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

第四讲 多元正态

2024.3.11

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。

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。

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

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。其兄许宝骙著文回忆：“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！”

多元正态分布

随机向量/相依随机变量的概率模型（即多元分布）不太丰富，最常见的是多元正态分布。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。 其它多元分布还有球对称分布、椭球分布、马氏随机场、 Copula等。

第二讲我们通过仿射变换，从球对称的多元标准正态分布得到了椭球多元正态分布，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$f(\mathbf{z})=\dfrac{1}{(2\pi)^{q/2}}\exp(-\|\mathbf{z}\|^2/2)$。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$\boldsymbol{\mu}$，$AA^{\mathrm{T}}$的多元正态分布，记作$\mathbf{x}\sim N_p(\boldsymbol{\mu},AA^{\mathrm{T}})$。

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

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

由此由$\mathbf{z}$的分布立得（非退化）$\mathbf{x}$的密度如定义2所示（Jacobi $J(\mathbf{z}\to\mathbf{x})=\left|\Sigma^{-1/2}\right|=1/|\Sigma|^{1/2}$）。

非退化的多元正态分布一般基于该密度定义：

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$\Sigma>0$.

定义1这种生成方式或者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$\boldsymbol{\mu}$，$\Sigma$

证：由$E(\mathbf{z}) = \mathbf{0}, \mathrm{var}(\mathbf{z}) = I_p$, 容易得到$\mathbf{x}$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：$\Sigma_{12} = 0$时，$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu} = \begin{pmatrix} \Sigma_{11}^{1/2}\mathbf{z}_1 \\ \Sigma_{22}^{1/2}\mathbf{z}_2 \end{pmatrix} + \boldsymbol{\mu}$

定理1. $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$的则矩母函数

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

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

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

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

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

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

注：这当然也可作为多元正态的一个等价定义。

定理2．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3：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.

定理3. 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$A_{q\times p} = (I_q, 0)$,由定理2，$\mathbf{x}_1 = A\mathbf{x} \sim N_q(A\boldsymbol{\mu}, A\Sigma A^T) = N_q(\boldsymbol{\mu}_1, \Sigma_{11})$.

(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$\mathrm{cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{0}$。由定理2

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

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

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

P9之后的内容是定理3的重新表述，即高斯图模型

我们（可能）不太熟悉的是，条件分布（定理3）也可由逆矩阵Σ^{-1}表述：

定理3*. 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

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

我们尤其关心$q=1$的情形,回归函数(定理3(3))

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

中$\mathbf{x}_2$的哪些分量与x_1无关即$\boldsymbol{\beta}=\Sigma_{22}^{-1}\Sigma_{21}$哪些分量为0? 这包含了计算$(p-1)\times(p-1)$矩阵$\Sigma_{22}$的逆，以及计算乘积$\Sigma_{22}^{-1}\Sigma_{21}$之后才能判定。

推论1给出了$\boldsymbol{\beta}$的另一种表达：$\boldsymbol{\beta}=\Sigma_{22}^{-1}\Sigma_{21}=-\Omega_{21}\Omega_{11}^{-1}$,

当$q=1$时，$\Omega_{11}^{-1}=1/\omega_{11}$是标量，因而为了判断那些$\boldsymbol{\beta}$分量等于0,

只需考察向量Ω_{21}的哪些分量为零。

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

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

推论1. 假设 $\mathbf{x} \sim N_p(\mu, \Omega^{-1})$，记$\mathbf{x}_{-i}$为$\mathbf{x}$中除了第$i$个分量之外的其它分量组成的向量：$\mathbf{x}_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)^\top$, 记 $\Omega = (\omega_{ij})$，则

$$x_i | \mathbf{x}_{-i} \sim N_1\left(\mu_i - \sum_{j \neq i} \omega_{ij}(x_j - \mu_j)/\omega_{ii}, 1/\omega_{ii}\right),$$

特别地，$\omega_{ij} = 0 \Leftrightarrow x_i \perp\!\!\!\perp x_j | \mathbf{x}_{-(i,j)}$ （条件独立/马氏性）

这是定理3*的直接推论，其中需要说明的是：
$\omega_{ij} = 0 \Rightarrow x_i | \mathbf{x}_{-i} = x_i | x_j, \mathbf{x}_{-(i,j)}$的分布与$x_j$无关，这表明在给定$\mathbf{x}_{-(i,j)}$的条件下，$x_i$与$x_j$条件独立。

下一讲我们将使用更简单的方式得到这个结果。