高等计量经济学基础

缪柏其　叶五一　编著

内容简介

高等计量经济学是金融工程的基础课程，本书从回归、时间序列、面板数据介绍了有关基本思想和内容，由于属性数据在金融数据的分析中大量出现，我们也简要介绍了有关属性数据相关和建模的一些内容。每章结合金融工程研究中的问题来介绍和解释有关的基本概念和内容，并配有一定量的习题。

本书可作为金融工程硕士研究生的教学用书，也可作为从事经济学、金融工程、证和基金等从业人员了解有关理论和应用的入门参考书。

目 录

前　言

自挪威经济学家 R. Frisch 1926 年提出经济计量学 (Econometriic) 以来至今已有 80 多年的历史了。它起源于对经济问题的定量研究。我国自改革开放以来学习西方国家的学者对经济学的定量研究，也加速了对经济计量学的教学和研究。在这个大背景下，我们也在 1995 年开办了应用经济学专业下的金融学，开始了计量经济学的教学与研究。

计量经济学是以数理统计和数学为工具对经济问题进行研究。因此，这是一门交叉学科，它不是简单地等于经济学和数理统计的叠加，而是一种交叉和相互融合的学科。它对定量研究经济和金融的重要性是不言而喻的，国内外的专家学者在文献中早已进行了大量的阐述。现在的问题是如何进行这方面的教学。一般在本科生和研究生阶段都会开设这门课程，关键是如何界定这两个阶段的教学内容和深浅，特别是研究生这个阶段。由于所研究的数据的规模越来越大，计算机硬件容量和速度的更新周期越来越短，数理统计学中有关统计方法和统计工具越来越多，可用来处理数据的软件也越来越多，计量经济学的研究范围也越来越大，所用到的统计方法、统计工具和计算机软件也越来越多。那么什么是学生在研究生阶段必须要掌握的？什么是应该了解的？国内学者也在不断地探索着这个问题。本教材是我们对这门课程在研究生阶段教学的一种尝试。

计量经济学中研究的变量一般而言都是随机的，有的变量还是定性变量，这是比较符合实际的。我们的任务就是用统计方法根据数据来建模，然后从所建的模型来解释经济变量之间的关系。由于经济是一个动态的复杂系统，很难用一种模型来拟合经济系统，所以我们就要尝试用不同的思路、不同的模型来尽可能好地拟合实际数据，这就需要计量经济学家要掌握较宽和较深的经济学和统计学的知识。这是我们对研究生阶段学生要求的基本认识，也是本书的一种定位。

在本书中，我们分为五个部分，第一部分是有关的数学和统计基础。为了能够对后面章节和文献中的定量统计方法有个基本的了解，一定的数学和统计学知识，特别是线性代数和多元统计的知识是必不可少的。而这部分内容对大部分研究生而言，在本科没有系统学过，或没有学到。所以我们用两章的篇幅来综述在今后章节中要用到的知识点。第二部分是有关回归分析的，包括建模和参数和非参数估计，这是计量经济学研究的基本方法之一。在这部

分我们用投影这样比较直观的叙述来介绍线性和非线性回归的内容。第三部分是有关时间序列的。经济是一个动态系统，就要用到动态的模型来拟合它，时间序列是一种有力的工具，我们从时间域和频率域两个方面简述了有关的内容，介绍了有用的差分方程。第四部分是面板数据分析。回归是在横截面分上分析经济变量之间的因果关系，时间序列主要是研究变量随时间的变化规律，没有做横截面分析，面板数据分析则综合了这两种分析，由于同时考虑经济变量和动态变化，所以能考虑更广的经济问题，当然也加大了分析的难度。第五部分是关于因变量为属性数据时的建模问题。

本教材虽然在研究生教学中使用多年，但是一定还有不少遗憾之处，请读者尽可能给我们指出，以便我们能不断完善我们的教材。

缪柏其　　叶五一

2010 年 8 月 08 日

于中国科学技术大学

第一章　线性代数和矩阵基本知识

§1.1　线性代数基本知识

§1.1.1　向量空间

我们知道，线性空间是线性代数的理论，线性空间是线性代数研究的最基本的几何对象，我们最关注的是有关线性空间的投影理论。矩阵是研究线性空间中各种几何问题的最有效的工具。我们假定已有线性代数的基本知识，所以本节我们仅对有关概念作一个简要介绍。

定义1.1. 设 S 是一个空间，有零元素 0，对空间的任意元素定义了加法运算和与实数的数乘运算。如果运算满足如下条件，则称 S 为线性空间：

设 $x \in S, y \in S, z \in S$, 有

$x+y=y+x$; $x+(y+z)=(x+y)+z$; $x+0=x$; 存在 $\xi \in S$，使得 $\xi+x=0$

对任意实数 c,

$$c(x+y)=cx+cy;\ (c_1+c_2)x=c_1x+c_2x;\ c_1(c_2x)=(c_1c_2)x;\ 1\cdot x=x$$

我们称 S 中的元素为向量。

例1.1. 我们通常遇到的 n 维欧氏空间就是线性空间，元素是起点为原点的行向量 $a=(a_1,a_2,\cdots,a_n)$ 或列向量 $a=(a_1,a_2,\cdots,a_n)'$，其中 $'$ 表示转置。

例1.2. 设 $L_2=\{r.v.X: EX^2<\infty\}$, 则容易验证 L_2 为线性空间。

定义1.2. 线性相关和线性无关

设 $a_1,a_2,\cdots,a_m$ 是 S 中的一组向量，如果存在 m 个不全为零的实数 $\lambda_1,\lambda_2,\cdots,\lambda_m$, 使得 $\lambda_1a_1+\lambda_2a_2+\cdots+\lambda_ma_m=0$, 则称这一组向量是线性相关的，否则称为线性无关的。

在向量组 $a_1,a_2,\cdots,a_m$ 中, 一定存在若干个向量, 不妨记为 $a_1,a_2,\cdots,a_d$, 它们之间是线性无关的, 而 $a_1,a_2,\cdots,a_m$ 中的其余向量都可以由 $a_1,a_2,\cdots,a_d$ 线性表出，这组向量称为原向量组的一个极大线性无关向量组，其中的向量个数 d 称为 $a_1,a_2,\cdots,a_m$ 的秩. 当然，极大线性无关向量组不是惟一的。

定义1.3. 线性空间的基和维数

设 S 是线性空间，若存在一组向量 $a_1, a_2, \cdots, a_n$, 它们之间是线性无关的，对 S 中的任一向量 b, 存在 n 个实数 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 使得 $b = \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_n a_n$, 即 b 可以表为 $a_1, a_2, \cdots, a_n$ 的线性组合，则称向量组 $a_1, a_2, \cdots, a_n$ 为线性空间 S 的一组基，n 称为空间的维数，通常用 $d(S)$ 来表示空间 S 的维数。

由定义，对任一 n 维线性空间，一定存在一组基 $e_1, \cdots, e_n$，基中向量的个数就是空间的维数 n。如果把每个基看成一个坐标，由于空间任一向量 x 可以表为一组基的线性组合 $x = a_1 e_1 + a_2 e_2 + \cdots + a_n e_n$，我们可以把这组系数 $(a_1, a_2, \cdots, a_n)$ 看成向量 x 的一个表达。一组基定下来以后，这种表达是惟一的。当然，同一个向量在不同的基下有不同的表达。

我们常见的 n 维欧氏空间的一组基向量为 $e_1, e_2, \cdots, e_n$, 其中向量 e_i 的第 i 个分量为 1, 其余分量为零，一般向量可以用 $a_1, a_2, \cdots, a_n$ 来表达。

注 1：基不惟一，例如把上面 n 维欧氏空间的一组基改为 $f_1, e_2, \cdots, e_n$, 其中向量 f_1 的第 1 和第 2 个分量为 1，其余分量为零，则 $f_1, e_2, \cdots, e_n$ 仍是一组基，上面的 x 在这组基下表为 $a_1 - a_2, , a_2, a_3, \cdots, a_n$。

注 2：不是所有的线性空间都是有限维的，如上面的线性空间 L_2 就不是有限维的。

定义1.4. 子空间

设 S 为线性空间，$x_1, x_2, \cdots, x_m$ 为 S 中的 m 个向量，记集合

$$\mu = \mu(x_1, x_2, \cdots, x_m) = \{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m : \forall(\alpha_1, \alpha_2, \cdots, \alpha_m) \in R^m\}$$

容易验证集合 μ 是线性空间，我们称它是由向量 $x_1, x_2, \cdots, x_m$ 生成的 S 的一个线性子空间，简称子空间。子空间的维数记为 $d(\mu)$, 显然，$d \leq n$

由线性代数知识可知，子空间的维数 d 就是向量组 $x_1, x_2, \cdots, x_m$ 的秩。

定义1.5. 内积空间

设 $x,y,z \in S$, t 为实数，定义映射 $g:(x,y) \longrightarrow R$, 满足

$$\begin{aligned}(x,y) &= (y,x)\\(x,x) &> 0, x \neq 0\\(x,x) &= 0 \Leftrightarrow x=0\\(tx,y) &= t(x,y)\\(x+y,z) &= (x,z)+(y,z)\end{aligned}$$

则二元实函数 (x,y) 称为线性空间 S 上的一个内积，带内积的线性空间称为内积空间。

由内积我们可以定义向量的模 $\|x\| = \sqrt{(x,x)}$, 若 $\|x\|=1$, 我们称为单位向量。

我们通常见到的 R^n 就是内积空间, 记 $x=(x_1,x_2,\cdots,x_n)', y=(y_1,y_2,\cdots,y_n)'$, 内积的定义为 $(x,y)=\sum_{i=1}^n x_i y_i$.

例1.3. 在 L_2 中，我们定义 $(X,Y)=EXY$, 可以验证这是一个内积，因此 L_2 是一个内积空间。$\| X \| \equiv \sqrt{(X,X)}$, 称为随机变量 X 的模。也可以把内积定义为 $(X,Y)=Cov(X,Y)=E(X-EX)(Y-EY)$, 即随机变量 X 和 Y 的协方差，这相当于把向量的起点都平移到原点。

在内积定义为协方差时，若 $(X,Y)=0$, 我们称 $X \perp Y$. 由内积定义，我们知向量 X,Y 的夹角余弦为

$$cos\theta = \frac{(X,Y)}{\|X\|\|Y\|}$$

在 L_2 中，两个随机向量之间的夹角余弦就是相关系数

$$\rho_{XY} = \frac{Cov(X,Y)}{Var(X)Var(Y)}$$

上面也就是相关系数的几何意义。

§1.1.2 Gram-Schmidt 正交化程序

设 $e_1,e_2,\cdots,e_n$ 是 n 维内积空间 S 的一组基，如果 $e_i \perp e_j, \forall i \neq j$, 且对一切 $i=1,2,\cdots,n$, $\|e_i\|=1$, 我们称 $e_1,e_2,\cdots,e_n$ 是 S 的一组标准正交基。对 n 维内积空间 S 的任一组基 $x_1,x_2,\cdots,x_n$，我们可以构造出一组标准正交基

$e_1, e_2, \cdots, e_n$ 来。这就是如下的 Gram-Schmidt 正交化程序：

$$
\begin{aligned}
&\widetilde{e}_1 = x_1,\ e_1 = x_1/\|x\| \\
&\widetilde{e}_2 = x_2 - (x_2, e_1)e_1, e_2 = \frac{\widetilde{e}_2}{\|\widetilde{e}_2\|} \\
&\widetilde{e}_3 = x_3 - (x_3, e_1)e_1 - (x_3, e_2)e_2,\ e_3 = \frac{\widetilde{e}_3}{\|\widetilde{e}_3\|}, \\
&\cdots\cdots \\
&\widetilde{e}_n = (x_n, e_1)e_1 - (x_n, e_2)e_2 - \cdots - (x_n, e_{n-1})e_{n-1},\ e_n = \frac{\widetilde{e}_n}{\|\widetilde{e}_n\|}
\end{aligned}
$$

由此可知，每个 $e_1, e_2, \cdots, e_n$ 都可以用基向量 $x_1, x_2, \cdots, x_n$ 的线性组合表达出来。

§1.1.3 向量的正交投影和 Bessel 不等式

一. 向量的正交投影

我们需要下面的正交分解定理。

定理1.1. *(正交分解定理) 设 $W \subset S$ 是 S 的一个 r 维子空间，$x \notin W$，则存在向量 y, z，满足 $z \in W,\ y \in S,\ y \neq 0$，且 $y \perp W$，使得 $x = y + z$，且该分解是惟一的。*

证. 设 $x_1, x_2, \cdots, x_r$ 为 W 的一组基，由于 x 不在空间 W 中，故 $x_1, x_2, \cdots, x_r, x$ 是线性无关的。由 Gram-Schimidt 正交化程序，由 $x_1, x_2, \cdots, x_r, x$ 可以构造一组正交基 $e_1, e_2, \cdots, e_r, e_{r+1}$，使得

$$x = \alpha_{r+1}e_{r+1} + (\alpha_1 e_1 + \cdots + \alpha_r e_r)$$

而

$$z = \alpha_1 e_1 + \cdots + \alpha_r e_r \in W,\ y = \alpha_{r+1}e_{r+1} \perp W$$

$z = \alpha_1 e_1 + \cdots + \alpha_r e_r$ 称为是向量 x 在子空间 W 上的投影。投影在统计中常常是一种估计量，而向量 y 常常被称为残差。它有如下重要性质：

$$\|y\| = \inf_{u \in W} \|x - u\|$$

注意到若 $e_1, e_2, \cdots, e_r$ 是 W 的一组标准正交基，内积 $(x, e_i) = \alpha_i$ 可视为向量 x 在 e_i 上的投影。设 x 有正交投影分解 $x = y + z,\ y \perp W,\ z \in W$，则 $\forall u \in W, (u - z, y) = 0$，故

$$
\begin{aligned}
&\|x - u\|^2 = (x - u, x - u) = (y + z - u, y + z - u) \\
&= (y, y) + 2(y, z - u) + (z - u, z - u) \\
&= (y, y) + \|z - u\|^2 \geq \|y\|^2
\end{aligned}
$$

等号成立当且仅当 $u=z$, 这说明向量 x 在子空间 W 上的投影是使残差最小的 W 中的向量。这个几何直观对理解统计中常用统计量是非常有帮助的。

二. Bessel 不等式

设 $f_1, f_2, \cdots, f_n$ 为内积空间 S（不必是有限维的）中的一组标准正交向量, $\beta \in S$, 则

$$|(f_1,\beta)|^2+\cdots+|(f_n,\beta)|^2 \leq \|\beta\|^2$$

等号成立 $\Leftrightarrow \beta \in \mu(f_1,\cdots,f_n)$

上面不等式的几何意义是向量长度的平方（相当于三维空间立方体中斜边的平方）不小于任意直角边的平方和。

证. 设 $\delta=\sum_{i=1}^n (f_i,\beta)f_i$, 则 δ 是 β 在 $\mu(f_1,\cdots,f_n)$ 上的投影。由向量的正交分解, $\beta=\delta+r$, $(\delta,r)=0$, 因此

$$\|\beta\|^2=\|\delta\|^2+\|r\|^2 \geq \|\delta\|^2=\sum^n |(f_i,\beta)|^2$$

注. 这儿空间 S 不必是有限维的，例如由 $(-\pi,\pi)$ 上的 $Fourier$ 正交基 $(1,\sin nx,\cos nx, n=1,2,\cdots)$ 表达的 $Herbert$ 空间是无穷维的。

由上面讨论知，若 S_1 为 S 的一个子空间，设 $S_2=\{x:\ x\perp S_1, x\in S\}$, 则我们称 $S_2 \perp S_1$, 或称为 S_1 的正交补，记为 $S_1^{\perp}$，此时，空间可以记为 $S=S_1\oplus S_1^{\perp}$, 称为子空间 S_1 与 $S_1^{\perp}$ 的直和。子空间维数与空间 S 维数的关系是 $d(S)=d(S_1)+d(S_1^{\perp})$.

我们也容易把直和的概念推广到 k 个子空间的直和，这儿不再细述了。

§1.2 矩阵的一般理论和性质

一、矩阵的运算

把 m 个 n 维的行向量 $a_1,a_2,\cdots,a_m$ 叠起来，就构成一个 $m\times n$ 的矩阵. 一般记为 $A=(a_{ij})_{m\times n}$. 我们也可以把 $m\times n$ 的矩阵看成 n 个 m 维的列向量 $b_1,b_2,\cdots,b_n$ 并列在一起。如果 $m=n$, 我们称为方阵，方阵中的每个 a_{ii} 称为对角元。如果方阵中非对角元都是零，我们称为对角阵, 常记为 $A=diag(a_{11},\cdots,a_{nn})$。对矩阵，我们可以定义加法、数乘和矩阵乘法：

加法（都是 $m\times n$ 的矩阵）

$$A+B=(a_{ij}+b_{ij})$$

它有如下性质:

$$A+B=B+A,\quad A+(B+C)=(A+B)+C$$

数乘：设 c 为任一常数，定义 $cA=(ca_{ij})$, 数乘满足分配率和结合律：

$$(c+d)A=cA+dA,\quad c(A+B)=cA+cB$$

矩阵乘法：设 $A=(a_{ij})_{p\times q},B=(b_{ij})_{q\times r}$, 定义 $C=AB=(c_{ij})_{p\times r}$, 其中

$$c_{ij}=\sum_{k=1}^{q}a_{ik}b_{kj}$$

矩阵乘法满足结合律和分配率：

$$A(BC)=(AB)C=ABC$$
$$A(B+C)=AB+AC,\quad (A+B)(C+D)=A(C+D)+B(C+D)$$

交换律一般不成立。如果矩阵中的每个元素都是零，我们称为零矩阵，记为 $0=(0)m\times n$, 如果对角阵中每个对角元都等于 1，我们称为单位阵，记为 I. 零矩阵和单位阵有如下显然的性质 (只要能进行运算)：

$$A+0=A,$$
$$IA=AI=A$$

矩阵的转置 (transpose of a matrix): 设 $A=(a_{ij})_{m\times n}$, 定义 $b_{ij}=a_{ji},B=(b_{ij})_{n\times m}$, 我们称矩阵 B 为矩阵 A 的转置，记为 A^T 或 A', 矩阵转置满足

$$(AB)'=B'A',\quad (ABC)'=C'B'A'$$

若实方阵 A 满足 $A'=A$, 我们称方阵 A 是对称阵

共轭转置，设矩阵 $A=(a_{ij})_{p\times q}$ 的元素 a_{ij} 是复数, 定义 $A^*=\bar{A}'=(\bar{a}_{ji})_{q\times p}$, 其中 $\bar{a}_{ji}$ 表示元素 a_{ji} 的共轭。共轭转置这一运算满足 $(AB)^*=B^*A^*$,$(ABC)^*=C^*B^*A^*$.

若复方阵 A 满足 $A^*=A$, 我们称方阵 A 是 $Hermit$ 阵.

逆矩阵：设 A 为方阵，若存在方阵 B, 使得 $AB=BA=I$, 则方阵 B 称为 A 的逆，记为 A^{-1}. 逆运算有如下性质：

$$(AB)^{-1}=B^{-1}A^{-1},\quad (ABC)^{-1}=C^{-1}B^{-1}A^{-1}$$
$$(A')^{-1}=(A^{-1})',\quad (A^*)^{-1}=(A^{-1})^*$$

注 1 $AB=0 \nRightarrow A=0$ 或 $B=0$

U 阵 (unitary matrix) 和正交阵 (orthogonal matrix): 一个复元素的方阵 A 称为 U 阵, 若 $A^*A=AA^*=I$, 此时, $A^*=A^{-1}$, 当 A 的元素为实数时, 称为正交阵, 此时 $A^*=A'$, 故正交阵满足 $A'A=AA'=I$.

分块矩阵：若矩阵 A 可以分为 $r\times s$ 个矩形小块, 如

$$A=\begin{pmatrix} P & Q \\ R & S \end{pmatrix}$$

则 A 称为一个分块矩阵. 设分块矩阵 B 为

$$B=\begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

只要运算能进行, 我们有分块矩阵的乘法

$$\begin{aligned} AB &= \begin{pmatrix} P & Q \\ R & S \end{pmatrix}\cdot\begin{pmatrix} E & F \\ G & H \end{pmatrix} \\ &= \begin{pmatrix} PE+QG & PF+QH \\ RE+SG & RF+SH \end{pmatrix} \end{aligned}$$

分块矩阵求逆：设分块矩阵

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

若 A 可逆, 且 A_{11}, A_{22} 都可逆, 记

$$A_{11.2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$$

$$A_{22.1}=A_{22}-A_{21}A_{11}^{-1}A_{12}$$

则

$$A^{-1}=\begin{pmatrix} P & Q \\ R & S \end{pmatrix}^{-1}=\begin{pmatrix} A_{11\cdot 2}^{-1} & -A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1} \\ -A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} & A_{22\cdot 1}^{-1} \end{pmatrix}$$

其中

$$A_{11\cdot 2}^{-1}=(A_{11}^{-1}-A_{12}A_{22}A_{11}^{-1}A_{21})^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1}$$

$$A_{22\cdot 1}^{-1}=(A_{22}^{-1}-A_{21}A_{11}^{-1}A_{12})^{-1}=A_{22}^{-1}+A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1}$$

例1.4. 在分块矩阵求逆公式中取 $A_{12}=a, A_{22}=-1, A_{21}=b'$, 其中 a,b 为列向量, 则我们有

$$(A_{11}+ab')^{-1}=A_{11}^{-1}-A_{11}^{-1}ab'A_{11}^{-1}/(1+b'A_{11}^{-1}a)$$

二. 矩阵的相抵

对矩阵 $A_{m\times n}$，我们可以作如下的初等变换：

某行乘以一个数 λ。取对角阵 $M_i(\lambda)=diag(1,\cdots,\lambda,1,\cdots,1)$, 即第 i 个对角元为 1，其余为 1. 则矩阵 $M_i(\lambda)$ 左乘 A 就完成此功能。

把 s 行的 λ 倍加到第 r 行。令 $P_{rs}(\lambda)=(p_{ij})_{m\times n}$, 其中 $p_{ii}=1, i=1,\cdots,m$, $p_{rs}=\lambda$, 其余元都为零。则矩阵 $P_{rs}(\lambda)$ 左乘 A 就完成此功能。

交换 r,s 行。令 $T_{rs}=(t_{ij})$, 其中 $t_{rs}=t_{sr}=1, t_{rr}=t_{ss}=0$, 矩阵的其余元为零。则矩阵 T_{rs} 左乘 A 就完成此功能。

注 2 我们也可以对矩阵中的列作上述三种变换, 此时只要把上述的三种初等变换矩阵右乘矩阵 A 就完成此功能。

注 3 可以验证，$M_i(\lambda)^{-1}=M_i(\lambda^{-1}), P_{rs}(\lambda)^{-1}=P_{rs}(-\lambda), T_{rs}^{-1}=T_{rs}$

由上面对行和列的三种初等变换，在线性代数中我们可以得到矩阵的相抵标准型：对任意的矩阵 $A_{m\times n}$, 存在 $m\times m$ 可逆方阵 P 和 $n\times n$ 可逆方阵 Q，使得

$$PAQ=\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

其中 r 称为矩阵的秩 (rank)，记为 $rank(A)$, 它可以看作矩阵 A 的 n 个列向量中极大线性无关组中向量的个数，也可以看作矩阵 A 的 m 个行向量中极大线性无关组中向量的个数。

三. 矩阵的相合

这是针对对称方阵 A 而言的, 设 A 为对称方阵，即 $A'=A$, 则存在可逆方阵 P, 使得

$$P'AP=\begin{pmatrix} I_r & 0 & 0 \\ 0 & -I_s & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.1}$$

其中 r 和 s 分别称为对称方阵 A 的正负惯性指数，上述的 $P'AP$ 称为对称方阵 A 的相合标准型。

设 A 为 n 阶实对称方阵, x 为 $n\times 1$ 实的列向量, 称 $x'Ax$ 为一个二次型。设 P 是把 A 变换为标准型的可逆矩阵, 记 $y=P^{-1}x$, 由于

$$x'Ax=(P^{-1}x)'(P'AP)(P^{-1}x)=y'\begin{pmatrix} I_r & 0 & 0 \\ 0 & -I_s & 0 \\ 0 & 0 & 0 \end{pmatrix}y$$
$$=y_1^2+\cdots+y_r^2-y_{r+1}^2-\cdots-y_{r+s}^2$$

其中每个 y_i 是向量 x 的一个线性组合。如果 $s=0$, 则二次型 $x'Ax$ 就是 r 个线性组合的平方和, 因此一定是非负的。当然, $r=0$ 就意味着二次型 $-x'Ax$ 就是 s 个线性组合的平方和, 所以二次型一定是非正的. 在许多实际应用中, 我们会遇到二次型求极值问题, 上面的分解能给我们求极值提供帮助。

如果 A 是一个复的 $Hermit$ 矩阵, 我们可以讨论 x^*Ax, 它有类似于二次型的性质。

四. 矩阵的相似

对任意方阵 A, 一定存在可逆方阵 P, 使得

$$P^{-1}AP=\begin{pmatrix} J_1(\lambda_1) & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & J_r(\lambda_r) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

其中

$$J_i(\lambda_i)=\begin{pmatrix} 1 & \lambda_i & \dots & 0 & 0 \\ 0 & 1 & \lambda_i & \cdots & 0 \\ 0 & 0 & \ddots & \vdots & 0 \\ 0 & 0 & 0 & 1 & \lambda_i \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

§1.3 矩阵的数字特征

矩阵是由 $m\times n$ 个数据构成的, 这对于研究矩阵本身带来一定的困难。因此一个想法是能否用一个数字来反映矩阵的某种特性? 这如同在研究随机变量时用均值、方差和中位数等来描述随机变量的某个特征一样。下面我们给出若干常用的描述矩阵特性的量。

一. 矩阵的秩

矩阵的秩 r 在上一节已介绍过, 它表示由矩阵 A 的列向量 (或 A 的行向量) 所组成的线性空间的维数, 也就是矩阵 A 的列向量 (或 A 的行向量) 中极大线性无关向量组中的向量个数, 它是初等变换下的不变量。

二. 方阵的行列式

如果 A 是一个 $n \times n$ 阶的方阵，我们在线性代数中可以定义 A 的行列式 $det(A)$，或记为 $|A|$。它在 n 维欧氏空间中的几何意义在线性代数中已有叙述，设 A 是一个线性变换，$y = Ax$, 以 $dx_1 \cdots dx_n$ 表示由向量 $dx_1, \cdots, dx_n$ 所构成的 n 维立方体的体积，则在线性变换 A 下，$dy_i, \cdots, dy_n$ 所构成的立方体的体积为 $|A| dx_1 \cdots dx_n$，即 $|A|$ 表示变换前后体积的比值. 由于在变换中仅仅考虑体积，而行列式有正负，即还考虑到向量的方向，所以我们用的是行列式的绝对值。

例1.5. 在计算二重积分 $\int\int f(x,y)dxdy$ 时，经常会用到变量代换，设

$$\begin{aligned} u &= \varphi(x,y) \\ v &= \phi(x,y) \end{aligned}$$

则

$$\begin{aligned} du &= \frac{\partial \varphi}{\partial x}dx + \frac{\partial \varphi}{\partial y}dy \\ dv &= \frac{\partial \phi}{\partial x}dx + \frac{\partial \phi}{\partial y}dy \end{aligned}$$

由此得到

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} \frac{\partial \varphi}{\partial x} & \frac{\partial \varphi}{\partial y} \\ \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \equiv A \begin{pmatrix} dx \\ dy \end{pmatrix}$$

其中 A 就是变量 dx, dy 到变量 du, dv 的线性变换，因此 $|A|$ 的绝对值就是 (u,v) 关于 (x,y) 的 $Jacobi$ 行列式的绝对值，即从 du, dv 构成的 u, v 平面上平行四边形的面积与矩形 $dxdy$ 面积的比值的绝对值。为方便记忆，可以把经变换后的二重积分写为

$$\int\int f(x,y)dxdy = \int\int f(x,y)\frac{dxdy}{dudv}dudv$$

其中 $\frac{dxdy}{dudv}$ 是 x, y 关于 u, v 的 $Jacobi$ 行列式，其值等于 $1/|A|$.

关于行列式，我们有如下的重要性质，设 A, B 都是 $n \times n$ 矩阵，则

$|AB| = |A||B|$

证： 首先我们证明上三角准对角块方阵行列式等于对角块行列式的乘积，即

$$\begin{vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{vmatrix} = |A_{11}||A_{22}|$$

记

$$A=\begin{pmatrix} a_{11} & \cdots & a_{1n} & a_{1n+1} & \cdots & a_{1n+m} \\ \vdots & & & \vdots & & \\ a_{n1} & \cdots & a_{nn} & a_{nn+1} & \cdots & a_{nn+m} \\ & 0 & & & A_{22} & \end{pmatrix}$$

由线性代数中关于行列式的定义，

$$det(A)=\sum\delta\begin{pmatrix} 1 & 2 & \cdots & n & n+1 & \cdots & n+m \\ i_1 & i_2 & \cdots & i_n & i_{n+1} & \cdots & i_{n+m}\end{pmatrix}$$
$$\times a_{1,i_1}\cdots a_{n,i_n}a_{n+1,i_{n+1}}\cdots a_{n+m,i_{n+m}}$$

其中 $i_1,\cdots i_{n+m}$ 是 $1,2,\cdots,n+m$ 的一个排列,δ 取 $+1$ 或 -1, 取决于 $1,2,\cdots,n+m$ 到给定排列所需对换的最小数是偶数还是奇数。但是当 $i_{n+1},\cdots,i_{n+m}$ 中有一个等于 $1,2,\cdots,n$ 中的一个时, 由于 A 为准上三角块, 故 $a_{n+1,i_{n+1}}\cdots a_{n+m,i_{n+m}}$ 的值为 0，所以在 $det(A)$ 的展开中不为零的项只能是 $i_{n+1},\cdots i_{n+m}$ 为 $n+1,\cdots n+m$ 的一个排列，于是

$$\begin{aligned} det(A) &= \sum\delta\begin{pmatrix} 1 & \cdots & n \\ i_1 & \cdots & i_n\end{pmatrix}\delta\begin{pmatrix} n+1 & \cdots & n+m \\ i_{n+1} & \cdots & i_{n+m}\end{pmatrix} \\ &\quad\times a_{1,i_1}\cdots a_{n,i_n}a_{n+1,i_{n+1}}\cdots a_{n+m,i_{n+m}} \\ &= \sum\delta\begin{pmatrix} 1 & \cdots & n \\ i_1 & \cdots & i_n\end{pmatrix}a_{1,i_1}\cdots a_{n,i_n} \\ &\quad\times\sum\delta\begin{pmatrix} n+1 & \cdots & n+m \\ i_{n+1} & \cdots & i_{n+m}\end{pmatrix}a_{n+1,i_{n+1}}\cdots a_{n+m,i_{n+m}} \\ &= det(A_{11})det(A_{22}) \end{aligned}$$

从线性代数教科书中我们知道行列式有如下性质：把一行乘以常数后加到另一行上去, 行列式的值不变, 交换行列式的两行或两列，行列式仅仅改变符号。方阵转置后的行列式等于原方阵的行列式，故下三角准对角块方阵的行列式等于对角块方阵行列式的乘积。由此，我们有

$$|A||B|=det\begin{pmatrix} A & 0 \\ -I & B\end{pmatrix}$$
$$=det\begin{pmatrix} A & AB \\ -I & 0\end{pmatrix}=(-1)^n det\begin{pmatrix} AB & A \\ 0 & -I\end{pmatrix}$$
$$=(-1)^n det(AB)\cdot(-1)^n=det(AB)$$

即 $|AB|=|A||B|$.

由上面的结论，当 A_{11}^{-1} 存在时，

$$\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I\end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix}=\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22\cdot 1}\end{pmatrix}$$

其中 $A_{22\cdot 1}=A_{22}-A_{21}A_{11}^{-1}A_{12}$, 所以我们有

$$|A|=|A_{11}||A_{22}-A_{21}A_{11}^{-1}A_{12}|$$

同理，当 $|A_{22}|\neq 0$ 时，

$$|A|=|A_{22}||A_{11}-A_{12}A_{22}^{-1}A_{21}|$$

若 A_{11} 与 A_{21} 可交换，可以推出 A_{11}^{-1} 与 A_{21} 可交换, 则

$$\begin{aligned}&|A_{22}-A_{21}A_{11}^{-1}A_{12}|=|A_{22}-A_{21}A_{12}A_{11}^{-1}|\\&=|A_{22}A_{11}-A_{21}A_{12}||A_{11}^{-1}|\end{aligned}$$

所以

$$|A|=|A_{22}A_{11}-A_{21}A_{12}|$$

同理, 若 A_{22} 与 A_{21} 可交换，则

$$|A|=|A_{11}A_{22}-A_{21}A_{12}|$$

例1.6. 设 $|D_{n\times n}|\neq 0, a, b$ 为 n 维列向量，则

$$\begin{vmatrix} D & a \\ -b' & 1 \end{vmatrix}=|D||1+b'D^{-1}a|=|D+ab'|$$

三. 特征根和特征向量

设 A 是 n 阶方阵, 我们称 $|A-\lambda I|=0$ 的根为方阵 A 的特征根, $|A-\lambda I|=0$ 称为特征方程。如果计算复根的话，n 阶方阵 A 有 n 个特征根。设 λ 为 A 的特征根，称满足 $Ax=\lambda x$ 的 n 维向量为特征根 λ 对应的特征向量。当 x 为特征向量时，由于对任一常数 c, cx 仍然是特征根 λ 对应的特征向量，由此知同一个特征根对应的特征向量不唯一，当特征向量的模为 1 时，我们称为单位特征向量。如果特征根 λ 是单根，其对应的单位特征向量的第一个非零分量为正，则该单位特征向量是特征根对应的惟一特征向量。

我们复习一下有关特征根和对应特征向量的有关性质。以下记 x_i 为特征根 λ_i 对应的特征向量。

1. 如果 $\lambda_i\neq\lambda_j$, 则对应的特征向量正交，即 $<x_i,x_j>=0$.

2. 若特征根 λ_i 为特征方程的 k 重根，则对应的特征向量构成一个 k 维线性子空间，从中可以选出 k 个正交的单位特征向量。

3. 若 A 为实对称方阵，则 A 的特征根全为实根，从而对应的特征向量也全部是实的。特征根可以按大小排列为 $\lambda_1(A) \geq \lambda_2(A) \cdots \geq \lambda_n(A)$。

4. A 和 A' 有相同的特征根，设 A 和 B 分别为 $p \times q$ 和 $q \times p$ 矩阵, 则 AB 和 BA 有相同的非零特征根。

证：由于 $|A' - \lambda I| = |A - \lambda I|$, 所以 A 和 A' 有相同的特征根。对第二个结论，由下式可以得到：

$$\begin{aligned} &\begin{vmatrix} \lambda I_p & A_{p\times q} \\ B_{q\times p} & I_q \end{vmatrix} = |\lambda I_p||I_q - B(\lambda I_p)^{-1}A| \\ &= \lambda^{p-q}|\lambda I_q - BA| = |I_q||\lambda I_p - AI_q^{-1}B| = |\lambda I_p - AB| \end{aligned}$$

5. 若 $rank(A) = r$, 则 A 有 r 个非零特征根，$n-r$ 个零特征根。

这可以从 A 的相似标准型得到。

6. $\lambda(A^{-1}) = (\lambda(A))^{-1}$

这由等式 $|A^{-1} - \lambda I| = |\lambda A^{-1}||\lambda^{-1}I - A| = 0$ 即得。

7. 设 $\lambda_1(A), \cdots, \lambda_n(A)$ 为 $A_{n\times n}$ 的特征根，f 为一多项式，则 n 阶方阵 $f(A)$ 的特征根为 $f(\lambda_1), \cdots, f(\lambda_n)$.

由定义可以直接验证，若 x 为 A 的特征根 λ 对应的特征向量时，$f(A)x = f(\lambda)x$, 所以结论成立。

例1.7. 设 a, b 为 n 维列向量，求 ab' 的非零特征根和对应的特征向量。

解：由于 ab' 和 $b'a$ 有相同的非零特征根，而 $b'a$ 是一个数，故矩阵 ab' 有非零特征根 $b'a$, 对应特征向量为 a.

8. 设 $A_{n\times n}$ 的特征根为 $\lambda_1, \cdots, \lambda_n$, 则 $|A| = \prod_{i=1}^n \lambda_i$.

这可以从方阵的相似标准型得到。

四. 方阵的迹 (trace)

设 $A = (a_{ij})_{n\times n}$, 特征根为 $\lambda_1, \cdots, \lambda_n$, 则方阵对角元之和称为该方阵的迹，记为 $tr(A)$, 且

$$tr(A) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i$$

这可以从特征方程根与系数的关系得到。

五. 矩阵的模

设 A 为 $m \times n$ 矩阵，称

$$||A|| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2} = \sqrt{tr(AA')}$$

为矩阵 A 的模，如果把矩阵看作 $mn \times 1$ 的向量，矩阵的模就是向量的长度。

六. 条件数

设 $A' = A$, 且其特征根为 $\lambda_1(A) \geq \cdots, \geq \lambda_n(A) > 0$, 称 $r = \frac{\lambda_1(A)}{\lambda_n(A)}$ 是矩阵阵 A 的条件数。条件数是衡量矩阵病态的一个量。一般，若 $r \geq 1000$, 称矩阵严重病态，当 $100 \leq r < 1000$ 时称为中等病态，$10 \leq r < 100$ 时为弱病态。

七. 广义逆

当矩阵 $A_{m\times n}$ 不是方阵，或 A 是方阵但是不可逆时，我们可以定义矩阵的广义逆。

定义1.6. *设 A 为 $n \times p$ 矩阵，若存在 $p \times n$ 矩阵 X, 使得 $AXA = A$, 则 X 称为 A 的一个广义逆，记为 A^-.*

注： 可以利用相抵标准型来证明广义逆一定存在，但是不惟一。一个常用的广义逆是加号逆 A^+（又称 Moore-Penrose 逆）, 其定义为

$$AA^+A = A, A^+AA^+ = A^+$$

加号逆是惟一的。有关性质可以参见线性代数或线性回归和线性模型的有关教科书。

§1.4 几类特殊的矩阵

在金融工程和计量经济学的理论和实际问题中，我们经常会遇到几类特殊的矩阵，如正交阵、非负定阵和投影阵等等。在本节中我们简要介绍有关内容。

一. 正交阵 (orthogonal matrix)

定义1.7. *若实矩阵 P 满足 $P'P = PP' = I$, 则 P 称为正交阵。*

由定义，$P^{-1} = P'$, 正交阵有如下特性：

1. 正交阵的所有特征根的模为 1. 设 λ 为正交阵 P 的特征根，x 为 λ 对应的非零特征向量，则 $Px = \lambda x$, 共轭转置得 $x^*P^* = \lambda^*x^*$, 注意到 P 是实的，故 $P^* = P'$, 两式相乘得 $x^*P'Px = x^*x = |\lambda|^2x^*x$, 约去非零常数 $x^*x = |x|^2$, 得 $|\lambda|^2 = 1$.

2. P 的 n 个行（或列）可看成 R^n 中的一组标准正交基。这由定义 $P'P = PP' = I$ 按行（或列）展开立得。

3. 正交阵表示 R^n 中的旋转变换，如二维空间中的正交阵为

$$P=\begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix}$$

三维空间中的正交阵为以下两种类型：

$$P=\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & cos\theta & sin\theta \\ 0 & -sin\theta & cos\theta \end{pmatrix}$$

注 正交阵在复数域上的推广是酉阵 U，它满足 $U^*U=UU^*=I, U$ 阵的特征根的模也是 1，它的行（或列）为复空间 Z^n 中的一组标准正交基。

二. 非负定阵、正定阵和负定阵 (non-nagetive definite matrix, positive definite matrix and nagetive definite matrix)

1. 非负定阵、正定阵和负定阵的定义

定义1.8. *若对称方阵 A 满足：对任意 $x\in R^n, x'Ax\geq 0$, 则称 A 为非负定阵，记为 $A\geq 0$, 若对称方阵 A 满足：对任意 $x\in R^n, x\neq 0$, 有 $x'Ax>0$, 则称 A 为正定阵，记为 $A>0$, 若 $-A>0$, 则称 A 为负定阵。*

$A\geq B$ 是指 A, B 都是对称阵，且 $A-B\geq 0$,

由线性代数知非负定阵和正定阵有下面的性质：

(1) $A\geq 0 \Leftrightarrow A$ 的一切主子行列式 ≥ 0,

$$A>0 \Leftrightarrow 一切\lambda(A)>0 \Leftrightarrow 一切顺序主子行列式>0,$$

以 A_{kk} 表示 A 的前 k 行和前 k 列构成的 $k\times k$ 矩阵的行列式，则

$$A<0 \Leftrightarrow 一切\lambda(A)<0 \Leftrightarrow A_{11}<0, A_{22}>0, \cdots, (-1)^j A_{jj}>0, \cdots$$

(2) 对任意矩阵 $B_{p\times q}$, 我们有 $BB'\geq 0, B'B\geq 0$

(3) $A>0 \Rightarrow A^{-1}>0$

(4) 设

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}>0, 则 A_{11}>0, A_{11\cdot 2}>0, A_{22\cdot 1}>0.$$

这儿 $A_{11.2}$ 和 $A_{22.1}$ 在本节前面定义。

(5) 存在正交阵 P, 使得 $P'AP=diag(\lambda_1,\cdots,\lambda_p)$

由性质 (5)，记 $B = Pdiag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_p})P'$, 则 B 称为 A 的平方根，记为 $B = A^{1/2}$。容易验证，$B^2 = A$ 及 $B \geq 0$。

(6) 由

$$
\begin{aligned}
A &= Pdiag(\lambda_1, \cdots, \lambda_p)P' = (e_1, \cdots, e_p)\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}\begin{pmatrix} e_1' \\ \vdots \\ e_p' \end{pmatrix} \\
&= (e_1, \cdots, e_p)\begin{pmatrix} \lambda_1 e_1' \\ \vdots \\ \lambda_p e_p' \end{pmatrix} = \sum_{i=1}^{p} \lambda_i e_i e_i'
\end{aligned}
$$

称为 A 的谱分解 (spectrum decomposition).

注： 性质 (5) 和 (6) 可以推广到 Hermite 阵 $H, H \geq 0$, 则存在酉阵 U, 使得 $U^* HU = diag(\lambda_1, \cdots, \lambda_p), \lambda_i \geq 0$.

2. 矩阵 A 的奇异值分解

设 $A = (a_{ij})_{m\times n}$，则 $AA' \geq 0$，故存在正交阵 P，使得 $P'AA'P = diag(\lambda_1^2, \cdots, \lambda_p^2, 0, \cdots, 0)$。设 $u_1, \cdots, u_p$ 分别为 $\lambda_1^2, \cdots, \lambda_p^2$ 对应的标准特征向量，$u_{p+1}, \cdots, u_m$ 是 $\mu^{\perp}(u_1, \cdots, u_p)$（即对应于特征根 0 的特征子空间）上的标准正交基，则

$$
u_1 u_1' + \cdots + u_p u_p' + u_{p+1} u_{p+1}' + \cdots + u_m u_m' = I_m,
$$

因此

$$
A = (\sum_{i=1}^{m} u_i u_i')A = (\sum_{i=1}^{p} u_i u_i')A \tag{1.2}
$$

这是由于对 $j > p$，u_j 是 AA' 对应于特征根 0 的特征向量，故 $AA'u_j = 0$, 从而 $u_j' AA' u_j = |A'u_j|^2 = 0$, 所以对所有的 $j > p, A'u_j = 0$, 令 n 维向量 $v_i = \lambda_i^{-1} A' u_i, i = 1, \cdots, p$, 则

$$
\begin{aligned}
v_i' v_i &= \lambda_i^{-2} u_i' AA' u_i = \lambda_i^{-2} \lambda_i^2 u_i' u_i = 1 \\
v_i' v_j &= \lambda_i^{-1} \lambda_j^{-1} u_i' AA' u_j = \lambda_i^{-1} \lambda_j^{-1} u_i' u_j = 0, i \neq j
\end{aligned}
$$

即 $(v_1, \cdots, v_p)$ 为一组标准正交基，所以由 (1.2) 式，

$$
\begin{aligned}
A &= (u_1 u_1' + \cdots + u_p u_p')A = \lambda_1 u_1 v_1' + \cdots + \lambda_p u_p v_p' \\
&= Udiag(\lambda_1, \cdots, \lambda_p)V
\end{aligned}
$$

其中 $U=(u_1,\cdots,u_p)_{m\times p}, V=(v_1,\cdots,v_p)_{n\times p}, \lambda_1,\cdots,\lambda_p$ 称为 A 的奇异值。容易验证 $v_1,\cdots,v_p$ 为 $A'A$ 中对应于特征根 $\lambda_1^2,\cdots,\lambda_p^2$ 的单位特征向量, 记 $v_{p+1},\cdots,v_n$ 为 $A'A$ 中对应于特征根 0 的标准特征向量, 令 $\Lambda=diag(\lambda_1,\cdots,\lambda_p)$, $P=(u_1,\cdots,u_p,u_{p+1},\cdots,u_m)_{m\times m}, Q=(v_1,\cdots,v_p,v_{p+1},\cdots,v_n)_{n\times n}$, 则 P,Q 分别为 $m\times m$ 和 $n\times n$ 正交阵, 且

$$A=P\begin{pmatrix}\Lambda & 0\\ 0 & 0\end{pmatrix}Q$$

注： 奇异值分解可以推广到复值矩阵，若 n 阶复方阵 $AA^*=A^*A$, 则 A 称为正规阵 (normal matrix)，此时存在酉阵 U 和实数 $\lambda_1,\cdots,\lambda_p$，使得 $A=U^*diag(\lambda_1,\cdots,\lambda_p)U$. 利用此分解，进行实值奇异值分解类似的讨论，可以得到复值矩阵的奇异值分解，我们不写出来了。

3. 相对特征根

定义1.9. 设 $A>0, B\geq 0$, 则 $|\lambda A-B|=0$ 称为 B 相对于 A 的特征方程，特征方程的根称为 B 关于 A 的相对特征根。

由于 $A>0$, 上面的定义等价于 $|\lambda I-A^{-1/2}BA^{-1/2}|=0$, 即 λ 为非负定矩阵 $A^{-1/2}BA^{-1/2}$ 的特征根。由于 AB 与 BA 的非零特征根相同，所以 λ 为 BA^{-1} 的特征根。注意到 λ 是非负实数, 而 BA^{-1} 不必对称。

$A^{-1/2}BA^{-1/2}, BA^{-1}$ 有相同的特征根，请考虑同一个特征根 λ 对应的特征向量有什么关系？

三. 投影阵

统计和金融工程中经常会遇到线性投影的问题，能否根据向量和投影空间的生成元 (一组基或生成该空间的一组向量) 直接写出该向量的投影？回答是肯定的。这涉及到投影阵的概念。

定义1.10. 设 A 为 $n\times n$ 实方阵，若满足 $A'=A, A^2=A$, 则我们称 A 为投影阵。

投影阵有如下性质：

(1) A 的特征根 $\lambda(A)=0$ 或 1。

设 λ 为 A 的任一特征根, x 为对应的非零特征向量, 则 $Ax=\lambda x\Rightarrow A^2x=\lambda Ax=\lambda^2x$, 由于 $x\neq 0$, $\Rightarrow \lambda^2=\lambda\Rightarrow\lambda=0$ 或 1.

(2) $tr(A)=A$ 的非零特征根的个数 $=rank(A)$

这是由于 $tr(A) = \sum \lambda_i = rank(A) =$ 非零特征根的个数

(3) $I - A$ 为投影阵, $A \oplus (I - A) = R^n$

证: $(I-A)' = I - A, (I-A)^2 = I - 2A + A^2 = I - A$

任意 $x \in R^n, x = Ax + (I-A)x$, 而 $Ax \perp (I-A)x$.

例1.8. $\frac{1}{n}\mathbf{1}\mathbf{1}'$ 和 $I - \frac{1}{n}\mathbf{1}\mathbf{1}'$ 都是投影阵。

这儿 $\mathbf{1}$ 是每个分量都是 1 的 n 维列向量。

设 $x = (x_1, \cdots, x_n)'$, 则

$$\frac{1}{n}\mathbf{1}\mathbf{1}'x = (\bar{x}, \cdots, \bar{x})'$$
$$(I - \frac{1}{n}\mathbf{1}\mathbf{1}')x = x - \mathbf{1}\bar{x} = (x_1 - \bar{x}, \cdots, x_n - \bar{x})'$$

即关于 x 的中心化。

例1.9. 设 $X_{n\times p}, n \geq p, rank(X) = p$, 我们可把 X 中的列视为 R^n 中的 p 个列向量。记 $P_x = X(X'X)^{-1}X'$, 则容易验证 P_x 为投影阵, 且 $rank(P_x) = p$

投影阵的几何意义: 设 X 的列生成的 p 维子空间为 $\mu(X)$, 则任意 R^n 中向量 y 在子空间 $\mu(X)$ 上的投影为 $X(X'X)^{-1}X'y$

回忆多元线性回归中系数的最小二乘解为 $\hat{\beta} = (X'X)^{-1}X'y \Rightarrow X\hat{\beta} = X(X'X)^{-1}X'y$, 其几何意义是 $X\hat{\beta}$ 为 y 在子空间 $\mu(X)$ 上的投影。

§1.5 二次型

设 $x \in R^n$, A 为任一实矩阵, $x'Ax$ 称为一个二次型。由于 $x'Ax = x'A'x = x'(\frac{A+A'}{2})x$, 而 $(A+A')/2$ 是对称阵, 所以下面我们讨论的二次型中的矩阵是实的对称阵。在 $n=2$ 和 $n=3$ 时, 二次型 $x'Ax = 1$ 的几何意义分别是平面和空间中的二次曲线和二次曲面。一般 $x'Ax = 1$ 表示 n 维空间中的一个 $n-1$ 维空间曲面。

由相合标准型知, 存在可逆矩阵 $C_{n\times n}$ 使得,

$$C'AC = \begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix}$$

在二次型中, 令 $y = C^{-1}x$, 则

$$x'Ax = y'(C'AC)y = y' \begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix} y$$
$$= y_1^2 + \cdots + y_p^2 - y_{p+1}^2 - \cdots - y_{p+q}^2$$

其中 p 和 q 为正负惯性指数. 如果 C 取为正交阵 P,(此时 $P' = P^{-1}$), 则 $P'AP$ 的标准型为

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

其中 $\lambda_1 \geq \cdots \geq \lambda_n$ 为对称阵 A 的特征根，即在旋转变换下，可以取到一个方向，使二次型 $x'Ax = 1$ 化为 $\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = 1$,(注意这儿有的 λ_i 可能为负数)。当所有的 $\lambda_i > 0$，即 A 为正定阵时，取 $y = P'x$, 我们有 $x'Ax = 1 \Rightarrow \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = 1$, 这是 n 维空间中的椭球面, $\lambda_1^{-1/2}$ 为该椭球面的半短轴, $\lambda_n^{-1/2}$ 为该椭球面的半长轴。

例1.10. 在极值问题中的应用：设 n 元函数 $f(x_1, \cdots, x_n)$ 有二阶连续偏导数，求 $f(x_1, \cdots, x_n)$ 的极值。

解： 记 $x = (x_1, \cdots, x_n)'$, $\frac{\partial f}{\partial x} = (\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n})'$, 由 *Tailor* 展开，在点 $x_0 \in R^n$ 附近有

$$\begin{aligned} & f(x_1, \cdots, x_n) = f(x_1, \cdots, x_n)|_{x=x_0} \\ & +(\frac{\partial f}{\partial x})'(x - x_0) + \frac{1}{2}(x - x_0)'H(x - x_0) + o(||x - x_0||^2) \end{aligned}$$

其中

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

称为 Hessaian 矩阵。若 x_0 为极值点，则 $\frac{\partial f}{\partial x}|_{x=x_0} = 0$，故

$$f(x) = f(x_0) + \frac{1}{2}(x - x_0)'H(x - x_0) + o(||x - x_0||^2)$$

若 $H > 0$, 则 $f(x) > f(x_0) \Rightarrow x_0$ 为极小值点, 若 $H < 0$, 则 $f(x) < f(x_0) \Rightarrow x_0$ 为极大值点, 若 H 既不是负定, 也不是正定阵, 则 x_0 不是极值点。

例1.11. 求 $f(x_1, x_2, x_3) = x_1^3 + x_2^2 + 3x_3^3 - 6x_1x_2 - 48x_1 - 9x_3 + 1$ 的极值。

解： 由上例, 先求驻点,

$$\begin{aligned} \frac{\partial f(x)}{\partial x} &= \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \frac{\partial f(x)}{\partial x_3} \end{pmatrix} \\ &= \begin{pmatrix} 3x_1^2 - 6x_2 - 48 \\ 2x_2 - 6x_1 \\ 9x_3^2 - 9 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

得 4 个驻点：

$$x^{(1)} = \begin{pmatrix} x_1 = -2 \\ x_2 = -6 \\ x_3 = 1 \end{pmatrix},\ x^{(2)} = \begin{pmatrix} x_1 = -2 \\ x_2 = -6 \\ x_3 = -1 \end{pmatrix},$$

$$x^{(3)} = \begin{pmatrix} x_1 = 8 \\ x_2 = 24 \\ x_3 = 1 \end{pmatrix},\ x^{(4)} = \begin{pmatrix} x_1 = 8 \\ x_2 = 24 \\ x_3 = -1 \end{pmatrix}.$$

在驻点的 Hessaian 矩阵为

$$\begin{pmatrix} 6x_1 & -6 & 0 \\ -6 & 2 & 0 \\ 0 & 0 & 18x_3 \end{pmatrix}$$

把 4 个解分别代入上面的 Hessaian 矩阵，只有解 3 是正定阵，其余既不是正定阵，也不是负定阵，所以 f 只有在 $(8, 24, 1)$ 处达到极小值 -493.

如何寻找正定阵的特征根 $\lambda_1 \geq \cdots \geq \lambda_n > 0$? 由线性代数中正定阵的几何意义可以得到，

$$\lambda_1 = max_{||x_1||=1} x_1' A x_1 = max \frac{x_1' A x_1}{x_1' x_1}$$

$$\lambda_2 = max_{||x_2||=1, x_2 \perp x_1} x_2' A x_2 = max_{x_2 \perp x_1} \frac{x_2' A x_2}{x_2' x_2}$$

$$\cdots$$

$$\lambda_{k+1} = max_{||x_{k+1}||=1, x_{k+1} \perp \mu(x_1, \cdots, x_k)} x_{k+1}' A x_{k+1}$$

$$\cdots$$

$$\lambda_n = max_{||x_n||=1, x_n \perp \mu(x_1, \cdots, x_{n-1})} x_n' A x_n$$

其中 $x_i, i = 1, \cdots, n$ 分别为 λ_i 对应的特征向量.

§1.6 矩阵的特殊运算与矩阵的微商

1. 拉直运算

设矩阵 $A_{p\times q} = (a_{ij})_{p\times q}$, 记 $a_i = (a_{i1}, \cdots, a_{iq})', i = 1, \cdots, p$, 称

$$\overrightarrow{A} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}_{pq\times 1}$$

为矩阵 A 按行的拉直。当然，也可以定义按列拉直，记 $\tilde{a}_j = (a_{1j}, \cdots, a_{pj})', j = 1, \cdots, q$, 则 $(\tilde{a}_1', \cdots, \tilde{a}_q')'_{pq\times 1}$ 称为矩阵 A 按列的拉直。

对数据矩阵

$$\overrightarrow{X}_{n\times p} \equiv \begin{pmatrix} x_1^{'} \\ \vdots \\ x_n^{'} \end{pmatrix}_{n\times p}, x_i \in R^p, i=1,\cdots,n$$

通常的拉直运算是

$$\overrightarrow{X^{'}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_{np\times 1}$$

注： 今后我们对矩阵的拉直运算是按行作拉直运算的。

2. 叉积运算（Kronecker 积、张量积、直积）

设 $A_{p\times q}=(a_{ij})_{p\times q}, B_{r\times s}$ 为任意两个矩阵，定义

$$A\otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \cdots & a_{pq}B \end{pmatrix}_{pr\times qs}$$

为矩阵 $A_{p\times q}, B_{r\times s}$ 的叉积。叉积在多元线性回归等方面有很重要的应用。根据定义，叉积有如下主要性质：

(1) $(A\otimes B)'=(A'\otimes B')$,

(2) $(A\otimes B)(C\otimes D)=(AC\otimes BD)$, 只要矩阵乘法有意义。

特别，当 A, B 都是同阶投影阵时，$(A\otimes B)^2=(A\otimes B)$,

(3) $(A\otimes B)^{-1}=(A^{-1}\otimes B^{-1})$,

(4) 设 $A_{n\times n}$ 的特征根为 $\lambda_i, i=1,\cdots,\lambda_n, B_{m\times m}$ 的特征根为 $\mu_i, i=1,\cdots,m$, 则 $(A\otimes B)_{nm\times nm}$ 的特征根为 $\lambda_i\mu_j, i=1,\cdots,n, j=1,\cdots,m$,

(5) $|A\otimes B)|=|A|^n|B|^m$,

(6) $tr(A\otimes B)=tr(A)\cdot tr(B), \ \ rank(A\otimes B)=rank(A)\cdot rank(B)$

(7) 与向量拉直运算的关系

按列拉直，我们有 $\overrightarrow{ABC'}=(C\otimes A)\vec{B}$, 特别，向量 $x_{p\times 1}y_{1\times p}^{'}=(y\otimes x)_{p^2\times 1}$.

按行拉直，我们有 $\overrightarrow{ABC'}=(A\otimes C)\vec{B}$,

任给 $A\in R^{m\times n}$, 存在与 A 无关的 mn 阶置换矩阵 H，使得 $H\vec{A'}=\vec{A}$.

3. 矩阵的微商

矩阵微商是通常微商的推广，它在多元函数求二阶微商和偏导数时非常有用，如可以应用在 ML 估计和 LS 估计中。

设

$$Y_{p\times q}=\begin{pmatrix} y_{11} & \cdots & y_{1q} \\ \vdots & \ddots & \vdots \\ y_{p1} & \cdots & y_{pq} \end{pmatrix}$$

(1) 若 $y_{ij}=y_{ij}(t)$, 定义

$$\frac{\partial Y}{\partial t}=\left(\frac{\partial y_{ij}(t)}{\partial t}\right)_{p\times q}$$

(2) 设 $y_{ij}=y_{ij}(x_1,\cdots,x_s)$,

(3) 设 $y_{ij}=y_{ij}(x_1^{'},\cdots,x_r^{'})=y_{ij}((\overrightarrow{X}_{r\times s})')$,

其中 (3) 中的 x_i 为 $s\times 1$ 列向量, $X'=(x_1,\cdots,x_r)_{s\times r}$。显然, 情况 (2) 是 (3) 的特例, 故我们仅仅考虑 (3). 记 $Y=F(X)$, 其中 F 是矩阵 $(f_{ij}(X))_{p\times q}, X$ 为 $m\times n$ 的矩阵。注意以下的拉直都是按行拉直。定义

$$\frac{\partial Y}{\partial X}|_{mn\times pq}\equiv\frac{\overrightarrow{\partial F(X)'}}{\partial\overrightarrow{X}}=\begin{bmatrix} \frac{\partial f_{11}}{\partial x_{11}} & \frac{\partial f_{12}}{\partial x_{11}} & \cdots & \frac{\partial f_{pq}}{\partial x_{11}} \\ \frac{\partial f_{11}}{\partial x_{12}} & \frac{\partial f_{12}}{\partial x_{12}} & \cdots & \frac{\partial f_{pq}}{\partial x_{12}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{11}}{\partial x_{mn}} & \frac{\partial f_{12}}{\partial x_{mn}} & \cdots & \frac{\partial f_{pq}}{\partial x_{mn}} \end{bmatrix}$$

$\frac{\partial F(X)}{\partial X}$ (简记为 $\frac{\partial F}{\partial X}$) 为 $F(X)$ 关于 X 的微商矩阵, (当 $m=n, X$ 对称时要另行处理)。

作为特例, $y=\varphi(x_1,\cdots,x_n), x=(x_1,\cdots,x_n)'$, 则

$$\frac{\partial\varphi}{\partial x'}=(\frac{\partial\varphi}{\partial x_1},\cdots,\frac{\partial\varphi}{\partial x_n}),$$
$$\frac{\partial\varphi}{\partial x}=(\frac{\partial\varphi}{\partial x_1},\cdots,\frac{\partial\varphi}{\partial x_n})'$$

按上面定义, 矩阵 $X\in R^{m\times n}$ 的数值函数 $\Phi(X)$ 对矩阵 X 的微商是

$$\frac{\partial\Phi}{\partial X}=(\frac{\partial\Phi}{\partial x_{11}},\frac{\partial\Phi}{\partial x_{12}},\cdots,\frac{\partial\Phi}{\partial x_{1n}}\frac{\partial\Phi}{\partial x_{21}},\cdots,\frac{\partial\Phi}{\partial x_{2n}},\cdots,\frac{\partial\Phi}{\partial x_{m1}},\cdots,\frac{\partial\Phi}{\partial x_{mn}})'$$

即为 $mn\times 1$ 的列向量, 但是习惯上记 $\frac{\partial\Phi}{\partial X}$ 为等价的矩阵形式, 即

$$\frac{\partial\Phi}{\partial X}=\left(\frac{\partial\Phi}{\partial x_{ij}}\right)_{m\times n}$$

我们还需要矩阵数值函数 $\Phi(X)$ 对矩阵 X 的二阶微商矩阵:

$$\frac{\partial^2\Phi(X)}{\partial X^2}=\frac{\partial}{\partial X}\left(\frac{\partial\Phi(X)}{\partial\overrightarrow{X}}\right)'=\frac{\partial^2\Phi(X)}{\partial x_{ij}\partial x_{k\ell}}\in R^{mn\times mn}$$

当 X 为向量 $x\in R^n$ 时,

$$\frac{\partial^2\Phi}{\partial x^2}=\left(\frac{\partial^2\Phi(x)}{\partial x_i\partial x_j}\right)_{n\times n}\in R^{n\times n}$$

矩阵微商的性质，只要下面所涉及的运算可行，我们有

(1) 设 C 为常数矩阵，$F(X)=C \Rightarrow \frac{\partial F(X)}{\partial X}=0$(矩阵)，$\frac{\partial X}{\partial X}=I_{mn}, X\in R^{m\times n}$,

(2) $\frac{\partial \sum \beta_i F_i(X)}{\partial X}=\sum \beta_i \frac{\partial F_i(X)}{\partial X}$ (线性性)

(3) 设 $F(X)\in R^{p\times q}, G(X)\in R^{q\times r}$, 则

$$\frac{\partial F(X)G(X)}{\partial X}=\frac{\partial F(X)}{\partial X}(I_p\otimes G(X))+\frac{\partial G(X)}{\partial X}(F'(X)\otimes I_r)$$

(4) 设 $F(G(X))$ 为两个矩阵函数的复合函数，则

$$\frac{\partial F(G(X))}{\partial X}=\frac{\partial G(X)}{\partial X}\cdot\frac{\partial F(G)}{\partial G}$$

推论：只要下面的运算可行，就有

(5) $\frac{\partial AXB}{\partial X}=A'\otimes B$

证：由性质 (3),

$$\begin{aligned}\frac{\partial AXB}{\partial X}&=\frac{\partial AX}{\partial X}(I\otimes B)+\frac{\partial B}{\partial X}(X'A'\otimes I)\\&=\left(\frac{\partial A}{\partial X}(I\otimes X)+\frac{\partial X}{\partial X}(A'\otimes I)\right)(I\otimes B)+0\\&=I(A'\otimes I)(I\otimes B)=A'\otimes B\end{aligned}$$

(6) $\frac{\partial X^{-1}}{\partial X}=-(X^{-1})'\otimes X^{-1}$

证：由于 $XX^{-1}=I$, 所以

$$0=\frac{\partial XX^{-1}}{\partial X}=\frac{\partial X}{\partial X}(I\otimes X^{-1})+\frac{\partial X^{-1}}{\partial X}(X'\otimes I)$$

注意到 $(X'\otimes I)^{-1}=(X')^{-1}\otimes I$, 故

$$\begin{aligned}&-I\otimes X^{-1}=\frac{\partial X^{-1}}{\partial X}(X'\otimes I)\Rightarrow\\&\frac{\partial X^{-1}}{\partial X}=-(I\otimes X^{-1})((X')^{-1}\otimes I)=-(X^{-1})'\otimes X^{-1}+\frac{\partial X^{-1}}{\partial X}(X'\otimes I)\end{aligned}$$

(7) 矩阵迹关于矩阵 X 的微商

$$\begin{aligned}&\frac{\partial}{\partial X}tr(AX+X'BX)=A'+BX+B'X\\&\frac{\partial}{\partial X}tr(AXBXCX')=B'X'A'XC'+AXBXC\end{aligned}$$

性质 (7) 的证明留给读者。

第二章　多元统计基本知识

本章复习有关多元统计量的分布和有关的数字特征，为计量经济学的学习提供相关的基础知识。

§2.1　随机向量的数字特征

我们知道，随机变量是随机会而取值的变量，描述它全部随机特性的是随机变量的分布（包括分布律、分布函数和概率密度函数），描述它某方面特性的量称为它的数字特征，均值、中位数、方差和相关系数是最常用的量。随机变量只能反映研究对象某一个指标的变化规律，如果要反映研究对象两个或两个以上指标同时变化的规律，就要用随机向量的分布来描述。在计量经济学中，同时研究两个或以上变量的情况是经常发生的，如同时研究国民经济收入和支出，同时研究两个以上不同地区经济发展的差异，教育与生产力的关系，研究若干股票之间的关联等等。随机向量 $X=(X_1,\cdots,X_p)'$ 是一个 p 维向量，它的每个分量都是随机变量。随机向量联合分布的定义为：

定义2.1. 设 $X=(X_1,\cdots,X_p)'$ 是一个 p 维随机向量，$x=(x_1,\cdots,x_p)'\in R^p$，我们称

$$F(x)=F(x_1,\cdots,x_p)=P(X_1\le x_1,\cdots,X_p\le x_p)\equiv F(X\le x)$$

为随机向量 X 的分布。

若 X 的每个分量都仅取有限多个或可数多个值，我们称 X 为离散型随机向量。设 X_i 的取值为 $x_{i1},x_{i2},\cdots,i=1,\cdots,p$。定义 X 的分布律为，

$$p(x)=p(x_1,\cdots,x_p)=P(X_1=x_1,\cdots,X_p=x_p)$$

其中 x_i 的取值范围分别是 $x_{i1},x_{i2},\cdots$.

如果存在一个非负函数 $f(x)\equiv f(x_1,\cdots,x_p)$，使得分布函数 $F(x)$ 可以用 $f(x)$ 的积分表达，即对任意 $x\in R^p$，

$$\begin{aligned}F(x)=F(x_1,\cdots,x_p)&=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_p}f(t_1,\cdots,t_p)dt_1\cdots dt_p\\&\equiv\int_{-\infty}^{x}f(t)dt\end{aligned}$$

我们称随机向量 X 为连续型随机向量，F 称为连续型分布函数，f 称为联合概率密度函数，简称密度函数。

随机向量的分布全面描述了随机向量的取值规律，但是一方面它不容易从数据中得到，另一方面我们有时候也不需要了解到如此细的程度。对照随机变量，我们引入随机向量的数字特征，其中最常用的是随机向量的均值向量、协方差矩阵和相关系数矩阵。

定义2.2. 设 X 和 Y 分别为 $p\times 1$ 和 $q\times 1$ 的随机向量，称 $p\times 1$ 的随机向量

$$\mu = (EX_1, \cdots, EX_p)' \equiv EX$$

为 X 的均值向量。称 $p\times q$ 矩阵

$$\begin{aligned} Cov(X,Y) &= E(X-EX)(Y-EY)' \\ &= \begin{pmatrix} Cov(X_1,Y_1) & \cdots & Cov(X_1,Y_q) \\ \vdots & \cdots & \vdots \\ Cov(X_p,Y_1) & \cdots & Cov(X_p,Y_q) \end{pmatrix} \end{aligned}$$

为 X 和 Y 的协方差矩阵，称 $Cov(X,X)$ 为随机向量 X 的协方差矩阵（也称为方差阵），记为 $Var(X)$.

由定义，均值向量由每个分量的均值组成，协方差矩阵描述了随机向量 X 每个分量和 Y 的每个分量之间的协方差，而随机向量 X 本身的协方差矩阵则描述了本身 p 个分量两两之间的协方差。

设 A 为 $m\times p$ 常数矩阵，b 为常数向量，X 为 $p\times 1$ 的随机向量，则 AX 为由随机向量 X 分量的线性组合所构成的 m 维的随机向量，所以它的均值向量和方差矩阵可以按随机变量线性组合来计算，具体地，我们有

$$\begin{aligned} &E(AX+b) = AEX+b, \\ &Cov(AX+b, CY+d) = ACov(X,Y)C' \end{aligned}$$

可以看出常数矩阵可以作为常数一样提到期望号外面，由于矩阵乘法没有交换性，所以原有的顺序不能变化。常数向量对协方差而言不起作用, 特别

$$Var(AX+b) = E(AX+b-E(AX+b))(AX+b-E(AX+b))' = AVar(X)A'$$

关于随机向量 X 的方差阵，容易看出它是对称阵。进一步，我们有如下的命题：

命题2.1.

$$Var(X) \geq 0$$

证： 对任给的 $a \in R^p$,

$$a'Var(X)a = a'E(X-EX)(X-EX)' = E[(a'(X-EX))(X-EX)'a]$$
$$= E[a'(X-EX)]^2 \geq 0$$

通常情况下，我们都假定 $Var(X) > 0$, 即 X 的分量之间没有严格的线性关系。

由协方差阵，我们可以定义随机向量之间的相关系数矩阵 R. 设 X 的分量 X_i 和 Y 的分量 Y_j 之间的相关系数为

$$\rho_{X_iY_j} = \frac{Cov(X_i, Y_j)}{\sqrt{Var(X_i)Var(Y_j)}}$$

定义 X 和 Y 之间的相关系数矩阵为

$$R_{XY} = (\rho_{X_iY_j})_{p\times q}$$

当 $Y = X$ 时，$R_{XX} \equiv R_X$ 为随机向量 X 的相关系数矩阵。相关系数矩阵和协方差矩阵之间有如下关系：

$$R_{XY} = \begin{pmatrix} \sigma_{X_1}^{-1} & & \\ & \ddots & \\ & & \sigma_{X_p}^{-1} \end{pmatrix} Cov(X,Y) \begin{pmatrix} \sigma_{Y_1}^{-1} & & \\ & \ddots & \\ & & \sigma_{Y_q}^{-1} \end{pmatrix}$$

其中 $\sigma_{X_i}^{-1} = \sqrt{Var(X_i)}$, $\sigma_{Y_j}^{-1} = \sqrt{Var(Y_j)}$.

下面讨论样本协方差矩阵。设 $X, X_1, \cdots, X_n$ 为独立同分布 (iid) 的 $p \times 1$ 随机向量，其公共协方差阵为 Σ, 记

$$\mathbf{X}' = (X_1, \cdots, X_n),$$
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\mathbf{X}'\mathbf{1}$$

随机向量 X 的样本协方差矩阵定义为

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})'$$
$$= \frac{1}{n-1}(X_1 - \overline{X}, \cdots, X_n - \overline{X}) \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_n - \overline{X} \end{pmatrix}$$
$$= \frac{1}{n-1}\left(\mathbf{X}' - \frac{1}{n}\mathbf{X}'\mathbf{1}\mathbf{1}'\right)\left(\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}\right)$$
$$= \frac{1}{n-1}\mathbf{X}'(I - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{X}$$

样本协方差矩阵 S 是总体协方差阵的无偏估计，不失一般性，不妨设总体均值向量 μ 为零向量，不然，在下面证明中分别用 $X_i-\mu, \bar{X}-\mu$ 代替 X_i 和 $\bar{X}$，注意到 $i \neq j$ 时，X_i 和 X_j 相互独立，它们之间的协方差为零，

$$\begin{aligned}
ES &= \frac{1}{n-1}\sum_{i=1}^{n} E(X_i-\bar{X})(X_i-\bar{X})' \\
&= \frac{1}{n-1}\sum_{i=1}^{n} E(X_iX_i'-2X_i\bar{X}+\bar{X}\bar{X}') = \frac{1}{n-1}(\Sigma-\frac{2}{n}\Sigma+\frac{1}{n}\Sigma) \\
&= \frac{1}{n-1}(1-\frac{1}{n})\Sigma = \Sigma
\end{aligned}$$

§2.2　多元正态及由它生成的统计量

一. 多元正态的定义及性质

在实际问题中遇到最多的是多元正态随机向量，许多统计量和电脑中的软件就是根据正态来的，所以我们必须对多元正态有较多的了解，这样才能面对实际数据来判断是否可以用多元正态来近似描述数据的统计特性。

定义2.3. 设 $Y=(Y_1,\cdots,Y_q)'$，其中 $Y_1,\cdots,Y_q$ 为 $iid, Y_i \sim N(0,1)$, $X=\mu+A'Y$, 其中 A 为 $q\times p$ 矩阵，μ 为 p 维列向量. 则 X 的分布称为 p 元正态分布，记为 $X \sim N_p(\mu, A'A)$。

由第一章关于随机向量期望的运算，可以得到，

$$\begin{aligned}
&EX = \mu \\
&Var(X) = Var(AY) = A'Var(Y)A = A'A
\end{aligned}$$

定义2.4. 设 X 为 p 维连续型随机向量，μ 为 p 维向量，$\Sigma > 0$, 若 X 的密度函数为

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

则称 X 服从均值向量为 μ，协方差矩阵为 Σ 的 p 维正态随机向量，记为 $X \sim N_p(\mu,\Sigma)$.

这是两个关于多元正态的定义，第一个定义要广一点，表现在多元正态的协方差矩阵不必为正定的。但是第二个定义是大家比较熟悉的，而且给出了密度函数。今后我们视问题来决定用哪个定义。

由第二个定义可推知，正态分布的均值向量等于 μ，协方差矩阵等于 Σ。因此正态分布完全由均值向量和协方差矩阵所惟一确定。利用这个特点，通过验证均值向量和协方差矩阵我们可以得到如下有用的结果：

设 $X, X_1, \cdots, X_n, iid.\ X \sim N(\mu, \Sigma), \Sigma > 0$, 则

$$\Sigma^{-1/2}(X-\mu) \sim N_p(0, I_p) \tag{2.1}$$

$$\overline{X} \sim N_p(\mu, \frac{1}{n}\Sigma), \quad \sqrt{n}\Sigma^{-1/2}(\overline{X}-\mu) \sim N_p(0, I_p) \tag{2.2}$$

定义2.5. 随机过程 $(X(t), t \in T)$, $EX^2(t) < \infty$ 称为 *Gauss* 过程，若对任意的 k 和任意 T 中的时刻 $t_1, \cdots, t_k$，$(X(t_1), \cdots, X(t_k))'$ 服从 k 维正态分布。

由定义，Gauss 过程由均值函数和协方差函数惟一确定。

命题2.2. 正态随机向量的线性组合仍是正态随机向量。

证： 设 $X \sim N(\mu, \Sigma)$, $Z = B'X + d$, 则

$$Z \sim N(B'\mu, B'\Sigma B)$$

我们用第一个定义来证明。取 $A = \Sigma^{1/2}$, 由定义，$X = \mu + A'Y$, 故

$$Z = B'X + d = B'(\mu + A'Y) + d = (B'\mu + d) + (AB)'Y$$

由定义 2.3 知结论成立。当然也可以用多元正态的第二个定义，不过要利用密度函数的积分来做，比较麻烦，这儿就不证了。

由推论可知，若 p 维正态随机向量 $X \sim N_p(\mu, \Sigma), \Sigma > 0$, 设

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},\ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},\ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中 X_1 为 q 维，Σ_{11} 为 $q \times q$ 矩阵，其他作相应分块，则

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

证： 在推论中取 $d = 0, B' = (I_q, 0)_{p\times q}$, 则

$$X_1 = B'X \sim N_q(B'\mu, B'\Sigma B) = N_q(\mu_1, \Sigma_{11})$$

同理可得 X_2 也是 $p-q$ 维正态随机向量。

例2.1. 设

$$X \sim N_p(\mu, \Sigma),\ \Sigma > 0,\ X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

μ 和 Σ 的分块同前，求给定 X_2 下随机向量 X_1 的条件密度函数 $f_{X_1|X_2}(x_1|x_2)$.

解： 由条件密度的定义，

$$
\begin{aligned}
&f_{X_1|X_2}(x_1|x_2) = \frac{f(x_1,x_2)}{f_{X_2}(x_2)} \\
&= \frac{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}}{(2\pi)^{-(p-q)/2}|\Sigma_{22}|^{-1/2}\exp\{-\frac{1}{2}(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2)\}} \\
&= (2\pi)^{-\frac{q}{2}}|\Sigma_{11\cdot 2}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}[(x-\mu)'\Sigma^{-1}(x-\mu)-(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2)]\}
\end{aligned}
$$

上式中的 $|\Sigma_{11\cdot 2}|^{-1/2}$ 由 $|\Sigma| = |\Sigma_{22}|\cdot|\Sigma_{11\cdot 2}|$ 得到, 把 Σ^{-1} 按分块矩阵求逆，经过较麻烦的计算，可得

$$
\begin{aligned}
&(x-\mu)'\Sigma^{-1}(x-\mu)-(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2) \\
&= [(x_1-\mu_1)-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)]'\Sigma_{11\cdot 2}^{-1}[(x_1-\mu_1)-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)]
\end{aligned}
$$

所以我们得到

$$
X_2|X_1 \sim N_q(\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2),\Sigma_{11\cdot 2})
$$

即在给定 X_2 下随机向量 X_1 的条件分布服从均值向量为 $\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)$, 协方差矩阵为 $\Sigma_{11\cdot 2}$ 的 q 维正态分布。

注意到 $\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)$ 是随机向量 $x_1-\mu_1$ 在随机向量 $x_2-\mu_2$ 上的线性投影，所以给定 $X_2=x_2$ 下随机向量 $X_1-\mu_1$ 的条件期望就是 $x_1-\mu_1$ 在随机向量 $x_2-\mu_2$ 上的线性投影。

注意到 $\Sigma_{11\cdot 2}\le\Sigma_{11}$,(等号成立当且仅当 $\Sigma_{12}=0$), 所以 $Var(X_1|X_2)\le Var(X_1)$. 矩阵 $\beta=\Sigma_{12}\Sigma_{22}^{-1}$ 称为 X_1 对 X_2 的回归系数矩阵, 而 $\mu_1+\beta(x_2-\mu_2)$ 称为 X_1 对 X_2 的回归函数。X_1 和 X_2 相互独立的充分必要条件是 X_1 和 X_2 的协方差矩阵 $\Sigma_{12}=0$。

§2.3 矩阵正态分布和 Wishart 分布

一. 矩阵正态的定义

设 $X_1,\cdots,X_n$ 是从正态总体 $N_p(\mu,\Sigma)$ 中得到的一组 p 维数据样本，把这 n 个 p 维数据放在一起，构成一个数据矩阵 $\mathbf{X}_{n\times p}$, 转置拉直记为

$$
\overrightarrow{\mathbf{X'}} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}_{np\times 1}
$$

数据矩阵 $\mathbf{X}_{n\times p}$ 和转置拉直后的均值和方差阵分别为

$$E\overrightarrow{\mathbf{X'}} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mathbf{1}_{n\times 1} \otimes \mu_{p\times 1}$$

$$E\mathbf{X}'_{p\times n} = \mu_{p\times 1}\mathbf{1}'_{1\times n}$$

$$Var(\mathbf{X}) \equiv Cov(\overrightarrow{\mathbf{X'}}, \overrightarrow{\mathbf{X'}}) = \begin{pmatrix} Var(X_1) & \cdots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \cdots & Var(X_n) \end{pmatrix}_{np\times np}$$

$$= \begin{pmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{pmatrix} = I_n \otimes \Sigma_{p\times p}$$

在理论上可以推广到更一般情况。

定义2.6. 设 $\boldsymbol{Y} = (y_{ij})_{m\times q}$, 其中 y_{ij} 为 $iid. \sim N(0,1), i = 1,\cdots,m, j = 1,\cdots,q$, $M_{n\times p}, A_{p\times q}, B_{n\times m}$ 为常数矩阵，称 $\boldsymbol{X} = M + B\boldsymbol{Y}A'$ 服从矩阵正态分布，记为 $X \sim N_{n\times p}(M, BB' \otimes AA')$.

由拉直运算知

$$\overrightarrow{\mathbf{X'}} = \overrightarrow{M'} + \overrightarrow{A\mathbf{Y'}B'} = \overrightarrow{M'} + (B \otimes A)\overrightarrow{\mathbf{Y'}}$$

$$E\overrightarrow{\mathbf{X'}} = \overrightarrow{M'},\ E\mathbf{X} = M$$

$$Var(\overrightarrow{\mathbf{X'}}) = Var((B \otimes A)\overrightarrow{\mathbf{Y'}})$$

$$= (B \otimes A)Var(\overrightarrow{\mathbf{Y'}})(B' \otimes A') = BB' \otimes AA'$$

矩阵正态是多元正态的推广，这从矩阵拉直后可以看出。统计上是由多元正态总体的样本 $X_1,\cdots,X_n$ 联合分布的研究所产生的。

二. Wishart 分布

Wishart 分布的研究起源于多元正态样本方程阵。下面我们先给出定义。

定义2.7. 设 $\boldsymbol{X}$ 服从矩阵正态 $N_{n\times p}(M_{n\times p}, I_n \otimes \Sigma_{p\times p})$, 即 $\boldsymbol{X}$ 的行向量 X_i 相互独立都服从 p 维正态，它们有公共的方差阵，但均值向量可以不同。$W = (\boldsymbol{X'X})_{p\times p}$ 称为非中心 *Wishart* 分布，记为 $W \sim W_p(n, \Sigma, \Delta)$, 其中 $\Delta = M'M$, $M = 0$ 时称为 *Wishart* 分布，记为 $W \sim W_p(n, \Sigma)$

由定义，记 $\mathbf{X}' = (X_1,\cdots,X_n)$, 则 $W = \mathbf{X}'\mathbf{X} = \sum_{i=1}^n X_i X_i'$, 此为 n 个独立 p 维正态的“平方和”。这种表示对理解 Wishart 分布很有帮助。考虑

Wishart 分布 $W \sim W_p(n, \Sigma)$. 当 $p = 1$ 时，即 $X_i \sim N(0, \sigma^2)$ 为一维零均值正态随机变量，$W = X'X = \sum_{i=1}^n X_i^2 \sim \sigma^2\chi_n^2$, 当 $n = 1$ 时，即 $\mathbf{X}$ 为一个样本 X_1 的 "平方和. 由 $X_1 \sim N_p(0, \Sigma)$ 推出 $Y_1 = \Sigma^{-1/2}X_1 \sim N_p(0, I_p)$, 从而

$$Y_1^{'}Y_1 = (\Sigma^{-1/2}X_1)'\Sigma^{-1/2}X_1 = X_1'\Sigma^{-1}X_1 \sim \chi_p^2,$$

由上可看出 χ^2 分布是 Wishart 分布的特例。

可以证明，多元正态样本协方差矩阵 S 满足 $(n-1)S \sim W_p(n-1, \Sigma)$。因此 Wishart 分布相当于一元正态生成的 χ^2 在 p 维正态时的作用。

三. Hoteling T^2 统计量

在一元正态均值的检验中，我们经常用到 t 统计量，在多元正态时相应的统计量就是 T^2 统计量。

定义2.8. 设 $A \sim W_p(n, I_p), u \sim N_p(\mu, I_p), A$ 与 u 相互独立，称

$$T^2 = nu'A^{-1}u$$

服从非中心 T^2 分布，记为 $T^2 \sim T^2_{p,n,\mu}$ ，当 $\mu = 0$ 时称 T^2 服从 T^2 分布，记为 $T^2 \sim T^2_{p,n}$

在定义中，若 $p = 1$, 则由上节知，$A \sim \sigma^2\chi_n^2, u/\sigma \sim N(0, 1)$, 从而

$$\begin{aligned} T^2_{1,n} &= n\frac{u}{\sigma}\left(\frac{\chi_n^2}{\sigma^2}\right)^{-1}\frac{u}{\sigma} \\ &= \left(\frac{\sqrt{n}u/\sigma}{\sqrt{\chi_n^2/(n\sigma^2)}}\right)^2 \sim t_n^2 \sim F_{1,n} \end{aligned}$$

由此看出 T^2 统计量是 t 统计量的推广。由于 $t_n^2 \sim F_{1,n}$, 故 T^2 分布也应该与一定自由度的 F 分布相对应。可以证明，

$$\begin{aligned} &\frac{n-p+1}{np}T^2_{p,n,\mu} \sim F_{p,n-p+1,\lambda}, \text{ 这儿}\lambda = \mu'\mu, \\ &\frac{n-p+1}{np}T^2_{p,n} \sim F_{p,n-p+1} \end{aligned}$$

这儿 $F_{p,n-p+1,\lambda}$ 称为非中心 F 分布，λ 为非中心参数。

注 1 若 $X_i \sim N(\mu_i, 1), i = 1, \cdots, n$ 相互独立，则 $\chi^2_{n,\lambda} = \sum_{i=1}^n X_i^2$ 的分布称为自由度为 n 的非中心 χ^2 分布，非中心参数为 $\lambda = \sum_{i=1}^n \mu_i^2$，记为 $\chi^2_{n,\lambda}$. 设 $X \sim \chi^2_{m,\lambda}, Y \sim \chi^2_n$, 且 X, Y 相互独立，则称比值 $F = \frac{X/m}{Y/n}$ 服从自由度为 m, n 的非中心 F 分布，记为 $F_{m,n,\lambda}$.

注 2 当 $A \sim W_p(n,\Sigma), u \sim N_p(\mu, c\Sigma), A$ 与 u 相互独立，则可以推出 $cnu'A^{-1}u \sim T^2_{p,n}$.

在一维正态均值的检验中，我们用 t 统计量 $\frac{\sqrt{n}(\bar{X}-\mu)}{S}$, 如何推广到多维场合？注意到

$$\left(\frac{\sqrt{n}(\bar{X}-\mu)}{S}\right)^2 = n(\bar{X}-\mu)\left(\frac{\sum_{i=1}^n (X_i-\mu)^2}{n-1}\right)^{-1}(\bar{X}-\mu)$$

而 $\sum_{i=1}^n (X_i-\mu)^2 \sim \sigma^2\chi^2_{n-1} \sim W_1(n,\sigma^2)$, 所以在 p 维正态时的相应统计量应该是 T^2 统计量 $n(X_i-\mu)'S^{-1}(X_i-\mu)$，其中

$$\begin{aligned} &S = \frac{1}{n-1}\mathbf{X}'(I-\frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{X} \equiv \frac{1}{n-1}A \\ &\Rightarrow n(n-1)(\bar{X}-\mu)'A^{-1}(\bar{X}-\mu) \sim T^2_{p,n-1} \end{aligned}$$

四. Wilks-Λ 统计量

在比较两个或多个一元正态总体均值的检验中，当它们有相同的方差时，我们可以根据两个（或多个）总体的样本构造 F 统计量，然后根据构造的 F 统计量来检验均值是否相等。也可以构造无偏的 χ^2 统计量来估计各自的方差，再根据样本方差构造 F 统计量来检验方差是否相等。当总体为 p 维正态时，相应的 χ^2 统计量换成了 Wishart 统计量，一般而言 Wishart 统计量是矩阵，如何构造相应的 F 统计量？由于统计量是个随机变量，所以要把 Wishart 统计量转换为一个随机变量。把一个矩阵用一个数表达，就是矩阵的某个数字特征。在多元统计中常用的数字特征是矩阵的行列式或最大特征根。为此我们引入如下的定义。

定义2.9. 设 $A \sim W_p(n,\Sigma)$, $B \sim W_p(m,\Sigma)$, A 与 B 独立，$\min(n,m) > p, \Sigma > 0$, 称

$$\Lambda = \frac{|A|}{|A+B|}$$

服从 *Wilks* 分布，记为 $\Lambda \sim \Lambda_{p,n,m}$.

设 $b \sim N_p(0,\Sigma)$, 则 $bb' \sim W_p(1,\Sigma)$, 此时由第一章线性代数知识得

$$\begin{aligned} &\Lambda = \frac{|A|}{|A+B|} = (1+b'A^{-1}b)^{-1} \\ &b'A^{-1}b \sim T^2, \ \Rightarrow \Lambda = \frac{1}{1+T^2} \Rightarrow \frac{1}{\Lambda} - 1 \sim T^2 \end{aligned}$$

§2.4 相关性分析和关联性分析

一. 相关关系和相关系数

两个变量之间的关系从大的方面来说，可以分为两大类：确定性关系和不确定性关系。确定性关系通常是有数学、物理、化学等学科给出的关系，如正方形面积是边长的平方。不确定性关系称为相关关系。在经济、金融和生活中大量存在的是相关关系，如一般而言商品的供给量和价格的关系，实施宽松的货币政策和通货膨胀的关系，家庭收入和家庭支出的关系，吸烟与患肺癌的关系等等。如何定量地描述这种相关关系或关联性？这取决于数据。不同的数据要用不同的方法来处理。

一般，数据分为名义数据和定量数据，如把人群分为公务员、教师、企业员工、工人和农民等，分别以 $1,2,3,4,5$ 来表示，则这些数据之间没有大小之分，称为名义数据。定量数据又可分为有序数据和比例数据。如对顾客满意度的调查，从最不满意到最满意分为 5 档，记为 $1,2,3,4,5$，则这些数据有大小之分，越大表示越满意，但是它们之间的差是没有意义的，我们称为有序数据。比例数据就是我们通常提到的数据，它们有零点，数据之间的差也是有意义的，如时间，金钱，路程等等。

1. 比例数据之间的相关系数

如果两个变量 X 和 Y 的取值都是比例数据，例如家庭的收入和支出，它们之间是有关系的。一般而言，收入多的家庭支出也会多一些。所以我们考虑随机向量 (X,Y), 现在我们有 n 组数据 $(x_1,y_1),\cdots,(x_n,y_n)$, 如何来反映它们之间的关系？我们希望这个量与变量的量纲无关，其值在 $[-1,1]$ 之间，值为零表示这两个量没有关系。从理论讲，可以把变量 X 和 Y 视为随机变量，用随机变量的相关系数

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

来表达它们之间的相关关系. 一般理论上的相关系数是不知道的，我们可以从数据来定义样本相关系数（又称为 Pearson 相关系数）

$$r_{XY}=\frac{\sum_{i=1}^n(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^n(x_i-\bar{x})^2\sum_{i=1}^n(y_i-\bar{y})^2}}$$

可以证明 ρ 和 r 的取值都在 $[-1,1]$ 之间，$\rho=0$ 和 $r=0$ 表明两个变量线性不相关，如果值大于零称为正相关，表示一个量随着另一个量的增加而增加，如果值小于零称为负相关，表示一个量随着另一个量的增加而减少。

注： ρ 和 r 仅仅刻画了两个变量之间的线性相关程度，即 $\rho=0$ 或 $r=0$ 仅仅说明两者之间没有线性关系，但是不代表两个量之间没有函数关系，如设 $X\sim R(-1,1)$, $Y=\cos(\pi X)$, 由计算得 $\rho_{XY}=0$.

比例数据之间的相关系数是我们用的最多的，但是一定要注意到只能对比例数据来做。

2. 有序数据对之间的相关系数和相关性分析

有序数据是经济和金融领域经常出现的数据，如顾客满意度和名次等等。由于它们之间的差没有意义，所以由它们得到的均值和方差都是没有意义的。如何来反映这类数据之间的相关性？我们在这儿介绍几种常用的相关系数和 γ 系数。

(i). Sperman 秩相关系数

设 X 和 Y 都是有序数据，取值为 $1,2,\cdots,k$. 我们有 n 组有序数据 $(x_1,y_1),\cdots,$

(x_n,y_n)，定义 Sperman 秩相关系数为

$$r=1-\frac{6\sum_{i=1}^n d_i^2}{n(n^2-1)} \tag{2.3}$$

这儿 $d_i=y_i-x_i.i=1,\cdots,n$, 数学上可以验证 Sperman 秩相关系数 r 的值在 $[-1,1]$ 之间. 也有和比例数据相关系数相类似的解释。如果把 X 和 Y 视为取值为 $1,2,\cdots,k$ 的两个离散随机变量，以 $rank(x_i)$ 表示 x_i 在数据 $x_1,\cdots,x_n$ 从小到大排列中的序位，$rank(y_i)$ 表示 y_i 在数据 $y_1,\cdots,y_n$ 从小到大排列中的序位，则 Sperman 秩相关系数 r 可以表达为

$$r=\frac{12}{n(n^2-1)}\sum_{i=1}^n(rank(x_i)-\frac{n+1}{2})(rank(y_i)-\frac{n+1}{2})$$

在理论上，可以定义离散型随机变量 X 和 Y 的相关系数

$$\rho_S(X,Y)=\rho(F_1(X),F_2(Y))$$

其中 F_1,F_2 分布为 X,Y 的分布函数。对应于相关系数 $\rho_S(X,Y)$ 的样本相关系数就是 Sperman 秩相关系数。

(ii).Kendell$-\tau$

设 X 和 Y 分别取值为 $1,2,\cdots,I$ 和 $1,2,\cdots,J$ 的有序数据，相关系数的符号决定了一个变量随另一个变量变化的方向，所以在理论上可以考虑事件 $\{(X_1-X_2)(Y_1-Y_2)>0\}$ 和 $\{(X_1-X_2)(Y_1-Y_2)<0\}$, 事件 $\{(X_1-$

$X_2)(Y_1 - Y_2) > 0\}$ 表示变量 X 变大时变量 Y 也跟着变大，X 变小时 Y 跟着变小，即它们是同向变化的，我们把 $(X_1, Y_1), (X_2, Y_2)$ 称为同向对，反之，$\{(X_1 - X_2)(Y_1 - Y_2) < 0\}$ 表示 Y 的变化方向与 X 的变化方向相反，我们称这样的对子 $(X_1, Y_1), (X_2, Y_2)$ 为异向对。据此我们把同向对和异向对概率的差定义为 X 和 Y 之间的相关系数，

$$\rho_\tau(X, Y) = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0)$$

与之相对应，根据 n 对数据 $(x_1, y_1), \cdots, (x_n, y_n)$，我们定义样本相关系数

$$\hat{\rho}_\tau(X, Y) = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i<j} sign[(X_i < X_j)(Y_i < Y_j)], \tag{2.4}$$

称为 Kendell$-\tau$。这儿一共有 n 对 $(x_i, y_i), i = 1, \cdots, n$, 从中任选两对，一共有 $\begin{pmatrix} n \\ 2 \end{pmatrix}$ 种不同的选法，则 $\sum_{i<j, sign[(X_i<X_j)(Y_i<Y_j)>0]}$ 是同向对的对子数，而 $\sum_{i<j, sign[(X_i<X_j)(Y_i<Y_j)<0]}$ 是异向对的对子数. 所以 Kendell$-\tau$ 是同向和异向对子数比例的差。

(iii). 系数 γ

设变量 X 有 I 个水平 $1, 2, \cdots, I$, 变量 Y 有 J 个水平 $1, 2, \cdots, J$, 设各水平是有序的，不妨设为升序，设它们的联合分布为 $P(X = i, Y = j) = \pi_{ij}, i = 1, \cdots, I, j = 1, \cdots, J$, 记

$$\begin{aligned}
\pi_{i\cdot} &= \sum_{j=1}^{J} \pi_{ij},\ \pi_{\cdot j} = \sum_{i=1}^{I} \pi_{ij}, \\
\pi_{\cdot\cdot} &= \sum_{i=1}^{I} \pi_{i\cdot} = \sum_{j=1}^{J} \pi_{\cdot j} = 1, \\
\Pi_c &= 2 \sum_i \sum_j \pi_{ij} (\sum_{h>i} \sum_{k>j} \pi_{hk}), \\
\Pi_d &= 2 \sum_i \sum_j \pi_{ij} (\sum_{h>i} \sum_{k<j} \pi_{hk}),
\end{aligned}$$

Π_c 和 Π_d 分别称为 X, Y 同向 (concordance) 的概率和异向 (discordance) 的概率，定义

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d} \tag{2.5}$$

γ 反映了有序变量间的关联程度，称为 γ 系数，由定义，$|\gamma| < 1$, $\gamma \to 1$ 表示 X, Y 越倾向于同向，$\gamma \to -1$ 表示 X, Y 越倾向于异向。$\Pi_d = 0$, 则 $\gamma = 1$, $\Pi_c = 0$, 则 $\gamma = -1$.

注 1： 联合概率函数 π_{ij} 和边缘分布 $\pi_{i\cdot}, \pi_{\cdot j}$, 可以由 (X,Y) 的 $I\times J$ 数据表中估出。(有一个表) 设事件 $\{X=i, Y=j\}$ 在 n 组数据中有 n_{ij} 个，则 $\pi_{ij}, \pi_{i\cdot}, \pi_{\cdot j}$ 的估计分别为

$$\hat{\pi}_{ij} = n_{ij}/n,\ \hat{\pi}_{i\cdot} = n_{i\cdot}/n = \sum_{j=1}^{J} n_{ij}/n,$$
$$\hat{\pi}_{\cdot j} = n_{ij}/n = \sum_{i=1}^{I} n_{ij}/n,$$

注 2： 经过计算，可以得到

$$\begin{aligned}\Pi_c &= 2\sum_i\sum_j \pi_{ij}P(X>i, Y>j) = 2P(X_2>X_1, Y_2>Y_1)\\ &= P((X-2-X_1)(Y_2-Y_1)>0),\\ \Pi_d &= 2\sum_i\sum_j \pi_{ij}P(X>i, Y<j) = P((X_2-X_1)(Y_2-Y_1)<0),\\ \Pi_c &+ \Pi_d \le 1\end{aligned}$$

其中 $(X_1,Y_1),(X_2,Y_2)$ 为 iid。

3. 名义数据之间的关联性分析

如果变量 X,Y 中至少有一个名义变量, 或两个都是名义变量时, 讨论两个变量是否同向或异向以及单调性就没有什么意义了。Goodman, Kruskel(1954) 和 Theil(1970) 分别定义了如下的关联性系数 τ 和 U。

大部分对名义数据关联性可解释的指标与回归中的可决系数 R^2 有相同的结构，其中 R^2 和相关系数 ρ 描述了边缘分布与响应条件分布之间变差比值的一种简约形式。根据这一想法，Goodman , Kruskel(1954) 和 Theil(1970) 对名义变量首先构造了变量取值概率差异的度量（这种取值概率的差异我们称之为差异），然后给出变量之间的关联性指标。设 $V(Y)$ 是对响应变量 Y 的边缘分布 $(\pi_{\cdot 1}, \cdots, \pi_{\cdot J})$ 变异程度的一种度量, $V(Y|i)$ 表示在解释变量 X 的第 i 水平上 Y 的条件分布 $\pi_{1|i}, \cdots, \pi_{J|i}$ 变异程度的度量，则在变异度量中的比值关系式为

$$\frac{V(Y) - E[V(Y|X]}{V(Y)}$$

当 X 为名义数据时，$EV(Y|X) = \sum_{i=1}^{I} \pi_{i\cdot} V(Y|i)$, 对名义响应变量 Y 的变异程度的一种度量为

$$V(Y) = \sum_{j=1}^{J} \pi_{\cdot j}(1-\pi_{\cdot j}) = 1 - \sum_{j=1}^{J} \pi_{\cdot j}^2.$$

这是来自于 Y 的边缘分布的两个独立观测值 Y_1, Y_2 落在不同类中的概率 $P(Y_1 \neq Y_2)$, 若存在 j, 使得 $\pi_{\cdot j} = 1$, 则 $V(Y) = 0$ 取到最小值。

第 i 行的条件变异为

$$V(Y|i) = 1 - \sum_{j=1}^{J} \pi_{j|i}^2$$

这是在给定 $X = i$ 下 Y_1, Y_2 条件独立的假定下得到的：

$$\begin{aligned} P(Y_1 \neq Y_2|i) &= \sum_{j=1}^{J} P(Y_1 \neq j, Y_2 = j|i) \\ &= \sum_{j=1}^{J} P(Y_1 \neq j|i) P(Y_2 = j|i) \\ &= \sum_{j=1}^{J} (1 - \pi_{j|i}) \pi_{j|i} = 1 - \sum_{j=1}^{J} \pi_{j|i}^2 \end{aligned}$$

对 $I \times J$ 列联表，设 X, Y 的联合分布为 $\pi_{ij}, i = 1, \cdots, I, j = 1, \cdots, J$，则平均变异为

$$EV(Y|i) = \sum_{i=1}^{I} \pi_{i\cdot}^2 EV(Y|i)$$

故

$$EV(Y|i) = 1 - \sum_{i=1}^{I} \pi_{i\cdot} \sum_{j=1}^{J} \pi_{j|i}^2 = 1 - \sum_{i} \sum_{j} \pi_{ij}^2 / \pi_{i\cdot}$$

所以，度量变异的比值为

$$\tau = \frac{\sum_i \sum_j \pi_{ij}^2 / \pi_{i\cdot} - \sum_j \pi_{\cdot j}^2}{1 - \sum_j \pi_{\cdot j}^2}, \tag{2.6}$$

称为集中 (concentration) 系数。这是 Goodman , Kruskel(1954) 提出的。

Theil(1970) 提出了另一种变异的度量：$V(Y) = \sum_{j=i}^{J} \pi_{\cdot j} \log \pi_{\cdot j}$, 由此得到如下的关联性指标 U,

$$U = \frac{-\sum_i \sum_j \pi_{ij} \cdot \log(\pi_{ij} / (\pi_{i\cdot} \pi_{\cdot j}))}{\sum_j \pi_{\cdot j} \cdot \log(\pi_{\cdot j})}. \tag{2.7}$$

§2.5 其他重要的多元分布 *

由于在经济和金融工程中遇到的大量问题是多元的，所以有必要把一元中的一些重要统计量推广到多元，例如我们已把一元正态推广到多元正态，但是如何把 t 分布和 F 分布等推广到多元？它们都与 χ^2 分布有关，而 χ^2 分布

到多元的推广是 $Wishart$ 分布，在统计理论中，已有很成熟的方法，称为多元 $Beta$ 分布，读者可参见张润楚的《多元统计分析》（2006，科学出版社）。其他一些常用分布有

1. 多元 t 分布

定义2.10. 如果随机向量 $X=(X_1,\cdots,X_p)'$ 有如下的概率密度函数

$$f_v(t;\mu,\Sigma)=c_p|\Sigma|^{-1/2}[1+v^{-1}(t-\mu)'\Sigma^{-1}(t-\mu)]^{-(v+p)/2}, \tag{2.8}$$

其中 t,μ 为 p 维向量，Σ 为 $p\times p$ 正定阵，$c_p=\Gamma((v+p)/2)(\pi v)^{-p/2}\Gamma^{-1/2}(v/2)$，则称 X 有 p 元 t 分布，v 称为分布的自由度。

当 $p=1$ 时，它就是一元 t 分布。

多元 t 分布也可以这样来定义。设 $X\sim N_p(\mu,\Sigma)$，$Y\sim\chi_v^2$，且相互独立，记 $U_i=X_i/(Y/v),i=1,\cdots,p$，$U=(U_1,\cdots,U_p)$，则称 U 为有参数 μ,Σ,v 的 p 元 t 分布。在这个定义中，显然 U 的每个分量都是 t 分布，p 个 t 分布随机变量构成的随机向量的分布就叫 p 维 t 分布。

2. 多元 $Cauchy$ 分布

在多元 t 分布定义中，如果 $v=1$，就称为 p 元 $Cauchy$ 分布。

3. $Dirichlet$ 分布

我们知道一元 $Beta$ 分布的密度函数是

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-x)^{\alpha-1}x^{\beta-1},\ 0<x\leq 1, \tag{2.9}$$

作如下推广就得到 Dilihelet 分布：

定义2.11. 若随机向量 $X=(X_1,\cdots,X_p)'$ 有如下的概率密度函数

$$\begin{aligned}&f(x_1,\cdots,x_p)=\frac{\Gamma(\sum_{i=0}^p a_i)}{\prod_{i=0}^p\Gamma(a_i)}\left(1-\sum_{i=1}^p x_i\right)^{a_0-1}\prod_{i=1}^p x_i^{a_i-1},\\&x_i>0,\ i=1,\cdots,p,\ \sum_{i=1}^p x_i\leq 1,\end{aligned}$$

其中 $\alpha_i>0,i=1,\cdots,p$ 为参数，则该分布称为 $Dirichlet$ 分布，记为 $D(a_0,a_1,\cdots,a_p)$。

这一分布在成分分析中非常有用。它是离散多项分布到连续型随机变量的推广。

4. 逆 $Wishart$ 分布

在多元统计分析中，经常会遇到缺失数据的问题，通常对这类数据的一个基本假定是它们服从正态分布 $N_p(\mu,\Sigma)$, 但是分布中的协方差矩阵一般也是不知道的，为了要生成数据，必须要假定协方差矩阵是从某个总体中抽取的一个样本。这个总体我们通常选用逆 $Wishart$ 分布。一个正定的 $p\times p$ 随机矩阵 Σ 称为逆 $Wishart$ 分布, 如果 Σ^{-1} 服从 $Wishart$ 分布。有关理论可以参见缺失数据分析的统计书籍。

5. 成分分布

成分数据分析的统计问题是一个富有挑战性的工作，由于它在经济和金融工程中有大量的应用，所以近年来受到人们的关注，有关内容可以参见张尧庭教授的《成分数据统计分析引论》（2000，科学出版社），也可参见《成分数据的统计分析》（J. Aitchison 著，中译本，1990 中国地质大学出版社）。我们仅仅在这儿介绍一点基本概念和有关的分布。

定义2.12. *设* $w=(w_0,w_1,\cdots,w_p)'\in R^{+(p+1)}, w\neq 0$, *记*

$$\begin{aligned} t &= \sum_{i=1}^{p} w_i, \\ x_i &= \frac{w_i}{t},\ i=1,\cdots,p \end{aligned}$$

则称 t *是* w *相应的总量 (total)*, $x=(x_1,\cdots,x_p)$ *称为* w *相应的成分 (composition)*。

易见

$$\begin{aligned} &w=tx,\ x=w/t \\ &t>0, x\geq 0, \mathbf{1}'x=1\ w\neq 0 \end{aligned}$$

当然，从数学的角度，可以把成分概念推广，如可以为负值等，但在经济和金融工程中，我们理解成分为一组和为 1 的非负实数，即通常我们讲的权。但是由于 $W>0$，拟合 W 的分布要在 R^{1+p} 中取值，如常用的多元正态就不合适了。解决问题的方法之一是对向量 w 和 x 取对数。取对数后的向量分别用 $\ln w$ 和 $\ln x$ 表示，即

$$\left\{\begin{array}{c} \ln w_0 \\ \ln w_1 \\ \vdots \\ \ln w_p \end{array}\right. , \quad \left\{\begin{array}{c} \ln x_0 \\ \ln x_1 \\ \vdots \\ \ln x_p \end{array}\right.$$

$\ln w$ 和 $\ln x$ 可以在 R^{1+p} 和 R^p 中取值, 能用的分布就广了。如果 $a'\mathbf{1}=0$, 我们称 $a'x$ 是 x 的一个对比 (contrast), 我们称 $\ln w$ 是 w 的一个对数对比 (logcontrast), 同样, 把 $\ln x$ 称为是 x 的一个对数对比.

定义2.13. *设 $w\in R^{+(p+1)}, w\neq 0$, 如果存在正值函数 G, 对一切 $c>0$, 有*

$$G(cw)=cG(w)$$

则称 G 是 w 的大小 (size), 称

$$z_G(w)=w/G(w)$$

是 w 的形状 (shape).

由定义，大小不是惟一的，如 $(\sum_{i=0}^p w_i^2)^{1/2}$ 和 $\max_{0\le i\le p} w_i$ 都是大小。但是它们对应的形状是成比例的，所以形状和成分仅仅相差一个比例因子。形状可以描述有多个指标的经济体。例如衡量一个地区的经济发展方式可以从第一产业，第二产业和第三产业这三个产业的经济量来考虑，如何比较两个地区的发展方式是否相同，相似还是完全相反？形状就可以比较形象地描述出来。《红楼梦》前八十回和后四十回是否是同一个作者？可以从单词使用的频数等来做比较，用一定量的单词来构造形状，从一个方面可以来描述是否为同一作者。

一般，是先假定知道 W 的联合分布，再由此导出成分 X 的联合分布，然后进行如何成分数据统计分析. 对于 W 的联合分布，有一些是假定 W 的各分量独立，有相同的分布类型。例如，设 W 的分量独立，W_i 分别服从反正态分布 (inverse Gaussian distribution)$IN(\mu_i,\mu_i^3/\lambda)$, 即 W_i 的密度函数为

$$g(t)=\sqrt{\frac{\lambda}{2\pi}}t^{-3/2}\exp-\frac{\lambda}{2\mu^2 t}(t-\mu)^2,\ t>0$$

其中参数 $\mu>0,\lambda>0$. 可以通过计算得到 $EW_i=\mu, Var(W_i)=\mu^3/\lambda$. 其所以命名为反正态分布是由于它的矩母函数

$$E\exp\{tW_i\}=\exp\{\frac{\lambda}{\mu}(1-\sqrt{1-2\mu^2t/\lambda})\}$$

而正态变量 X 的分布 $N(\mu,\sigma^2)$ 的矩母函数为 $\theta=E\exp\{tX\}=\exp\{\mu t+\sigma^2t^2/2\}$, 其反函数为

$$t=-\frac{\mu}{\sigma^2}(1\pm\sqrt{1-2\sigma^2\theta/\mu^2})$$

很显然，上式与反正态分布的矩母函数的对数完全吻合，在这个意义上我们可以说，它是正态的反函数。如果记 $\beta = \lambda/\mu^2$，则 $IN(\mu, \mu^3/\lambda)$ 可以写为 $IN(\mu, \mu/\beta)$. 我们有如下命题

命题2.3. 若 $X_i, i = 1, \cdots, p, iid. \sim IN(\mu, \mu^3/\lambda)$，则

$$S = \sum_{i=1}^{p} X_i \sim IN(p\mu, p\mu^3/\lambda)$$

如果 $X_1, \cdots, X_p$ 独立，$X_i \sim IN(\mu_i, \mu_i/\beta, i = 1, \cdots, p)$，则

$$S = \sum_{i=1}^{p} X_i \sim IN(\sum_{i=1}^{p} \mu_i, \sum_{i=1}^{p} \mu_i/\beta)$$

注意在此地 $\beta = \lambda_i/\mu_i^2 = EX_i/Var(X_i)$ 与 i 无关, 所以参数用 (μ, β) 比较合适。

例2.2. 多元对数正态分布

设 $X \sim N(\mu, \Sigma)$ 为 p 维正态，记 $Y = (Y_1, \cdots, Y_p)'$，$Y_i = \exp\{X_i\}, i = 1, \cdots, p$，即 $X = \ln Y$，则 Y 称为多元对数正态，记为 $Y \sim \Lambda(\mu, \Sigma)$，其密度函数为

$$f(y) = \begin{cases} \left(\frac{1}{2\pi}|\Sigma|^{-1/2}(\prod_{i=1}^{p} y_i)^{-1} \exp\{-\frac{1}{2}(\ln y - \mu)'\Sigma^{-1}(\ln y - \mu)\}\right)^p, & y > 0, \\ 0, & \text{其他}. \end{cases}$$

经过比较复杂的计算可得

$$EY_i = \exp\{\mu_i + \sigma_{ii}/2\}, i = 1, \cdots, p, \tag{2.10}$$

$$Cov(Y_i, Y_j) = \exp\{\mu_i + \mu_j + (\sigma_{ii} + \sigma_{jj})/2\}(\exp\{\sigma_{ij}\} - 1) \tag{2.11}$$

进一步我们可以考虑如果 $W = (W_0, W_1, \cdots, W_p)' \sim \Lambda(\mu, \Sigma)$, 则 p 维成分向量 X 的分布是什么？请参阅张尧庭教授的《成分数据统计分析引论》。如果 $W = (W_0, W_1, \cdots, W_p)'$ 服从 $Dirichlet$ 分布，则总量 T 和成分向量 X 的分布会是什么？

命题2.4. 设 $W_i, i = 0, 1, \cdots, p$ 相互独立，分别服从参数为 a_i, b 的伽马分布 $(W_i \sim \Gamma(a_i, b))$，则

(1) 总量 $T = \sum_{i=0}^{p} W_i$ 和成分向量 $X = (X_1, \cdots, X_p)'$ 相互独立，

(2) $T \sim \Gamma(a, b)$，其中 $a = \sum_{i=0}^{p} a_i$，

(3) $X \sim D(a_0, a_1, \cdots, a_p)$，

若记 $x_0 = 1 - \sum_{i=1}^p x_i \geq 0, x = (x_0, x_1, \cdots, x_p)'$, 则 *Dirichlet* 分布的密度函数可以写为对称的形式：

$$\begin{aligned} &f(x_0, x_1, \cdots, x_p) = \frac{\Gamma(\mathbf{1}'a)}{\prod_{i=0}^p \Gamma(a_i)} \prod_{i=0}^p x_i^{\alpha_i - 1}, \\ &x > 0, \mathbf{1}'x = 1, \end{aligned}$$

由上面的表达式我们可以得到 X 的均值向量和协方差矩阵. 其中的元素为

$$\begin{aligned} &EX_i = \frac{a_i}{\mathbf{1}'a},\ i = 1, \cdots, p \\ &E(X_iX_j) = \frac{a_ia_j}{(\mathbf{1}'a+1)\mathbf{1}'a},\ i \neq j, \\ &EX_i^2 = \frac{a_i(a_i+1)}{(\mathbf{1}'a+1)\mathbf{1}'a},\ i = 1, \cdots, p \\ &Var(X_i) = \frac{a_i(\mathbf{1}'a - a_i)}{(\mathbf{1}'a+1)(\mathbf{1}'a)^2},\ i = 1, \cdots, p \\ &Cov(X_i, X_j) = -\frac{a_ia_j}{(\mathbf{1}'a+1)(\mathbf{1}'a)^2},\ i \neq j. \end{aligned}$$

§2.6 大数定律和中心极限定理

这儿我们介绍几个多元统计量常用的极限定理。设 p 维随机向量 X 来自总体 $(\mu_p, \Sigma)_{p\times p}$, 这儿 μ, Σ 分别为总体的均值向量和协方差矩阵, 则当 $n \to \infty$ 时, 有

$$(1)\ \overline{X} \to \mu,\ inProb.(\text{均值弱大数定律}), \tag{2.12}$$

$$(2)\ S = \frac{1}{n}A \to \Sigma,\ inProb.(\text{协方差弱大数定律}), \tag{2.13}$$

$$(3)\ \sqrt{n}(\bar{X} - \mu) \to N_p(0, \Sigma),\ (\text{均值中心极限定理}), \tag{2.14}$$

$$(4)\ \text{设}E||X||^4 < \infty, \text{则}\sqrt{n}(S - \Sigma) \to N_{p^2}(0, V)(\text{以分布收敛}). \tag{2.15}$$

其中 $A = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$, V 为 $(X_1 - \bar{X})(X_1 - \bar{X})'$ 拉直后的协方差矩阵。

证明的方法有很多, 常用的有两种, 一种是用特征函数, 另一种是把多维的随机向量 X 化为随机变量 $a'X$, 其中 $a \in R^p$ 是任意的, 然后用一元的有关结果。

第三章　时间序列分析

§3.1　差分方程

要对时间序列作分析，理论基础之一就是差分方程，所以在本章我们将对有关的内容作一介绍。

§3.1.1　常系数齐次差分方程

定义3.1. 如果序列 $\{a_n, n=0,1,\cdots,\}$ 满足下面的递推关系：

$$a_n = \beta_1 a_{n-1} + \cdots + \beta_p a_{n-p},\ n = p, p+1, \cdots \tag{3.1}$$

$$x^p - \beta_1 x^{p-1} - \cdots - \beta_p = 0 \tag{3.2}$$

其中 $\beta_1, \cdots, \beta_p$ 为常数，则序列 $\{a_n\}$ 称为 p 阶常系数齐次差分方程。p 次多项式 *(3.2)* 称为对应的特征方程。

定义中的齐次表示递推式中没有常数项和其它的变量。我们有如下的命题

命题3.1. 设 $x_1, \cdots, x_p$ 为 *(3.2)* 的 p 个根（包括复根），则 *(3.1)* 的通解为

$$a_n = \alpha_1 x_1^n + \alpha_2 x_2^n + \cdots + \alpha_p x_p^n,\ n = p, p+1, \cdots, \tag{3.3}$$

其中系数 $\alpha_1, \cdots, \alpha_p$ 由初值 $a_0, a_1, \cdots, a_{p-1}$ 给定。

证： 我们先证明齐次差分方程的通解有线性性，即如果 $\{b_n\}$ 和 $\{c_n\}$ 都是 (3.1) 的解，则线性组合 $\gamma_1 b_n + \gamma_2 c_n$ 仍是 (3.1) 的解，由假定，

$$b_n = \beta_1 b_{n-1} + \cdots + \beta_p b_{n-p},$$
$$c_n = \beta_1 c_{n-1} + \cdots + \beta_p c_{n-p}$$

因此

$$\gamma_1 b_n + \gamma_2 c_n = \alpha_1(\gamma_1 b_{n-1} + \gamma_2 c_{n-1}) + \cdots + \alpha_p(\gamma_1 b_{n-p} + \gamma_2 c_{n-p})$$

即满足 (3.1). 设 x_1 是特征方程 (3.2) 的任意一个根，令 $a_n = x_1^n, n \geq p$, 由于

$$x_1^p = \beta_1 x_1^{p-1} + \cdots + \beta_{p-1} x_1 + \beta_p = 0$$
$$\Rightarrow x_1^n = \beta_1 x_1^{n-1} + \cdots + \beta_{p-1} x^{n-p+1} + \beta_p x_1^{n-p} = 0$$

即 $a_n = x_1^n$ 是齐次差分方程 (3.1) 的解，由于解有线性性，所以 (3.3) 为 (3.1) 的解。进一步可以证明，如果给定初值 $a_0, a_1, \cdots, a_{p-1}$，则解是惟一的。

注： 当特征方程有重根时，例如设 $x_1 = \cdots = x_q$, 则 $x_1^n, nx_1^n, \cdots, n(n-1)\cdots(n-q+2)x_1^n$ 都是 (3.1) 的解，从而对应于 q 重根 x_1 的解是上面 q 个解的线性组合。大体证明如下：设 x_1 为特征方程 (3.2) 的 q 重根，即特征方程为 $(x-x_1)^q f(x) = 0$, 从而 $(x-x_1)^q = 0$, 展开得

$$\sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x^{q-k} x_1^k (-1)^k = 0,$$

两边乘以 x^{n-q}, 有

$$\sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x^{n-k} x_1^k (-1)^k = 0,$$

定义后移算子 B, $Ba_n = a_{n-1}, B^2 a_n = a_{n-2}, \cdots$, 则由特征方程知 a_n 满足 $(1-x_1B)^q a_n = 0$, 所以

$$\sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x_1^k (-1)^k a_{n-k} = 0,$$

当 $a_n = x_1^n$ 时得

$$\sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x_1^k (-1)^k x_1^{n-k} = \sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x_1^n (-1)^k = 0,$$

要验证 $x_1^n, nx_1^n, \cdots, n(n-1)\cdots(n-q+2)x_1^n$ 都是 (3.1) 的解，例如 $n(n-1)\cdots(n-q+2)x_1^n$ 是 (3.1) 的解。只要验证

$$(1-x_1B)^q a_n = \sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x_1^k (-1)^k a_{n-k} = 0,$$

把 $a_n = n(n-1)\cdots(n-q+2)x_1^n$ 代入上式，只要验证

$$\begin{aligned} &\sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x_1^k (-1)^k (n-k)(n-k+1)\cdots(n-k-q+2)x_1^{n-k} \\ &= x_1^n \sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} (-1)^k (n-k)(n-k+1)\cdots(n-k-q+2) = 0, \end{aligned}$$

令 $g(x) = x^{n-q}(x-1)^q$, 则

$$\begin{aligned} &g^{(q-1)}(x)\,|_{x=1} \\ &= \sum_{k=0}^{q} \begin{pmatrix} q \\ k \end{pmatrix} x^{n-k-q+1} (-1)^k (n-k)\cdots(n-k-q+2)x_1^{n-k}\,|_{x=1} = 0 \end{aligned}$$

即 $a_n = n(n-1)\cdots(n-q+2)x_1^n$ 是 (3.1) 的根。对 $g(x)$ 在 $x=1$ 处求 $j(j<q)$ 次导数可得 $a_n = n(n-1)\cdots(n-j+1)x_1^n$, $j \le q$ 都是 (3.1) 的解。由解的线性性得到 (3.1) 关于重根 x_1 的解的一般形式为 $(c_0 + c_1 n + \cdots + c_{q-1}n^{q-1})x_1^n$.

例3.1. 菲比那契级数。一对大兔子，假设一个月后生一对小兔子，小兔子一个月后再生一对小兔子，老兔子每个月生一对小兔子。问第 n 个月时的兔子总数。

解： 设 a_n 表示第 n 个月的兔子数，由题意，$a_0 = 1, a_1 = 1, a_n = a_{n-1} + a_{n-2}$，这是二阶常系数齐次差分方程，对应的特征方程为 $x^2 - x - 1 = 0$，其根为

$$x = \frac{1 \pm \sqrt{5}}{2}$$

（注意 $\frac{\sqrt{5}-1}{2} \approx 0.618$ 为黄金分割数）所以

$$a_n = c_1 \left(\frac{1+\sqrt{5}}{2}\right)^n + c_2 \left(\frac{1-\sqrt{5}}{2}\right)^n$$

由初值 $a_0 = a_1 = 1$ 最后可以得到菲比那契级数的解为

$$a_n = \frac{1}{\sqrt{5}} \left[\left(\frac{\sqrt{5}+1}{2}\right)^{n+1} + (-1)^n \left(\frac{\sqrt{5}-1}{2}\right)^{n+1} \right]$$

我们再仔细讨论 $p = 2$ 这一常见情况。设二阶实系数齐次差分方程和对应的特征方程为

$$a_n = \beta_1 a_{n-1} + \beta_2 a_{n-2}$$
$$x^2 - \beta_1 x - \beta_2 = 0$$

特征方程的根为

$$x = \frac{\beta_1 \pm \sqrt{\beta_1^2 + 4\beta_2}}{2}$$

(1) $\beta_1^2 + 4\beta_2 = 0$, 方程有重根 $x_1 = \pm\beta_1/2$, 故差分方程解为

$$a_n = (c_0 + c_1 n)\left(\frac{\beta_1}{2}\right)^n.$$

(2) $\beta_1^2 + 4\beta_2 > 0$, 特征方程有两个不同的实根 x_1, x_2, 则差分方程解为

$$a_n = c_1 x_1^n + c_2 x_2^n.$$

(3) $\beta_1^2 + 4\beta_2 < 0$，特征方程有一对共轭复根，$x_1, x_2 = \alpha_1 \pm i\alpha_2 \equiv \rho \exp\{\pm i\theta\}$, 由解的线性性，$\rho^n \cos n\theta, \rho^n \sin n\theta$ 是差分方程的两个不同的实解，再由线性性，得到通解

$$a_n = \rho^n (c_1 \cos n\theta + c_2 \sin n\theta).$$

§3.1.2 非齐次差分方程

(1) 带常数的非齐次差分方程

设序列 $a_n, n \geq 0$ 满足

$$a_n = \beta_0 + \beta_1 a_{n-1} + \cdots + \beta_p a_{n-p} \tag{3.4}$$

我们研究方程 (3.4) 解的结构。

命题3.2. *设非齐次差分方程 (3.4) 对应的齐次差分方程*

$$b_n = \beta_1 b_{n-1} + \cdots + \beta_p b_{n-p}$$

的通解为 $b_n = c_1\lambda_1^n + \cdots + c_p\lambda_p^n$, $n \geq 0$, c_n 为 (3.4) 的一个特解, 则 (3.4) 的通解 a_n 为

$$a_n = b_n + c_n = c_n + c_1\lambda_1^n + \cdots + c_p\lambda_p^n,\ n \geq 0.$$

此命题说明带常数的非齐次差分方程通解是对应的齐次差分方程通解加 (3.4) 的一个特解。

证明是非常简单的，只要按定义验证即可。

(2) 带常系数的非齐次差分方程的特解

由上面的命题知，关键问题是导出 (3.4) 的一个特解。现在我们来导出它的一个特解。

先假设特解为最简单的常数 c, 看我们能否定出这个常数，如果定不出来，说明特解不可能是常数。如果特解是常数 c，则由 (3.4),

$$c = \beta_0 + \sum_{i=1}^{p} \beta_i c$$

当 $\beta_1 + \beta_2 + \cdots + \beta_p \neq 1$ 时，解出

$$c = \frac{\beta_0}{1 - \sum_{i=1}^{p} \beta_i}.$$

当 $\beta_1 + \beta_2 + \cdots + \beta_p = 1$ 时，对应于 (3.4) 的齐次差分方程的特征方程有等于 1 的根，设 (3.4) 有形如 αn 的特解（由上面讨论知，常数特解是不可能有了），代入 (3.4) 式，

$$\alpha n = \beta_0 + \alpha \sum_{i=1}^{p} \beta_i (n - i) = \beta_0 + \alpha n - \alpha \sum_{i=1}^{p} i\beta_i,$$

因此,

$$\alpha=\frac{\beta_0}{\sum_{i=1}^{p} i\beta_i},$$

即此时 (3.4) 的通解的形式为

$$a_n=c_0+\frac{\beta_0}{\sum_{i=1}^{p} i\beta_i}n+c_2\lambda_1^n+\cdots+c_p\lambda_p^n.$$

(3) 带变系数的非齐次差分方程

设序列 $\{a_n, n\geq 0\}$ 满足

$$a_n=b_n+\beta_1 a_{n-1}+\cdots+\beta_p a_{n-p} \tag{3.5}$$

其中 b_n, $n\geq 0$ 是已知的一个数列,我们研究方程 (3.5) 解的结构。

命题3.3. *设非齐次差分方程 (3.5) 对应的齐次差分方程*

$$a_n=\beta_1 a_{n-1}+\cdots+\beta_p a_{n-p}$$

的通解为 $a_n=c_1\lambda_1^n+\cdots+c_p\lambda_p^n$, $n\geq 0$, c_n 为 (3.5) 的一个特解, 则 (3.5) 的通解 a_n 为

$$a_n=c_n+c_1\lambda_1^n+\cdots+c_p\lambda_p^n,\ n\geq 0,$$

所以求 (3.5) 的通解的关键是要求出它的一个特解。下面我们用算子法来导出特解。后移算子 B 的定义同前。先讨论 $p=1$, 即一阶变系数差分方程。为方便起见,定义 $b_k=0, k<0$, 则 (3.5) 可用算子表达为

$$a_n=\beta_1 a_{n-1}+b_n,$$
$$(1-\beta_1 B)a_n=b_n,$$

故

$$a_n=\frac{1}{1-\beta_1 B}b_n=\sum_{k=0}^{\infty}\beta_1^k B^k b_n=\sum_{k=0}^{\infty}\beta_1^k b_{n-k},$$

注意到当 $n-k<0$, 即 $k>n$ 时, $b_{n-k}=0$, 故上面的无穷级数实际上仅有有限项, $a_n=\sum_{k=0}^{n}\beta_1^k b_{n-k}$ 是 (3.5) 的一个特解, 当 $p>1$ 时,由算子法把 (3.5) 式表为

$$(1-\beta_1 B-\beta_2 B^2-\cdots-\beta_p B^p)a_n=b_n,$$
$$a_n=\frac{1}{1-\beta_1 B-\beta_2 B^2-\cdots-\beta_p B^p}b_n,$$

形式上，由微积分中的部分分式表达，当方程 $1-\beta_1 B-\beta_2 B^2-\cdots-\beta_p B^p=0$ 的根 $\lambda_1,\cdots,\lambda_p$ 全部不等时，存在常数 $c_1,\cdots,c_p$，使得

$$\frac{1}{1-\beta_1 B-\beta_2 B^2-\cdots-\beta_p B^p}=\sum_{j=1}^{p}\frac{c_j}{1-\lambda_j B}.$$

所以

$$a_n=\sum_{j=1}^{p}\frac{c_j}{1-\lambda_j B}b_n,$$

其中 $c_j\frac{1}{1-\lambda_j B}b_n$ 化为上面讨论的一阶情况，因此可以得到 (3.5) 式的一个特解。当 $1-\beta_1 B-\beta_2 B^2-\cdots-\beta_p B^p=0$ 的根有重根时，也有部分分式的分解，不过要复杂一点，这儿就不讨论了，有兴趣的读者可以参阅部分分式的有关教科书（如《复分析》）。

§3.2 平稳过程的定义

当我们研究的某个经济指标为随机变量 X 时，刻画它统计特性的是它的分布 F，当经济指标为多个（记为 p 个），即可以把这若干个指标用随机向量 $X=(X_1,\cdots,X_p)'$ 来表达，此时刻画该随机向量统计特性的是它们的联合分布 F。如果我们要刻画某个经济指标在一段时间中的变化规律，如上一年人民币和美元的兑换率，由于汇率时刻在变化，可以把时间看成连续变化的，经济指标可以用 $\{X_t,t\in T\}$ 来表达，此时 $\{X_t,t\in T\}$ 称为一个随机过程 (stochastic process)，随机过程在每个时刻 t，都是随机变量。如何来刻画随机过程？下面的有限维分布族全面刻画了随机过程的统计特性。

定义3.2. 设 $X_t,t\in T$ 为一个随机过程，如果对任一 n 和任意 n 个时刻 $t_1,\cdots,t_n$，以 $F_{t_1,\cdots t_n}(x_{t_1},\cdots,x_{t_n})$ 记随机向量 $X_{t_1},\cdots,X_{t_n}$ 的联合分布函数，则称

$$\mathfrak{F}=\{F_{t_1,\cdots,t_n}(x_{t_1},\cdots,x_{t_n}),\forall n\in N,\forall t_1,\cdots,t_n\in T\}$$

为随机过程 $X_t,t\in T$ 的有限维分布族。

注： 这儿 T 可以是离散时间，也可以是连续时间段，也可以不是时间，如地面的海平面高度。

一般，随机过程的有限维分布族不容易得到，现实生活中我们也不需要知道有限维分布族。其次，对我们有意义的随机过程有一些特有的性质，特别一类重要的随机过程是平稳随机过程。

定义3.3. 设 $X_t, t \in T$ 为一个随机过程，如果对任一 n 和任意 n 个时刻 $t_1, \cdots, t_n \in T$, 对任意满足 $t_i + h \in T, i = 1, \cdots, n$ 的 h,

$$(X_{t_1}, \cdots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \cdots, X_{t_n+h}),$$

则称 $X_t, t \in T$ 为一个严格平稳的随机过程。

这儿 $\stackrel{d}{=}$ 表示两个随机向量的分布函数相同。但是严格平稳的定义也不好验证，在实际中能够验证的是随机过程的一阶矩和二阶矩，所以我们有如下较弱的定义：

定义3.4. 设 $X_t, t \in T$ 为一个随机过程，$EX^2(t) < \infty, \forall t \in T$, 且

$$EX_t = \mu,$$

$$Cov(X_t, X(t+h)) = R(h)$$

则称 $X_t, t \in T$ 为弱平稳过程（有些书上称为宽平稳过程）。

这儿要求所涉及到的下标都在定义域 T 中。

由弱平稳过程的定义知，随机过程的均值函数是不随时间变化的常数，而协方差函数仅与时间的差有关，与时间起点无关。今后我们研究的平稳过程就是指的弱平稳过程。由于矩的关系，两个随机过程的定义没有包含关系，如果随机过程的二阶矩存在，则严格平稳过程可以推出一定是弱平稳过程。在弱平稳过程中有几类我们感兴趣的特殊随机过程。

(1) Gauss 过程

定义3.5. 设 $\{X_t, t \in T\}$ 是一个随机过程，且 $EX^2(t) < \infty$. 如果它的任一有限维分布都是多元正态分布，则称 $\{X_t, t \in T\}$ 是一个 *Gauss* 过程。

由定义，由于多元正态分布函数完全由它的均值函数和协方差函数所惟一确定，所以 Gauss 过程的统计性质完全由它本身的均值函数和协方差函数惟一确定。因此若 Gauss 过程是一个严格平稳过程，则一定是弱平稳过程，反之，如果 Gauss 过程满足 $EX_t = \mu$, $Cov(X_t, X(t+h) = R(h)$, 即是一个弱平稳过程，则一定是一个严格平稳过程。因此，对 Gauss 过程而言，严格平稳和弱平稳是等价的。

(2) 独立增量过程

定义3.6. 设 $\{X_t, t \in T\}$ 是一个随机过程，若对任意 T 中的 $t_1, \cdots, t_n$, $X(t_2) - X(t_1), \cdots, X(t_n) - X(t_{n-1})$ 相互独立，则称该随机过程为独立增量过程。

(3) 独立平稳增量过程

定义3.7. 设 $\{X_t, t \in T\}$ 是一个独立增量过程, 而且对任意 T 中的 $t_1, t_2, t_1 + h, t_2 + h$,

$$(X_{t_1+h} - X_{t_1}) \stackrel{d}{=} (X_{t_2+h}, X_{t_2}),$$

则称该随机过程为独立平稳增量过程。

独立平稳增量过程是在经济和金融工程中用得最多的随机过程。

例3.2. 白噪声 $WN(0, \sigma^2)$ 和 $Gauss$ 白噪声序列

如果随机变量序列 $\{\varepsilon_n, n = 0, \pm 1, \pm 2, \cdots\}$, 满足

$$E\varepsilon_n = 0,\ E\varepsilon_n^2 = \sigma^2,$$
$$Cov(\varepsilon_i, \varepsilon_j) = 0,\ i \neq j,$$

则称 $\{\varepsilon_n\}$ 为白噪声序列, 记为 $\varepsilon_n \sim WN(0, \sigma^2)$. 进一步, 若对一切 n, $\varepsilon_n \sim N(0, \sigma^2)$, 则称为 $Gauss$ 白噪声序列.

§3.3 线性时间序列

时间序列是经济和金融工程中最常用的建模工具。其中最简单也是最重要的是线性时间序列。

§3.3.1 常用时间序列的定义

定义3.8. 设 $\{\varepsilon_t, t = 0, \pm 1, \pm 2, \cdots\}$ 是一列白噪声序列, $E\varepsilon_t = 0$, $Var(\varepsilon_t) = \sigma^2$, 若时间序列

$$X_t = \sum_{k=0}^{\infty} \varphi_k \varepsilon_{t-k},\ t = 0, \pm 1, \pm 2, \cdots, \tag{3.6}$$

其中 $\{\varphi_k\}$ 满足 $\sum_{k=0}^{\infty} |\varphi_k| < \infty$, 则 $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ 称为线性过程

线性过程中最重要的是平稳滑动平均过程 $MA(q)$, 平稳自回归过程 $AR(p)$, 平稳自回归滑动平均混合过程 $ARMA(p, q)$ 和单整自回归滑动平均混合过程 $ARIMA(p, d, q)$。理论上, 时间可以从 $-\infty$ 算起, 但在实际中最常见的情况是时间有起点, 所以下面我们主要假定 $t = 0, 1, 2, \cdots$, 但是为了推导方便起见不排除从 $-\infty$ 算起. 下面分别作一简要介绍。

以下没有特殊声明, 则都默认 $\{\varepsilon_t\}$ 是方差为 σ^2 的白噪声序列。

(1) 滑动平均过程 $MA(q)$

若线性过程 $\{X_t, t=0,1,2,\cdots\}$ 有表达式：

$$X_t = \varphi_t - \beta_1\varepsilon_{t-1} + \cdots - \beta_q\varepsilon_{t-q}, \tag{3.7}$$

则时间序列 $\{X_t, t=0,1,2,\cdots\}$ 称为 q 阶滑动平均 (moving average) 过程，记为 $MA(q)$。引入后移算子 B，记 $\phi(B) = 1 - \beta_1 B - \cdots - \beta_q B^q$, 则 $MA(q)$ 过程可以写为

$$X_t = \phi(B)\varepsilon_t,$$

(2) 自回归过程 $AR(p)$

若线性过程 $\{X_t, t=0,1,2,\cdots\}$ 有表达式：

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varphi_t, \tag{3.8}$$

则时间序列 $\{X_t, t=0,1,2,\cdots\}$ 称为 p 阶自回归 (auto-regressive) 过程，记为 $AR(p)$。引入后移算子 B，记 $\varphi(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^p$, 则 $AR(p)$ 过程可以写为

$$\varphi(B)X_t = \varepsilon_t,$$

(3) 自回归滑动平均混合过程 $ARMA(p,q)$

若线性过程 $\{X_t, t=0,1,2,\cdots\}$ 有表达式：

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varphi_t - \beta_1\varepsilon_{t-1} + \cdots - \beta_q\varepsilon_{t-q}, \tag{3.9}$$

则时间序列 $\{X_t, t=0,1,2,\cdots\}$ 称为 (p,q) 阶自回归滑动平均混合过程，记为 $ARMA(p,q)$。引入后移算子 B，则 $ARMA(p,q)$ 过程可以写为

$$\varphi(B)X_t = \phi(B)\varepsilon_t, \tag{3.10}$$

形式上

$$X_t = \frac{\phi(B)}{\varphi(B)}\varepsilon_t, \tag{3.11}$$

再把有理分式 $\phi(B)/\varphi(B)$ 展开为 B 的幂级数，注意到 $B^k\varepsilon_t = \varepsilon_{t-k}$，所以 $AR(p)$ 和 $ARMA(p,q)$ 都是线性过程。

(4) 单整自回归滑动混合过程 $ARIMA(p,d,q)$

令

$$\begin{aligned}
&\triangle X_t = X_t - x_{t-1} = (1-B)X_t,\\
&\triangle^2 X_t = X_t - 2x_{t-1} + X_{t-2} = (1-B)^2 X_t\\
&\cdots\cdots\\
&\triangle^d X_t = \triangle(\triangle^{d-1}X_t) = (1-B)^d X_t \equiv y_t
\end{aligned}$$

若上述 Y_t 满足 $\varphi(B)Y_t = \phi(B)\varepsilon_t$, 则 X_t 称为 d 阶单整 (integration) 自回归滑动混合过程, 记为 $ARIMA(p,d,q)$。形式上 $\varphi(B)Y_t = \varphi(B)(1-B)^d X_t = \phi(B)\varepsilon_t$.

时间序列是一类特殊的离散随机过程, 一个重要的问题是在什么情况下它是平稳的? 在统计上, 如果是平稳的, 则有关参数 $(\alpha_1,\cdots,\alpha_p,\beta_1,\cdots,\beta_q,p,d,q)$ 如何估计?

§3.3.2 ARMA(p,q) 的平稳性

由于 $MA(q), AR(p)$ 是 $ARMA(p,q)$ 的特殊情况, 所以我们主要研究 $ARMA$ (p,q) 的平稳性。

由后移算子表达式 (3.10), 设

$$\varphi(B) = (1-\lambda_1 B)(1-\lambda_2 B)\cdots(1-\lambda_p B), \tag{3.12}$$

由部分分式分解和 $Tailor$ 展开,

$$\begin{aligned}
X_t &= \frac{\phi(B)}{\varphi(B)}\varepsilon_t = \frac{1-\beta_1 B - \cdots - \beta_q B^q}{(1-\lambda_1 B)(1-\lambda_2 B)\cdots(1-\lambda_p B)}\varepsilon_t\\
&= \sum_{k=0}^{\infty}\varphi_k B^k \varepsilon_t = \sum_{k=0}^{\infty}\varphi_k \varepsilon_{t-k},
\end{aligned} \tag{3.13}$$

即 X_t 可以表为 $\cdots\varepsilon_{-1},\varepsilon_0,\varepsilon_1\cdots$ 的线性组合, 我们都将以这点为平台对时间序列进行研究。设 $\sum_{k=0}^{\infty}|\varphi_k| < \infty$ (绝对可和) 以及 $EX_t = 0$, 则交换期望号和求和号, 有

$$\begin{aligned}
Var(X_t) &= \sum_k \sum_\ell \varphi_k\varphi_\ell E\varepsilon_{t-k}\varepsilon_{t-\ell}\\
&= \sigma^2 \sum_{k=0}^{\infty}\varphi_k^2 \equiv \sigma_X^2 = R(0)\\
Cov(X_{t+\ell},X_t) &= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} E\varepsilon_{t+\ell-k}\varepsilon_{t-j}\\
&= \sigma^2 \sum_{k-j=\ell}\varphi_k\varphi_j = \sigma^2\sum_{j=0}^{\infty}\varphi_{j+\ell}\varphi_j = R(\ell),
\end{aligned}$$

由于 $\sum_{k=0}^{\infty}|\varphi_k|<\infty$ 可以推出 $\sum_{k=0}^{\infty}|\varphi_k|^2<\infty$, 而 $\varphi_k\varphi_{k_\ell}\le\frac{1}{2}(\varphi_k^2+\varphi_{k+\ell}^2)$, 所以

$$|\sum_{k=0}^{\infty}\varphi_{k+\ell}\varphi_k|<\infty$$

且 $Cov(X_t,X_{t+\ell})$ 与 t 无关，因此是弱平稳的。

由前几节的讨论我们知道，时间序列 $ARMA(p,q)$ 的通解是对应的齐次差分方程 $x_t=\alpha_1x_{t-1}+\cdots+\alpha_px_{t-p}$ 的通解加上特解 (3.12), 设 $\varphi(B)$ 有表达式 (3.11), $ARMA(p,q)$ 对应的齐次差分方程的特征方程为 $\varphi(x^{-1})=0$, 所以特征方程的根为 $\lambda_1^{-1},\cdots,\lambda_p^{-1}$, 对应的齐次差分方程的通解形如 $c_1\lambda_1^{-t}+\cdots+c_p\lambda_p^{-t}$, 通常我们希望当 $t\to\infty$ 时通解趋于零，即我们要求时间序列是渐近平稳的, 这就要求 $|\lambda_i|>1,i=1,\cdots,p$。从上面讨论中知，这相当于 $\varphi(B)=0$ 的根在单位圆外。此时时间序列 $ARMA(p,q)$ 的平稳性取决于特解是否是平稳的。由上面讨论知，$\sum_{k=0}^{\infty}|\varphi_k|<\infty$ 能保证特解 (3.10) 的平稳性。什么时候我们有 $\sum_{k=0}^{\infty}|\varphi_k|<\infty$? 设有理分式 $\varphi(B)/\varphi(B)$ 有如下的分解：

$$\begin{aligned}\frac{\varphi(B)}{\varphi(B)}&=\frac{1-\beta_1B-\cdots-\beta_qB^q}{(1-\lambda_1B)(1-\lambda_2B)\cdots(1-\lambda_pB)}\\&=(\gamma_0+\cdots+\gamma_{(q-p)\vee0}B^{(q-p)\vee0})+\frac{c_1}{1-\lambda_1B}+\cdots+\frac{c_p}{1-\lambda_pB}\end{aligned}$$

这儿 $(q-p)\vee0=\max\{q-p,0\}$, γ_i 是展开得到的常数. 把 $(1-\lambda_iB)^{-1}$ 展开，得到

$$\begin{aligned}&\frac{\varphi(B)}{\varphi(B)}=\sum_{k=0}^{\infty}\varphi_kB^k\\&\varphi_k=\gamma_kI(k\le(q-p)\vee0)+c_1\lambda_1^k+\cdots+c_p\lambda_p^k,\end{aligned}\tag{3.14}$$

当所有的 $|\lambda_i|<1$ 时，$\sum_{k=0}^{\infty}|\varphi_k|<\infty$, 如果有某个 $|\lambda_j|\ge1$, 则由 φ_k 的表达式 (3.12) 知 $\sum_{k=0}^{\infty}|\varphi_k|=\infty$，即发散。

综上讨论可知，当 $\varphi(B)=0$ 的根在单位圆外时，时间序列 $ARMA(p,q)$ 是平稳的，否则不是平稳的。由此也可以推知，单整序列有单位根，所以 $ARIMA(p,d,q)$ 过程 $\{X_t\}$ 本身不是平稳的。

例3.3. 讨论二阶自回归过程的平稳性

设二阶自回归过程为

$$X_t=\alpha_1X_{t-1}+\alpha_2X_{t-2}+\varepsilon_t$$

讨论系数满足什么条件二阶自回归过程是平稳性的。

解： 二阶自回归过程对应的 $\varphi(B)=0$ 的根在单位圆外等价于 $x^2-\alpha_1 x-\alpha_2=0$ 的根 λ_1,λ_2 在单位圆内。当自回归过程是平稳时，λ_1,λ_2 满足

$$\lambda_1+\lambda_2=\alpha_1,\quad \lambda_1\lambda_2=-\alpha_2,$$
$$|\lambda_i|<1,\ i=1,2,$$

由此推出系数 α_1,α_2 应该满足条件 $\alpha_2\pm\alpha_1<1,\ |\alpha_2|<1.$

§3.3.3 时间域上平稳 $ARMA(p,q)$ 的研究

对平稳 $ARMA(p,q)$ 的研究，可以在时间域上进行，研究的对象主要是自相关函数和偏相关函数，在频率域上主要是对谱密度的研究。

(1) 自相关函数 (ACF)

设平稳 $ARMA(p,q)$ 过程 $\{X_t\}$ 有表达式 (3.7)，$X_t=\sum_{k=0}^{\infty}\varphi_k\varepsilon_{t-k},\varphi_0=1$，其中 $\varepsilon_t\sim WN(0,\sigma_\varepsilon^2)$，由表达式知 X_t 为时刻 t 及之前误差 $\varepsilon_t,\varepsilon_{t-1},\cdots$ 的线性组合，由于 $\{\varepsilon_t\}$ 为白噪声，故 $\{\varepsilon_{t+1},\varepsilon_{t+2},\cdots\}$ 与 X_t 是不相关的。这是今后研究问题的出发点，经计算后可以得到

$$EX_t=0,$$
$$EX_t\varepsilon_t=\sigma^2,$$
$$E\varepsilon_{t-k}X_t=\varphi_k\sigma_\varepsilon^2,\ k\geq 0.$$

定义 $ARMA(p,q)$ 过程 $\{X_t\}$ 的自协方差函数 $\{R(k)\}$ 和自相关函数 (autocorrelation function)$\{\rho(k)\}$ 分别为

$$R(k)=Cov(X_{t+k},X_t),\quad k=0,\pm1,\pm2,\cdots \tag{3.15}$$
$$\rho(k)=\frac{R(k)}{R(0)},\quad k=0,\pm1,\pm2,\cdots \tag{3.16}$$

其中 $R(0)$ 就是过程 $\{X_t\}$ 的方差 $Var(X_t)$，记为 σ_X^2，自协方差函数和自相关函数与 $\{\varphi_k\}$ 有如下关系：

$$R(k)=R(-k)=\sigma_\varepsilon^2\sum_{j=0}^{\infty}\varphi_j\varphi_{j+k},$$
$$\rho(k)=\frac{\sum_{j=0}^{\infty}\varphi_j\varphi_{j+k}}{\sum_{j=0}^{\infty}\varphi_j^2},$$
$$\rho(0)=1$$

下面给出三个平稳时间序列的自协方差函数 $\{R(k),k=0,\pm1,\pm2,\cdots\}$

(i) $MA(q)$

滑动平均过程 $X_t=\varepsilon_t-\beta_1\varepsilon_{t-1}-\cdots-\beta_q\varepsilon_{t-q}$, 因此，我们有

$$\varphi(0)=1,\varphi_i=-\beta_i,i=1,\cdots,q,\varphi_{i+q}=0,i>0$$

于是

$$\begin{aligned}R(k)&=\sigma_\varepsilon^2\sum_{j=0}^{\infty}\varphi_j\varphi_{j+k}=\sigma_\varepsilon^2\sum_{j=0}^{(q-k)\vee 0}\varphi_j\varphi_{j+k}\\&=\begin{cases}\sigma_\varepsilon^2(1+\beta_1^2+\cdots+\beta_q^2) & k=0\\ \sigma_\varepsilon^2(-\beta_k+\beta_1\beta_{k+1}+\cdots+\beta_{q-k}\beta_k) & 0<|k|\le q\\ 0 & |k|>q\end{cases}\end{aligned}$$

因此 $MA(q)$ 是自相关函数截尾的，这是判断模型是否为滑动平均模型的一个判据。

(ii) $AR(p)$

自回归过程满足的方程是

$$X_t=\alpha_1X_{t-1}+\cdots+\alpha_pX_{t-p}+\varepsilon_t,$$

由此经推导可以得到

$$\begin{aligned}R(0)&=EX_t^2=EX_t(\alpha_1X_{t-1}+\cdots+\alpha_pX_{t-p}+\varepsilon_t)\\&=\alpha_1R(1)+\cdots+\alpha_pR(p)+\sigma_\varepsilon^2,\end{aligned}\tag{3.17}$$

$$\begin{aligned}R(k)&=EX_tX_{t-k}=EX_{t-k}(\alpha_1X_{t-1}+\cdots+\alpha_pX_{t-p}+\varepsilon_t)\\&=\sum_{i=1}^{p}\alpha_iR(k-i),\end{aligned}\tag{3.18}$$

$$\rho(k)=\sum_{i=1}^{p}\alpha_i\rho(k-i),k=\pm1,\pm2,\cdots\tag{3.19}$$

由于 $\rho(0)=1,\rho(-k)=\rho(k)$, 由此由 (3.18) 式得到著名的 $Yule-Walker$ 方程：

$$\begin{cases}\rho(1) & =\alpha_1+\alpha_2\rho(1)+\cdots+\alpha_p\rho_{p-1}\\ \rho(2) & =\alpha_1\rho(1)+\alpha_2+\cdots+\alpha_p\rho_{p-2}\\ \cdots & \quad\cdots\\ \rho(p) & =\alpha_1\rho(p-1)+\alpha_2\rho(p-2)+\cdots+\alpha_p\end{cases}\tag{3.20}$$

可以用向量和矩阵更简洁地表达：记 $b_p=(\rho(1),\rho(2),\cdots,\rho(p))'$, $\alpha(p)=(\alpha_1,\alpha_2,\cdots,\alpha_p)'$,

$$\Gamma_p=\begin{pmatrix}1 & \rho(1) & \rho(2) & \cdots & \rho(p-1)\\ \rho(1) & 1 & \rho(1) & \cdots & \rho(p-2)\\ \rho(2) & \rho(1) & 1 & \cdots & \rho(p-3)\\ \vdots & & & \ddots & \vdots\\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1\end{pmatrix}\tag{3.21}$$

注意到 $\rho(-k)=\rho(k)$, $Yule-Walker$ 方程可写为

$$b_p=\Gamma_p\alpha(p), \tag{3.22}$$

其中 Γ_p 称为 $Yule-Walker$ 矩阵, 反解 (3.20) 得

$$\alpha(p)=\Gamma_p^{-1}b_p, \tag{3.23}$$

由 (3.18) 知自回归序列的自相关函数是拖尾的。

(iii) $ARMA(p,q)$

从 $ARMA(p,q)$ 的表达式 (3.7), 我们可以计算得出

$$\sigma_X^2=\sigma_\varepsilon^2\sum_{k=0}^{\infty}\varphi_k^2,$$
$$R(k)=\sigma_\varepsilon^2\sum_{j=0}^{\infty}\varphi_j\varphi_{j+k},$$
$$\rho(k)=\frac{R(k)}{\sigma_X^2}$$

因此 $ARMA(p,q)$ 的自相关函数也是拖尾的。

(2) 偏自相关函数 (PACF)

由于 $AR(p)$ 和 $ARMA(p,q)$ 的自相关函数都是拖尾的，如何识别它们？为此我们引入给定 $X_{t-1},X_{t-2},\cdots,X_{t-k+1}$ 下 X_t 和 X_{t-k} 的偏自相关函数 (partial autocorrelation function).

定义 $\hat{E}(X_t|x_{t-1},X_{t-2},\cdots,X_{t-k+1})$ 为 $X_{t-1},X_{t-2},\cdots,X_{t-k+1}$ 的线性组合，这可以看成 X_t 在由 $(X_{t-1},X_{t-2},\cdots,X_{t-k+1})$ 所张成的空间上的线性投影。即设

$$\hat{E}(X_t|X_{t-1},X_{t-2},\cdots,X_{t-k+1})=\theta_1X_{t-1}+\cdots+\theta_{k-1}X_{t-k+1}$$

而且满足 $E[X_t-(\theta_1X_{t-1}+\cdots+\theta_{k-1}X_{t-k+1})]^2=\min$, 用最小二乘法和对向量求导运算可以算得

$$\hat{E}(X_t|x_{t-1},X_{t-2},\cdots,X_{t-k+1})=\mu_t+\Sigma_{12}\Sigma_{22}^{-1}\begin{pmatrix}X_{t-1}-\mu_{t-1}\\ \vdots\\ X_{t-k+1}-\mu_{t-k+1}\end{pmatrix}$$

其中 $\mu_j=EX_j$, $j=t-k+1,t-k+2,\cdots,t$, Σ_{12} 是 X_t 与 $(X_{t-1},X_{t-2},\cdots,X_{t-k+1})'$ 的 $1\times(k-1)$ 协方差行向量, Σ_{22} 是 $(X_{t-1},X_{t-2},\cdots,X_{t-k+1})'$ 的 $(k-1)\times(k-1)$

协方差矩阵, $\Sigma_{21}=(\Sigma_{12})'$. 同理可以计算 $\hat{E}(X_{t-k}|X_{t-1},X_{t-2},\cdots,X_{t-k+1})$, 因此给定 $X_{t-1},X_{t-2},\cdots,X_{t-k+1}$ 下 X_t 与 X_{t-k} 的偏自相关系数就是

$$(X_t-\hat{E}(X_t|X_{t-1},X_{t-2},\cdots,X_{t-k+1})$$

和

$$X_{t-k}-\hat{E}(X_{t-k}|X_{t-1},X_{t-2},\cdots,X_{t-k+1}))$$

之间的相关系数.

当 $X_t,X_{t-k},X_{t-1},X_{t-2},\cdots,X_{t-k+1}$ 为联合正态时, 给定 $X_{t-1},X_{t-2},\cdots,X_{t-k+1}$ 下 X_t,X_{t-k} 的条件期望就是 X_t,X_{t-k} 在 $X_{t-1},X_{t-2},\cdots,X_{t-k+1}$ 张成的子空间上的线性投影, 设

$$Var[(X_t,X_{t-k},X_{t-1},X_{t-2},\cdots,X_{t-k+1})']=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$$

其中 $\Sigma_{11}=Var(X_t,X_{t-k}),\Sigma_{22}=Var(X_{t-1},X_{t-2},\cdots,X_{t-k+1})$, Σ_{12} 为它们之间的 $2\times(k-1)$ 协方差矩阵, 则偏自相关系数矩阵就是与条件协方差矩阵 $\Sigma_{11\cdot 2}$ 相对应的相关系数矩阵。因此我们在这儿就是要计算自回归时间序列下的 $\Sigma_{11\cdot 2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$。容易算得,

$$\Sigma_{11}=R(0)\begin{pmatrix}1 & \rho(k)\\ \rho(k) & 1\end{pmatrix},\quad \Sigma_{12}=R(0)\begin{pmatrix}\rho(1) & \cdots & \rho(k-1)\\ \rho(k-1) & \cdots & \rho(1)\end{pmatrix}$$

$$\Sigma_{22}=R(0)\begin{pmatrix}1 & \rho(1) & \cdots & \rho(k-2)\\ \rho(1) & 1 & \cdots & \rho(k-1)\\ \vdots & \ddots & \ddots & \vdots\\ \rho(k-2) & \rho(k-3) & \cdots & 1\end{pmatrix}$$

在 $Yule-Walker$ 方程 (3.19) 中, 取 $k>p$, 即假设自回归序列的阶比真正的阶要高。记

$$\alpha(k)=\begin{pmatrix}\alpha_{1k}\\ \cdots\\ \alpha_{pk}\\ \alpha_{(p+1)k}\\ \cdots\\ \alpha_{kk}\end{pmatrix}\quad b_k=\begin{pmatrix}\rho_1\\ \rho_2\\ \cdots\\ \rho_k\end{pmatrix}\quad \beta_k=\begin{pmatrix}\rho_k\\ \rho_{k-1}\\ \cdots\\ \rho_1\end{pmatrix}$$

当自回归序列的阶为 $p<k$ 时, $\alpha_{(p+1)k}=\cdots=\alpha_{kk}=0$, 则由 (3.22),

$$b_k=\Gamma_k\alpha(k),\quad \alpha(k)=\Gamma_k^{-1}b_k$$

注意到 Γ_k 对称，各次对角线元素相同，记

$$\alpha_p(k) = \begin{pmatrix} 0 & \cdots & 1 \\ & \vdots & \\ 1 & & 0 \end{pmatrix}$$

则 H 为投影阵 (满足 $H^2 = H,\ H^{-1} = H,\ H' = H$)，且

$$H\alpha(k) = (\alpha_{kk}, \cdots, \alpha_{(p+1)k}, \alpha_p, \cdots, \alpha_1),$$

$$\beta_k = Hb_k = H\Sigma_k\alpha(k) = (H\Gamma_k H)H\alpha(k) = \Gamma_k H\alpha(k)$$

所以 $\beta_k = \Gamma_k H\alpha(k)$, 下面计算 $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$,

$$\begin{aligned}
\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} &= R(0)\begin{pmatrix} b'_{k-1} \\ (Hb_{k-1})' \end{pmatrix}\Gamma_{k-1}^{-1}\begin{pmatrix} b_{k-1} & Hb_{k-1} \end{pmatrix} \\
&= R(0)\begin{pmatrix} b'_{k-1} \\ (Hb_{k-1})' \end{pmatrix}\begin{pmatrix} \alpha_{k-1} & H\alpha_{k-1} \end{pmatrix} \\
&= \begin{pmatrix} \alpha_1\rho_1 + \cdots + \alpha_p\rho_p & \alpha_1\rho_{k-1} + \cdots + \alpha_p\rho_{k-p} \\ \alpha_1\rho_{k-1} + \cdots + \alpha_p\rho_{k-p} & \alpha_1\rho_1 + \cdots + \alpha_p\rho_p \end{pmatrix}
\end{aligned}$$

由 $Yule-Walker$ 方程 (3.19)，$\alpha_1\rho_{k-1} + \cdots + \alpha_p\rho_{k-p} = \rho_k$, 因此当 $k > p$ 时，

$$\begin{aligned}
\Sigma_{11\cdot 2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\
&= \begin{pmatrix} 1 - \alpha_1\rho_1 - \cdots - \alpha_p\rho_p & 0 \\ 0 & 1 - \alpha_1\rho_1 - \cdots - \alpha_p\rho_p \end{pmatrix}
\end{aligned} \tag{3.24}$$

因此当 $k > p$ 时，给定 $X_{t-1}, X_{t-2}, \cdots, X_{t-k+1}$ 下 X_t, X_{t-k} 的偏自相关系数为零，当 $k \le p$ 时，$\alpha_1\rho_{k-1} + \cdots + \alpha_p\rho_{k-p} \ne \rho_k$, 这说明 $AR(p)$ 的偏自相关系数有截尾性，而 $MA(q)$ 和 $ARMA(p,q)$ 的偏自相关系数没有截尾性，这是判断序列是否为自相关的重要判据。$ARMA(p,q)$ 模型的自相关函数和偏自相关函数都没有截尾性。这样我们就能把这三个常用的平稳线性序列区分开。

注 1：对 $ARMA(p,q)$ 序列, 及它的后移算子表达式为 $\varphi(B)X_t = \psi(B)\varepsilon_t$, 设 $\varphi(B) = \prod_{i=1}^{p}(1-\lambda_i B)$, 当 $i \ne j$ 时, $\lambda_i \ne \lambda_j$, 则由有理分式分解和 $(1-\lambda_i B)^{-1}$ 的幂级数展开，有

$$\begin{aligned}
\frac{\psi(B)}{\varphi(B)}\varepsilon_t &= \left(\psi(B)\sum_{i=1}^{p}\prod_{j\ne i}\frac{\lambda_i}{\lambda_i - \lambda_j}\frac{1}{1-\lambda_i B}\right)\varepsilon_t \\
&= \sum_{k=0}^{\infty}\left(\sum_{i=1}^{p}\prod_{j\ne i}\frac{\lambda_i}{\lambda_i - \lambda_j}\lambda_i^k\psi(B)\varepsilon_{t-k}\right) \\
&\equiv \sum_{k=0}^{\infty}\varphi_k\varepsilon_{t-k},
\end{aligned}$$

因此在所有的 $|\lambda_i| < 1$ 时，φ_k 以指数速度衰减，即以前数据遗忘的速度非常快，所有自相关系数也是以指数速度趋于零。如果 $\varphi(x) = 0$ 有重根，也有同样的结论。

注 2： 单整自回归滑动平均混合模型的后移算子表达式为 $\varphi(B)X_t = \varphi_1(B)(1-B)^d X_t = \psi(B)\varepsilon_t, \varphi(B) = 0$ 有单位根 1，所以它不是平稳序列，这表明 X_t 经过 d 次差分后的序列 $Y_t = (1-B)^d X_t$ 是平稳自回归滑动平均混合模型。

§3.3.4 频率域上平稳 $ARMA(p,q)$ 的研究

一个函数，或者一个序列，可以在时间域中进行研究，为了研究周期等性质，也可以在频率域中研究，所谓频率域，就是函数或序列的 Fourier 变换，我们已经知道，在一般性条件下，函数与它的 Fourier 变换是一一对应的。我们这儿的时间序列频率域研究是关于自相关函数 Fourier 变换有关性质的研究。

(1) 谱密度函数的定义

设 $R(k)$ 是 $ARMA(p,q)$ 的自相关函数，$k = 0, \pm 1, \pm 2, \cdots$，$\sum_{-\infty}^{\infty} |R(k)| < \infty$，定义

$$S(\omega) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} R(k) \exp\{-i\omega k\}$$

为平稳 $ARMA(p,q)$ 过程的功率谱密度，其中 $i = \sqrt{-1}$。则

$$R(k) = \int_{-\pi}^{\pi} S(\omega) \exp\{i\omega k\} d\omega = \int_{-\pi}^{\pi} S(\omega) \cos k\omega d\omega$$

即 $S(\omega)$ 和 $R(k)$ 是一对 Fourier 变换，由定义知，

$$\sigma_X^2 = R(0) = \int_{-\pi}^{\pi} S(\omega) d\omega$$

称为平均功率，

$$F(\omega) = \int_{-\pi}^{\omega} S(t) dt$$

称为谱分布，$S(\omega)$ 称为谱密度函数。

由 $ARMA(p,q)$ 的表达式 (3.13) 和后移算子表达式 $X_t = \frac{\psi(B)}{\varphi(B)}\varepsilon_t$，再注意

到过程的方差和自协方差函数有限，我们得到

$$
\begin{aligned}
S(\omega) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} R(k)\exp\{-ik\omega\} = \frac{\sigma_\varepsilon^2}{2\pi}\sum_{k=-\infty}^{\infty}\left(\sum_{j=0}^{\infty}\varphi_j\varphi_{j+k}\right)\exp\{-ik\omega\} \\
&= \frac{\sigma_\varepsilon^2}{2\pi}\sum_{j=0}^{\infty}\varphi_j\exp\{-ij\omega\}\sum_{k=-\infty}^{\infty}\varphi_{j+k}\exp\{-i(j+k)\omega\} \\
&= \frac{\sigma_\varepsilon^2}{2\pi}\sum_{j=0}^{\infty}\varphi_j\exp\{-ij\omega\}\sum_{k=0}^{\infty}\varphi_k\exp\{-ik\omega\} \\
&= \frac{\sigma_\varepsilon^2}{2\pi}|\sum_{k=0}^{\infty}\varphi_k\exp\{-ik\omega\}|^2 = \frac{\sigma_\varepsilon^2}{2\pi}|\sum_{k=0}^{\infty}\varphi_k(\exp\{-i\omega\})^k|^2
\end{aligned}
$$

注意到

$$
\begin{aligned}
X_t &= \frac{\psi(B)}{\varphi(B)}\varepsilon_t = \sum_{k=0}^{\infty}\varphi_k\varepsilon_{t-k} \\
&= (\sum_{k=0}^{\infty}\varphi_k B^k)\varepsilon_t \equiv h(B)\varepsilon_t,
\end{aligned}
$$

所以

$$
\begin{aligned}
S(\omega) &= \frac{\sigma_\varepsilon^2}{2\pi}|h(e^{-i\omega})|^2 \\
&= \frac{\sigma_\varepsilon^2}{2\pi}\left|\frac{\psi(e^{-i\omega})}{\varphi(e^{-i\omega})}\right|^2, \qquad (3.25)
\end{aligned}
$$

例3.4. *求 $AR(p)$ 的谱密度.*

解： 在 (3.25) 式中，$h(B) = \frac{1}{\varphi(B)}$，所以

$$
S(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\frac{1}{|\varphi(e^{i\omega})|^2}
$$

特别，当 $p=1$ 和 $p=2$ 时，

$$
\begin{aligned}
S_1(\omega) &= \frac{\sigma_\varepsilon^2}{2\pi}\frac{1}{1+\alpha_1^2-2\alpha_1\cos\omega}, \\
S_2(\omega) &= \frac{\sigma_\varepsilon^2}{2\pi}\frac{1}{1+\alpha_1^2+\alpha_2^2-2\alpha_1\cos\omega-2\alpha_2\cos 2\omega+2\alpha_1\alpha_2\cos\omega}
\end{aligned}
$$

(2) 周期图

考虑谱密度的估计问题。设 $\{x_t, t=1,2,\cdots,n\}$ 是时间序列的一个路径，令

$$
z(\omega) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(x_t-\bar{x})e^{-it\omega},
$$

称

$$
I(\omega) = \frac{1}{2\pi}|z(\omega)|^2 = \frac{1}{2\pi n}|\sum_{t=1}^{n}(x_t-\bar{x})e^{-it\omega}|^2 \qquad (3.26)
$$

为时间序列 $\{x_1, \cdots, x_n\}$ 的周期图, 这是谱密度函数的一个估计。定义

$$\hat{R}(k) = \hat{R}(-k) = \frac{1}{n}\sum_{t=1}^{n-|k|}(x_t - \bar{x})(x_{t+|k|} - \bar{x}), k = 0, 1, \cdots, n-1 \qquad (3.27)$$

为自相关函数 $R(k)$ 的估计。如果我们知道时间序列均值为零, 则去掉上面式中的 $\bar{x}$. 我们有

性质 1: $I(\omega)$ 和 $\hat{R}(k)$ 是一对 Fourier 变换, 即

$$I(\omega) = \frac{1}{2\pi}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\omega} = \bar{I}(\omega), \qquad (3.28)$$

$$\hat{R}(k) = \int_{-\pi}^{\pi} I(\omega)e^{ik\omega}d\omega,\ k = 0, \pm 1, \cdots. \qquad (3.29)$$

这儿 $\hat{I}$ 表示 I 的共轭。

证: 不失一般性, 设 $\bar{x} = 0$, 由 $I(\omega)$ 的定义,

$$\begin{aligned} I(\omega) &= \frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n}x_t x_s e^{-i(t-s)\omega} \\ &= \frac{1}{2\pi n}\sum_{k=-n+1}^{n-1}\sum_{t=1}^{n-|k|}x_t x_{t+|k|}e^{-ik\omega}, \\ &= \sum_{k=-n+1}^{n-1}\hat{\rho}(k)e^{-ik\omega}, \end{aligned}$$

反之,

$$\int_{-\pi}^{\pi} I(\omega)e^{ik\omega}d\omega = \frac{1}{2\pi n}\sum_{j=-n+1}^{n-1}\sum_{t=1}^{n-|j|}x_t x_{t+|j|}\int_{-\pi}^{\pi}e^{-i(j-k)\omega}d\omega$$

$$= \frac{1}{n}\sum_{t=1}^{n-|k|}x_t x_{t+|k|} = \hat{\rho}(k).$$

性质 2: 若 $EX_n = 0, \sum |R(k)| < \infty$, 则

$$\lim_{n\to\infty} EI(\omega) = S(\omega),$$

其中 $S(\omega)$ 是平稳序列 X_n 的谱密度函数。

这个性质说明 $I(\omega)$ 是 $S(\omega)$ 的一致估计 (consistent estimator).

证: 由 (3.27), 注意到 $EX_t = 0$, 我们有

$$\begin{aligned} EI(\omega) &= E\frac{1}{2\pi}\sum_{k=-n+1}^{n-1}\hat{R}(k)e^{-ik\omega} \\ &= \frac{1}{2\pi}\sum_{k=-n+1}^{n-1}(E\hat{R}(k))e^{-ik\omega}, \end{aligned}$$

$$E\hat{R}(k) = E\frac{1}{n}\sum_{t=1}^{n-|k|}x_t x_{t+|k|} = \frac{n+|k|}{n}R(k),$$

所以

$$\begin{aligned}EI(\omega)&=E\frac{1}{2\pi}\sum_{k=-n+1}^{n-1}(1-\frac{|k|}{n})R(k)e^{-ik\omega}\\&=\frac{1}{2\pi}\sum_{k=-n+1}^{n-1}R(k)e^{-ik\omega}-\frac{1}{2\pi}\sum_{k=-n+1}^{n-1}\frac{|k|}{n}R(k)e^{-ik\omega},\end{aligned}$$

由于 $\sum|R(k)|<\infty$, 所以

$$\begin{aligned}&S(\omega)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}R(k)e^{-ik\omega}=\frac{1}{2\pi}\sum_{k=-n+1}^{n-1}R(k)e^{-ik\omega}+o(1),\\&\left|\sum_{k=-n+1}^{n-1}\frac{|k|}{n}R(k)e^{-ik\omega}\right|\\&=|\sum_{k=-N+1}^{N-1}\frac{|k|}{n}R(k)e^{-ik\omega}|+|\sum_{N\le|k|<n-1}\frac{|k|}{n}R(k)e^{-ik\omega}|,\end{aligned}$$

取 N 满足 $\sum_{|k|\ge N}|R(k)|<\varepsilon$, 再取 n, 使得 $\frac{N}{n}\sum|R(k)|<\varepsilon$, 则

$$\begin{aligned}&|\sum_{k=-N+1}^{N-1}\frac{|k|}{n}R(k)e^{-ik\omega}|\le\frac{N}{n}\sum|R(k)|<\varepsilon,\\&|\sum_{N\le|k|<n-1}\frac{|k|}{n}R(k)e^{-ik\omega}|\le\sum_{|k|\ge N}|R(k)|<\varepsilon,\end{aligned}$$

由于 ε 可以任意小, 所以 $n\to\infty$ 时,

$$EI(\omega)\to\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}R(k)e^{-ik\omega}=S(\omega).$$

(3) 白噪声序列的谱密度

白噪声序列是时间序列的基础, 因此有必要对它的性质作更多的讨论。设 $\varepsilon_t\sim WN(0,\sigma^2)$, 则

$$R(k)=\begin{cases}\sigma^2 & k=0\\0 & k\neq0\end{cases}$$

$$\sum|R(k)|<\infty$$

所以白噪声序列 $\{\varepsilon_t\}$ 的谱密度为

$$S(\omega)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}R(k)e^{-it\omega}=\frac{\sigma^2}{2\pi}.$$

若以 $I_\varepsilon(\omega)$ 记正态白噪声序列 $\{\varepsilon_t\}$ 的周期图, 则由上可知

$$EI_\varepsilon(\omega)=S(\omega)=\frac{\sigma^2}{2\pi}.\tag{3.30}$$

下面计算周期图的方差 $Var(I_\varepsilon(\omega))$。我们有

$$\lim_{n\to\infty} Var(I_\varepsilon(\omega)) = S(\omega) = \frac{\sigma^2}{2\pi} = \begin{cases} \frac{2\sigma^4}{(2\pi)^2}, & \omega = 0 \text{ 或 } \omega = \pm\pi \\ \frac{\sigma^4}{(2\pi)^2}, & -\pi < \omega < \pi \text{ 且 } \omega \neq 0 \end{cases} . \quad (3.31)$$

证： 设 $-\pi \le \omega \le \pi$,

$$\begin{aligned} &Var(I_\varepsilon(\omega)) = EI_\varepsilon^2(\omega) - (EI_\varepsilon(\omega))^2 \\ &= E\left\{\frac{1}{2\pi n}|\sum_{t=1}^n \varepsilon_t e^{-it\omega}|^2\right\}^2 - \left(\frac{\sigma^2}{2\pi}\right)^2 \\ &E\left|\sum_{t=1}^n \varepsilon_t e^{-it\omega}\right|^4 = \sum_{t_1,t_2,t_3,t_4=1}^n E\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4} e^{i(t_1-t_2)\omega + i(t_3-t_4)\omega}, \quad (3.32) \end{aligned}$$

而

$$E\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4} = \begin{cases} 3\sigma^4, & t_1 = t_2 = t_3 = t_4 \\ \sigma^4, & t_1 = t_2 \neq t_3 = t_4 \\ & t_1 = t_3 \neq t_2 = t_4 \\ & t_1 = t_4 \neq t_2 = t_3 \\ 0, & \text{其它} \end{cases}$$

把 (3.32) 中求和号按上面 4 种情况分开计算，有

$$E\left|\sum_{t=1}^n \varepsilon_t e^{-it\omega}\right|^4 = (2n^2)\sigma^4 + \sigma^4\left|\sum_{t=1}^n e^{2it\omega}\right|^4$$

由于

$$\sum_{t=1}^n e^{2it\omega} = \begin{cases} n, & \omega = 0, \pm\pi \\ e^{2it\omega}\frac{1-e^{2in\omega}}{1-e^{2i\omega}} & -\pi < \omega < \pi, \omega \neq 0 \end{cases}$$

故

$$\lim_{n\to\infty} \frac{1}{n^2}\left|\sum_{t=1}^n e^{2it\omega}\right|^2 = \begin{cases} 1, & \omega = 0, \pm\pi \\ 0 & -\pi < \omega < \pi, \omega \neq 0 \end{cases}$$

§3.4 平稳线性序列的参数估计

本节讨论如何估计平稳线性序列，如 $MA(q), AR(p), ARMA(p,q)$，的有关参数，其中平稳线性序列的均值 μ，自协方差函数 $R(k)$ 及自相关函数 $\rho(k)$ 是最关心的参数。在实际中应用最多的是自回归模型，其中自回归系数 $\alpha_1, \cdots, \alpha_p$，过程方差 σ_X^2 和阶 p 是我们关注的参数。

§3.4.1 平稳过程均值 μ 的估计

根据遍历性定理，平稳过程的均值 μ 可以用一条样本路径来估计，即

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} x_t = \bar{X}$$

定理3.1. 均值估计 $\hat{\mu}$ 有如下的优良性质：

(1) $E\hat{\mu} = \mu$, (无偏性)

(2) 若当 $n \to \infty, R(k) \to 0$, 则

$$\lim_{n\to\infty} E(\hat{\mu} - \mu)^2 = 0, \text{ (均方收敛)} \tag{3.33}$$

(3) 设平稳线性过程为

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \varphi_j \varepsilon_{t-j},$$
$$\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty, \sum_{j=-\infty}^{\infty} \varphi_j \neq 0,$$

其中 ε_t 为正态白噪声，方差为 σ^2, 则

$$\sqrt{n}(\hat{\mu} - \mu) \to (d)N(0, v^2), \text{ (渐近正态性)} \tag{3.34}$$

其中

$$v^2 = \sigma^2 \sum_{j=-\infty}^{\infty} \varphi_j^2$$

证： 性质 (1) 是简单的，关于性质 (2), 展开 $Var(\hat{\mu})$, 有

$$\begin{aligned}
Var(\hat{\mu}) &= E(\bar{X} - \mu)^2 \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} E(X_t - \mu)(X_s - \mu) = \frac{1}{n^2}\sum_{t,s=1}^{n} R(t-s) \\
&= \frac{1}{n^2}\sum_{\ell=-n+1}^{n-1} (n - |\ell|)R(\ell) = \frac{\sigma_X^2}{n}\sum_{\ell=-n+1}^{n-1} (1 - \frac{|\ell|}{n})\rho(\ell),
\end{aligned}$$

在定理的条件下，上式 $\leq 2\frac{\sigma_X^2}{n}\sum_{\ell=0}^{n-1}|\rho(\ell)|$, 取极限，根据 $Stoltz$ 引理，有

$$\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=0}^{n-1} |\rho(\ell)| = \lim_{n\to\infty} \rho(n) = 0,$$

这就是平稳时间序列的遍历性定理，或称 $\hat{\mu}$ 为 μ 的一致估计。

性质 (3) 的证明不难但是有点繁，我们不在这儿展开了, 而且 $\{\varepsilon_t\}$ 不必要求是正态白噪声。

由上面的三条性质，我们知道 $\hat{\mu}$ 是 μ 的一个性质非常好的估计。

§3.4.2 平稳过程自协方差和自相关函数的估计

这儿我们假定 $\{X_t\}$ 是零均值的平稳序列, 要估计它的协方差和自相关函数。通常有如下两种估计 c_k 和 $\hat{\gamma}_k$, 其中

$$c_k = c_{-k} = \frac{1}{n-k}\sum_{t=1}^{n-k} x_t x_{t+k},\ k = 0, 1, \cdots$$

$$\hat{\gamma}_k = \hat{\gamma}_{-k} = \frac{1}{n}\sum_{t=1}^{n-k} x_t x_{t+k},\ k = 0, 1, \cdots, \tag{3.35}$$

其中 $\{\hat{\gamma}_k\}$ 称为样本自协方差函数。记

$$\hat{\Gamma}_n = (\hat{\gamma}_{i-j})_{1\le i,j\le n}, \tag{3.36}$$

$$C_n = (c_{i-j})_{1\le i,j\le n}$$

它们都是自协方差矩阵 $\Gamma_n = (\gamma_{i-j})_{1\le i,j\le n}$ 的估计。注意到 $\Gamma_n \ge 0$, 而 $\hat{\Gamma}_n = \frac{1}{n}Q_n'Q_n \ge 0$, 其中

$$Q_n' = \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots & 0 \\ 0 & x_1 & \cdots & x_{n-1} & & 0 \\ 0 & \cdots & \ddots & & \ddots & \vdots \\ 0 & 0 & \cdots & x_1 & \cdots & x_n \end{pmatrix}_{n\times 2n}$$

而 C_n 中的元 c_k 有系数 $\frac{1}{n-k}$ 而不是 $\frac{1}{n}$, 因此没有 $\hat{\Gamma}_n$ 的那种形状, 也不必是非负定的。

实际中, 当 n 不太大时用 C_n, n 较大时用 $\hat{\Gamma}_n$。

有关样本自协方差函数的大样本性质, 我们有如下的结果。设

$$X_t = \sum_{k=-\infty}^{\infty} \varphi_k \varepsilon_{t-k},\quad \sum_{k=-\infty}^{\infty} |\varphi_k| < \infty,$$

$$E\varepsilon^4 < \infty,\quad \varepsilon_t \sim \quad WN(0, \sigma^2)$$

则

$$Cov(\hat{\gamma}_k, \hat{\gamma}_j) = \frac{\mu_4 - 3\sigma^4}{n}\gamma_k\gamma_j + o(\frac{1}{n}),$$

$$\sqrt{n}(\hat{\gamma}_0 - \gamma_0, \hat{\gamma}_1 - \gamma_1, \cdots, \hat{\gamma}_k - \gamma_k)' \to (d) N_{k+1}(0, G),$$

其中协方差矩阵

$$G = (g_{ij})_{(k+1)\times(k+1)}$$

$$g_{ij} = \sum_{s=-\infty}^{\infty} (\gamma_{s+i}\gamma_{s+j} + \gamma_{s-i}\gamma_{s-j}) + \frac{\mu_4 - 3\sigma^4}{n}\gamma_i\gamma_j,$$

$$0 \le i, j \le k.$$

当过程均值不为零时，

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}),$$

可以证明，对 $\sum|\varphi_k| < \infty$ 的线性过程 $\left\{\sum_{k=-\infty}^{\infty}\varphi_k\varepsilon_{t-k}\right\}$ 而言，

$$\bar{X} \to (L_2)\ \mu,\ \ \hat{\gamma}_k \to (L_2)\ \gamma_k.$$

§3.4.3 自回归模型的参数估计

若 $\{X_t\}$ 为零均值平稳 $AR(p)$ 模型，(若均值不为零，用 $X_t - \bar{X}$ 代替 X_t 来讨论)

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t,\ t > p,$$

其中 $\{\varepsilon_t \sim WN(0, \sigma^2)\}$, 我们的任务是估计参数 $p, \alpha_1, \cdots, \alpha_p$ 和 σ^2。先假定 p 已知，记

$$y = \begin{pmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{pmatrix},\quad X = \begin{pmatrix} x_p & x_{p-1} & \cdots & x_1 \\ x_{p+1} & x_p & \cdots & x_2 \\ \vdots & \vdots & & \vdots \\ x_n & x_{n-1} & \cdots & x_{n-p} \end{pmatrix},\quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}$$

(1) 最小二乘估计 (LSE)

最小二乘就是使残差 $||y - X\alpha||^2 = \min$, 由此得到

$$\hat{\alpha} = (X'X)^{-1}X'y, \tag{3.37}$$

以及残差平方和

$$SS_e = y'(I - X(X'X)^{-1}X')y,$$

由于 $Rank(X) = p$, 所以误差方差的估计是

$$\hat{\sigma}^2 = \frac{1}{n-p}SS_e = \frac{1}{n-p}y'(I - X(X'X)^{-1}X')y,$$

形式上与回归模型中回归系数和误差方差的估计一样。

(2) $Yule - Walker$ 估计

自回归模型的自协方差函数满足 $Yule - Walker$ 方程：

$$\begin{aligned}
\gamma_1 &= \alpha_1\gamma_0 + \alpha_2\gamma_1 + \cdots + \alpha_p\gamma_{p-1}, \\
\gamma_2 &= \alpha_1\gamma_1 + \alpha_2\gamma_0 + \cdots + \alpha_p\gamma_{p-2}, \\
&\vdots \\
\gamma_p &= \alpha_1\gamma_{p-1} + \alpha_2\gamma_{p-2} + \cdots + \alpha_p\gamma_0,
\end{aligned}$$

用 $\hat{\gamma}_k$ 代替上面方程中的 γ_k，解联立方程得到的解称为 $Yule-Walker$ 估计，

$$\begin{aligned}\hat{\Gamma}_p\hat{\alpha} &= \hat{b}_p,\\ \hat{\alpha} &= (\hat{\Gamma}_p)^{-1}\hat{b}_p,\end{aligned} \tag{3.38}$$

其中关于 Γ_p, b_p 的关系见 (3.22) 和 (3.23) 式，这儿 $\hat{\Gamma}$ 和 $\hat{b}$ 中的元素是用 $\hat{\gamma}_k$ 代替 γ_k。

(3) 极大似然估计 (MLE)

在误差序列为正态白噪声时，我们可以用极大似然方法来估计有关参数。一般而言，用极大似然方法得到的估计有许多优良性质，所以极大似然方法在估计参数方面有广泛的应用。下面我们以自回归模型来介绍。

设 $\varepsilon_t \sim N(0, \sigma^2)$，由于 $AR(p)$ 序列 $\{X_t\}$ 可以表为

$$X_t = \sum_{k=0}^{\infty} \varphi_k \varepsilon_{t-k},$$

其中 $\sum |\varphi_k| < \infty$，即 X_t 为无穷个正态随机变量的线性组合，而且方差有限，所以 X_t 仍是正态随机变量，因此

$$(X_1, X_2, \cdots, X_n)' \sim N_n(0, \Gamma_n),$$

其中协方差矩阵 $\Gamma_n = (\gamma_{i-j})_{n\times n}$，于是 $(X_1, X_2, \cdots, X_n)'$ 的对数似然函数为

$$L(\alpha_1, \cdots, \alpha_p, \sigma^2; x_1, \cdots, x_n) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Gamma_n| - \frac{1}{2}x'\Gamma_n^{-1}x, \tag{3.39}$$

其中 $x' = (x_1, x_2, \cdots, x_n)'$. 由 (3.17) 和 $Yule-Walker$ 方程 (3.20) 知，γ_k 是 $\alpha_1, \cdots, \alpha_p, \sigma^2$ 的函数，因此似然方程也是 $\alpha_1, \cdots, \alpha_p, \sigma^2$ 的函数。取 $\hat{\alpha}_1, \cdots, \hat{\alpha}_p, \hat{\sigma}^2$ 使 (3.39) 最大，我们就得到 $AR(p)$ 参数的 MLE，当然，方程一般没有显式解，但是有相关的求 $AR(p)$ 参数的 MLE 程序可用。

§3.4.4 自回归模型阶数 p 的估计

自回归模型的阶数 p 是一个重要的参数，只有在知道 p 以后，我们才能进一步对有关参数作估计。对自回归模型而言，一般有如下方法可用来定模型的阶 p。

(1) 基于 LSE 残差平方和的估计

由于在一般的自回归模型中，p 不是很大，不妨设存在常数 L，而 $p \le L$，所以我们对给定的 p，可以按上节叙述的三种方法之一来估计 $\alpha_1, \cdots, \alpha_p, \sigma^2$，然后计算残差平方和 $SS_e(p) = \sum_{t=p+1}^{n}(x_t - \hat{\alpha}_1 x_{t-1} - \cdots - \hat{\alpha}_p x_{t-p})^2$，这是 p

的函数，若 p_0 使 $SS_e(p)=\min$，我们就把 p_0 作为 p 的一个估计，这是最常用的。

(2) 基于偏自相关函数截尾性质的估计

由于 $AR(p)$ 的偏自相关函数有截尾性质，因此我们通过 (3.23) 式可以计算

$\alpha(k)=\Gamma_k^{-1}b_k$，其中 $\alpha(k)=(\alpha_{1k},\alpha_{2k},\cdots,\alpha_{kk})'$。在具体估计时，我们先用 (3.35) 式估计 R_k，再由 (3.24) 式估计

$$\hat{\rho}_{k+1,k+1}=(\hat{\rho}_{k+1}-\sum_{j=1}^{k}\hat{\rho}_{k+1-j}\hat{\alpha}_{jk})(1-\sum_{j=1}^{k}\hat{\rho}_j\hat{\alpha}_{jk})^{-1}, \tag{3.40}$$

$$\hat{\rho}_{1,1}=\hat{\rho}_1 \tag{3.41}$$

对 $AR(p)$ 而言，当 $k>p$ 时，$\alpha_{kk}=0$。但是 $\hat{\alpha}_{kk}$ 仅仅是 $\alpha_{kk}=0$ 的一个估计，大部分情况下不可能为 0，由于 $\hat{\alpha}_{kk}$ 是对 0 的估计，故只要绘图 $\hat{\alpha}_{kk}=0$，寻找 $\hat{\alpha}_{kk}$ 接近于 0 值的起始位子。

(3) AIC 准则

在方法 (1) 的估计中，我们仅知道 $p\le L$，不知道 p 到底有多大。但 L 是我们给出的，若 L 给得过大，容易多出一些虚假参数 $\hat{\alpha}_{p+1},\cdots,\hat{\alpha}_{\hat{p}}$，反之 $\hat{p}$ 太小可能会失真。为了克服这个缺陷，日本学者赤池 (Akaike) 引入如下准则 (Akaike 信息论准则)：

$$AIC(k)=\log\hat{\sigma}^2(k)+\frac{2k}{n},\ k=0,1,\cdots, \tag{3.42}$$

$$k=\arg\min(AIC(k)), \tag{3.43}$$

其中 $\hat{\sigma}^2(k)$ 是模型阶取 $p=k$ 时对 σ^2 的估计，AIC 准则本质上是利用残差来估计。但是 $\log\hat{\sigma}^2(k)$ 随 k 的增大而减小, 没有最小值，所以要得到 p 的估计，必须对阶 k 进行惩罚，所以在 AIC 准则中加上了惩罚项 $\frac{2k}{n}$, 这一项随 k 的增加而增加，这样，在 AIC 准则下，一定有个最小值，我们把取到这个最小值的 k 作为模型阶 p 的估计。模型中的 $\hat{\sigma}^2(k)$ 可以用 $Yule-Walker$ 方程来估计

$$\hat{\sigma}^2(0)=\hat{\gamma}_0,$$

$$\hat{\sigma}^2(k)=\hat{\gamma}_0-\hat{b}_k\hat{\Gamma}_k^{-1}\hat{b}_k,$$

AIC 准则是实际中对阶的估计时用得最多的一种。但是理论上可以证明，AIC 准则是不相合的，所以在实际应用中有一些改进，如 BIC 准则 (Bayes

信息论准则),

$$BIC(k) = \log \hat{\sigma}^2(k) + \frac{k \log n}{n},\ k = 0, 1, \cdots, \tag{3.44}$$

我国的赵林城教授进一步给出了相合的类似于 AIC 准则的一般信息论准则：

$$GIC(k) = \log \hat{\sigma}^2(k) + c\frac{k \log \log n}{n},\ k = 0, 1, \cdots, \tag{3.45}$$

其中 c 为任一常数。这一准则在信息论中得到广泛的应用。

对 $MA(q)$ 和 $ARMA(p, q)$ 的参数估计和阶的估计就不再这儿一一介绍了，有兴趣的读者可参见《时间序列分析》(安鸿志)。