时间序列分析引言

程腾飞

广西大学

2023.09.25

目录

1. 时间序列简介

1.1 时间序列

- 时间序列是以观测值按照时间顺序排列而成的数列。
- 许多数据集都是以时间序列的形式表现的，比如工厂船运物品数量的月度序列、道路事故数量的周序列、日降雨量、化工过程每小时产量的观测值等。
- 相邻数据相互依赖是时间序列的内在特征，时间序列分析侧重的正是分析这种依赖性的技术。

1.1 时间序列

时间序列分析在许多不同的领域都有广泛的应用。以下是时间序列分析的一些常见应用领域：

- 经济与商业
- 气象学
- 地球物理
- 工程
- 社会科学
- ……

1.2 五个重要的应用领域

本书将在下面五个重要的应用领域介绍这些时间序列和动态模型的使用

- 通过时间序列的现在值与过去值来预测将来值。
- 确定惯性系统的转换函数，即确定一个动态输入-输出模型，这个模型可以表达任意给定的输入序列对系统输出的影响。
- 在转换函数模型中使用指示性输入变量，来表现和评定意外干扰事件对事件序列变化的影响。
- 检测几个相关的收益时间序列变量间的相互关系，以确定恰当的多元动态模型来体现这些变量随着时间变化的共同关系。
- 设计简单的控制机制，通过这个控制机制，系统输出值与期望目标值的背离很可能通过调整序列输入值而被抵消掉。

1.2 五个重要的应用领域

本书将在下面五个重要的应用领域介绍这些时间序列和动态模型的使用

- 通过时间序列的现在值与过去值来预测将来值。
- 确定惯性系统的转换函数，即确定一个动态输入-输出模型，这个模型可以表达任意给定的输入序列对系统输出的影响。
- 在转换函数模型中使用指示性输入变量，来表现和评定意外干扰事件对事件序列变化的影响。
- 检测几个相关的收益时间序列变量间的相互关系，以确定恰当的多元动态模型来体现这些变量随着时间变化的共同关系。
- 设计简单的控制机制，通过这个控制机制，系统输出值与期望目标值的背离很可能通过调整序列输入值而被抵消掉。

1.2 五个重要的应用领域

本书将在下面五个重要的应用领域介绍这些时间序列和动态模型的使用

- 通过时间序列的现在值与过去值来预测将来值。
- 确定惯性系统的转换函数，即确定一个动态输入-输出模型，这个模型可以表达任意给定的输入序列对系统输出的影响。
- 在转换函数模型中使用指示性输入变量，来表现和评定意外干扰事件对事件序列变化的影响。
- 检测几个相关的收益时间序列变量间的相互关系，以确定恰当的多元动态模型来体现这些变量随着时间变化的共同关系。
- 设计简单的控制机制，通过这个控制机制，系统输出值与期望目标值的背离很可能通过调整序列输入值而被抵消掉。

1.2 五个重要的应用领域

本书将在下面五个重要的应用领域介绍这些时间序列和动态模型的使用

- 通过时间序列的现在值与过去值来预测将来值。
- 确定惯性系统的转换函数，即确定一个动态输入-输出模型，这个模型可以表达任意给定的输入序列对系统输出的影响。
- 在转换函数模型中使用指示性输入变量，来表现和评定意外干扰事件对事件序列变化的影响。
- 检测几个相关的收益时间序列变量间的相互关系，以确定恰当的多元动态模型来体现这些变量随着时间变化的共同关系。
- 设计简单的控制机制，通过这个控制机制，系统输出值与期望目标值的背离很可能通过调整序列输入值而被抵消掉。

1.2 五个重要的应用领域

本书将在下面五个重要的应用领域介绍这些时间序列和动态模型的使用

- 通过时间序列的现在值与过去值来预测将来值。
- 确定惯性系统的转换函数，即确定一个动态输入-输出模型，这个模型可以表达任意给定的输入序列对系统输出的影响。
- 在转换函数模型中使用指示性输入变量，来表现和评定意外干扰事件对事件序列变化的影响。
- 检测几个相关的收益时间序列变量间的相互关系，以确定恰当的多元动态模型来体现这些变量随着时间变化的共同关系。
- 设计简单的控制机制，通过这个控制机制，系统输出值与期望目标值的背离很可能通过调整序列输入值而被抵消掉。

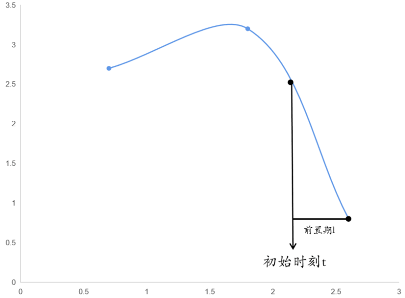
2. 五个重要的时间序列问题概念

2.1 时间序列预测

定义

利用一个时间序列在 t 时刻的观测值去预测它在 t+l 时刻的将来值。

预测通常需要提前一段时间提出，这段时间被称为**前置期**l（lag），即预测距离当前时间（初始时间）l 之后的状态。



2.1 时间序列预测

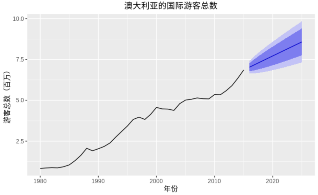
假设可以得到不相关且等时间间隔出现的预测值。比如，在销售预测问题中，可用当月 t 的销量 z_t 和之前几个月的销量 z_{t-1}，z_{t-2},z_{t-3},... 来预测前置期 l=1,2,3,...,12 的销量。
用 $z_t(l)$ 来表示将来时刻 t+l 的销量，即初始时刻为 t、前置期为 l 的销量。函数 $z_t(l)$ 被称为初始时刻为 t 的预测函数。它可以使真实值 z_{t+l} 和预测值 $z_t(l)$ 之间的偏差尽量小。

2.1 时间序列预测

预测区间的值表示的是预测中的不确定性。如果我们只做点预测，就无法得到这些预测的准确程度。

因此，除了要计算最优预测值之外，设定精确度也是很有必要的，比如，设定精确度以后可以计算基于预测值的决策风险。

预测区间给出了一定置信度下的置信区。这些置信限可以由一个特定的概率给出，比如 80% 或 90%。在下图的例子中，以颜色代表置信区，真实值落在深紫色区域的概率为 80%，落在浅紫色区域的概率是 90%。推导一个用来计算将来之落在两个特殊值之间的概率模型是可行的，这种模型被称为概率模型或随机模型。

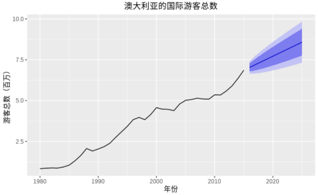


2.1 时间序列预测

预测区间的值表示的是预测中的不确定性。如果我们只做点预测，就无法得到这些预测的准确程度。

因此，除了要计算最优预测值之外，设定精确度也是很有必要的，比如，设定精确度以后可以计算基于预测值的决策风险。

预测区间给出了一定置信度下的置信区。这些置信限可以由一个特定的概率给出，比如 80% 或 90%。在下图的例子中，以颜色代表置信区，真实值落在深紫色区域的概率为 80%，落在浅紫色区域的概率是 90%。推导一个用来计算将来之落在两个特殊值之间的概率模型是可行的，这种模型被称为概率模型或随机模型。



2.2 转换函数估计

转换函数由输入时间变量 X_t 和输出时间变量 Y_t 构成。若能够确定两个时间变量 X_t 和 Y_t 之间的动态关系得到转换函数，两个序列的过去值就可能用于预测序列 Y_t

假设有一个已知的权重序列，该序列被称为脉冲响应函数 v_j(j=0,1,2,...)，该函数通过输入变量 X_t 和输出变量 Y_t 的动态线性关系 $Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + ...$（线性滤波模型）确定系统的转换函数。

2.2 转换函数估计

转换函数由输入时间变量 X_t 和输出时间变量 Y_t 构成。若能够确定两个时间变量 X_t 和 Y_t 之间的**动态关系**得到转换函数，两个序列的过去值就可能用于预测序列 Y_t
假设有一个已知的权重序列，该序列被称为**脉冲响应函数** v_j(j=0,1,2,...)，该函数通过输入变量 X_t 和输出变量 Y_t 的动态线性关系 $Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \ldots$(**线性滤波模型**）确定系统的转换函数。

2.3 异常干扰事件对系统的影响分析

众所周知，一些特殊的外部事件，即干扰事件，可能会对时间序列 z_t 造成影响，这些干扰事件可能是：

- 特殊的促销活动
- 罢工
- 经济政策变化
- 新的环境规则
- ……

在这种情况下，可以使用更复杂的转换函数模型来解释干扰事件对时间序列 z_t 的影响。

2.4 多元时间序列分析

将单个序列看作多变量或向量时间序列的元素，并且对几个序列进行联合分析，经常是一种信息更丰富而且更高效的分析方法。多元时间序列分析可以在许多领域中应用，如经济学、金融学、气象学、市场研究等。它可以帮助我们理解变量之间的相互作用和影响，提供更准确的预测和决策依据。

2.5 离散型控制系统

控制系统

对时间序列数据进行监控和调整，以确保数据的质量和稳定性。

在化学工业过程中，已经有各种形式的反馈和前馈调节被用于工程过程控制中。
当测量一个可观测但是不能改变的输入变量的波动时，对其他的控制变量做合理的补偿性改变是可行的，这被称为前馈控制。例如，在自动驾驶汽车中，根据预测的路况和交通状况提前调整车辆的速度和方向。另外，还可以利用与目标值的偏离或“错误信号”本身的输出特征来计算对控制变量合适的补偿性改变，这被称为反馈控制。例如，在温度控制中，通过不断测量温度并与设定值进行比较，来调整加热器的功率。

2.5 离散型控制系统

以大物实验中，用天平与砝码称量物体重量为例，解释前馈控制与反馈控制的区别

假设物体重量 100g（未知），放置砝码的重量为输入 X，天平的表现为输出 Y，预期结果为天平平衡.那么如何通过两种方法调整砝码重量，从而使天平平衡呢？

前馈控制方法：得到的转换函数为：X=100 时天平平衡。那么直接将 100g 砝码放在天平上即可。
反馈控制方法：得到的转换函数只能显示是否达到预期或超出预期。先放入 50g 砝码，转换函数显示没有达到预期结果；再放入 100g 砝码，共 150g，转换函数显示超出预期结果；再卸下 50g 砝码，共 100g，显示天平平衡。
两种方法中，转换函数的不同即代表着现实问题中输入数据特点的不同。

3. 时间序列数学模型

3.1 用于平稳与非平稳过程预测和控制的随机模型

在假设过程平衡时所建的均衡模型中，平稳过程的概率性质不会随时间变化而改变，并且有固定的均值与有限的方差。但是，实际领域中的许多时间序列经常是非平稳的，这些非平稳过程一般具有时变的均值与方差，即均值与方差会随时间的推移而发生变化。
平稳与非平稳的过程充分体现了许多时间序列的类型，它们提供了广阔的建模空间。

一些简单算子

滞后算子 B是一类简单的算子，定义为 $Bz_t = z_{t-1}$，以此类推有 $B^m z_t = z_{t-m}$。另一个重要的算子是向后差分算子 ∇，定义为 $\nabla z_t = z_t - z_{t-1}$，将其写成滞后算子 B 的形式为

$$\nabla z_t = z_t - z_{t-1} = (1 - B) z_t$$

3.1.1 平稳模型

1. 线性滤波模型
2. 自回归模型
3. 移动平均模型
4. 自回归移动平均模型

3.1.1.1 线性滤波模型

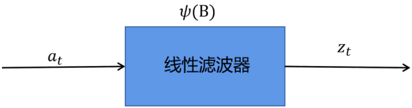
独立的随机变量 $a_t, a_{t-1}, a_{t-2}, \ldots$ 形成随机扰动序列，假设 a_t 经由线性滤波器转化为过程 z_t。
简单的线性滤波是随机扰动前期值得加权和，所以有

$$z_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \ldots = \mu + \Psi(B) a_t \tag{1}$$

通常，μ 是一个确定过程的“水平”参数，并且

$$\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \ldots$$

是把 a_t 转化成 z_t 的线性算子，称为滤波的转换函数。其关系如下图：



图：时间序列的线性滤波器输出图

3.1.1.2 自回归模型

在自回归模型中，过程的现值会被表示成有限个过程过去值的线性加权和，再加上一个随机扰动 a_t 的形式。
设时点对应的过程取值为 $z_t, z_{t-1}, z_{t-2}, \ldots$，并且令 $\tilde{z}_t = z_t - \mu$ 表示与 μ 的偏离序列。得到

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p} + a_t$$

此模型称为p 阶自回归（AR）模型。
代入滞后算子，$Bz_{t-n} = z_{t-(n+1)}$，得

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

那么，自回归模型式可以简写为

$$\phi(B)\tilde{z}_t = a_t \tag{2}$$

3.1.1.2 自回归模型

在自回归模型中，过程的现值会被表示成有限个过程过去值的线性加权和，再加上一个随机扰动 a_t 的形式。
设时点对应的过程取值为 $z_t, z_{t-1}, z_{t-2}, \ldots$，并且令 $\tilde{z}_t = z_t - \mu$ 表示与 μ 的偏离序列。得到

$$\tilde{z}_t = \phi_1 \tilde{z_{t-1}} + \phi_2 \tilde{z_{t-2}} + \ldots + \phi_p \tilde{z_{t-p}} + a_t$$

此模型称为p 阶自回归（AR）模型。
代入滞后算子，$Bz_{t-n} = z_{t-(n+1)}$，得

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

那么，自回归模型式可以简写为

$$\phi(B)\tilde{z}_t = a_t \qquad (2)$$

3.1.1.3 移动平均模型

在对时间序列的观察中，另一种在实践中非常重要的模型是有限的移动平均过程。这里，$\tilde{z}_t$ 线性依赖于有限的 q 个前期扰动 $a_{t-1}, a_{t-2}, \ldots, a_{t-q}$，可以表示为

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{q-1}$$

该式被称为q 阶移动平均（MA）过程。
再代入滞后算子，可将模型简写为：

$$\tilde{z}_t = \theta(B) a_t \tag{3}$$

3.1.1.4 自回归移动平均模型

为了拟合实际中更加复杂的时间序列，有些时候模型中既含有自回归项，也含有移动平均项，能更好地表达时间序列的规律。
自回归移动平均（ARMA）模型设定为：

$$\tilde{z}_t = \phi_1 \tilde{z_{t-1}} + \phi_2 \tilde{z_{t-2}} + \ldots + \phi_p \tilde{z_{t-p}} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{q-1}$$

化简得，

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

此模型还可以转化成线性滤波形式：

$$\tilde{z}_t = \phi^{-1}(B)\theta(B)a_t = \Psi(B)a_t \tag{4}$$

其中，$\Psi(B) = \phi^{-1}(B)\theta(B)$

3.1.2 非平稳模型

由自回归模型的公式可知，

$$\phi(B)\tilde{z}_t = a_t$$

$$\varphi(B)\tilde{z}_t = a_t$$

该式中的 $\phi(B)$ 为平稳模型自回归算子，$\varphi(B)$ 为非平稳模型自回归算子。

在该书后续的章节中会推导：如果 $\phi(B) = 0$ 的根在单位圆外，一个 ARMA 过程是平稳的;

如果根在单位圆内，它会表现出爆炸性的非平稳行为;

唯一剩下的就是 $\phi(B) = 0$ 的根在单位圆上这种情况。

设 $\varphi(B)$ 为非平稳自回归算子，该算子有 d 个单位根，其余的根位于单位圆之外。

$\phi(B)$ 为平稳自回归算子，那么可以得到两算子之间的关系：

$$\varphi(B) = \phi(B)(1 - B)^d$$

3.1.2 非平稳模型

那么具有单位根特征的非平稳模型可以写作：

$$\varphi(B)\tilde{z}_t = \phi(B)(1-B)^d\tilde{z}_t = \theta(B)a_t \tag{5}$$

令 $\omega_t = (1-B)^d\tilde{z}_t = \nabla^d\tilde{z}_t$
且对于 $d \geq 1$，有 $(1-B)^d\tilde{z}_t = (1-B)^d z_t$，即

$$\omega_t = \nabla^d z_t \tag{6}$$

(6) 式模型的意义是：该系列的第 d 个差分可以表示为一个平稳可逆的 ARMA 过程。代入（5）式得，

$$\phi(B)\omega_t = \theta(B)a_t \tag{7}$$

3.1.3 平稳和非平稳时间序列模型

此模型可以描述 d 阶积分过程（p,d,q），即ARIMA(p,d,q) 过程。
这个过程是：

$$\phi(B)\nabla^d z_t = \theta(B)a_t$$

d=0　时，该模型为自回归移动平均模型；
d=0,q=0 时，该模型为自回归平均模型；
d=0,p=0 时，该模型为移动平均模型；
因此，ARIMA(p,d,q) 包含四种模式形态。

3.2 转换函数模型

由自回归移动平均模型可知，连接输入 X 和输出之间 Y 的**转换函数**可以表示为：

$$Y_t = v(B)X_t$$

其中，$v(B) = \delta^{-1}(B)\Omega(B)$

然而，由于**噪音**（干扰）的存在，破坏了输入和输出之间的关系，假设噪声 N_t 也能通过自回归移动平均模型来描述，就得到下面的式子：

$$N_t = \Psi(B)a_t$$

其中，$\Psi(B) = \varphi^{-1}(B)\theta(B)$

在原有的转换函数中叠加噪声模型，根据线性滤波模型，得到：

$$Y_t = v(B)X_t + N_t = \delta^{-1}(B)\Omega(B)X_t + \varphi^{-1}(B)\theta(B)a_t \qquad (8)$$

3.3 离散控制系统模型

- 设 t 时刻，输出 Y_t 与目标 T 之间的偏差为 $\varepsilon_t = Y_t - T$。
- 控制的目的是要选一个控制方程:$x_t = X_t - X_{t-1}$，使得偏差 ε 尽可能小。

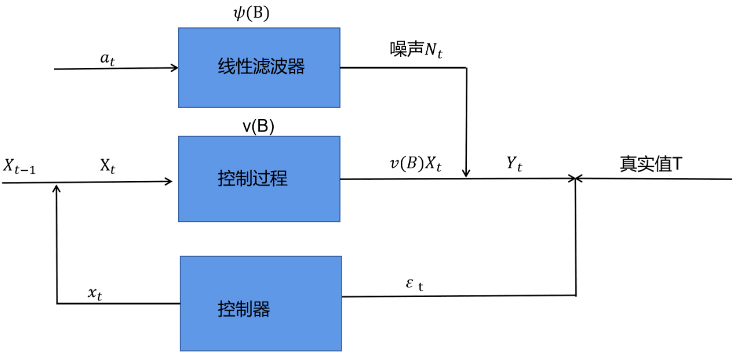


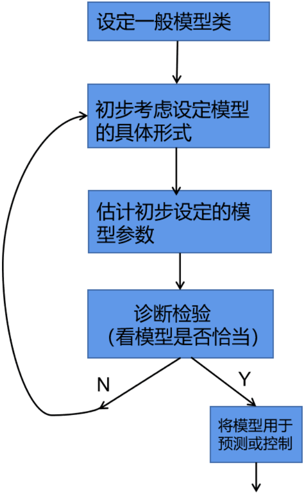
图: 反馈控制步骤图

4. 建模思想

4.1 简洁性

由于模型中的参数需要利用数据去估计，所以，所要用到的数学模型需要控制参数个数。参考前面的建模过程中，滞后算子的使用，以及 ARIMA(p,d,q) 的四种模式形态，均能反应模型的简洁性。

4.2 模型选择的试验阶段



考虑对当前目标有效的一类模型.

通过选择有效的方法对模型进行识别，并得到模型参数的初步估计。

选择合适的初始值与估计方法。

若拟合不足，就要诊断拟合不足的原因；若发现模型不恰当，就要放弃模型，然后重复前面的步骤。

图: 模型选择流程图

请批判指正！