，这种距离应能够体现各个变量在变差大小上的不同，以及有时存在的相关性，还要求距离与各变量所

用的单位无关。看来，我们选择的距离要依赖于样本方差和协方差。因此，采用“统计距离”这个术语，以区别通常习惯用的欧氏距离。

下面先介绍统计距离。

设 $P=(x_1, x_2, \cdots, x_p)$，$Q=(y_1, y_2, \cdots, y_p)$，且 $Q$ 的坐标是固定的，$P$ 的坐标相互独立地变化。用 $S_{11}$，$S_{22}$，…，$S_{pp}$ 表示 $p$ 个变量 $X_1$，$X_2$，…，$X_p$ 的 $n$ 次观测的样本方差。将坐标的各维度除以相应变量的样本标准差 $\sqrt{S_{ii}}$，得到标准化的坐标，其中各变量的样本标准差的倒数可以看作坐标各维度的权重系数，则 $P$ 到 $Q$ 的统计距离为：

$$d(P,Q)=\sqrt{\frac{(x_1-y_1)^2}{S_{11}}+\frac{(x_2-y_2)^2}{S_{22}}+\cdots+\frac{(x_p-y_p)^2}{S_{pp}}} \tag{1.18}$$

所有与点 $Q$ 的距离平方为常数的点 $P$ 构成一个椭球，其中心在点 $Q$，其长短轴平行于坐标轴。容易看到：

(1) 在式 (1.18) 中，令 $y_1=y_2=\cdots=y_p=0$，得到点 $P$ 到原点 $O$ 的距离。

(2) 如果 $S_{11}=S_{22}=\cdots=S_{pp}$，则用欧氏距离式 (1.17) 是方便可行的。

还可以利用旋转变换的方法得到合理的距离。考虑点 $P=(x_1, x_2, \cdots, x_p)$ 和点 $Q=(y_1, y_2, \cdots, y_p)$，这里 $Q$ 为固定点，而 $P$ 的坐标是变化的，且彼此相关，$O=(0, 0, \cdots, 0)$ 为坐标原点，则 $P$ 到 $O$ 和 $Q$ 的距离分别为：

$$\begin{aligned} d(O,P)&=(a_{11}x_1^2+a_{22}x_2^2+\cdots+a_{pp}x_p^2+2a_{12}x_1x_2+\cdots+2a_{p-1,p}x_{p-1}x_p)^{1/2} \\ &=(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})^{1/2} \end{aligned} \tag{1.19}$$

和

$$\begin{aligned} d(P,Q)=&[a_{11}(x_1-y_1)^2+a_{22}(x_2-y_2)^2+\cdots+a_{pp}(x_p-y_p)^2 \\ &+2a_{12}(x_1-y_1)(x_2-y_2)+\cdots \\ &+2a_{p-1,p}(x_{p-1}-y_{p-1})(x_p-y_p)]^{1/2} \\ =&[(\boldsymbol{X}-\boldsymbol{Y})'\boldsymbol{A}(\boldsymbol{X}-\boldsymbol{Y})]^{1/2} \end{aligned} \tag{1.20}$$

这里

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}, \quad \boldsymbol{X}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad \boldsymbol{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

且 $\boldsymbol{A}$ 为对称阵，满足条件：对任意的 $\boldsymbol{X}$，恒有 $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\geqslant \boldsymbol{0}$，且等号成立当且仅当 $\boldsymbol{X}=\boldsymbol{0}$，即 $\boldsymbol{A}$ 为正定方阵。

最常用的一种统计距离是印度统计学家马哈拉诺比斯 (Mahalanobis) 于 1936 年引入的，称为马氏距离。下面先用一个一维的例子说明欧氏距离与马氏距离在概率上的差异。设有两个一维正态总体 $G_1$：$N(\mu_1, \sigma_1^2)$ 和 $G_2$：$N(\mu_2, \sigma_2^2)$。若有一个样品，其值在点 $A$ 处，点 $A$ 距离哪个总体近些呢？如图 1－2 所示。

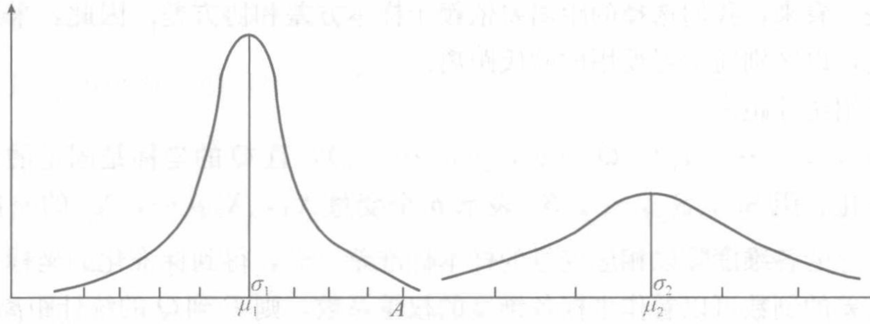


图 1－2

由图 1－2 可看出，从绝对长度来看，点 $A$ 距左面总体 $G_1$ 近些，即点 $A$ 到 $\mu_1$ 比点 $A$ 到 $\mu_2$ 要近一些（这里用的是欧氏距离，比较的是点 $A$ 坐标与 $\mu_1$ 到 $\mu_2$ 值之差的绝对值），但从概率观点来看，点 $A$ 在 $\mu_1$ 右侧约 $4\sigma_1$ 处，点 $A$ 在 $\mu_2$ 的左侧约 $3\sigma_2$ 处，若以标准差的观点来衡量，点 $A$ 离 $\mu_2$ 比离 $\mu_1$ 要近一些。显然，后者是从概率角度来考虑的，因而更为合理，它是用坐标差平方除以方差（或说乘以方差的倒数），从而转化为无量纲数的，推广到多维就要乘以协方差阵 $\boldsymbol{\Sigma}$ 的逆矩阵 $\boldsymbol{\Sigma}^{-1}$，这就是马氏距离的概念。以后将会看到，这一距离在多元分析中起着十分重要的作用。

有了上面的讨论，现在可以定义马氏距离了。

设 $\boldsymbol{X}$，$\boldsymbol{Y}$ 是从均值向量为 $\boldsymbol{\mu}$，协方差阵为 $\boldsymbol{\Sigma}$ 的总体 $G$ 中抽取的两个样品，定义 $\boldsymbol{X}$，$\boldsymbol{Y}$ 两点之间的马氏距离为：

$$d_m^2(\boldsymbol{X},\boldsymbol{Y})=(\boldsymbol{X}-\boldsymbol{Y})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{Y}) \tag{1.21}$$

定义 $\boldsymbol{X}$ 与总体 $G$ 的马氏距离为：

$$d_m^2(\boldsymbol{X},G)=(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \tag{1.22}$$

设 $E$ 表示一个点集，$d$ 表示距离，它是 $E\times E$ 到 $[0,\infty)$ 的函数，可以证明，马氏距离符合如下距离的四条基本公理：

(1) $d(x, y)\geqslant 0$，$\forall x, y\in E$；

(2) $d(x, y)=0$，当且仅当 $x=y$；

(3) $d(x, y)=d(y, x)$，$\forall x, y\in E$；

(4) $d(x, y)\leqslant d(x, z)+d(z, y)$，$\forall x, y, z\in E$。

## 1.3 多元正态分布

多元正态分布是一元正态分布的推广。迄今为止，多元分析的主要理论都是建立在多元正态总体基础上的，多元正态分布是多元分析的基础。另外，许多实际问题的分布常是多元正态分布或近似正态分布，或虽然本身不是正态分布，但它的样本均值近似于多元正态分布。

本节将介绍多元正态分布的定义，并简要给出它的基本性质。

### 1.3.1 多元正态分布的定义

在概率论中学过，一元正态分布的密度函数为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \sigma>0$$

上式可以改写成

$$f(x)=(2\pi)^{-1/2}\sigma^{-1}\exp\left[-\frac{1}{2}(x-\mu)'(\sigma^2)^{-1}(x-\mu)\right] \tag{1.23}$$

式（1.23）用 $(x-\mu)'$代表 $(x-\mu)$ 的转置。由于 $x$，$\mu$ 均为一维的数字，转置与否都相同，所以可以这样写。

当遵从一元正态分布的随机变量 $X$ 的概率密度函数改写为式（1.23）时，我们就可以将其推广，给出多元正态分布的定义。

**定义 1.5** 若 $p$ 元随机向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$的概率密度函数为：

$$f(x_1,x_2,\cdots,x_p)=\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \quad \boldsymbol{\Sigma}>\boldsymbol{0} \tag{1.24}$$

则称 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$遵从 $p$ 元正态分布，也称 $\boldsymbol{X}$ 为 $p$ 元正态变量，记为：

$$\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$$

$|\boldsymbol{\Sigma}|$ 为协方差阵 $\boldsymbol{\Sigma}$ 的行列式。

式（1.24）实际是在 $|\boldsymbol{\Sigma}|\neq 0$ 时定义的。若 $|\boldsymbol{\Sigma}|=0$，则不存在通常意义下的密度，但可以在形式上给出一个表达式，使有些问题可以利用这一形式对 $|\boldsymbol{\Sigma}|\neq 0$及 $|\boldsymbol{\Sigma}|=0$ 的情况给出统一的处理。

当 $p=2$ 时，可以得到二元正态分布的密度公式。

设 $\boldsymbol{X}=(X_1, X_2)'$遵从二元正态分布，则

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}=\begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2 r\\ \sigma_2\sigma_1 r & \sigma_2^2\end{bmatrix}, \quad r\neq\pm 1$$

这里 $\sigma_1^2$，$\sigma_2^2$ 分别是 $X_1$ 与 $X_2$ 的方差，$r$ 是 $X_1$ 与 $X_2$ 的相关系数。此时

$$|\boldsymbol{\Sigma}|=\sigma_1^2\sigma_2^2(1-r^2)$$

$$\boldsymbol{\Sigma}^{-1}=\frac{1}{\sigma_1^2\sigma_2^2(1-r^2)}\begin{bmatrix}\sigma_2^2 & -\sigma_1\sigma_2 r\\ -\sigma_2\sigma_1 r & \sigma_1^2\end{bmatrix}$$

故 $X_1$ 与 $X_2$ 的密度函数为：

$$f(x_1, x_2)=\frac{1}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2r\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

这与我们学过的概率统计中的结果是一致的。

如果 $r=0$，那么 $X_1$ 与 $X_2$ 是独立的；若 $r>0$，则 $X_1$ 与 $X_2$ 趋于正相关；若 $r<0$，则 $X_1$ 与 $X_2$ 趋于负相关。

**定理 1.1** 设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则

$$E(\boldsymbol{X})=\boldsymbol{\mu}, \quad D(\boldsymbol{X})=\boldsymbol{\Sigma}$$

定理 1.1 将正态分布的参数 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 赋予了明确的统计意义。有关这个定理的证明可参见参考文献 [3]。

多元正态分布不止定义 1.5 一种形式，更广泛的可采用特征函数来定义，也可用一切线性组合均为正态的性质来定义等。有关这些定义的方式参见参考文献 [3]。

### 1.3.2 多元正态分布的性质

(1) 如果正态随机向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$ 的协方差阵 $\boldsymbol{\Sigma}$ 是对角阵，则 $\boldsymbol{X}$ 的各分量是相互独立的随机变量。证明参见参考文献 [4]。

(2) 多元正态随机向量 $\boldsymbol{X}$ 的任何一个分量子集（多变量 $(X_1, X_2, \cdots, X_p)'$ 中的一部分变量构成的集合）的分布（称为 $\boldsymbol{X}$ 的边缘分布）仍然遵从正态分布。反之，若一个随机向量的任何边缘分布均为正态分布，并不能导出它是多元正态分布。

例如，设 $\boldsymbol{X}=(X_1, X_2)'$ 有分布密度

$$f(x_1,x_2)=\frac{1}{2\pi}\mathrm{e}^{-\frac{1}{2}(x_1^2+x_2^2)}\left[1+x_1x_2\mathrm{e}^{-\frac{1}{2}(x_1^2+x_2^2)}\right]$$

容易验证，$X_1 \sim N(0, 1)$，$X_2 \sim N(0, 1)$，但 $(X_1, X_2)$ 显然不遵从正态分布。

(3) 多元正态向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)'$ 的任意线性变换仍然遵从多元正态分布。

即设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，而 $m$ 维随机向量 $\boldsymbol{Z}_{m\times 1}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{b}$，其中 $\boldsymbol{A}=(a_{ij})$ 是 $m\times p$ 阶的常数矩阵，$\boldsymbol{b}$ 是 $m$ 维的常向量，则 $m$ 维随机向量 $\boldsymbol{Z}$ 也是正态的，且 $\boldsymbol{Z} \sim N_m(\boldsymbol{A\mu}+\boldsymbol{b}, \boldsymbol{A\Sigma A}')$。即 $\boldsymbol{Z}$ 遵从 $m$ 元正态分布，其均值向量为 $\boldsymbol{A\mu}+\boldsymbol{b}$，协方差阵为 $\boldsymbol{A\Sigma A}'$。

(4) 若 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则

$$d^2=(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi^2(p)$$

$d^2$ 若为定值，随着 $\boldsymbol{X}$ 的变化，其轨迹为一椭球面，是 $\boldsymbol{X}$ 的密度函数的等值面。若 $\boldsymbol{X}$ 给定，则 $d^2$ 为 $\boldsymbol{X}$ 到 $\boldsymbol{\mu}$ 的马氏距离。

### 1.3.3 条件分布和独立性

设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$p \geqslant 2$，将 $\boldsymbol{X}$，$\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 剖分如下：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{bmatrix}, \quad \boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}^{(1)}\\ \boldsymbol{\mu}^{(2)}\end{bmatrix}, \quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix} \tag{1.25}$$

其中，$\boldsymbol{X}^{(1)}$，$\boldsymbol{\mu}^{(1)}$ 为 $q\times 1$ 维，$\boldsymbol{\Sigma}_{11}$ 为 $q\times q$ 维，我们希望求给定 $\boldsymbol{X}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件分布，即 $(\boldsymbol{X}^{(1)} \mid \boldsymbol{X}^{(2)})$ 的分布。下一个定理指出：正态分布的条件分布仍为正态分布。

**定理 1.2** 设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>0$，则

$$(\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)})\sim N_q(\boldsymbol{\mu}_{1\cdot 2},\boldsymbol{\Sigma}_{11\cdot 2})$$

其中

$$\boldsymbol{\mu}_{1\cdot 2}=\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)}) \tag{1.26}$$

$$\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{1.27}$$

证明参见参考文献［3］。

该定理告诉我们，$\boldsymbol{X}^{(1)}$的分布与（$\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)}$）的分布均为正态，它们的协方差阵分别为$\boldsymbol{\Sigma}_{11}$与$\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$。由于$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\geqslant\boldsymbol{0}$，故$\boldsymbol{\Sigma}_{11}\geqslant\boldsymbol{\Sigma}_{11\cdot 2}$，等号成立当且仅当$\boldsymbol{\Sigma}_{12}=\boldsymbol{0}$。协差阵是用来描述指标之间关系及散布程度的，$\boldsymbol{\Sigma}_{11}\geqslant\boldsymbol{\Sigma}_{11\cdot 2}$，说明已知$\boldsymbol{X}^{(2)}$的条件下，$\boldsymbol{X}^{(1)}$散布的程度比不知道$\boldsymbol{X}^{(2)}$的情况下减小了，只有当$\boldsymbol{\Sigma}_{12}=\boldsymbol{0}$时，两者相同。还可以证明，$\boldsymbol{\Sigma}_{12}=\boldsymbol{0}$，等价于$\boldsymbol{X}^{(1)}$和$\boldsymbol{X}^{(2)}$独立，这时，即使给出$\boldsymbol{X}^{(2)}$，对$\boldsymbol{X}^{(1)}$的分布也是没有影响的。

**定理 1.3** 设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>0$，将$\boldsymbol{X}$，$\boldsymbol{\mu}$，$\boldsymbol{\Sigma}$剖分如下：

$$\boldsymbol{X}=\begin{array}{c}\begin{bmatrix}\boldsymbol{X}^{(1)}\\\boldsymbol{X}^{(2)}\\\boldsymbol{X}^{(3)}\end{bmatrix}\end{array}\begin{array}{l}r\\s\\t\end{array},\quad \boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}^{(1)}\\\boldsymbol{\mu}^{(2)}\\\boldsymbol{\mu}^{(3)}\end{bmatrix}\begin{array}{l}r\\s\\t\end{array},\quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11}&\boldsymbol{\Sigma}_{12}&\boldsymbol{\Sigma}_{13}\\\boldsymbol{\Sigma}_{21}&\boldsymbol{\Sigma}_{22}&\boldsymbol{\Sigma}_{23}\\\boldsymbol{\Sigma}_{31}&\boldsymbol{\Sigma}_{32}&\boldsymbol{\Sigma}_{33}\end{bmatrix}\begin{array}{l}r\\s\\t\end{array} \tag{1.28}$$

则$\boldsymbol{X}^{(1)}$有如下的条件均值和条件协方差阵的递推公式：

$$E(\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)},\boldsymbol{X}^{(3)})=\boldsymbol{\mu}_{1\cdot 3}+\boldsymbol{\Sigma}_{12\cdot 3}\boldsymbol{\Sigma}_{22\cdot 3}^{-1}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}_{2\cdot 3}) \tag{1.29}$$

$$D(\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)},\boldsymbol{X}^{(3)})=\boldsymbol{\Sigma}_{11\cdot 3}-\boldsymbol{\Sigma}_{12\cdot 3}\boldsymbol{\Sigma}_{22\cdot 3}^{-1}\boldsymbol{\Sigma}_{21\cdot 3} \tag{1.30}$$

其中 $\boldsymbol{\Sigma}_{ij\cdot k}=\boldsymbol{\Sigma}_{ij}-\boldsymbol{\Sigma}_{ik}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\Sigma}_{kj}$，　$i$，$j$，$k=1$，2，3

$$\boldsymbol{\mu}_{i\cdot 3}=\boldsymbol{E}(\boldsymbol{X}^{(i)}\mid\boldsymbol{X}^{(3)}),\quad i=1,2$$

证明参见参考文献［3］。

定理1.2和定理1.3在20世纪70年代中期国家标准部门制定服装标准时有成功的应用（参见参考文献［3］)。在制定服装标准时需抽样进行人体测量，现从某年龄段女性测量取出部分结果：

$X_1$：身高，$X_2$：胸围，$X_3$：腰围，$X_4$：上体长，$X_5$：臀围。已知它们遵从$N_5(\boldsymbol{\mu},\boldsymbol{\Sigma})$，其中

$$\boldsymbol{\mu}=\begin{bmatrix}154.98\\83.39\\70.26\\61.32\\91.52\end{bmatrix},\quad \boldsymbol{\Sigma}=\begin{bmatrix}29.66&&&&\\6.51&30.53&&&\\1.85&25.54&39.86&&\\9.36&3.54&2.23&7.03&\\10.34&19.53&20.70&5.21&27.36\end{bmatrix}$$

若取 $\boldsymbol{X}^{(1)}=(X_1, X_2, X_3)'$，$\boldsymbol{X}^{(2)}=(X_4)$，$\boldsymbol{X}^{(3)}=(X_5)$，则由式（1.26）和式（1.27）得

$$E\left[\begin{matrix}X_1\\X_2\\X_3\\X_4\end{matrix}\middle| X_5\right]=\begin{bmatrix}154.98\\83.39\\70.26\\61.32\end{bmatrix}+\begin{bmatrix}10.34\\19.53\\20.70\\5.21\end{bmatrix}(27.36)^{-1}(X_5-91.52)$$

$$=\begin{bmatrix}154.98+0.38(X_5-91.52)\\83.39+0.71(X_5-91.52)\\70.26+0.76(X_5-91.52)\\61.32+0.19(X_5-91.52)\end{bmatrix}$$

$$D\left[\begin{matrix}X_1\\X_2\\X_3\\X_4\end{matrix}\middle| X_5\right]=\begin{bmatrix}29.66&6.51&1.85&9.36\\6.51&30.53&25.54&3.54\\1.85&25.54&39.86&2.23\\9.36&3.54&2.23&7.03\end{bmatrix}$$

$$-\begin{bmatrix}10.34\\19.53\\20.70\\5.21\end{bmatrix}(27.36)^{-1}(10.34,19.53,20.70,5.21)$$

$$=\begin{bmatrix}25.76&-0.86&-5.97&7.39\\-0.86&16.59&10.76&-0.18\\-5.97&10.76&24.19&-1.72\\7.39&-0.18&-1.72&6.04\end{bmatrix}$$

再利用式（1.30）得

$$D\left[\begin{matrix}X_1\\X_2\\X_3\end{matrix}\middle|\begin{matrix}X_4\\ \\X_5\end{matrix}\right]=\begin{bmatrix}25.76&-0.86&-5.97\\-0.86&16.59&10.76\\-5.97&10.76&24.19\end{bmatrix}$$

$$-\begin{bmatrix}7.39\\-0.18\\-1.72\end{bmatrix}(6.04)^{-1}(7.39,-0.18,-1.72)$$

$$=\begin{bmatrix}16.72&-0.64&-3.87\\-0.64&16.58&10.71\\-3.87&10.71&23.71\end{bmatrix}$$

此时看到

$$D(X_1|X_4,X_5)=16.72<29.66=D(X_1)$$
$$D(X_2|X_4,X_5)=16.58<30.53=D(X_2)$$
$$D(X_3|X_4,X_5)=23.71<39.86=D(X_3)$$

这说明，若已知一个人的上体长和臀围，则身高、胸围和腰围的条件方差比原来的方差大大减小。

在定理 1.2 中，我们给出了对 $\boldsymbol{X}$，$\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 做形如式（1.25）的剖分时条件协方差阵 $\boldsymbol{\Sigma}_{11\cdot 2}$ 的表达式及其与非条件协方差阵的关系。令 $\sigma_{ij\cdot q+1,\cdots,p}$ 表示 $\boldsymbol{\Sigma}_{11\cdot 2}$ 的元素，则可以定义偏相关系数的概念如下：

**定义 1.6** 当 $\boldsymbol{X}^{(2)}$ 给定时，$X_i$ 与 $X_j$ 的偏相关系数为：

$$r_{ij\cdot q+1,\cdots,p}=\frac{\sigma_{ij\cdot q+1,\cdots,p}}{(\sigma_{ii\cdot q+1,\cdots,p}\sigma_{jj\cdot q+1,\cdots,p})^{1/2}}$$

在上面制定服装标准的例子中，给出 $X_4$ 和 $X_5$ 时，$X_1$ 与 $X_2$，$X_1$ 与 $X_3$，$X_2$ 与 $X_3$ 的偏相关系数为：

$$r_{12\cdot 45}=\frac{-0.643}{\sqrt{16.717\times 16.582}}=-0.0386$$

$$r_{13\cdot 45}=\frac{-3.873}{\sqrt{16.717\times 23.707}}=-0.195$$

$$r_{23\cdot 45}=\frac{10.707}{\sqrt{16.582\times 23.707}}=0.540$$

**定理 1.4** 设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，将 $\boldsymbol{X}$，$\boldsymbol{\mu}$，$\boldsymbol{\Sigma}$ 按同样方式剖分为：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}^{(1)}\\ \vdots\\ \boldsymbol{X}^{(k)}\end{bmatrix},\quad \boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}^{(1)}\\ \vdots\\ \boldsymbol{\mu}^{(k)}\end{bmatrix},\quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1k}\\ \vdots & & \vdots\\ \boldsymbol{\Sigma}_{k1} & \cdots & \boldsymbol{\Sigma}_{kk}\end{bmatrix}$$

其中，$\boldsymbol{X}^{(j)}$：$S_j\times 1$，$\boldsymbol{\mu}^{(j)}$：$S_j\times 1$，$\boldsymbol{\Sigma}_{jj}$：$S_j\times S_j$（$j=1$，…，$k$），则 $\boldsymbol{X}^{(1)}$，$\boldsymbol{X}^{(2)}$，…，$\boldsymbol{X}^{(k)}$ 相互独立当且仅当 $\boldsymbol{\Sigma}_{ij}=\boldsymbol{0}$，对一切 $i\neq j$。

证明参见参考文献［3］。

因为 $\boldsymbol{\Sigma}_{12}=\mathrm{cov}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$，该定理同时指出对多元正态分布而言，“$\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 不相关”等价于“$\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 独立”。

## 1.4 均值向量和协方差阵的估计

上节已经给出了多元正态分布的定义和有关的性质，在实际问题中，通常可以假定研究对象是多元正态分布，但分布中的参数 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 是未知的，一般的做法是通过样本来估计。

在一般情况下，如果样本资料阵为：

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_p)=\begin{bmatrix} \boldsymbol{X}'_{(1)} \\ \boldsymbol{X}'_{(2)} \\ \vdots \\ \boldsymbol{X}'_{(n)} \end{bmatrix}$$

设样品 $\boldsymbol{X}_{(1)}$，$\boldsymbol{X}_{(2)}$，…，$\boldsymbol{X}_{(n)}$ 相互独立，同遵从于 $p$ 元正态分布 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，而且 $n>p$，$\boldsymbol{\Sigma}>\mathbf{0}$，则总体参数均值 $\boldsymbol{\mu}$ 的估计量为：

$$\hat{\boldsymbol{\mu}}=\overline{\boldsymbol{X}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{(i)}=\frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i2} \\ \vdots \\ \sum_{i=1}^{n} x_{ip} \end{bmatrix}=\begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_p \end{bmatrix} \tag{1.31}$$

即均值向量 $\boldsymbol{\mu}$ 的估计量就是样本均值向量。这可由极大似然法推导出来。很显然，当样本资料选取的是 $p$ 个指标的数据时，当然 $\hat{\boldsymbol{\mu}}=\overline{\boldsymbol{X}}$ 也是 $p$ 维向量。

总体参数协方差阵 $\boldsymbol{\Sigma}$ 的极大似然估计为：

$$\hat{\boldsymbol{\Sigma}}_p=\frac{1}{n}\boldsymbol{L}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_{(i)}-\overline{\boldsymbol{X}})(\boldsymbol{X}_{(i)}-\overline{\boldsymbol{X}})'$$

$$=\frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n}(x_{i1}-\overline{X}_1)^2 & \cdots & \sum_{i=1}^{n}(x_{i1}-\overline{X}_1)(x_{ip}-\overline{X}_p) \\ & \sum_{i=1}^{n}(x_{i2}-\overline{X}_2)^2 \cdots & \sum_{i=1}^{n}(x_{i2}-\overline{X}_2)(x_{ip}-\overline{X}_p) \\ & & \vdots \\ & & \sum_{i=1}^{n}(x_{ip}-\overline{X}_p)^2 \end{bmatrix}$$

其中 $\boldsymbol{L}$ 是离差阵，它是每一个样品（向量）与样本均值（向量）的离差积形成的 $n$ 个 $p\times p$ 阶对称阵的和。同一元相似，$\hat{\boldsymbol{\Sigma}}_p$ 不是 $\boldsymbol{\Sigma}$ 的无偏估计。为了得到无偏估计，我们常用样本协方差阵 $\hat{\boldsymbol{\Sigma}}=\frac{1}{n-1}\boldsymbol{L}$ 作为总体协方差阵的估计。

可以证明，$\overline{\boldsymbol{X}}$ 是 $\boldsymbol{\mu}$ 的无偏估计，是极小极大估计，是强相合估计，$\overline{\boldsymbol{X}}$ 还是 $\boldsymbol{\mu}$ 的充分统计量；$\hat{\boldsymbol{\Sigma}}$ 是 $\boldsymbol{\Sigma}$ 的强相合估计，但用 $\hat{\boldsymbol{\Sigma}}$ 估计 $\boldsymbol{\Sigma}$ 是有偏的，$\frac{1}{n-1}\boldsymbol{L}$ 才是 $\boldsymbol{\Sigma}$ 的无偏估计。在实际应用中，当 $n$ 不是很大时，人们常用$\frac{1}{n-1}\boldsymbol{L}$ 来估计 $\boldsymbol{\Sigma}$，但当 $n$ 比较大时，用 $\hat{\boldsymbol{\Sigma}}$ 或$\frac{1}{n-1}\boldsymbol{L}$ 差别不大。关于估计量 $\overline{\boldsymbol{X}}$，$\hat{\boldsymbol{\Sigma}}$ 的统计性质的证明，有兴趣的读者可参见参考文献［3］。

## 1.5 常用分布及抽样分布

多元统计研究的是多指标问题，为了解总体的特征，通过对总体抽样得到代表总体的样本，但因为信息是分散在每个样本上的，就需要对样本进行加工，把样本的信息浓缩到不包含未知量的样本函数中，这个函数称为统计量，如前面介绍的样本均值向量 $\overline{\boldsymbol{X}}$、样本离差阵 $\boldsymbol{L}$ 等都是统计量。统计量的分布称为抽样分布。

在数理统计中常用的抽样分布有 $\chi^2$ 分布、$t$ 分布和 $F$ 分布。在多元统计中，与之对应的分布分别为 Wishart 分布、$T^2$ 分布和 Wilks 分布。

### 1.5.1 $\chi^2$ 分布与 Wishart 分布

在数理统计中，若 $X_i \sim N(0,\ 1)(i=1,\ 2,\ \cdots,\ n)$，且相互独立，则 $\sum\limits_{i=1}^{n} X_i^2$ 所遵从的分布为自由度为 $n$ 的 $\chi^2$ 分布（chi-squared distribution），记为 $\chi^2(n)$。

$\chi^2(n)$ 分布是刻画正态变量二次型的一个重要分布，在一元统计分析中有着十分重要的地位，在对有关样本均值、样本方差的假设检验或非参数检验中经常用到 $\chi^2$ 统计量。

$\chi^2(n)$ 分布的均值和方差分别为：

$$E(\chi^2(n))=n$$
$$D(\chi^2(n))=2n$$

$\chi^2(n)$ 分布有两个重要的性质：

（1）若 $\chi_i^2 \sim \chi^2(n_i)(i=1,\ 2,\ \cdots,\ k)$，且相互独立，则

$$\sum_{i=1}^{k} \chi_i^2 \sim \chi^2\left(\sum_{i=1}^{k} n_i\right)$$

称为相互独立的 $\chi^2$ 变量具有可加性。

（2）设 $X_i \sim N(0,\ 1)(i=1,\ 2,\ \cdots,\ n)$，且相互独立，$\boldsymbol{A}_j\ (j=1,\ 2,\ \cdots,\ m)$ 为 $n$ 阶对称阵，且 $\sum\limits_{j=1}^{m} \boldsymbol{A}_j = \boldsymbol{I}_n$（$n$ 阶单位阵），记 $\boldsymbol{X}=(X_1,\ X_2,\ \cdots,\ X_n)'$，$Q_j=\boldsymbol{X}'\boldsymbol{A}_j\boldsymbol{X}$，则 $Q_1$，$Q_2$，$\cdots$，$Q_m$ 为相互独立的 $\chi^2$ 分布变量的充要条件为 $\sum\limits_{j=1}^{m} \mathrm{rank}(\boldsymbol{A}_j)=n$。此时 $Q_j \sim \chi^2(n_j)$，$n_j=\mathrm{rank}(\boldsymbol{A}_j)$。

这个性质称为 Cochran 定理，在方差分析和回归分析中起着重要作用。

从一元正态总体 $N(\mu,\ \sigma^2)$ 抽取容量为 $n$ 的随机样本 $X_1$，$X_2$，$\cdots$，$X_n$，其样本均值 $\overline{X}$ 和样本方差 $S^2=\dfrac{l_{xx}}{n-1}=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$ 的抽样分布有如下结果：

（1）$\overline{X}$ 和 $S^2$ 相互独立。

（2）$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$和$\frac{(n-1)S^2}{\sigma^2}=\frac{l_{xx}}{\sigma^2} \sim \chi^2(n-1)$ 相互独立。

以上两个结论在数理统计中有着重要的应用。

在多元统计中，$\chi^2$ 分布发展为 Wishart 分布。Wishart 分布是由统计学家威沙特（Wishart）为研究多元样本离差阵 $\boldsymbol{L}$ 的分布于 1928 年推导出来的，有人就将这个时间作为多元分析诞生的时间。Wishart 分布在多元统计中的作用与 $\chi^2$ 分布在一元统计中类似，它可以由遵从多元正态分布的随机向量直接得到，同时它也是构成其他重要分布的基础。

**定义 1.7** 设 $\boldsymbol{X}_{(\alpha)}(\alpha=1, 2, \cdots, n)$ 相互独立，且 $\boldsymbol{X}_{(\alpha)} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$，记 $\boldsymbol{X}=(\boldsymbol{X}_{(1)}, \boldsymbol{X}_{(2)}, \cdots, \boldsymbol{X}_{(n)})'$，则随机矩阵

$$\boldsymbol{W}=\boldsymbol{X}'\boldsymbol{X}=\sum_{\alpha=1}^{n} \boldsymbol{X}_{(\alpha)}\boldsymbol{X}'_{(\alpha)} \tag{1.32}$$

所遵从的分布称为自由度为 $n$ 的 $p$ 维中心 Wishart 分布，记为 $\boldsymbol{W} \sim \boldsymbol{W}_p(n, \boldsymbol{\Sigma})$。其中，$n \geqslant p$，$\boldsymbol{\Sigma}>\boldsymbol{0}$。

由 Wishart 分布的定义知，当 $p=1$ 时，$\boldsymbol{\Sigma}$ 退化为 $\sigma^2$，此时中心 Wishart 分布就退化为 $\sigma^2\chi^2(n)$，由此可以看出，Wishart 分布实际上是 $\chi^2$ 分布在多维正态情形下的推广。

下面不加证明地给出 Wishart 分布的五条重要性质：

（1）若 $\boldsymbol{X}_{(\alpha)}(\alpha=1, 2, \cdots, n)$是从 $p$ 维正态总体 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 中抽取的 $n$ 个随机样本，$\bar{\boldsymbol{X}}$ 为样本均值，样本离差阵为 $\boldsymbol{L}=\sum_{\alpha=1}^{n}(\boldsymbol{X}_{(\alpha)}-\bar{\boldsymbol{X}})(\boldsymbol{X}_{(\alpha)}-\bar{\boldsymbol{X}})'$，则

1）$\bar{\boldsymbol{X}}$ 和 $\boldsymbol{L}$ 相互独立。

2）$\bar{\boldsymbol{X}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$，$\boldsymbol{L} \sim \boldsymbol{W}_p(n-1, \boldsymbol{\Sigma})$。

（2）若 $\boldsymbol{W}_i \sim \boldsymbol{W}_p(n_i, \boldsymbol{\Sigma})(i=1, 2, \cdots, k)$且相互独立，则

$$\sum_{i=1}^{k} \boldsymbol{W}_i \sim \boldsymbol{W}_p\left(\sum_{i=1}^{k} n_i, \boldsymbol{\Sigma}\right)$$

（3）若 $\boldsymbol{W} \sim \boldsymbol{W}_p(n, \boldsymbol{\Sigma})$，$\boldsymbol{C}_{q\times p}$为非奇异阵，则

$$\boldsymbol{CWC}' \sim \boldsymbol{W}_q(n, \boldsymbol{C\Sigma C}')$$

（4）若 $\boldsymbol{W} \sim \boldsymbol{W}_p(n, \boldsymbol{\Sigma})$，$\boldsymbol{a}$ 为任一 $p$ 元常向量，满足 $\boldsymbol{a}'\boldsymbol{\Sigma a} \neq 0$，则$\frac{\boldsymbol{a}'\boldsymbol{Wa}}{\boldsymbol{a}'\boldsymbol{\Sigma a}} \sim \chi^2(n)$。

（5）若 $\boldsymbol{W} \sim \boldsymbol{W}_p(n, \boldsymbol{\Sigma})$，$\boldsymbol{a}$ 为任一 $p$ 元非零常向量，比值

$$\frac{\boldsymbol{a}'\boldsymbol{\Sigma}^{-1}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{W}^{-1}\boldsymbol{a}} \sim \chi^2(n-p+1)$$

特别地，设 $w_{ii}$ 和 $\sigma_{ii}$ 分别为 $\boldsymbol{W}^{-1}$ 和 $\boldsymbol{\Sigma}^{-1}$ 的第 $i$ 个对角元，则

$$\frac{\sigma_{ii}}{w_{ii}} \sim \chi^2(n-p+1)$$

### 1.5.2 $t$ 分布与 $T^2$ 分布

在数理统计中，若 $X\sim N(0,\ 1)$，$Y\sim\chi^2(n)$，且 $X$ 与 $Y$ 相互独立，则称 $T=\dfrac{X}{\sqrt{\dfrac{Y}{n}}}$ 遵从自由度为 $n$ 的 $t$ 分布，又称为学生分布（student distribution），记为 $T\sim t(n)$。如果将 $T$ 平方，即 $T^2=n\dfrac{X^2}{Y}$，则 $T^2\sim F(1,\ n)$，即 $t(n)$ 分布变量的平方遵从第一自由度为 1、第二自由度为 $n$ 的中心 $F$ 分布。

将上述 $F$ 分布的定义改写成

$$F=nX'Y^{-1}X$$

式中，用 $X'$ 表示 $X$ 的转置。由于 $X$ 为一维数字，转置与否都相同，所以可以这样写。

在多元统计中，仿照上式推广得到 $T^2$ 分布的定义如下：

**定义 1.8** 设 $\boldsymbol{W}\sim \boldsymbol{W}_p(n,\ \boldsymbol{\Sigma})$，$\boldsymbol{X}\sim N_p(\boldsymbol{0},\ c\boldsymbol{\Sigma})$，$c>0$，$n\geqslant p$，$\boldsymbol{\Sigma}>\boldsymbol{0}$，$\boldsymbol{W}$ 与 $\boldsymbol{X}$ 相互独立，则称随机变量

$$T^2=\frac{n}{c}\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X} \tag{1.33}$$

所遵从的分布为第一自由度为 $p$、第二自由度为 $n$ 的中心 $T^2$ 分布，记为 $T^2\sim T^2(p,\ n)$。

$T^2$ 分布是霍特林（Hotelling）于 1931 年由一元统计推广而来的，故 $T^2$ 分布又称为 Hotelling $T^2$ 分布。其作用相当于数理统计学中的 $t$ 分布。

中心 $T^2$ 分布可化为中心 $F$ 分布，其关系可表示为：

$$\frac{n-p+1}{pn}T^2(p,n)=F(p,n-p+1)$$

显然，当 $p=1$ 时，有 $T^2(1,\ n)=F(1,\ n)$。

下面不加证明地给出 $T^2$ 分布的两条重要性质：

(1) 设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，$\boldsymbol{W}\sim W_p(n,\ \boldsymbol{\Sigma})$，且 $\boldsymbol{X}$ 与 $\boldsymbol{W}$ 相互独立，则

$$n(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{W}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\sim T^2(p,n)$$

**推论** 设 $\boldsymbol{X}_{(\alpha)}=(X_{\alpha 1},\ X_{\alpha 2},\ \cdots,\ X_{\alpha p})'$（$\alpha=1,\ 2,\ \cdots,\ n$）是从 $p$ 维正态总体 $N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$ 中抽取的 $n$ 个随机样本，$\bar{\boldsymbol{X}}$ 为样本均值，样本离差阵为 $\boldsymbol{L}=\sum\limits_{\alpha=1}^{n}(\boldsymbol{X}_{(\alpha)}-\bar{\boldsymbol{X}})(\boldsymbol{X}_{(\alpha)}-\bar{\boldsymbol{X}})'$，则

$$n(n-1)(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{L}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})\sim T^2(p,n-1)$$

或 $$n(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})\sim T^2(p,n-1)$$

其中，$\boldsymbol{S}=\dfrac{1}{n-1}\boldsymbol{L}$，为样本的协方差矩阵。

(2) 设 $\boldsymbol{X}_i\sim N_p(\boldsymbol{\mu}_i,\ \boldsymbol{\Sigma})$（$i=1,\ 2$），从总体 $X_1$，$X_2$ 中取得容量分别为 $n_1$，$n_2$ 的两个随机样本，若 $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2$，则

$$\frac{n_1 n_2}{n_1+n_2}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)'\boldsymbol{S}_p^{-1}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)\sim T^2(p,n_1+n_2-2)$$

或 $$(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)'\boldsymbol{S}_p^{-1}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)\sim\frac{n_1+n_2}{n_1 n_2}T^2(p,n_1+n_2-2)$$

其中，$\bar{\boldsymbol{X}}_1$，$\bar{\boldsymbol{X}}_2$ 为两样本的均值向量；$\boldsymbol{S}_p=\frac{(n_1-1)\boldsymbol{S}_1+(n_2-1)\boldsymbol{S}_2}{n_1+n_2-2}$；$\boldsymbol{S}_1$，$\boldsymbol{S}_2$ 分别为两样本的协方差阵。

这个性质在下章的假设检验中有重要应用。

### 1.5.3 中心 *F* 分布与 Wilks 分布

在一元统计中，若 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且 $X$ 与 $Y$ 相互独立，则称 $F=\frac{X/m}{Y/n}$所遵从的分布为第一自由度为 $m$、第二自由度为 $n$ 的中心 $F$ 分布，记为 $F\sim F(m,\ n)$。$F$ 分布本质上是从正态总体 $N(\mu,\ \sigma^2)$ 中随机抽取的两个样本方差的比。

$F$ 分布能否推广到多元呢？由于 $F$ 分布由两个方差比构成，而多元总体$N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$的变异由协方差阵确定，它不是一个数字，这就产生了如何用一个与协方差阵 $\boldsymbol{\Sigma}$ 有关的量来描述总体 $N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$的变异的问题，它是将 $F$ 分布推广到多元情形的关键。

描述 $N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$的变异度的统计参数称为广义方差。围绕这一问题产生了许多方法，有的用行列式，有的用迹，主要的方法有以下几种：

（1）广义方差$\triangleq|\boldsymbol{\Sigma}|$；

（2）广义方差$\triangleq \mathrm{tr}(\boldsymbol{\Sigma})=\sum_{i=1}^{p}\sigma_i^2=\sigma_1^2+\sigma_2^2+\cdots+\sigma_p^2$，其中 $\mathrm{tr}(\boldsymbol{\Sigma})$ 为 $\boldsymbol{\Sigma}$ 的迹，等于 $\boldsymbol{\Sigma}$ 主对角线元素之和；

（3）广义方差$\triangleq \mathrm{tr}(\boldsymbol{\Sigma})=\prod_{i=1}^{p}\sigma_i^2$；

（4）广义方差$\triangleq |\boldsymbol{\Sigma}|^{\frac{1}{p}}$；

（5）广义方差$\triangleq (\mathrm{tr}(\boldsymbol{\Sigma}))^{\frac{1}{2}}=\sqrt{\sigma_1^2+\sigma_2^2+\cdots+\sigma_p^2}$；

（6）广义方差$\triangleq\max\{\lambda_i\}$，其中 $\lambda_i$ 为 $\boldsymbol{\Sigma}$ 的特征根；

（7）广义方差$\triangleq\min\{\lambda_i\}$，其中 $\lambda_i$ 为 $\boldsymbol{\Sigma}$ 的特征根。

在以上各种广义方差的定义中，目前使用最多的是第一种，它是 T. W. 安德森（T. W. Anderson）于 1958 年提出的。

下面根据第一种广义方差，仿照 $F$ 分布的定义给出多元统计中两个广义方差之比的统计量，称为 Wilks $\Lambda$ 分布。

**定义 1.9** 设$\boldsymbol{W}_1\sim\boldsymbol{W}_p(n_1,\ \boldsymbol{\Sigma})$，$\boldsymbol{W}_2\sim\boldsymbol{W}_p(n_2,\ \boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>\boldsymbol{0}$，$n_1>p$，且$\boldsymbol{W}_1$ 与$\boldsymbol{W}_2$ 相互独立，则

$$\Lambda=\frac{|\boldsymbol{W}_1|}{|\boldsymbol{W}_1+\boldsymbol{W}_2|}\tag{1.34}$$

所遵从的分布称为维数为 $p$，第一自由度为 $n_1$，第二自由度为 $n_2$ 的 Wilks $\Lambda$ 分布，记为

$\Lambda \sim \Lambda(p, n_1, n_2)$。

由上述定义，$\Lambda$ 分布为两个广义方差之比。

由于 $\Lambda$ 分布在多元统计中的重要性，关于它的近似分布和精确分布不断有学者进行研究。当 $p$ 和 $n_2$ 中的一个比较小时，$\Lambda$ 分布可化为 $F$ 分布，表 1-2 列举了常见的情况。

表 1-2　　$\Lambda \sim \Lambda(p, n_1, n_2)$ 与 $F$ 分布的关系，$n_1 > p$

| $p$ | $n_2$ | 统计量 $F$ | $F$ 的自由度 |
|---|---|---|---|
| 任意 | 1 | $\frac{1-\Lambda}{\Lambda}\frac{n_1-p+1}{p}$ | $p$，$n_1-p+1$ |
| 任意 | 2 | $\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{n_1-p}{p}$ | $2p$，$2(n_1-p)$ |
| 1 | 任意 | $\frac{1-\Lambda}{\Lambda}\frac{n_1}{n_2}$ | $n_2$，$n_1$ |
| 2 | 任意 | $\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{n_1-1}{n_2}$ | $2n_2$，$2(n_1-1)$ |

当 $p$，$n_2$ 不属于表 1-2 所列举的情况时，巴特莱特（Bartlett）指出可用 $\chi^2$ 分布来近似表示，即

$$V=-\left(n_1+n_2-\frac{p+n_2+1}{2}\right)\ln\Lambda(p,n_1,n_2)$$

近似遵从 $\chi^2(pn_2)$。

拉奥（Rao）后来又研究用 $F$ 分布来近似，即

$$R=\frac{1-\Lambda^{\frac{1}{s}}}{\Lambda^{\frac{1}{s}}}\frac{ts-2\lambda}{pn_2}$$

近似遵从 $F(pn_2, ts-2\lambda)$，其中

$$\begin{cases} t=n_1+n_2-\dfrac{p+n_2+1}{2} \\ s=\sqrt{\dfrac{p^2 n_2^2-4}{p^2+n_2^2-5}} \\ \lambda=\dfrac{pn_2-2}{4} \end{cases}$$

$ts-2\lambda$ 不一定是整数，用与它最近的整数来作为 $F$ 分布的第二自由度。

若 $n_2 < p$，有 $\Lambda(p, n_1, n_2)=\Lambda(n_2, p, n_1+n_2-p)$。该结论说明，在使用 $\Lambda$ 统计量时也可考虑 $n_2 > p$ 的情形，有关 $\Lambda$ 统计量的其他性质参见参考文献［1］。

## 参考文献

［1］张尧庭，方开泰．多元统计分析引论．北京：科学出版社，1982.

[2] Robb J. Muirhead. Aspects of Multivariate Statistical Theory. John Wiley, 1982.

[3] 方开泰. 实用多元统计分析. 上海：华东师范大学出版社，1989.

[4] 王学仁. 地质数据的多变量统计分析. 北京：科学出版社，1986.

[5] 王国梁，何晓群. 多变量经济数据统计分析. 西安：陕西科学出版社，1993.

[6] G. A. F. Seber. Multivariate Observations. John Wiley & Sons, Inc., 1984.

[7] 王静龙. 多元统计分析. 北京：科学出版社，2008.

## 思考与练习

1. 在数据处理时，为什么通常要进行标准化处理？

2. 欧氏距离与马氏距离的优缺点是什么？

3. 当变量 $X_1$ 和 $X_2$ 方向上的变差相等，且 $X_1$ 与 $X_2$ 互相独立时，采用欧氏距离与统计距离是否一致？

4. 如果正态随机向量 $\boldsymbol{X}=(x_1, x_2, \cdots, x_p)'$ 的协方差阵 $\boldsymbol{\Sigma}$ 是对角阵，证明 $\boldsymbol{X}$ 的分量是相互独立的随机变量。

5. $y_1$ 与 $y_2$ 是相互独立的随机变量，且 $y_1 \sim N(0, 1)$，$y_2 \sim N(3, 4)$。

(a) 求 $y_1^2$ 的分布。

(b) 如果 $\boldsymbol{y}=\begin{bmatrix} y_1 \\ (y_2-3)/2 \end{bmatrix}$，写出 $\boldsymbol{y}'\boldsymbol{y}$ 关于 $y_1$ 与 $y_2$ 的表达式，并写出 $\boldsymbol{y}'\boldsymbol{y}$ 的分布。

(c) 如果 $\boldsymbol{y}=\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ 且 $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，写出 $\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ 关于 $y_1$ 与 $y_2$ 的表达式，并写出 $\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ 的分布。