第一章 多元分布

1.0 一些预备知识

- **随机变量**

　　设变量*X*在不同的条件下，由于偶然因素影响可能取各种不同的值，具有不确定性和随机性，但这些取值落在某个范围的概率是一定的，则称为**随机变量**。

- 如果随机变量*X*只能取有限个或可数个值，并且取这些不同值的概率是确定的，则称*X*为**离散型随机变量**。

设*X*的取值为 $x_1, x_2, \cdots$，相应的概率为 $p_i = P\{X = x_i\},\ \ i = 1, 2, \cdots$，则有：

1) $p_i \geq 0,\ i = 1, 2, \cdots;$

2) $\sum\limits_{i=1}^{\infty} p_i = 1.$

- 如果随机变量*X*取值充满某个区间，并且*X*的值落在任何一个子区间的概率是确定的，则称*X*为**连续型随机变量。**

对一个连续型的随机变量*X*，如果存在一个非负可积函数$f(x)$使得

$$P\{a < X \leq b\} = \int_a^b f(x)\,dx,$$

对一切 $-\infty < a < b < \infty$ 成立，则称$f(x)$ 为*X*的**概率密度函数**。

- **分布函数**：设*X*为随机变量，令

$$F(x) = P\{X \leq x\}, -\infty < x < \infty,$$

则称 $F(x)$ 为*X*的**分布函数**。

- 当*X*为离散型随机变量时，

$$F(x) = \sum_{i:\ x_i \le x} p_i,$$

当*X*为连续型随机变量时，

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

- 随机变量的期望、矩、方差:
- 期望：

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} tf(t)\,dt, & 若X连续; \\ \sum_i x_i P\{X = x_i\}, & 若X离散, \end{cases}$$

- 矩：

$$E(X^k) = \begin{cases} \int_{-\infty}^{\infty} t^k f(t)\,dt, & 若X连续; \\ \sum_i x_i^k P\{X = x_i\}, & 若X离散, \end{cases}$$

- 方差： $Var(X) = E[(X - E(X))^2] = E(X^2) - (E(X))^2.$

- 设 $X_1, X_2, \cdots, X_p$ 为*p*个随机变量，它们组成的向量 $X=(X_1, X_2, \cdots, X_p)'$ 称为**随机向量**。

- 随机向量的**联合分布函数** F 定义为

$$
\begin{aligned}
F(x_1, x_2, \cdots, x_p) &= P\{X_1 \leq x_1, X_2 \leq X_2, \cdots, X_p \leq x_p\} \\
&= P\{X \leq x\},
\end{aligned}
$$

其中 $x=(x_1, x_2, \cdots, x_p)'$。

- **联合概率密度函数：**如果存在非负函数 $f(x_1, x_2, \cdots, x_p)$，使得对任意 $x_1, x_2, \cdots, x_p$ 有

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

则称 $f(x_1, x_2, \cdots, x_p)$为 X 的联合概率密度函数。

- X 的 q 个 $(q < p)$ 分量 $X^{(1)} = (X_1, X_2, \cdots, X_q)'$ 的分布称为**边缘分布**，即

$$\begin{aligned} P\{X^{(1)} \le u\} &= P\{X_1 \le u_1, \ldots, X_q \le u_q\} \\ &= F(u_1, \cdots, u_q, +\infty, \cdots, +\infty). \end{aligned}$$

- $X^{(1)}$ 的**边缘概率密度函数**定义为：$g(u) = \int_{R^{p-q}} f(u, v)\, dv.$

- 若 $X = (X^{(1)\prime}, X^{(2)\prime})'$ 有概率密度函数 $f(x) = f(x^{(1)}, x^{(2)})$，$X^{(1)}$ 有密度函数 $g(u)$，则 $X^{(2)}$ 在给定 $X^{(1)} = x^{(1)}$ 的**条件密度**为

$$f(x^{(2)}|X^{(1)} = x^{(1)}) = \frac{f(x^{(1)}, x^{(2)})}{g(x^{(1)})}.$$

- $X_1, X_2, \cdots, X_p$ 相互独立，当且仅当

$$F(x_1, \ldots, x_p) = \prod_{i=1}^{p} F_i(x_i), \forall (x_1, \ldots, x_p)' \in R^p,$$

其中 F_i 是 X_i 的边缘分布函数，$1 \le i \le p$ 。

- 多元随机变量（随机向量）矩的性质:
- 期望 $E(X)=(E(X_1),E(X_2),\cdots,E(X_p))'$
- 协方差
$$\begin{aligned} Cov(X) &= \big(E[(X_i-E(X_i))(X_j-E(X_j))]\big)_{p\times p} \\ &= (E[(X-E(X))(X-E(X))']). \end{aligned}$$
- 若 $X_{p\times 1},Y_{q\times 1}$ 为随机向量，则它们的协方差为

$$\begin{aligned} Cov(X,Y) &= \big(E[(X_i-E(X_i))(Y_j-E(Y_j))]\big)_{p\times q} \\ &= (E[(X-E(X))(Y-E(Y))']). \end{aligned}$$

- 其它一些重要的运算

$$
\begin{aligned}
&E(tr(AXB)) = tr(A(E(X))B),\\
&Cov(AX) = ACov(X)A';\\
&E(X'AX) = (E(X))'A(E(X)) + tr(ACov(X));\\
&Cov(AX, BY) = ACov(X, Y)B'.
\end{aligned}
$$

- 多元特征函数

随机向量 $X = (X_1, \cdots, X_p)'$ 的特征函数为:

$$\phi(t) = \phi(t_1, \cdots, t_p) = E[e^{\mathrm{i}(t_1X_1+\cdots+t_pX_p)}] = E[e^{\{\mathrm{i}t'X\}}],$$

其中，$t = (t_1, \cdots, t_p)' \in R^p$， i 是虚数单位， $\mathrm{i}^2 = -1$.

特征函数与概率分布函数是一一对应的.

- **特征函数的一些性质：**

性质1：对正整数 $k_1, \cdots, k_p$, 如果 $E(X_1^{k_1} \cdots X_p^{k_p})$ 存在，则

$$E(X_1^{k_1} \cdots X_p^{k_p}) = (-\mathrm{i})^{k_1+\cdots+k_p} \left[\frac{\partial^{k_1+\cdots+k_p} \phi(t_1, \cdots, t_p)}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} \right]_{t_1=\cdots=t_p=0}.$$

特别地，若期望 $E(X_j)$ 存在，则

$$E(X_j) = (-\mathrm{i}) \left[\frac{\partial \phi(t_1, \cdots, t_p)}{\partial t_j} \right]_{t_1=\cdots=t_p=0};$$

若二阶矩 $E(X_j^2)$存在，则

$$E(X_j^2) = -\left[\frac{\partial^2 \phi(t_1, \cdots, t_p)}{\partial t_j^2}\right]_{t_1=\cdots=t_p=0};$$

若二阶混合矩 $E(X_j X_k)$ 存在，则

$$E(X_j X_k) = -\left[\frac{\partial^2 \phi(t_1, \cdots, t_p)}{\partial t_j \partial t_k}\right]_{t_1=\cdots=t_p=0}.$$

性质2： 对 $0 < k < p$，分量 $X^{(1)} = (X_1, \cdots, X_k)'$ 的特征函数为 $\phi(t_1, \cdots, t_k, 0, \cdots, 0)$.

性质3：记 $X_1, \cdots, X_p$ 的边缘特征函数分别为 $\phi_1(t_1), \cdots, \phi_p(t_p)$, 记 $X = (X_1, \cdots, X_p)'$ 的特征函数为 $\phi(t_1, \cdots, t_p)$, 则 $X_1, \cdots, X_p$ 相互独立的充分必要条件是：

$$\phi(t_1, \cdots, t_p) = \phi_1(t_1) \cdots \phi_p(t_p).$$

性质4：设 p 维随机向量 $Y_1, \cdots, Y_m$ 的特征函数分别为 $\phi^{(1)}(t), \cdots, \phi^{(m)}(t)$, 如果 $Y_1, \cdots, Y_m$ 相互独立，则随机向量和 $Y_1 + \cdots + Y_m$ 的特征函数为

$$\phi(t) = \phi^{(1)}(t) \cdots \phi^{(p)}(t).$$

- 分块矩阵的运算

假设矩阵$\mathbf{A}$可以剖分为

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

其中 $\mathbf{A}_{11}$ 是非退化的方阵.

记$\mathbf{A}_{2|1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, 则有

$$|\mathbf{A}| = |\mathbf{A}_{11}| \cdot |\mathbf{A}_{2|1}|,$$

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{2|1}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{pmatrix}.$$

如果记

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix},$$

则有

$$\begin{aligned}
\mathbf{B}_{11} &= (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}, \\
\mathbf{B}_{22} &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}, \\
\mathbf{B}_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}, \\
\mathbf{B}_{21} &= -(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}.
\end{aligned}$$

1.1 一元正态分布

1. 1. 1 一元正态分布密度

定义1: 若随机变量X的概率密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

其中, $-\infty < x, \mu < \infty$, $\sigma^2 > 0$, 则称随机变量X服从正态分布.

记为 $X \overset{d}{\sim} N(\mu, \sigma^2)$, 其中$\mu$是均值,$\sigma^2$是方差.

正态分布是最常见的连续分布.

1.2 多元正态分布

1.2.1 多元正态分布密度

多元正态分布重要性:

1）许多多元统计技术基于多元正态假设；
2）正态分布数学上易于处理，形式简洁；
3）众多实际问题中，总体分布是正态分布或近似正态分布；
4）即使总体分布非正态，许多统计量的分布渐近为正态分布。

定义2: 称 p 元随机向量 X 服从参数为 μ 和 Σ 的多元正态分布，如果其概率密度函数为

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

其中 $\mu \in R^p$, Σ 为 p 阶正定矩阵. 记为 $X \overset{d}{\sim} N_p(\mu, \Sigma)$.

- 称 $N_p(0, I_p)$ 为p元**标准正态分布**，其中 I_p 是 $p \times p$ 的单位矩阵.

- **定理1.** 设p元随机向量 $X = \mu + AY$，其中 $\mu \in R^k, A$ 为 $k \times p$ 的行满秩矩阵, $k \leq p$，随机向量 $Y \overset{d}{\sim} N_p(0, I_p)$，则

$$X \overset{d}{\sim} N_k(\mu, \Sigma),$$

  其中 $\Sigma = AA' > 0$.

  证明： 利用特征函数.

  如何产生 $N_p(\mu, \Sigma)$ 的(伪)随机数?

1.2.2 多元正态分布的性质

性质1： 密度函数

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

性质2：（特征函数）$X \overset{d}{\sim} N(\mu, \Sigma)$，则

$$E(\exp\{\mathrm{i}t'X\}) = \exp\left\{\mathrm{i}t'\mu - \frac{t'\Sigma t}{2}\right\}.$$

性质3： 若 $X \overset{d}{\sim} N_p(\mu, \Sigma)$, 则 $E(X) = \mu$, $Cov(X) = \Sigma$.

性质4： （线性变换）若 $X \overset{d}{\sim} N_p(\mu, \Sigma)$, $Y = \eta + AX$, $\eta \in R^k$, A 是 $k \times p$ 的矩阵, 则

$$Y \overset{d}{\sim} N_k(\eta + A\mu, A\Sigma A').$$

性质5：设 $X_1, \cdots, X_k$ 相互独立, $X_i \overset{d}{\sim} N_p(\mu_i, \Sigma_i)$, $1 \leq i \leq k$, 则

$$\sum_{i=1}^{k} a_i X_i \overset{d}{\sim} N_p(\sum_{i=1}^{k} a_i \mu_i, \sum_{i=1}^{k} a_i^2 \Sigma_i).$$

性质6： 若 $X \overset{d}{\sim} N_p(\mu, \Sigma)$, $\Sigma > 0$, 则

$$(X - \mu)' \Sigma^{-1} (X - \mu) \overset{d}{\sim} \chi_p^2,$$

其中 χ_p^2 是自由度为 p 的卡方分布.

性质7：若 $X \overset{d}{\sim} N_p(\mu, \Sigma)$, 有分解

$$X = \begin{pmatrix} X_1^{(q)} \\ X_2^{(p-q)} \end{pmatrix}, \ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

则 $X_1^{(q)} \overset{d}{\sim} N_q(\mu_1, \Sigma_{11}), \ X_2^{(p-q)} \overset{d}{\sim} N_{p-q}(\mu_2, \Sigma_{22})$.

性质8.（分量独立性） 设 $X \overset{d}{\sim} N_p(\mu, \Sigma)$, 有分解

$$X = \begin{pmatrix} X_1^{(q_1)} \\ \vdots \\ X_k^{(q_k)} \end{pmatrix}, \ \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} \\ \vdots & \ddots & \vdots \\ \Sigma_{k1} & \cdots & \Sigma_{kk} \end{pmatrix},$$

则 $X_i^{(q_i)}, X_j^{(q_j)}$ $(1 \le i < j \le k)$ 相互独立的充分必要条件是

$$Cov(X_i^{(q_i)}, X_j^{(q_j)}) = \Sigma_{ij} = 0.$$

性质9.（条件分布）同上假设，

则 $(X_1|X_2=x_2)\overset{d}{\sim}N_q(\mu_{1|2},\Sigma_{1|2})$, 其中

$$\mu_{1|2}=E(X_1|X_2=x_2)=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2),$$
$$\Sigma_{1|2}=Cov(X_1|X_2=x_2)=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

注意： $\Sigma_{1|2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\leq\Sigma_{11}.$

性质10.（变量的独立分解）同上假设，令

$$Y_1=X_1,\ Y_2=X_2-\Sigma_{21}\Sigma_{11}^{-1}X_1;$$
$$Z_2=X_2,\ Z_1=X_1-\Sigma_{12}\Sigma_{22}^{-1}X_2,$$

则 Y_1 与 Y_2 相互独立, Z_1 与 Z_2 相互独立, 且

$$Y_1\overset{d}{\sim}N_q(\mu_1,\Sigma_{11}),\ Y_2\overset{d}{\sim}N_{p-q}(\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1,\Sigma_{2|1});$$
$$Z_2\overset{d}{\sim}N_{p-q}(\mu_2,\Sigma_{22}),\ Z_1\overset{d}{\sim}N_q(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2,\Sigma_{1|2});$$
$$\Sigma_{2|1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

性质10的证明：令

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} I_q & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{p-q} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

不难计算得

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \overset{d}{\sim} N_p\left(\begin{pmatrix} \mu_1 \\ \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{2|1} \end{pmatrix}\right),$$

因此由**性质8**知 Y_1 与 Y_2 相互独立, 同理可证 Z_1 与 Z_2 相互独立. #

有关多元正态分布条件期望和条件标准差的算例：

成年男子上衣的8个人体部位尺寸的均值与标准差

部位	均值	标准差
身高	167.48	6.09
颈椎点高	142.91	5.60
腰围高	100.58	4.44
坐姿颈椎点高	65.61	2.67
颈围	36.83	2.11
胸围	87.53	5.55
后肩横弧	43.24	2.75
臂全长	54.53	3.04

成年男子上衣的8个人体部位尺寸的协方差阵

	身高	颈椎点高	腰围高	坐姿颈椎点高	颈围	胸围	后肩横弧	臂全长
身高	37.115							
颈椎点高	33.069	31.314						
腰围高	24.631	22.624	19.739					
坐姿颈椎点高	12.364	11.506	7.119	7.131				
颈围	2.695	2.593	1.217	1.575	4.437			
胸围	11.155	11.177	6.163	5.334	7.013	30.784		
后肩横弧	7.367	7.075	4.030	3.229	2.084	7.472	7.554	
臂全长	12.597	11.911	9.322	3.573	0.577	4.049	2.340	9.246

由于样本量足够大，我们就假定成年男子上衣的8个人体部位服从一个8维的正态分布，均值和协方差阵如上。

1) 计算给定身高下其它分量的的条件标准差

利用**性质9**中的条件协方差公式计算

	标准差	给定身高的条件标准差
身高	6.09	–
颈椎点高	5.60	1.36
腰围高	4.44	1.84
坐姿颈椎点高	2.67	1.74
颈围	2.11	2.06
胸围	5.55	5.24
后肩横弧	2.75	2.47
臂全长	3.04	2.23

身高对胸围基本无影响

2) 计算给定身高和胸围下的条件标准差

	标准差	给定身高和胸围的条件标准差
身高	6.09	–
颈椎点高	5.60	1.34
腰围高	4.44	1.83
坐姿颈椎点高	2.67	1.71
颈围	2.11	1.68
胸围	5.55	–
后肩横弧	2.75	2.25
臂全长	3.04	2.23

身高和胸围基本能代表上衣尺寸

身高和胸围对后肩横弧的影响不大

3) 计算给定身高和胸围下的条件期望

利用**性质9**中的条件期望公式计算

	给定身高和胸围下的条件期望
颈椎点高	-7.985+0.877×身高 + 0.0451 × 胸围
腰围高	-8.881+0.677×身高 − 0.0452 × 胸围
坐姿颈椎点高	7.623 +0.315×身高 + 0.059 × 胸围
颈围	16.252+0.0047×身高 + 0.226 × 胸围
后肩横弧	2.863 +0.141×身高 + 0.192 × 胸围
臂全长	-2.667+0.337×身高 + 0.0096 × 胸围

可以用身高和胸围来预测其它6个尺寸

1.3 相关系数

1.3.1 相关系数

设随机向量 $X = (X_1, \cdots, X_p)'$, $Cov(X) = \Sigma = (\sigma_{ij})_{p\times p}$,

则 X_i 与 X_j $(1 \le i < j \le p)$ 的相关系数 ρ_{ij} 为

$$\rho_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)}\sqrt{Var(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}.$$

1.3.2 相关矩阵

$$\mathbf{R} = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1p} \\ \vdots & & \vdots \\ \rho_{p1} & \cdots & \rho_{pp} \end{pmatrix} = diag(\sigma_{11}^{-1/2}, \cdots, \sigma_{pp}^{-1/2}) \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} diag(\sigma_{11}^{-1/2}, \cdots, \sigma_{pp}^{-1/2}).$$

相关性： 正（负）相关性的意义

1.3.3 偏相关系数

设 $X \overset{d}{\sim} N_p(\mu, \Sigma), \Sigma > 0$, 有分解

$$X = \begin{pmatrix} X_1^{(q)} \\ X_2^{(p-q)} \end{pmatrix}, \ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

则 $(X_1^{(q)}|X_2^{(p-q)}) \overset{d}{\sim} N_q(\mu_{1|2}, \Sigma_{1|2})$, 其中

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2),$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \overset{\text{记成}}{=} (\sigma_{(ij)|(1|2)})_{q\times q}.$$

在给定 $X_2^{(p-q)}$ 的条件下, X_i 与 X_j $(1 \le i < j \le q)$ 的条件相关系数

$$\rho_{(ij)|(1|2)} = \frac{\sigma_{(ij)|(1|2)}}{\sqrt{\sigma_{(ii)|(1|2)}}\sqrt{\sigma_{(jj)|(1|2)}}}.$$

条件相关系数也称为偏相关系数。

1.3.4 精度矩阵

设随机向量 X, 有 $Cov(X) = \Sigma > 0$, 那么称 $K = \Sigma^{-1}$ 为 X 的精度矩阵.

性质1: 若 $X = (X_1, \cdots, X_p)' \overset{d}{\sim} N_p(\mu, \Sigma)$, $\Sigma > 0$, $K = \Sigma^{-1} = (k_{ij})_{p\times p}$, 则

$$k_{ii} = (Var(X_i|X_{(-i)}))^{-1},\ 1 \le i \le p,$$

其中 $X_{(-i)} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_p)'$, $1 \le i \le p$.

设 $X \overset{d}{\sim} N_p(\mu, \Sigma), \Sigma > 0$, 有如下分解

$$X = \begin{pmatrix} X_1^{(p_1)} \\ X_2^{(p_2)} \\ X_3^{(p_3)} \end{pmatrix},\ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix},\ K = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix}.$$

性质2: 在 X_3 给定的条件下, X_1 与 X_2 相互条件独立的充要条件是 $K_{12}=0.$

证明： 由1.2中的**性质9**，有

$$
\begin{aligned}
\Sigma_{12|3} &\overset{\text{定义}}{=} Cov((X_1,X_2)|X_3) \\
&\overset{\text{性质9}}{=} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} - \begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} \Sigma_{33}^{-1} \begin{pmatrix} \Sigma_{31} & \Sigma_{32} \end{pmatrix} \\
&\overset{\text{矩阵运算}}{=} \begin{pmatrix} \Sigma_{11}-\Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{12}-\Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} \\ \Sigma_{21}-\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{22}-\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32} \end{pmatrix} \\
&\overset{\text{分块矩阵运算}}{=} \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}^{-1}.
\end{aligned}
$$

由1.2中的**性质8**知, X_1 与 X_2 条件独立的充要条件为

$$\Sigma_{12}-\Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32}=0 \Longleftrightarrow K_{12}=0. \qquad \#$$

推广: $K_{13}=0,\ K_{23}=0.$

由于精度矩阵的数值与变量的量纲有关, 作类似于相关系数阵的标准化处理. 令

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & \ddots & \vdots \\ c_{p1} & \cdots & c_{pp} \end{pmatrix} \triangleq diag(k_{11}^{-1/2}, \cdots, k_{pp}^{-1/2}) \begin{pmatrix} k_{11} & \cdots & k_{1p} \\ \vdots & \ddots & \vdots \\ k_{p1} & \cdots & k_{pp} \end{pmatrix} diag(k_{11}^{-1/2}, \cdots, k_{pp}^{-1/2}),$$

C 与量纲无关, 并将之作相应分块处理

$$C = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix}.$$

因此, 给定 X_3 的条件下, X_1 与 X_2 条件独立的充要条件为 $C_{12} = 0$.

1.3.5 偏相关系数的应用:

由**性质**2和**性质**3, 可以通过偏相关系数判别多元正态随机向量分量之间的条件独立性。

偏相关系数是**图模型**和**因果推断**中的重要统计量。

1.3.6. 有关相关系数和偏相关系数矩阵的算例:

某种水泥在凝固时释放的热量 y 与水泥中下列4种化学成分有关:

x_1: 为 $3CaO{\cdot}Al_2O_3$ 的质量分数(%);
x_2: 为 $3CaO{\cdot}SiO_2$ 的质量分数(%);
x_3: 为 $4CaO{\cdot}Al_2O_3 \cdot Fe_2O_3$ 的质量分数(%);
x_4: 为 $2CaO{\cdot}SiO_2$ 的质量分数(%).

问题：如何利用4种化学成分的观测值预测水泥凝固时释放的热量?

13组实验数据为

样本编号	x_1	x_2	x_3	x_4	y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

1）计算样本协方差阵

假设有n个观测样本$X_1, \cdots, X_n$, 则样本均值和样本协方差阵为

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n},$$

$$V_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)'}{n-1}.$$

利用上面的公式可以计算得该组数据的样本均值和协方差阵

$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$	$\bar{y}$
7.462	48.154	11.769	30.000	95.423

	x_1	x_2	x_3	x_4	y
x_1	31.941				
x_2	19.314	223.515			
x_3	-28.663	-12.811	37.870		
x_4	-22.308	-233.923	2.923	258.615	
y	56.689	176.381	-47.556	-190.900	208.905

2） **计算给定(x_1, x_2, x_3, x_4)下y的条件期望**

利用正态分布的条件期望公式得

$$
\begin{aligned}
\hat{E}(y|x_1, x_2, x_3, x_4) \quad = \quad & 95.423 \\
& -2.069(x_1 - 7.462) \\
& -2.836(x_2 - 48.154) \\
& -3.515(x_3 - 11.769) \\
& -3.442(x_4 - 30.000).
\end{aligned}
$$

注: 都是基于样本均值和样本协差阵得到.

3）计算样本相关（系数）阵

利用1.3.2中的定义计算得该组数据的样本相关系数矩阵为

	x_1	x_2	x_3	x_4	y
x_1	1.000				
x_2	0.229	1.000			
x_3	-0.824	-0.139	1.000		
x_4	-0.245	-0.973	0.030	1.000	
y	0.731	0.816	-0.535	-0.821	1.000

利用正态分布的性质：x_3和x_4很有可能相互独立

4）计算样本精度矩阵

对样本协方差阵求逆得

	x_1	x_2	x_3	x_4	y
x_1	1.859				
x_2	1.329	1.209			
x_3	1.247	1.156	1.240		
x_4	1.037	1.093	1.119	1.098	
y	-0.421	-0.139	-0.028	0.040	0.272

利用本节的性质2：在给定(x_1, x_2)下，y与(x_3, x_4)很有可能相互独立.

5）**预测**

利用上面相关性分析，可以用样本协方差阵

	x_1	x_2	y
x_1	31.941		
x_2	19.314	223.515	
y	56.689	176.381	208.905

计算 y 关于(x_1, x_2)的条件期望，从而用(x_1, x_2)预测 y.

$$\hat{E}(y|x_1, x_2) \quad = \quad 95.423 + 1.369(x_1 - 7.462) + 0.671(x_2 - 48.154).$$

矩阵拉直和Kronecker积

矩阵拉直: 记$X=(x_1,\cdots,x_p)$是$n\times p$的矩阵。矩阵拉直运算就是将矩阵按列拉直为向量，拉直后的向量记为$vec(X)$, 有

$$vec(X)=\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix},$$

即$vec(X)$是一个$(np)\times 1$的向量。

Kronecker积: 令$A=(a_{ij})_{n\times p}$和B分别是$n\times p$和$m\times q$的矩阵。矩阵A和B的Kronecker积记为$A\otimes B$, 有

$$A\otimes B=(a_{ij}B),$$

所以$A\otimes B$是$(nm)\times(pq)$的矩阵。

拉直运算和Kronecker积的性质

性质1: 对任意实数λ, 有$(\lambda A)\otimes B = A\otimes(\lambda B)=\lambda(A\otimes B)$.

性质2: $A\otimes(B+C)=A\otimes B+A\otimes C$, $(B+C)\otimes A=B\otimes A+C\otimes A$.

性质3: $(A\otimes B)\otimes C=A\otimes(B\otimes C)$.

性质4: $(A\otimes B)'=A'\otimes B'$.

性质5: $(A\otimes B)(C\otimes D)=(AC)\otimes(BD)$.

性质6: 若A和B都是非奇异的方阵，则$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$.

性质7: $tr(A\otimes B)=tr(A)\cdot tr(B)$, $tr(C'D)=(vec(C))'(vec(D))$.

性质8: 若A和B分别是n和m阶方阵，则$|A\otimes B|=|A|^m\cdot|B|^n$.

性质9: 若A,Y和B分别是$n\times p, p\times q$和$q\times m$的矩阵，则

$$vec(AYB)=(B'\otimes A)vec(Y).$$

1.4 矩阵多元正态分布

设 $X_1, \cdots, X_n$ $i.i.d.$, $X_i = (x_{i1}, \cdots, x_{ip})' \overset{d}{\sim} N_p(\mu, \Sigma)$, 即

$X_1, \cdots, X_n$ 是来自 p 元正态总体 $N_p(\mu, \Sigma)$的独立样本.

记 $X = (X_1, \cdots, X_n)$, 则X是一个$p \times n$的随机矩阵.

随机矩阵的期望：$E(X) = (\mu, \cdots, \mu) = \mu \cdot \mathbf{1}_n'$,

其中 $\mathbf{1}_n = (1, \cdots, 1)'$.

矩阵的拉直运算： $vec(X) = vec((X_1, \cdots, X_n)) = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}_{(np)\times 1}$,

矩阵的拉直运算即是将矩阵依列拉直后形成一个向量。

随机矩阵的协方差阵：$Cov(X) = Cov(vec(X))$.

1.4.1 矩阵分布

随机矩阵的分布：随机矩阵拉直后的随机向量的分布。

矩阵 X 的运算：由于 $X = (X_1, \cdots, X_n)$，有

$$E(vec(X)) = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, Cov(vec(X)) = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma \end{pmatrix},$$

即 $E(vec(X)) = \mathbf{1}_n \otimes \mu$, $Cov(vec(X)) = \mathbf{I}_n \otimes \Sigma$, 其中 $\mathbf{I_n}$ 是 $\mathbf{n}$ 阶单位阵.

1.4.2 随机矩阵拉直运算的性质

性质1：对 $n \times m$ 的随机矩阵Y, 若有

$$E(vec(Y)) = \alpha \otimes \beta,\ Cov(vec(Y)) = A \otimes B,$$

其中, α, β分别是m和 n维列向量，A和B 分别是m和 n阶方阵, 则

$$E(vec(Y')) = \beta \otimes \alpha,\ Cov(vec(Y')) = B \otimes A.$$

由性质1，对上述随机矩阵 X 有

$$E(vec(X')) = \mu \otimes \mathbf{1}_n,\ Cov(vec(X')) = \Sigma \otimes \mathbf{I}_n.$$

因此，对由 n个p 维正态总体的独立样本组成的随机矩阵X,

$$vec(X) \overset{d}{\sim} N_{np}(\mathbf{1}_n \otimes \mu,\ \mathbf{I}_n \otimes \Sigma),$$
$$vec(X') \overset{d}{\sim} N_{np}(\mu \otimes \mathbf{1}_n,\ \Sigma \otimes \mathbf{I}_n).$$

1.4.3 矩阵正态分布

若 $vec(X) \overset{d}{\sim} N_{np}(\mathbf{1}_n \otimes \mu,\ \mathbf{I}_n \otimes \Sigma)$ 或 $vec(X') \overset{d}{\sim} N_{np}(\mu \otimes \mathbf{1}_n,\ \Sigma \otimes \mathbf{I}_n)$ ，则称随机矩阵X和X'分别服从**矩阵正态分布**，记为

$$
\begin{aligned}
X &\overset{d}{\sim} N_{p\times n}(\mu \cdot \mathbf{1}_n', \mathbf{I}_n \otimes \Sigma), \\
X' &\overset{d}{\sim} N_{n\times p}(\mathbf{1}_n \cdot \mu', \Sigma \otimes \mathbf{I}_n).
\end{aligned}
$$

一般地，记 $n \times p$的正态随机矩阵为 $X \overset{d}{\sim} N_{n\times p}(B, \Sigma \otimes V)$，其中

$$B = E(X),\ \Sigma \otimes V = Cov(vec(X)),$$

Σ 和 V分别是 p和 n阶方阵。

若 $X \overset{d}{\sim} N_{n\times p}(B, \Sigma \otimes V)$, 则 $vec(X) \overset{d}{\sim} N_{np}(vec(B), \Sigma \otimes V)$.

1.4.4 矩阵正态分布的密度函数

若 $X \overset{d}{\sim} N_{p\times n}(B, V\otimes\Sigma)$, Σ和V均为正定的方阵,

则由 $vec(X) \overset{d}{\sim} N_{np}(vec(B), V\otimes\Sigma)$，可以导出矩阵 X 的密度函数如下：

$$\frac{1}{(2\pi)^{(np)/2}\sqrt{|V\otimes\Sigma|}}\exp\left\{-\frac{(vec(X-B))'(V\otimes\Sigma)^{-1}(vec(X-B))}{2}\right\},$$

上式等价于：

$$\frac{1}{(2\pi)^{(np)/2}|V|^{p/2}|\Sigma|^{n/2}}\exp\left\{-\frac{tr[(X-B)'\Sigma^{-1}(X-B)V^{-1}]}{2}\right\}.$$

证明： 利用Kronecker乘积的一个**重要性质：**

$$vec(AXB) = (B'\otimes A)vec(X),$$

其中 A, X 和 B 分别是 $n\times p$, $p\times q$ 和 $q\times m$ 的矩阵.

1.4.5 矩阵正态分布的线性变换

性质2. 设 $p \times n$ 的矩阵 X 服从矩阵正态分布 $X \overset{d}{\sim} N_{p\times n}(B, V \otimes \Sigma)$, 有 $\Sigma_{p\times p} \geq 0, V_{n\times n} \geq 0$. 令

$$Y = C + AX\Gamma,$$

其中 $C_{q\times m}, A_{q\times p}$ 和 $\Gamma_{n\times m}$ 是常数矩阵，则

$$Y \overset{d}{\sim} N_{q\times m}((C + AB\Gamma), (\Gamma' V \Gamma) \otimes (A\Sigma A')).$$