实用预测学

霍俊编著

第三册 时间序列预测分析

中国发明创造者基金会
中国预测研究会

实　用　预　测　学

第 三 册

时间序列预测分析

霍　俊　编著

中国发明创造者基金会
中 国 预 测 研 究 会

1984.9

目 录

说　　明

大多数与模型有关的统计方法都是假设观察数据是统计上独立或不相关的。这些方法不能直接应用于所考虑的数据是统计相关的情况。但是，在经济、工程和自然科学方面以时间序列形式出现，即按时间顺序排列的大量观察数据是统计相关的。事实上，在这种情况下，观察数据之间的相互依存或相关性是最重要和有用的特性。利用观察数据之间的相关性可以建立相应的数学模型来描述客观对象的动态特征，从而可以利用过去的观察数据对未来值进行预测和控制。

用于分析这种相关的观察序列的主要方法称为时间序列分析。时间序列分析是概率统计学中的一个重要分支。它是用概率统计方法分析随时间变化的随机数据序列。频谱分析包括一类在频域中的时间序列分析方法，是时间序列分析的一个重要部分。本书只讨论在时域中，建立离散时间序列的统计模型及其在经济和企业预测方面的应用。

时间序列模型可以认为是一类特殊的单方程回归模型。因此，讨论单方程回归模型的大部分经济计量方法可以用于分析时间序列模型。本册不介绍这方面的内容。在阅读本册涉及到这方面的内容时，请读者查阅本书第二册。

全书共分五章。第一章介绍随机时间序列的基本概念和基本性质。第二章详细讨论了时间序列模型。包括确定型时间序列模型和随机型时间序列模型。重点讨论了随机时间序列模型的自回归模型、移动平均模型、自回归——移动平均混合模型以及非平稳时间序列模型。介绍了这些模型的基本性质和基本特征。第三章是讨论随机时间序列模型的建立。包括模型识别、模型参数估计和模型的诊断检验。第四章是计算最小均方差预测、预测误差和预测的置信区间。最后一章是介绍时间序列模型的建立和应用实例。建立了一些经济变量的时间序列模型，并将这些模型用于短期预测。这一章的最后，通过实例说明，如何建立时间序列与回归分析的综合模型。

本册是在我们一九八〇年编译的《时间序列预测技术》一书基础上经过重大修改和补充之后写成的。对于初学者最好还是由本书第一册、第二册读起较为方便。

蔡福元参加了第三册的编著工作。

第一章　时间序列的基本性质

第一节　时间序列的基本概念

一、时间序列

在经济、工程、自然科学和社会科学等领域的实际工作者和研究人员都要和一系列的观察数据打交道，我们把按时间顺序产生和排列的观察数据序列称为时间序列。如果数据序列是连续的，称为连续时间序列，可见图 1.1.1。

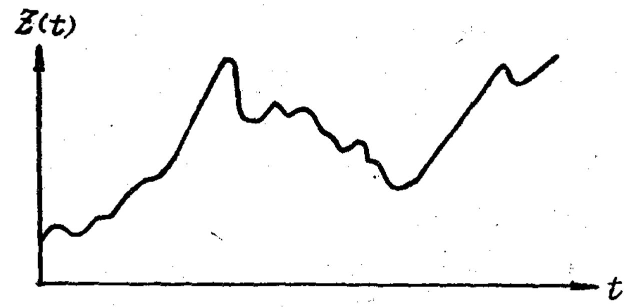

图 1.1.1　连续时间序列

如果数据序列是离散的，则称为离散时间序列，可见图1.1.2。

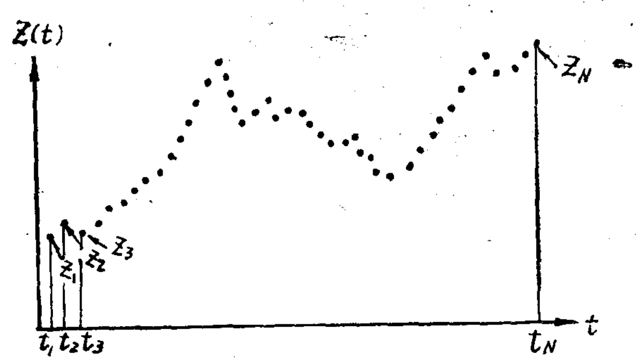

图 1.1.2.　离散时间序列

在时间t_1 t_2 t_3………t_N所获得的离散时间序列的观察值可以用$Z(t_1)$，$Z(t_2)$……$Z(t_t)$，……$Z(t_N)$表示。本书只考虑以均匀时间间隔h获取的离散时间序列。如果用于分析的N个连续观察值，则可以用Z_1 Z_2……Z_t……Z_N表示在相同时间间隔t_0+h，$t_0+{}_2h$，……t_0+th……t_0+Nh的观察值。在许多情况下，t_0和h的值是不重要的。如果需要完全确定观察时间，可以具体规定这两个参数的值。如果把t_0作为时间的起始点，把h作为时间单位，就可以把Z_t作为时间t的观

察值。

如果一个时间序列的未来数值可以由某一数学函数，如

$$Z_t = Ae^{rt} \tag{1.1.1}$$

来确定，称为确定型时间序列。如果时间序列的未来数值，只能用一概率分布来描述，则称为随机时间序列。

二、随机时间序列

随机时间序列是由一串随机变量Z_t（$t=\cdots\cdots1, 2, 3\cdots\cdots$）所构成的序列，对每个固定的时刻t，$Z_t$是一随机变量。应当指出，一个随机变量$Z_t$与它的一个样本值Z，是两种不同意义的量。前者是这一随机现象（或随机试验结果）的总称，后者是对这个随机现象取得的一个具体的观测值（或试验值）。如果一个描述随机现象的随机变量，可以通过重复试验，获得它的多个样本值，就能利用这些样本值来求出随机现象的各种统计量的值。但是对于随时间变化的随机过程，则只能获得所观察的随机现象的有限数据，即只能获得相应的随机过程的一个样本值序列的一段数据，不可能得到一个无限长的完整的样本序列。当然，要得到两个或多个样本序列更是难以想象的。这是通常的数理统计与时间序列分析的一个主要不同之点。下面将会看到，上述不足之处并不妨碍利用一段有限的样本值来寻求某些随机时间序列的各种统计量。

三、随机时间序列的概率分布

一个随机变量的统计特性完全由它的概率分布所确定，同样，一个随机时间序列的统计特性也完全由它的概率分布所确定。随机变量的概率分布可用概率分布函数来描述。随机时间序列是由无穷多个随机变量构成的。一个随机序列 Z_t（$t=\cdots1, 2, 3\cdots$）的概率分布，是指对于任意有穷多个时刻$t_1, t_2\cdots t_m$，相应的随机变量Z_{t1}，$Z_{t2}\cdots Z_{tm}$的联合分布函数$F_{t_1t_2\cdots t_m}$（Z_1，Z_2，$\cdots Z_m$）都是被给定的，而且它们之间不能矛盾，即由高维联合分布推出的低维联合分布与原给定的低维分布相同。如果对于任何有穷个不同时刻t_1，t_2，$\cdots t_m$，相应的Z_{t1} Z_{t2}，$\cdots Z_{tm}$是相互独立的随化变量，则

$$F_{t12\cdots ttm}(Z_1, Z_2\cdots Z_m) = F_{t1}(Z_1)\cdot F_{t2}(Z_2)\cdots F_{tm}(Z_m) \tag{1.1.2}$$

这时，称Z_t为独立的随机时间序列。如果一个随机时间序列Z_t的任意有穷维概率分布具有以下性质$F_{t_1t_2\cdots t_m}(Z_1, Z_t, \cdots Z_m) = F_{t_1+\tau, t_2+\tau\cdots t_m+\tau}(Z_1, Z_2, \cdots Z_m)$ (1.1.3)
其中τ是任意间隔时间，就称Z_t与狭义平稳时间序列（或严格的平稳时间序列）。也就是说，对于任意的m值和时刻t_1, t_2，$\cdots t_m$以及任意时间间隔τ，$Z_{t_1}Z_{t_2}\cdots t_{tm}$和$Z_{t_1+\tau}$，$Z_{t_2+\tau}\cdots Z_{tm+\tau}$具有相同的联合概率分布。

四、随机时间序列的参数表征

上面已经指出，随机时间序列的概率分布决定了该序列的全部统计特性。但是要完全确定某一时间序列的概率分布函数几乎是不可能的。在许多实际应用中，并不需要确定随机时间序列的概率分布函数，只要掌握某些参数表征就够了。下面介绍与本书内容关系密切的三个参数表征的定义。

1.均值函数

对每一个 t 而言，如果把随机变量Z_t的均值记做$E[Z_t]=\mu_t$，则随机时间序列 Z_t 的均值函数就是

$$\mu_t \quad (t=\cdots1, 2, 3, \cdots) \tag{1.1.4}$$

若$F_t(Z)$和$P_t(Z)$分别为Z_t的概率分布函数和概率密度函数，则

$$\mu_t = E[Z_t] = \int Z dF_t(Z) = \int Z P_t(Z) dZ \tag{1.1.5}$$

如果对于任意时间t，μ_t是一常数，则称μ为时间序列Z_t的均值。

2. 自协方差函数

由（1.1.5）式可知，随机时间序列的均值函数只和随机时间序列的一维概率分布有关。为了分析随机时间序列Z_t在不同时刻取值的统计关系，需要考虑Z_t与Z_s之间的相关值。令

$$\gamma_{ts} = E[(Z_t - E[Z_t])(Z_s - E[Z_s])]$$
$$= \int\int (Z_t - \mu_t)(Y_t - \mu_s) P_{ts}(Z, Y) dZdY \tag{1.1.6}$$

γ_{ts}称为随机时间序列Z_t的自协方差函数。若S＝t，则

$$\gamma_{tt} = E[(Z_t - E[Z_t])^2] = E[(Z_t - \mu_t)^2] \tag{1.1.7}$$

γ_{tt}称为随机时间序列的方差函数，或简称方差。

如果一个随机时间序列的任意有穷维概率分布都是正态分布，就称为正态随机时间序列。其概率分布密度为

$$P_{t_1t_2\cdots\cdots t_m}(Z_1, Z_2, \cdots Z_m)$$
$$= (2\pi|\Gamma_m|)^{-\frac{m}{2}} \cdot \exp\left\{-\frac{1}{2}(Z_m - \mu_m)^{\tau}\Gamma_m^{-1}(Z_m - \mu_m)\right\} \tag{1.1.8}$$
$$= N(\mu_m, \Gamma_m)$$

其中：$Z_m^{\tau} = (Z_1, Z_2 \cdots Z_m)$

$\mu_m^{\tau} = (\mu_{t1}, \mu_{t2} \cdots \mu_{tm})$；

$$\Gamma_m = \begin{pmatrix} \gamma_{t_1t_1} & \gamma_{t_1t_2} & \cdots & \gamma_{t_1tm} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{tmt_1} & \gamma_{tmt_2} & \cdots & \gamma_{tmtm} \end{pmatrix}$$

很多实际应用的随机时间序列可以近似地视为正态随机时间序列。对于正态随机序列的数学处理比较方便。

3. 自相关函数

时间序列的自相关函数ρ_{ts}定义为

$$\rho_{ts} = \frac{\gamma_{ts}}{\sqrt{\gamma_{tt} \cdot \gamma_{ss}}} \tag{1.1.9}$$

ρ_{ts}表示随机时间序列Z_t在不同时刻取值的线性相关程度。

由（1.1.5），（1.1.6）和（1.1.9）三式可以看出，μ_t、γ_{ts}和ρ_{ts}是由随机时间序列Z^t的概率分布所唯一决定的。但是，一般来说，由μ_t、γ_{ts}和ρ_{ts}则不能唯一地确定随机时间序列Z_t的概率分布。也就是说，具有不同概率分布的随机时间序列可以有相同的均值函数、自协方差函数和自相关函数。虽然如此，对于大量的实际应用而言，用以上三个函数量来表征随机时间序列的统计特性就够了。当然，对于某些实际问题的处理，也可能要用到随机时间序列的概率分布。

五、时间序列分析

对以均匀时间间隔获得的离散时间序列进行分析的统计方法叫做时间序列分析。和其他

的统计分析方法不同，时间序列分析的特点是观察值排列顺序的重要性和前后观察值之间是统计相关的。数据的统计相关性是用观察值之间的相关或自相关函数来表示。因此现有的时间序列分析方法几乎都是建立在经验的或估计的即样本自相关函数或它的富氏变换（自频谱）的基础上。而估计的样本自相关函数是理论自相关函数的一种比较差的估计。这就使得基于这种估计自相关函数的时间序列分析方法变得困难而复杂。这个困难可以通过利用线性系统分析的方法来说明时间序列数据间的相关性来解决。随机时间序列可以看成是一个随机系统对于不相关或独立的"白噪声"输入的反应的一个现实。这样该随机动态系统的数学模型就可以把不独立或相关的时间序列输入转化为独立的或不相关的输入。这样，时间序列分析就可以归结为寻求一种模型，它能实现把不独立的观察数据变成独立数据的转换。然后利用对于独立观察值的统计方法进行估计、预测和控制。

第二节　时间序列与随机过程

按概率统计规律随时间变化的随机现象，也就是取决于时间的随机变量的全体，称为随机过程。随机时间序列可以看成是由所研究系统的基本概率机理产生的一个具体现实。也就是说，在分析随机时间序列的时候，可以把它看成是随机过程的一个现实，如图（1.2.1）所示

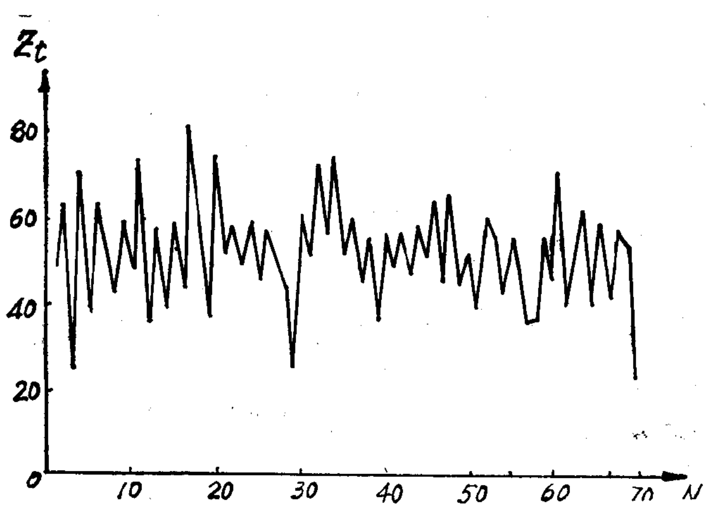

图1.2.1　观察随机时间序列

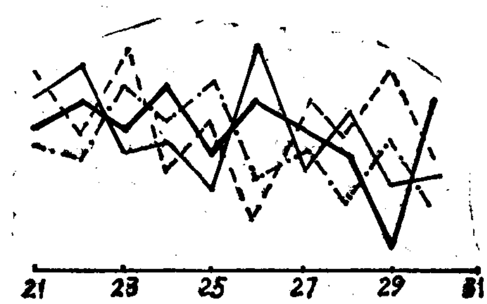

图 1.2.2　来自同一随机过程的4个时间序列的抽样现实

N个观察值的时间序列$Z_t=(Z_1, Z_2, \cdots Z_N)$是来自一无限母体的样本现实。而这个母体是由随机过程产生的。图1.2.2表示由基本随机过程所确定的时间序列母体获得的四个时间序列的抽样现实。因此，在时间t（如$t=25$）的观察值Z_t是具有概率密度函数$P(Z_t)$的随机变量Z_t的一个现实。同样，在任意两个时刻t_1和t_2（如$t_1=25$，$t_2=27$）的观察值Z_{t_1}和Z_{t_2}是具有联合概率密度函数$P(Z_{t_1}, Z_{t_2})$的两个随机变量Z_{t_1}和Z_{t_2}的一个现实。一般来说，具有联合概率密度函数$P(Z_{t_1}, Z_{t_2}, \cdots Z_N)$的N维随机变量$(Z_1, Z_2, \cdots Z_N)$可以用来描述均匀时间间隔采样的随机时间序列。统计分析的主要目的是，要从观察的抽样现实来推导母体的性质。例如，为了预测，在已知过去样本数据的情况下，从母体来推断未来观察数值的概率

分布。描述时间序列概率结构的模型叫做随机过程。

第三节　平稳随机时间序列

随机时间序列包括很广泛的随机现象。它的结构可以是很复杂的。这一节介绍最重要和最常用的一类随机时间序列——平稳随机时间序列。

一、平稳随机时间序列的基本概念

如果时间序列的统计特性不受时间起点的影响，则称为狭义（或严格的）平稳时间序列。也就是说，在时刻t_1、t_2、$\cdots t_m$观察的时间序列Z_{t_1}，Z_{t_2}，$\cdots Z_{t_m}$的联合概率分布函数，与在时刻t_1+K、t_2+K、$\cdots t_m+K$观察的时间序列Z_{t_1+K}、Z_{t_2+K}、$\cdots Z_{t_m+K}$的联合概率分布函数是相同的。即，如果

$$F_{t_1t_2\cdots t_m}(Z_1,\ Z_2,\ \cdots Z_m)$$
$$=F_{t_1+K,t_2+K,\cdots t_m+K}(Z_1,\ Z_2,\ \cdots Z_m) \qquad (1.3.1)$$

则时间序列Z_t称为严格的平稳时间序列。这个性质表明，对于严格的平稳时间序列，任意的时间平移不会改变时间序列的统计性质。但是，这个性质在实际应用中不容易验证，因为要对随机时间序列的所有有限维概率分布函数都要验证平稳条件（1.3.1）。这一点通常是很难做到的，而且在许多理论研究和实际应用中也不需要这样严格的性质。因此，将平稳性的要求放宽，不要求它的一切统计特性都具有平稳性，而只要求它的部分统计特性具有平稳性。于是提出了广义（或弱）平稳性概念。

如果一随机时间序列Z_t的均值函数μ_t和自协方差函数γ_{ts}满足下列条件：

$$\mu_t=\mu=\text{常数}$$
$$\gamma_{ts}=\gamma_{t-s} \qquad (1.3.2)$$

则称该随机时间序列为广义（或弱）平稳时间序列。

二、平稳随机时间序列的均值与方差

根据平稳性定义，对所有时刻t，随机时间序列的概率密度函数$P(Z_t)$是相同的，可以写成$P(Z)$，所以平稳随机时间序列的均值μ和方差σ_Z^2都是常数。

$$\mu=E[Z_t]=\int_{\infty}^{\infty}ZP(Z)dZ$$
$$\sigma_Z^2=E[(Z_t-\mu)^2]=\int_{-\infty}^{\infty}(Z-\mu)^2P(Z)dZ \qquad (1.3.3)$$

均值μ表示随机时间序列波动的水平线。方差σ^2表示随机时间序列沿均值μ上下的离散程度。

因为，对任意时刻t的概率密度函数$P(Z_t)$是相同的，因此可以由观察的时间序列Z_1，Z_2，$\cdots Z_N$构成的频率曲线来推断$P(Z)$的形状。随机时间序列的均值可以用观察时间序列的均值$\bar{Z}$来估计

$$\bar{Z}=\frac{1}{N}\sum_{t=1}^{N}Z_t \qquad (1.3.4)$$

随机时间序列的方差σ_Z^2可以用观察时间序列的方差来估计

$$\hat{\sigma}_Z^2 = \frac{1}{N}\sum_{t=1}^{N}(Z_t - \bar{Z})^2 \tag{1.3.5}$$

三、平稳随机时间序列的自协方差和自相关系数

根据平稳性定义，对于任意两个具有相同时间间隔的时刻t_1，t_2，随机时间序列的二维联合概率密度函数$P(Z_{t_1}, Z_{t_2})$都是相同的。$P(Z_{t_1}, Z_{t_2})$的性质可以由时间序列数据（Z_t，Z_{t+K}）（K是任意常数）所绘制的分布图来推断。图1.3.1是滞后K＝1和K＝2时，所绘制的（Z_t，Z_{t+2}）和（Z_t，Z_{t+1}）数据点的分布图。由图1.3.1可以看出，随机时间序列相邻数据之间是相关的。图1.3.1（a）表示Z_t与Z_{t+1}之间是负相关，图1.3.1（b）表示Z_t与Z_{t+2}之间是正相关。

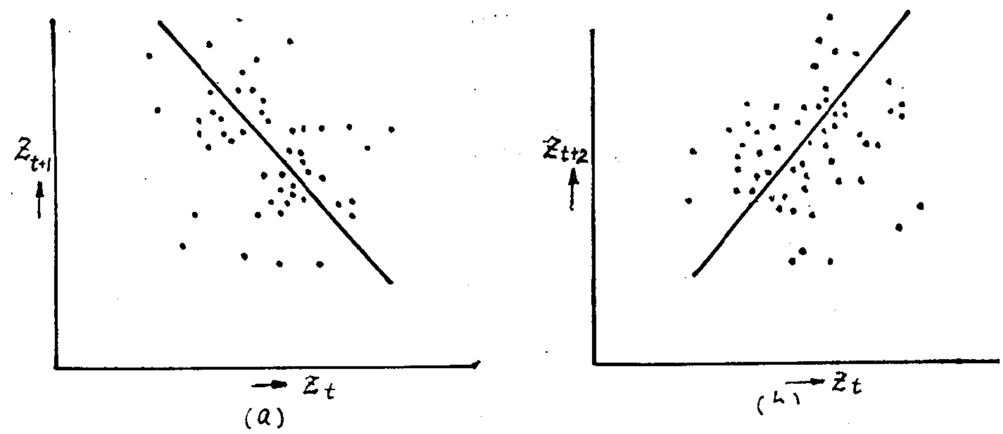

图1.3.1 k＝1(a)，k＝2(b)的时间序列数据(Z_t，Z_{t+k})的分布图

Z_t与相隔K个采样周期的Z_{t+K}之间的协方差称为滞后K的自协方差γ_K，定义为

$$\gamma_K = COV[Z_t, Z_{t+K}] = E[(Z_t - \mu)(Z_{t+K} - \mu)] \tag{1.3.6}$$

类似的，滞后K的自相关系数

$$\rho_K = \frac{E[(Z_t - \mu)(Z_{t+K} - \mu)]}{\sqrt{E[(Z_t - \mu)^2 \cdot E[(Z_{t+K} - \mu)^2]}}$$

$$= \frac{E[(Z_t - \mu)(Z_{t+K} - \mu)]}{\sigma_Z^2} \tag{1.3.7}$$

因为，对于平稳随机时间序列，在时刻t和t＋K的方差是相同的，即$\sigma_Z^2 = \gamma_0$，所以其自相关系数

$$\rho_K = \frac{\gamma_K}{\gamma_0} \tag{1.3.8}$$

当$K = 0$，$\rho_0 = 1$。

四、自协方差矩阵和正定性

对于在N个连续时间点上进行观察的平稳随机时间序列(Z_1，Z_2，…Z_N)，其自协方差矩阵是

$$\Gamma_n=\begin{pmatrix}\gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1}\\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2}\\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0\end{pmatrix}=\sigma_Z^2\begin{pmatrix}1 & \rho_1 & \rho_2\cdots\rho_{n-1}\\ \rho_1 & 1 & \rho_1\cdots\rho_{n-2}\\ \rho_2 & \rho_1 & 1\cdots\rho_{n-1}\\ \cdots & \cdots & \cdots\\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3}\cdots 1\end{pmatrix}=\sigma_Z^2 P_n \tag{1.3.9}$$

Γ_n称为自协方差矩阵。它是一任意对角线都是相同的元素且对称的矩阵。P_n称为自 相 关 矩阵。

下面考虑随机变量Z_t，$Z_{t-1}\cdots Z_{t-n+1}$的任意线性函数

$$L_t=l_1Z_t+l_2Z_{t-1}+\cdots+l_nZ_{t-n+1} \tag{1.3.10}$$

因为平稳随机时间序列的协方差是

$$COV[Z_i\ \ Z_j]=\gamma_{|j-i|} \tag{1.3.11}$$

所以，如果（1.3.10）式的系数不全都等于0的话，则L_t的方差

$$Var[L_t]=\sum_{i=1}^{n}\sum_{j=1}^{n} l_i l_j \gamma_{|j-i|} \tag{1.3.12}$$

必定大于0。由此可以得出结论，任何平稳随机时间序列的自协方差矩阵Γ_n和自相关矩阵 P_n二者都是正定的。

（一）平稳随机时间序列的自相关系数应满足的条件。

自相关矩阵的正定性意味着，矩阵P_n的行列式及其全部主子式都必须 大 于 0。例 如，$n=2$，则要求行列式

$$\begin{bmatrix}1 & \rho_1\\ \rho_1 & 1\end{bmatrix}>0,\quad 即\ 1-\rho_1^2>0$$

因此， $-1<\rho_1<1$

类似的，当$n=3$时．自相关系数必须满足的条件是

$$\begin{bmatrix}1 & \rho_1\\ \rho_1 & 1\end{bmatrix}>0,\quad \begin{bmatrix}1 & \rho_2\\ \rho_2 & 1\end{bmatrix}>0$$

$$\begin{pmatrix}1 & \rho_1 & \rho_2\\ \rho_1 & 1 & \rho_1\\ \rho_2 & \rho_1 & 1\end{pmatrix}>0$$

因此，
$$-1<\rho_1<1$$
$$-1<\rho_2<1$$
$$-1<\frac{\rho_2-\rho_1^2}{1-\rho_1^2}<1$$

因为对于n为任意值的自相关矩阵P_n都是正定的，所以当观察数n很大时，平稳 随 机 时间序列的自相关系数必须满足大量的条件。虽然这些条件能够为随机时间序列的平稳性提供解析检验，但是在实际应用中，通常是对时间序列本身和样本自相关函数的直观检查，来判断平稳性。在某些应用情况下，只要记住，对于平稳随机时间序 列，$K>0$时，满 足 条 件$-1<\rho_K<1$就足够了。

（二）线性函数的平稳性。

根据平稳性定义，对平稳随机时间序列Z_t进行线性变换后的序列L_t[见（1.3.10）式]也是平稳的。特别是，平稳随机序列Z_t的一次差分$\Delta Z_t = Z_t - Z_{t-1}$和高次差分$\Delta^d Z_t$都是平稳随机时间序列。

（三）正态随机时间序列的平稳性。

如果一个随机时间序列Z_t的任意有穷维概率分布都是多变量正态分布，则称Z_t为正态随机时间序列。多变量正态分布的特性是由它的一阶和二阶矩所完全确定。如果对于所有n值，存在一个固定的均值μ和（1.3.9）式所示的自协方差矩阵Γ_n，就能充分保证该正态随机时间序列的平稳性。

五、平稳随机时间序列的自协方差函数和自相关函数

前面已经说明，自协方差系数γ_K是时间间隔（即滞后K）的两个数值Z_t和Z_{t+K}之间的协方差。γ_K随K的变化用K的函数γ_K表示，称为随机时间序列的自协方差函数。同样，自相关系数P_K随K的变化用K的函数P_K表示，称为随机时间序列的自相关函数。对于平稳随机时间序列，$\gamma_k = \rho_k \sigma_z^2$,所以只要已知自相关函数[$P_K$]和方差$\sigma_z^2$,就等于已知自协方差函数[$\gamma_K$]。图（1.3.2）表示，由自相关矩阵的对角线系数的数值绘制的自相关函数[P_K]的图形。

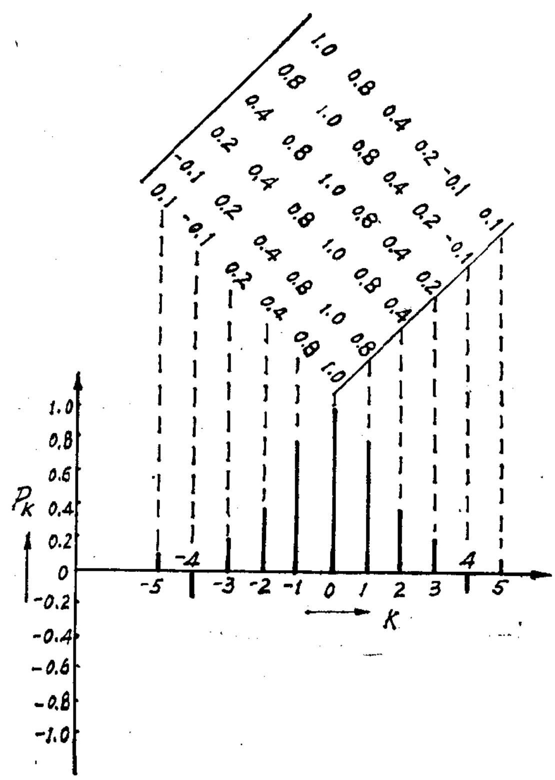

图1.3.2 由自相关矩阵产生的自相关函数

上图揭示了，一个平稳随机时间序列的任意两个数值之间的相关程度是随它们之间时间间隔的变化而变化的规律。因为$P_K = P_{-K}$，所以自相关函数P_K是沿K = 0点对称的。因此，只要

画出其正半部分，如图（1.3.3）所示本书下面在讨论中，涉及到自相关函数 P(K) 时，通常是指其正半部分。

前面已经说明，一个正态平稳随机时间序列，可以用它的均值μ和自协方差函数（γ_K）来完全描述其特征，也就是可以用它的均值μ，方差σ_z^2和自相关函数[P_K]来完全描述其特征。

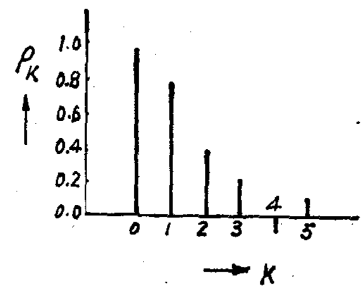

图1.3.3 图1.3.2中自相关函数的正半部分

六、自协方差函数和自相关函数的估计

前面讨论的只是描述一个概念性随机时间序列的理论自相关函数。实际上，由一个观察次数为N的有限时间序列Z_1，Z_2，…Z_N，只能获得自相关函数的估计值，称为样本自相关函数。自相关函数和自协方差函数的估计值ρ_K和γ_K的计算公式如下

$$\hat{\rho}_K = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \qquad (1.3.13)$$

$$\hat{\gamma} = \frac{1}{N}\sum_{t=1}^{N-K}(Z_t - \bar{Z})(Z_{t+K} - \bar{Z}), \qquad K = 0,1,2 \qquad (1.3.14)$$

其中$\bar{Z}$是时间序列的平均值。作为一个例子，利用表1.3.1中时间序列$t = 1 \sim 10$的数据来计算$\hat{\rho}_1$

表1.3.1时间序列数据

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Z_t	47	64	23	71	38	64	55	41	59	48	71	35	57	40	58	44	80	55	37
t	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Z_t	74	51	57	50	60	45	57	50	45	25	59	50	71	56	74	50	58	45	54
t	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
Z_t	36	54	48	55	45	57	50	62	44	64	43	52	38	59	55	41	53	49	34
t	58	59	60	61	62	63	64	65	66	67	68	69	70						
Z_t	35	54	45	68	38	50	60	39	59	40	57	54	23						

$$\bar{Z} = \frac{1}{10}\sum_{t=1}^{10} Z_t = 51$$

$$\hat{\gamma}_1=\frac{1}{10}\sum_{t=1}^{9}(Z_t-Z)(Z_{t+1}-Z)=\frac{-1497}{10}=-149.7$$

$$\hat{\gamma}_0=\frac{1}{10}\sum_{t=1}^{10}(Z_t-Z)^2=189.6$$

因此，
$$\hat{\rho}_1=\frac{r_1}{r_0}=\frac{-149.7}{189.6}=-0.79$$

对于大多数实际目的来说，估计自相关函数取两位数就够了。上面的计算只是为了说明的目的。实际上，为了获得自相关函数的有效估计，至少需要50个观察值，并且计算自相关系数ρ_K的个数不能大于N/4。

利用表1.3.1中的70个观察值，求出了K＝1～15的15个自相关系数列于表1.3.2，并将其绘成图1.3.4。

表1.3.2估计自相关函数

K	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	−0.39	0.30	−0.17	0.07	−0.10	−0.05	0.04	0.04
K	9	10	11	12	13	14	15	
$\hat{\rho}_k$	−0.01	0.01	0.11	−0.07	0.15	0.04	−0.01	

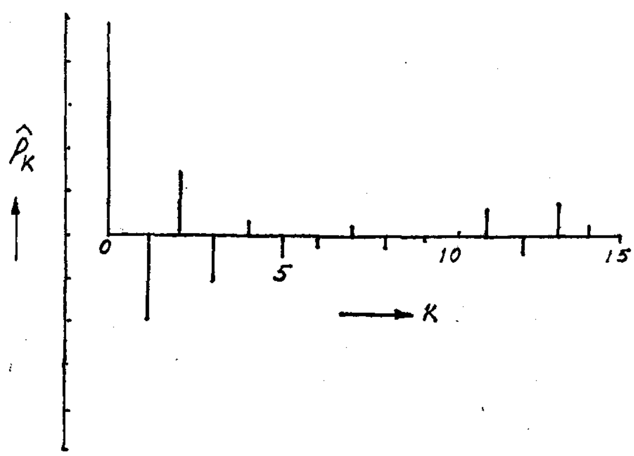

图 1.3.4 估计自相关函数

第二章　时间序列模型

与单方程回归模型和多方程模拟模型不同，时间序列模型不是根据与其他变量的因果关系来预测一个变量的未来变化，而是根据该变量过去的变化规律来预测其未来的变化。例如下图2.1.1中所示的$Z(t)$表示某经济或企业变量（如股票市场指数、利率、生产指数、或某商品的日销售量）的历史数据，但是不能基于经济理论或直观推断来解释该变量为什么如此变化。如果$Z(t)$表示某商品的销售量根据价格、个人收入和利率的变化而上下起伏，但是，影响其变化的许多因素可能是无法说明的，例如天气变化、顾客爱好的变化，或者只是顾客消费的季节性周期变化等等因素。因此，在许多情况下，很难或不可能利用变量之间的因果关系模型（即结构模型）来说明$Z(t)$的变化。可能有这种情况，在回归模型中，无法获得影响预测变量$Z(t)$的那些说明性变量（即自变量）的有效数据，或者是，即使能获得说明性变量的数据，但是由于回归模型估计的标准误差太大，致使大多数估计系数是不显著的，而因预测的标准误差大得无法接受。在某些情况下，即使能够估计一个$Z(t)$的统计显著的回归方程，其结果对于预测来说也可能是无效的。为了由回归方程求得$Z(t)$的预测值，必须首先对那些无滞后的说明性变量进行预测，对这些说明性变量进行预测可能比对$Z(t)$本身进行预测还要更为困难。利用说明性变量的未来预测值求得的$Z(t)$的预测值的标准误差可能比较小，如果回归方程拟合得很好的话，但是说明性变量未来值的预测误差太大而使得$Z(t)$的总误差大得无法接受。

很明显，存在不能用结构模型来说明Z_t的情况。对于这种情况，时间序列模型则是用于预测的一个很有效的工具。这种模型不象回归模型那样是根据因果关系，而是根据被预测变量过去的变化规律来建立模型，然后利用这个模型来预测该变量未来的变化。因此，时间序列模型是一个先进的外推方法。利用时间序列模型不需要知道影响预测变量的因果关系。

通常是在下述情况下选用时间序列模型：关于影响预测变量的决定性因素的信息很少；有足够多的数据量可以用来构成一个合理长度的时间序列。

时间序列模型包括确定型时间序列模型和随机型时间序列模型。确定型时间序列模型包括一些简单外推方法。随机型时间序列模型与简单外推的不同是，预测的时间序列是由随机过程产生的。随机时间序列模型能描述产生观察样本的随机过程的偶然性。这种描述，不是象回归模型那样根据因果关系，而是根据随机过程中所包含的偶然性。由于简单外推不能说明时间序列的随机性质，因此，随机时间序列分析要比简单外推方法更加完善，随机时间序列模型比确定型时间序列模型能提供更多的信息，所以能改善预测。

第一节　确定型时间序列模型

确定型时间序列模型，可用于根据时间序列的过去变化特征来预测其将来的变化特征。其所以是确定型的，因为它不涉及时间序列的随机性根源或随机性质。这种简单外推方法，

在经济预测和企业预测中，尽管精度不如随机模型，但是作为标准工具已应用多年了，因此有必要进行回顾和评价。

经常遇到的绝大多数时间序列在时间上都不是连续的，它们都是由时间间隔有规律而又不连续的观察所组成。图2.1.1所示系一种典型的离散时间序列。用Z_t表示该序列，这样，Z_1就代表第一次观察值，Z_2代表第二次观察值，Z_T是该序列的最后一次观察值。研究的目的是建立能够描述序列y_t的模型，并且用这个模型预测最后观察Z_T以后的Z_t。用$\hat{Z}_{T+1}$表示一个周期以后的预测，$\hat{Z}_{T+2}$表示两个周期以后的预测，$\hat{Z}_{T+l}$表示l个周期以后的预测。

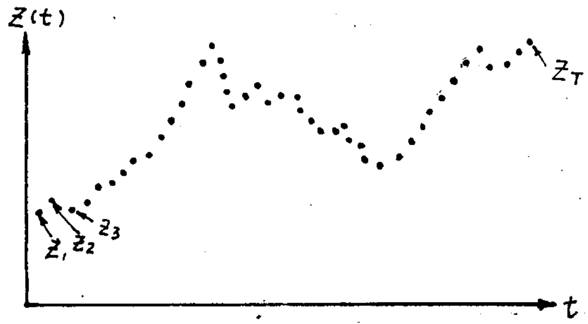

图 2.1.1 离散时间序列

如果观察的次数不太多，那么就可以用一个多项式来最简单和最全面地说明Z_t，也就是用时间的连续函数f(t)来描述Z_t，

$$f(t) = a_0 + a_1 + a_2t^2 + \cdots + a_nt^n \qquad (2.1.1)$$

式中，t为时间，a是系数，n = T − 1。这个多项式，只要系数a选择得当，可以穿过离散时间序列Z_t的各点。因而，就能知道在每一个时间t，由1到T，f(t)必等于Z_t。但是，能不能相信，通过f(t)对一个周期或几个周期以后的Z_t所作的预测将完全接近Z_t的实际未来值呢？例如下式

$$f(T+1) = a_0 + a_1(T+1) + a(T+1)^2 + \cdots a_{T-1}(T+1)^{T-1} = \hat{Z}_{T+1}$$

其预测是否接近实际未来值Z_{T+1}呢？遗憾的是，无法知道对这个问题的答案，除非有额外更重要的信息。2.1.1式的模型，其难点在于它不能描述Z_t，只能再现Z_t。它没有捕获到那些可能在未来自行重复的Z_t的任何特征。因此，f(t)虽与Z_t相关得很好，但对预测Z_t的用处不大。

下面简单介绍两种常用于预测的确定型时间序列模型。

一、简单外推模型

可以描述时间序列Z_t的基本特征之一是它的长期增长模式。尽管有短期的上下起伏运动，但Z_t可能显示明显的向上发展趋势。如果相信这种向上趋势存在并且将继续下去的话（且不问其理由），那么，就可以建立一种简单模型，描述该趋势，并用于预测，即外推Z_t。

最简单的外推模型是线性趋势模型。如果相信序列Z_t在每一个时间周期以恒定的绝对数量增长的话，那么，就可以通过拟合下述趋势线来预测未来的Z_t

$$Z_t = c_1 + c_2t \qquad (2.1.2)$$

式中，Z_t是在时间t时，Z的数值；c_1、c_2为系数。

通常选择在基础周期（第一次观察）时，令t等于零，而在以后各周期每次增加1。例如，设由回归确定下式

$$Z_t = 27.5 + 3.2t \qquad (2.1.3)$$

则可预测在周期t + 1时的Z值将比前一值高3.2单位。

更为现实的是，可以假定序列Z_t是以恒定百分数增加，而不是以恒定绝对值增加。这种趋势模型称为指数增长模型，如

$$Z_t = f(t) = Ae^{rt} \qquad (2.1.4)$$