

Ch2 多元正态分布及参数的估计

§2.1 随机向量

Def: n 个样品排为 $n \times p$ 矩阵:
$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n) = \begin{pmatrix} \mathbf{z}_1' \\ \mathbf{z}_2' \\ \vdots \\ \mathbf{z}_n' \end{pmatrix}$$

Def: 联合分布函数: $F(x_1, \dots, x_p) = P\{\mathbf{z}_1 \leq x_1, \mathbf{z}_2 \leq x_2, \dots, \mathbf{z}_p \leq x_p\}$ ($\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_p)'$ 为 p 维随机向量)

连续型随机向量: $\exists f(x_1, x_2, \dots, x_p) \geq 0$, s.t. $F(x_1, x_2, \dots, x_p) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_p} f(x_1, x_2, \dots, x_p) dx_1 \dots dx_p$
 \downarrow
 \mathbf{z} 的联合概率密度函数

边缘分布有 $C_1^p + C_2^p + \dots + C_p^p = 2^p - 1$ 个

计算 (x_1, x_2, x_3) 中边缘分布 $f_{22}(x_1, x_2) = P(\mathbf{z}_1 \leq x_1, \mathbf{z}_2 \leq x_2, \mathbf{z}_3 \in \mathbb{R}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} (\int_{-\infty}^{+\infty} f(u, v, w) dw) du dv$
 $= \int_{-\infty}^{+\infty} f(x_1, x_2, w) dw$

条件分布 $\mathbf{z} = \begin{pmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \end{pmatrix}$, 给定 $\mathbf{z}^{(2)}$, 称 $\mathbf{z}^{(1)}$ 的分布为条件分布

$$f_1(\mathbf{x}^{(1)} | \mathbf{x}^{(2)}) = f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) / f_2(\mathbf{x}^{(2)})$$

独立: $F(x_1, \dots, x_p) = F_1(x_1) \dots F_p(x_p)$ $f(x_1, \dots, x_p) = f_1(x_1) \dots f_p(x_p)$ (连续型)

数字特征

$$E(\mathbf{z}) = \begin{pmatrix} E(\mathbf{z}_1) \\ \vdots \\ E(\mathbf{z}_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

$$\Sigma = D(\mathbf{z}) = E[(\mathbf{z} - E(\mathbf{z}))(\mathbf{z} - E(\mathbf{z}))'] = \begin{pmatrix} \text{Cov}(\mathbf{z}_1, \mathbf{z}_1) & \text{Cov}(\mathbf{z}_1, \mathbf{z}_2) & \dots & \text{Cov}(\mathbf{z}_1, \mathbf{z}_p) \\ \text{Cov}(\mathbf{z}_2, \mathbf{z}_1) & \text{Cov}(\mathbf{z}_2, \mathbf{z}_2) & \dots & \text{Cov}(\mathbf{z}_2, \mathbf{z}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\mathbf{z}_p, \mathbf{z}_1) & \text{Cov}(\mathbf{z}_p, \mathbf{z}_2) & \dots & \text{Cov}(\mathbf{z}_p, \mathbf{z}_p) \end{pmatrix} = (\sigma_{ij})_{p \times p} \quad \text{对称非负定}$$

$$\text{Cov}(\mathbf{z}, \mathbf{y}) = E[(\mathbf{z} - E(\mathbf{z}))(\mathbf{y} - E(\mathbf{y}))'] = \begin{pmatrix} \text{Cov}(\mathbf{z}_1, \mathbf{y}_1) & \text{Cov}(\mathbf{z}_1, \mathbf{y}_2) & \dots & \text{Cov}(\mathbf{z}_1, \mathbf{y}_q) \\ \text{Cov}(\mathbf{z}_2, \mathbf{y}_1) & \text{Cov}(\mathbf{z}_2, \mathbf{y}_2) & \dots & \text{Cov}(\mathbf{z}_2, \mathbf{y}_q) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\mathbf{z}_p, \mathbf{y}_1) & \text{Cov}(\mathbf{z}_p, \mathbf{y}_2) & \dots & \text{Cov}(\mathbf{z}_p, \mathbf{y}_q) \end{pmatrix} \quad \text{若 } \text{Cov}(\mathbf{z}, \mathbf{y}) = 0 \text{ 则 } \mathbf{z}, \mathbf{y} \text{ 不相关}$$

相关阵 $R = (r_{ij})_{p \times p}$, $r_{ij} = \frac{\text{Cov}(\mathbf{z}_i, \mathbf{z}_j)}{\sqrt{\text{Var}(\mathbf{z}_i)} \sqrt{\text{Var}(\mathbf{z}_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}$, ($\text{Var}(\mathbf{z}_i) = \text{Cov}(\mathbf{z}_i, \mathbf{z}_i) \triangleq \sigma_{ii}$)

Thm: \mathbf{z}, \mathbf{y} 为随机向量, A, B 为常数矩阵

$$E(A\mathbf{z}) = AE(\mathbf{z}), E(A\mathbf{z}B) = AE(\mathbf{z})B, D(A\mathbf{z}) = AD(\mathbf{z})A', \text{Cov}(A\mathbf{z}, B\mathbf{y}) = ACov(\mathbf{z}, \mathbf{y})B'$$

$$E(\mathbf{z}'A\mathbf{z}) = E(\mathbf{z}')AE(\mathbf{z}) + \text{tr}(A \cdot \text{Cov}(\mathbf{z}))$$

$$\begin{aligned} E(\mathbf{z}'A\mathbf{z}) &= E[(\mathbf{z} - E(\mathbf{z}))' + E(\mathbf{z})]'A[(\mathbf{z} - E(\mathbf{z})) + E(\mathbf{z})]) \\ &= E[(\mathbf{z} - E(\mathbf{z}))'A(\mathbf{z} - E(\mathbf{z}))] + E[(\mathbf{z} - E(\mathbf{z}))'A]E(\mathbf{z}) + E[E(\mathbf{z})'A(\mathbf{z} - E(\mathbf{z}))] + E(\mathbf{z})'AE(\mathbf{z}) \\ &= E[\text{tr}((\mathbf{z} - E(\mathbf{z}))A(\mathbf{z} - E(\mathbf{z})))] + E(\mathbf{z})'AE(\mathbf{z}) \\ &= E[\text{tr}(A(\mathbf{z} - E(\mathbf{z}))(\mathbf{z} - E(\mathbf{z}))')] + E(\mathbf{z})'AE(\mathbf{z}) \\ &= \text{tr}(ACov(\mathbf{z})) + E(\mathbf{z})'AE(\mathbf{z}) \end{aligned}$$

§2.2 多元正态分布的定义和基本性质

Def: $U=(U_1, \dots, U_q)'$ 为随机向量, U_1, \dots, U_q 相互独立且同 $N(0,1)$ 分布;

设 μ 为 p 维常数向量, A 为 $p \times q$ 常数矩阵,

则称 $X=AU+\mu$ 的分布为 p 元正态分布, 或称 X 为 p 维正态随机向量, 记为 $X \sim N_p(\mu, AA')$

Thm: 设 $U=(U_1, \dots, U_q)'$ 为随机向量, U_1, \dots, U_q 相互独立且同 $N(0,1)$ 分布;

令 $X=AU+\mu$, 则 X 的特征函数: $\phi_X(t) = \exp[it'\mu - \frac{1}{2}t'AA't]$

$$E(e^{it'X}) = E(e^{it'(AU+\mu)}) = e^{it'\mu} E(e^{i(A't)U})$$

$$\phi_X(t) = E(e^{it'X}) = \exp[it'\mu - \frac{1}{2}t'AA't]$$

$$= e^{it'\mu} E(\prod_{j=1}^q e^{i s_j U_j}) = e^{it'\mu} \prod_{j=1}^q E(e^{i s_j U_j}) = e^{it'\mu} \prod_{j=1}^q e^{i s_j \cdot 0 - \frac{1}{2} s_j^2} = \exp[it'\mu - \frac{1}{2} s' s] = \exp[it'\mu - \frac{1}{2} t' A' A t]$$

Def: 若 p 维随机向量 X 的特征函数为 $\phi_X(t) = \exp[it'\mu - \frac{1}{2}t'\Sigma t]$ ($\Sigma \succ 0$)

则称 X 服从 p 元正态分布, 记为 $X \sim N_p(\mu, \Sigma)$

Thm: 设 $X \sim N_p(\mu, \Sigma)$, B 为 $s \times p$ 常数矩阵, d 为 s 维常向量,

令 $Z=BX+d$, 则 $Z \sim N_s(B\mu+d, B\Sigma B')$

即: 正态随机向量的任意线性组合仍服从正态分布

\Rightarrow 设 $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}_{p+r} \sim N_p(\mu, \Sigma)$, 将 μ 写为 $\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}_{p+r}$, Σ 写为 $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{p+r}$

则 $X^{(1)} \sim N_r(\mu^{(1)}, \Sigma_{11})$, $X^{(2)} \sim N_{p+r}(\mu^{(2)}, \Sigma_{22})$

Thm: 若 $X \sim N_p(\mu, \Sigma)$, 则 $E(X) = \mu$, $D(X) = \Sigma$

Thm: 设 $X = (X_1, \dots, X_p)'$ 为 p 维随机向量

则 X 服从 p 元正态分布 \Leftrightarrow 对任一 p 维实向量 a , $\zeta = a'X$ 是一维正态随机变量

Def: 若 p 维随机向量 X 的任意线性组合均服从一元正态分布

则称 X 为 p 维正态随机向量

Thm: 设 $X \sim N_p(\mu, \Sigma)$ 且 $\Sigma \succ 0$, 则 X 的联合密度函数为 $f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)]$

Def: 若 p 维随机向量 $X = (X_1, X_2, \dots, X_p)'$ 的联合密度函数为

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)],$$
 其中 μ 为 p 维实向量, $\Sigma \succ 0$ 时,

则称 X 为 p 维正态随机向量, $X \sim N_p(\mu, \Sigma)$

Thm: $X \sim N_p(\mu, \Sigma)$, $\Sigma \succ 0$, 则 $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(p)$

令 $\zeta = X-\mu$, 则 $\zeta \sim N_p(0, \Sigma)$, $\eta = \Sigma^{-1/2}\zeta \sim N_p(0, I_p)$, $\eta'\eta \sim \chi^2(p)$

§2.3 条件分布和独立性

独立性 Thm: $X^{(1)}$ 与 $X^{(2)}$ 相互独立 $\Leftrightarrow \Sigma_{12} = 0$

$X^{(1)}, \dots, X^{(k)}$ 相互独立 $\Leftrightarrow \Sigma_{ij} = 0$

条件分布 Thm: 设 $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}_{p+r} \sim N_p(\mu, \Sigma)$ ($\Sigma \succ 0$),

则当 $X^{(2)}$ 给定, $X^{(1)}$ 的条件分布为 $(X^{(1)} | X^{(2)}) \sim N_r(\mu_{12}, \Sigma_{11.2})$

$$E(X^{(1)} | X^{(2)}) = \mu_{12} = \mu^{(1)} + \Sigma_{12} \Sigma_{22}^{-1} (X^{(2)} - \mu^{(2)}), \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \stackrel{\text{条件期望}}{=} (\delta_{ij-r+1, \dots, p})_{r \times r} \quad (i, j=1, \dots, r)$$

回归系数 B

$$\text{偏相关系数: } r_{ij, r+1, \dots, p} = \frac{\delta_{ij-r+1, \dots, p}}{\sqrt{\delta_{ii-r+1, \dots, p} \delta_{jj-r+1, \dots, p}}}$$

$$\text{全相关系数: } R = \left(\frac{\Sigma_{y2} \Sigma_{22}^{-1} \Sigma_{2y}}{\delta_{yy}} \right)^{1/2}$$

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix}_p \sim N_{p+1} \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \delta_{yy} \end{pmatrix} \right)$$

§2.4 随机阵的正态分布

Def: p 元正态总体: $\mathcal{X} = (\mathcal{X}_{ij})_{n \times p} = (\mathcal{X}_{(1)}, \mathcal{X}_{(2)}, \dots, \mathcal{X}_{(n)})_{n \times p}$ $\mathcal{X}_{(i)} = (x_{i1}, \dots, x_{ip})'$

$$\text{Vec}(\mathcal{X}') = \begin{pmatrix} \mathcal{X}_{(1)} \\ \mathcal{X}_{(2)} \\ \vdots \\ \mathcal{X}_{(n)} \end{pmatrix}_{np \times 1} \sim N_{np} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \begin{pmatrix} \Sigma & & \\ & \Sigma & \\ & & \ddots \\ & & & \Sigma \end{pmatrix} \right) = N_{np} (1_n \otimes \mu, 1_n \otimes \Sigma)$$

矩阵正态分布, 记作 $\mathcal{X} \sim N_{n \times p}(\mu, 1_n \otimes \Sigma)$, $\text{Vec}(1_n \otimes \mu) = 1_n \otimes \mu$

Thm: $\mathcal{X} \sim N_{n \times p}(\mu, 1_n \otimes \Sigma)$, $Z = A\mathcal{X}B' + D$, 则 $Z \sim N(AMB' + D, (AA') \otimes (BZB'))$

§2.5 多元正态分布的参数估计

μ, Σ 的最大似然估计 相合性

样本均值向量: $\bar{\mathcal{X}} = \frac{1}{n} \sum_{i=1}^n \mathcal{X}_{(i)} = (\bar{x}_1, \dots, \bar{x}_p)' = \frac{1}{n} \mathcal{X}' 1_n$

样本离差阵: $A = \sum_{i=1}^n (\mathcal{X}_{(i)} - \bar{\mathcal{X}})(\mathcal{X}_{(i)} - \bar{\mathcal{X}})' = \mathcal{X}' \mathcal{X} - n \bar{\mathcal{X}} \bar{\mathcal{X}}' = \mathcal{X}' [1_n - \frac{1}{n} 1_n 1_n'] \mathcal{X} \triangleq (a_{ij})_{p \times p}$

样本协方差阵: $S = \frac{1}{n-1} A$

样本相关阵: $R = (r_{ij})_{p \times p}$, $r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}} = \frac{a_{ij}}{\sqrt{a_{ii} a_{jj}}}$

$$\begin{aligned} (*) &: \sum_{i=1}^n (\mathcal{X}_{(i)} \mathcal{X}_{(i)}' - \mathcal{X}_{(i)} \bar{\mathcal{X}}' - \bar{\mathcal{X}} \mathcal{X}_{(i)}' + \bar{\mathcal{X}} \bar{\mathcal{X}}') & \bar{\mathcal{X}} &= \frac{1}{n} \sum_{i=1}^n \mathcal{X}_{(i)} = \frac{\mathcal{X}' 1_n}{n} \\ &= \sum_{i=1}^n \mathcal{X}_{(i)} \mathcal{X}_{(i)}' - \frac{(\sum_{i=1}^n \mathcal{X}_{(i)}) \bar{\mathcal{X}}'}{n} - \bar{\mathcal{X}} (\sum_{i=1}^n \mathcal{X}_{(i)}') + n \bar{\mathcal{X}} \bar{\mathcal{X}}' & \bar{\mathcal{X}} &= \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix}_{1 \times p} \\ &= \sum_{i=1}^n \mathcal{X}_{(i)} \mathcal{X}_{(i)}' - n \bar{\mathcal{X}} \bar{\mathcal{X}}' = \mathcal{X}' \mathcal{X} - n \bar{\mathcal{X}} \bar{\mathcal{X}}' \end{aligned}$$

参数估计

$P(\mathcal{X}_1 = x_1, \dots, \mathcal{X}_n = x_n)$

$\rightarrow P(x_1 \leq \mathcal{X}_1 \leq x_1 + dx_1, \dots, x_n \leq \mathcal{X}_n \leq x_n + dx_n)$

$$= \prod_{i=1}^n P(x_i \leq \mathcal{X}_i \leq x_i + dx_i)$$

$$= \prod_{i=1}^n \int_{x_i}^{x_i + dx_i} f(x_i) dx_i = \prod_{i=1}^n f(x_i) dx_i = \left(\prod_{i=1}^n f(x_i) \right) \left(\prod_{i=1}^n dx_i \right) \rightarrow \prod_{i=1}^n f(x_i; \theta)$$

似然函数 $L(\mu, \Sigma)$

1. 似然函数 $L(\mu, \Sigma)$
把随机数据 X 按行拉直后形成的 np 维列向量 $\text{Vec}(X')$ 的联合密度函数看成未知参数 μ, Σ 的函数, 并称为样本 $X_{(i)} (i=1, \dots, n)$ 的似然函数, 记为 $L(\mu, \Sigma)$;
$$L(\mu, \Sigma) = \prod_{i=1}^n \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x_{(i)} - \mu)' \Sigma^{-1} (x_{(i)} - \mu) \right]$$
$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (x_{(i)} - \mu)' \Sigma^{-1} (x_{(i)} - \mu) \right]$$
$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \text{tr}((x_{(i)} - \mu)' \Sigma^{-1} (x_{(i)} - \mu)) \right]$$
$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \text{tr}(\Sigma^{-1} (x_{(i)} - \mu) (x_{(i)} - \mu)') \right]$$
$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left[\text{tr} \left(-\frac{1}{2} \Sigma^{-1} \sum_{i=1}^n (x_{(i)} - \mu) (x_{(i)} - \mu)' \right) \right]$$
$$\frac{dL}{d\Sigma} = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left[-\frac{1}{2} \Sigma^{-1} \sum_{i=1}^n (x_{(i)} - \mu) (x_{(i)} - \mu)' \right] \cdot \left(-\frac{n}{2} \Sigma^{-1} \right)$$

其中 $\sum_{i=1}^n (x_{(i)} - \mu) (x_{(i)} - \mu)'$
$$= \sum_{i=1}^n (x_{(i)} - X + X - \mu) (x_{(i)} - X + X - \mu)'$$
$$= \sum_{i=1}^n (x_{(i)} - X) (x_{(i)} - X)' + n(X - \mu) (X - \mu)'$$
$$= A + n(X - \mu) (X - \mu)'$$

由于 $\ln x$ 是 x 的单调函数, $L(\mu, \Sigma)$ 与 $\ln L(\mu, \Sigma)$ 有相同的最大值点, 以下只须讨论 $\ln L(\mu, \Sigma)$ 的最大值问题.

$$\begin{aligned} \ln L(\mu, \Sigma) &= -\frac{np}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| \\ &\quad - \frac{1}{2} \text{tr} \left[\Sigma^{-1} \sum_{i=1}^n (x_{(i)} - \mu) (x_{(i)} - \mu)' \right] \\ &= C - \frac{1}{2} \text{tr}[\Sigma^{-1} A + n \Sigma^{-1} (X - \mu) (X - \mu)'] \\ &= C - \frac{1}{2} \text{tr}(\Sigma^{-1} A) - \frac{n}{2} [(X - \mu)' \Sigma^{-1} (X - \mu)] \\ &\leq C - \frac{1}{2} \text{tr}(\Sigma^{-1} A). \end{aligned}$$

以上不等式仅当 $\mu = X$ 时等号成立, 即对于固定的 $\Sigma > 0$, 有 $\ln L(X, \Sigma) = \max_{\mu} \ln L(\mu, \Sigma)$.

$$\Sigma: \max_{\Sigma > 0} L(\mu, \Sigma) = \max_{\Sigma > 0} L(\bar{X}, \Sigma)$$

$$\begin{aligned} L(\bar{X}, \Sigma) &= (2\pi)^{-np/2} |\Sigma|^{-n/2} e^{-\frac{1}{2} \text{tr}(A \Sigma^{-1})} \\ &= (2\pi)^{-np/2} |A|^{-n/2} |\Sigma^{-1}|^{n/2} e^{-\frac{1}{2} \text{tr}(A \Sigma^{-1})} \\ &= |A \Sigma^{-1}|^{-n/2} \exp \left\{ \frac{n}{2} \text{tr}(\Sigma^{-1} A) \right\} \end{aligned}$$

$$\Sigma^{-1} A \Sigma^{-1} = nI \text{ 时}$$
$$\Sigma = \frac{A}{n} \text{ 时, 取 max}$$

引理: A 为正定阵, $f(A) = |A|^{-n/2} e^{-\frac{1}{2} \text{tr}(A)}$
 $f(A)$ 在 $A = nI$ 时达到 max

△看课本

