一、多元正态分布

(一). 密度函数

p维随机向量 $X \in \mathbb{R}^p$ 服从多元正态分布，当且仅当其联合密度函数为：

$$f(X) = (2\pi)^{-\frac{p}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right\}$$

记作 $X \sim N_p(\mu, \Sigma)$ 其中 $\mu \in \mathbb{R}^p$ $\Sigma \in \mathbb{R}^{p\times p}$ $(\Sigma > 0)$

证明积分为 1：

记 $f(x_1, x_2, \cdots, x_p) = K \cdot \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$

由 Σ (Σ^{-1}) 的正定性 $\Rightarrow$ $(x-\mu)^T \Sigma^{-1} (x-\mu) \geq 0$

故 $f(x_1, x_2, \cdots, x_p) \leq K \cdot \exp(0) = K$ 有界

记 $K^* = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\} dx_1 dx_2 \cdots dx_p$

【引理】 对于正定阵 Σ，一定存在非奇异阵 A，使得

$$A^T \Sigma A = I$$

回到原题：令 $x-\mu = Ay$ $A \in \mathbb{R}^{p\times p}$ $y \in \mathbb{R}^p$ 且 $A^T \Sigma^{-1} A = I$

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

雅克比行列式 $|J| = |A|$

于是 $K^* = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^T y} \cdot \bmod |A| \cdot dy$ （mod 表示绝对值）

$$= \bmod |A| \cdot \prod_{i=1}^{p} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2} dy_i$$

$$= \bmod |A| \cdot \prod_{i=1}^{p} \sqrt{2\pi}$$ （欧拉积分）

而 $A^T \Sigma^{-1} A = I$ 故 $|A^T||\Sigma^{-1}||A| = |I| = 1$ $\Rightarrow$ $|A|^2 = |\Sigma|$

故 $\bmod |A| = |\Sigma|^{\frac{1}{2}}$ 代入可证 $\square$

(二). 一元正态构造多元正态

设随机变量 $Y_1, Y_2, \cdots, Y_p$ iid $\sim N(0,1)$ 则有：

$$Y = [Y_1, Y_2, \cdots, Y_p]^T \sim N_p(0, I)$$

记 $X = \mu + AY$ 其中 A 非奇异 $A \in \mathbb{R}^{p\times p}$ $Y \in \mathbb{R}^p$

于是： $Y = A^{-1}(X-\mu)$ $\quad J(Y \to X) = |A^{-1}|$

那么 X 的分布：

$$g(x) = f(y(x)) \bmod J$$

$$= (2\pi)^{-\frac{p}{2}}\, 1^{-\frac{1}{2}}\, |A^{-1}| \cdot \exp\left\{ -\frac{1}{2}(X-\mu)^T (A^{-1})^T A^{-1}(X-\mu)\right\}$$

记 $A^TA = \Sigma$

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

于是 $X \sim N_p(\mu, A^TA)$.

(三). 二元正态分布

例 4.1 （二元正态密度）

根据参数 $\mu_1 = E(X_1), \mu_2 = E(X_2), \sigma_{11} = \mathrm{Var}(X_1), \sigma_{22} = \mathrm{Var}(X_2)$ 和 $\rho_{12} = \sigma_{12}/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}) = \mathrm{Corr}(X_1, X_2)$，计算 $p=2$ 元正态密度.

利用结论 2A.8，我们求得协方差矩阵

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

的逆是

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

由 $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$，引入相关系数 ρ_{12}，我们得到 $\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$，且距离平方变为 $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$

$$= [x_1-\mu_1, x_2-\mu_2]\frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}\begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$= \frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}$$

$$= \frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right] \tag{4-5}$$

最后表达式是依据标准化值 $(x_1-\mu_1)/\sqrt{\sigma_{11}}$ 和 $(x_2-\mu_2)/\sqrt{\sigma_{22}}$ 写出的.

接下来，由于 $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$，我们可以替换式(4-4)中的 $\boldsymbol{\Sigma}^{-1}$ 和 $|\boldsymbol{\Sigma}|$ 而得到包含各个参数 $\mu_1, \mu_2, \sigma_{11}, \sigma_{22}$ 和 ρ_{12} 的二元($p=2$)正态密度：

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}}$$

$$\times \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\} \quad (4\text{-}6)$$

(四) 常数密度轮廓线

满足 $(x-\mu)^T\Sigma^{-1}(x-\mu)=c^2$ 的所有 x ：即中心点在 $\mu\in\mathbb{R}^p$ 的椭球的表面。

定理 若 Σ 正定，手是 Σ^{-1} 存在，则

$$\Sigma e=\lambda e \quad \text{意味着} \quad \Sigma^{-1}e=\lambda^{-1}e$$

所以 (λ, e) 是 Σ 的一个特征值－特征向量对，对应于 Σ^{-1} 的特征值－特征向量对为 (λ^{-1}, e)

将 $(x-\mu)$ 视为 e，可总结为：

对 p 维正态分布，常数密度的轮廓线是由 x 确定的椭球面，x 满足

$$(x-\mu)^T\Sigma^{-1}(x-\mu)=c^2$$

它的中心在 $\mu\in\mathbb{R}^p$，而轴为 $\pm c\sqrt{\lambda_i}\,e_i$ 其中 $\Sigma e_i=\lambda_i e_i$

即 (λ_i, e_i) 为 Σ 的一个特征值－特征向量对 $(i=1,2,\cdots,p)$

（e_i 已归一化 $e_i^Te_i=1$）

(五) 多元正态分布的可加性

① 若 $X \in \mathbb{R}^p$ 且 $X \sim N_p(\mu, \Sigma)$ 则多个变量的线性组合：

$$a^T X = \sum_{i=1}^{p} a_i X_i$$

则满足新的正态(一元)分布：

$$a^T X \sim N(a^T\mu,\ a^T \Sigma a)$$

② 更进一步，$X \in \mathbb{R}^p$，而其 q 个线性组合：

$$a_i^T X = \sum_{j=1}^{p} a_{ij} X_j \in \mathbb{R}$$

构成新的随机向量：

$$AX = \begin{bmatrix} a_1^T X \\ a_2^T X \\ \vdots \\ a_q^T X \end{bmatrix} \in \mathbb{R}^q$$

则 $AX \in \mathbb{R}^q$ 满足：

$$AX \sim N_q(A\mu,\ A\Sigma A^T)$$

> 特别地： $X + d \sim N(\mu + d, \Sigma)$

③ X 的所有子集都是正态分布。若记 X 的均值向量为 μ，协方差矩阵为 Σ 划分：

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

其中 $X^{(1)} \in \mathbb{R}^q$ $X^{(2)} \in \mathbb{R}^{p-q}$ ； $\mu_1 \in \mathbb{R}^q$ $\mu_2 \in \mathbb{R}^{p-q}$ ；

$\Sigma_{11} \in R^{q\times q}$ $\Sigma_{22} \in R^{(p-q)\times(p-q)}$ $\Sigma_{12}, \Sigma_{21}^T \in R^{q\times(p-q)}$。

则有：

$$X^{(1)} \sim N_q(\mu_1, \Sigma_{11})$$

④ （边际分布 marginal distribution）已知 X 是 X_1, X_2 的联合分布（joint distribution）服从如下联合正态分布：

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

则有 $X^{(1)}, X^{(2)}$ 的边际分布：

$$X^{(1)} \sim N(\mu_1, \Sigma_{11}) \qquad X^{(2)} \sim N(\mu_2, \Sigma_{22})$$

更进一步，如果 $\Sigma_{12} = \Sigma_{21}^T = 0$，则有 $X^{(1)} \perp X^{(2)}$

⑤ （条件分布 conditional distribution）已知 X 是 X_1, X_2 的联合分布（joint distribution）服从如下联合正态分布：

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

且 Σ_{22} 可逆，则

$$X_1 | X_2 \sim N(\mu_1 + \Sigma_{12}\cdot\Sigma_{22}^{-1}\cdot(X_2 - \mu_2), \quad \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

即给定 X_2 时 X_1 的分布.

Proof: 记 $Z = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ 且记 $A = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 则

$$\begin{bmatrix} X_2 \\ Z \end{bmatrix} = \begin{bmatrix} O_{(n-m)\times m} & I_{n-m} \\ I_m & -\Sigma_{12}\Sigma_{22}^{-1} \end{bmatrix} \cdot X \sim N\left(\begin{bmatrix} \mu_2 \\ \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{22} & O \\ O & A \end{bmatrix}\right)$$

由 Theorem 3 副对角矩阵为 0 故 $Z \perp X_2$

且 $Z \sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\ A)$

故 $Z|X_2 = Z \sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\ A)$

$\therefore\ X_1|X_2 = Z + \Sigma_{12}\Sigma_{22}^{-1}X_2 \mid X_2$

而 X_2 给定，$\Sigma_{12}\Sigma_{22}^{-1}X_2$ 为常数

$\therefore\ X_1|X_2 \sim N(\mu_1 + \Sigma_{12}\cdot\Sigma_{22}^{-1}\cdot(X_2-\mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

二、多元正态分布 二次型

(一). 正态随机向量的二次型

$X \in \mathbb{R}^n$ 且 $X \sim N(\mu, \Sigma)$ 且 Σ 可逆，则有

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

(2) $P\left((x-\mu)^T \cdot \Sigma^{-1}(x-\mu) \leq \chi^2_\alpha(n)\right) = 1-\alpha$

三、样本

$X_1, X_2, \cdots, X_n \in \mathbb{R}^p$ 是 n 个样本（非随机）来自总体 $N_p(\mu, \Sigma)$ 则考虑：

$$V_1 = \sum_{i=1}^{n} c_i X_i \in \mathbb{R}^p$$

$$V_2 = \sum_{i=1}^{n} b_i X_i \in \mathbb{R}^p$$

则有：$\mathrm{Cov}(V_1, V_2) =$

$$\begin{bmatrix} c^T c \cdot \Sigma & b^T c \cdot \Sigma \\ b^T c \cdot \Sigma & b^T b \cdot \Sigma \end{bmatrix}$$

Proof: 对 $\forall X_i \in \mathbb{R}^p$ 其分量为 $X_{i1}\ X_{i2} \cdots X_{ip} \in \mathbb{R}$

记

$$X = [X_{11}\ X_{12} \cdots X_{1p}\ \ X_{21}\ X_{22} \cdots X_{2p} \cdots X_{n1}\ X_{n2} \cdots X_{np}]^T$$

则有 $X \in \mathbb{R}^{np}$ (np维向量)，而 $\forall X_{ij} \sim N(\mu_j, \sigma_j^2)$ $j \in \{1,2,\cdots,p\}$

有 $X \sim N_{np}(\mu_X, \Sigma_X)$ 其中：

$$\mu_X = [\mu_1^T\ \mu_2^T \cdots \mu_n^T]^T \in \mathbb{R}^{np} \quad 其中 \quad \mu_i \in \mathbb{R}^p$$

$$\Sigma_X = \begin{bmatrix} \Sigma & & & \\ & \Sigma & & \\ & & \ddots & \\ & & & \Sigma \end{bmatrix}_{np \times np} \quad 其中 \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & & \sigma_p^2 \end{bmatrix} \in \mathbb{R}^{p \times p}$$

设 $A = \begin{bmatrix} c_1 I & c_2 I & \cdots & c_n I \\ b_1 I & b_2 I & \cdots & b_n I \end{bmatrix} \in \mathbb{R}^{2p \times np}$

$I \in \mathbb{R}^{p \times p}$ 就有

$$AX = \begin{bmatrix} \sum_{i=1}^n c_i X_i \\ \sum_{i=1}^n b_i X_i \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \in \mathbb{R}^{2p}$$

则有

$$AX \sim N_{np}(A\mu_X, A\Sigma_X A^T)$$

关于 $A\Sigma_X A^T$ 第一个分块对角线：

$$[c_1 I\ \ c_2 I \cdots c_n I]\, \Sigma\, [c_1 I\ \ c_2 I \cdots c_n I]^T$$

$$= \sum_{i=1}^n c_i^2 \cdot \Sigma = c^T c\, \Sigma \quad 为 \ \mathrm{Cov}(V_1, V_1)$$

非对角线：

$$[c_1 I\ \ c_2 I \cdots c_n I] \cdot \Sigma \cdot [b_1 I\ \ b_2 I \cdots b_n I]^T$$

$$= c^T b \cdot \Sigma$$

故 $Cov(V_1, V_2) = c^T b \cdot \Sigma$ $\square$

性质： $c^T b = 0$ 时 $V_1 \perp V_2$

Proof： 由 $AX = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \sim N_{np}(A\mu_x, A\Sigma_x A^T)$

而 $Cov(V_1, V_2) = c^T b \Sigma = 0$

$\Rightarrow V_1 \perp V_2$

四. 其他

(一). 拉直运算

将矩阵平展为一个长向量，通过它建立了矩阵和向量之间的联系。令 A 为 $n \times p$ 维的矩阵，用 A 的列向量 $a_1, a_2, \cdots, a_p \in R^n$ 组成长为 np 的长向量：

$$Vec(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \in R^{np}$$

同理，可将 A 的行向量 $a_{(1)}, a_{(2)}, \cdots, a_{(n)} \in R^p$ 组成长为 np 的长向量：

$$Vec(A^T) = \begin{bmatrix} a_{(1)} \\ a_{(2)} \\ \vdots \\ a_{(n)} \end{bmatrix} \in R^{np}$$

(二). Kronecker 积

设 $A=(a_{ij})\in\mathbb{R}^{n\times p}$ 和 $B=(b_{ij})\in\mathbb{R}^{m\times q}$ 两个矩阵的Kronecker 积为：

$$A\otimes B=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1p}B \\ a_{21}B & a_{22}B & \cdots & a_{2p}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{np}B \end{bmatrix}\in\mathbb{R}^{mn\times pq}$$

(三). 矩阵正态分布

设 $Y=(Y_{ij})\in\mathbb{R}^{m\times q}$ 为一随机矩阵，即每个 Y_{ij} 为一个随机变量。Y_{ij} 之间相互独立，同分布于 $N(0,1)$

记矩阵 $M\in\mathbb{R}^{n\times p}$ $A\in\mathbb{R}^{p\times q}$ $B\in\mathbb{R}^{n\times m}$ 为常矩阵，则令

$$X_{n\times p}=M_{n\times p}+B_{n\times m}\cdot Y_{m\times q}\cdot A^T_{p\times q}$$

则称 X 服从矩阵正态分布为 $N_{n\times p}(M,\ (BB^T)\otimes(AA^T))$

推论：设 $X\sim N_{n\times p}(M,\ W\otimes V)$ 其中 $W=BB^T$ $V=AA^T$，记

$$Z_{m\times q}=C_{m\times n}X_{n\times p}D^T_{q\times p}+U_{m\times q}$$

则有

$$Z\sim N_{m\times q}(CMD^T+U,\ (CWC^T)\otimes(DVD^T))$$

示例：对 $X\in\mathbb{R}^{n\times p}$ 随机矩阵，其中 X 的行向量 $X_1,X_2,\cdots,X_n\in\mathbb{R}^p$ 满足 $X_i\sim N_p(\mu,\Sigma)$。由性质易知：$\mu\in\mathbb{R}^p$ $\Sigma\in\mathbb{R}^{p\times p}$

$$X_i = \mu + A Y_i$$ 其中 $AA^T = \Sigma$ $A \in \mathbb{R}^{p \times p}$

现在考虑 Y 的矩阵分布: 对 $\forall Y_i \sim N_p(0, I_p)$

$$Y \sim N_{n\times p}(0, I_n \otimes I_p)$$

由

$$X = \begin{bmatrix} \mu^T \\ \mu^T \\ \vdots \\ \mu^T \end{bmatrix} + \begin{bmatrix} Y_1^T \\ Y_2^T \\ \vdots \\ Y_n^T \end{bmatrix} \cdot A^T \sim N_{n\times p}(\Lambda_{n\times p}, I_n \otimes \Sigma)$$

$$\Lambda = \begin{bmatrix} \mu^T \\ \mu^T \\ \vdots \\ \mu^T \end{bmatrix} = 1_{n\times 1} \cdot \mu^T \in \mathbb{R}^{n\times p} \quad AA^T = \Sigma \in \mathbb{R}^{p\times p}$$