第四章　多元统计基础

§4.1　多元分布概念

§4.1.1 随机向量

讨论多变量总体，数据是同时观测p个变量，进行n次观测得到的，把这p个指标表示为$X_1, X_2, \cdots, X_p$，用向量$X = (X_1, X_2, \cdots, X_p)^T$表示对同一个体观测的$p$个变量。

观测n个个体，数据如下表所示，称每一个个体的p个变量为一个样品，而全体n个样品形成一个样本。

变量	X_1	X_2	…	X_p
1	x_{11}	x_{12}	…	x_{1p}
2	x_{21}	x_{21}	…	x_{2p}
…	…	…		…
n	x_{n1}	x_{n2}	…	x_{np}

记$X_{(a)}=(x_{a1},x_{a2},\cdots,x_{ap})^T$， $a=1,2,\ldots,n$

表示第a个样品的观测值。

第j列的元素$X_j=(x_{1j},x_{2j},\cdots,x_{nj})^T$， $j=1,2,\ldots,p$

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

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}=(X_1,X_2,\cdots,X_p)=\begin{bmatrix} X_{(1)}^T \\ X_{(2)}^T \\ \vdots \\ X_{(n)}^T \end{bmatrix}$$

定义4.1.1 设$X_1,X_2,\cdots,X_p$为p个随机变量，由它们组成的向量$X=(X_1,X_2,\cdots,X_p)^T$称为随机向量。

§4.1.2　分布函数与密度函数

定义4.1.2 设随机向量$X=(X_1,X_2,\cdots,X_p)^T$，多元分布函数为

$$F(x)=F(x_1,x_2,\cdots,x_p)=P(X_1\le x_1,X_2\le x_2,\ldots,X_p\le x_p)$$

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

定义4.1.3 $X \sim F(x) = F(x_1, x_2, \cdots, x_p)$，若存在一个非负的函数$f(\cdot)$，使得

$$F(x) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p} f(t_1, t_2, \cdots, t_p)\, dt_1\, dt_2 \cdots dt_p$$

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

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$f(x) \geq 0, \forall x \in R^p$，且$\int_{R^p} f(x) dx = 1$。

§4.1.3　多元变量的独立性

定义4.1.4 两个随机向量X和Y称为互相独立的，若$P(X \le x,\ Y \le y) = P(X \le x)P(Y \le y)$对一切$x, y$均成立。

若$F(x, y)$为(X, Y)的联合分布函数，$G(x)$和$H(y)$分别为X和Y的分布函数，则X和Y独立当且仅当

$$F(x, y) = G(x)H(y)$$

若$(\boldsymbol{X},\boldsymbol{Y})$有密度$\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y})$，用$\boldsymbol{g}(\boldsymbol{x})$和$\boldsymbol{h}(\boldsymbol{y})$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

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

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

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

§4.1.4　随机向量的数字特征

随机向量X的均值

设$X=\left(X_1,X_2,\cdots,X_p\right)^T$有$p$个分量。

若$E(X_i)=\mu_i(i=1,2,\cdots,p)$存在，定义随机向量$X$的均值为

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

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

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

（1） $\boldsymbol{E(AX) = AE(X)}$

（2） $\boldsymbol{E(AXB) = AE(X)B}$

随机向量X的协方差阵

$$\Sigma = cov(X,X) = E(X-EX)(X-EX)^T = D(X)$$

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

$$= (\sigma_{ij})$$

称它为p维随机向量X的协方差阵。称$|cov(X,X)|$为X的广义方差，是协方差阵的行列式的值。

随机向量X和Y的协方差阵

设$X=(X_1,X_2,\cdots,X_p)^T$和$Y=(Y_1,Y_2,\cdots,Y_q)^T$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$cov(X_i,Y_j)$，即

$$cov(X,Y)=\left(cov\left(X_i,Y_j\right)\right),i=1,2,\cdots,p;j=1,2,\cdots,q$$

若$cov(X,Y)=0$，称X和Y是不相关的。

当$\boldsymbol{A}$和$\boldsymbol{B}$为常数矩阵时，由定义可推出如下性质

（1） $\boldsymbol{D(AX) = AD(X)A^T = A\Sigma A^T}$

（2） $\boldsymbol{cov(AX, BY) = Acov(X, Y)B^T}$

（3）设$\boldsymbol{X}$为$\boldsymbol{p}$维随机向量，期望和协方差存在，记

$\boldsymbol{\mu = E(X), \Sigma = D(X)}$， $\boldsymbol{A}$为$\boldsymbol{p \times p}$常数阵，则

$$\boldsymbol{E(X^T AX) = tr(A\Sigma) + \mu^T A\mu}$$

随机向量$\boldsymbol{X = \left(X_1, X_2, \cdots, X_p\right)^T}$，其协方差阵$\boldsymbol{\Sigma}$都是对称阵，同时总是半正定的。大多数情形下是正定的。

随机向量X的相关阵

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

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

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

r_{ij}也称分量X_i和X_j之间的（线性）相关系数。

在数据处理时，为克服指标量纲不同的影响，使用统计分析方法前，将每个指标“标准化”：

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

$$X^* = (X_1^*, X_2^*, \cdots, X_p^*)$$

于是 $E(X^*) = 0,\ D(X^*) = corr(X) = R$

即标准化数据的协方差矩阵正好是原指标的相关阵：

$$R = \frac{1}{n-1} {X^*}^T X^*$$

§4.2　多元正态分布

§4.2.1 多元正态分布的定义

回顾二元正态分布

$\boldsymbol{x}=(x_1,x_2)^T \sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$，其密度函数为

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

其中$\sigma_1^2>0,\sigma_2^2>0,|\rho|<1$。特征函数为

$$\varphi(t_1,t_2)=E\left[e^{i(t_1x_1+t_2x_2)}\right]$$

$$=exp\left\{i(t_1\mu_1+t_2\mu_2)-\frac{1}{2}\left(\sigma_1^2t_1^2+2\rho\sigma_1\sigma_2t_1t_2+\sigma_2^2t_2^2\right)\right\}$$

若令$\mu=(\mu_1,\mu_2)^T,t=(t_1,t_2)^T$，以及

$$V=Var(x)=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$$

则分别有

$$p(x) = (2\pi)^{-\frac{2}{2}}|V|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(x-\mu)^T V^{-1}(x-\mu)\right\}$$

$$\varphi(t) = E\left(e^{it^T x}\right) = exp\left\{it^T\mu - \frac{1}{2}t^T V t\right\}$$

上述两个表达式主要由二元正态向量x的数学期望$E(x)=\mu$和协方差阵$Var(x)=V=(\sigma_{ij})$决定，易推广到多维随机向量的场合,可引出下述多元正态分布的密度函数定义和特征函数定义。

定义4.2.1（密度函数定义） 若p维随机向量（随机变量）$X=\left(X_1,X_2,\cdots,X_p\right)^T$（联合）概率密度函数为

$$f\left(x_1,x_2,\cdots,x_p\right)=\frac{1}{(2\pi)^{\frac{p}{2}}|V|^{\frac{1}{2}}}exp\left\{-\frac{1}{2}(X-\mu)^TV^{-1}(X-\mu)\right\}$$

则称随机向量X为p维正态随机向量，其中μ称为均值向量，V为协方差矩阵(协差阵)，且$V>0$。

对于一般情形$\boldsymbol{V} \geq \mathbf{0}$，仍可定义多维正态随机向量，记为$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{V})$。当$\boldsymbol{V} > \mathbf{0}$时，$\boldsymbol{X}$由前面密度表示，而当$|\boldsymbol{V}| = 0$时，$\boldsymbol{X}$的分布是退化的正态分布。

定义4.2.2（特征函数定义） 若p维随机向量（随机变量）$X=(X_1,X_2,\cdots,X_p)^T$特征函数为

$$\varphi(t)=E\left(e^{it^Tx}\right)=exp\left\{it^T\mu-\frac{1}{2}t^TVt\right\}$$

则称随机向量X为p维正态随机向量，其中μ称为均值向量，V为协方差矩阵(协差阵)，且$V\geq 0$。

定义4.2.3 若对任何非零向量$\boldsymbol{a} \in \boldsymbol{R}^p$， $\boldsymbol{X}$的线性组合$\boldsymbol{a}^T\boldsymbol{X}$服从一元正态分布$\boldsymbol{N}(\boldsymbol{a}^T\boldsymbol{\mu}, \boldsymbol{a}^T\boldsymbol{V}\boldsymbol{a})$，则称$\boldsymbol{X}$服从$\boldsymbol{p}$元正态分布$\boldsymbol{N}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{V})$。

利用特征函数的性质可以证明定义4.2.2与定义4.2.3是等价的。

§4.2.2 多元正态分布的基本性质

（1）$\boldsymbol{p}$维正态分布由其均值向量和协方差阵唯一确定。

（2）对于任一$\boldsymbol{p}$维向量$\boldsymbol{\mu}$及$\boldsymbol{p}$阶非负定矩阵$\boldsymbol{V}$，必存在$\boldsymbol{p}$维正态随机向量$\boldsymbol{X} \sim \boldsymbol{N_p}(\boldsymbol{\mu}, \boldsymbol{V})$。

（3）设$\boldsymbol{X} \sim \boldsymbol{N_p}(\boldsymbol{\mu}, \boldsymbol{V})$，$\boldsymbol{A}$是$\boldsymbol{m} \times \boldsymbol{p}$常数矩阵，$\boldsymbol{b}$是$\boldsymbol{m}$维向量，若令$\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$，则$\boldsymbol{Y} \sim \boldsymbol{N_p}(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{AVA^T})$。

证明 首先计算的Y特征函数

$$\varphi_Y(t)=E\left(e^{it^TY}\right)=E\left[e^{it^T(AX+b)}\right]=e^{it^Tb}E\left(e^{it^TAX}\right)$$

$$=e^{it^Tb}\varphi_X(A^Tt)=e^{it^Tb}exp\left\{it^TA\mu-\frac{1}{2}t^TAVA^Tt\right\}$$

$$=exp\left\{it^T(A\mu+b)-\frac{1}{2}t^T(AVA^T)\,t\right\}$$

因此由多元正态分布的特征函数定义4.2.2，有

$Y{\sim}N_p(A\mu+b,AVA^T)$

（4）X为p维正态随机向量的充要条件为对任一p维向量c，$c^T X$是一维正态随机变量。

（5）设$X=\left(X_1^T, X_2^T\right)^T$为多维正态随机向量，则$X_1$与$X_2$互不相关的充要条件是$X_1$与$X_2$相互独立。

注：若$Cov(X, Y)=0$，则称X与Y互不相关。

（6）设$X \sim N_p(\mu, V)$，则$rank(V)=m$的充要条件是存在$m \times p$的矩阵$B\left(BB^T=V\right)$使得$X=BY+\mu$，其中$Y \sim N_m(0, I_m)$。

证明 充分性由性质3立得。下证必要性。

由于V是秩为m的非负定阵，则必存在正交矩阵U使得

$$
U^TVU=\begin{pmatrix}
\lambda_1 & \cdots & 0 & 0 & \cdots & 0\\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\
0 & \cdots & \lambda_m & 0 & \cdots & 0\\
0 & \cdots & 0 & 0 & \cdots & 0\\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\
0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix}
$$

其中$\lambda_i>0, i=1,\cdots,m$。

令$\boldsymbol{\Lambda}=\begin{pmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_m\end{pmatrix}$，则有

$$\begin{pmatrix}\boldsymbol{\Lambda}^{-\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{p-m}\end{pmatrix}\boldsymbol{U}^T\boldsymbol{V}\boldsymbol{U}\begin{pmatrix}\boldsymbol{\Lambda}^{-\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{p-m}\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_m & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\triangleq\boldsymbol{D}$$

令$\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{W}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\Lambda}^{-\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{p-m}\end{pmatrix}\boldsymbol{U}^T(\boldsymbol{X}-\boldsymbol{\mu})$

则由性质3知$\boldsymbol{Z}\sim N_p(\boldsymbol{0},\boldsymbol{D})$，且$\boldsymbol{Y}\sim N_m(\boldsymbol{0},\boldsymbol{I}_m)$，由上式

可得$\boldsymbol{X}-\boldsymbol{\mu}=\boldsymbol{U}\begin{pmatrix}\boldsymbol{\Lambda}^{\frac{1}{2}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{p-m}\end{pmatrix}\boldsymbol{Z}=\boldsymbol{U}\begin{pmatrix}\boldsymbol{\Lambda}^{\frac{1}{2}}\\ \boldsymbol{0}\end{pmatrix}\boldsymbol{Y}$，若记

$\boldsymbol{B}=\boldsymbol{U}\begin{pmatrix}\boldsymbol{\Lambda}^{\frac{1}{2}}\\ \boldsymbol{0}\end{pmatrix}$，它是$p\times m$矩阵，即有$\boldsymbol{X}=\boldsymbol{B}\boldsymbol{Y}+\boldsymbol{\mu}$

（7）若$X \sim N_p(\mu, V)$，且$|V| \neq 0$，则

$\eta \triangleq (X-\mu)^T V^{-1}(X-\mu) \sim \chi^2(p)$.

证明　由$|V| \neq 0$可知V是正定矩阵，所以$V^{-\frac{1}{2}}$存在且为对称矩阵，这样

$$\eta = (X-\mu)^T V^{-\frac{1}{2}} V^{-\frac{1}{2}} (X-\mu) = \left(V^{-\frac{1}{2}}(X-\mu)\right)^T \left(V^{-\frac{1}{2}}(X-\mu)\right)$$

令$Y = V^{-\frac{1}{2}}(X-\mu)$，则$\eta = Y^T Y$，且$Y \sim N_p(\mathbf{0}, I_p)$.

由性质3知$\boldsymbol{Y}$的每个分量$\boldsymbol{Y_i}$服从标准正态分布，且相互独立，故由χ^2 −分布的定义知$\boldsymbol{Y^T Y \sim \chi^2(p)}$.

例4.2.1 设$X \sim N_p(\mu, V), V > 0$，将$X$分块为$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \begin{matrix} r \\ p-r \end{matrix}$，

相应$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$，

证明$X_1 - V_{12}{V_{22}}^{-1}X_2$与$X_2$相互独立，$X_2 - V_{21}{V_{11}}^{-1}X_1$与$X_1$也相互独立。

证明 令$A = \left(I_r, -V_{12}{V_{22}}^{-1}\right), B = \left(0, I_{p-r}\right)$，则

$$rank(A) = r,\ rank(B) = p - r$$

$$X_1 - V_{12}V_{22}{}^{-1}X_2 = AX, X_2 = BX$$

且由性质（3）知，$\begin{pmatrix} A \\ B \end{pmatrix} X$服从$p$维正态分布。由于

$$Cov(AX, BX) = ACov(X, X)B^T = (I_r \quad -V_{12}V_{22}{}^{-1})V\begin{pmatrix} 0 \\ I_{p-r} \end{pmatrix}$$

所以由性质（5）知，AX与BX相互独立，即$X_1 - V_{12}V_{22}{}^{-1}X_2$与$X_2$相互独立。同理可证$X_2 - V_{21}V_{11}{}^{-1}X_1$与$X_1$也相互独立。

§4.2.3　多元正态条件分布

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{V}), \boldsymbol{V} > \mathbf{0}$，将$\boldsymbol{X}$分块为$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \begin{matrix} k \\ p-k \end{matrix}$，同时对$\boldsymbol{\mu}$和$\boldsymbol{V}$相应分块

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix}$$

本小节主要介绍在给定$\boldsymbol{X}_2$时$\boldsymbol{X}_1$的条件分布。

定理4.2.1 设$X \sim N_p(\mu, V), V > 0$，则给定$X_2$时$X_1$的条件分布为$N_k(\mu_{1\cdot2}, V_{11\cdot2})$，其中

$$\mu_{1\cdot2} = \mu_1 + V_{12}{V_{22}}^{-1}(X_2 - \mu_2)$$

$$V_{11\cdot2} = V_{11} - V_{12}{V_{22}}^{-1}V_{21}$$

同样，给定X_1时X_2的条件分布为$N_{p-k}(\mu_{2\cdot1}, V_{22\cdot1})$，其中

$$\mu_{2\cdot1} = \mu_2 + V_{21}{V_{11}}^{-1}(X_1 - \mu_1)$$

$$V_{22\cdot1} = V_{22} - V_{21}{V_{11}}^{-1}V_{12}$$

证明 由$V>0$，可知$V_{11}>0, V_{22}>0, V_{11\cdot 2}>0$。令

$$A=\begin{pmatrix} I_k & -V_{12}{V_{22}}^{-1} \\ 0 & I_{p-k} \end{pmatrix},$$

$$Y=AX=\begin{pmatrix} X_1-V_{12}{V_{22}}^{-1}X_2 \\ X_2 \end{pmatrix} \triangleq \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

则由4.2.2性质（3）知，$Y_1 \sim N_k\left(\mu_1-V_{12}{V_{22}}^{-1}\mu_2, V_{11\cdot 2}\right)$

由例4.2.1可知Y_1与$Y_2=X_2$相互独立，因此

$$f_Y(Y)=f_{Y_1}(Y_1)\,f_{Y_2}(Y_2)$$

注意到变换雅可比行列式$J(Y,X)=|A|=1$，则X的联合密度可以表示为

$$f_X(X)=f_Y(Y)|J(Y,X)|=f_Y(Y)$$

$$=f_{Y_1}(Y_1)\,f_{Y_2}(Y_2)=f_{Y_1}(Y_1)\,f_{X_2}(X_2)$$

故给定X_2时X_1的条件密度函数为

$$f(X_1|X_2)=\frac{f_X(X)}{f_{X_2}(X_2)}=f_{Y_1}(Y_1)$$

$$= (2\pi)^{-\frac{k}{2}}|V_{11\cdot2}|^{-\frac{1}{2}}exp\left\{-\frac{1}{2}\left(Y_1 - \mu_1 + V_{12}V_{22}{}^{-1}\mu_2\right)^T V_{11\cdot2}{}^{-1}(Y_1 - \mu_1 + V_{12}V_{22}{}^{-1}\mu_2)\right\}$$

将$Y_1 = X_1 - V_{12}V_{22}{}^{-1}X_2$代入，可得

$$f(X_1|X_2) = (2\pi)^{-\frac{k}{2}}|V_{11\cdot2}|^{-\frac{1}{2}}exp\left\{-\frac{1}{2}(X_1 - \mu_{1\cdot2})^T V_{11\cdot2}{}^{-1}(X_1 - \mu_{1\cdot2})\right\}$$

即给定X_2时，$X_1 \sim N_k(\mu_{1\cdot2}, V_{11\cdot2})$。

例4.2.2 设$X \sim N_3(\mu, V)$,其中$\mu = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$, $V = \begin{pmatrix} 16 & -4 & 2 \\ -4 & 4 & -1 \\ 2 & -1 & 4 \end{pmatrix}$

在给定$X_1 + 2X_3$时，求$\begin{pmatrix} X_2 - X_3 \\ X_1 \end{pmatrix}$的条件分布。

解 令$Y_1 = \begin{pmatrix} X_2 - X_3 \\ X_1 \end{pmatrix}$, $Y_2 = X_1 + 2X_3$，则

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

由**4.2.2**性质（**3**）， $Y \sim N_3(E(Y), Var(Y))$，其中

$$E(Y) = E\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$$

$$Var(Y) = Var\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}\begin{pmatrix} 16 & -4 & 2 \\ -4 & 4 & -1 \\ 2 & -1 & 4 \end{pmatrix}\begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}^T$$

$$Var(Y)=\begin{pmatrix}10 & -6 & -16\\ -6 & 16 & 20\\ -16 & 20 & 40\end{pmatrix}$$

所以，由定理4.2.1，在给定X_1+2X_3时，

$$\begin{pmatrix}X_2-X_3\\ X_1\end{pmatrix}\sim N_3\left(E(Y_1|Y_2),Var(Y_1|Y_2)\right)$$，其中

$$E(Y_1|Y_2)=\begin{pmatrix}2\\1\end{pmatrix}+\begin{pmatrix}-16\\20\end{pmatrix}\times\frac{1}{40}\times(Y_2+3)=\begin{pmatrix}-\frac{2}{5}X_1-\frac{4}{5}X_3+\frac{4}{5}\\ \frac{1}{2}X_1+X_3+\frac{5}{2}\end{pmatrix}$$

$$Var(Y_1|Y_2)=\begin{pmatrix}10 & -6\\ -6 & 16\end{pmatrix}-\begin{pmatrix}-16\\20\end{pmatrix}\times\frac{1}{40}\times(-16\quad 20)=\begin{pmatrix}3.6 & 2\\ 2 & 6\end{pmatrix}$$

当X_1与X_2存在相关关系时，条件数学期望

$E(X_1|X_2)=\mu_{1\cdot 2}=\mu_1+V_{12}{V_{22}}^{-1}(X_2-\mu_2)$可以被视为$X_1$对$X_2$的多元回归函数，$V_{12}{V_{22}}^{-1}$可以称为$X_1$对$X_2$的回归系数矩阵。

条件协方差矩阵$V_{11\cdot 2}=V_{11}-V_{12}{V_{22}}^{-1}V_{21}$又称为协方差矩阵。它是$X_1$对$X_2$回归后的剩余方差，是从$X_1$的方差中扣除能被线性表出的部分（即

$\mu_1+V_{12}{V_{22}}^{-1}(X_2-\mu_2)$）的方差之后的剩余部分。

例4.2.3 设$X=(X_1,X_2,X_3)^T\sim N_3(\mu,V)$,其中$\mu=\begin{pmatrix}10\\4\\7\end{pmatrix}$,$V=\begin{pmatrix}9&-3&-3\\-3&5&1\\-3&1&5\end{pmatrix}$。求$X_1$对$(X_1,X_2)^T$的回归函数。

解 X_1对$(X_1,X_2)^T$的回归函数即$E(X_1|(X_1,X_2)^T)$。根据条件期望计算公式可得

$$E(X_1|(X_1,X_2\ )^T)=10+(-3\quad -3)\begin{pmatrix}5 & 1\\ 1 & 5\end{pmatrix}^{-1}\begin{pmatrix}X_2-4\\ X_3-7\end{pmatrix}$$

$$=15.5-0.5X_2-0.5X_3$$

因此，X_1对$(X_1,X_2\ )^T$的回归函数为

$$X_1=15.5-0.5X_2-0.5X_3$$

相关性是指两个随机变量之间存在线性关联，可以用简单相关系数$\rho_{12}=\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$来进行度量。

但两个统计相关的随机变量不一定有因果关系，X_1与X_2关系有可能会受到第三个因素的影响X_3。

把给定X_3的条件下X_1与X_2的相关系数称为的偏相关系数。偏相关系数也可以解释成在扣除了X_3的影响之后X_1与X_2的相关系数。

定义4.2.4 设$X=\begin{pmatrix}X_1\\X_2\end{pmatrix}\begin{matrix}k\\p-k\end{matrix}\sim N_p(\mu,V),V>0$。在给定$X_2$条件下，$X_1$的偏协方差矩阵为

$$Var(X_1|X_2)=V_{11\cdot 2}=\left(\sigma_{ij\cdot k+1\cdots p}\right)_{k\times k}$$

在给定X_2条件下， X_1与X_2的偏相关系数定义为

$$\rho_{ij\cdot k+1\cdots p}=\frac{\sigma_{ij\cdot k+1\cdots p}}{\sqrt{\sigma_{ii\cdot k+1\cdots p}\sigma_{jj\cdot k+1\cdots p}}}$$

由偏相关系数组成的矩阵

$$R(X_1|X_2) = R_{11\cdot 2} = \left(\rho_{ij\cdot k+1\cdots p}\right)_{k\times k}$$

称为给定X_2条件下X_1的偏相关矩阵。

设总体$X=\begin{pmatrix}X_1\\X_2\end{pmatrix}\begin{matrix}k\\p-k\end{matrix}\sim N_p(\mu,V),V>0$。在给定$X_2$条件下，$X_1$的偏协方差矩阵为

$$Var(X_1|X_2)=V_{11\cdot 2}=\left(\sigma_{ij\cdot k+1\cdots p}\right)_{k\times k}$$

在给定X_2条件下，X_1与X_2的偏相关系数定义为

$$\rho_{ij\cdot k+1\cdots p}=\frac{\sigma_{ij\cdot k+1\cdots p}}{\sqrt{\sigma_{ii\cdot k+1\cdots p}\sigma_{jj\cdot k+1\cdots p}}}$$

§4.3　均值向量与协方差阵的估计

在实际问题中，多元正态分布的均值$\boldsymbol{\mu}$和协方差阵$\boldsymbol{V}$往往未知，需要利用样本对它们进行估计和检验，这些内容可以看成是一元正态分布参数$\boldsymbol{\mu}$和$\boldsymbol{\sigma}^2$的估计和检验问题在多元情况下的推广。

下面介绍多元正态总体参数$\boldsymbol{\mu}$和$\boldsymbol{V}$的估计。

设$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$是来自多元正态总体$N_p(\boldsymbol{\mu},\boldsymbol{V})$的简单样本，样本容量$n>p$。

为简单起见，在此仅考虑$\boldsymbol{V}>0$，即非退化情形。令

$$\boldsymbol{X}_{n\times p}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)^T=\begin{pmatrix}\boldsymbol{x}_1{}^T\\ \vdots\\ \boldsymbol{x}_n{}^T\end{pmatrix}$$

则样本的联合分布密度函数为

$$p(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{V})=\prod_{i=1}^{n}p(\boldsymbol{x}_i;\boldsymbol{\mu},\boldsymbol{V})$$

化简后

$$p(X;\mu,V)$$

$$=(2\pi)^{-\frac{np}{2}}|V|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}tr\{V^{-1}[S+n(\bar{x}-\mu)(\bar{x}-\mu)^T]\}\right\}$$

其中，

$\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ ——样本均值向量

$S=\sum_{k=1}^{n}(X_k-\bar{X})(X_k-\bar{X})^T$——样本离差阵

为求出μ和V的极大似然估计，先介绍下列引理。

引理4.3.1 设A是$m\times m$阶正定矩阵，则当$A=nI_m$的时候，函数$f(A)=|A|^{\frac{n}{2}}exp\left\{-\frac{1}{2}tr(A)\right\}$，达最大值$n^{\frac{mn}{2}}e^{-\frac{mn}{2}}$。

定理4.3.1 设$X_1, X_2, \cdots, X_n (n > p)$是多元正态总体$N_p(\mu, V)$的简单样本，且$V > 0$，则$\bar{X}$是$\mu$的极大似然估计，$\frac{1}{n}S$是$V$的极大似然估计。

多元正态总体$N_p(\mu, V)$参数μ和V的估计量有如下性质

性质4.3.1 设$X_1, X_2, \cdots, X_n (n > p)$是多元正态总体$N_p(\mu, V)$的简单样本，且$V > 0$，则$\bar{X}$是$\mu$的一致最小方差无偏估计，$\frac{1}{n-1}S$是$V$的一致最小方差无偏估计。

性质4.3.2 $\bar{X} \sim N_p\left(\mu, \frac{1}{n}V\right)$，且$\bar{X}$与$S$相互独立。

性质4.3.3 样本离差阵S服从自由度为$n-1$的Wishart分布，即$S \sim W_p(n-1, V)$。

定义4.3.1 设$X=(X_1,X_2,\cdots,X_n)^T,X_i\sim N_p(\mu_i,V)(i=1,2,\cdots,n)$，且$X_1,X_2,\cdots,X_n$相互独立，记$M=E(X)=(\mu_1,\mu_2,\cdots,\mu_n)^T$，称$A=X^TX=\sum_{i=1}^{n}X_iX_i^T$所服从的分布是自由度为$n$的非中心Wishart分布，记为$A\sim W_p(n,V,\tau)$，其中$\tau=M^TM$为非中心参数。当$\mu_1=\mu_2=\cdots=\mu_n=0$时有$\tau=0$，此时称$A$服从中心Wishart分布，记为$W_p(n,V)$。

§4.4　均值向量的检验

§ 4.4.1 协差阵$\boldsymbol{V}$ 已知时，均值$\boldsymbol{\mu}$ 检验

设$\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n (n > p)$是多元正态总体$N_p(\boldsymbol{\mu}, \boldsymbol{V})$的简单样本，其中$\boldsymbol{V}$已知。考虑假设检验问题

$$H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0, H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

首先导出检验似然比。当$\boldsymbol{V}$已知时，参数空间$\boldsymbol{\Theta}$为p维欧式空间$\boldsymbol{R}^p$，原假设H_0的参数空间为$\boldsymbol{\Theta}_0 = \{\boldsymbol{\mu}_0\}$。

由于样本的联合密度函数为

$$p(\boldsymbol{X};\boldsymbol{\mu}) = (2\pi)^{-\frac{np}{2}}|\boldsymbol{V}|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}tr\{\boldsymbol{V}^{-1}[\boldsymbol{S}+n(\bar{\boldsymbol{X}}-\boldsymbol{\mu})(\bar{\boldsymbol{X}}-\boldsymbol{\mu})^T]\}\right\}$$

从而

$$\sup_{\boldsymbol{\mu}\in\boldsymbol{\Theta}}\{p(\boldsymbol{X};\boldsymbol{\mu})\} = p(\boldsymbol{X};\bar{\boldsymbol{X}})$$

$$= (2\pi)^{-\frac{np}{2}}|\boldsymbol{V}|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}tr(\boldsymbol{V}^{-1}\boldsymbol{S})\right\}$$

$$\sup_{\boldsymbol{\mu}\in\boldsymbol{\Theta}_0}\{\boldsymbol{p}(\boldsymbol{X};\boldsymbol{\mu})\}=\boldsymbol{p}(\boldsymbol{X};\boldsymbol{\mu}_0)$$

$$=(2\pi)^{-\frac{np}{2}}|V|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}tr\{V^{-1}[S+n(\bar{X}-\mu_0)(\bar{X}-\mu_0)^T]\}\right\}$$

因此检验假设检验问题的似然比为

$$\lambda(X)=\frac{\sup\limits_{\mu\in\Theta}\{p(X;\mu)\}}{\sup\limits_{\mu\in\Theta_0}\{p(X;\mu)\}}=n(\bar{X}-\mu_0)^T V^{-1}(\bar{X}-\mu_0)$$

令$D=n(\bar{X}-\mu_0)^T V^{-1}(\bar{X}-\mu_0)$，则可以证明当$H_0$成立时，即$\mu=\mu_0$时，$D\sim\chi^2(p)$

而当H_0不成立时，D有偏大的趋势。因此，对给定的显著性水平α，当

$$D=n(\bar{X}-\mu_0)^T V^{-1}(\bar{X}-\mu_0)\geq\chi^2_{1-\alpha}(p)$$

时拒绝H_0，否则接受H_0，即拒绝域为

$$W=\left\{D: D\geq\chi^2_{1-\alpha}(p)\right\}$$

实际上，检验统计量D可以看成是相应一维情形的推广。在一维情形中，当σ^2已知，使用统计量$z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$作为检验统计量。当$H_0$成立时，$z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$，平方有$z^2=n(\bar{X}-\mu_0)^T V^{-1}(\bar{X}-\mu_0)\sim\chi^2(1)$。在多元情形，$\bar{X}-\mu_0$变成列向量，而总体方差$\sigma^2$变为协方差矩阵$V$，就可得到检验统计量$D$。

§ 4.4.2 协差阵V 未知时，均值μ 检验

设$X_1, X_2, \cdots, X_n (n > p)$是多元正态总体$N_p(\mu, V)$的简单样本，其中$V$未知。考虑假设检验问题

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$$

仍根据似然比原则导出检验统计量及检验的拒绝域。

当V未知时，样本的联合密度函数为

$$p(X;\mu,V)=(2\pi)^{-\frac{np}{2}}|V|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}tr\{V^{-1}[S+n(\bar{X}-\mu)(\bar{X}-\mu)^T]\}\right\}$$

参数空间为$\Theta=\{(\mu,V):\mu\in R^p,V>0\}$，由定理4.3.1知，$\mu$和$V$的极大似然估计分别为$\bar{X}$和$\frac{1}{n}S$。

$$\sup_{\theta\in\Theta}\{p(X;\mu,V)\}=p\left(X;\bar{X},\frac{1}{n}S\right)=(2\pi)^{-\frac{np}{2}}|S|^{-\frac{n}{2}}n^{\frac{np}{2}}exp\left\{-\frac{np}{2}\right\}$$

原假设H_0的参数空间为$\Theta_0 = \{(\mu_0, V): V > 0\}$，与定理4.3.1证明过程类似，可得$V$的极大似然估计为

$$\hat{V}_0 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)^T$$

$$= \frac{1}{n}[S + n(\bar{X} - \mu_0)(\bar{X} - \mu_0)^T]$$

$$\sup_{\theta\in\Theta_0}\{p(X;\mu,V)\}=p\left(X;\mu_0,\hat{V}_0\right)$$

$$=(2\pi)^{-\frac{np}{2}}\left|S+n(\bar{X}-\mu_0)(\bar{X}-\mu_0)^T\right|^{-\frac{n}{2}}n^{\frac{np}{2}}exp\left\{-\frac{np}{2}\right\}$$

进而检验假设H_0的似然比为

$$\lambda(X)=\frac{\sup_{\theta\in\Theta}\{p(X;\mu,V)\}}{\sup_{\theta\in\Theta_0}\{p(X;\mu,V)\}}=\frac{|S|^{-\frac{n}{2}}}{\left|S+n(\bar{X}-\mu_0)(\bar{X}-\mu_0)^T\right|^{-\frac{n}{2}}}$$

$$=\left[1+(\bar{X}-\mu_0)^TS^{-1}(\bar{X}-\mu_0)\right]^{\frac{n}{2}}$$

令$F=\frac{n}{p}(n-p)(\bar{X}-\mu_0)^T S^{-1}(\bar{X}-\mu_0)$，则可以证明

当H_0成立时，即$\mu=\mu_0$时，$F \sim F(p,n-p)$

而当H_0不成立时，F有偏大的趋势。因此，对给定的显著性水平α，当

$$F=\frac{n}{p}(n-p)(\bar{X}-\mu_0)^T S^{-1}(\bar{X}-\mu_0) \geq F_{1-\alpha}(p,n-p)$$

时拒绝H_0，否则接受H_0，即拒绝域为

$$W=\{F: F \geq F_{1-\alpha}(p,n-p)\}$$

§4.4.3 两个正态总体均值相等的检验

设$\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{n_1} (n_1 > p)$是多元正态总体$N_p(\boldsymbol{\mu}_1, \boldsymbol{V})$的简单样本， $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_{n_2} (n_2 > p)$是多元正态总体$N_p(\boldsymbol{\mu}_2, \boldsymbol{V})$的简单样本，且两个样本相互独立，协方差阵$\boldsymbol{V} > 0$。考虑假设检验问题

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, H_1: \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

根据协方差阵$\boldsymbol{V}$已知和未知分两种情形：

（1）V已知

在一维情形，当σ^2已知时，检验两个正态总体均值相等

的检验统计量为$z=\frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2}}$，平方并重新整理有

$$z^2=\frac{n_1 n_2}{n_1+n_2}(\bar{X}-\bar{Y})(\sigma^2)^{-1}(\bar{X}-\bar{Y})$$

推广到p维情形，$\bar{X}-\bar{Y}$变为向量，σ^2变为协方差阵V。

检验统计量$D^*=\frac{n_1 n_2}{n_1+n_2}(\bar{X}-\bar{Y})^T V^{-1}(\bar{X}-\bar{Y})$

可以证明当H_0成立时，即$\mu_1=\mu_2$时，

$$D^*=\frac{n_1n_2}{n_1+n_2}(\bar{X}-\bar{Y})^TV^{-1}(\bar{X}-\bar{Y})\sim\chi^2(p)$$

而当H_0不成立时，D^*有偏大的趋势。因此，对给定的显著性水平α，当

$$D^*=\frac{n_1n_2}{n_1+n_2}(\bar{X}-\bar{Y})^TV^{-1}(\bar{X}-\bar{Y})\geq\chi^2_{1-\alpha}(p)$$

时拒绝H_0，否则接受H_0，即拒绝域为

$$W=\{D^*:D^*\geq\chi^2_{1-\alpha}(p)\}$$

（2）V未知

在一维情形，当σ^2未知时，检验两个正态总体均值相

等的检验统计量为$t=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_w^2}}$，其中

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

$$S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2,$$

$$S_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$$

即用作估计σ^2的S_w^2是样本方差S_1^2和S_2^2的加权平均。对t平方并重新整理有$t^2=\frac{n_1n_2}{n_1+n_2}(\bar{X}-\bar{Y})(S_w^2)^{-1}(\bar{X}-\bar{Y})$

推广到p维情形，$\bar{X}-\bar{Y}$变为向量，用样本协方差阵

$$\hat{V}_1 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})(X_i - \bar{X})^T, \hat{V}_2$$

$$= \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})(Y_i - \bar{Y})^T$$

分别替代一维时的样本方差S_1^2和S_2^2，S_w^2变为样本联合

协方差阵$\hat{V} = \frac{(n_1-1)\hat{V}_1+(n_2-1)\hat{V}_2}{n_1+n_2-2}$

检验统计量$F^* = \frac{n_1 n_2 (n_1 + n_2 - p - 1)}{p(n_1 + n_2)(n_1 + n_2 - 2)} (\bar{X} - \bar{Y})^T \hat{V}^{-1} (\bar{X} - \bar{Y})$

其中$\hat{V} = \frac{1}{n_1 + n_2 - 2}(S_1 + S_2)$是协方差阵$V$的估计量。

可以证明当H_0成立时，即$\mu_1 = \mu_2$时，

$$F^* \sim F(p, n_1 + n_2 - p - 1)$$

而当H_0不成立时，F^*有偏大的趋势。因此，对给定的显著性水平α，拒绝域为

$$W = \{F^* : F^* \geq F_{1-\alpha}(p, n_1 + n_2 - p - 1)\}$$