时间序列分析

Tims Series Analysis

作者：组队学习

组织：Datawhale

时间：Sep 15, 2021

版本：9.15

ElegantLATEX Program

Victory won't come to us unless we go to it. — M. Moore

目录

第一章 统计与时间序列分析基础

内容提要

- ❑ 数据的统计描述
- ❑ 统计中几个重要的概率分布
- ❑ 正态总体统计量的分布
- ❑ 参数估计与假设检验
- ❑ 时间序列的基本定义
- ❑ 时间序列分析流程

1.1 数据的统计描述

数理统计研究的对象是受随机因素影响的数据（以下简称**数理统计**或**统计**），统计是以概率论为基础的一门应用学科。

人们往往希望通过少数几个包含最多相关信息的数据来描述数据总体的规律。描述性统计就是搜集、整理、加工和分析统计数据，使之系统化、条理化，以显示出数据资料的趋势、特征和数量关系。它是统计推断的基础，实用性较强，在统计工作中经常使用。面对一批数据如何进行描述与分析，需要掌握参数估计和假设检验这两个数理统计的最基本方法。

定义 1.1. 总体、个体与样本

1. **总体**是人们研究对象的全体，又称**母体**，如工厂一天生产的全部产品 (按合格品及废品分类)，学校全体学生的身高。
2. 总体中的每一个基本单位称为**个体**，个体的特征用一个变量 (如 x) 来表示，如一件产品是合格品记 $x=0$ ，是废品记 $x=1$ ；一个身高 170(cm) 的学生记 $x=170$ 。
3. 从总体中随机产生的若干个个体的集合称为**样本**，或子样，如 n 件产品，100 名学生的身高，或者一根轴直径的 10 次测量。实际上这就是从总体中随机取得的一批数据，不妨记作 $x_1, x_2, \cdots, x_n$，其中 n 称为**样本容量**。 ♣

简单地说，**统计的任务是由样本推断总体。**

一组数据 (样本) 往往是杂乱无章的，做出它的频数表和直方图，可以看作是对这组数据的一个初步整理和直观描述。

将数据的取值范围划分为若干个区间，然后统计这组数据在每个区间中出现的次数，称为**频数**，由此得到一个频数表。以数据的取值为横坐标，频数为纵坐标，画出一个阶梯形的图，称为直方图，或频数分布图。

假设有一个容量为 n 的样本 (即一组数据)，记作 $x=(x_1, x_2, \cdots, x_n)$ ，需要对它进行一定的加工，才能提出有用的信息，用作对总体 (分布) 参数的估计和检验。**统计量**就是加工出来的、反映样本数量特征的函数，它不含任何未知量。

下面我们介绍几种常用的统计量。

1. 表示位置的统计量—算术平均值和中位数

定义 1.2. 总体、个体与样本

算术平均值 (简称均值) 描述数据取值的平均位置，记作 $\bar{x}$,

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

中位数是将数据由小到大排序后位于中间位置的那个数值。♣

2. 表示变异程度的统计量—标准差、方差和极差

定义 1.3. 标准差、方差与极差

标准差 s 定义为

$$s=[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2]^{\frac{1}{2}}$$

它是各个数据与均值偏离程度的度量，这种偏离不妨称为变异，而方差是标准差的平方 s^2；极差是 $x=(x_1,x_2,\cdots,x_n)$ 的最大值与最小值之差。♣

你可能注意到标准差 s 的定义中，对 n 个 (x_i-x) 的平方求和，却被 $(n-1)$ 除，这是出于无偏估计的要求。如果强制采用 n 作为样本的估计会出现什么问题? 这会导致我们得到的结果是原样本的有偏估计。直观来说，样本的自由度为 $n-1$ 是因为我们减去了均值，导致样本数少 1，而当估计总体时，$(n-1)\sim n$ 是几乎相等的，所以当用 n 作为除数时意味着我们对整体进行估计。当然，一组数据的均值可能并不在其中，因此这种说法存在很多问题，下面我们用公式来推导一下，假设 $E(S_1^2)$ 为除以 n 的标准差，那么我们有:

$$\begin{aligned}
E\left(S_1^2\right) &= \frac{1}{n}\sum_{i=1}^{n}E\left(\left(X_i-\bar{X}\right)^2\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i-\mu+\mu-\bar{X}\right)^2\right)\\
&= \frac{1}{n}E\left(\sum_{i=1}^{n}\left(\left(X_i-\mu\right)^2-2\left(X_i-\mu\right)\left(\bar{X}-\mu\right)+\left(\bar{X}-\mu\right)^2\right)\right)\\
&= \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i-\mu\right)^2-2\sum_{i=1}^{n}\left(X_i-\mu\right)\left(\bar{X}-\mu\right)+n(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i-\mu\right)^2-2n(\bar{X}-\mu)(\bar{X}-\mu)+n(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i-\mu\right)^2-n(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}\left(\sum_{i=1}^{n}E\left(\left(X_i-\mu\right)^2\right)-nE\left(\left(\bar{X}-\mu\right)^2\right)\right)\\
&= \frac{1}{n}(n\operatorname{Var}(X)-n\operatorname{Var}(\bar{X}))\\
&= \operatorname{Var}(X)-\operatorname{Var}(\bar{X})=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2,
\end{aligned}$$

而 $\frac{n-1}{n}\sigma^2$ 并不等于 σ^2，因此我们需要调整系数才使得我们的估计为无偏估计。

3 中心矩、表示分布形状的统计量—偏度和峰度

随机变量 x 的 r 阶中心矩为 $E(x-Ex)^r$。

随机变量 x 的偏度和峰度指的是 x 的标准化变量 $(x-Ex)/\sqrt{Dx}$ 的三阶中心矩和四阶中心矩:

$$\nu_1 = E[(\frac{x-E(x)}{\sqrt{D(x)}})^3] = \frac{E[(x-E(x))^3]}{(D(x))^{3/2}}$$

$$\nu_2 = E[(\frac{x-E(x)}{\sqrt{D(x)}})^4] = \frac{E[(x-E(x))^4]}{(D(x))^2}$$

定义 1.4. 偏度与峰度

1. 偏度反映分布的对称性，ν_1 >0 称为右偏态，此时数据位于均值右边的比位于左边的多;ν_1 <0 称为左偏态，情况相反; 而 ν_1 接近 0 则可认为分布是对称的。
2. 峰度是分布形状的另一种度量，正态分布的峰度为 3，若 2 比 3 大得多，表示分布有沉重的尾巴，说明样本中含有较多远离均值的数据，因而峰度可以用作衡量偏离正态分布的尺度之一。

♣

统计量中最重要、最常用的是均值和标准差，由于样本是随机变量，它们作为样本的函数自然也是随机变量，当用它们去推断总体时，有多大的可靠性就与统计量的概率分布有关，因此我们需要知道几个重要分布的简单性质。

1.2 统计中几个重要的概率分布

随机变量的特性完全由它的 (概率) 分布函数或 (概率) 密度函数来描述。设有随机变量 X，其分布函数定义为 $X \leqslant x$ 的概率，即 $F(x) = P\{X \leqslant x\}$。若 X 是连续型随机变量，则其密度函数 $p(x)$ 与 $F(x)$ 的关系为

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x.$$

上 α 分位数是下面常用的一个概念，其定义为: 对于 $0<\alpha<1$，使某分布函数 $F(x) = 1-\alpha$ 的 x，称为这个分布的上 α 分位数，记作 x_α。

统计中几个重要的概率分布:

- **正态分布**

正态分布随机变量 X 的密度函数曲线呈中间高两边低、对称的钟形，期望 (均值) $EX=\mu$，方差 $DX=\sigma^2$，记作 $X \sim N(\mu,\sigma^2)$，σ 称均方差或标准差，当 $\mu=0,\sigma=1$ 时称为标准正态分布，记作 $X \sim N(0,1)$。正态分布完全由均值 和方差 σ^2 决定，它的偏度为 0，峰度为 3。正态分布可以说是最常见的 (连续型) 概率分布，成批生产时零件的尺寸，射击中弹着点的位置，仪器反复量测的结果，自然界中一种生物的数量特征等，多数情况下都服从正态分布，这不仅是观察和经验的总结，而且有着深刻的理论依据，即在大量相互独立的、作用差不多大的随机因素影响下形成的随机变量，其极限分布为正

态分布。

- **χ^2 分布**

若 $X_1, X_2, \cdots, X_n$ 为相互独立的、服从标准正态分布 $N(0,1)$ 的随机变量，则它们的平方和 $Y = \sum\limits_{i=1}^{n} X_i^2$ 服从 χ^2 分布，记作 $Y \sim \chi^2(n)$，n 称自由度，它的期望 $EY = n$，方差 $DY = 2n$。

- **t 分布**

若 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且相互独立, 则 $T = \frac{X}{\sqrt{Y/n}}$ 服从 t 分布, 记作 $T \sim t(n)$, n 称自由度。t 分布又称**学生氏 (Student) 分布**。

t 分布的密度函数曲线和 $N(0,1)$ 曲线形状相似。理论上 $n \to \infty$ 时，$T \sim t(n) \to N(0,1)$，实际上当 $n > 30$ 时它与 $N(0,1)$ 就相差无几了。

- **F 分布**

若 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$, 且相互独立，则 $F = \frac{X/n_1}{Y/n_2}$ 服从 F 分布，记作 $F \sim F(n_1, n_2)$，(n_1, n_2) 称自由度。

1.3 正态总体统计量的分布

用样本来推断总体，需要知道样本统计量的分布，而样本又是一组与总体同分布的随机变量，所以样本统计量的分布依赖于总体的分布。当总体服从一般的分布时，求某个样本统计量的分布是很困难的，只有在总体服从正态分布时，一些重要的样本统计量(均值、标准差) 的分布才有便于使用的结果。另一方面，现实生活中需要进行统计推断的总体，多数可以认为服从 (或近似服从) 正态分布，所以统计中人们在正态总体的假定下研究统计量的分布，是必要的与合理的。

设总体 $X \sim N(\mu, \sigma^2)$，其中 $x_1, x_2, \cdots, x_n$ 为一容量 n 的样本，其均值 $\bar{x}$ 和标准差 s 由公式确定，则用 $\bar{x}$ 和 s 构造的下面几个分布在统计中是非常有用的。

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \text{或} \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

设有两个总体 $X \sim N(\mu_1, \sigma^2)$ 和 $Y \sim N(\mu_2, \sigma^2)$，及由容量分别为 n，n 的两个样本确定的均值 $\bar{x}, \bar{y}$ 和标准差 s_1, s_2，则

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n}} \sim N(0,1)$$

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{s_w\sqrt{1/n_1+1/n_2}} \sim t\,(n_1+n_2-2)$$

其中，$s_w^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$，

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

对于上式，假定 $\sigma_1 = \sigma_2$，但它们未知，于是用 s 代替。在下面的统计推断中我们要反复用到这些分布。

1.4 参数估计

利用样本对总体进行统计推断的一类问题是参数估计，即假定已知总体的分布，通常是 $X \sim N(\mu, \sigma^2)$，估计有关的参数，如 μ, σ^2 。参数估计分点估计和区间估计两种。

- **点估计**

点估计是用样本统计量确定总体参数的一个数值。评价估计优劣的标准有无偏性、最小方差性、有效性等，估计的方法有矩法、极大似然法等。

最常用的是对总体均值 μ 和方差 σ^2 (或标准差 σ) 作点估计。让我们暂时抛开评价标准，当从一个样本算出样本均值 $\bar{x}$ 和方差 s^2 后，对 μ 和 σ^2(或 σ) 一个自然、合理的点估计显然是 (在字母上加ˆ示它的估计值)

$$\hat{\mu} = \bar{x}, \hat{\sigma}^2 = s^2, \hat{\sigma} = s$$

- **区间估计**

点估计虽然给出了待估参数的一个数值，却没有告诉我们这个估计值的精度和可信程度。一般地，总体的待估参数记作 θ(如 μ, σ^2)，由样本算出的 θ 的估计量记作 $\hat{\theta}$ ，人们常希望给出一个区间 $[\hat{\theta}_1, \hat{\theta}_2]$，使 θ 以一定的概率落在此区间内。若有

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha, 0 < \alpha < 1$$

则 $[\hat{\theta}_1, \hat{\theta}_2]$ 称为 θ 的置信区间，$\hat{\theta}_1, \hat{\theta}_2$ 分别称为置信下限和置信上限，$1 - \alpha$ 称为**置信概率**或**置信水平**，α 称为**显著性水平**。

给出的置信水平为 1-α 的置信区间 $[\hat{\theta}_1, \hat{\theta}_2]$，称为 θ 的**区间估计**。置信区间越小，估计的精度越高; 置信水平越大，估计的可信程度越高。但是这两个指标显然是矛盾的，通常是在一定的置信水平下使置信区间尽量小。通俗地说，区间估计给出了点估计的误差范围。

1.5 假设检验

统计推断的另一类重要问题是假设检验问题。在总体的分布函数完全未知或只知其形式但不知其参数的情况，为了推断总体的某些性质，提出某些关于总体的假设。例如，提出总体服从泊松分布的假设，又如对于正态总体提出数学期望等于 μ_0 的假设等。假设检验就是根据样本对所提出的假设做出判断: 是接受还是拒绝。这就是所谓的假设检验问题。

1.5.1 单个总体 $N(\mu, \sigma^2)$ 均值 μ 的检验

假设检验有三种:

双边检验:$H_0 : \mu = \mu_0$，$H_1 : \mu \neq \mu_0$;

右边检验:H_0 :$\mu \leqslant \mu_0$，H_1 :$\mu > \mu_0$;

左边检验:H_0 :$\mu \geqslant \mu_0$，H_1 :$\mu < \mu_0$;

- σ^2 已知，关于 μ 的检验 (**Z** 检验)

在 Matlab 中 **Z** 检验法由函数 ztest 来实现，命令为

$$[h, p, ci] = ztest(x, mu, sigma, alpha, tail)$$

其中输入参数 x 是样本，mu 是 H_0 中的 μ_0，sigma 是总体标准差 σ，alpha 是显著性水平 α(alpha 缺省时设定为 0.05)，tail 是对备选假设 H_1 的选择：H_1 为 $\mu \neq \mu_0$ 时用 tail=0(可缺省); H_1 为 $\mu > \mu_0$ 时用 tail=1; H_1 为 $\mu < \mu_0$ 时用 tail=-1。输出参数 h=0 表示接受 H_0，h=1 表示拒绝 H_0，p 表示在假设 H_0 下样本均值出现的概率，p 越小 H_0 越值得怀疑，ci 是 μ_0 的置信区间。

- σ^2 未知，关于 μ 的检验 (t 检验)

在 Matlab 中 t 检验法由函数 ttest 来实现，命令为

$$[h, p, ci] = ttest(x, mu, alpha, tail)$$

1.5.2 分布拟合检验

在实际问题中，有时不能预知总体服从什么类型的分布，这时就需要根据样本来检验关于分布的假设。下面介绍 χ^2 检验法和专用于检验分布是否为正态的“偏峰、峰度检验法”。

- **χ^2 检验法**

什么是 χ^2 检验？χ^2 检验就是在总体的分布未知的情况下，根据样本 $X_1, X_2, \cdots, X_n$ 来检验关于总体分布的假设。一般情况下，我们会设置原假设：H_0：总体 X 的分布函数为 $F(x)$，进而得到备择假设 H_1：总体 X 的分布函数不是 $F(x)$，然后运用统计学方法进行检验。值得注意的是，若总体 X 为离散型，则上述假设写为 H_0：总体 X 的分布律为 $P\{X = t\} = p, i = 1, 2, \cdots$. 若总体 X 为连续型，则上述假设写为 H_0：总体 X 的概率密度为 $f(x)$. 此外，在使用 χ^2 检验法检验假设 H_0 时，若 $F(x)$ 的形式已知，但有参数值未知，需要先用最大似然估计法估计参数，然后作检验.

χ^2 检验的整体思路，就是局部样本的某些统计学特征和整体样本相似。将随机试验可能结果的全体 Ω 分为 m 个互不相容的事件 $A_1, A_2, \cdots, A_m(\sum\limits_{i=1}^{m} A_i = \Omega, A_iA_j = \Phi, i \neq j, i, j = 1, 2, \cdots, m)$. 于是在假设 H_0 下，我们可以计算 $p_{i0} = P(A_i), i = 1, 2, \cdots, m$. 在 n 次试验中，事件 A_i 出现的频率 $\frac{N_i}{n}$ 与 p_{i0} 往往有差异，但一般来说，若 H_0 为真，且试验次数又多时，这种差异不应很大.

练习 1.1 把一颗骰子重复抛掷 300 次，结果如下：
试证明这颗骰子的六个面并不匀称 (取 $\alpha = 0.05$)。

出现的点数	1	2	3	4	5	6
出现的频数	40	70	48	60	52	30

证明 根据题意需要检验假设：

$$H_0: \quad P\{X=i\}=\frac{1}{6} \quad (i=1,2,\cdots,6)$$

其中 X 表示抛掷这骰子一次所出现的点数 (可能值只有 6 个)，在 H_0 为真的前提下，

$$p_{i0}=\frac{1}{6} \quad (i=1,2,3\cdots,6)$$

$$\chi_n^2=\sum_{i=1}^{6}\frac{(N_i-np_{i0})^2}{np_{i0}}=\frac{\left(40-300\times\frac{1}{6}\right)^2}{300\times\frac{1}{6}}+\cdots+\frac{\left(30-300\times\frac{1}{6}\right)^2}{300\times\frac{1}{6}}=20.16$$

自由度为 6 - 1 = 5，而 $\chi_{0.05}^2=11.07, \chi^2=20.16$，所以拒绝原假设，这颗骰子的六个面并不均匀。

✍ **练习 1.2** 某盒中装有白球和黑球，现做下面的试验，用返回式抽取方式从盒子中取球，直到白球为止，记录下抽取的次数，重复如此的试验 100 次，其结果为:

抽取次数	1	2	3	4	5 以上
频数	43	31	15	6	5

试证明该盒中的白球与黑球的个数相等 (取 α = 0.05)。

证明 从题意可知，该总体服从几何分布，

$$P\{X=k\}=(1-p)^{k-1}p, \quad k=1,2,\cdots,$$

若黑球白球个数相等，则 p=1/2, 因此

$$P\{X=1\}=\tfrac{1}{2}, P\{X=2\}=\tfrac{1}{4}, P\{X=3\}=\tfrac{1}{8}$$
$$P\{X=4\}=\tfrac{1}{16}, P\{X\geq 5\}=\tfrac{1}{16}$$

由此可知，检验的问题是

$$H_0: \boldsymbol{p}_1=\frac{1}{2}, \boldsymbol{p}_2=\frac{1}{4}, \boldsymbol{p}_3=\frac{1}{8}, \boldsymbol{p}_4=\frac{1}{16}, \boldsymbol{p}_5=\frac{1}{16}$$

计算皮尔逊统计量可得:

$$\chi_n^2=\sum_{i=1}^{m}\frac{(N_i-np_{i0})^2}{np_{i0}}=3.2$$

查表可得：$\chi_{0.05}^2(4)=9.488$，显然 $3.2<9.488$，因而接受原假设，黑球白球个数相等。

第二章 手算时间序列

内容提要

- 简单移动平均法
- 加权移动平均法
- 趋势移动平均法
- 指数平滑法

时间序列数据本质上反映的是某个或某些随机变量时间不断变化的趋势，而时间序列方法的核心就是从数据中挖掘出这种规律，并利用其对将来的数据做出估计。随着社会的进步和计算机技术的发展，时间序列分析的应用越来越广泛，已在经济、气象、地质、水文、军事等领域产生了显著的经济效益和社会效益。

时间序列也称**动态序列**，是指将某种现象的指标数值按照时间顺序排列而成的数值序列。时间序列的范围极为广泛，在日常生活中我们也常常可以见到其身影，例如某家商店要统计未来一段时间内需要进货多少；研究股票市场的走势；分析人口数量等都可能会用到时间序列分析。其应用方向可大致分成三个部分，分别是**描述过去**、**分析规律**和**预测未来**。

通常，影响时间序列的因素有四个：**长期变动趋势 (T)**、**季节变动规律 (S)**、**周期变动规律 (C)** 与**不规则变动 (I)**（也称随机扰动项）。有时这些变动会同时出现在一个时间序列里面，有时也可能只出现一种或几种，这是由引起各种变动的影响因素决定的。正是由于变动组合的不确定性，时间序列的数值变化才那么千变万化。四种变动与指标数值最终变动的关系可能是叠加关系，也可能是乘积关系。

定理 2.1. 叠加模型和乘积模型

- 如果四种变动之间是相互独立的关系，那么叠加模型可以表示为:

$$Y = T + S + C + I$$

- 如果四种变动之间存在相互影响关系，那么应该使用乘积模型:

$$Y = T \times S \times C \times I$$

下面我们将从日常生活中最基础的时间序列分析入手，逐步拆解复杂时间序列模型的数学原理。

2.1 简单移动平均法

移动平均法是根据时间序列资料逐渐推移，依次计算包含一定项数的时序平均数，以反映长期趋势的方法。当时间序列的数值由于受周期变动和不规则变动的影响，起伏较大，不易显示出发展趋势时，可用移动平均法，消除这些因素的影响，分析、预测序列的长期趋势。移动平均法有**简单移动平均法**，**加权移动平均法**，**趋势移动平均法**等。

如果数据具有一定的规律性，且一直在某一个值附近波动，那么我们可以用该组数据的平均值来近似拟合下一次可能出现的值。例如，从第 1 期开始一直到第 10 期我们的数据如下：

$$2,\quad 3,\quad 3,\quad 4,\quad 2,\quad 3,\quad 3,\quad 5,\quad 2,\quad 3$$

假设当前观测值只受到过去 4 期数值影响，可以计算 $\hat{y}_5 = \frac{1}{4}(2+3+3+4) = 3, \hat{y}_6 = \frac{1}{4}(3+3+4+2) = 3, \cdots$ 如此计算出所有的预测值，并与观测值之间对比误差，可得：

表 2.1: 预测值与实际值对比

期数	5	6	7	8	9	10
预测值:	3	3	3	3	3.25	3.25
真实值:	2	3	3	5	2	3
误差:	1	0	0	-2	1.25	0.25

通过误差分析发现：该种方法预测的准确度较高，因此我们可以预测第 11 期数值：$\hat{y}_{11} = \frac{1}{4}(3+5+2+3) = 3.25$，进而我们可以总结出更为一般的标准误计算公式：

定义 2.1. 简单移动平均法误差计算公式

设 N 为移动的平均项数，$\hat{y}_t$ 为第 t 期预测值，则有标准误：

$$S = \sqrt{\frac{\sum\limits_{t=N+1}^{T} (\hat{y}_t - y_t)^2}{T - N}}$$

最近 N 期序列值的平均值作为未来各期的预测结果，一般 N 取值范围：$5 \leqslant \mathrm{N} \leqslant 200$。 ♣

当历史序列的基本趋势变化不大且序列中随机变动成分较多时，N 的取值应较大一些。否则 N 的取值应小一些。在有确定的季节变动周期的资料中，移动平均的项数应取周期长度。选择最佳 N 值的一个有效方法是，比较若干模型的预测误差。预测标准误差最小者为好。简单移动平均法只适合做近期预测，而且是预测目标的发展趋势变化不大的情况。如果目标的发展趋势存在其它的变化，采用简单移动平均法就会产生较大的预测偏差和滞后。

练习 2.1 某企业 1 月 ∼11 月份的销售收入时间序列如表 2 示。试用一次简单滑动平均法预测第 12 月份的销售收入。

表 2.2: 销售收入时间序列表

月份 t	1	2	3	4	5	6
销售收入 y_t	533.8	574.6	606.9	649.8	705.1	772
月份 t	7	8	9	10	11	
销售收入 y_t	816.4	892.7	963.9	1015.1	1102.7	

解：分别取 N =4,N =5 的预测公式

$$\hat{y}_{t+1}^{(1)} = \frac{y_t + y_{t-1} + y_{t-2} + y_{t-3}}{4}, t = 4, 5, \cdots, 11$$

$$\hat{y}_{t+1}^{(2)} = \frac{y_t + y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}}{5}, t = 5, \cdots, 11$$

当 N=4 时，预测值 $\hat{y}_{12}^{(1)} = 993.6$，预测的标准误差为

$$S_1 = \sqrt{\frac{\sum\limits_{t=5}^{11} (\hat{y}_t^{(1)} - y_t)^2}{11 - 4}} = 150.5$$

当 N=5 时，预测值 $\hat{y}_{12}^{(2)} = 182.4$，预测的标准误差为

$$S_2 = \sqrt{\frac{\sum\limits_{t=6}^{11} (\hat{y}_t^{(2)} - y_t)^2}{11 - 5}} = 958.2$$

计算结果表明，N=4 时，预测的标准误差较小，所以选取 N=4. 预测的第 12 月销售收入为 993.6。

2.2 加权移动平均法

在简单移动平均公式中，每期数据在求平均时的作用是等同的。但是，每期数据所包含的信息量不一样，近期数据包含着更多关于未来情况的信息。因此，把各期数据等同看待是不尽合理的，应考虑各期数据的重要性，对近期数据给予较大的权重，这就是加权移动平均法的基本思想。

定义 2.2. 加权移动平均法

设时间序列为 $y_1, y_2, \cdots, y_t, \cdots$；加权移动平均公式为

$$M_{tw} = \frac{w_1 y_t + w_2 y_2 + \cdots + w_N y_{t-N+1}}{w_1 + w_2 + \cdots + w_N}$$

利用加权移动平均数来做预测，其预测公式为

$$\hat{y}_{t+1} = M_{tw}$$

即以第 t 期加权移动平均数作为第 t +1 期的预测值。♣

在加权移动平均法中，w_t 的选择，同样具有一定的经验性。一般的原则是: 近期数据的权数大，远期数据的权数小。至于大到什么程度和小到什么程度，则需要按照预测者对序列的了解和分析来确定。

2.3 趋势移动平均法

简单移动平均法和加权移动平均法，在时间序列没有明显的趋势变动时，能够准确反映实际情况。但当时间序列出现直线增加或减少的变动趋势时，用简单移动平均法和加权移动平均法来预测就会出现滞后偏差。因此，需要进行修正，修正的方法是作二次

移动平均，利用移动平均滞后偏差的规律来建立直线趋势的预测模型。这就是趋势移动平均法。

定义 2.3. 一次移动平均与二次移动平均法

一次移动的平均数为：

$$M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \cdots + y_{t-N+1})$$

在一次移动平均的基础上再进行一次移动平均就是二次移动平均，其计算公式为

$$M_t^{(2)} = \frac{1}{N}(M_t^{(1)} + \cdots + M_{t-N+1}^{(1)}) = M_{t-1}^{(2)} + \frac{1}{N}(M_t^{(1)} - M_{t-N}^{(1)})$$

♣

下面讨论如何利用移动平均的滞后偏差建立直线趋势预测模型。设时间序列 $\{y_t\}$ 从某时期开始具有直线趋势，且认为未来时期也按此直线趋势变化，则可设此直线趋势预测模型为：

$$\hat{y}_{t+m} = a_t + b_t m, \quad m = 1, 2, \cdots$$

其中 t 为当前时期数; m 为由 t 至预测期的时期数; a_t 为截距; b_t 为斜率。两者又称为平滑系数。举个例子，假如我们有六个观测到的数据分别为

$$2, \quad 4, \quad 6, \quad 8, \quad 10, \quad 12.$$

可由平滑系数公式：

$$\begin{cases} a_t = 2M_t^{(1)} - M_t^{(2)} \\ b_t = \frac{2}{N-1}(M_t^{(1)} - M_t^{(2)}) \end{cases}$$

分别计算这条直线的斜率、截距。值得注意的是，这种做法会大幅度减少可用的时间序列个数，在进行拟合、预测时需要大量的数据才可以进行。

✍ **练习 2.2** 我国 1965~1985 年的发电总量如下表，试预测 1986 年和 1987 年的发电总量。

表 2.3: 我国发电量及一、二次移动平均值计算表

年份	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
t	1	2	3	4	5	6	7	8	9	10	11
发电总量	676	825	774	716	940	1159	1384	1524	1668	1668	1958
一次 (N=6)						848.3	966.3	1082.8	1231.8	1393.8	1563.5
二次 (N=6)											1181.1

年份	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
t	12	13	14	15	16	17	18	19	20	21
发电总量	2031	2234	2566	2820	3006	3093	3277	3514	3770	4107
一次 (N=6)	1708.8	1850.5	2024.2	2216.2	2435.8	2625	2843.7	3046	3246.7	3416.2
二次 (N=6)	1324.5	1471.9	1628.8	1792.8	1966.5	2143.4	2330.7	2530	2733.7	2941.2

解： 通过对数据可视化，可见发电总量基本呈直线上升趋势，可用趋势移动平均法来预测。取 N = 6 ，分别计算一次和二次移动平均值并列于表中。

$$M_{21}^{(1)} = 3461.2, \quad M_{21}^{(2)} = 2941.2$$

代入平滑系数公式，得：

$$a_{21} = 2M_{21}^{(1)} - M_{21}^{(2)} = 3981.1$$

$$b_{21} = \frac{2}{6-1}(M_{21}^{(1)} - M_{21}^{(2)}) = 208$$

于是可以得到 $t = 21$ 时直线趋势预测模型为

$$\hat{y}_{21+m} = 3981.1 + 208m$$

预测未来两年分别为 4192.1 和 4397.1.

趋势移动平均法对于同时存在直线趋势与周期波动的序列，是一种既能反映趋势变化，又可以有效地分离出来周期变动的方法。

2.4 指数平滑法

一次移动平均实际上认为最近 N 期数据对未来值影响相同，都加权 $\frac{1}{N}$; 而 N 期以前的数据对未来值没有影响，加权为 0。但是，二次及更高次移动平均数的权数却不是 $\frac{1}{N}$ ，且次数越高，权数的结构越复杂，但永远保持对称的权数，即两端项权数小，中间项权数大，不符合一般系统的动态性。一般说来历史数据对未来值的影响是随时间间隔的增长而递减的。所以，更切合实际的方法应是对各期观测值依时间顺序进行加权平均作为预测值。指数平滑法可满足这一要求，而且具有简单的递推形式。指数平滑法根据平滑次数的不同，又分为一次指数平滑法、二次指数平滑法和三次指数平滑法等。

2.4.1 一次指数平滑法

定义 2.4. 一次指数平滑法

设时间序列为 $y_1, y_2, \cdots, y_t, \cdots, \alpha$ 为加权系数，0<α<1，一次指数平滑公式为：

$$S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha(y_t - S_{t-1}^{(1)})$$

♣

这是由移动平均公式改进而来的，移动平均数的递推公式为

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N}$$

以 $M_{t-1}^{(1)}$ 作为 y_{t-N} 的最佳估计，可以推导一次指数平滑法公式。将一次指数平滑公式展开，有：

$$S_t^{(1)} = \alpha y_t + (1-\alpha)[\alpha y_{t-1} + (1-\alpha)S_{t-2}^{(1)}] = \alpha \sum_{j=0}^{\infty}(1-\alpha)^j y_{t-j}$$

该式表明，$S_t^{(1)}$ 是全部历史数据的加权平均，加权系数分别为 $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \cdots$;

显然有

$$\alpha\sum_{j=0}^{\infty}(1-\alpha)^j=\frac{\alpha}{1-(1-\alpha)}=1$$

由于加权系数符合指数规律，又具有平滑数据的功能，故称为指数平滑。以这种平滑值进行预测，就是一次指数平滑法。预测模型为

$$\hat{y}_{t+1}=S_t^{(1)}$$

即:

$$\hat{y}_{t+1}=\alpha y_t+(1-\alpha)\hat{y}_t$$

也就是以第 t 期指数平滑值作为 t +1 期预测值。

在进行指数平滑时，加权系数的选择是很重要的。由上式可以看出，α 的大小规定了在新预测值中新数据和原预测值所占的比重。α 值越大，新数据所占的比重就愈大，原预测值所占的比重就愈小，反之亦然。若把公式改写为

$$\hat{y}_{t+1}=\hat{y}_t+\alpha(y_t-\hat{y}_t)$$

则从上式可看出，新预测值是根据预测误差对原预测值进行修正而得到的。α 的大小则体现了修正的幅度，α 值愈大，修正幅度愈大;α 值愈小，修正幅度也愈小。若选取 α =0，则 $\hat{y}_{t+1}=\hat{y}_t$，即下期预测值就等于本期预测值，在预测过程中不考虑任何新信息;若选取 α=1，则 $\hat{y}_{t+1}=y_t$ ，即下期预测值就等于本期观测值，完全不相信过去的信息。这两种极端情况很难做出正确的预测。因此，α 值应根据时间序列的具体性质在 0~1 之间选择。具体如何选择一般可遵循下列原则:

定理 2.2. α 的选择

1. 如果时间序列波动不大，比较平稳，则 α 应取小一点，如 (0.1~0.5)。以减少修正幅度，使预测模型能包含较长时间序列的信息;
2. 如果时间序列具有迅速且明显的变动倾向，则 α 应取大一点，如 (0.6~0.8)。使预测模型灵敏度高一些，以便迅速跟上数据的变化。

在实用上，类似移动平均法，多取几个 α 值进行试算，看哪个预测误差小，就采用哪个。

用一次指数平滑法进行预测，除了选择合适的 α 外，还要确定初始值 $s_0^{(1)}$ 。初始值是由预测者估计或指定的。当时间序列的数据较多，比如在 20 个以上时，初始值对以后的预测值影响很少，可选用第一期数据为初始值。如果时间序列的数据较少，在 20 个以下时，初始值对以后的预测值影响很大，这时，就必须认真研究如何正确确定初始值。一般以最初几期实际值的平均值作为初始值。

练习 2.3 某市 1976~1987 年某种电器销售额如表 4 所示。试预测 1988 年该电器销售额。

表 2.4: 某种电器销售额及指数平滑预测值计算表（单位：万元）

年份	t	实际销售额 y_t	预测值 $\hat{y}_t, \alpha = 0.2$	预测值 $\hat{y}_t, \alpha$ =0.5	预测值 $\hat{y}_t, \alpha$ =0.8
1976	1	50	51	51	51
1977	2	52	50.8	50.5	50.2
1978	3	47	51.04	51.25	51.64
1979	4	51	50.23	49.13	47.93
1980	5	49	50.39	50.06	50.39
1981	6	48	50.11	49.53	49.28
1982	7	51	49.69	48.77	48.26
1983	8	40	49.95	49.88	50.45
1984	9	48	47.96	44.94	42.09
1985	10	52	47.97	46.47	46.82
1986	11	51	48.77	49.24	50.96
1987	12	59	49.22	50.12	50.99

采用指数平滑法，并分别取 α = 0.2,0.5 和 0.8 进行计算，初始值

$$S_0^{(1)} = \frac{y_1 + y_2}{2} = 51, \quad \text{即：} \hat{y}_1 = S_0^{(1)} = 51$$

按照预测模型 $\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$，从表中可以看出，$\alpha$= 0.2,0.5 和 0.8 时，预测值是很不相同的。究竟 α 取何值为好，可通过计算它们的预测标准误差 S，选取使 S 较小的那个 α 值。预测的标准误差为:

表 2.5: 预测的标准误差

α	0.2	0.5	0.8
S	4.5029	4.5908	4.8426

计算结果表明:α =0.2 时，S 较小，预测 1988 年该电器销售额为 y_{1988} =51.1754。

2.4.2 二次指数平滑法

一次指数平滑法虽然克服了移动平均法的缺点。但当时间序列的变动出现直线趋势时，用一次指数平滑法进行预测，仍存在明显的滞后偏差。因此，也必须加以修正。修正的方法与趋势移动平均法相同，即再作二次指数平滑，利用滞后偏差的规律建立直线趋势模型。这就是二次指数平滑法。

定义 2.5. 二次指数平滑法

二次指数平滑法其计算公式为

$$S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)}$$
$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)}$$

♣

式中 $S_t^{(1)}$ 为一次指数的平滑值; $S_t^{(2)}$ 为二次指数的平滑值。当时间序列 $\{y_t\}$, 从某时期开始具有直线趋势时, 类似趋势移动平均法, 可用直线趋势模型: $\hat{y}_{t+m} = a_t + b_t m, m = 1, 2, \cdots$, 计算公式为

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \frac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)}) \end{cases}$$

练习 2.4 仍以我国 1965~1985 年的发电总量资料为例，试用二次指数平滑法预测 1986 年和 1987 年的发电总量。

解：取 $\alpha = 0.3$，初始值 $S_0^{(1)}$ 和 $S_0^{(2)}$ 都取序列的首项数值，即 $S_0^{(1)} = S_0^{(2)}$=676，计算 $S_0^{(1)}, S_0^{(2)}$ 列于表中。得到 $S_{21}^{(1)}$=3523.1， $S_{21}^{(2)}$ =3032.6，代入系数公式：

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} = 4013.7 \\ b_t = \frac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)}) = 210.24 \end{cases}$$

表 2.6: 我国发电量及一、二次移动平均值计算表

年份	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
t	1	2	3	4	5	6	7	8	9	10	11
发电总量 y_t	676	825	774	716	940	1159	1384	1524	1668	1668	1958
一次平滑值	676	720.7	736.7	730.5	793.3	903	1047.3	1190.3	1333.6	1439.9	1595.4
二次平滑值	676	689.4	703.6	711.7	736.2	786.2	864.6	962.3	1073.7	1183.6	1307.1
y_{t+1} 的估计值		676	765.4	784	757.4	875	1069.9	1308.4	1516.1	1705	1806.1

年份	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
t	12	13	14	15	16	17	18	19	20	21
发电总量 y_t	2031	2234	2566	2820	3006	3093	3277	3514	3770	4107
一次平滑值	1726.1	1878.4	2084.7	2305.3	2515.5	2688.8	2865.2	3059.9	3272.9	3523.1
二次平滑值	1432.8	1566.5	1722	1897	2082.5	2264.4	2444.6	2629.2	2822.3	3032.6
y_{t+1} 的估计值	2007.2	2145	2324.1	2602.9	2888.6	3134.1	3295	3466.1	3675.1	3916.6

于是，得 t = 21 时直线趋势方程为：y_{21+m} =4013.7+210.24m，预测 1986 年和 1987 年的发电总量为 (单位: 亿度)

$$y_{1986} = y_{22} = y_{21+1} = 4223.95$$

$$y_{1987} = y_{23} = y_{21+2} = 4434.19$$

2.4.3 三次指数平滑法

当时间序列的变动表现为二次曲线趋势时，则需要用三次指数平滑法。三次指数平滑是在二次指数平滑的基础上，再进行一次平滑，计算式定义如下：

定义 2.6. 三次指数平滑法

三次指数平滑法其计算公式为

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \end{cases}$$

♣

式中 $S_{t-1}^{(3)}$ 为三次指数平滑值。三次指数平滑法的预测模型为：$\hat{y}_{t+m} = a_t + b_t m + C_t m^2, m = 1, 2, \cdots$，其中

$$\begin{cases} a_t = 3S_t^{(1)} 3S_t^{(2)} + S_t^{(3)} \\ b_t = \frac{\alpha}{2(1-\alpha)^2}[(65\alpha)S_t^{(1)} 2(54\alpha)S_t^{(2)} + (43\alpha)S_t^{(3)}] \\ c_t = \frac{\alpha^2}{2(1-\alpha)^2}[S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}] \end{cases}$$

✍ **练习 2.5** 某省 1978~1988 年全民所有制单位固定资产投资总额如下表所示，试预测 1989 年和 1990 年固定资产投资总额。

表 2.7: 某省全民所有制单位固定资产投资总额及一、二、三次指数平滑值计算表 (单位: 亿元)

年份	t	投资总额 y_t	一次平滑	二次平滑	三次平滑	y_{t+1} 的估计值
1978	1	20.04	21.37	21.77	21.89	21.94
1979	2	20.06	20.98	21.53	21.78	20.23
1980	3	25.72	22.4	21.79	21.78	19.56
1981	4	34.61	26.06	23.07	22.17	24.49
1982	5	51.77	33.78	26.28	23.4	34.59
1983	6	55.92	40.42	30.52	25.54	53.89
1984	7	80.65	52.49	37.11	29.01	64.58
1985	8	131.11	76.07	48.8	34.95	89.3
1986	9	148.58	97.83	63.51	43.52	142.42
1987	10	162.67	117.28	79.64	54.35	176.09
1988	11	232.26	151.77	101.28	68.43	196.26

对数据进行可视化，可见投资总额呈二次曲线上升，可用三次指数平滑法进行预测。取 α=0.3，初始值 $S_1^{(0)} = S_2^{(0)} = S_3^{(0)} = \frac{y_1+y_2+y_3}{3} = 21.94$，计算 $S_t^{(1)}, S_t^{(2)}, S_t^{(3)}$ 列于表中，得到得到

$$S_{11}^{(1)} = 151.77 S_{11}^{(2)} = 101.28 S_{11}^{(3)} = 68.43$$

可得到当 t =11 时，

$$a_{11} = 219.91, b_{11} = 38.38, c_{11} = 1.62$$

于是，得 t =11 时预测模型为 $\hat{y}_{11+m} = 219.91 + 38.38m + 1.62m^2$，预测 1989 年和

1990 年的固定资产投资总额为 (单位: 亿元):

$$\hat{y}_{1989} = \hat{y}_{12} = \hat{y}_{11+1} = a_{11} + b_{11} + c_{11} = 259.91$$

$$\hat{y}_{1990} = \hat{y}_{13} = \hat{y}_{11+2} = a_{11} + 2b_{11} + 2^2 c_{11} = 303.16$$

因为国家从 1989 年开始对固定资产投资采取压缩政策，这些预测值显然偏高，应作适当的修正，以消除政策因素的影响。

指数平滑预测模型是以时刻 t 为起点，综合历史序列的信息，对未来进行预测的。选择合适的加权系数 α 是提高预测精度的关键环节。根据实践经验，α 的取值范围一般以 0.1~0.3 为宜。α 值愈大，加权系数序列衰减速度愈快，所以实际上 α 取值大小起着控制参加平均的历史数据的个数的作用。α 值愈大意味着采用的数据愈少。因此，可以得到选择 α 值的一些基本准则。

1. 如果序列的基本趋势比较稳，预测偏差由随机因素造成，则 α 值应取小一些，以减少修正幅度，使预测模型能包含更多历史数据的信息。
2. 如果预测目标的基本趋势已发生系统的变化，则 α 值应取得大一些。这样，可以偏重新数据的信息对原模型进行大幅度修正，以使预测模型适应预测目标的新变化。

另外，由于指数平滑公式是递推计算公式，所以必须确定初始值 $S_0^{(1)}, S_0^{(2)}, S_0^{(3)}$。可以取前 3~5 个数据的算术平均值作为初始值。

第 三 章 常用时间序列模型

内容提要

- 差分指数平滑法
- 自适应滤波法
- 趋势外推预测方法
- 平稳时间序列
- 平稳序列自协方差函数与自相关函数的估计

3.1 差分指数平滑法

在上节我们已经讲过，当时间序列的变动具有直线趋势时，用一次指数平滑法会出现滞后偏差，其原因在于数据不满足模型要求。因此，我们也可以从数据变换的角度来考虑改进措施，即在运用指数平滑法以前先对数据作一些技术上的处理，使之能适合于一次指数平滑模型，以后再对输出结果作技术上的返回处理，使之恢复为原变量的形态。差分方法是改变数据变动趋势的简易方法。下面我们讨论如何用差分方法来改进指数平滑法。当时间序列呈直线增加时，可运用一阶差分指数平滑模型来预测。其公式如下:

$$\nabla y_t = y_t - y_{t-1}$$
$$\nabla \hat{y}_{t+1} = \alpha \nabla y_t + (1-\alpha)\nabla \hat{y}_t$$
$$\hat{y}_{t+1} = \nabla \hat{y}_{t+1} + y_t$$

其中，∇ 为差分符号，上述公式是等价的。

练习 3.1 某工业企业 1977~1986 年锅炉燃料消耗量资料如下表所示，试预测 1987 年的燃料消耗量。

表 3.1: 某企业锅炉燃料消耗量的差分指数平滑法计算表 (= 0.4)(单位: 百吨)

年份	t	燃料消耗量	差分	差分指数平滑值	预测值
1977	1	24			
1978	2	26	2		
1979	3	27	1	2	28
1980	4	30	3	1.6	28.6
1981	5	32	2	2.16	32.16
1982	6	33	1	2.1	34.1
1983	7	36	3	1.66	34.66
1984	8	40	4	2.19	38.19
1985	9	41	1	2.92	42.92
1986	10	44	3	2.15	43.15
1987	11			2.49	46.49

解：由资料可以看出，燃料消耗量，除个别年份外，逐期增长量大体在 200 吨左右，即呈直线增长，因此可用一阶差分指数平滑模型来预测。我们取 $\alpha = 0.4$，初始值为差分序列首项值，计算结果列于表中。预测 1987 年燃料消耗量为

$$\hat{y}_{1987} = 2.49 + 44 = 46.49\,(\text{百吨}).$$

当时间序列呈现二次曲线增长时，可用二阶差分指数平滑模型来预测，计算公式如下:

$$\begin{aligned}
&\nabla y_t = y_t - y_{t-1} \\
&\nabla^2 y_t = \nabla y_t - \nabla y_{t-1} \\
&\nabla^2 \hat{y}_{t+1} = \alpha \nabla^2 y_t + (1-\alpha)\nabla^2 \hat{y}_t \\
&\hat{y}_{t+1} = \nabla^2 \hat{y}_{t+1} + \nabla y_t + y_t
\end{aligned}$$

其中 ∇^2 表示二阶差分。

在前面我们已分析过，指数平滑值实际上是一种加权平均数。因此把序列中逐期增量的加权平均数 (指数平滑值) 加上当前值的实际数进行预测，比一次指数平滑法只用变量以往取值的加权平均数作为下一期的预测更合理。从而使预测值始终围绕实际值上下波动，从根本上解决了在有直线增长趋势的情况下，用一次指数平滑法所得出的结果始终落后于实际值的问题。

差分方法和指数平滑法的联合运用，除了能克服一次指数平滑法的滞后偏差之外，对初始值的问题也有显著的改进。因为数据经过差分处理后，所产生的新序列基本上是平稳的。这时，初始值取新序列的第一期数据对于未来预测值不会有多大影响。其次，它拓展了指数平滑法的适用范围，使一些原来需要运用配合直线趋势模型处理的情况可用这种组合模型来取代。但是，对于指数平滑法存在的加权系数 α 的选择问题，以及只能逐期预测问题，差分指数平滑模型也没有改进。

3.2 自适应滤波法

自适应滤波法与移动平均法、指数平滑法一样，也是以时间序列的历史观测值进行某种加权平均来预测的，它要寻找一组“最佳”的权数，其办法是先用一组给定的权数来计算一个预测值，然后计算预测误差，再根据预测误差调整权数以减少误差。这样反复进行，直至找出一组“最佳”权数，使误差减少到最低限度。由于这种调整权数的过程与通讯工程中的传输噪声过滤过程极为接近，故称为自适应滤波法。自适应滤波法的基本预测公式为

$$\hat{y}_{t+1} = w_1 y_t + w_2 y_{t-1} + \cdots + w_N y_{t-N+1} = \sum_{i=1}^{N} w_i y_{t-i+1}$$

$\hat{y}_{t+1}$ 为第 $t+1$ 期的预测值，w_i 为第 $ti+1$ 期的观测值权数，y_{t-i+1} 为第 $t-i+1$ 期的观测值，N 为权数的个数。其调整权数的公式为

$$w_i' = w_i + 2k \cdot e_{i+1} y_{t-i+1}$$

式中，$i=1,2,\cdots t=N,N+1,\cdots,n,n$ 为序列数据的个数，w_i 为调整前的第 i 个权数，w' 为调整后的第 i 个权数，k 为学习常数，e_{i+1} 为第 $t+1$ 期的预测误差。上式表明: 调整后的一组权数应等于旧的一组权数加上误差调整项，这个调整项包括预测误差、原观测值和学习常数等三个因素。学习常数 k 的大小决定权数调整的速度。例如，下面有 10 期数据，试用自适应滤波法，以两个权数来求第 11 期的预测值。

表 3.2: 某时间序列数据表

时期 t	1	2	3	4	5	6	7	8	9	10
观测值 y_t	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

本例中 $N=2$。取初始权数 w_1 = 0.5，w_2 = 0.5，并设 k = 0.9。t 的取值由 N = 2 开始，当 t = 2 时:

- 按预测公式，求第 $t+1=3$ 期的预测值。$\hat{y}_{t+1}=\hat{y}_3=w_1y_2+w_2y_1=0.15$
- 计算预测误差。$e_{t+1}=e_3=y_3-\hat{y}_3=0.3-0.15=0.15$
- 根据调整权数公式：$w_1'=w_1+2ke_3y_2=0.554$，$w_2'=w_2+2ke_3y_1=0.527$

上面我们完成了第一次轮调整，把现有的新权数作为初始权数，重新开始 $t=2$ 的过程。这样反复进行下去，到预测误差 (指新一轮的预测总误差) 没有明显改进时，就认为获得了一组"最佳"权数，能实际用来预测第 11 期的数值。本例在调整过程中，可使得误差降为零，而权数达到稳定不变，最后得到的"最佳"权数为

$$w_1'=2.0,\quad w_2'=1.0$$

用"最佳"权数预测第 11 期的取值

$$\hat{y}_{11}=w_1'y_{10}+w_2'y_9=1.1$$

在实际应用中，权数调整计算工作量可能很大，必须借助于计算机才能实现。

在开始调整权数时，首先要确定权数个数 N 和学习常数 k 。一般说来，当时间序列的观测值呈季节变动时，N 应取季节性长度值。如序列以一年为周期进行季节变动时，若数据是月度的，则取 $N=12$ ，若季节是季度的，则取 $N=4$ 。如果时间序列无明显的周期变动，则可用自相关系数法来确定，即取 N 为最高自相关系数的滞后时期。

k 的取值一般可定为 $1/N$ ，也可以用不同的 k 值来进行计算，以确定一个能使 S 最小的 k 值。

初始权数的确定也很重要，如无其它依据，也可用 $1/N$ 作为初始权系数用，即

$$w_i=\frac{1}{N}(i=1,2,\cdots,N)$$

自适应滤波法有两个明显的优点: 一是技术比较简单，可根据预测意图来选择权数的个数和学习常数，以控制预测。也可以由计算机自动选定。二是它使用了全部历史数据来寻求最佳权系数，并随数据轨迹的变化而不断更新权数，从而不断改进预测。

由于自适应滤波法的预测模型简单，又可以在计算机上对数据进行处理，所以这种预测方法应用较为广泛。

3.3 趋势外推预测方法

趋势外推法是根据事物的历史和现时资料，寻求事物发展规律，从而推测出事物未来状况的一种比较常用的预测方法。利用趋势外推法进行预测，主要包括六个阶段: (a) 选择应预测的参数;(b) 收集必要的数据;(c) 利用数据拟合曲线;(d) 趋势外推;(e) 预测说明;(f) 研究预测结果在进行决策中应用的可能性。

趋势外推法常用的典型数学模型有: 指数曲线、修正指数曲线、生长曲线、包络曲线等。

3.3.1 指数曲线法

一般来说，技术的进步和生产的增长，在其未达饱和之前的新生时期是遵循指数曲线增长规律的，因此可以用指数曲线对发展中的事物进行预测。

指数曲线的数学模型为

$$y = y_0 e^{Kt}$$

其中系数 y_0 和 K 值由历史数据利用回归方法求得。对上式取对数可得

$$\ln y = \ln y_0 + Kt$$

令

$$Y = \ln y, \quad A = \ln y_0$$

则

$$Y = A + Kt$$

其中 A, K 可以用最小二乘法求得。

3.3.2 修正指数曲线法

利用指数曲线外推来进行预测时，存在着预测值随着时间的推移会无限增大的情况。这是不符合客观规律的。因为任何事物的发展都是有一定限度的。例如某种畅销产品，在其占有市场的初期是呈指数曲线增长的，但随着产品销售量的增加，产品总量接近于社会饱和量时。这时的预测模型应改用修正指数曲线。

$$\hat{y}_t = K + ab^t$$

在此数学模型中有三个参数 K,a 和 b 要用历史数据来确定。

修正指数曲线用于描述这样一类现象。

1. 初期增长迅速，随后增长率逐渐降低。
2. 当 $K > 0, \quad a < 0, \quad 0 < b < 1$ 时，$t \to \infty, ab^t \to 0$，即 $\hat{y}_t \to K$

当 K 值可预先确定时，采用最小二乘法确定模型中的参数。而当 K 值不能预先确定时，应采用三和法。把时间序列的 n 个观察值等分为三部分，每部分有 m 期，即 n = 3m 。

- $y_1, y_2, \cdots, y_m$；

- $y_{m+1}, y_{m+2}, \cdots, y_{2m}$;
- $y_{2m+1}, y_{2m+2}, \cdots, y_{3m}$

令每部分的趋势值之和等于相应的观察值之和，由此给出参数估计值。三和法步骤如下:

记观察值的各部分之和

$$S_1=\sum_{t=1}^{m} y_t, \quad S_2=\sum_{t=m+1}^{2m} y_t, \quad S_3=\sum_{t=2m+1}^{3m} y_t$$

且

$$\begin{cases} S_1=\sum\limits_{t=1}^{m}\hat{y}_t=\sum\limits_{t=1}^{m}\left(K+ab^t\right)=mK+ab\left(1+b+b^2+\cdots b^{m-1}\right) \\ S_2=\sum\limits_{t=m+1}^{2m}\hat{y}_t=\sum\limits_{t=m+1}^{2m}\left(K+ab^t\right)=mK+ab^{m+1}\left(1+b+b^2+\cdots b^{m-1}\right) \\ S_3=\sum\limits_{t=2m+1}^{3m}\hat{y}_t=\sum\limits_{t=2m+1}^{3m}\left(K+ab^t\right)=mK+ab^{2m+1}\left(1+b+b^2+\cdots b^{m-1}\right) \end{cases}$$

由于

$$\left(1+b+b^2+\cdots b^{m-1}\right)(b-1)=b^m-1$$

有

$$\begin{cases} S_1=mK+ab\frac{b^m-1}{b-1} \\ S_2=mK+ab^{m+1}\frac{b^m-1}{b-1} \\ S_3=mK+ab^{2m+1}\frac{b^m-1}{b-1} \end{cases}$$

解上式，可得

$$\begin{cases} b=\left(\frac{S_3-S_2}{S_2-S_1}\right)^{\frac{1}{m}} \\ a=\left(S_2-S_1\right)\frac{b-1}{b(b^m-1)^2} \\ K=\frac{1}{m}\left[S_1-\frac{ab(b^m-1)}{b-1}\right] \end{cases}$$

值得注意的是，并不是任何一组数据都可以用修正指数曲线拟合。采用前应对数据进行检验，检验方法是看给定数据的逐期增长量的比率是否接近某一常数 b 。即

$$\frac{y_{t+1}-y_t}{y_t-y_{t-1}}\approx b$$

练习 3.2 根据统计资料，某厂收音机连续 15 年的销售量如下表，试用修正指数曲线预测 1986 年的销售量。

表 3.3: Add caption

时间 (年)	1969	1970	1971	1972	1973	1974	1975	1976
销售量 (万部)	42.1	47.5	52.7	57.7	62.5	67.1	71.5	75.7
时间 (年)	1977	1978	1979	1980	1981	1982	1983	
销售量 (万部)	79.8	83.7	87.5	91.1	94.6	97.9	101.1	

解：经计算可知

$$\frac{y_{t+1}-y_t}{y_t-y_{t-1}} \in [0.9429, 0.9762]$$

可以认定这组数据可以采用修正指数曲线拟合。现将以上 15 个数据分为三部分，每部分 5 个数据，即 n = 15 ，m = 5 ，并以 1969 年作为开始年份 t = 1，经过计算，有：

$$S_1 = 262.5, \quad S_2 = 377.8, \quad S_3 = 472.2$$

于是可以计算出

$$b = 0.9608, \quad a = -143.2063, \quad K = 179.7162$$

故修正指数曲线的数学模型为

$$y = 179.7162 - 143.2063 \times 0.9608^t$$

预测 1986 年的产量时，t =19861969+1=18。所以

$$y_{1986} = 179.7162 - 143.2063 \times 0.9608^{18} = 110$$

3.3.3 Logistic 曲线 (生长曲线)

生物的生长过程经历发生、发展到成熟三个阶段，在三个阶段生物的生长速度是不一样的，例如南瓜的重量增长速度，在第一阶段增长的较慢，在发展时期则突然加快，而到了成熟期又趋减慢，形成一条 S 形曲线，这就是有名的 Logistic 曲线 (生长曲线)，很多事物，如技术和产品发展进程都有类似的发展过程，因此 Logistic 曲线在预测中有相当广泛的应用。

Logistic 曲线的一般数学模型是

$$\frac{dy}{dt} = ry\left(1 - \frac{y}{L}\right)$$

式中 y 为预测值，L 为 y 的极限值，r 为增长率常数，$r > 0$。解此微分方程得

$$y = \frac{L}{1 + ce^{-rt}}$$

记 Logistic 曲线的一般形式为

$$y_t = \frac{1}{K + ab^t}, \quad K > 0, \quad a > 0, \quad 0 < b \neq 1$$

检验能否使用 Logistic 曲线的方法，是看给定数据倒数的逐期增长量的比率是否接近某一常数 b 。即

$$\frac{1/y_{t+1} - 1/y_t}{1/y_t - 1/y_{t-1}} \approx b$$

其中

$$\begin{cases} b = \left(\frac{S_3 - S_2}{S_2 - S_1}\right)^{\frac{1}{m}} \\ a = (S_2 - S_1)\frac{b-1}{b(b^m-1)^2} \\ K = \frac{1}{m}\left[S_1 - \frac{ab(b^m-1)}{b-1}\right] \end{cases}$$

3.4 平稳时间序列

一个时间序列，如果均值没有系统的变化（无趋势）、方差没有系统变化，且严格消除了周期性变化，就称之是**平稳的**。本部分，我们指的平稳是指宽平稳，其特性是序列的统计特性不随时间的平移而变化，即均值和协方差不随时间的平移而变化。

3.4.1 平稳随机序列 (平稳时间序列)

定义 3.1. 均值函数

给定随机过程 $\{X_t, t \in T\}$。固定 t, X_t 是一个随机变量，设其均值为 μ_t，当 t 变动时，此均值是 t 的函数，记为

$$\mu_t = E(X_t)$$

称为随机过程的**均值函数**。♣

固定 t，设 X_t 的方差为 σ_t^2。当 t 变动时，这个方差也是 t 的函数，记为

$$\sigma_t^2 = Var(X_t) = E[(X_t - \mu_t)^2]$$

称为随机过程的**方差函数**。方差函数的平方根 σ_t 称为随机过程的**标准差函数**，它表示随机过程 X_t 对于均值函数 μ_t 的偏离程度。

定义 3.2. 均值函数

给定随机过程 $\{X_t, t \in T\}$。固定 t, X_t 是一个随机变量，设其均值为 μ_t，当 t 变动时，此均值是 t 的函数，记为

$$\mu_t = E(X_t)$$

对随机过程 $\{X_t, t \in T\}$，取定 $t, s \in T$，定义其**自协方差函数**为

$$\gamma_{t,s} = Cov(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)]$$

♣

为刻画 $\{X_t, t \in T\}$ 在时刻 t 与 s 之间的相关性，还可将 $\gamma_{t,s}$ 标准化，即定义自相关函数

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}}\sqrt{\gamma_{s,s}}} = \frac{\gamma_{t,s}}{\rho_t \rho_s}$$

因此，自相关函数 $\rho_{t,s}$ 是标准化自协方差函数。

定义 3.3. 平稳时间序列

设随机序列 $\{X_t, t = 0, \pm1, \pm2, \cdots\}$ 满足

- $E(X_t) = \mu =$ 常数
- $\gamma_{t+k,t} = \gamma_k (k = 0, \pm1, \pm2, \cdots)$ 与 t 无关

则称 X_t 为平稳随机序列 (平稳时间序列)，简称**平稳序列**。♣

定义 3.4. 平稳白噪声序列

设平稳序列 $\{\varepsilon_t, t=0,\pm1,\pm2,\cdots\}$ 的自协方差函数 γ_k 是

$$\gamma_k=\sigma^2\delta_{k,0}=\begin{cases}0, & k\neq 0\\ \sigma^2, & k=0\end{cases}$$

其中 $\delta_{k,0}$ 当 $k=0$ 时为 1，当 $k\neq 1$ 时为 0，则称该序列为**平稳白噪声序列**。♣

平稳白噪声序列的方差是常数 σ^2 ，因为 $\gamma_k=0(k\neq 0)$，则 ε_t 的任意两个不同时点之间是不相关的。平稳白噪声序列是一种最基本的平稳序列。

既然平稳白噪声序列各种统计参数都不变，那么我们如何分辨一列数据是否是白噪声还是真正的时间序列？这里我们用到时间序列的另一个特性：前后相关性，白噪声是没有前后相关性的，因此我们借此来进行检验：即 Daniel 检验。Daniel 检验方法建立在 Spearman 相关系数的基础上。

Spearman 相关系数是一种秩相关系数。先讲样本的秩的概念。设 $x_1,x_2,\cdots,x_n$ 是从一元总体抽取的容量为 n 的样本，其顺序统计量是 $x_{(1)},x_{(2)},\cdots,x_{(n)}$。若 $x_i=x_{(k)}$，则称 k 是 x_i 在样本中的**秩**，记作 R_i ，对每一个 $i=1,2,\cdots,n$ ，称 R_i 是第 i 个秩统计量。$R_1,R_2,\cdots,R_n$ 总称为**秩统计量**。例如，对样本数据

$$-0.8,\quad -3.1,\quad 1.1,\quad -5.2,\quad 4.2$$

顺序统计量是

$$-5.2,\quad -3.1,\quad -0.8,\quad 1.1,\quad 4.2$$

而秩统计量是

$$3,\quad 2,\quad 4,\quad 1,\quad 5$$

你可认为：秩就是“秩序”，就是原来样本数据在顺序统计量中的排位数，例如-0.8 在顺序排位中排名第 3，因此原数据的秩统计量为 3，以此类推。对于二维总体 (X,Y) 的样本观测数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，可得各分量 X,Y 的一元样本数据 $x_1,x_2,\cdots,x_n$ 与 $y_1,y_2,\cdots,y_n$ 。设 $x_1,x_2,\cdots,x_n$ 的秩统计量是

$$R_1,\quad R_2,\quad \cdots,\quad R_n$$

$y_1,y_2,\cdots,y_n$ 的秩统计量是

$$S_1,\quad S_2,\quad \cdots,\quad S_n$$

当 X,Y 联系比较紧密时，这两组秩统计量联系也是紧密的。Spearman 相关系数定义为这两组秩统计量的相关系数，即 Spearman 相关系数是

$$q_{xy}=\frac{\sum\limits_{i=1}^{n}\left(R_i-\bar{R}\right)\left(S_i-\bar{S}\right)}{\sqrt{\sum\limits_{i=1}^{n}\left(R_i-\bar{R}\right)^2}\sqrt{\sum\limits_{i=1}^{n}\left(S_i-\bar{S}\right)^2}}$$

其中 $\bar{R}=\frac{1}{n}\sum\limits_{i=1}^{n}R_i, \bar{S}=\frac{1}{n}\sum\limits_{i=1}^{n}S_i$，经过运算，可证

$$q_{xy}=1-\frac{6}{n\left(n^2-1\right)}\sum_{i=1}^{n}d_i^2$$

其中 $d_i=R_iS_i,\quad i=1,2,\cdots,n$。

对于 Spearman 相关系数，作假设检验

$$H_0:\rho_{XY}=0,\quad H_1:\rho_{XY}\neq 0$$

可以证明，当 (X,Y) 是二元正态总体，且 H_0 成立时，统计量

$$t=\frac{q_{xy}\sqrt{n-2}}{\sqrt{1-q_{xy}^2}}$$

服从自由度为 $n2$ 的 t 分布 $t(n2)$。

对于给定的显著水平 α，通过 t 分布表可查到统计量 t 的临界值 $t_{\alpha/2}(n2)$ ，当 $t\leqslant t_{\alpha/2}(n2)$ 时，接受 H_0; 当 $t>t_{\alpha/2}(n2)$ 时，拒绝 H_0。

对于时间序列的样本 $X_1,X_2,\cdots,X_n$，记 X_t 的秩为 $R_t=R(X_t)$，考虑变量对 $(t,R_t),t=1,2,\cdots,n$ 的 Spearman 相关系数 q_s，有

$$q_s=1-\frac{6}{n\left(n^2-1\right)}\sum_{i=1}^{n}\left(t-R_t\right)^2$$

构造统计量

$$T=\frac{q_s\sqrt{n-2}}{\sqrt{1-q_s^2}}$$

作下列假设检验:

$$H_0:\text{序列}X_t\text{平稳},\quad H_1:\text{序列}X_t\text{非平稳 (存在上升或下降趋势)}$$

Daniel 检验方法: 对于显著水平 α, 由时间序列 X_t 计算 $(t,R_t),t=1,2,\cdots,n$ 的 Spearman 秩相关系数 q_s，若 $|T|>t_{\alpha/2}(n2)$，则拒绝 H_0，认为序列非平稳。且当 $q_s>0$ 时，认为序列有上升趋势;$q_s<0$ 时，认为序列有下降趋势。又当 $|T|\leqslant t_{\alpha/2}(n2)$ 时，接受 H_0 ，可以认为 X_t 是平稳序列。

3.5 平稳序列自协方差函数与自相关函数的估计

对平稳序列 X_t 的样本 $X_1,X_2,\cdots,X_n$，可以用样本均值估计随机序列的均值

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_t=\bar{X}$$

γ_k 通常有如下两种估计

$$\hat{\gamma}_k=\frac{1}{n}\sum_{t=1}^{n-k}\left(X_{t+k}-\bar{X}\right)\left(X_t-\bar{X}\right),0\leq k\leq K$$

且 $\hat{\gamma}_{-k}=\hat{\gamma}_k$，或者

$$\hat{\gamma}_k=\frac{1}{n-k}\sum_{t=1}^{n-k}\left(X_{t+k}-\bar{X}\right)\left(X_t-\bar{X}\right),0\leq k\leq K$$

ρ_k 的估计为

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, 0 \le k \le K$$

第四章 ARMA 时间序列模型与预测

内容提要

- ❏ 时间序列分析流程
- ❏ ARMA 时间序列
- ❏ ARMA 序列的相关特性

自回归滑动平均模型（Autoregressive moving average model，简称：ARMA 模型）。是研究时间序列的重要方法，由自回归模型（简称 AR 模型）与移动平均模型（简称 MA 模型）为基础“混合”构成。在市场研究中常用于长期追踪资料的研究，如：Panel 研究中，用于消费行为模式变迁研究；在零售研究中，用于具有季节变动特征的销售量、市场规模的预测等。

4.1 时间序列分析流程

时间序列常以其复杂的数学原理著称，一些复合时间序列模型更是用到了大量专业的统计知识。本部分我们将以一般的 ARMA 模型分析过程为例，探究时间序列的一些直观特点与主要流程。

时间序列的分析过程主要包括以下问题：

1. 分析的数据是否有价值？是否为白噪声序列？
2. 分析的时间序列是否为平稳时间序列？如果不是平稳时间序列该怎么办？
3. 当前的观测值受到之前几期数据的影响？分别受到 AR、MA 模型的几期影响？
4. 如何检验时间序列模型的好坏？

下面，我们将分别从上述几个方面进行探究。

4.1.1 是否是白噪声？

并不是所有随时间变化的数据都是值得分析的时间序列，看起来规律的点也可能是毫无意义的“白噪声”，

定义 4.1. 白噪声特点

白噪声序列通常有三个特点：

1. 均值为 0；
2. 方差存在且为常数；
3. 数据前后不相关.

♣

在进行时间序列模型分析之前，我们首先要判断这列点是否是一组“白噪声”序列。

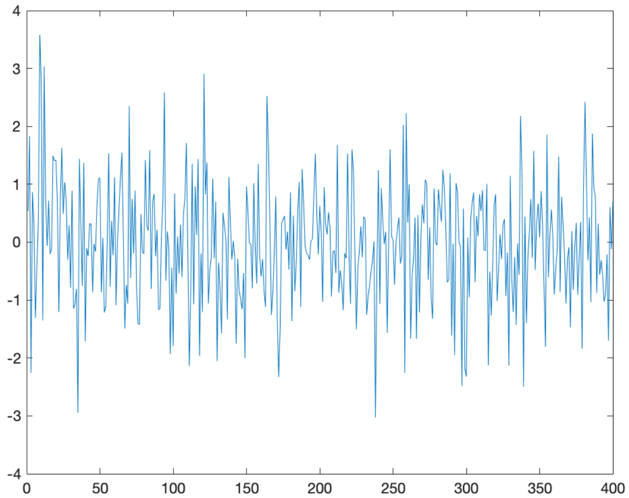

图 4.1: “白噪声”or “时间序列”?

仅通过分析关系图我们没办法分别该序列是否为“白噪声”，因此需要进行检验分析：

定理 4.1. 区别白噪声与平稳时间序列的方法

平稳时间序列通常有三个特点：

1. 均值为固定常数；
2. 方差存在且为常数；
3. 协方差只与时间间隔有关.

通常可以判断一组数值的自相关系数（协方差）来断定该组数据是平稳时间序列还是白噪声。

什么是**协方差**？直观理解为：协方差即为自变量 x 与因变量 y 的协同性，若 x，y 同大于期望值，则协方差为正；若 x，y 分别处在期望值两侧，则协方差为负。

4.1.2 是否是平稳时间序列?

时间序列包括平稳与非平稳时间序列，然而在现实分析中我们常常只考虑平稳时间序列模型。如何处理那些非平稳时间序列？通过**差分**变换。

什么是差分？假如我们有一组数据：

$$1,\quad 2,\quad 4,\quad 7,\quad 11,\quad 16,\quad 22,\quad 29$$

这组数据可能看起来毫无规律，但我们做差分变换：用第 $i+1$ 个值减去第 i 个值，可得：

$$1,\quad 2,\quad 3,\quad 4,\quad 5,\quad 6,\quad 7$$

规律则十分明显，我们对上述数据再进行一次差分即可得到一组全为 1 的值，因此上述数据经过两次差分后可得平稳序列。ARMA 模型要求我们只考虑平稳时间序列，对非平稳时间序列（非白噪声），我们需要进行差分处理。

4.1.3 当前数据受之前几期数据的影响?

由于时间序列前后具有较强的相关性，该部分我们需要判断当前数据究竟受前面多少期观测值的影响。通常有两种方法：1. 比较直观的 ACF、PACF 图判断；2. 较为精确的单位根检验判断。

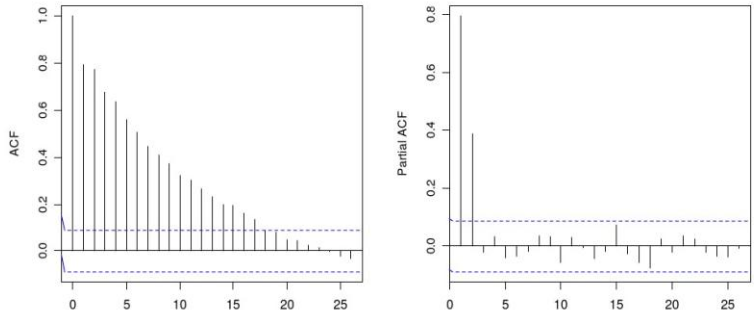

图 4.2: 拖尾与截尾

如果数据受到前面 n 期数据影响的效果逐渐递减，如上图 ACF 所示，我们称为“拖尾”；若数据受前面 n 期影响出现断崖式下跌，我们称为“截尾”。对于常见的截尾与拖尾出现情况，我们将其化为不同模型：

表 4.1: 由（偏）自相关图确定 ARMA 模型

ACF	PACF	选用模型
截尾	拖尾	MA
拖尾	截尾	AR
拖尾	拖尾	ARMA

第一个处在两倍误差范围内的观测值前一期作为 ARMA（p,q）内 p 与 q 的值。如果 ACF 与 PACF 同时出现了截尾该选用什么模型？直观上来说，我们之所以按照上表进行划分，依据是拖尾更符合“相关性”的概念，因此只有拖尾出现在某个关系图中，该关系图所代表的模型才具有更好的拟合效果；若 ACF 与 PACF 都同时出现了截尾现象，则说明 AR 与 MA 模型都不适用于该种情况，我们需要额外考虑其他方法。

4.2 ARMA 时间序列

下面介绍一种重要的平稳时间序列—ARMA 时间序列。ARMA 时间序列分为三种类型:

- AR 模型，即自回归序列 (Auto Regressive Model);
- MA 序列，即滑动平均序列 (Moving Average Model);
- ARMA 序列，即自回归滑动平均序列 (Auto Regressive Moving Average Model)。

可证 ARMA 时间序列具有遍历性，因此可以通过它的一个样本估计自协方差函数及自相关函数。

4.2.1 AR(p) 序列

设 $\{X_t, t = 0, \pm1, \pm2, \cdots\}$ 是零均值平稳序列，满足下列模型:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_n X_{t-p} + \varepsilon_t$$

其中 ε_t 是零均值、方差是 σ_ε^2 的平稳白噪声。则称 X_t 是阶数为 p 的自回归序列，简记为 AR(p) 序列，而

$$\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_p)^T$$

称为自回归参数向量，其分量 $\varphi_j, j = 1, 2, \cdots, p$ 称为**自回归系数**。

引进后移算子对描述式比较方便。算子 B 定义如下:

$$BX_t \equiv X_{t-1}, \quad B^k X_t \equiv X_{t-k}$$

记算子多项式

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$$

则 AR(p) 可以改写为

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

4.2.2 MA(q) 序列

设 $\{X_t, t = 0, \pm1, \pm2, \cdots\}$ 是零均值平稳序列，满足下列模型:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q},$$

其中 ε_t 是零均值、方差是 σ_ε^2 的平稳白噪声。则称 X_t 是阶数为 q 的滑动平均序列，简记为 MA(q) 序列，而

$$\theta = (\theta_1, \theta_2, \cdots, \theta_q)^T$$

称为滑动平均参数向量，其分量 $\theta_j, j = 1, 2, \cdots, q$ 称为**滑动平均系数**。

对于线性后移算子 B ，有

$$B\varepsilon_t \equiv \varepsilon_{t-1}, \quad B^k \varepsilon_t \equiv \varepsilon_{t-k};$$

再引进算子多项式

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

则 MA(q) 可以改写为

$$X_1 = \theta(B) \varepsilon_1$$

4.2.3 ARMA(p, q) 序列

设 $\{X_t, t = 0, \pm1, \pm2, \cdots\}$ 是零均值平稳序列，满足下列模型:

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

其中 ε_t 是零均值、方差是 σ_ε^2 的平稳白噪声，则称 X_t 是阶数为 p, q 的**自回归滑动平均序列**，简记为 ARMA(p, q) 。当 $q = 0$ 时，它是 AR(p) 序列; 当 $p = 0$ 时，它为 MA(q) 序列。

应用算子多项式 $\varphi(B),\theta(B)$，可以将上式写为

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

对于一般的平稳序列 $\{X_t,t=0,\pm1,\pm2,\cdots\}$，设其均值 $E(X_t)=\mu$，满足下列模型:

$$(X_t-\mu)-\varphi_1(X_{t-1}-\mu)-\cdots-\varphi_p(X_{t-p}-\mu)=\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-9}$$

其中 ε_t 是零均值、方差是 σ_ε^2 的平稳白噪声，利用后移算子 $\varphi(B),\theta(B)$，可表为:

$$\varphi(B)(X_t-\mu)=\theta(B)\varepsilon_t$$

关于算子多项式 $\varphi(B),\theta(B)$ 通常还要作下列假定:

- $\varphi(B)$ 和 $\theta(B)$ 无公共因子，又 $\varphi_p\neq0,\theta_q\neq0$
- $\varphi(B)=0$ 的根全在单位圆外，这一条件称为模型的平稳性条件;
- $\theta(B)=0$ 的根全在单位圆外，这一条件称为模型的可逆性条件。

4.3 ARMA 序列的相关特性

我们要阐述 AR，MA，ARMA 序列的相关特性，这些特性对于序列的建模起着重要的作用。相关特性包括相关函数，以及下面将要阐述的偏相关函数。

4.3.1 MA(q) 序列的自相关函数

设 $\{X_t,t=0,\pm1,\pm2,\cdots\}$ 是 MA(q) 序列

$$X_t=\theta(B)\varepsilon_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\cdots-\theta_q\varepsilon_{t-q}$$

其中 ε_t 是零均值、方差是 σ_ε^2 的平稳白噪声，则 X_t 是零均值平稳序列，其自协方差函数为

$$\gamma_k=\begin{cases}\sigma_\varepsilon^2\left(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2\right), & k=0\\ \sigma_\varepsilon^2\left(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q\right), & 1\leq k\leq q\\ 0, & k>q\end{cases}$$

自相关函数为

$$\rho_k=\frac{\gamma_k}{\gamma_0}=\begin{cases}1, & k=0\\ \frac{-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q}{1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2}, & 1\leq k\leq q\\ 0, & k>q\end{cases}$$

注意到，当 $k>q$ 时

$$\gamma_k=0,\quad\rho_k=0$$

这表明对于 MA(q) 序列，当 $ts>q$ 时，X_t 与 X_s 不相关，这一性质称为自协方差函数与自相关函数的 **q 步截尾性**。

4.3.2 AR(p) 序列的自相关函数

设 X_t 是 AR(p) 序列

$$X_t=\varphi_1X_{t-1}+\varphi_2X_{t-2}+\cdots+\varphi_pX_{t-p}+\varepsilon_t$$

$\mathrm{Var}(\varepsilon_t)=\sigma_\varepsilon^2$，用 $X_{t-k}(k>0)$ 乘上式两端，求均值得

$$E\left(X_tX_{t-k}\right)=\varphi_1E\left(X_{t-1}X_{t-k}\right)+\varphi_2E\left(X_{t-2}X_{t-k}\right)+\cdots+\varphi_pE\left(X_{t-p}X_{t-k}\right)+E\left(\varepsilon_tX_{t-k}\right)$$

且 AR 模型有一个极为重要的公式

$$E(X_{t-k}\varepsilon_t)=0,k\leqslant 1$$

故得

$$\gamma_k=\varphi_1\gamma_{k-1}+\varphi_2\gamma_{k-2}+\cdots+\varphi_p\gamma_{k-p},\quad k>0,$$

即

$$\varphi(B)\gamma_k=0,\quad k>0$$

这是自协方差函数 γ_k 满足的差分方程。自相关函数 ρ_k 也满足同样的差分方程

$$\varphi(B)\rho_k=0,\quad k>0$$

4.3.3 ARMA(p,q) 序列的自相关函数

设 $\{X_t,t=0,\pm1,\pm2,\cdots\}$ 是 ARMA(p,q) 序列

$$X_t-\varphi_1X_{t-1}-\cdots-\varphi_pX_{t-p}=\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q}$$

则其自协方差函数满足

$$\varphi(B)\gamma_k=\begin{cases}0, & k>q\\ -\theta_q\sigma_\varepsilon^2, & k=q\end{cases}$$

取 $k=q+1,q+2,\cdots,q+p$，得到方程组

$$\begin{cases}\gamma_{q+1}=\varphi_1\gamma_q+\varphi_2\gamma_{q-1}+\cdots+\varphi_p\gamma_{q+1-p}\\ \gamma_{q+2}=\varphi_1\gamma_{q+1}+\varphi_2\gamma_q+\cdots+\varphi_p\gamma_{q+2-p}\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \gamma_{q+p}=\varphi_1\gamma_{q+p-1}+\varphi_2\gamma_{q+p-2}+\cdots+\varphi_p\gamma_q\end{cases}$$

将其写成矩阵形式有

$$\begin{bmatrix}\gamma_{q+1}\\ \gamma_{q+2}\\ \vdots\\ \gamma_{q+p}\end{bmatrix}=\begin{bmatrix}\gamma_q & \gamma_{q-1} & \cdots & \gamma_{q+1-p}\\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q+2-p}\\ \vdots & \vdots & & \vdots\\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q\end{bmatrix}\begin{bmatrix}\varphi_1\\ \varphi_2\\ \vdots\\ \varphi_p\end{bmatrix}$$

解此方程组，可得自回归系数向量 $(\varphi_1,\varphi_2,\cdots,\varphi_p)^T$.

其次，令 $Y_t=\varphi(B)X_t=X_t-\varphi_1X_{t-1}-\cdots-\varphi_pX_{t-p}$，则 $\{Y_t,t=0,\pm1,\pm2,\cdots\}$

是参数为 $(\theta_1,\theta_2,\cdots,\theta_q)^T$ 的 MA(q) 序列，$\{Y_t,t=0,\pm1,\pm2,\cdots\}$ 的自协方差函数

$$\gamma_k^{(Y)}=E\left(Y_{t+k}Y_t\right)=E\left[\left(-\sum_{i=0}^{p}\varphi_iX_{t+k-i}\right)\left(-\sum_{j=0}^{p}\varphi_jX_{t-j}\right)\right]=\sum_{i,j=0}^{p}\varphi_i\varphi_j\gamma_{k+j-i},\quad(\varphi_0=-1)$$

即

$$\sum_{i,j=0}^{p}\varphi_i\varphi_j\gamma_{k+j-i}=\begin{cases}\sigma_\varepsilon^2\left(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2\right), & k=0\\ \sigma_\varepsilon^2\left(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q\right), & 1\le k\le q\end{cases}$$

解此方程组，得滑动平均系数向量 $(\theta_1,\theta_2,\cdots,\theta_q)^T$ 及 σ_ε^2.

可以证明，ARMA(p,q) 序列的自协方差函数 (自相关函数) 为拖尾的。

4.3.3.1 偏相关函数及 AR(p) 序列偏相关函数的截尾性

设 $\{X_t,t=0,\pm1,\pm2,\cdots\}$ 是零均值平稳序列从时间序列，从预报的角度引出偏相关函数的定义。如果已知 $\{X_{t-1},X_{t-2},\cdots,X_{t-k}\}$ 的值，要求对 X_t 作出预报。此时，可以考虑由 $\{X_{t-1},X_{t-2},\cdots,X_{t-k}\}$ 对 X_t 的线性最小均方估计，即选择系数 $\varphi_{k,1},\varphi_{k,2},\cdots,\varphi_{k,k}$，使得

$$\delta=E\left[\left(X_t-\sum_{j=1}^{k}\varphi_{k,j}X_{t-j}\right)^2\right]=\min$$

将 δ 展开，得

$$\delta=\gamma_0-2\sum_{j=1}^{k}\varphi_{k,j}\gamma_j+\sum_{j=1}^{k}\sum_{i=1}^{k}\varphi_{k,j}\varphi_{k,i}\gamma_{j-i}$$

令 $\frac{\partial\delta}{\partial\varphi_{k,j}}=0,\quad j=1,2,\cdots,k$，得

$$-\gamma_j+\sum_{i=1}^{k}\varphi_{k,i}\gamma_{j-i}=0,\quad j=1,2,\cdots,k$$

两端同除 γ_0 并写成矩阵形式，可知 $\varphi_{k,j}$ 应满足下列线性方程组

$$\begin{bmatrix}1 & \rho_1 & \cdots & \rho_{k-1}\\ \rho_1 & 1 & \cdots & \rho_{k-2}\\ \vdots & \vdots & & \vdots\\ \rho_{k-1} & \rho_{k-2} & \cdots & 1\end{bmatrix}\begin{bmatrix}\varphi_{k,1}\\ \varphi_{k,2}\\ \vdots\\ \varphi_{k,k}\end{bmatrix}=\begin{bmatrix}\rho_1\\ \rho_2\\ \vdots\\ \rho_k\end{bmatrix}$$

该式称为 Yule-Walker 方程。

我们称 $\varphi_{k,k},k=1,2,\cdots$ 为 X_t 的偏相关函数。该式系数矩阵是 Toeplitz 矩阵，这一线性方程组的解 $\varphi_{k,1},\varphi_{k,2},\cdots,\varphi_{k,k}$ 有下列递推公式

$$\begin{cases}\varphi_{1,1}=\rho_1\\ \varphi_{k+1,k+1}=\left(\rho_{k+1}-\sum_{j=1}^{k}\rho_{k+1-j}\varphi_{k,j}\right)\left(1-\sum_{j=1}^{k}\rho_j\varphi_{k,j}\right)^{-1}\\ \varphi_{k+1,j}=\varphi_{k,j}-\varphi_{k+1,k+1}\varphi_{k,k+1-j},\quad j=1,2,\cdots,k\end{cases}$$

AR(p) 序列的偏相关函数 $\{\varphi_{k,k}, k=1,2,\cdots\}$ 是 p 步截尾的，且有

$$\begin{cases} \varphi_{k,k}=\varphi_k, & 1\le k\le p \\ \varphi_{k,k}=0, & k>p \end{cases}$$

对于 MA(q) 和 ARMA(p,q) 序列，其偏相关函数 $\{\varphi_{k,k}, k=1,2,\cdots\}$ 具有拖尾性。

4.3.4 ARMA 时间序列的建模与预报

在实际问题中，若考察的时间序列是 ARMA 序列

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

首先要进行模型的识别与定阶，即要判断是 AR(p)，MA(q)，ARMA(p, q) 模型的类别，并估计阶数 p, q 。其实，这都归结到模型的定阶问题。当模型定阶后，就要对模型参数 $\varphi=(\varphi_1,\varphi_2,\cdots,\varphi_p)^T$ 及 $\theta=(\theta_1,\theta_2,\cdots,\theta_q)^T$ 进行估计。定阶与参数估计完成后，还要对模型进行检验，即要检验 ε_t 是否为平稳白噪声。若检验获得通过，则 ARMA 时间序列的建模完成。作为时间序列建模之后的一个重要应用，我们还要讨论 ARMA 时间序列的预报。

4.4 ARIMA 时间序列估计与预测

20 世纪 70 年代初，Box 和 Jenkins 提出了 ARIMA(p,d,q) 模型 (差分自回归移动平均模型). 其中：AR 是自回归，p 为自回归项数；MA 为移动平均，q 为移动平均项数，d 为时间序列平稳时做的差分次数。

对时间序列处理前，须平稳化处理非平稳时间序列，然后，仅对因变量的滞后值以及随机误差项的现值和滞后值进行回归. 事实上，时间序列数据常常带有一定的时间趋势或季节特征，这些不满足 ARIMA 对平稳性的要求，因此需要进行差分处理. 对于平稳时间序列，可建立 ARIMA 模型。

$$X_t=\phi_1X_{t-1}+\cdots+\phi_pX_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q} \tag{4.1}$$

其中: 前半部为自回归部分，非负整数 p 为自回归阶数，$\phi_1,\phi_2,\cdots,\phi_p$ 为自回归系数; 后半部为滑动平均部分，非负整数 q 为滑动平均阶数，$\theta_1,\theta_2,\cdots,\theta_1$ 为滑动平均系数;

上文中，我们已经考虑过一般的 AR 模型：AR(1)：$y_t=\rho\cdot y_{t-1}+u_t$，更一般地，有：AR(p)：$y_t=\rho_1\cdot y_{t-1}+\rho_2\cdot y_{t-2}+\cdots+\rho_p\cdot y_{t-p}+u_t$，其自相关系数（ACF）为：

$$\rho_s=\frac{Cov(y_t,t_{t+s})}{Var[y_t]} \tag{4.2}$$

偏自相关系数（PACF）为：

$$\begin{cases} y_t=a_{11}y_{t-1}+u_t \\ y_t=a_{21}y_{t-1}+a_{22}y_{t-2}+u_t \\ \cdots \\ y_t=a_{k1}y_{t-1}+a_{k2}y_{t-2}+\cdots+a_{kk}y_{t-k}+u_t \end{cases} \tag{4.3}$$

可以理解为：PACF 为 $\{a_{11}, a_{22}, a_{33}, \cdots, a_{kk}\}$，相当于控制其他滞后项的影响后，得到的净相关系数。

4.4.1 山西省教育人口逆差的 ARIMA 拟合

下面我们对山西省 2019-2025 年教育人口逆差数进行 ARIMA 时间序列建模，体会 ARIMA 建模的整个过程。时间序列分析分为以下五步：

- ARIMA 模型要求序列满足平稳性，查看 ADF 检验结果，根据分析 t 值，分析其是否可以显著地拒绝序列不平稳的假设（p<0.05 或 0.01）
- 查看差分前后数据对比图，判断是否平稳（上下波动幅度不大），同时对时间序列进行偏（自相关分析），根据截尾情况估算其 p、q 值
- ARIMA 模型要求模型具备纯随机性，即模型残差为白噪声，查看模型检验表，根据 Q 统计量的 p 值（p 值大于 0.01 为白噪声，严格则需大于 0.05）对模型白噪声进行检验，也可以结合信息准则 AIC 和 BIC 值进行分析（越低越好），也可以通过模型残差 ACF/PACF 图进行分析
- 根据模型参数表，得出模型公式
- 结合时间序列分析图进行综合分析，得到向后预测的阶数结果

表 4.2: 教育人口流出模型 ADF 检验表

变量	差分阶数	t	p	lag	AIC	临界值		
						1%	5%	10%
教育人口流出	0	0.093	0.966	3	-5.382	-4.665	-3.367	-2.803
	1	-4.422	0.000**	0	24.519	-4.332	-3.233	-2.749
	2	-17.93	0.000***	3	-22.683	-5.354	-3.646	-2.901

上表为 ADF 检验结果，包括变量、差分阶数、T 检验结果、AIC 值等，用于检验时间序列是否平稳. 该序列检验的结果显示，基于总人数，在差分为 1 阶时，显著性 P 值为 0.001**，水平上呈显著性，拒绝原假设，该序列为平稳的时间序列.

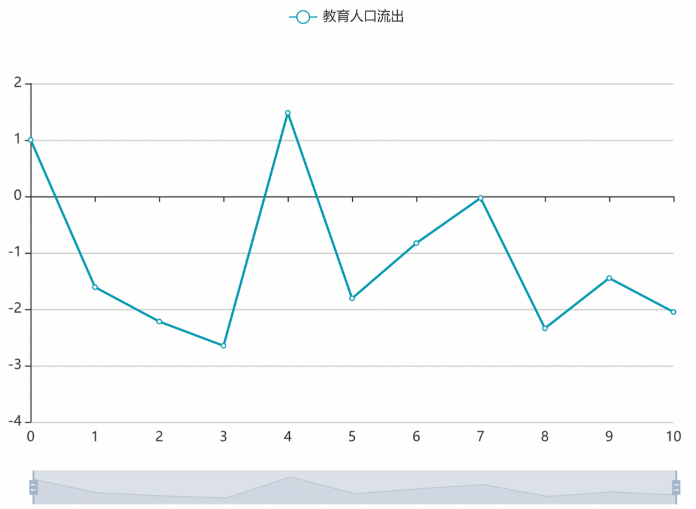

图 4.3: 最终差分序列图

上图展示了原始数据 1 阶差分后的时序图，下面我们分析 ACF 与 PACF. 自相关图 (ACF) 包括系数、置信上限和置信下限. 其中横轴代表延迟数目，纵轴代表自相关系数. 偏自相关图 (PACF) 包括系数、置信上限和置信下限.

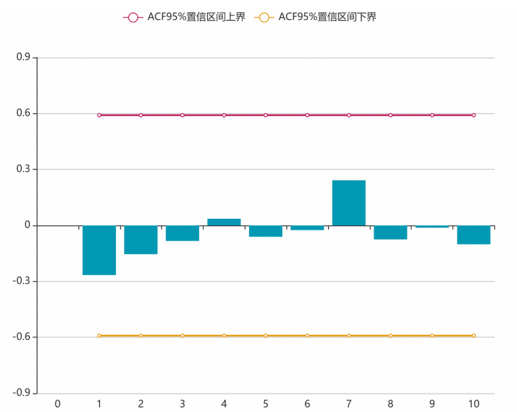

图 4.4: 数据自相关图 (ACF)

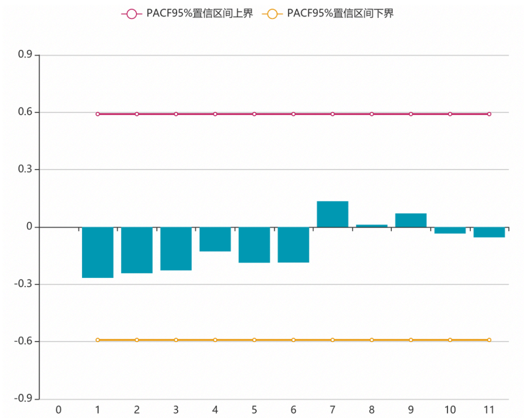

图 4.5: 数据偏自相关图 (PACF)

表 4.3 展示了本次模型检验结果，包括样本数、自由度、Q 统计量和信息准则模型的拟合优度. 值得注意的是，ARIMA 模型要求模型的残差不存在自相关性，即模型残差为白噪声，查看模型检验表，根据 Q 统计量的 p 值（p 值大于 0.1 为白噪声）对模型白噪声进行检验；根据信息准则 AIC 和 BIC 值用于多次分析模型对比（越低越好）；R^2 代表时间序列的拟合优度，越接近 1 越好.

表 **4.3:** ARIMA 模型 (0, 1, 1) 检验表

项	符号	值
	Df Residuals	9
样本数量	N	21
Q 统计量	Q6(p 值)	0.128(0.720)
信息准则	AIC	38.987
	BIC	40.181
拟合优度	R^2	0.915

基于 AIC、BIC 信息准则可知最优参数的模型结果为 ARIMA(0, 1, 1)，基于总人数，从 Q 统计量结果分析可得：Q_6 在水平上不呈现显著性，不能拒绝模型的残差为白噪声序列的假设，同时模型的拟合优度 R^2 为 0.915，拟合效果较好，模型基本满足要求.

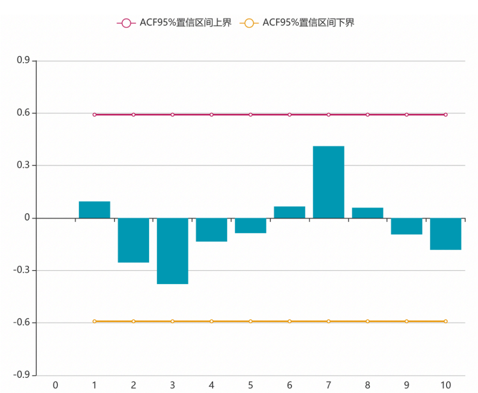

图 **4.6:** 残差自相关图 (ACF)

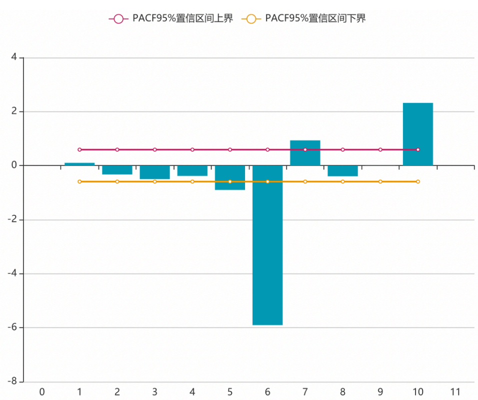

图 **4.7:** 残差偏自相关图 (PACF)

上述两幅图分别为残差自相关图 (ACF) 与残差偏自相关图 (PACF)，都包括系数、置信上限和置信下限；其中 ACF 图中横轴代表延迟数目，纵轴代表自相关系数；（ACF 中）若相关系数均在虚线内，AR 残差为白噪声序列，时间序列要求模型残差为白噪声序列；（PACF 中）若相关系数均在虚线内，MA 残差为白噪声序列，时间序列要求模型残差为白噪声序列； 模型结果为：ARIMA(0, 1, 1)，且基于 d 差分数据，模型公式如下：

$$y(t) = -1.133 = 0.998 * \varepsilon(t-1) \tag{4.4}$$

4.4.2 模型结果与分析

上面的例子只是为了说明时间序列的一般流程。事实上，上述 ARIMA 模型效果并不好，原因有三：

- ARIMA(0，1，1) 模型可直接用差分后的 MA（1）来等效替换；

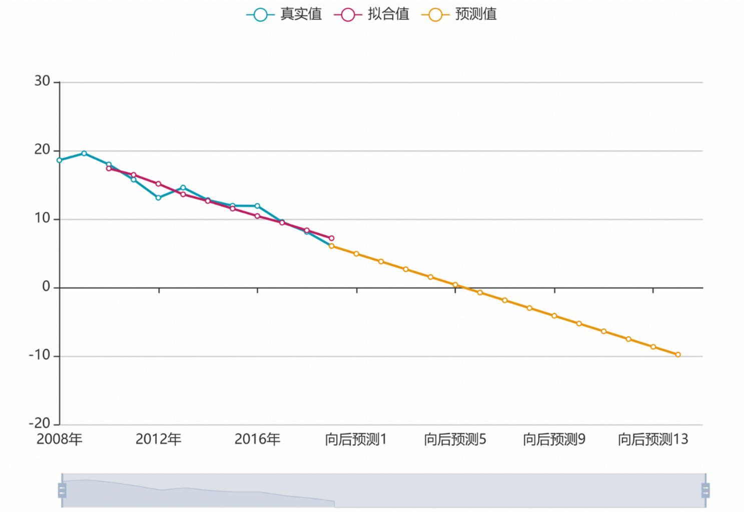

图 4.8: 2019-2033 年教育人口流失数量的时间序列预测图

- ACF、PACF 图并未表现出十分明显的拖尾/截尾现象，该问题并不十分适合时间序列模型；
- 从拟合图像来看，该问题可以用回归分析来拟合.