第3章 矩阵代数

3.1 引言

矩阵代数是多元回归（第 4 章）的关键，因此这里只介绍一些基本内容. 第 4 节讨论正定矩阵（positive definite matrix），并在第 5 节简要介绍正态分布. 矩阵是数字的一个矩形排列.（本书仅考虑实数矩阵.）例如，M 为一个 3×2 矩阵，有 3 行 2 列，而 b 是一个 2×1 的列向量（column vector）：

$$M=\begin{bmatrix}3&-1\\2&-1\\-1&4\end{bmatrix},\quad b=\begin{bmatrix}3\\-3\end{bmatrix}.$$

M 的第 ij 个元素记为 M_{ij}，例如 $M_{32}=4$，类似地，$b_2=-3$. 矩阵可以（逐元素）乘以一个数量. 同阶矩阵可以相加（也是逐元素）. 比如

$$2\times\begin{bmatrix}3&-1\\2&-1\\-1&4\end{bmatrix}=\begin{bmatrix}6&-2\\4&-2\\-2&8\end{bmatrix},$$

$$\begin{bmatrix}3&-1\\2&-1\\-1&4\end{bmatrix}+\begin{bmatrix}3&2\\4&-1\\-1&1\end{bmatrix}=\begin{bmatrix}6&1\\6&-2\\-2&5\end{bmatrix}.$$

一个 $m\times n$ 的矩阵 A 可以被一个 $n\times p$ 的矩阵 B 乘. 乘积为一个 $m\times p$ 矩阵，其第 ik 个元素为 $\sum_j A_{ij}B_{jk}$. 例如

$$Mb=\begin{bmatrix}3\times3+(-1)\times(-3)\\2\times3+(-1)\times(-3)\\(-1)\times3+4\times(-3)\end{bmatrix}=\begin{bmatrix}12\\9\\-15\end{bmatrix}.$$

矩阵乘法**不可**交换. 这乍看上去有些难以捉摸，但你会习惯的. 练习 1～2（下面）将给出一点解释.

矩阵 $0_{m\times n}$ 为一个 $m\times n$ 的矩阵，其所有元素都为 0. 例如

$$0_{2\times3}=\begin{bmatrix}0&0&0\\0&0&0\end{bmatrix}.$$

一个 $m\times m$ 的单位矩阵写成 I_m 或 $I_{m\times m}$. 这个矩阵在对角线上都是 1，而对角线外都是 0：

$$I_{3\times3}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}.$$

如果 A 是 $m\times n$，那么 $I_{m\times m}\times A=A=A\times I_{n\times n}$.

一个 $m\times n$ 矩阵 A 能够被“转置，”结果是一个 $n\times m$ 矩阵，用 A' 或 A^{T} 表示，例如，

$$\begin{bmatrix} 3 & -1 \\ 2 & -1 \\ -1 & 4 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 3 & 2 & -1 \\ -1 & -1 & 4 \end{bmatrix}.$$

如果 $A'=A$，则 A 是对称的（symmetric）.

如果 u 和 v 是 $n\times 1$ 的列向量，内积（inner product）或点积（dot product）为 $u\cdot v=u'\times v$. 如果这是 0，那么 u 和 v 是正交的（orthogonal）：记为 $u\perp v$. u 的模（norm）或长度（length）为 $\|u\|$，这里 $\|u\|^2=u\cdot u$. 人们常将模或长度写成 $|u|$ 而不是 $\|u\|$. 内积 $u\cdot v$ 等于 u 的长度乘上 v 的长度再乘以两个向量的夹角余弦. 如果 $u\perp v$，则夹角为 $90°$，而 $\cos(90°)=0$.

对于正方矩阵，迹（trace）为对角线元素的和：

$$\operatorname{trace}\begin{bmatrix} 1 & 2 \\ 5 & 3 \end{bmatrix} = 4.$$

练习组 A

1. 假定 A 为 $m\times n$，B 为 $n\times p$. 对于 i 和 j，$1\leqslant i\leqslant m$，$1\leqslant j\leqslant p$，令 r_i 为 A 的第 i 行，令 c_j 为 B 的第 j 列. r_i 的大小（即维数）是多少？c_j 呢？$r_i\times c_j$ 是如何和 $A\times B$ 的第 ij 个元素相关联的？

2. 假定 A 为 $m\times n$，而 u，v 为 $n\times 1$，α 为一个数量. 表明 $Au\in R^m$，$A(\alpha u)=\alpha Au$，$A(u+v)=Au+Av$. 一般来说，A 是一个从 R^n 到 R^m 的线性映射，这里 R^n 为 n 维欧式空间.（例如，R^1 是直线，R^2 是平面.）

3. 如果 A 为 $m\times n$ 矩阵，核对：$A+0_{m\times n}=0_{m\times n}+A=A$.

对练习 4 和 5，令

$$M=\begin{bmatrix} 3 & -1 \\ 2 & -1 \\ 1 & -4 \end{bmatrix}.$$

4. 表明 $I_{3\times 3}M=M=MI_{2\times 2}$.

5. 计算 $M'M$ 及 MM'. 分别求 $M'M$ 及 MM' 的迹.

6. 计算下面定义的 u 和 v 的长度. 这两个向量正交吗？计算外积（outer product）$u\times v'$. 该外积的迹是多少？

$$u=\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix},\qquad v=\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}.$$

3.2 行列式及逆

为了得到回归的估计值及其标准误差，需要矩阵求逆. 一种求逆方法是通过行列式（determinant）. 正方矩阵的行列式是通过归纳法计算的：

$$\det(4)=4,\quad \det\begin{bmatrix} 1 & 2 \\ 5 & 3 \end{bmatrix}=(1\times 3)-(2\times 5)=-7,$$

$$\det\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\\ 0 & 1 & 1\end{pmatrix} = 1\times\det\begin{pmatrix}3 & 1\\ 1 & 1\end{pmatrix} - 2\times\det\begin{pmatrix}2 & 3\\ 1 & 1\end{pmatrix} + 0\times\det\begin{pmatrix}2 & 3\\ 3 & 1\end{pmatrix}$$
$$= 1\times(3-1) - 2\times(2-3) + 0\times(2-9) = 4.$$

这里，我们沿着第一列往下，把经过目前位置的行和列去掉来得到更小的矩阵．这些行列式都有一个额外的符号，它们交替地为＋和－．有额外符号的行列式称为余子式（cofactor）．比如，对于一个 4×4 矩阵，额外的符号为

$$\begin{pmatrix}+ & - & + & -\\ - & + & - & +\\ + & - & + & -\\ - & + & - & +\end{pmatrix}.$$

一个矩阵的行列式为 $\sum_{i=1}^{n} a_{i1}c_{i1}$，这里 a_{ij} 为矩阵的第 ij 个元素，而 c_{ij} 为余子式．（注意：行列式既有自己的符号，也有上面表明的额外的符号．）实际上，你能够用任何行或列，而不一定非用第一列来计算行列式．由于符号问题，人们常用 $|A|$ 而不是 $\det A$ 来表示 A 的行列式．

令 $v_1, v_2, \cdots, v_k$ 为 $n\times1$ 向量．如果 $c_1v_1+c_2v_2+\cdots+c_kv_k=0_{n\times1}$ 意味着 $c_1=\cdots=c_k=0$，那么这些向量是线性独立的[㊀]（linearly independent）．一个矩阵的秩（rank）是矩阵中线性独立列的个数（或者独立行的个数，它们一定相同）．如果 $n>p$，$n\times p$ 矩阵 X 的秩为 p，那么称它满秩（full rank）；否则，X 不满秩（rank deficient）．一个 $n\times n$ 矩阵 A 满秩，当且仅当 $\det A\neq 0$．那么，该矩阵有一个逆 A^{-1}：

$$A\times A^{-1} = A^{-1}\times A = I_{n\times n}.$$

这样的矩阵是可逆的（invertible）或者非奇异的（non-singular）．逆是唯一的，这可由存在性证明．相反地，如果 A 是可逆的，那么 $\det(A)\neq0$ 而且 A 的秩为 n．

逆能够如下计算：

$$A^{-1} = \mathrm{adj}(A)/\det(A),$$

这里 adj(A) 为余子式矩阵的转置．（这是经典的伴随（矩阵）（classical adjoint）．）例如

$$\mathrm{adj}\begin{pmatrix}1 & 2\\ 5 & 3\end{pmatrix} = \begin{pmatrix}3 & -2\\ -5 & 1\end{pmatrix},$$

$$\mathrm{adj}\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\\ 0 & 1 & 1\end{pmatrix} = \begin{pmatrix}a & b & c\\ d & e & f\\ g & h & i\end{pmatrix},$$

这里

$$a = \det\begin{pmatrix}3 & 1\\ 1 & 1\end{pmatrix},\quad b = -\det\begin{pmatrix}2 & 3\\ 1 & 1\end{pmatrix},\quad c = \det\begin{pmatrix}2 & 3\\ 3 & 1\end{pmatrix}\cdots$$

㊀ 也称为线性无关．——译者注．

练习组 B

对于下面练习 1～7，令

$$A=\begin{pmatrix}1 & 2\\5 & 3\end{pmatrix},\quad B=\begin{pmatrix}1 & 2 & 3\\2 & 3 & 1\\0 & 1 & 1\end{pmatrix},\quad C=\begin{pmatrix}1 & 2\\2 & 4\\3 & 6\end{pmatrix}.$$

1. 求 $\mathrm{adj}(B)$. 这里利用定义来计算，以后用计算机来做这类事情.

2. 表明 $A\times\mathrm{adj}\,A=\mathrm{adj}\,A\times A=\det(A)\times I_n$. 对 B 再做一遍. 在每种情况，n 是多少？

3. 求 A 的秩和迹. 对 B 再做一遍.

4. 求 C 的秩.

5. 如果可能，求 C 的迹和行列式. 如果不行，为什么？

6. 如果可能，求 A^2. 如果不行，为什么？（提示：$A^2=A\times A$.）

7. 如果可能，求 C^2. 如果不行，为什么？

8. 如果 M 是 $m\times n$，N 是 $m\times n$，表明 $(M+N)'=M'+N'$.

9. 假定 M 是 $m\times n$，N 是 $n\times p$.

(a) 证明 $(MN)'=N'M'$.

(b) 假定 $m=n=p$，而且 M，N 两个都是可逆的. 证明：$(MN)^{-1}=N^{-1}M^{-1}$ 及 $(M')^{-1}=(M^{-1})'$.

10. 假定 X 是 $n\times p(p\leqslant n)$ 矩阵. 如果 X 的秩为 p，表明 $X'X$ 的秩为 p，并证明其逆命题也成立. 提示：假定 X 的秩为 p，而且 c 为 $p\times 1$ 向量，那么 $X'Xc=0_{p\times 1}\Rightarrow c'X'Xc=0\Rightarrow\|Xc\|^2=0\Rightarrow Xc=0_{n\times 1}$.

注意. 矩阵 $X'X$ 为 $p\times p$ 矩阵. $X'X$ 的秩为 p 当且仅当它是可逆的. 符号“$\Rightarrow$”代表“意味着.”

11. 如果 A 是 $m\times n$ 矩阵，B 是 $n\times m$ 矩阵，表明 $\mathrm{trace}(AB)=\mathrm{trace}(BA)$. 提示：$AB$ 的第 ii 个元素为 $\sum_j A_{ij}B_{ji}$，而 BA 的第 jj 个元素为 $\sum_i B_{ji}A_{ij}$.

12. 如果 u 和 v 是 $n\times 1$，表明 $\|u+v\|^2=\|u\|^2+\|v\|^2+2u\cdot v$.

13. 如果 u 和 v 是 $n\times 1$，表明 $\|u+v\|^2=\|u\|^2+\|v\|^2$ 当且仅当 $u\perp v$.（这是 n 维的勾股定理[1].）

14. 假定 X 是 $n\times p$ 矩阵，并且矩阵秩 $p<n$. 假定 Y 是 $n\times 1$. 令 $\hat{\beta}=(X'X)^{-1}X'Y$ 以及 $e=Y-X\hat{\beta}$.

(a) 表明 $X'X$ 为 $p\times p$，而 $X'Y$ 为 $p\times 1$.

(b) 表明 $X'X$ 为对称的. 提示：见练习 9(a).

(c) 表明 $X'X$ 为可逆的. 提示：见练习 10.

(d) 表明 $(X'X)^{-1}$ 为 $p\times p$，因而 $\hat{\beta}=(X'X)^{-1}X'Y$ 为 $p\times 1$.

(e) 表明 $(X'X)^{-1}$ 为对称的. 提示：见练习 9(b).

(f) 表明 $X\hat{\beta}$ 及 $e=Y-X\hat{\beta}$ 为 $n\times 1$.

---

[1] 西方称 Pythagoras 定理. ——译者注.

(g) 表明 $X'X\hat{\beta}=X'Y$，并因此 $X'e=0_{p\times 1}$.

(h) 表明 $e\perp X\hat{\beta}$，因而 $\|Y\|^2=\|X\hat{\beta}\|^2+\|e\|^2$.

(i) 如果 γ 为 $p\times 1$，表明 $\|Y-X\gamma\|^2=\|Y-X\hat{\beta}\|^2+\|X(\hat{\beta}-\gamma)\|^2$. 提示：$Y-X\gamma=Y-X\hat{\beta}+X(\hat{\beta}-\gamma)$.

(j) 表明 $\|Y-X\gamma\|^2$ 在 $\gamma=\hat{\beta}$ 时达到最小值.

(k) 如果 $\tilde{\beta}$ 为 $p\times 1$，而且 $Y-X\tilde{\beta}\perp X$，表明 $\tilde{\beta}=\hat{\beta}$. 符号：$v\perp X$ 意味着 v 与 X 的每一列正交. 提示：$X'(Y-X\tilde{\beta})$ 是什么？

(l) XX'是可逆的吗？提示：根据 $p<n$ 的假定. 你能够找到一个 $n\times 1$ 向量 $c\neq 0_{n\times 1}$ 使得 $c'X=0_{1\times p}$ 吗？

(m) 是否 $(X'X)^{-1}=X^{-1}(X')^{-1}$？

注意. $\hat{\beta}$ 称为"OLS 估计"，这里 OLS 为"通常最小二乘"(ordinary least squares) 的缩写. 这个练习发展了大量关于 OLS 估计量的定理. 简述其几何意义：$X'e=0_{p\times 1}$ 意味着 e 正交或垂直于 X 的每一列. 因此 $\hat{Y}=X\hat{\beta}$ 为 Y 到 X 列上的投影，而且是在列空间中最接近 Y 的一点. (j) 部分是多元回归的 Gauss 定理.

15. 在练习 14 中，假定 $p=1$，这样，X 为一个列向量. 表明 $\hat{\beta}=X\cdot Y/\|X\|^2$.

16. 在练习 14 中，假定 $p=1$ 而且 X 为一列"1". 表明 $\hat{\beta}$ 为 Y 的均值. 这和练习 2B12 (c)，即第 2 章练习组 B 练习 12 的 (c) 部分，有什么关系？

17. 这个练习解释了在练习 14 中计算 $\hat{\beta}$ 的一个逐步过程. 虽然有提示，但还有些事情要做. 令 M 为 X 的前 $p-1$ 列，因此 M 为 $n\times(p-1)$. 令 N 为 X 的最后一列，因此 N 为 $n\times 1$.

(ⅰ) 令 $\hat{\gamma}_1=(M'M)^{-1}M'Y$ 及 $f=Y-M\hat{\gamma}_1$.

(ⅱ) 令 $\hat{\gamma}_2=(M'M)^{-1}M'N$ 及 $g=N-M\hat{\gamma}_2$.

(ⅲ) 令 $\hat{\gamma}_3=f\cdot g/\|g\|^2$ 及 $e=f-g\hat{\gamma}_3$.

表明 $e\perp X$. (提示：先核对 $f\perp M$ 及 $g\perp M$.) 最后表明

$$\hat{\beta}=\begin{pmatrix}\hat{\gamma}_1-\hat{\gamma}_2\hat{\gamma}_3\\ \hat{\gamma}_3\end{pmatrix}.$$

注意. 这个过程概括为：(ⅰ) 在 M 上对 Y 回归，(ⅱ) 在 M 上对 N 回归，(ⅲ) 在第二组残差上对第一组残差回归.

18. 假定 u，v 为 $n\times 1$，没有一个是全为 0 的. $u\times v$ 的秩是多少？

3.3 随机向量

令 $U=\begin{pmatrix}U_1\\U_2\\U_3\end{pmatrix}$ 为 3×1 的随机变量列向量. 那么 $E(U)=\begin{pmatrix}E(U_1)\\E(U_2)\\E(U_3)\end{pmatrix}$ 为 3×1 的数值列向量. 另一方面，$\mathrm{cov}(U)$ 为 3×3 实数矩阵：

$$\mathrm{cov}(U)=E\left\{\begin{pmatrix}U_1-E(U_1)\\U_2-E(U_2)\\U_3-E(U_3)\end{pmatrix}\begin{pmatrix}U_1-E(U_1) & U_2-E(U_2) & U_3-E(U_3)\end{pmatrix}\right\}.$$

因此，cov 不仅仅作用于数据，而且作用于随机向量（“cov” 为协方差（covariance）的缩写）. 同样定义可用于任意大小的向量.

人们有时对随机变量做相关：比如，在 U_1 和 U_2 之间的相关为 $\mathrm{cov}(U_1, U_2)/\sqrt{\mathrm{var}(U_1)\mathrm{var}(U_2)}$.

练习组 C

1. 表明 $\mathrm{cov}(U)$ 的第 1，1 个元素等于 $\mathrm{var}(U_1)$，而其第 2，3 个元素等于 $\mathrm{cov}(U_2, U_3)$.

2. 表明 $\mathrm{cov}(U)$ 是对称的.

3. 如果 A 是固定的（即非随机的）$n\times 3$ 矩阵，而 B 是固定的 $1\times m$ 矩阵，表明

$$E(AUB) = AE(U)B.$$

4. 表明 $\mathrm{cov}(AU)=A\mathrm{cov}(U)A'$.

5. 如果 c 是固定的 3×1 向量，表明 $\mathrm{var}(c'U)=c'\mathrm{cov}(U)c$ 及 $\mathrm{cov}(U+c)=\mathrm{cov}(U)$.

评论. 如果 V 是一个 $n\times 1$ 随机向量，C 是一个固定的 $m\times n$ 矩阵，而且 D 是一个固定的 $m\times 1$ 向量，那么 $\mathrm{cov}(CV+D)=C\mathrm{cov}(V)C'$.

6. 在 $\overline{U}=(U_1+U_2+U_3)/3$ 和 $E(U)$ 之间的区别是什么？

7. 假定 ξ 及 ζ 为两个 7×1 的随机向量. 如果 $\xi'\zeta=0$，ξ 和 ζ 独立吗？那么逆命题呢：如果 ξ 和 ζ 独立，$\xi'\zeta=0$ 吗？

8. 假定 ξ 及 ζ 为两个随机变量，满足 $E(\xi)=E(\zeta)=0$. 表明 $\mathrm{var}(\xi)=E(\xi^2)$ 以及 $\mathrm{cov}(\xi, \zeta)=E(\xi\zeta)$.

注意. 更一般地，$\mathrm{var}(\xi)=E(\xi^2)-[E(\xi)]^2$ 及 $\mathrm{cov}(\xi, \zeta)=E(\xi\zeta)-E(\xi)E(\zeta)$.

9. 假定 ξ 为 $n\times 1$ 随机向量，$E(\xi)=0$. 表明 $\mathrm{cov}(\xi)=E(\xi\xi')$.

注意. 一般地，$\mathrm{cov}(\xi)=E(\xi\xi')-E(\xi)E(\xi')$ 及 $E(\xi')=[E(\xi)]'$.

10. 假定对于 $i=1, \cdots, n$，ξ_i，ζ_i 为随机变量. 作为一对随机变量，它们关于 i 是独立同分布的. 令 $\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$，对 ζ 也类似. 判断下面论述的对错，并给出解释：

(a) $\mathrm{cov}(\xi_i, \zeta_i)$ 对每个 i 都相同.

(b) $\mathrm{cov}(\xi_i, \zeta_i)=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})(\zeta_i-\bar{\zeta})$.

11. 随机变量 X 在直线上有密度 f，σ 和 μ 是实数. $\sigma X+\mu$ 的密度是什么？X^2 的密度呢？回忆：如果 X 有密度 f，那么 $P(X<x)=\int_{-\infty}^{x}f(u)\mathrm{d}u$.

3.4 正定矩阵

这一节的内容将在讨论广义最小二乘时（5.3 节）会用到. 证明的细节将不会涉及. 一个 $n\times n$ 的正交矩阵（orthogonal matrix，或 unitary）R 有 $R'R=I_{n\times n}$. 一定有 $RR'=I_{n\times n}$. 在几何上，R 是个旋转，它保持夹角和距离不变，R 还能保持某些方向. 一个对角（diagonal）矩阵 D 是个方阵，其主对角线外的元素均为零：例如 D_{11} 和 D_{22} 可能为非零，但 $D_{12}=D_{21}=0$. 一个 $n\times n$ 矩阵 G 称为非负定的（non-negative definite），如果满足

（ⅰ）G 为对称的，而且

（ⅱ）对任何 n 维向量 x，$x'Gx\geqslant 0$.

矩阵 G 称为正定的（positive definite），如果对任何 n 维向量 x，除了 $x=0_{n\times 1}$ 之外，都有 $x'Gx\geqslant 0$.（非负定矩阵也称为半正定的（positive semi-definite).）

定理 1 矩阵 G 为非负定的当且仅当有一个对角线矩阵 D，它的对角线元素非负，并且存在一个正交矩阵 R，使得 $G=RDR'$. 矩阵 G 为正定的当且仅当 D 的所有对角线元素都是正的.

R 的列为 G 的特征向量（eigenvector），而 D 的对角线元素为特征值（eigenvalue）. 例如，如果 c 为 R 的第一列，而且 $\lambda=D_{11}$，那么 $Gc=c\lambda$.（这是因为 $GR=RD$.）由定理 1 可得，一个非负定的 G 有一个非负定的平方根，$G^{1/2}=RD^{1/2}R'$，这里 D 的平方根是逐元素取平方根. 一个正定的 G 有一个正定的逆 $G^{-1}=RD^{-1}R'$.（看下面练习.）如果 G 为非负定的而不是正定的，即对某个 $x\neq 0$，$x'Gx=0$，那么 G 是不可逆的. 定理 1 是"谱定理"（spectral theorem）的一个初等形式.

练习组 D

1. 下面哪个矩阵是正定的？非负定的？

$$\begin{pmatrix}2 & 0\\ 0 & 1\end{pmatrix}\qquad\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix}\qquad\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\qquad\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$$

提示：计算出

$$[u,v]\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}=[u,v]\left[\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}\right].$$

2. 假定 X 为 $n\times p$ 矩阵，秩为 $p\leqslant n$.

(a) 表明 $X'X$ 为 $p\times p$ 正定的. 提示：如果 c 为 $p\times 1$，那么 $c'X'Xc$ 是多少？

(b) 表明 XX' 为 $n\times n$ 非负定的.

对于练习 3～6，假定 R 为 $n\times n$ 正交矩阵，D 为 $n\times n$ 对角矩阵，对所有的 i，$D_{ii}>0$. 令 $G=RDR'$. 不用定理 1，直接做练习.

3. 表明对任意 $n\times 1$ 向量 x，$\|Rx\|=\|x\|$.

4. 表明 D 和 G 为正定的.

5. 令 $\sqrt{D}$ 为 $n\times n$ 矩阵，其第 ij 个元素为 $\sqrt{D_{ij}}$. 表明 $\sqrt{D}\sqrt{D}=D$. 另外证明 $R\sqrt{D}R'R\sqrt{D}R'=G$.

6. 令 D^{-1} 的非对角线元素均为 0，其第 ii 个元素为 D_{ii}^{-1}. 表明 $D^{-1}D=I_{n\times n}$ 及 $RD^{-1}R'G=I_{n\times n}$.

7. 假定 G 为正定的. 表明：

(a) G 是可逆的，而且 G^{-1} 是正定的.

(b) G 有正定的平方根 $G^{1/2}$.

(c) G^{-1} 有正定的平方根 $G^{-1/2}$.

8. 令 U 为 3×1 随机向量. 表明：cov(U) 为非负定的，而且除非有一个 3×1 固定的（即非随机的）向量 c，使得以概率 1 有 $c'U=c'E(U)$，则 cov(U) 是正定的. 提示：你能够从 cov(U) 计算 var$(c'U)$ 吗？如果这个提示不够，试试 $E(U)=0_{3\times 1}$ 的情况. **评论.** 如果以概率 1 有 $c'U=c'E(U)$，则 $U-E(U)$ 聚集在一个固定的超平面上.

3.5 正态分布

这是一个快速的回顾，将不会给出证明．如果一个随机变量 X 有均值 μ 和方差 σ^2 的正态分布，那么说它是 $N(\mu,\sigma^2)$．X 的密度为

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right],(\text{这里符号 }\exp(t)=e^t).$$

如果 X 为 $N(\mu,\sigma^2)$，那么 $(X-\mu)/\sigma$ 为 $N(0,1)$，即 $(X-\mu)/\sigma$ 为标准正态的（standard normal）．标准正态密度为

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right).$$

如果随机变量 X_1，…，X_n 的所有线性组合都有正态分布，则它们为联合正态的（jointly normal）．如果 X_1，X_2 是独立正态变量，则它们是联合正态的，这是因为对于任意实数 a_1 和 a_2，$a_1X_1+a_2X_2$ 是正态分布．后面一两个例子将涉及联合正态变量，而下面的定理将是有帮助的．（如果你想要构造正态变量，看下面练习 1 的方法．）

定理 2 联合正态随机变量的分布是由均值向量 α 和协方差矩阵 G 确定的，后者必须是非负定的．如果 G 是正定的，那么随机变量在 x 的密度为

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n\frac{1}{\sqrt{\det G}}\exp\left[-\frac{1}{2}(x-\alpha)'G^{-1}(x-\alpha)\right].$$

对任何一对随机变量 X_1，X_2，无论正态与否，如果 X_1 和 X_2 独立，那么 $\mathrm{cov}(X_1,X_2)=0$．虽然可能设计出反例，但其逆命题一般不对．对于正态随机变量，逆命题是对的：如果 X_1，X_2 为联合正态，而且 $\mathrm{cov}(X_1,X_2)=0$，那么 X_1 和 X_2 独立．

中心极限定理． 对于大的样本，和（或均值）的概率分布将接近正态分布．更正式地，假定 X_1，X_2，…为独立同分布的，有 $E(X_i)=\mu$ 及 $\mathrm{var}(X_i)=\sigma^2$．那么 $S_n=X_1+X_2+\cdots+X_n$ 有期望值 $n\mu$ 及方差 $n\sigma^2$．为标准化，减去期望值并除以标准误差（方差的平方根）：

$$Z_n=\frac{S_n-n\mu}{\sigma\sqrt{n}}.$$

中心极限定理说，如果 n 很大，Z_n 的分布接近标准正态．例如：

$$P\{|S_n-n\mu|<\sigma\sqrt{n}\}=P\{|Z_n|<1\}\to\frac{1}{\sqrt{2\pi}}\int_{-1}^{1}\exp\left(-\frac{1}{2}x^2\right)dx\doteq 0.682\,7.$$

该定理有许多推广．例如，具有不同分布的独立变量的和，只要和中的每一项仅是总和的一小部分，则近似于正态分布．还有对随机向量的中心极限定理．Feller（1971）对此有认真的论述与证明，其他概率课本也一样．

术语．（ⅰ）对称性是正定矩阵定义中固有的．（ⅱ）正交矩阵有正交的行，而且每一行的长度为 1．这些行称为"标准正交化的"（orthonormal）．类似的结论可用于列．（ⅲ）"多元正态"为联合正态的同义词．（ⅳ）虽然会产生混淆，"联合正态"有时被简化为"正态"．（ⅴ）"渐近地"意味着样本量变得很大，即在和中的项数目变得很大．

练习组 E

1. 假定 G 为 $n\times n$ 非负定的，α 为 $n\times 1$.
   (a) 求一个均值为 0，协方差 $\mathrm{cov}(U)=G$ 的正态随机变量的 $n\times 1$ 向量 U. 提示：令 V 为独立 $N(0, 1)$ 变量的 $n\times 1$ 向量，然后令 $U=G^{1/2}V$.
   (b) 你如何修正构造过程使得 $E(U)=\alpha$?
2. 假定 R 为一个正交的 $n\times n$ 矩阵. 如果 U 为 IID $N(0, \sigma^2)$ 变量的一个 $n\times 1$ 向量，表明 RU 为 IID $N(0, \sigma^2)$ 变量的一个 $n\times 1$ 向量. 提示：$E(RU)$ 是什么？$\mathrm{cov}(RU)$ 呢？（“IID”为“独立同分布”的缩写.）
3. 假定 ξ 和 ζ 为两个随机变量，如果 $E(\xi\zeta)=E(\xi)E(\zeta)$，$\xi$ 和 ζ 是独立的吗？那么逆命题：如果 ξ 和 ζ 是独立的，是否 $E(\xi\zeta)=E(\xi)E(\zeta)$?
4. 如果 U 和 V 为随机变量，表明 $\mathrm{cov}(U, V)=\mathrm{cov}(V, U)$ 及 $\mathrm{var}(U+V)=\mathrm{var}(U)+\mathrm{var}(V)+2\mathrm{cov}(U, V)$. 提示：$[(U-\alpha)+(V-\beta)]^2$ 是什么？
5. 假定 ξ 和 ζ 为联合正态变量，$E(\xi)=\alpha$，$\mathrm{var}(\xi)=\sigma^2$，$E(\zeta)=\beta$，$\mathrm{var}(\zeta)=\tau^2$，$\mathrm{cov}(\xi, \zeta)=\rho\sigma\tau$. 求 $\xi+\zeta$ 的均值和方差. $\xi+\zeta$ 是正态的吗？

评论. 练习 6～8 是为下一章做准备的. 练习 6 被（例如）Freedman-Pisani-Purves（2007）的第 18 章所涵盖，练习 7 和 8 被其第 20～21 章所涵盖.

6. 抛一枚硬币 1000 次. 利用中心极限定理来近似得到 475～525 次（包括端点）正面的机会.
7. 一个盒子有红色和蓝色的弹子. 红色的比例 p 未知. 随机地抽取 250 个弹子，每次取完都放回，结果有 102 个是红色的. 估计 p. 再对你的估计给出一个标准误差.
8. 令 $\hat{p}$ 为练习 7 中的估计值.
   (a) $\hat{p}$ 和 p 之间的差别大约有多大？
   (b) 你能够找到 p 的近似的 95%置信区间吗？
9. “误差函数” Ψ 定义如下：
$$\Psi(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-u^2)\,\mathrm{d}u.$$
   表明 Ψ 是 $|W|$ 的分布函数，这里 W 为 $N(0, \sigma^2)$. 求出 σ^2. 如果 Z 为 $N(0, 1)$，如何从 Ψ 计算 $P(Z<x)$?
10. 如果 U，V 为 IID $N(0, 1)$，表明 $(U+V)/\sqrt{2}$，$(U-V)/\sqrt{2}$ 为 IID $N(0, 1)$.

3.6 关于矩阵代数的书

Blyth TS，Robertson EF（2002）. *Basic Linear Algebra*. 2nd ed. Springer. 清楚，数学味的.

Strang G（2005）. *Linear Algebra and Its Applications*. 4th ed. Brooks Cole. 爱它或者恨它.

Mayer CD（2001）. *Matrix Analysis and Applied Linear Algebra*. SIAM. 比较通常类型的教科书.

Lax PD（2007）. *Linear Algebra and Its Applications*. 2nd ed. Wiley. 研究生水平的教材.