第七章　多元分析初步

7.1 多元正态分布的定义及性质

7.2 多元正态分布参数的估计与假设检验

7.1 多元正态分布的定义及性质

一、多元正态分布的定义

二、多元正态分布的性质

一、多元正态分布的定义

定义7.1 若p维随机变量$X = (X_1, X_2, \cdots, X_p)^T$的概率密度为：

$$f(x_1, x_2, \cdots, x_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\} \tag{7.1}$$

则称X服从p维正态分布,也称X为p维正态变量,记为$X \sim N_p(\mu,\Sigma)$,其中μ为X的均值向量，Σ为X的协方差矩阵(简称为协差阵),Σ为正定阵，记为$\Sigma > 0$.

验证（7.1）式定义的$f(x_1, x_2, \cdots, x_p)$确为概率密度函数，需要证明：

(1) 非负性 $f(x_1, x_2, \cdots, x_p) \geq 0$,

(2) 规范性 $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \cdots, x_p) dx_1 dx_2 \cdots dx_p = 1$

（1）显然成立，下面证明（2）成立

因为$\Sigma > 0$,故存在Q使得$\Sigma^{-1} = Q^T Q$.

做变换$y = Q(x-\mu)$,则

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = (x-\mu)^T Q^T Q (x-\mu) = y^T y,$$

这个变换的雅可比行列式为$|Q|^{-1}$.因$|\Sigma^{-\frac{1}{2}}|=|\Sigma^{-1}|^{\frac{1}{2}}=\pm|Q|$,

此处当$|Q|>0$时取正号，否则取负号，于是

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1, x_2,\cdots, x_p)dx_1dx_2\cdots dx_p$$

$$=\frac{|\Sigma|^{-1/2}}{(2\pi)^{p/2}}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}dx_1\cdots dx_p$$

$$=\frac{|\Sigma|^{-1/2}|Q|^{-1}}{(2\pi)^{p/2}}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^Ty}dy_1dy_2\cdots dy_p$$

$$=\prod_{i=1}^{p}\int_{-\infty}^{+\infty}\frac{1}{(2\pi)^{1/2}}e^{-\frac{1}{2}y_i^2}dy_i=1$$

得证。

随机向量的数字特征

1、数学期望：均值

定义：

$$\mathbf{X}=\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1q} \\ X_{21} & X_{22} & \cdots & X_{2q} \\ \vdots & \vdots & & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pq} \end{bmatrix}$$

是由随机变量构成的随机矩阵，定义X的数学期望为

$$E(\mathbf{X})=\begin{bmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1q}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2q}) \\ \vdots & \vdots & & \vdots \\ E(X_{p1}) & E(X_{p2}) & \cdots & E(X_{pq}) \end{bmatrix}$$

特别当时$q=1$，便可得到随机向量$X=(X_1,X_2,\cdots,X_p)'$的数学期望为

$$E(X)=(E(X_1),E(X_2),\cdots,E(X_p))'$$

随机变量X的均值为

$$E(X)\ =\ (EX_1,\ EX_2,\cdots,\ EX_p)^T=\ (\mu_1,\ \mu_2,\cdots,\ \mu_p)^T=\mu$$

其中 μ 是 p 维列向量，称为均值向量。 (7.2)

定义随机变量 X 的方差为

$$D(\boldsymbol{X}) = E[(\boldsymbol{X}-E\boldsymbol{X})(\boldsymbol{X}-E\boldsymbol{X})^{\mathrm{T}}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \stackrel{def}{=} \boldsymbol{\Sigma} \tag{7.3}$$

其中 $\sigma_{ij} = E[(X_i - EX_i)(X_j - EX_j)]$, $i, j = 1,2,\cdots, p$.

上式给出的 X 的方差实际上是协方差阵, 由协方差阵的定义可知,任何随机变量 X 的协方差阵皆为对称阵, 且总是非负定的, 大多数情况下是正定的. 将"Σ 是正定阵"简记为 $\Sigma > 0$, 而将"Σ是非负定阵"简记为 $\Sigma \geq 0$.

由线性代数知识可知，若协方差 $\boldsymbol{\Sigma} > \mathbf{0}$，则 $\boldsymbol{\Sigma}$ 的特征根均为正数，$\boldsymbol{\Sigma}$ 的行列式 $|\boldsymbol{\Sigma}| > 0$，且 $\boldsymbol{\Sigma}$ 能分解成 $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$,其中 $\boldsymbol{A}$为非奇异对称阵，$\boldsymbol{A}^T = \boldsymbol{A}$.

若记$\boldsymbol{A} = \boldsymbol{\Sigma}^{\frac{1}{2}}$,则有$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{A}^2$.

二、多元正态分布的性质

性质1 若X服从p维正态分布$N_p(\mu,\Sigma)$,

则$E(X)=\mu, D(X)=\Sigma$.

即p维正态分布由其均值向量和协方差阵唯一确定.

性质2 对于任意p维向量μ及$p\times p$非负定对称矩阵Σ,
必有p维正态变量X服从$N_p(\mu,\Sigma)$分布.

$$\begin{pmatrix}\mu\\ \Sigma\end{pmatrix} \Leftrightarrow N_p(\mu,\Sigma)$$

性质3*** 若C为$m\times p$矩阵,b为$m\times 1$向量, $m\geq 1$,

$X\sim N_p(\mu,\Sigma),Y=CX+b$,则$Y$服从$m$维正态分布,

且$EY=C\mu+b,\ Cov(Y,Y)\ =C\Sigma C^T$,

即$Y\sim N_m(C\mu+b,C\Sigma C^T)$.

性质 3 说明, 多维正态随机变量的任意线性组合仍为正态分布.

例　设三维随机变量$X=(X_1,X_2,X_3)^T$,且

$$X=\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix}\sim N_3\left(\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}1&1&0\\1&2&0\\0&0&3\end{pmatrix}\right)$$

若$C=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$, $b=\begin{pmatrix}1\\2\\1\end{pmatrix}$,

则$Y=CX+b\sim$

$$N\left(\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix}+\begin{pmatrix}1\\2\\1\end{pmatrix},\ \begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}\begin{pmatrix}1&1&0\\1&2&0\\0&0&3\end{pmatrix}\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}^T\right)$$

若$C'=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$，$b'=\begin{pmatrix}2\\ 1\end{pmatrix}$，

则$Y=C'X+b'\sim$

$$N\left(\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1\\ 1\\ 0\end{pmatrix}+\begin{pmatrix}2\\ 1\end{pmatrix},\ \begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 1 & 0\\ 1 & 2 & 0\\ 0 & 0 & 3\end{pmatrix}\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}^{T}\right),$$

$$若\boldsymbol{C''}=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\\3&1&2\end{pmatrix},\ \boldsymbol{b''}=\begin{pmatrix}1\\2\\3\\4\end{pmatrix},$$

$$则\boldsymbol{Y}=\boldsymbol{C''X}+\boldsymbol{b''}\sim$$

$$\boldsymbol{N}\left(\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\\3&1&2\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix}+\begin{pmatrix}1\\2\\3\\4\end{pmatrix},\ \begin{pmatrix}0&1&0\\0&0&1\\1&0&0\\3&1&2\end{pmatrix}\begin{pmatrix}1&1&0\\1&2&0\\0&0&3\end{pmatrix}\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\\3&1&2\end{pmatrix}^{T}\right),$$

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{X}_2\\\boldsymbol{X}_3\\\boldsymbol{X}_1\end{pmatrix}$$的分布？

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{X}_2\\\boldsymbol{X}_3\\\boldsymbol{X}_1\end{pmatrix}=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}\begin{pmatrix}\boldsymbol{X}_1\\\boldsymbol{X}_2\\\boldsymbol{X}_3\end{pmatrix}$$

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{X}_2\\\boldsymbol{X}_3\\\boldsymbol{X}_1\end{pmatrix}\sim \boldsymbol{N}\left(\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix},\ \begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}\begin{pmatrix}1&1&0\\1&2&0\\0&0&3\end{pmatrix}\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}^{\boldsymbol{T}}\right)$$

性质4　X为p维正态变量$\Leftrightarrow$对任一p维向量C, $Y=C^T X$是一维正态变量.

性质5　$\mathbf{X}=\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\begin{matrix} m \\ p-m \end{matrix}$ 为p维正态变量, 则$\mathbf{X}_1$, $\mathbf{X}_2$互不相关(指$\mathbf{X}_1$的任一分量与$\mathbf{X}_2$的任一分量均不相关)的充要条件是$\mathbf{X}_1$与$\mathbf{X}_2$独立.

$$\Sigma=\begin{bmatrix} \Sigma_{11} & \mathbf{0}_{m\times(p-m)} \\ \mathbf{0}_{(p-m)\times m} & \Sigma_{22} \end{bmatrix}\begin{matrix} m \\ p-m \end{matrix}$$

性质6 $X \sim N_p(\mu,\Sigma)$，且Σ的秩为$m(m \le p)$

$\Leftrightarrow$

X可以表示为

$$X = \mu + B_{p\times m}Y \qquad (BB^T = \Sigma)$$

其中Y为m维$N_m(0,I)$变量.

即X可表示为m个相互独立的$N(0,1)$分布变量与常数的线性组合.

证明　必要性.

因为$\boldsymbol{\Sigma}$是秩为m的对称非负定阵,所以必有正交阵$\boldsymbol{U}_{p\times p}$,使

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{U}=\begin{bmatrix}\lambda_1 & & & & & \\ & \ddots & & & \boldsymbol{0} & \\ & & \lambda_m & & & \\ & & & 0 & & \\ & \boldsymbol{0} & & & \ddots & \\ & & & & & 0\end{bmatrix}\xlongequal{\text{def}}\begin{bmatrix}\boldsymbol{\Lambda} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}$$

$$\lambda_i>0,1\leq i\leq m,$$

其中$\boldsymbol{\Lambda}$为$m\times m$对角阵,则

$$\begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} U^T \Sigma U \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \overset{def}{=} D \qquad (7.4)$$

其中 I_m 为 m 阶单位阵, $\Lambda^{-\frac{1}{2}} = \mathrm{diag}(\lambda_1^{-\frac{1}{2}}, \lambda_2^{-\frac{1}{2}}, \cdots, \lambda_m^{-\frac{1}{2}})$，记

$$Z = \begin{bmatrix} Y \\ W \end{bmatrix} = \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} U^T (X - \mu)$$

$$= \underbrace{\begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} U^T}_{C} X - \underbrace{\begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} U^T \mu}_{b} = CX - b,$$

$$EZ = CEX - b = \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} U^T \mu - \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} U^T \mu = 0,$$

$$Cov(Z, Z) = C\Sigma C^T = \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} U^T \Sigma U \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} = D,$$

则由性质3及式(7.4)知Z服从$N(0, D)$分布，特别地由Z的前m个分量构成的m维向量Y服从$N_m(0, I_m)$分布，而W以概率1为零向量.

另一方面，由上式不难推出

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

$$=\boldsymbol{U}\begin{bmatrix}\Lambda^{\frac{1}{2}}\\ 0\end{bmatrix}\boldsymbol{Y}.$$

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

充分性． 设$X=\mu+BY, BB^T=\Sigma, Y\sim N_m(0,I)$,则对任意p维常数向量$l$,有

$$l^T X = l^T(\mu+BY)\sim N(l^T\mu, l^T BB^T l) = N(l^T\mu, l^T\Sigma l),$$

即$l^T X\sim N(l^T\mu, l^T\Sigma l)$,

这说明$l^T X$是一维正态变量，由性质4知，

$$X\sim N_p(\mu,\Sigma).$$

性质7 若$X=\begin{array}{c}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\begin{array}{l} m \\ p-m \end{array}\end{array} \sim N_p(\mu,\Sigma),\Sigma>0$,则$\mu$和$\Sigma$也有相应的分块表示

$$\mu=\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\begin{array}{l} m \\ p-m \end{array}, \quad \Sigma=\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\begin{array}{l} m \\ p-m \end{array},$$

有 $X_1 \sim N_m(\mu_1,\Sigma_{11}),X_2 \sim N_{p-m}(\mu_2,\Sigma_{22})$,

进一步，$\mathbf{X}_1$关于$\mathbf{X}_2=x_2$的条件分布为$\mathbf{N}_m(\mu_{1,2},\Sigma_{11,2})$,

$$\mu_{1,2}=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2), \quad \Sigma_{11,2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

可看出，条件均值是x_2的线性函数，条件方差与x_2无关.

性质7表明

多元正态分布的任何低维边缘分布仍为正态分布。反之，若随机变量的任何边缘分布均为正态分布，但不一定能导出该随机向量服从多元正态分布。

反例

令 (X,Y) 的联合密度函数为

$$p(x,y)=\frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}}(1+\sin x\sin y),$$

显然, (X,Y) 不服从正态分布 ,但是

$$p_X(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},\quad p_Y(y)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2}}.$$

因此边缘分布均为正态分布的随机变量,其联合分布不一定是二维正态分布.

例　设三维随机变量$X=(X_1,X_2,X_3)^T$,且

$$X=\begin{pmatrix} X_1 \\ \hline X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ \hline 1 \\ 0 \end{pmatrix}, \left(\begin{array}{c|cc} 1 & 1 & 0 \\ \hline 1 & 2 & 0 \\ 0 & 0 & 3 \end{array}\right)\right)$$

则有　$X_1 \sim N(1,1)$,

$$\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\right).$$

例 7.1 二元正态分布

解 设$\mu^T=(\mu_1,\mu_2)^T$,

$$\Sigma=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix},\text{其中}\sigma_1^2>0,\sigma_2^2>0,-1<\rho<1,$$

因$|\Sigma|=\sigma_1^2\sigma_2^2(1-\rho^2)>0$,即$\Sigma$正定，故$\Sigma^{-1}$存在。

$$\Sigma^{-1}=\frac{1}{1-\rho^2}\begin{bmatrix}\frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2}\\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}\end{bmatrix},$$

设$X^T=(X_1,X_2)^T\sim N_2(\mu,\Sigma)$,由于

$$(X-\mu)^T\Sigma^{-1}(X-\mu)$$

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

于是的概率密度函数为

$$f(x_1,x_2)=\left[\frac{1}{(2\pi)^2\sigma_1^2\sigma_2^2(1-\rho^2)}\right]^{-\frac{1}{2}}$$

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

容易验证,ρ是X_1和X_2的相关系数.当$\rho=0$时,Σ为对角阵,此时X_1和X_2相互独立.

由性质3知，X_i的边缘分布为$N(\mu_i,\ \sigma_i^2),\ i=\ 1,2.$

由性质7知，给定$X_2=x_2,X_1$的条件分布为

$N_1(\mu_{1,2},\Sigma_{11,2})$，其中

$$\mu_{1,2}=\mu_1+\rho\left(\frac{\sigma_1}{\sigma_2}\right)(x_2-\mu_2),\quad \Sigma_{11,2}=\sigma_1^2(1-\rho^2)$$

同理可得，在给定$X_1=x_1$时X_2的条件分布为$N(\mu_{2,1},\Sigma_{22,1})$，其中

$$\mu_{2,1}=\mu_2+\rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_1-\mu_1),\qquad \Sigma_{22,1}=\sigma_2^2(1-\rho^2)$$

性质8　若X服从p维$N_p(\mu,\Sigma)$分布，且$|\Sigma|>0$，则

$$\eta \overset{def}{=} (x-\mu)^T\Sigma^{-1}(x-\mu) \sim \chi^2(p)$$

其中$\chi^2(p)$表示自由度为p的χ^2分布.

证明 若令$Y \overset{def}{=} \Sigma^{-1/2}(X-\mu)$，则由性质3,

$Y=(Y_1, Y_2,\cdots, Y_p)^T$服从$p$维$N_p(0,I)$分布，因此

$$Y^TY=\sum_{i=1}^{p}Y_i^2$$

为p个相互独立服从$N(0,1)$分布的随机变量的平方和，

所以Y^TY服从$\chi^2(p)$分布，

另一方面

$$Y^T Y = (X-\mu)^T \Sigma^{-1/2} \Sigma^{-1/2} (X-\mu)$$

$$= (X-\mu)^T \Sigma^{-1} (X-\mu) = \eta$$

所以$\eta \sim \chi^2(p)$.

Thank you