第六节　多元正态分布

本节的向量都是列向量, 例如：$\mu=(\mu_1,\mu_2,\cdots,\mu_n)^T$, $X=(X_1,X_2,\cdots,X_n)^T$.　以Σ表示矩阵

$$\Sigma=\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

以Σ^{-1}表示Σ的逆矩阵，$\det\Sigma$表示Σ的行列式的值.

多元正态分布有三种形式的定义：密度函数定义, 特征函数定义、标准正态分布线性变换形式的定义, 本节采用第三种形式的定义.

一　n元正态分布

从最简单的情况讨论．设$Y_1,Y_2,\cdots,Y_n$是定义在同一概率空间$(\Omega,\mathbb{F},P)$上的相互独立的随机变量, 并服从$N(0,1)$，因而n维随机向量$Y=(Y_1,Y_2,\cdots,Y_n)^T$的联合密度函数为

$$p(y)=\frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}(y_1^2+y_2^2+\cdots+y_n^2)\right\}$$

$$=\frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}y^T I_n y\right\},\qquad y\in R^n.$$

显然, Y的取值范围是整个n维空间R^n，且$EY=0$, 协方差矩阵为单位矩阵I_n，我们称之为标准n元正态分布，记作$N(0,I_n)$．易求得标准n元正态分布的特征

函数为

$$f(t)=f_1(t_1)\cdots f_n(t_n)=\exp\left\{-\frac{1}{2}t^T t\right\},\quad \forall t\in R^n.$$

一般地，我们有如下定义.

定义4.6.1 设n 维随机向量$Y\sim N(0,I_n)$，A为任意n 阶实矩阵，$\mu=(\mu_1,\mu_2,\cdots,\mu_n)^T$是$n$ 维实向量，则将n 维随机向量

$$X=AY+\mu$$

所服从的分布称为n元正态分布，记作$N(\mu,\Sigma)$，其中$\Sigma=\mathrm{cov}(X,X)=AA^T$是非负定矩阵. 显然$X$的取值范围是$R^n$或其子集.

定理4.6.1 n元正态分布$X \sim N(\mu, \Sigma)$的特征函数为

$$f(t) = \exp\left\{ \mathrm{i}\mu^T t - t^T \Sigma t / 2 \right\}, \quad \forall t \in R^n.$$

证明 标准n元正态分布Y的特征函数为

$$f_Y(t) = \exp\left\{ -t^T t / 2 \right\},$$

从而$X = AY + \mu$的特征函数为

$$f(t) = E\exp\left\{ \mathrm{i}t^T X \right\} = E\exp\left\{ \mathrm{i}t^T (AY + \mu) \right\}$$

$$= \exp\left\{ \mathrm{i}\mu^T t \right\} E\exp\left\{ \mathrm{i}(t^T A)Y \right\}$$

$$= \exp\left\{ \mathrm{i}\mu^T t \right\} \exp\left\{ -(t^T A)(A^T t)/2 \right\}$$

$$= \exp\left\{ \mathrm{i}\mu^T t - t^T \Sigma t / 2 \right\}$$

由上面的特征函数表达式知, n元正态分布$N(\mu,\Sigma)$完全由其一阶和二阶矩决定，由于其协方差矩阵为$\Sigma = AA^T$，根据线性代数的知识，半正定矩阵Σ的秩等于矩阵A的秩，因而Σ是正定矩阵当且仅当A是非奇异矩阵. 当Σ是正定矩阵时，n元正态分布$N(\mu,\Sigma)$具有密度函数.

定理4.6.2 若$X \sim N(\mu,\Sigma)$，其中Σ是n 阶正定矩阵，则X的密度函数为

$$p(x) = \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

定理3.3.2 如果对于$(Y_1,Y_2,\cdots,Y_n)$的每一组可能的取值, 方程组 （存在唯一的解）

$$\begin{cases} y_1 = g_1(x_1,x_2,\cdots,x_n) \\ \cdots\cdots\cdots\cdots\cdots \\ y_n = g_n(x_1,x_2,\cdots,x_n) \end{cases} \quad \begin{cases} x_1 = h_1(y_1,y_2,\cdots,y_n) \\ \cdots\cdots\cdots\cdots\cdots \\ x_n = h_n(y_1,y_2,\cdots,y_n) \end{cases}$$

其中每个$h_j(y_1,y_2,\cdots,y_n)$都有一阶连续偏导数, 那么随机向量$(Y_1,Y_2,\cdots,Y_n)$是连续的，具有密度函数

$$q(y_1,\cdots,y_n)=\begin{cases} p(h_1,\cdots,h_n)|J|, & \text{当}(y_1,\cdots,y_n)\text{属于值域时}, \\ 0, & \text{其它}. \end{cases}$$

其中

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

证明　因$X \sim N(\mu, \Sigma)$，则存在$Y \sim N(0, I_n)$，使得$X = AY + \mu$，其中$AA^T = \Sigma$. 而线性方程组

$$x = Ay + \mu$$

的解为

$$y = A^{-1}(x - \mu)$$

变换的雅可比行列式为$\det A^{-1}$,其绝对值为$(\det \Sigma)^{-1/2}$.

由于Y的密度函数为

$$q(y) = \frac{1}{(2\pi)^{n/2}} \exp\left\{ -y^T I_n y / 2 \right\}$$

因而X的密度函数为

$$p(x) = \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{1/2}} \exp\left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

当Σ不是正定矩阵时，有$\det\Sigma = 0$. 称为退化正态分布或奇异正态分布.不妨设$\text{rank}(\Sigma) = r < n$，此时有$\text{rank}(A) = r$，因而$X = AY + \mu$的值域是$r$维超平面. 这时概率分布集中于$R^n$的$r$ 维超平面$\left\{Ay + \mu : y \in R^n\right\}$上.

例如: 对二维正态分布$N(\mu,\Sigma)$，当$r = 0$时，$X = \mu$退化为一个点$\mu = (\mu_1, \mu_2)^T$；当$r = 1$时，不妨记

$$A = \begin{pmatrix} a & b \\ ca & cb \end{pmatrix}$$

则 $X_1 = aY_1 + bY_2 + \mu_1$，$X_2 = caY_1 + cbY_2 + \mu_2$，因而

$$X = (X_1, X_2)^T$$

是直线$x_2 - \mu_2 = c(x_1 - \mu_1)$上的退化分布.

二　n元正态分布的性质

下面的讨论中，我们总是假定随机向量$X=(X_1,X_2,\cdots,X_n)^T$服从n元正态分布$N(\mu,\Sigma)$.

定理4.6.3　X的任一m维子向量$(X_{k_1},X_{k_2},\cdots,X_{k_m})^T$服从正态分布$N(\tilde{\mu},\tilde{\Sigma})$.

其中$\tilde{\mu}=(\mu_{k_1},\mu_{k_2},\cdots,\mu_{k_m})^T$，$\tilde{\Sigma}$为保留$\Sigma$的第$k_1,k_2,\cdots,k_m$行和第$k_1,k_2,\cdots,k_m$列的$m$阶矩阵.

证明　只须在X的特征函数中，对一切不等于$k_1,k_2,\cdots,k_m$的j，令$t_j=0$，即得$(X_{k_1},X_{k_2},\cdots,X_{k_m})^T$的特征函数

$$\tilde{f}(\tilde{t})=\exp\left\{\mathrm{i}\tilde{\mu}^T\tilde{t}-\frac{1}{2}\tilde{t}^T\tilde{\Sigma}\tilde{t}\right\}$$

这里$\tilde{t}=(t_{k_1},t_{k_2},\cdots,t_{k_m})^T$，这正是$N(\tilde{\mu},\tilde{\Sigma})$的特征函数.

定理4.6.3表明：多元正态分布的边际分布还是正态分布.

定理4.6.4　$X_1,X_2,\cdots,X_n$相互独立的充要条件是它们两两不相关.

证明　必要性显然.

下证充分性.

若$X_1,X_2,\cdots,X_n$两两不相关，即对一切$j\neq k$，有

$$\sigma_{jk}=E(X_j-EX_j)(X_k-EX_k)=0$$

因此协方差Σ是对角形矩阵，从而特征函数为

$$f(t)=\exp\left\{\mathrm{i}\mu^T t-\frac{1}{2}t^T\Sigma t\right\}$$

$$=\prod_{k=1}^{n}\exp\left\{\mathrm{i}\mu_k t_k-\frac{1}{2}\sigma_{kk}t_k^2\right\}=\prod_{k=1}^{n}f_{X_k}(t_k)$$

由多元特征函数的性质可知相互独立.

定理4.6.5 若$X=\begin{pmatrix}X_1\\X_2\end{pmatrix}$，这里$X_1$与$X_2$是$X$的子向量，

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

其中Σ_{11}及Σ_{22}分别是X_1与X_2的协方差矩阵，Σ_{12}是由X_1与X_2的相应分量的协方差构成的相互协方差矩阵，则X_1与X_2独立的充要条件是$\Sigma_{12}=0$.

证明 必要性 若X_1与X_2独立, 则X_1的任一分量与X_2的任一分量独立，因此其协方差为0，从而由它们构成的矩阵$\Sigma_{12}=\mathbf{0}$.

充分性 由$\sigma_{ij}=\sigma_{ji}$，因此$\Sigma_{21}=(\Sigma_{12})^T=\mathbf{0}$，若$\boldsymbol{t}=\begin{pmatrix}\boldsymbol{t}_1\\ \boldsymbol{t}_2\end{pmatrix}$

这里$\boldsymbol{t}_1$与X_1有相同维数，$\boldsymbol{t}_2$与X_2有相同维数，则

$$\boldsymbol{t}^T\Sigma\boldsymbol{t}=\boldsymbol{t}_1^T\Sigma_{11}\boldsymbol{t}_1+2\boldsymbol{t}_1^T\Sigma_{12}\boldsymbol{t}_2+\boldsymbol{t}_2^T\Sigma_{22}\boldsymbol{t}_2=\boldsymbol{t}_1^T\Sigma_{11}\boldsymbol{t}_1+\boldsymbol{t}_2^T\Sigma_{22}\boldsymbol{t}_2$$

记$\mu=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$，其中$\mu_1$，$\mu_2$分别为$X_1$与$X_2$的数学期望, 则

$$f(\boldsymbol{t})=\exp\left\{\mathrm{i}\mu^T\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}^T\Sigma\boldsymbol{t}\right\}$$

$$= \exp\left\{ \mathrm{i}\mu_1^T t_1 + \mathrm{i}\mu_2^T t_2 - \frac{1}{2} t_1^T \Sigma_{11} t_1 - \frac{1}{2} t_2^T \Sigma_{22} t_2 \right\}$$

$$= \exp\left\{ \mathrm{i}\mu_1^T t_1 - \frac{1}{2} t_1^T \Sigma_{11} t_1 \right\} \exp\left\{ \mathrm{i}\mu_2^T t_2 - \frac{1}{2} t_2^T \Sigma_{22} t_2 \right\}$$

$$= f_{X_1}(t_1) f_{X_2}(t_2)$$

由多元特征函数的性质6可知X_1与X_2独立.

类似可以证明，若X的子向量两两不相关，则它们也相互独立.

三　线性变换

服从正态分布的随机向量在线性变换下具有许多特殊的性质，这些性质有很大的理论和使用价值, 下面讨论一些最基本的性质.

下面考虑n 元正态分布在线性变换下的分布.

定理4.6.6　随机向量$X=(X_1,X_2,\cdots,X_n)^T$服从n元正态分布$N(\mu,\Sigma)$的充要条件是它的任何一个线性组合 $Y=l^TX=\sum_{j=1}^{n}l_jX_j$ 服从一元正态分布

$$N(\sum_{j=1}^{n}l_j\mu_j,\sum_{j,k=1}^{n}l_jl_k\sigma_{jk}).$$

证明　必要性：若$X \sim N(\mu, \Sigma)$，则

$$f_X(t) = E\exp\left\{\mathrm{i}t^T X\right\} = \exp\left\{\mathrm{i}\mu^T t - \frac{1}{2}t^T \Sigma t\right\}$$

特别取$t = (t_1, t_2, \cdots, t_n)^T = (ul_1, ul_2, \cdots, ul_n)^T = ul$,
这里u是任意实数，则随机变量Y的特征函数为

$$f_Y(u) = E\exp\left\{\mathrm{i}uY\right\} = E\exp\left\{\mathrm{i}ul^T X\right\}$$

$$= \exp\left\{\mathrm{i}(\mu^T l)u - \frac{1}{2}(l^T \Sigma l)u^2\right\}$$

这说明随机变量　$Y \sim N(\sum_{j=1}^{n} l_j \mu_j, \sum_{j,k=1}^{n} l_j l_k \sigma_{jk})$.

充分性：若对任意的l，$Y = l^T X \sim N(l^T \mu, l^T \Sigma l)$,

则随机变量Y的特征函数为

$$f_Y(u) = E\exp\left\{\mathrm{i}uY\right\} = E\exp\left\{\mathrm{i}ul^T X\right\}$$

$$= \exp\left\{\mathrm{i}(\mu^T l)u - \frac{1}{2}(l^T \Sigma l)u^2\right\}$$

在上面的式子中取$\mathrm{u} = 1$，得

$$E\exp\left\{\mathrm{i}l^T X\right\} = f_Y(1) = \exp\left\{\mathrm{i}\mu^T l - \frac{1}{2}l^T \Sigma l\right\}$$

由的任意性l，这说明$X \sim N(\mu, \Sigma)$.

定理4.6.7 若随机向量$X = (X_1, X_2, \cdots, X_n)^T$服从$n$元正态分布$N(\mu, \Sigma)$，而$C$为$m \times n$任意阵，则$Y = CX$服从$m$元正态分布$N(C\mu, C\Sigma C^T)$.

证明　因为对任意m维实值列向量t,

$$f_Y(t) = E\exp\left\{\mathrm{i}t^T CX\right\} = E\exp\left\{\mathrm{i}(C^T t)^T X\right\}$$

$$= \exp\left\{\mathrm{i}\mu^T (C^T t) - \frac{1}{2}(C^T t)^T \Sigma (C^T t)\right\}$$

$$= \exp\left\{\mathrm{i}(C\mu)^T t - \frac{1}{2} t^T (C\Sigma C^T) t\right\}$$

因而$Y = CX$服从m元正态分布$N(C\mu, C\Sigma C^T)$.

可以用线性变换的定义来证明

若$X \sim N(\mu, \Sigma)$，设$X = AZ + \mu$，其中$Z \sim N(0, I_n)$,

$AA^T = \Sigma$. 则　$Y = CX = CAZ + C\mu$,

(1) $m = n$,　(2) $m < n$,　(3) $m > n$. . . . . .

定理表明正态变量在线性变换下还是正态变量，简称为正态变量的线性变换不变性.

推论1　若$X \sim N(\mu,\Sigma)$，则存在一个正交变换U，使得$Y=UX$是一个具有独立正态分布分量的随机向量，它的数学期望为$U\mu$，而它的方差分量为Σ的特征值.

证明　根据矩阵理论知：对实对称矩阵Σ，存在正交阵U，使得$U\Sigma U^T=D$，其中$D=diag(d_1,d_2,\cdots,d_n)$，这里$d_1,d_2,\cdots,d_n$是$\Sigma$的特征值. 若$\Sigma$的秩为$r$，则有$r$个特征值不为零. 根据定理4.6.7知：$Y \sim N(U\mu,D)$.

从推论1得知：若Σ的秩为r，则Y有$n-r$个分量的方差为零，这$n-r$个分量退化为常数，也即Y退化到一个r维超平面上, 这正是我们前面已有的结论.

推论1说明，对于正态变量，可以对其进行正交变换, 使其既保持正交性不变, 又让各分量独立，这种方法在数理统计中十分有用.

第3章第3节的例10给出了坐标旋转(正交变换的一种)把二维正态随机变量化为独立分量的例子.

推论2　若n元正态变量的各分量相互独立，并且具有相同的方差, 则它在正交变换下, 仍然保持上述性质.

证明　设$X \sim N(\mu,\sigma^2 I)$，$Y=UX$，其中U是正交阵.
根据定理4.6.7知：

$$D(Y)=U(\sigma^2 I)U^T=\sigma^2 I$$

因此Y的各分量相互独立且具有相同的方差.

推论3　若$X \sim N(\mu,\Sigma)$，其中Σ是正定矩阵，则

$$(X-\mu)^T \Sigma^{-1}(X-\mu) \sim \chi^2(n).$$

证明　不妨设$X=AY+\mu$，其中$Y \sim N(0,I_n)$，$AA^T=\Sigma$. 而线性方程组$x=Ay+\mu$的解为

$$y=A^{-1}(x-\mu)$$

从而 $(X-\mu)^T \Sigma^{-1}(X-\mu)$

$$=(X-\mu)^T (AA^T)^{-1}(X-\mu)=Y^T Y \sim \chi^2(n).$$

四　条件分布

定理4.6.8　若$X=\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$，服从$n$元正态分布$N(\mu,\Sigma)$，$EX_1=\mu_1$，$EX_2=\mu_2$．这里$X_1$与$X_2$是$X$的子向量，记

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

其中Σ_{11}及Σ_{22}分别是X_1与X_2的协方差矩阵，Σ_{12}是由X_1与X_2的相应分量的协方差构成的相互协方差矩阵，则在给定$X_1=x_1$的条件下，X_2的条件分布还是正态分布，其条件数学期望为

$$\mu_{2\cdot 1}=E(X_2 \mid X_1=x_1)=\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1)$$

其条件方差

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

证明　作如下的线性变换

$$\begin{cases}\boldsymbol{Y}_1=\boldsymbol{X}_1\\ \boldsymbol{Y}_2=-\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{X}_1+\boldsymbol{X}_2\end{cases}$$

则$\boldsymbol{Y}_1$与$\boldsymbol{Y}_2$都服从正态分布，且

$$\boldsymbol{EY}_1=\boldsymbol{EX}_1=\boldsymbol{\mu}_1$$

$$\boldsymbol{EY}_2=-\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{EX}_1+\boldsymbol{EX}_2=\boldsymbol{\mu}_2-\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1$$

$$\boldsymbol{DY}_1=\boldsymbol{DX}_1=\Sigma_{11}$$

$$\boldsymbol{DY}_2=\boldsymbol{E}(\boldsymbol{Y}_2-\boldsymbol{EY}_2)(\boldsymbol{Y}_2-\boldsymbol{EY}_2)^T$$

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

$$COV(Y_1,Y_2)=E(Y_1-EY_1)(Y_2-EY_2)^T$$

$$=E(X_1-\mu_1)\left[(X_2-\mu_2)-\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1)\right]^T$$

$$=\Sigma_{12}-\Sigma_{11}(\Sigma_{21}\Sigma_{11}^{-1})^T=0$$

因而Y_1与Y_2相互独立, 又变换的雅可比行列式为**1.**
所以

$$p_X(x_1,x_2)=p_Y(y_1,y_2)=p_{Y_1}(y_1)p_{Y_2}(y_2).$$

显然$p_{X_1}(x_1)=p_{Y_1}(y_1)$，因而给定$X_1=x_1$下，$X_2$的条件密度函数为

$$p(x_2|X_1=x_1)=\frac{p_X(x_1,x_2)}{p_{X_1}(x_1)}=\frac{p_Y(y_1,y_2)}{p_{Y_1}(y_1)}$$

$$= p_{Y_2}(y_2) = p_{Y_2}(x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1)$$

因为Y_2服从正态分布$N(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$，

所以给定$X_1 = x_1$下，X_2的条件分布是正态分布，而且

$$\mu_{2\cdot 1} = E(X_2 \mid X_1 = x_1) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1),$$

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