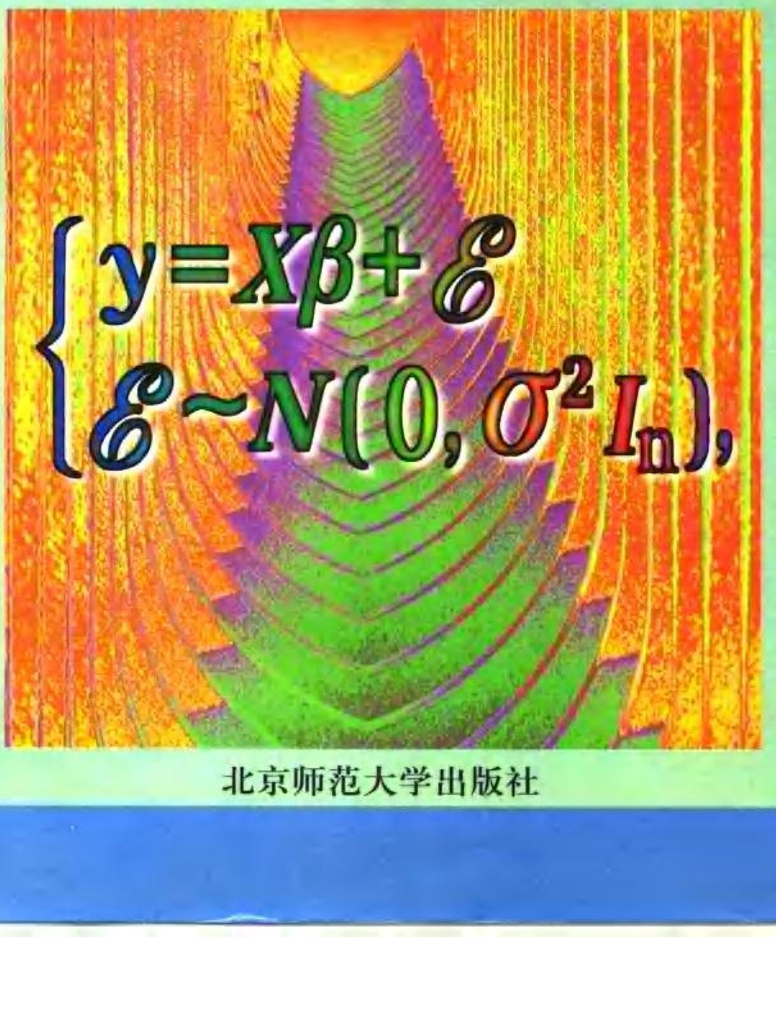
线性模型引论

杨文礼　编著



北京师范大学出版社

线性模型引论

杨文礼 编著

北京师范大学出版社

图书在版编目（CIP）数据

线性模型引论/杨文礼编著. 一北京：北京师范大学出版社，1998. 4 ISBN 7-303-04606-2

Ⅰ. 线… Ⅱ. 杨… Ⅲ. 线性-模型论 Ⅳ. 0141. 4

中国版本图书馆 CIP 数据核字（97）第 25892

北京师范大学出版社出版发行
（100875 北京新街口外大街 19 号）
北京师范大学印刷厂印刷 全国新华书店经销
开本:850×1168 1/32 印张:8.625 字数:209 千
1998 年 5 月北京第 1 版 1998 年 5 月北京第 1 次印刷
印数:1～3 000 册
定价:10.00 元

前 言

线性统计模型(简称为线性模型)是非常重要的一种统计模型,这一方面是由于它包括了一系列统计模型,诸如线性回归模型、方差分析与协方差分析模型、方差分量模型等等,这些统计模型在国民经济的发展中都有着广泛而重要的应用;另一方面,线性模型的基本理论与方法也为其它统计理论与方法提供了基本的工具。

正因为线性模型在理论与应用上具有十分重要的作用,近年来,许多高等学校将其列为必选课或有关课程的基本内容。为了满足统计专业研究生及数学系高年级学生教学之需要,我们编写了“线性统计模型讲义”,并在我校数学系多次使用。经过不断修改、充实,才逐步形成这本“线性模型引论”。它可供学生一个学期(每周4学时)教学课程之用,也可供有关专业教师及科技工作者参考。

线性模型的内容十分丰富,作为一个学期使用的教材,只能讲述其中最基本的内容。在这种意义下,本书又可视为“线性模型简明教程”。

本书的编写力求深入浅出、重点突出、讲述详细、论证清楚,以使读者逐步掌握线性模型的基本理论与方法。

本书虽经多次使用,并不断完善提高,但限于作者的水平,肯定会有不少缺点与错误,恳请同行专家与广大读者批评指正。

最后我要对王松桂教授的热情鼓励、积极支持,北京师大出版社的慷慨资助及张其友同志的热情具体的帮助,表示衷心的感谢。

杨文礼

1997.5.12 于北京师范大学数学系

目　录

第一章　一般线性模型，矩阵的广义逆

§1. 引言与模型

研究一个（或一组）变量与另一个（或另一组）变量之间的关系，在几乎所有的科学分支中都是非常重要的. 而变量之间的关系大体上可分为两类，一类是确定性关系，即数学上的函数关系，例如 $S=v_ot+\frac{1}{2}at^2$ 就是匀加速直线运动中路程 S 与时间 t 之间的确定性关系，其中 v_o 为初始速度，a 为匀加速度；另一类是非确定性关系，又叫统计相关关系. 例如某种农作物在单位面积上的产量 Y 与施肥量 X_1、浇水量 X_2、种籽 X_3（定性变量）就有着非常密切的关系，但对给定的 $X_1=x_1, X_2=x_2, X_3=x_3$ 来说，Y 还是不确定的. 事实上，Y 是一随机变量，其概率分布与自变量 $(X_1,X_2,X_3)'\triangleq X$ 所取的值 $(x_1,x_2,x_3)'\triangleq x$ 密切相关. 作为自变量的 X 可以是一维的，也可以是多维的；可以是定量变量，也可以是定性变量；可以是非随机变量，也可以是随机变量（甚至一部分分量随机，另一部分分量不随机）. 因变量（又叫响应变量）Y 与自变量（又叫预测变量或预报变量）$X=(X_1,\cdots,X_k)'$（在上面的例子中 $k=3$）之间的这种统计相关关系可表示为

$$Y=\mu(x_1,\cdots,x_k)+\varepsilon,$$

其中 $\mu(x_1,\cdots,x_k)$ 是在给定 $X\triangleq(X_1,\cdots,X_k)'=(x_1,\cdots,x_k)'$ 的条件下，因变量 Y 的平均值（条件数学期望），即

$$\mu(x_1,\cdots,x_k)=E(Y\,|\,X_1=x_1,\cdots,X_k=x_k)$$

且 $\mu(x_1,\cdots,x_k)$ 中含有若干个（比如说 p 个）未知参数 $\beta_1,\cdots,\beta_p$，而 ε 为随机误差，满足 $E\varepsilon=0$.

特别，当$\mu(x_1,\cdots,x_k)$为未知参数$\beta_1,\cdots,\beta_p$的线性函数，即$\mu(x_1,\cdots,x_k)=\sum_{j=1}^{p}\beta_j f_j(x_1,\cdots,x_k)$时，称因变量$Y$与自变量$X=(X_1,\cdots,X_k)'$之间的统计相关关系为线性统计模型，简称为线性模型.这样，一个线性模型乃是因变量Y与自变量$X=(X_1,\cdots,X_k)'$之间如下类型的一种统计相关关系：

$$Y=\sum_{j=1}^{p}\beta_j f_j(x_1,\cdots,x_k)+\varepsilon,$$

其中β_j为未知参数，$f_j(x_1,\cdots,x_k)$为$(x_1,\cdots,x_k)'\triangleq x$的某种已知函数(不含未知参数)，$j=1,\cdots,p$，$\varepsilon$为随机误差，满足$E\varepsilon=0$，人们通常还假定$D\varepsilon\triangleq\sigma^2>0$.

下面是线性模型的一些例子.

1. 在研究儿子的身高Y与父母亲的平均身高X之间的关系时，下面的模型是适用的：

$$Y=\alpha+\beta x+\varepsilon,$$

其中α,β为未知参数，ε为随机误差.

2. 若用t表示时间，Y是要研究的某种动物(或植物)的体重(相应地，高度)的增加量，则下述类型的增长曲线模型是适用的：

$$Y=\alpha+\beta t+\gamma t^2+\varepsilon.$$

注意：这是关于未知参数α,β,γ的线性模型.

3. 在平炉炼钢中，由于矿石与炉气的氧化作用，铁水的含碳量在不断降低.一炉钢在冶炼初期总的去碳量Y与所加的两种矿石量x_1,x_2及熔化时间x_3之间服从如下的线性模型：

$$Y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon,$$

其中$\beta_0,\beta_1,\beta_2,\beta_3$为未知参数，$\varepsilon$为随机误差.

4. 在农业科学试验中，要考察某种农作物的p个品种$A_1,\cdots,$

A_p 对单产(单位面积上的产量)的影响是否有显著性的差异，考虑下述模型是适宜的：

$$Y_i=\mu_i+\varepsilon_i, i=1,\cdots,p.$$

其中 Y_i 是播种 A_i 品种情况下的单产，μ_i 为未知参数，ε_i 为随机误差，满足 $E\varepsilon_i=0, i=1,\cdots,p$.
这里响应变量 $Y_i, i=1,\cdots,p$，就不是一个而是一组(p 个)了，μ_1，$\cdots$，μ_p 为未知参数，$\varepsilon_1,\cdots,\varepsilon_p$ 为随机误差.

若要考虑 p 个品种 $A_1,\cdots,A_p$ 及 q 种肥料 $B_1,\cdots,B_q$ 各自(以及它们之间的交互作用)对单产的影响是否显著，考虑下述模型是适宜的：

$$Y_{ij}=\mu_{ij}+\varepsilon_{ij}, i=1,\cdots,p; j=1,\cdots,q.$$

其中 Y_{ij}是播种 A_i 品种、施 B_j 种肥料情况下的单产，μ_{ij}为未知参数，ε_{ij}为随机误差.

§2. 一般线性模型

如前所述，线性模型乃是如下类型的一种统计相关关系：

$$Y=\sum_{j=1}^{p}\beta_j f_j(x_1,\cdots,x_k)+\varepsilon, \tag{2.1}$$

其中 ε 为随机误差，满足

$$E\varepsilon=0, D\varepsilon\triangleq\sigma^2>0, \tag{2.2}$$

$\beta_1,\cdots,\beta_p$ 及 σ^2 为未知参数，$f_j(x_1,\cdots,x_k)$是不含未知参数的关于$(x_1,\cdots,x_k)'\triangleq x$ 的某种已知函数.

线性模型要解决的问题主要有两个：一是未知参数的估计，二是关于未知参数的假设检验. 当然对不同类型的线性模型来说，所要研究解决的问题其侧重点会有所不同，例如，在线性回归模型中，未知参数$(\beta_1,\cdots,\beta_p)'\triangleq\beta$ 及其线性函数的估计是最重要的，而

在方差分析模型中，侧重点则在假设检验上．为了解决线性模型中未知参数的估计与假设检验问题，必须进行试验（或观测），以获得必要的数据．假设进行了 n 次独立观测，得 Y 的 n 个观测值 Y_1，$\cdots$，Y_n，则它们必须满足(2.1)，即应有

$$Y_i=\sum_{j=1}^{p}\beta_j f_j(\tilde{x}_{i1},\cdots,\tilde{x}_{ik})+\varepsilon_i,i=1,\cdots,n. \qquad (2.3)$$

其中 Y_i 是当 $X_1=\tilde{x}_{i1},\cdots,X_k=\tilde{x}_{ik}$ 时，因变量 Y 的观测值，ε_i 是在第 i 次观测中所产生的随机误差，$i=1,\cdots,n$，独立，且 $E\varepsilon_i=0,D\varepsilon_i=\sigma^2>0,i=1,\cdots,n$.

若记

$$y=\begin{bmatrix}Y_1\\ \vdots\\ Y_n\end{bmatrix},X=\begin{bmatrix}x_{11}\cdots & x_{1p}\\ \vdots\ \ddots & \vdots\\ x_{n1}\cdots & x_{np}\end{bmatrix},\beta=\begin{bmatrix}\beta_1\\ \vdots\\ \beta_p\end{bmatrix},\varepsilon=\begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{bmatrix},$$

其中 $x_{ij}=f_j(\tilde{x}_{i1},\cdots,\tilde{x}_{ik}),i=1,\cdots,n;j=1,\cdots,p.$

则上面的 n 个方程（即(2.3)）可以合写成如下一个矩阵方程

$$y=X\beta+\mathscr{E}, \qquad (2.3)'$$

其中随机误差向量 $\mathscr{E}$ 满足

$$E\mathscr{E}=0,D\mathscr{E}=\sigma^2 I_n. \qquad (2.4)$$

人们通常也称(2·3)′、(2·4)为线性模型或线性模型的数据形式，或一般线性模型．易见，一般线性模型乃是§1 所述线性模型的数据表现形式，有人称其为完全线性模型，而称其原型（即§1 所述的线性模型）为理论线性模型．但人们通常并不将理论线性模型与一般线性模型严格加以区别，而统称它们为线性模型．

一般地，称模型

$$\begin{cases} y=X\beta+\mathscr{E}, \\ E\mathscr{E}=0, D\mathscr{E}=\Sigma \end{cases}$$

为线性统计模型，简称为线性模型. 其中 n 维向量 y 称为观测向量. 而称已知矩阵 $\underset{n\times p}{X}$ 为模型的设计矩阵. β 为 p 维未知参数向量. $\mathscr{E}$ 为 n 维随机误差向量，其均值为 $E\mathscr{E}=0$. 协方差阵 $\Sigma=D\mathscr{E}$. 通常假定 $\Sigma=\sigma^2 I_n$ 或 $\Sigma=\sigma^2 V$，或其它适当的假定，其中 $\sigma^2>0$ 未知，$V\geqslant 0$ 已知.

更一般地，可将 X 理解成一个线性变换（线性算子），而将 Σ 理解成对称、非负定映射（称为协差算子）. 换言之，我们有如下更一般的

定义 设$(\Omega,\mathscr{F},P)$，$P\in\mathscr{P}$，是一概率空间族，y 是 $\Omega\to\mathscr{H}$（一个向量空间）的一个可测变换（随机元）. 若存在线性子空间 $\mathscr{L}\subseteq\mathscr{H}$ 及 $\mathscr{H}\to\mathscr{H}$ 的对称非负定映射的集合 $\mathscr{Q}$，使得

i) $E_P y\in\mathscr{L}$，$\forall\ P\in\mathscr{P}$；

ii) $\mathrm{Span}\{E_P y: P\in\mathscr{P}\}=\mathscr{L}$；

iii) $\mathrm{Cov}_P(y)\in\mathscr{Q}$，$\forall\ P\in\mathscr{P}$，

则称 y 服从线性模型 $M(\mathscr{L},\mathscr{Q})$.

注：设$(\Omega,\mathscr{F},P)$是一概率空间，$\mathscr{H}$ 是带有内积$\langle\cdot,\cdot\rangle$的一个向量空间，$\mathscr{A}$ 是 $\mathscr{H}$ 上的一个 σ—代数，则称 $\Omega\to\mathscr{H}$ 的一个可测变换 y（即 $y^{-1}(\mathscr{A})\subseteq\mathscr{F}$）为 $\mathscr{H}$ 值随机元（$r.e.$）.

若存在 $\mu\in\mathscr{H}$，使$\forall\ A\in\mathscr{H}$，有

$$\mathrm{E}\langle A,y\rangle=\langle A,\mu\rangle,$$

则称 μ 为随机元 y 的数学期望，记作 $Ey=\mu$.

若存在 $\mathscr{H}\to\mathscr{H}$ 的对称、非负定映射 Σ，使$\forall\ A,B\in\mathscr{H}$ 有

$$\mathrm{Cov}(\langle A,y\rangle,\langle B,y\rangle)=\langle A,\Sigma B\rangle,$$

则称 Σ 为随机元 y 的协差算子，记作 $\mathrm{Cov}(y)=\Sigma$，或 $D(y)=\Sigma$.

其中 $\mathscr{H}\to\mathscr{H}$ 的映射 Σ 是对称的，记作 $\Sigma=\Sigma'$，意指：$\langle A,\Sigma B\rangle=\langle\Sigma A,B\rangle,\forall\ A,B\in\mathscr{H}$.

而 $\mathscr{H}\to\mathscr{H}$ 的对称映射 Σ 是非负定的，记作 $\Sigma\geqslant0$，意指：$\langle A,\Sigma A\rangle\geqslant0,\forall\ A\in\mathscr{H}$.

例. 设 $\xi=(\xi_1,\cdots,\xi_p)'$ 为在 $\mathbf{R}^p$ 中取值的 $r.v.$（即为 p 维实 $r.v.$），则 $\forall\ a\in\mathbf{R}^p$，有

$$E\langle a,\xi\rangle=Ea'\xi=E\sum_{i=1}^{p}a_i\xi_i=\sum_{i=1}^{p}a_iE\xi_i$$

$$=(a_1,\cdots,a_p)\begin{bmatrix}E\xi_1\\ \vdots\\ E\xi_p\end{bmatrix}=\langle a,\mu\rangle$$

其中 $a=(a_1,\cdots,a_p)'\in\mathbf{R}^p,\mu=(E\xi_1,\cdots,E\xi_p)'$.

所以 $\mu=E\xi$，即 $E\xi=(E\xi_1,\cdots,E\xi_p)'$.

这与先前随机向量的数学期望的定义完全一致.

又 $\forall\ a,b\in\mathbf{R}^p$ 来说，

$$\mathrm{Cov}(\langle a,\xi\rangle,\langle b,\xi\rangle)=\mathrm{Cov}(a'\xi,b'\xi)=a'\mathrm{Cov}(\xi,\xi)b$$

$$=a'\mathrm{Cov}(\xi)b=a'\Sigma b=\langle a,\Sigma b\rangle,\text{其中 }\mathrm{Cov}(\xi)=D\xi=\Sigma.$$

故新定义的协差算子与原先定义的完全一致.

线性模型是非常重要的统计模型；它包括了一系列统计模型，诸如线型回归模型，方差分析模型，协方差分析模型，方差分量模型等，线性模型的研究也为其它的统计技术，诸如试验设计、判别分析、生物统计、增长曲线分析、多元分析、时间序列分析等提供了基本的理论工具. 因此，线性模型无论在理论上还是在应用上都起着非常重要的作用.

§3. 随机向量二次型的均值与方差

为了以后叙述的方便，我们先将以后经常要用到的一些记号

列举如下：

n 维 $r.v.$ $y \triangleq (Y_1, \cdots, Y_n)'$，其中每一个 Y_i 为一维 $r.v.$；而 $n \times p$ 阶随机矩阵 $(r.m)$ $Y \triangleq (Y_{ij})$，其中每一 Y_{ij} 为一维 $r.v.$ "$\triangleq$"意为"定义为"或"记作". n 维 $r.v.$ $y = (Y_1, \cdots, Y_n)'$ 的数学期望定义为 $Ey = (EY_1, \cdots, EY_n)'$. $n \times p$ 阶随机矩阵 $(r.m.)$ $Y = (Y_{ij})$ 的数学期望定义为 $EY = (EY_{ij})$. n 维 $r.v.$ $y = (Y_1, \cdots, Y_n)'$ 与 m 维 $r.v.$ $z = (Z_1, \cdots, Z_m)'$ 的(互)协方差阵定义为 $\mathrm{Cov}(y, z) = E(y - Ey)(z - Ez)'$. 特别 y 的(自)协方差阵为 $\mathrm{Cov}(y, y) \triangleq \mathrm{Cov}(y) \triangleq D(y) = E(y - Ey)(y - Ey)'$. 易见，$\mathrm{Cov}(y, z) = Eyz - EyEz' = (\mathrm{Cov}(Y_i, Z_j))$. 在上面的记号下，有下面的

定理 1 设 Y 为 $r.m.$；x, y, z 为 $r.v.$；A, B, C, D 为常数矩阵(或向量)，则

i) $E(AYB + C) = A(EY)B + C$，特别有

$E(AY) = AEY$；

ii) $\mathrm{Cov}(y + C, z + D) = \mathrm{Cov}(y, z)$，特别有

$\mathrm{Cov}(y + c) = \mathrm{Cov}(y) \triangleq D(y) \triangleq \mathrm{Var}(y)$；

iii) $\mathrm{Cov}(x + y, z) = \mathrm{Cov}(x, z) + \mathrm{Cov}(y, z)$；

iv) $\mathrm{Cov}(Ay, Bz) = A\mathrm{Cov}(y, z)B'$，特别有

$\mathrm{Cov}(Ay) \triangleq D(Ay) = A(Dy)A'$.

在上面各式中，当然要求矩阵(或向量)间的运算有意义.

定理 2 若 y 为 n 维 $r.v.$，A 为 n 阶对称矩阵，则

$$Ey'Ay = \mathrm{tr}[A\mathrm{Cov}(y)] + Ey'AEy.$$

其中 $tr(\cdot)$ 为矩阵$(\cdot)$的迹，即矩阵$(\cdot)$的主对角元素之和.

证：
$$\begin{aligned} Ey'Ay &= E\,\mathrm{tr}(y'Ay) = E\,\mathrm{tr}(Ayy') = \mathrm{tr}(EAyy') \\ &= \mathrm{tr}(AEyy') = \mathrm{tr}\,A[\mathrm{Cov}(y) + EyEy'] \\ &= \mathrm{tr}[A\,\mathrm{Cov}(y)] + \mathrm{tr}(AEyEy') \\ &= \mathrm{tr}[A\mathrm{Cov}(y)] + Ey'AEy \quad \square \end{aligned}$$

习题：证明：若 $\underset{n\times p}{Y}=(y_1,\cdots,y_n)'$ 中 $y_1,\cdots,y_n\overset{iid}{\sim}(\mu,\Sigma)$，$\underset{n\times n}{A}=A'$. 则 $EY'AY=\Sigma tr(A)+EY'AEY$

定理 3 若 $y=\theta+\mathscr{E}$，其中 $\mathscr{E}=(\varepsilon_1,\cdots,\varepsilon_n)'$ 的诸分量独立，且满足 $E\varepsilon_i=0,E\varepsilon_i^2=\sigma^2,E\varepsilon_i^3=\mu_3,E\varepsilon_i^4=\mu_4$（存在，有限）又 $A'=\underset{n\times n}{A}\triangleq(a_{ij}),B'=\underset{n\times n}{B}\triangleq(b_{ij})$则

$$\begin{aligned}\mathrm{Cov}(y'Ay,y'By)&=(\mu_4-3\sigma^4)tr(AdiagB)+2\sigma^4tr(AB)\\&\quad+2\mu_3\theta'Ba+2\mu_3\theta'Ab+4\sigma^2\theta'AB\theta\\&=(\mu_4-3\sigma^4)a'b+2\sigma^4tr(AB)+2\mu_3\theta'(Ab+Ba)+\\&\quad 4\sigma^2\theta'AB\theta,\end{aligned}$$

特别有

$$\begin{aligned}\mathrm{Var}(y'Ay)&=(\mu_4-3\sigma^4)\mathrm{tr}(AdiagA)+2\sigma^4\mathrm{tr}(A^2)\\&\quad+4\mu_3\theta'Aa+4\sigma^2\theta'A^2\theta\\&=(\mu_4-3\sigma^4)a'a+2\sigma^4\mathrm{tr}(A^2)+4\mu_3\theta'Aa\\&\quad+4\sigma^2\theta'A^2\theta,\end{aligned}$$

其中 $\theta=(\theta_1,\cdots,\theta_n)'$，$\mathrm{diag}A=\begin{pmatrix}a_{11} & & 0\\ & \ddots & \\ 0 & & a_{nn}\end{pmatrix}$，$a=(a_{11},\cdots,a_{nn})'$ 为 $A\triangleq(a_{ij})$的主对角元素所构成的列向量，b 为 B 的主对角元素所构成的列向量.

而当 $A\theta=0,B\theta=0$ 时，

$$\begin{aligned}\mathrm{Cov}(y'Ay,y'By)&=(\mu_4-3\sigma^4)\mathrm{tr}(\mathrm{Adiag}B)+2\sigma^4\mathrm{tr}(AB)\\&=\gamma_2\sigma^4\mathrm{tr}(\mathrm{Adiag}B)+2\sigma^4\mathrm{tr}(AB),\end{aligned}$$

其中 $\gamma_2=\dfrac{E\varepsilon_i^4}{\sigma^4}-3$ 为峰度系数.

证： 由于 $y=\theta+\mathscr{E}$，故

$$\begin{aligned}\mathrm{Cov}(y'Ay,y'By)&=\mathrm{Cov}((\theta+\mathscr{E})'A(\theta+\mathscr{E}),(\theta+\mathscr{E})'B(\theta+\mathscr{E}))\\&=\mathrm{Cov}(\mathscr{E}'A\mathscr{E},\mathscr{E}'B\mathscr{E})+2\theta'A\mathrm{Cov}(\mathscr{E},\mathscr{E}'B\mathscr{E})\\&\quad+2\mathrm{Cov}(\mathscr{E}'A\mathscr{E},\mathscr{E})B\theta+4\theta'A\mathrm{Cov}(\mathscr{E},\mathscr{E})B\theta,\end{aligned}$$

其中

$$
\begin{aligned}
\mathrm{Cov}(\mathscr{E}'A\mathscr{E},\mathscr{E}'B\mathscr{E}) &= \mathrm{Cov}\Big(\sum_{i,j=1}^{n} a_{ij}\varepsilon_i\varepsilon_j,\ \sum_{k,l=1}^{n} b_{kl}\varepsilon_k\varepsilon_l\Big) = \sum_{i,jk,l}\sum a_{ij}b_{kl}\mathrm{Cov}\\
&\quad(\varepsilon_i\varepsilon_j,\varepsilon_k\varepsilon_l)\\
&= \sum_{i,jk,l}\sum a_{ij}b_{kl}(E\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l - E\varepsilon_i\varepsilon_j E\varepsilon_k\varepsilon_l)\\
&= \sum_{i,jk,l}\sum a_{ij}b_{kl}E\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l - \sum_{i,j}a_{ij}E\varepsilon_i\varepsilon_j\sum_{k,l}b_{kl}E\varepsilon_k\varepsilon_l\\
&= \sum_{i=1}^{n}a_{ii}b_{ii}\mu_4 + \sum_{i\neq k}a_{ii}b_{kk}\sigma^4 + \sum_{i\neq j}a_{ij}b_{ij}\sigma^4 + \sum_{i\neq j}a_{ij}b_{ji}\sigma^4\\
&\quad - \sum_{i=1}^{n}a_{ii}\sigma^2\sum_{k=1}^{n}b_{kk}\sigma^2\\
&= \mu_4 a'b + \sigma^4\Big(\sum_{i,k=1}^{n}a_{ii}b_{kk} + \sum_{i,j=1}^{n}a_{ij}b_{ij} + \sum_{i,j=1}^{n}a_{ij}b_{ij}\Big)\\
&\quad - \sigma^4\Big(\sum_{i=1}^{n}a_{ii}b_{ii} + \sum_{i=1}^{n}a_{ii}b_{ii} + \sum_{i=1}^{n}a_{ii}b_{ii}\Big) - \sigma^4\mathrm{tr}(\boldsymbol{A})\\
&\quad \cdot \mathrm{tr}(\boldsymbol{B})\\
&= (\mu_4 - 3\sigma^4)a'b + 2\sigma^4\mathrm{tr}(\boldsymbol{AB})\\
&= (\mu_4 - 3\sigma^4)\mathrm{tr}(A\mathrm{diag}B) + 2\sigma^4\mathrm{tr}(\boldsymbol{AB})
\end{aligned}
$$

其中用到

$$
E\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l = \begin{cases} \mu_4, i=j=k=l \\ \sigma^4, i=j\neq k=l,\text{或 } i=k\neq j=l,\text{或 } i=l\neq j=k \\ 0,\text{其它} \end{cases}
$$

而

$$
\begin{aligned}
\mathrm{Cov}(\mathscr{E}'\boldsymbol{A}\mathscr{E},\mathscr{E}) &= E\mathscr{E}'\boldsymbol{A}\mathscr{E}\mathscr{E}' - E\mathscr{E}'\boldsymbol{A}\mathscr{E}E\mathscr{E}' = E\mathscr{E}'\boldsymbol{A}\mathscr{E}\mathscr{E}' - 0\\
&= \Big(\sum_{i,j}a_{ij}E\varepsilon_i\varepsilon_j\varepsilon_1,\cdots,\sum_{i,j}a_{ij}E\varepsilon_i\varepsilon_j\varepsilon_n\Big)\\
&= (a_{11}\mu_3,\cdots,a_{nn}\mu_3) = \mu_3 a'
\end{aligned}
$$

$$
\mathrm{Cov}(\mathscr{E},\mathscr{E}'\boldsymbol{B}\mathscr{E}) = [\mathrm{Cov}(\mathscr{E}'\boldsymbol{B}\mathscr{E},\mathscr{E})]' = (\mu_3 b')' = \mu_3 b \cdot
$$

$$
\mathrm{Cov}(\mathscr{E},\mathscr{E}) = D(\mathscr{E}) = \sigma^2\boldsymbol{I}_n.
$$

代回整理即得所要的结果. □

定理 4 设 $y\sim N(\mu,\sigma^2 I_n),\sigma^2>0,\underset{n\times n}{\boldsymbol{A}}=A',\underset{n\times n}{\boldsymbol{B}}=\boldsymbol{B}'$. 则

$$
y'\boldsymbol{A}y \text{ 与 } y'\boldsymbol{B}y \text{ 独立} \Leftrightarrow \boldsymbol{BA}=0.
$$

为完成定理的证明,先来证明下面的

引理 若 $y\sim N_n(\mu,\Sigma)$，$\Sigma>0$，$A=A'$. 则 $y'Ay$ 的特征函数为

$$\psi_A(t)=|I-2\mathrm{i}tA\Sigma|^{-\frac{1}{2}}exp\{\mathrm{i}t\mu'A(I-2\mathrm{i}t\Sigma A)^{-1}\mu\}$$

$$=|I-2\mathrm{i}tA\Sigma|^{-\frac{1}{2}}exp\{\mathrm{i}t\mu'(I-2\mathrm{i}tA\Sigma)^{-1}A\mu\}$$

而 $$Ee^{iy'Ay}=|I-2\mathrm{i}A\Sigma|^{-\frac{1}{2}}exp\{\mathrm{i}\mu'A(I-2\mathrm{i}\Sigma A)^{-1}\mu\}$$

$$=|I-2\mathrm{i}A\Sigma|^{-\frac{1}{2}}exp\{\mathrm{i}\mu'(I-2\mathrm{i}A\Sigma)^{-1}A\mu\}.$$

证：先看 $\Sigma=I_n$ 的情况. 此时存在正交阵 Γ 使得

$$A=\Gamma\begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}\Gamma'，其中\ \lambda=\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_r\end{pmatrix}，\lambda_i\neq 0，i=1,\cdots,r.$$

$r=\mathrm{rk}(A)$.

于是 $\Gamma'y\triangleq z=(z_1,\cdots,z_n)'\sim N_n(\nu,I_n)$，其中 $\nu=\Gamma'\mu\triangleq(\nu_1,\cdots,\nu_n)'$. 从而

$z_j\sim N(\nu_j,1)$，$z_j^2\sim\chi^2(1,\nu_j^2)$（自由度为 1，非中心参数为 ν_j^2 的 χ^2－分布），

其特征函数为 $(1-2\mathrm{i}t)^{-\frac{1}{2}}exp\{\frac{\mathrm{i}t\nu_j^2}{1-2\mathrm{i}t}\}$. 从而

$$y'Ay=y'\Gamma\begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}\Gamma'y=z'\begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}z=\sum_{j=1}^r\lambda_jz_j^2\ 的特征函数为$$

$$\psi_A(t)=Ee^{ity'Ay}=\prod_{j=1}^r(1-2\mathrm{i}t\lambda_j)^{-\frac{1}{2}}exp\{\frac{\mathrm{i}t\lambda_j\nu_j^2}{1-2\mathrm{i}\lambda_jt}\}$$

$$=|I-2\mathrm{i}tA|^{-\frac{1}{2}}exp\{\mathrm{i}t\mu'A(I-2\mathrm{i}tA)^{-1}\mu\}$$

$$=|I-2\mathrm{i}tA|^{-\frac{1}{2}}exp\{\mathrm{i}t\mu'(I-2\mathrm{i}tA)^{-1}A\mu\},$$

其中用到

$$|I-2\mathrm{i}tA|=|I-2\mathrm{i}t\Gamma\begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}\Gamma'|=|I-2\mathrm{i}t\begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}|$$

$$=\prod_{j=1}^r(1-2\mathrm{i}t\lambda_j)，而\quad\sum_{j=1}^r\frac{\lambda_j\nu_j^2}{1-2\mathrm{i}\lambda_jt}=\sum_{j=1}^r\lambda_j\nu_j^2(1-2\mathrm{i}\lambda_jt)^{-1}$$