课程主页：http://staff.ustc.edu.cn/~ynyang/vector

第四讲　多元正态分布

2025.3.10

$N_p(\boldsymbol{\mu}, \Sigma)$

内容

- 多元生成分布：球对称正态、Boltzmann分布
- 椭球分布
- 多元正态分布
- 高斯图模型
- 概率图模型：马氏随机场
- Hammersley-Clifford定理

许宝騄 （**Pao-Lu Hsu,** 1910年9月1日－1970年12月18日），字闲若，浙江省杭州府仁和县人，生于北京，中国现代数学家，概率论、数理统计学科的开创者。中央研究院院士。

许宝騄 P. L. Hsu	
出生	1910年9月1日 大清北京
逝世	1970年12月18日（60岁） 中国北京市
国籍	中华人民共和国
科学生涯	
研究领域	数学
机构	北京大学数学系
博士导师	埃贡·皮尔逊 耶日·内曼[1]
博士生	方开泰

幼年经历

许宝騄出身于浙江杭州的名门世家，生于北京。幼年随父亲赴任，曾经寓居天津、杭州等地，10岁起学作文言文。1928年考入燕京大学理学院。 1929年，转入清华大学数学系。当时，他的老师有熊庆来、孙光远、杨武之等人，共同学习的有华罗庚、柯召等人。1933年，许宝騄毕业获得理学士学位，经考试录取赴英国留学，但在体检时发现他体重太轻而不合格，故未能成行。

1934年，许宝騄出任北京大学数学系助教，担任正在北京大学访问的美国哈佛大学教授奥斯古德的助教，任职两年至1936年。在此两年中，许宝騄做了许多数学分析方面的习题，并掌握了矩阵这一工具，精于分块演算。

留学英国

1936年，许宝騄再度考取赴英国留学，派赴伦敦大学学院统计系，师从奈曼和皮尔逊，学习数理统计并攻读博士学位。从1936年至1940年，许宝騄一直是伦敦大学学院的研究生。1938年，许宝騄发表了3篇论文。当时，伦敦大学学院规定数理统计方向取得哲学博士学位，必须寻一新的统计量，编制一张统计量的临界值表，但许宝騄成绩优异，研究成果突出，成为首位破格采用统计实习的口试代替统计量临界值表的学生。1938年，许宝騄获哲学博士学位。

1938年，系主任耶日·内曼应聘赴美国加州大学伯克利分校执教，内曼推荐许宝騄升为伦敦大学学院讲师，接替他在伦敦大学学院授课。1939年，许宝騄发表两篇论文，1940年发表3篇论文。其中两篇论文是数理统计学的重要文献，为多元统计分析以及内曼-皮尔逊理论中的奠基性工作，由此许宝騄于1940年获得科学博士学位。

西南联大、哥大和北卡罗莱纳

抗日战争爆发后，许宝騄决定回国，1940年抵达昆明，成为北京大学数学系教授，任教于西南联合大学，一直任至1945年。钟开莱、王寿仁、徐利治等人都是他的学生。

1945年秋，许宝騄应邀赴美国加州大学伯克利分校及哥伦比亚大学担任访问教授，各授课一个学期，学生中包括T. W. Anderson、Erich Leo Lehmann等人。1946年，许宝騄来到北卡罗莱纳大学任教。1947年许宝騄回到北京大学。

文革期间

此次回国后不久，许宝騄发现自己已经患肺结核。 文革开始许宝騄与他的同事都遭大巨大的折磨，他的学生张尧庭表示，文革中期，许宝騄老师仍在研究组合数学中的问题。1966年开始，许宝騄拖着病躯，一步三歇应命前往指定地点接受批斗，如是三年。捱至1970年底终于不支死去。其兄许宝騤著文回忆：“文化大革命中，宝騄身蒙灾难，兢兢自克。亲属往访，辄闭门不纳；偶或遥遥一望，便挥手令去，不交一言，其精神痛苦可想见矣！宝騄偶染肺炎，不数日，竟在家属毫无所闻情况下溘然长逝。.....数页写着未竟的残稿散落在地，见之凄然掩涕，怆然神伤而已！”

链接：杨振宁谈许宝騄

多元生成模型：球对称正态、Boltzmann分布

球对称分布的仿射变换可以生成椭球分布，最重要的是$N_n(0,\sigma^2 I_n)$。Boltzmann分布是物理学中的重要概率模型，众多分布与其有关，我们也把它视作多元离散分布的生成模型。

球对称正态$N_n(0,\sigma^2 I_n)$

$\mathbf{x}=(x_1,\ldots,x_n)^\top \sim N_n(0,\sigma^2 I_n)$ （$\sigma^2=1$时是标准正态）

$\Leftrightarrow x_1,\ldots,x_n$ iid $\sim N(0,\sigma^2)$，

$\Leftrightarrow$ 概率密度函数$f(\mathbf{x})=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2\right)$

$\Leftrightarrow \frac{\mathbf{x}}{\|\mathbf{x}\|}\sim U(S^{n-1})$，$\frac{\mathbf{x}}{\|\mathbf{x}\|}\perp\!\!\!\perp \|\mathbf{x}\|$，$\|\mathbf{x}\|\sim\sigma\chi_n$，概率密度:

$$p_{n,\sigma}(r)=\frac{\sigma^{-n}}{2^{n/2-1}\Gamma(n/2)}r^{n-1}e^{-r^2/2\sigma^2},\quad r>0$$

Boltzmann (Gibbs)分布

封闭物理系统在热力学温度T时处于能量状态E_i的概率(或密度)

$$p(E_i)=\frac{1}{Z}\exp(-E_i/kT),$$

其中配分(partition)函数 $Z=\sum_i\exp(-E_i/kT)$，Boltzmann常数$k=1.380649\times10^{-23}$焦耳/开尔文。

$\sigma\chi_3$ 分布

分子速率的麦克斯韦-玻尔兹曼分布(Maxwell–Boltzmann)：

$$p_{3,\sigma}(v) = \sqrt{\frac{2}{\pi\sigma^3}} v^2 e^{-v^2/2\sigma^2},\ v > 0, \sigma = \sqrt{\frac{kT}{m}}$$

这是$\sigma\chi_3$ 分布，其中m: 质量, T:温度, k:Boltzmann常数。

麦克斯韦(1860)和玻尔兹曼考虑热平衡状态下理想气体分子（或粒子）速度向量$\mathbf{v} = (v_1, v_2,\ v_3)^\top$和速率$v = \sqrt{v_1^2 + v_2^2 + v_3^2}$ 的概率分布。

- 麦克斯韦假设$\mathbf{v}$各向同性且分量独立$\Rightarrow v_1, v_2,\ v_3$ iid $\sim N(0, \sigma^2)$。Hershell(1850)更早发现了这一结果。
- 玻尔兹曼从热力学角度推导出球对称正态，得到 $\sigma^2 = \frac{kT}{m}$,（是一般的Boltzmann能量分布的基础）

从而速率$v = \sqrt{v_1^2 + v_2^2 + v_3^2}\ \sim\sigma\chi_3$ ，$\sigma = \sqrt{kT/m}$

由此可得能量$E = mv^2/2$服从$(\frac{kT}{2})\chi_3^2$

Maxwell-Hershell定理：随机向量$\mathbf{x}$球对称分布且所有分量独立$\Leftrightarrow \mathbf{x} \sim N_n(0, \sigma^2 I_n)$

- 球对称多元正态的仿射变换生成（椭球）多元正态分布，这是本课程的主要模型；

  常见的统计模型，比如线性回归模型、误差正态的时间序列、纵向数据模型实际上是具有特殊结构的多元正态模型。

- Boltzmann分布应用于特定场合，假定能量的特殊形式，生成一些物理学中重要的多元离散分布或模型。

  Ising model, Boltzmann machine, restricted Boltzmann machine, energy-based models and deep Boltzmann machine

椭球分布

假设$\mathbf{x}$服从方差为I_p的球对称分布，密度函数为$f(\mathbf{x}) = h(\|\mathbf{x}\|)$, $\mathbf{x} \in R^p$。给定常数向量$\boldsymbol{\mu} \in R^p$和 p 阶正定矩阵Σ，则$\mathbf{x}$的仿射变换的分布称为椭球分布：

$$\mathbf{y} = \Sigma^{1/2}\,\mathbf{x} + \boldsymbol{\mu} \sim E_p(\boldsymbol{\mu}, \Sigma)$$

其概率密度函数为

$$p(\mathbf{y}) = |\Sigma|^{-1/2}\, h((\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}))$$

因为$\mathbf{x}$均值为0，方差为I_p, $E_p(\boldsymbol{\mu}, \Sigma)$的均值为$\boldsymbol{\mu}$，协方差为$\Sigma$。

最重要的椭球分布是多元正态分布。

例1. 假设$\mathbf{x}$服从标准的多元正态

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}} \exp\left(-\|\mathbf{x}\|^2/2\right),$$

则$\mathbf{y} = \Sigma^{1/2}\,\mathbf{x} + \boldsymbol{\mu}$ 的概率密度函数为

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp(-(\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})/2)$$

这称为多元正态分布， $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$.

例2. 球内均匀分布 $f(\mathbf{x}) = \frac{1}{|B^n|}, \|\mathbf{x}\| \le 1,$

椭球均匀分布：

$$g(\mathbf{y}) = \frac{1}{|B^n||\Sigma|^{1/2}}, \mathbf{y} \in E_n = \{\mathbf{y}: (\mathbf{y}-\boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{y}-\boldsymbol{\mu}) \le 1\}$$

椭球E_n的体积$|E_n| = |B^n||\Sigma|^{1/2}$。

椭圆$E_2 = \left\{(x_1, x_2): \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} \le 1\right\}$的面积$V_2 = \pi a_1 a_2$，周长没有初等表达

$$S_{E_2} = 4\int_0^{\pi/2} \sqrt{a_1^2 \sin^2(\theta) + a_2^2 \cos^2(\theta)}\, d\theta = 2\pi a_1 \sum_{k=0}^{\infty} \frac{\Gamma(k-\frac{1}{2})\Gamma(k+\frac{1}{2})}{\Gamma(-\frac{1}{2})\Gamma(k+1)} \frac{\alpha^k}{k!}, \alpha = 1 - a_2^2/a_1^2$$

椭圆积分

椭球$E_n = \left\{(x_1, \dots, x_n): \sum_{i=1}^n \frac{x_i^2}{a_i^2} \le 1\right\}$的体积$V_n = |B^n| a_1 \cdots a_n$，而表面积$S_n$没有初等表达， Rivin（2007）给出了一个有趣表达：

$$S_{E_n} = \frac{n\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} E\sqrt{\sum_{i=1}^n X_i^2/a_i^2}, \quad \mathrm{X}_1, \dots, \mathrm{X}_n \text{ iid} \sim N(0, 1/2)$$

Igor Rivin. Surface area and other measures of ellipsoids. Advances in Applied Mathematics 39 (2007) 409–427

多元正态分布

基于球对称的多元标准正态分布生成椭球多元正态分布，这其实是多元正态分布的一种定义方式：

多元正态定义1

定义1（多元正态，许宝騄）：假设随机向量$\mathbf{z}_{q\times 1} \sim N_q(\mathbf{0}, I_q)$,即$\mathbf{z}$的联合密度为$f(\mathbf{z}) = \frac{1}{(2\pi)^{q/2}} \exp(-\|\mathbf{z}\|^2/2)$。假设$A$是$p\times q$常数矩阵，$\boldsymbol{\mu}$是$p\times 1$常数向量，则称$\mathbf{x} = A\mathbf{z} + \boldsymbol{\mu}$服从参数为$\boldsymbol{\mu}$，$AA^{\mathsf{T}}$的多元正态分布，记作$\mathbf{x} \sim N_p(\boldsymbol{\mu}, AA^{\mathsf{T}})$。

若A不是行满秩的，则AA^{T}不可逆，此时称为退化多元正态分布。退化多元正态没有恰当的概率密度，但与非退化正态类似。

定义1最常见的情况是取$q=p, A=\Sigma^{1/2}$, 其中$\Sigma>0$，此时

$$\mathbf{x}=\Sigma^{1/2}\mathbf{z}+\boldsymbol{\mu}\sim N_p(\boldsymbol{\mu},\Sigma)$$

由$\mathbf{z}$的分布立得$\mathbf{x}$的密度（如定义2所示）

Jacobian $J\left(\mathbf{z}\to\mathbf{x}=\Sigma^{1/2}\mathbf{z}+\boldsymbol{\mu}\right)$
$=\left|\Sigma^{-1/2}\right|=1/|\Sigma|^{1/2}$

多元正态定义2

定义2(非退化多元正态)：若p维随机向量$\mathbf{x}$的概率密度函数为

$$f(\mathbf{x})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),\quad \mathbf{x}\in R^p$$

其中参数$\boldsymbol{\mu}\in R^p$，Σ为$p\times p$正定矩阵，则称$\mathbf{x}$服从p元正态分布，记作$\mathbf{x}\sim N_p(\boldsymbol{\mu},\Sigma)$。

非退化情形下定义1和定义2等价。

下面除非特别声明，总假设$\Sigma>0$.

我们已熟知多元正态分布的性质1-6（p13-16），简单列举如下。

定义1这种生成/表示方式使得我们很容易地从标准多元正态的性质推导出一般多元正态的性质。上述密度求解即是一例 。再如：

性质1：$N_p(\boldsymbol{\mu}, \Sigma)$的均值和方差为$\boldsymbol{\mu}$， Σ

证：由$\mathbf{z} \sim N_p(\mathbf{0}, I_p)$，$E(\mathbf{z}) = \mathbf{0}, \mathrm{var}(\mathbf{z}) = I_p$, 容易得到$\mathbf{x} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}$的均值和方差 $E(\mathbf{x}) = \boldsymbol{\mu}$，$\mathrm{var}(\mathbf{x}) = \mathrm{var}\left(\Sigma^{1/2}\mathbf{z}\right) = \Sigma$.

性质2： $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，划分$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$， $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$，则 $\mathbf{x}_1 \perp\!\!\!\perp \mathbf{x}_2 \Leftrightarrow \Sigma_{12} = 0$，

证： $\mathbf{z} \sim N_p\left(\mathbf{0}, I_p\right)$，$\Sigma_{12} = 0$时，

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu} = \begin{pmatrix} \Sigma_{11}^{1/2}\mathbf{z}_1 \\ \Sigma_{22}^{1/2}\mathbf{z}_2 \end{pmatrix} + \boldsymbol{\mu}，\ \boldsymbol{z}_1 \perp\!\!\!\perp \boldsymbol{z}_2 \Rightarrow \mathbf{x}_1 \perp\!\!\!\perp \mathbf{x}_2$$

性质3. $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$的则矩母函数

$$E\exp(\mathbf{t}^{\top}\mathbf{x}) = \exp\left(\mathbf{t}^{\top}\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^{\top}\Sigma\mathbf{t}\right), \ \forall \mathbf{t} \in R^p.$$

注：这也可作为多元正态的一个等价定义。

证明：假设$\mathbf{z} \sim N_p(0, I_p)$，$\mathbf{x} = \Sigma^{-1/2}\mathbf{z} + \boldsymbol{\mu}$，则$\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$,

对任何$\mathbf{s} \in R^p$，已知对一元正态$z_i \sim N(0,1)$：$E\exp(tz_i) = e^{t^2/2}$.

$$E\exp(\mathbf{s}^{\top}\mathbf{z}) = \prod_{i=1}^{p} E\exp(s_i z_i) = \exp(\mathbf{s}^{\top}\mathbf{s}/2)$$

$$\Rightarrow \ E\exp(\mathbf{t}^{\top}\mathbf{x}) = E\exp(\mathbf{t}^{\top}(\Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu})) = \exp(\mathbf{t}^{\top}\boldsymbol{\mu})E\exp((\Sigma^{1/2}\mathbf{t})^{\top}\mathbf{z})$$

$$= \exp(\mathbf{t}^{\top}\boldsymbol{\mu})\exp((\Sigma^{1/2}\mathbf{t})^{\top}\Sigma^{1/2}\mathbf{t}/2) = \exp(\mathbf{t}^{\top}\boldsymbol{\mu} + \mathbf{t}^{\top}\Sigma\mathbf{t}/2)。$$

性质4. 若 $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，则对任意常数矩阵$A_{q\times p}$和常数向量$\mathbf{b}_{q\times 1}$，有

$$A\mathbf{x}+\mathbf{b} \sim N_q\left(A\boldsymbol{\mu}+\mathbf{b}, A\Sigma A^{\mathsf{T}}\right)。$$

证：由定义1，$\mathbf{x}$可表示为$\mathbf{x}=\Sigma^{1/2}\mathbf{z}+\boldsymbol{\mu}$，其中$\mathbf{z} \sim N_p(0, I_p)$，

所以$A\mathbf{x}+\mathbf{b}=A\Sigma^{1/2}\mathbf{z}+A\boldsymbol{\mu}+\mathbf{b}$,由定义1，$A\mathbf{x}+\mathbf{b} \sim N_q\left(A\boldsymbol{\mu}+\mathbf{b}, A\Sigma^{1/2}\Sigma^{1/2}A^{\mathsf{T}}\right)$。

性质5：设 $\mathbf{x} \sim N_n(\boldsymbol{\mu}, \Sigma), \Sigma=(\sigma_{ij})$,若$\sigma_{ij}=0$，则 x_i与x_j独立.

证明：由定理2，取$A=I_p$的第i, j行组成的$2\times p$矩阵，则

$$A\mathbf{x}=\begin{pmatrix} x_i \\ x_j \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \begin{pmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ji} & \sigma_{jj} \end{pmatrix}\right),$$

若$\sigma_{ij}=0$，则x_i，x_j独立.

性质6. 若$\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，划分 $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$，$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$，$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} > 0$,

其中$\mathbf{x}_1$长度为q,则边际分布和条件分布依然是（多元）正态分布：

(1) $\mathbf{x}_1 \sim N_q(\boldsymbol{\mu}_1, \Sigma_{11}), \mathbf{x}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \Sigma_{22})$.

(2) $\mathbf{x}_1 \mid \mathbf{x}_2 \sim N_q\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11\bullet 2}\right)$，其中$\Sigma_{11\bullet 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$。

证：(1) 取$A_{q\times p} = (I_q, 0)$,由性质4，$\mathbf{x}_1 = A\mathbf{x} \sim N_q(A\boldsymbol{\mu}, A\Sigma A^T) = N_q(\boldsymbol{\mu}_1, \Sigma_{11})$.

(2) 令 $\mathbf{y}_1 = \mathbf{x}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2$，$\mathbf{y}_2 = \mathbf{x}_2$，容易验证$\mathrm{cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{0}$。由性质4

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} I_q & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-q} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$$

$$\sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11\bullet 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right),$$

由(1)和(2)知$\mathbf{y}_1 = \mathbf{x}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2 \sim N_q(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2, \Sigma_{11\bullet 2})$，且与$\mathbf{y}_2 = \mathbf{x}_2$独立.

所以给定$\mathbf{x}_2$条件下，$\mathbf{x}_1 = \mathbf{y}_1 + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2 \sim N_q(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2 + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2, \Sigma_{11\bullet 2})$.

Σ^{-1}称为精度矩阵(precision matrix)

条件分布（性质6）也可由逆矩阵Σ^{-1}表述：

性质6*. 若$\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，划分 $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$，$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$，$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} > 0$

$\Omega = \Sigma^{-1} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$，其中$\mathbf{x}_1$，$\boldsymbol{\mu}_1$，$\Sigma_{11}$，$\Omega_{11}$都是$q$阶，则

$$\mathbf{x}_1 \mid \mathbf{x}_2 \sim N_q\left(\boldsymbol{\mu}_1 - \Omega_{11}^{-1}\Omega_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Omega_{11}^{-1}\right)。$$

证：由$\Omega\Sigma = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} I_q & 0 \\ 0 & I_{p-q} \end{pmatrix} \Rightarrow \begin{cases} \Omega_{11}\Sigma_{12} + \Omega_{12}\Sigma_{22} = 0 \\ \Omega_{11}\Sigma_{11} + \Omega_{12}\Sigma_{21} = I_q \end{cases}$，

第一式$\Rightarrow \Sigma_{12}\Sigma_{22}^{-1} = -\Omega_{11}^{-1}\Omega_{12} \Rightarrow \Omega_{12} = -\Omega_{11}\Sigma_{12}\Sigma_{22}^{-1}$

第二式$\Rightarrow I_q = \Omega_{11}\Sigma_{11} + \Omega_{12}\Sigma_{21} = \Omega_{11}\Sigma_{11} - \Omega_{11}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$$\Rightarrow \Omega_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} = \Sigma_{11\bullet 2}^{-1}$$

因此条件分布

$$\mathbf{x}_1 \mid \mathbf{x}_2 \sim N_q\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11\bullet 2}\right) = N_q\left(\boldsymbol{\mu}_1 - \Omega_{11}^{-1}\Omega_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Omega_{11}^{-1}\right).$$

注：以$\Omega=\Sigma^{-1}$表达条件分布有利于考察条件独立性（马氏性质）

对于$q=1$情形。我们对比性质6和性质6*

(1) 性质6表明回归函数

$$E(x_1\mid \mathbf{x}_2)=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)\triangleq\mu_1+\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_2-\boldsymbol{\mu}_2)$$

我们尤其关心$\boldsymbol{\beta}=\Sigma_{12}\Sigma_{22}^{-1}$哪些分量为0，这是因为

$$\beta_j=0\Leftrightarrow x_1\mid\mathbf{x}_2\text{的分布与}x_j\text{无关}\Leftrightarrow x_1\text{与}x_j\text{条件独立。}$$

这需要计算乘积$\Sigma_{12}\Sigma_{22}^{-1}$才能判定。

(2) 性质6*给出了另外一个表达：$\boldsymbol{\beta}=\Sigma_{22}^{-1}\Sigma_{21}=-\Omega_{21}\Omega_{11}^{-1}$，$\Omega_{11}^{-1}=1/\omega_{11}$是标量

$$\beta_j=0\Leftrightarrow(\Omega_{21})_j=0$$

简言之，$\boldsymbol{\beta}=\Sigma_{12}\Sigma_{22}^{-1}=-\Omega_{11}^{-1}\Omega_{12}$，

$\Sigma_{12}\Sigma_{22}^{-1}$不容易判断哪些分量为0，而$\Omega_{11}^{-1}\Omega_{12}$容易判断

$1\times(p-1)$　$(p-1)\times(p-1)$　　1×1　$1\times(p-1)$

附录：统计学的指数族分布尝试以Boltzman分布统一概率分布

$$f(x,\theta)=h(x)\exp\big(\eta(\theta)x-A(\theta)\big)=\frac{1}{Z(\theta)}\exp(\eta(\theta)x-B(x)\,),$$

$Z(\theta)$=exp$(A(\theta))$ 包含了f的所有信息，矩母函数：

$$E\exp(tx)=\frac{Z(\theta+\mathrm{t})}{Z(\theta)}=\exp\big(A(\theta+t)-A(\theta)\big),$$

一般，$Z(\theta)$作为光滑函数没有实数零点，也不包含突变信息，但其虚数根在极限情况下，却有实际意义（包含了分布突变/相变的信息）。

例如Bernouli：$x\sim B(1,p), P(x=1)=p, P(x=0)=1-p$

$$p(x)=p^x(1-p)^{1-x}=\exp\left(x\log\left(\frac{p}{1-p}\right)+\log(1-p)\,\right)$$

令$\theta=\log\left(\frac{p}{1-p}\right)$，则$p=\frac{e^\theta}{1+e^\theta}$，$\log(1-p)=-\log\big(1+e^\theta\big)$

$$p(x)=\exp(x\theta-\log\big(1+e^\theta\big))=e^{\theta x}/\mathrm{Z}(\theta),$$

其中$\mathrm{Z}(\theta)=1+e^\theta$。

$$E\exp(tx)=\frac{Z(\theta+\mathrm{t})}{Z(\theta)}=\frac{1+e^{\theta+t}}{1+e^\theta}$$