

## **Part-II**

## 第 11 章

# 趋势的估计

在经济时间序列等数据中，常常可以观察到长期持续增加或减少的倾向。本章前半部分将介绍多项式回归模型，作为包含这种变动倾向的时间序列的分析方法。接着，为了估计多项式回归模型等简单参数模型无法表达的复杂趋势，引入趋势模型。表达参数随机变化的趋势分量模型，是后续各章介绍的各种非平稳时间序列模型的基础。

### 11.1 多项式回归模型

图 1.1(e) 的 WHARD 数据在整个区间内呈现增长趋势。这种在经济时间序列等中经常出现的长期趋势变动称为**趋势** (trend)。不言而喻，为了把握时间序列的走势和进行预测，估计趋势是必要的。即使目的是分析更短期的变动分量，也不宜将 AR 模型等平稳模型直接应用于原序列，而是通常先用某种方法去除趋势后再对序列进行分析。

本节将介绍作为估计趋势最简单方法的**多项式回归模型** (polynomial trend model)。多项式回归模型假设时间序列  $y_n$  可以表示为多项式与残差之和

$$y_n = t_n + w_n \quad (11.1)$$

其中， $w_n$  为均值 0、方差  $\sigma^2$  的正态白噪声， $t_n$  为  $x_n$  的多项式

$$t_n = a_0 + a_1 x_n + \cdots + a_m x_n^m \quad (11.2)$$

由于该模型是回归模型的特殊情形，可以用第 5 章的方法方便地求解。为了求多项式趋势模型，在 (5.1) 中令第  $j$  个解释变量为  $x_{nj} = x_n^{j-1}$ ，构造

$$X = \begin{bmatrix} 1 & x_1 & \cdots & x_1^m & y_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & x_N & \cdots & x_N^m & y_N \end{bmatrix} \quad (11.2)$$

即可。此时通过 Householder 变换  $U$ , 缩约为上三角矩阵

$$UX = \begin{bmatrix} S \\ O \end{bmatrix} = \begin{bmatrix} s_{11} & \cdots & s_{1,m+2} \\ & \ddots & \vdots \\ & & s_{m+2,m+2} \\ & & & 0 \end{bmatrix} \quad (11.2)$$

则  $j$  阶多项式趋势模型的残差方差  $\hat{\sigma}_j^2$  可由

$$\hat{\sigma}_j^2 = \frac{1}{N} \sum_{i=j+2}^{m+2} s_{i,m+2}^2 \quad (11.2)$$

求得。此外,  $j$  阶模型具有  $j+1$  个回归系数和残差方差作为参数, 其 AIC 为

$$AIC_j = N \log 2\pi \hat{\sigma}_j^2 + N + 2(j+2) \quad (11.2)$$

使其最小的  $j$  即为最优阶数。对于这样确定的阶数  $j$ , 通过求解一次方程

$$\begin{bmatrix} s_{11} & \cdots & s_{1,j+1} \\ & \ddots & \vdots \\ & & s_{j+1,j+1} \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_j \end{bmatrix} = \begin{bmatrix} s_{1,m+2} \\ \vdots \\ s_{j+1,m+2} \end{bmatrix} \quad (11.2)$$

可求得回归系数的最大似然估计值  $\hat{a}_0, \dots, \hat{a}_j$ 。由于矩阵  $S$  为上三角矩阵, 该一次方程可通过后向代入简单地求解。

### 例：最高气温数据

R 的 TSSS 程序包中的函数 `polreg` 用于对时间序列拟合多项式回归模型。通过参数 `order` 指定多项式回归的最高阶数, 即可用 AIC 确定最优阶数。函数的输出如下:

```
order.maice:  AIC 最小的趋势阶数
sigma2:      阶数为  $0 \leq m \leq M$  的模型的残差方差
aic:         阶数为  $0 \leq m \leq M$  的模型的 AIC
daic:       AIC - 最小 AIC
coef:       回归系数  $a(i, m)$ ,  $i = 0, \dots, m$ ,  $m = 1, \dots, M$ 
trend:      AIC 最小模型的趋势估计值
```

利用函数 `polreg` 的输出, 例如要绘制 AIC 的变化图, 可以将 `polreg` 的输出保存到 `z` 中, AIC 的信息通过 `z$aic` 获取。

```

> # 东京最高气温数据
> data( Temperature )
> polreg( Temperature, order = 13 )
>
> # 显示多项式回归模型 AIC 的变化
> z <- polreg( Temperature, order = 13 )
> plot( z$aic,type="b",lwd=2 )
>
> # 批发额数据
> data( WHARD )
> y <- log10( WHARD )
> polreg( y, order = 15 )

```

表 11.1 是对图 1.1(c) 所示的最高气温数据拟合多项式回归模型时的残差方差  $\hat{\sigma}_j^2$  和  $AIC_j$  的汇总。残差方差  $\hat{\sigma}_j^2$  单调递减，而 AIC 在 6 阶时取得最小值。

表 11.1 最高气温数据

$j$	$\hat{\sigma}_j^2$	$AIC_j$	$j$	$\hat{\sigma}_j^2$	$AIC_j$
0	60.09	3373.76	7	8.96	2463.09
1	58.89	3366.02	8	8.96	2465.02
2	33.61	3095.35	9	8.91	2464.24
3	23.74	2928.47	10	8.89	2465.29
4	10.18	2518.95	11	8.89	2467.27
5	9.65	2494.72	12	8.86	2467.22
6	8.97	2461.42	13	8.86	2469.18

同样，表 11.2 是 WHARD 数据对数值的情况。在此情况下，10 阶时 AIC 取得最小值。

表 11.2 WHARD 数据

$j$	$\hat{\sigma}_j^2$	$AIC_j$	$j$	$\hat{\sigma}_j^2$	$AIC_j$
0	0.02752	-113.04	7	0.00122	-582.11
1	0.00163	-548.73	8	0.00115	-589.46
2	0.00150	-560.34	9	0.00112	-591.45
3	0.00149	-559.11	10	0.00107	-595.98
4	0.00147	-559.56	11	0.00107	-594.18
5	0.00142	-562.75	12	0.00106	-594.38
6	0.00123	-583.37	13	0.00105	-593.59

图 11.1 的 (a) 和 (b) 分别展示了各时间序列及使用 AIC 最小阶数的多项式回归模型估计的趋势。(a) 的最高气温数据的情况下，捕捉到了气温的缓慢变化。(b) 的

WHARD 数据的情况下, 整体上获得了平滑的估计值, 但在数据的开始部分趋势呈现急剧变化, 且在对应于 1974~75 年的  $n = 95$  附近的急剧下降未能被很好地捕捉。对于这种问题的改进趋势估计方法将在 11.3 节以后说明。

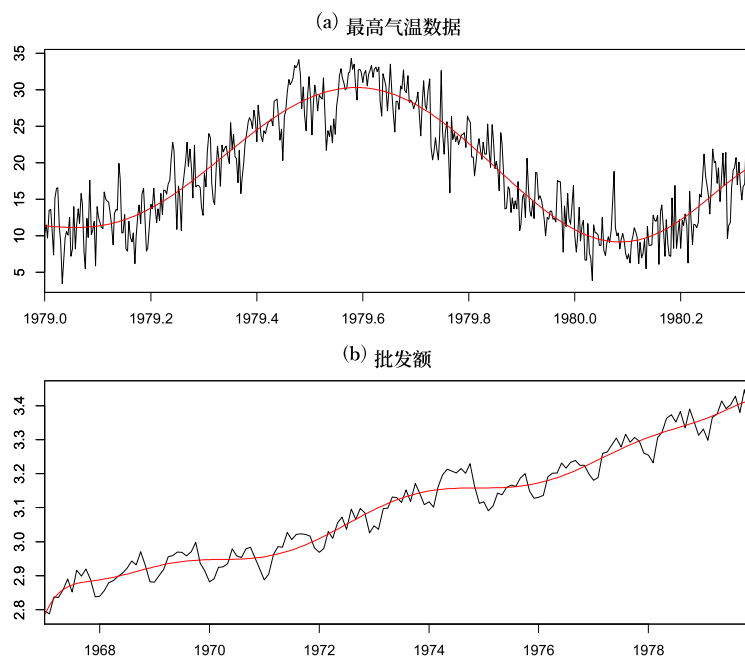


图 11.1 最高气温数据和 WHARD 数据的多项式趋势

## 11.2 趋势分量模型 – 结构随机变化的模型

前节的多项式趋势模型可以表示为

$$y_n \sim N(t_n, \sigma^2) \quad (11.2)$$

因此该模型可以说是假设方差恒定、但均值按多项式  $t_n$  变化的模型。这种参数模型在真实趋势确实遵循多项式、且  $w_n$  可视为白噪声的情况下能给出非常好的估计值，但在其他情况下，可能无法很好地捕捉趋势的变化，或者相反地过度追随数据而拾取偶然变动。这种现象尤其在数据区间的两端经常出现。下面我们考虑为了改善这些问题，将多项式回归模型扩展为更灵活的模型的方法。作为具体例子，考虑表示 1 次多项式（直线）的方程

$$t_n = an + b \quad (11.2)$$

此时，将  $\Delta \equiv 1 - B$  定义为由  $\Delta t_n = t_n - t_{n-1}$  给出的时间差分算子则

$$\Delta t_n = a, \quad \Delta^2 t_n = 0 \quad (11.2)$$

成立。这表明 1 次多项式是 2 阶差分方程初值问题

$$\Delta^2 t_n = 0, \quad \Delta t_0 = a, \quad t_0 = b \quad (11.2)$$

的解。一般地,  $k-1$  次多项式可以看作  $k$  阶差分方程

$$\Delta^k t_n = 0 \quad (11.2)$$

的解。

作为将多项式扩展为更灵活函数的方法, 可以考虑用  $t_n$  的  $k$  阶差分近似为 0 来代替差分方程, 即要求  $\Delta^k t_n \approx 0$  成立。这可以通过引入如下  $k$  阶随机差分方程

$$\Delta^k t_n = v_n \quad (11.2)$$

来实现。其中,  $v_n$  为均值 0、方差  $\tau^2$  的正态分布  $N(0, \tau^2)$  的白噪声。以下将 (11.2) 称为**趋势分量模型** (trend component model)。

由于  $\Delta^k t_n = 0$  的解为  $k-1$  次多项式,  $k$  阶趋势分量模型可以看作  $k-1$  次多项式的扩展。当噪声方差  $\tau^2$  较小时, 趋势分量模型的实现值在局部上是类似多项式的非常光滑的函数。但是, 与多项式的显著区别在于, 从整体上看可以表达非常灵活的函数。

**例** 当  $k=1$  时, 该模型变为**随机游走模型** (random walk model)

$$t_n = t_{n-1} + v_n, \quad v_n \sim N(0, \tau^2) \quad (11.2)$$

表达了趋势在局部近似取恒定值  $t_n \approx t_{n-1}$  的模型。此外, 当  $k=2$  时

$$t_n = 2t_{n-1} - t_{n-2} + v_n \quad (11.2)$$

表达了趋势在局部近似线性变化  $t_n - 2t_{n-1} + t_{n-2} \approx 0$  的模型。

一般地, 趋势分量模型 (11.2) 由于

$$\Delta^k = (1-B)^k = \sum_{i=0}^k {}_k C_i (-B)^i \quad (11.2)$$

通过  $c_i = (-1)^{i+1} {}_k C_i$  定义系数  $c_i$ , 可以表示为

$$t_n = \sum_{i=1}^k c_i t_{n-i} + v_n \quad (11.2)$$

当  $k=1$  时  $c_1 = 1$ , 当  $k=2$  时  $c_1 = 2$ ,  $c_2 = -1$ 。该模型不满足平稳性, 但形式上可以看作  $k$  阶 AR 模型, 因此令

$$\begin{aligned} x_n &= \begin{bmatrix} t_n \\ t_{n-1} \\ \vdots \\ t_{n-k+1} \end{bmatrix}, \quad F = \begin{bmatrix} c_1 & c_2 & \cdots & c_k \\ 1 & & & \\ & \ddots & & \\ & & 1 & \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ H &= [1 \quad 0 \quad \cdots \quad 0] \end{aligned} \quad (11.3)$$

可得趋势分量模型的状态空间表示

$$\begin{aligned}x_n &= Fx_{n-1} + Gv_n \\t_n &= Hx_n\end{aligned}\quad (11.2)$$

例 当  $k=1$  时

$$x_n = t_n, \quad F = G = H = 1 \quad (11.2)$$

即可。此外, 当  $k=2$  时

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \end{bmatrix}, \quad F = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H = [1 \ 0] \quad (11.2)$$

或者

$$x_n = \begin{bmatrix} t_n \\ -t_{n-1} \end{bmatrix}, \quad F = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H = [1 \ 0] \quad (11.2)$$

即可。(11.2) 给出了第 9 章所示的标准形。

此外, 当  $k=2$  时, 将状态向量定义为  $x_n = (t_n, \delta t_n)^T$ , 令

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad H = [1 \ 0] \quad (11.2)$$

也可以。在该模型中  $\delta t_n \equiv \Delta t_n$  成立, 可以简单地验证这与 (11.2) 是等价的表示。

但是, 使用该表示可以将趋势分量模型自然地进一步扩展。即令

$$v_n = \begin{bmatrix} v_{n1} \\ v_{n2} \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad H = [1 \ 0] \quad (11.2)$$

则趋势分量满足

$$\begin{aligned}\delta t_n &= \delta t_{n-1} + v_{n2} \\t_n &= t_{n-1} + \delta t_{n-1} + v_{n1} + v_{n2} \\&= t_{n-1} + \delta t_n + v_{n1}\end{aligned}\quad (11.2)$$

这一模型。在这种情况下  $\Delta t_n = \delta t_n + v_{n1}$ , 与之前使用的趋势分量模型具有一维系统噪声不同, 这是一个在趋势的水平 and 斜率两者上都加入独立噪声的更灵活的模型。

### 11.3 趋势模型

趋势表示时间序列的中长期变动倾向。反过来说, 实际观测到的时间序列是趋势上叠加了各种变动分量的结果。其中最简单的情况是以  $w_n$  为白噪声

$$y_n = t_n + w_n \quad (11.1)$$

可以这样表示的情况。这是表示获得观测值机制的最简单模型，可以看作第 9 章观测模型的一个例子。为了从时间序列  $y_n$  中估计趋势  $t_n$ ，可以考虑将趋势分量模型和观测模型组合而成的如下模型。将其称为**趋势模型** (trend model)。

$$\Delta^k t_n = v_n \quad (11.2)$$

$$y_n = t_n + w_n \quad (11.3)$$

其中， $v_n$  与 (11.2) 相同，为服从均值 0、方差  $\tau^2$  正态分布的白噪声， $w_n$  为服从均值 0、方差  $\sigma^2$  正态分布的白噪声。(11.3) 是将时间序列  $y_n$  看作在趋势上叠加了独立噪声后的观测结果的观测模型。另一方面，(11.2) 是对趋势变化方式进行建模的趋势分量模型。实际上，通常不具备这样简单的结构，在那种情况下需要如下一章所示的更复杂的建模。

利用前节的状态空间模型的状态空间表示，趋势模型可以用如下状态空间模型表示。

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \\ y_n &= Hx_n + w_n \end{aligned} \quad (11.3)$$

其中，状态向量为适当定义的  $k$  维向量， $F$ 、 $G$ 、 $H$  分别为从 (11.2)、(11.3) 适当确定的  $k \times k$  矩阵、 $k$  维列向量、 $k$  维行向量。与趋势分量模型的区别仅在于包含了观测噪声。例如，当  $k=2$  时

$$\begin{aligned} x_n &= \begin{bmatrix} t_n \\ t_{n-1} \end{bmatrix}, & F &= \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, & G &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ H &= [1 \quad 0] \end{aligned} \quad (11.4)$$

即可。

趋势模型的阶数  $k$  以及方差  $\tau^2$ 、 $\sigma^2$  确定后，通过卡尔曼滤波和平滑算法可以求得状态向量的平滑值  $x_{1|N}, \dots, x_{N|N}$ 。此时，由于状态向量的第 1 个分量是  $t_n$ ， $x_{n|N}$  的第 1 个分量即  $Hx_{n|N}$  就是趋势的平滑值  $t_{n|N}$ 。

### 例：最高气温数据的趋势

TSSS 程序包的函数 `trend` 用于对时间序列拟合趋势模型并估计趋势。需要趋势的阶数 `trend.order`、系统噪声方差的初始值 `tau2.ini` 以及方差  $\tau^2$  的最优化搜索步长。当 `tau2=NULL` 时，即未给定初始值的情况下，该函数在  $\tau^2 = 2^{-k}$ ,  $k=1, 2, \dots$  的范围内比较 AIC 值，确定  $\tau^2$  的粗略估计值。以该估计值为初始值反复计算，即可求得高精度的估计值。

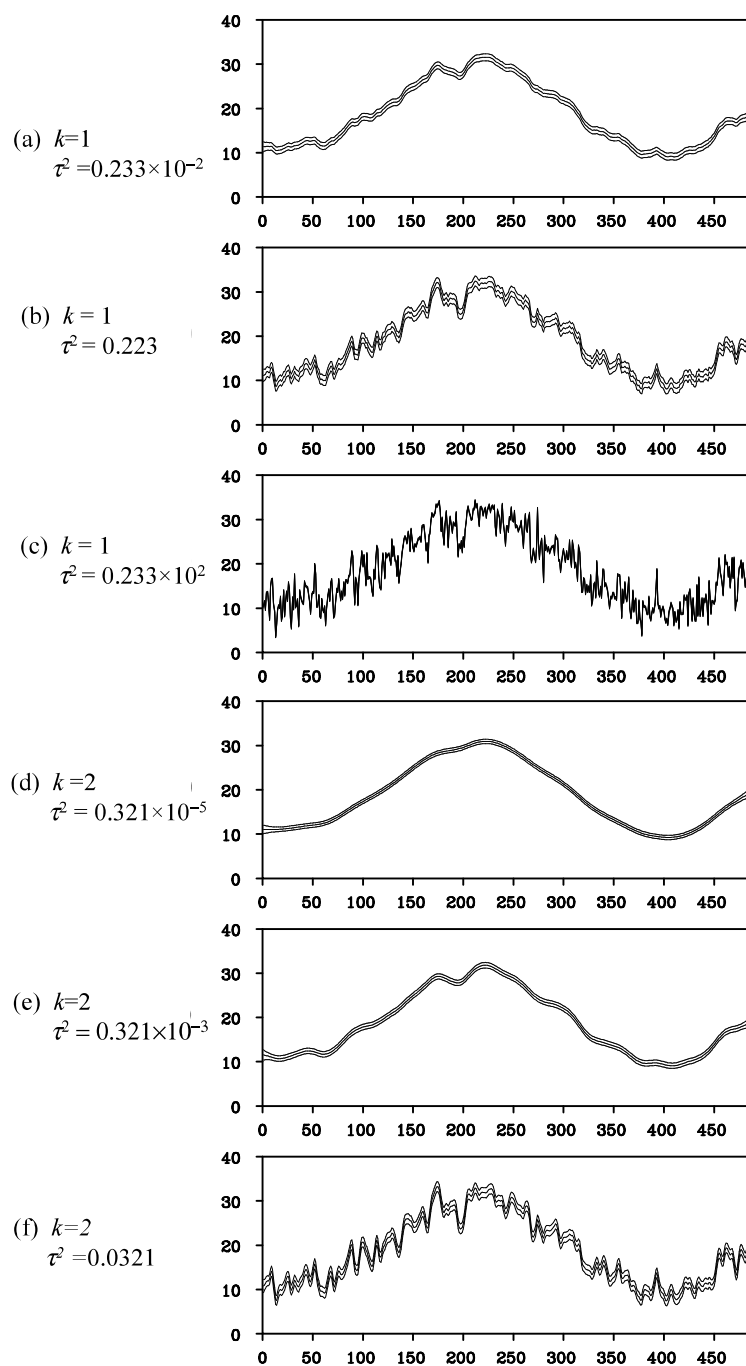
图 11.2 使用不同  $k$  和  $\tau^2$  的趋势模型估计的气温数据趋势

表 11.3 趋势模型的 AIC

$k = 1$		$k = 2$	
$\tau^2$	AIC	$\tau^2$	AIC
$0.223 \times 10^{-2}$	2690	$0.321 \times 10^{-5}$	2556
0.223	2448	$0.321 \times 10^{-3}$	2506
$0.223 \times 10^2$	2528	0.0321	2562

该函数的输出包括趋势的估计值 (trend)、残差 (residual)、系统噪声方差  $\tau^2$  (tau2)、观测噪声方差  $\sigma^2$  (sigma2)、模型的对数似然 (llhood) 以及 AIC (aic)。

```

> # 东京日最高气温
> data( Temperature )
> # 1 阶趋势模型
> trend( Temperature, trend.order = 1, tau2.ini = 0.223, delta = 0.001 )
>
> # 2 阶趋势模型
> trend( Temperature, trend.order = 2 )

```

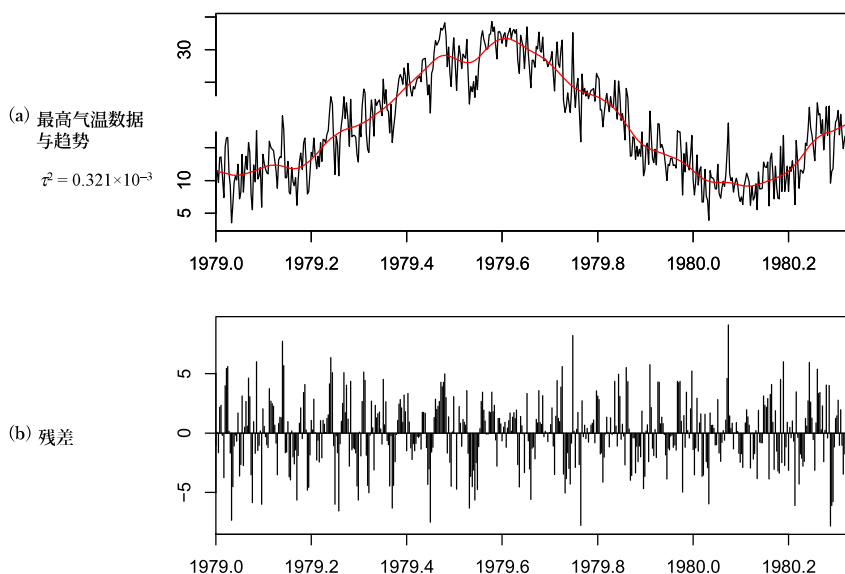
图 11.2 展示了对气温数据使用不同阶数  $k$  和系统噪声方差  $\tau^2$  的 6 个模型求得趋势估计值, 其中  $\sigma^2$  使用了最大似然法估计的值。(a) 是  $k = 1$ 、 $\tau^2 = 0.223 \times 10^{-2}$  的情况。估计的趋势基本捕捉到了气温的年周期变化, 与多项式趋势模型的估计值类似。当  $k = 1$ 、 $\tau^2 = 0.223$  时的 (b) 中, 趋势估计值再现了更详细的变动。(c) 是  $k = 1$ 、 $\tau^2 = 0.223 \times 10^2$  的情况。趋势只是在追随数据的变动。

另一方面, 当  $k = 2$  时,  $\tau^2 = 0.321 \times 10^{-5}$  的 (d) 过于简单,  $\tau^2 = 0.0321$  的 (f) 过于复杂, 而  $\tau^2 = 0.321 \times 10^{-3}$  的 (e) 则获得了恰到好处的估计值。进一步比较 (b) 和 (e), (e) 获得了更平滑的估计值。

趋势模型中包含阶数  $k$  和方差  $\sigma^2$  与  $\tau^2$  作为参数。如上例所示, 根据这些值的不同, 可以获得各种不同的趋势估计值。因此, 为了获得好的趋势估计值, 需要将这些值设定为适当的值。 $\sigma^2$  和  $\tau^2$  可以通过最大似然法求出最优值, 利用 9.6 节的方法, 实际上只需选择  $\lambda = \frac{\tau^2}{\sigma^2}$  的值,  $\sigma^2$  的值就会自动确定。由此可知, 决定趋势估计值的是系统噪声与观测噪声的方差比, 称为折中参数 (trade-off parameter)。

趋势模型的阶数可以通过 AIC 来确定, 通常多使用  $k = 2$ 。但是, 当趋势变化幅度较大时,  $k = 1$  可能更好。

表 11.3 汇总了图 11.2 所示模型的 AIC 值。在  $k = 1$  中 (b) 最小, 在  $k = 2$  中 (e) 最

图 11.3 函数 `trend` 的输出示例

小。此外，(b) 和 (e) 相比，(b) 明显更小，表明 (b) 是所有模型中最好的估计值。注意，(b) 和 (e) 的  $\tau^2$  值是通过最大似然法求得的最优值。

根据 AIC，比起  $k=2$  所求得平滑趋势，选择了  $k=1$  的变动较大的趋势。这表明在趋势模型中  $k=1$  更好，但需要注意的是，假设  $w_n$  为白噪声的趋势模型本身可能存在问题。图 11.3 展示了对最高气温数据使用  $k=2$  时函数 `trend` 的输出。得到了  $\hat{\tau}^2 = 0.321 \times 10^{-3}$ 。在这种情况下，残差序列呈现序列相关，难以认为是白噪声。

如下一章所示，在这种情况下，通过使用季节调整模型将时间序列分解为趋势、AR 分量和观测噪声三个分量，可以获得既具有平滑趋势又对数据有良好拟合的模型。这里附带说明，用于估计季节调整模型的函数在不包含季节分量的情况下也可以使用。

### ● 章末习题

1. 根据股价的随机游走假说，股价  $y_n$  服从随机游走模型  $y_n = y_{n-1} + v_n$ , ( $v_n \sim N(0, \sigma^2)$ )。
  - (1) 当某日  $n$  的日经平均股价为  $y_n = 17,000$  且  $\sigma^2 = 40,000$  时，求  $k$  日后 ( $k = 1, \dots, 5$ ) 的股价预测值及其预测误差的方差。

- (2) 求 4 日后达到 17,400 日元以上的概率。
- (3) 实际股价是否满足随机游走假说？如果存在问题，应如何改进模型？
- (4) 尝试估计实际股价数据的趋势。

## 第 12 章

# 季节调整模型

本章将说明作为趋势模型扩展的一个例子——季节调整模型。经济时间序列中存在每年同一时期呈现相同变动模式的特有模式。对这类时间序列进行走势判断或预测时，如果不考虑该特征，就有可能得出错误的结论。季节调整模型是用于分析以固定时间间隔重复相同模式的时间序列的模型。通过季节调整模型，可以将时间序列  $y_n$  分解为趋势  $t_n$ 、季节分量  $s_n$  和白噪声  $w_n$  三个分量，表示为  $y_n = t_n + s_n + w_n$ 。同时也将说明包含平稳 AR 分量和星期效应项的扩展模型。

### 12.1 季节分量模型

时间序列变动中每年重复出现的变动模式部分  $s_n$  称为**季节分量** (seasonal component)。以下考虑一个周期内有  $p$  个观测值的情况。月度数据的情况下  $p = 12$ ，季度数据的情况下  $p = 4$ 。此时，将季节分量记为  $s_n$ ，则  $s_n$

$$s_n = s_{n-p} \quad (12.0)$$

近似成立。利用时间延迟算子  $B$  来表示，由  $B^p s_n = s_{n-p}$  可知

$$(1 - B^p)s_n = 0 \quad (12.0)$$

近似成立。

与第 11 章的随机趋势分量模型的情况类似，作为随时间缓慢变化的季节分量的模型，可以考虑

$$(1 - B^p)^\ell s_n = v_{n2}, \quad v_{n2} \sim N(0, \tau_2^2) \quad (12.0)$$

特别地，令  $\ell = 1$  时，得到假设  $\{s_{pn+i}\}$  ( $n = 1, 2, \dots$ ) 为随机游走的模型

$$s_n = s_{n-p} + v_{n2} \quad (12.0)$$

因此，将前章的模型作为趋势分量  $t_n$  的模型，由 (11.15)、(12.1) 以及  $y_n = t_n + s_n + w_n$ ，即可得到季节调整的基本模型。

然而, 作为季节分量模型最自然的 (12.1) 在实际中并不能很好地适用。其原因在于, 比较趋势分量模型和季节分量模型的算子表示 (11.18) 和 (12.1), 可以发现

$$(1 - B^p)^\ell = (1 - B)^\ell (1 + B + \cdots + B^{p-1})^\ell$$

因此包含  $(1 - B)^q$  ( $q \geq 1$ ) 作为公共因子。此时令  $e_n$  为差分方程

$$(1 - B)^q e_n = 0 \quad (12.0)$$

的任意解。当  $q = 1$  时  $e_n$  取任意常数即可。这里定义

$$\begin{aligned} t'_n &= t_n + e_n \\ s'_n &= s_n - e_n \end{aligned}$$

则它们分别满足 (11.15) 和 (12.1), 而且

$$y_n = t'_n + s'_n + w_n \quad (12.-2)$$

成立。因此, 对于完全相同的  $v_{n1}$ 、 $v_{n2}$ 、 $w_n$ , 存在多种分解。而且, 由于与这些分解对应的模型的似然由  $v_{n1}$ 、 $v_{n2}$ 、 $w_n$  确定, 因此无法判断这些分解的优劣。这样, 在时间序列分解中使用具有公共因子的分量时, 表示的唯一性就会丧失。

保证分解唯一性的一个简单方法是, 使各分量模型不具有公共因子。因此, 注意到要使  $(1 - B^p)^\ell = 0$  成立, 只需

$$(1 + B + \cdots + B^{p-1})^\ell = 0 \quad (12.-2)$$

成立即可。为了使  $S_n \approx S_{n-p}$ , 只需

$$\sum_{i=0}^{p-1} B^i s_n \approx 0 \quad (12.-2)$$

成立即可。因此, 作为允许时间变化的季节变动分量的随机模型, 可以考虑

$$\left( \sum_{i=0}^{p-1} B^i \right)^\ell s_n = v_{n2}, \quad v_{n2} \sim N(0, \tau_2^2) \quad (12.-2)$$

本书将该模型称为阶数  $\ell$ 、周期  $p$  的**季节分量模型** (seasonal component model)。但是, 除了季节分量呈现显著趋势性变化的情况外, 通常使用  $\ell = 1$  的一阶模型

$$\sum_{i=0}^{p-1} s_{n-i} = v_{n2}, \quad v_{n2} \sim N(0, \tau_2^2) \quad (12.-2)$$

为了求得季节分量模型的状态空间表示, 首先将

$$\left( \sum_{i=0}^{p-1} B^i \right)^\ell = 1 - \sum_{i=1}^{\ell(p-1)} d_i B^i \quad (12.-2)$$

展开。当  $\ell = 1$  时  $d_i = -1$ , 当  $\ell = 2$  时  $d_i = -i - 1$ , ( $i \leq p - 1$ ),  $d_i = i + 1 - 2p$ , ( $p \leq i \leq 2(p - 1)$ )。在这种情况下, 季节分量模型形式上也是 AR 模型, 因此其状态空间表示为

$$x_n = \begin{bmatrix} s_n \\ s_{n-1} \\ \vdots \\ s_{n-\ell(p-1)+1} \end{bmatrix}, \quad F = \begin{bmatrix} d_1 & d_2 & \cdots & d_{\ell(p-1)} \\ 1 & & & \\ & \ddots & & \\ & & & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$H = [1, 0, \cdots, 0] \quad (12.-2)$$

## 12.2 标准季节调整模型

在标准的季节调整中, 将时间序列  $y_n$

$$y_n = t_n + s_n + w_n \quad (12.-2)$$

分解为趋势分量  $t_n$ 、季节分量  $s_n$ 、观测噪声  $w_n$  三个分量。利用前面推导的趋势分量模型和季节分量模型, 标准季节调整模型为

$$y_n = t_n + s_n + w_n \quad (\text{观测模型}) \quad (12.-1)$$

$$\Delta^k t_n = v_{n1} \quad (\text{趋势分量模型}) \quad (12.0)$$

$$\left( \sum_{i=0}^{p-1} B^i \right)^\ell s_n = v_{n2} \quad (\text{季节分量模型}) \quad (12.1)$$

其中  $w_n \sim N(0, \sigma^2)$ ,  $v_{n1} \sim N(0, \tau_1^2)$ ,  $v_{n2} \sim N(0, \tau_2^2)$ 。将该模型称为标准的**季节调整模型** (seasonal adjustment model)。

趋势分量模型的阶数为  $k$ , 周期为  $p$  时, 定义  $k + (p - 1)\ell$  维状态向量  $x_n = (t_n, \cdots, t_{n-k+1}, s_n, s_{n-1}, \cdots, s_{n-(p-1)\ell+1})^T$ , 噪声向量  $v_n = (v_{n1}, v_{n2})^T$ , 令

$$F = \begin{bmatrix} F_1 & \mathbf{O} \\ \mathbf{O} & F_2 \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & \mathbf{O} \\ \mathbf{O} & G_2 \end{bmatrix}, \quad H = [H_1 \ H_2] \quad (12.1)$$

则标准季节调整模型可由

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \\ y_n &= Hx_n + w_n \end{aligned} \quad (12.1)$$

这一状态空间模型表示。其中,  $F_1$ 、 $G_1$  和  $H_1$  是趋势分量模型状态空间表示中使用的矩阵和向量,  $F_2$ 、 $G_2$  和  $H_2$  是季节分量模型状态空间表示中使用的矩阵和向量。

例如, 当  $k = 2$ 、 $\ell = 1$ 、 $p = 4$  时

$$F = \begin{bmatrix} 2 & -1 & & & & \\ 1 & 0 & & & & \\ & & -1 & -1 & -1 & \\ & & & 1 & & \\ & & & & 1 & \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (12.2)$$

$$H = [1 \ 0 \ 1 \ 0 \ 0 \ 0]$$

即可。

### 例：WHARD 批发额数据的季节调整

R 程序包 TSSS 中的函数 `season` 用于拟合季节调整模型，估计趋势分量和季节分量。通过指定以下参数，可以求得各种季节调整模型：

<code>trend.order:</code>	趋势模型的阶数 (0, 1, 2, 3)
<code>seasonal.order:</code>	季节阶数 (0, 1, 2)
<code>period:</code>	当时间序列 $y$ 的属性为 NULL 时，需要指定季节周期的长度
<code>tau2.ini:</code>	系统噪声方差 $\tau^2$ 的初始估计值
<code>filter:</code>	指定滤波起始时点和终点的向量 $c(x1, x2)$
<code>predict:</code>	长期预测的最终时点 ( $\geq x2$ )。
<code>log:</code>	逻辑变量。为 TRUE 时使用数据 $y$ 的对数变换
<code>minmax:</code>	观测值的最小值和最大值。其他值视为异常值
<code>plot:</code>	逻辑变量。为 TRUE (默认) 时，绘制趋势和季节分量图

季节调整法不仅适用于月度数据的年周期，还可以应用于各种周期。例如，季度数据设 `period=4`，月度数据设=12，日数据的周周期设=5 或 7，小时数据的日周期设=24 即可。函数 `season` 给出以下输出：

```
tau2:      系统噪声的方差
sigma2:    观测噪声的方差
llkhood:   模型的对数似然
aic:       模型的 AIC
trend:     趋势分量 (trend.order > 0 时)
seasonal:  季节分量 (seasonal.order > 0 时)
noise:     噪声分量
cov:       平滑估计值的方差协方差矩阵
```

```
> # 批发额数据
> data( WHARD )
> season( WHARD, trend.order =2, seasonal.order =1, log =TRUE )
```

图 12.1 是对 WHARD 数据的对数值应用标准季节调整模型，分解为趋势、季节分量和噪声的结果。最大似然法估计的参数为  $\hat{\tau}_1^2 = 0.0248$ ， $\hat{\tau}_2^2 = 0.11 \times 10^{-7}$ ， $\hat{\sigma}^2 = 0.156 \times 10^{-3}$ ， $AIC = -728.50$ 。趋势整体呈现稳定的增长倾向，但在 1974~75 年前后出现了急剧下降。此外，季节分量在整个期间内非常稳定。

函数 `season` 还可以利用估计的季节调整模型进行长期预测。为此需要通过参数

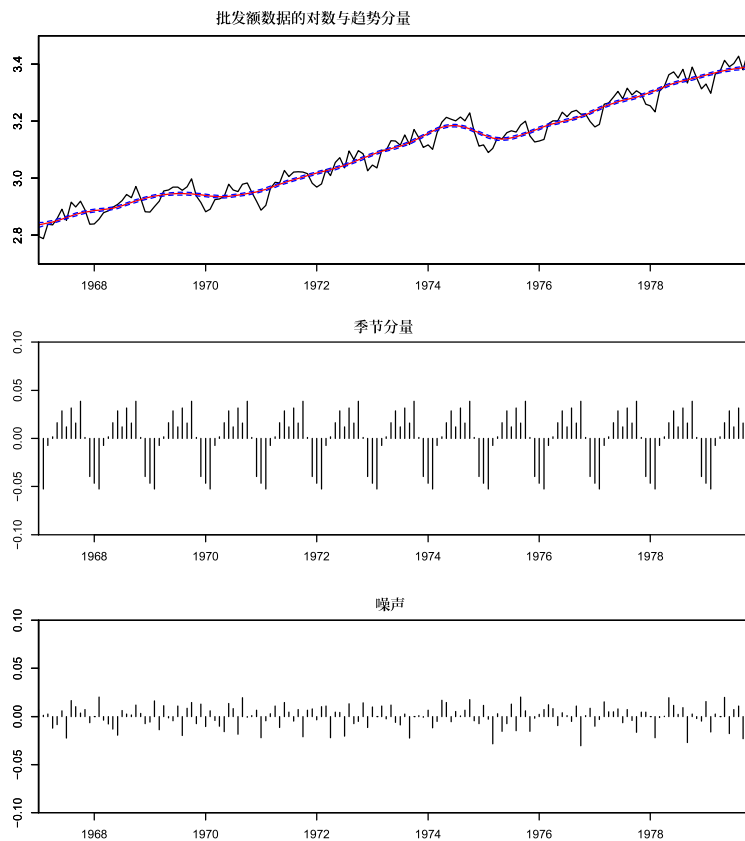


图 12.1 标准季节调整模型对 WHARD 数据的季节调整

`filter` 指定用于模型参数估计的数据区间。在以下示例中, 指定 `filter=c(1,132)`, 季节调整模型使用时间序列数据  $y_1, \dots, y_{132}$  进行估计, 其余数据  $y_{133}, \dots, y_N$  由估计的模型进行预测。

```
> # 批发额数据
> data( WHARD )
> season( WHARD, trend.order = 2, seasonal.order = 1, filter = c(1,132), log =
TRUE )
```

图 12.2 展示了 156 个数据中仅使用前 132 个观测值, 对剩余 2 年 (24 个) 时间序列用第 9 章的方法进行长期预测的结果。图中绘制了预测分布的均值  $y_{132+j|132}$  以及  $\pm 1$  个标准差的预测误差范围, 即  $y_{132+j|132} \pm \sqrt{d_{132+j|132}}$  ( $j = 1, \dots, 24$ )。同时也显示了实际的时间序列值, 由于该数据的趋势和季节分量都很稳定, 长期预测的均值与实际值重叠, 表明获得了极好的预测值。

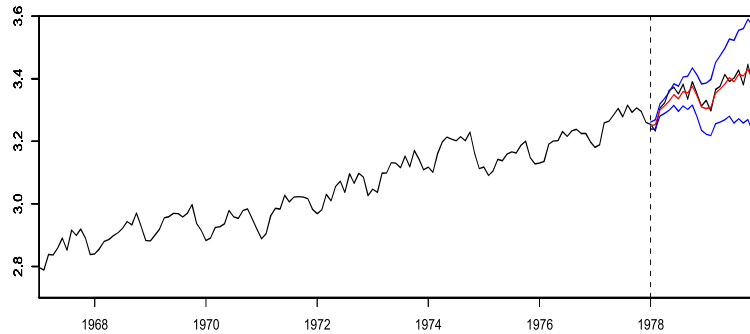


图 12.2 季节调整模型的长期预测

### 12.3 包含平稳 AR 分量的分解

这里考虑标准季节调整法的扩展。在标准季节调整法中，时间序列被分解为趋势分量、季节分量和观测噪声三个分量，各分量分别服从 (12.15)、(12.16) 等模型，且假设观测噪声为白噪声。因此，当严重偏离这些假设时，标准季节调整法得到的分解可能是不恰当的。图 12.3 是对图 1.1(d) 所示的食品产业从业人数数据，取  $k=2$ 、 $\ell=1$ 、 $p=12$  进行季节调整的结果。在这种情况下，估计的趋势与图 12.1 不同，在数据的中间部分和末端呈现波动。

```
> data( BLSALLFOOD )
> season( BLSALLFOOD, trend.order = 2, seasonal.order = 1 )
```

```
> data( BLSALLFOOD )
> season( BLSALLFOOD, trend.order = 2, seasonal.order = 1, filter = c(1,132) )
```

首先考虑得到这样的趋势时的问题。图 12.4 是食品产业从业人数数据 156 个数据中，与图 12.2 类似地仅使用前 132 个观测值对剩余 2 年 (24 个) 时间序列进行长期预测的结果。在这种情况下，预测值  $y_{132+j|132}$  看似很好地预测了实际时间序列  $y_{132+j}$ ，但从预测误差范围爆发性扩大可以推断，实际上该数据的情况下，利用该模型进行预测几乎没有可信度。

图 12.5 是以  $n=132$  为中心，前后各半年即  $n=126 \sim 138$  的各点为预测起始点，对  $n=156$  进行长期预测的 13 种预测值的叠加图。在图 12.3 中趋势估计值下降的  $n=130$  以前为起始时点进行长期预测时，预测值将大幅向下方偏移。另一方

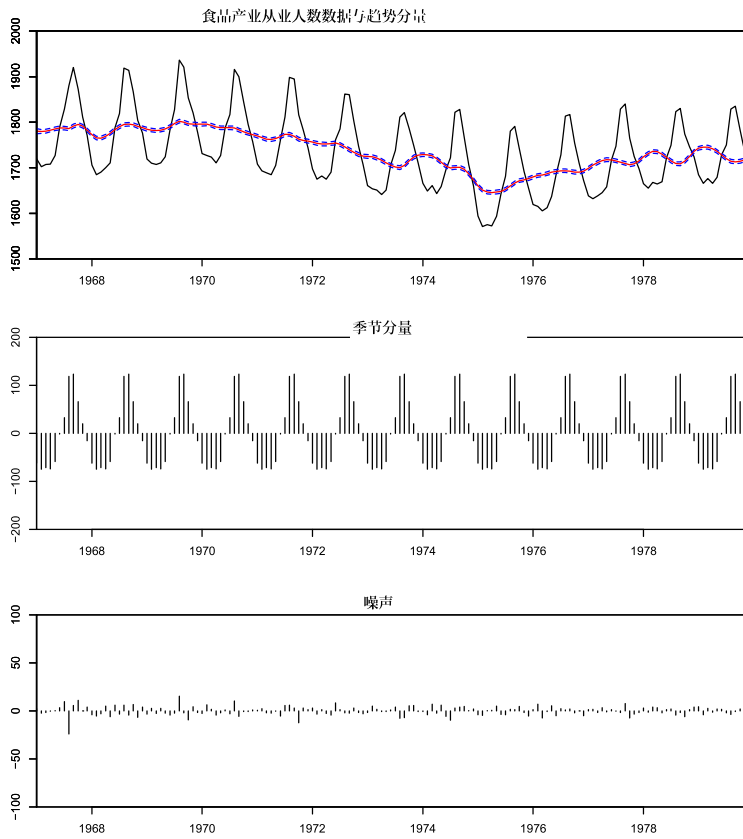
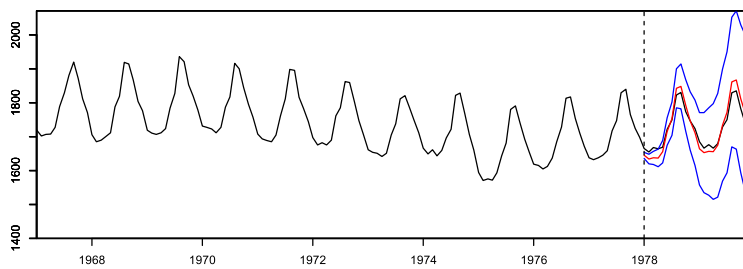


图 12.3 标准季节调整模型对食品产业从业人数数据的季节调整

图 12.4 食品产业从业人数数据的长期预测 (预测起始时点:  $n = 132$ )

面, 从趋势估计值呈上升趋势的  $n = 135$  以后开始长期预测时, 预测值反而大幅向上方偏移。预测误差范围迅速扩大反映了这一情况。季节调整模型中含有噪声项的趋势分量模型虽能灵活表现复杂趋势, 但在长期预测中其均值  $t_{n+j|n}$  与简单按差分方程

$$\Delta^k t_{n+j|n} = 0 \quad (12.1)$$

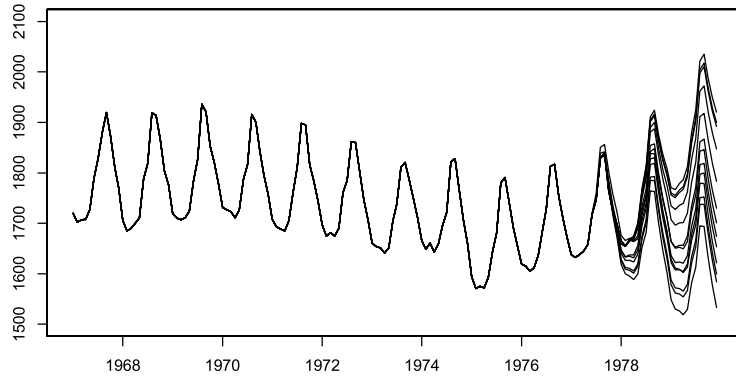


图 12.5 移动预测起始点时食品产业从业人数数据的长期预测 (预测起始点:  $n = 126 \sim 138$ )

计算的结果一致。因此，根据趋势最终阶段的倾向，预测值会向下或向上偏移。这样，当估计的趋势呈现波动时，利用标准季节调整法求得的模型来预测未来的走势是不恰当的。

如果仅以 1 年或 2 年以上的长期预测为目的，通过将  $\tau_1^2$  即趋势分量系统噪声的方差设定为比最大似然法估计的最优值更小的值，可以得到更平滑的趋势，从而可能获得较好的预测值。但是，这种方法不仅难以合理确定  $\tau_1^2$  的值，而且长期预测改善的同时，1 期前预测等短期预测能力会显著下降。

为了获得在短期预测和长期预测两方面都表现优异的模型，需要在标准季节调整模型考虑的三个分量之外，加入新的分量  $p_n$

$$y_n = t_n + s_n + p_n + w_n \tag{12.1}$$

考虑四分量的分解。其中  $p_n$  称为**平稳 AR 分量** (stationary AR component)，假设服从自回归模型

$$p_n = \sum_{i=1}^{m_3} a_i p_{n-i} + v_{n3} \tag{12.1}$$

其中， $v_{n3}$  为均值 0、方差  $\tau_3^2$  的正态白噪声。该模型表示的不是趋势分量那样的长期倾向，而是循环变动等短期变动分量。在 (12.3) 的模型中，标准季节调整中用一个趋势分量表示的部分，被自动分解为长期趋势曲线  $t_n$  和短期变动  $p_n$  两部分。如第 9 章所示，AR 模型令  $x_n = (p_n, p_{n-1}, \dots, p_{n-m_3+1})^T$ ,

$$F_3 = \begin{bmatrix} a_1 & a_2 & \cdots & a_{m_3} \\ 1 & & & \\ & \ddots & & \\ & & 1 & \end{bmatrix}, \quad G_3 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{12.2}$$

$$H_3 = [ 1 \ 0 \ \cdots \ 0 ], \quad Q_3 = \tau_3^2$$

可得状态空间模型。因此，利用第9章的模型合成方法，要进行(12.3)的分解，可使用由

$$\begin{aligned}
 F &= \begin{bmatrix} F_1 & & \\ & F_2 & \\ & & F_3 \end{bmatrix}, & G &= \begin{bmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & G_3 \end{bmatrix} \\
 H &= [H_1 \quad H_2 \quad H_3], & Q &= \begin{bmatrix} \tau_1^2 & 0 & 0 \\ 0 & \tau_2^2 & 0 \\ 0 & 0 & \tau_3^2 \end{bmatrix}
 \end{aligned} \tag{12.1}$$

定义的状态空间模型即可。

**例：包含AR分量的模型的季节调整法** 函数 `season` 中，要用包含AR分量的模型进行季节调整，只需如下通过参数 `ar.order` 指定AR阶数即可。

```
> data(BLSALLFOOD)
> season(BLSALLFOOD, trend.order = 2, seasonal.order = 1, ar.order = 2)
```

图12.6是将食品产业从业人数数据用该方法分解为趋势分量、季节分量、平稳AR分量和观测噪声的结果。估计的趋势与图12.1的情况类似，是非常平滑的曲线。另一方面，短期变动以平稳AR分量的形式表示，(趋势分量) + (平稳AR分量)与图12.3的趋势分量非常相似。该模型的AIC为1336.54。标准季节调整模型的AIC为1369.30，因此该模型明显更好。

```
> data(BLSALLFOOD)
> season(BLSALLFOOD, trend.order = 2, seasonal.order = 1, ar.order = 2, filter =
c(1,132))
```

图12.7是使用该模型在  $n = 132$  时点进行长期预测的结果。在这种情况下，无论短期还是长期都获得了良好的预测结果。

## 12.4 包含星期效应项的分解

在记录百货商店每月销售额的月度数据等中，由于一个月中包含的各星期天数的差异，可能导致销售额出现显著差异。这是因为周日和周六顾客较多，而周一等特定星期设有定休日。这种现象不仅出现在经济时间序列中，在观测  $\text{NO}_x$  和  $\text{CO}$  等环境数据中也经常出现。

为了消除这种星期差异的影响而提出的方法是**星期调整** (trading-day adjustment)。为了用时间序列模型实现星期调整，需要引入服从适当模型的星期效应

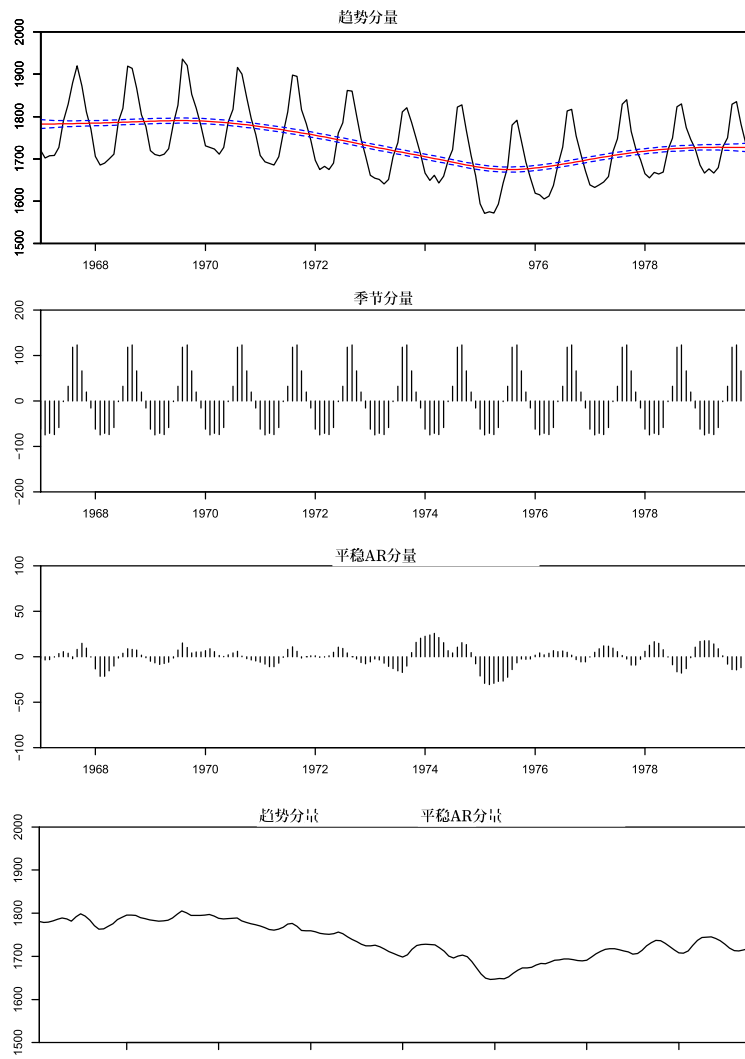


图 12.6 包含平稳 AR 分量的模型对 BLSALLFOOD 数据的季节调整

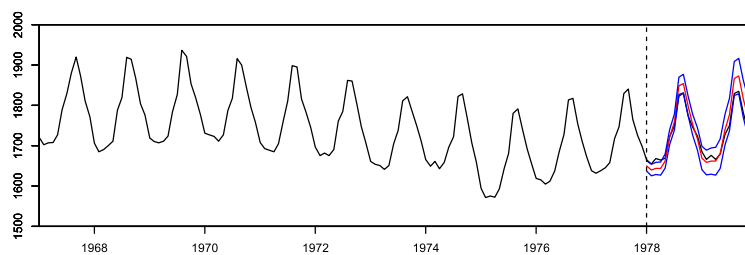


图 12.7 包含平稳 AR 分量的模型对食品产业从业人数数据的长期预测

项。将与时间序列  $y_n$  对应的第  $n$  个月中周日到周六的天数记为  $d_{n1}^*, \dots, d_{n7}^*$ 。各  $d_{ni}^*$  取 4 或 5 的值。此时假设星期效应为

$$td_n = \sum_{i=1}^7 \beta_{ni} d_{ni}^* \quad (12.1)$$

$\beta_{ni}$  是第  $i$  个星期日的增减对  $y_n$  的影响系数。但为了保证时间序列分解的唯一性, 需要附加所有系数之和为 0 的条件, 即

$$\beta_{n1} + \dots + \beta_{n7} = 0 \quad (12.1)$$

此时, 由于  $\beta_{n7} = -(\beta_{n1} + \dots + \beta_{n6})$ , 星期效应实际上仅用  $\beta_{n1}, \dots, \beta_{n6}$  即可表示为

$$\begin{aligned} td_n &= \sum_{i=1}^6 \beta_{ni} (d_{ni}^* - d_{n7}^*) \\ &\equiv \sum_{i=1}^6 \beta_{ni} d_{ni} \end{aligned} \quad (12.1)$$

其中  $d_{ni} \equiv d_{ni}^* - d_{n7}^*$  是各星期天数减去周六天数。

作为星期效应项的模型, 为简单起见假设  $\beta_{ni}$  本身随时刻  $n$  按一阶趋势分量模型

$$\Delta \beta_{ni} = v_{n4}^{(i)}, \quad v_{n4}^{(i)} \sim N(0, \tau_4^2), \quad i = 1, \dots, 6 \quad (12.1)$$

变化时, 表示星期效应项的状态空间模型为

$$\begin{aligned} F_{n4} &= G_{n4} = I_6, \quad H_{n4} = [d_{n1}, \dots, d_{n6}] \\ x_{n4} &= \begin{bmatrix} \beta_{n1} \\ \vdots \\ \beta_{n6} \end{bmatrix}, \quad Q = \begin{bmatrix} \tau_4^2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \tau_4^2 \end{bmatrix} \end{aligned} \quad (12.1)$$

在这种情况下,  $H_{n4}$  为依赖于时刻  $n$  的向量。此外, 为简单起见假设  $Q$  为所有对角元素相等的对角矩阵。但通常使用星期效应系数不随时间变化而恒定, 即

$$\beta_{ni} \equiv \beta_i \quad (12.1)$$

的模型。此时, 令  $\tau_4^2 = 0$  或  $G = 0$  即可。此时,

$$\begin{aligned} x_n &= \begin{bmatrix} x_{n1} \\ x_{n2} \\ x_{n3} \\ x_{n4} \end{bmatrix}, \quad F = \begin{bmatrix} F_1 & & & \\ & F_2 & & \\ & & F_3 & \\ & & & F_4 \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & G_3 \\ 0 & 0 & 0 \end{bmatrix} \\ H &= [H_1 \quad H_2 \quad H_3 \quad H_{n4}], \quad Q = \begin{bmatrix} \tau_1^2 & 0 & 0 \\ 0 & \tau_2^2 & 0 \\ 0 & 0 & \tau_3^2 \end{bmatrix} \end{aligned} \quad (12.1)$$

则可以构成将时间序列

$$y_n = t_n + s_n + p_n + td_n + w_n \quad (12.1)$$

分解为 5 个分量的模型。

### 例：包含星期效应项的季节调整

R 函数 `season` 中，要进行星期调整，只需如下设定 `trade=TRUE` 即可。

```
> season( WHARD, trend.order = 2, seasonal.order = 1, ar.order = 0, trade = TRUE,
log = TRUE )
```

图 12.8 是对 WHARD 批发额数据的对数值应用包含星期效应项的季节调整模型，分解为趋势、季节分量、星期效应项和噪声的结果。趋势和季节分量的估计值与图 12.1 的标准分解情况几乎相同。星期效应项看起来不太大，但从图 12.9 中（季节分量）+（星期效应项）的图来看，可以发现原始数据的细微变动也被很好地再现。顺便提一下，标准模型的 AIC 为  $-728.50$ ，包含星期效应项时为  $-778.18$ 。当星期效应项的系数  $\beta_i$  被估计后， $d_{n1}, \dots, d_{n6}$  对于将来的  $n$  也是确定的，因此可以实现高精度的预测。

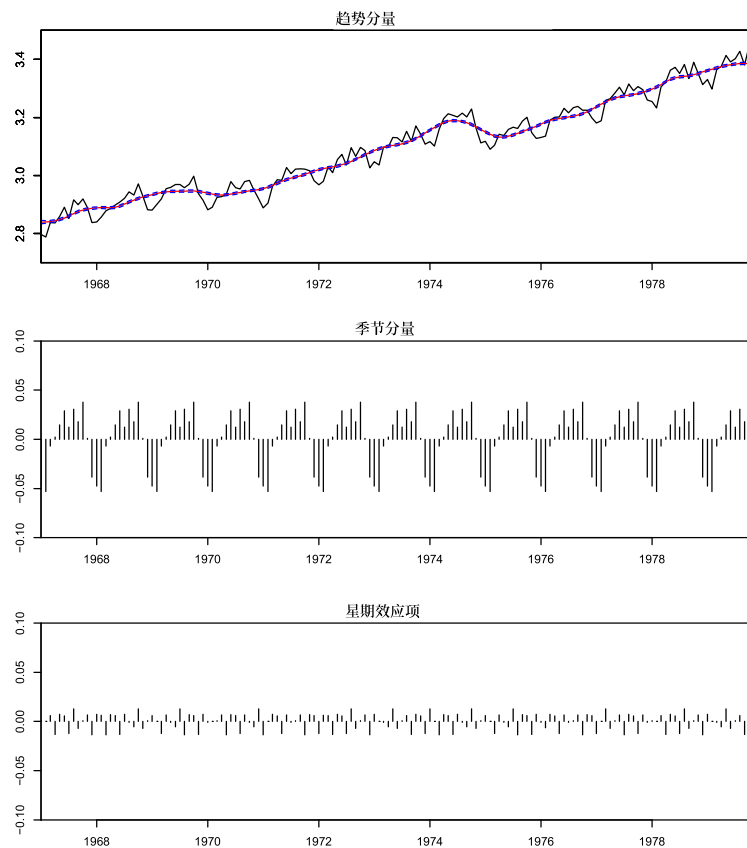


图 12.8 WHARD 数据的星期调整

函数 `season` 可以将时间序列分解为趋势、季节分量和星期效应项。在以下 R 代码中，这些分量分别由参数 `z$trend`、`z$season`、`z$day.effect` 给出。图 12.9 展示了

WHARD 数据对数的季节分量、星期效应项、季节分量 + 星期效应项以及 WHARD 数据对数减去趋势后的结果。虽然星期效应项的大小较小，但季节分量和星期效应项之和与数据减去趋势后的结果非常相似。这表明标准季节调整模型的噪声分量中相当大的部分实际上是星期效应项。

```

> data(WHARD)
> z <- season(WHARD, trend.order = 2, seasonal.order = 1, ar.order = 0, trade =
TRUE, log = TRUE)
> a1 <- z$seasonal + z$day.effect
> a2 <- log10( as.numeric(WHARD) )
> a3 <- a2 - z$trend
> par(mar=c(2,2,3,1)+0.1)
> par( mfrow=c(4,1) )
> plot( z$seasonal,type="l", ylim=c(-0.1,0.1), main="seasonal component" )
> plot( z$day.effect,type="l", ylim=c(-0.1,0.1), main="trading-day effect" )
> plot( a1,type="l", ylim=c(-0.1,0.1), main="seasonal + trading-day effect" )
> plot( a3,type="l", ylim=c(-0.1,0.1), main="data - trend" )

```

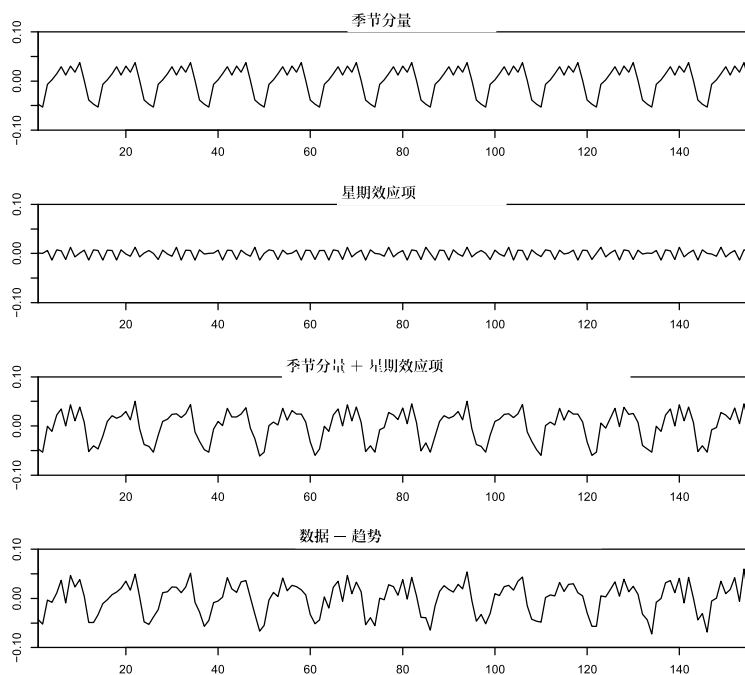


图 12.9 批发额数据的星期调整

1. 说明在包含平稳 AR 分量的分解中需要注意的事项。
2. 考虑与 12.1 节不同的季节分量模型。
3. 在星期调整中, 给出以下假设下的模型。
  - (1) 仅考虑周末和其他星期两种效应的情况
  - (2) 周日、周六和其他星期三种情况效应不同的情况
4. 在季节调整中, 说明本书未提及的其他可能因素。

## 第 13 章

# 时变系数 AR 模型

典型的非平稳时间序列中，有一类是平均值随时间变化的，另一类则是围绕平均值变动的方式发生变化的。在后一种情况下，时间序列的方差、自协方差函数以及功率谱会随时间变化。本章将针对此类时间序列，说明随时间变化的方差估计方法，以及拟合系数随时间变化的时变系数 AR 模型的方法。

### 13.1 时变方差模型

设  $y_n$  是均值为 0、方差为  $\sigma_n^2$  的正态白噪声。此时，假设  $\sigma_{2m-1}^2 = \sigma_{2m}^2$ ，通过

$$s_m = y_{2m-1}^2 + y_{2m}^2 \quad (13.1)$$

生成长度减半的时间序列  $s_1, \dots, s_{N/2}$ 。由于  $s_m$  服从自由度为 2 的  $\chi^2$  分布（即指数分布）， $s_m$  的密度函数为：

$$f(s) = \frac{1}{2\sigma^2} e^{-s/2\sigma^2} \quad (13.2)$$

因此，进一步进行变换：

$$z_m = \log\left(\frac{s_m}{2}\right) \quad (13.3)$$

则  $z_m$  的密度函数变为：

$$g(z) = \frac{1}{\sigma^2} \exp\left\{z - \frac{e^z}{\sigma^2}\right\} = \exp\left\{(z - \log \sigma^2) - e^{(z - \log \sigma^2)}\right\} \quad (13.4)$$

这表明可以表示为：

$$z_m = \log \sigma^2 + w_m \quad (13.5)$$

其中， $w_m$  是双指数分布，其密度函数由下式给出：

$$h(w) = \exp\{w - e^w\} \quad (13.6)$$

因此, 通过状态空间模型

$$\begin{aligned}\Delta^k t_m &= v_m \\ z_m &= t_m + w_m\end{aligned}\quad (13.7)$$

求得  $z_m$  的趋势项, 即可估计原序列  $y_n$  方差的对数值。问题在于  $w_m$  的分布并非正态分布, 但由于双指数分布的均值为  $-\gamma$  (其中  $\gamma=0.57722$  为欧拉常数), 方差为  $\frac{\pi^2}{6}$ , 因此通过进行如下正态分布近似:

$$w_m \sim N\left(-\gamma, \frac{\pi^2}{6}\right)\quad (13.8)$$

即可利用卡尔曼滤波进行趋势估计。此时,  $t_{m|M} + \gamma$  ( $M = \frac{N}{2}$ ) 即为  $\log \sigma_m^2$  的估计值。此外, 如果在第 14 章的非高斯型滤波器中直接使用式 (13.6) 的密度函数, 则可以在不使用正态近似的情况下求得趋势。

实际上, 在式 (13.1) 中使用相邻两个项的和并非硬性要求, 主要是为了使  $w_n$  的分布更接近正态分布。因此, 在使用非高斯型滤波器时, 也可以取  $s_n = y_n^2$ , 这种情况下则不需要  $\sigma_{2m-1}^2 = \sigma_{2m}^2$  的假设。

#### 例: 地震数据的时变方差

R 语言包 TSSS 中的函数 `tvvar` 用于估计时变方差。参数 `trend.order` 是趋势成分模型的阶数, `tau2.ini` 和 `delta` 分别指定系统噪声方差  $\tau^2$  最大似然估计的初始估计值和搜索步长。如果未明确给出这些参数 (即 `tau2.ini = NULL` 时), 则会在  $\tau^2 = 2^{-k}, k = 1, 2, \dots$  的范围内选择最优值。如果希望更精确地估计  $\tau^2$ , 可以将该输出作为初始值重复计算。

函数 `tvvar` 返回以下输出:

```
tvv:      时变方差
sm:       按式 (13.1) 变换后的数据
trend:    变换后数据 sm 的趋势项
noise:    残差
nordata:  标准化后的数据
tau2:     系统噪声的方差
sigma2:   观测噪声的方差
llkhood:  模型的对数似然
aic:      模型的 AIC
tsname:   时间序列  $y_n$  的名称
```

```
> data( MYE1F )
> tvvar( MYE1F, trend.order = 2, tau2.ini = 6.6e-06, delta = 1.0e-06 )
```

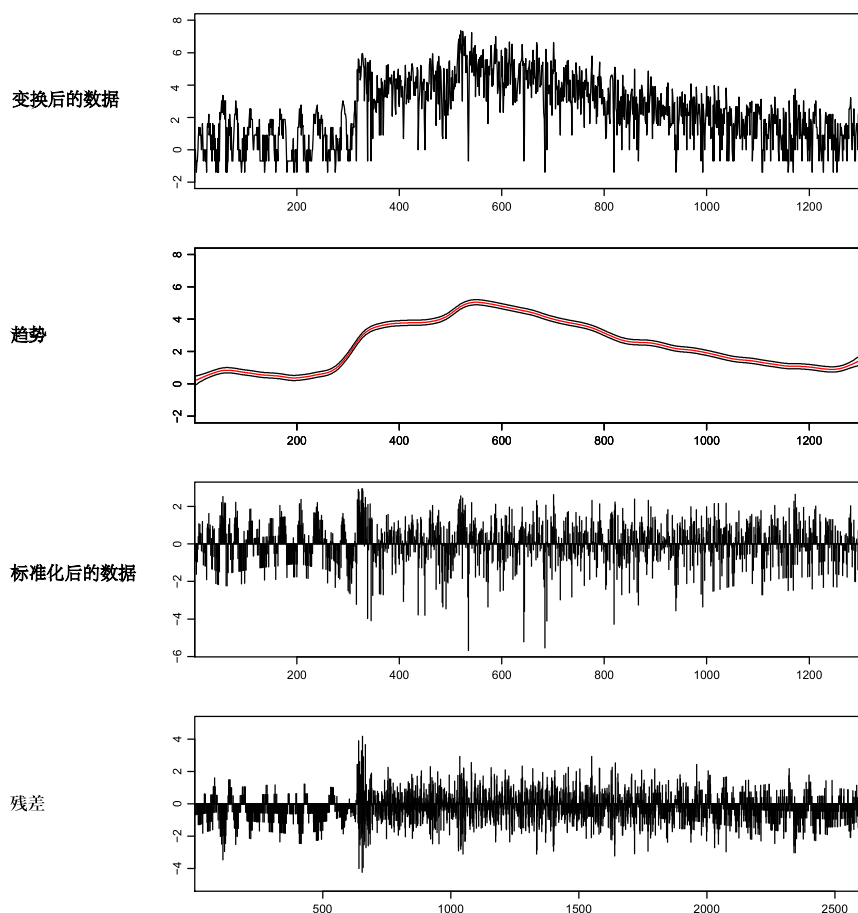


图 13.1 地震波数据的时变方差估计与时间序列标准化

图 13.1 显示了对图 2.1 所示的地震波数据应用函数 `tvvar` 的结果，其中趋势阶数为 2，初始方差估计值为  $\tau^2 = 0.66 \times 10^{-5}$ 。由上至下依次为：按式 (13.1) 变换后的数据  $s_m$ 、 $s_m$  的趋势、方差对数的估计值  $\log \hat{\sigma}^2$ 、通过  $\tilde{y}_n = \frac{y_n}{\sqrt{\hat{\sigma}^2}}$  标准化后的数据以及残差。地震波数据虽非白噪声，但通过此方法可以变换为方差近似为 1 的时间序列。

## 13.2 时变系数 AR 模型

平稳时间序列的变动特征由自协方差函数或功率谱刻画。因此，对于变动特征本身随时间变化的非平稳时间序列，其自协方差函数或功率谱也会随时间变化。在平稳时间序列中，通过适当选择 AR 模型或 ARMA 模型等时间序列模型的阶数和系数，可以表达各种自协方差函数和功率谱。因此，当变动方式随时间变化时，时间序列模型本身也会发生变化。本节针对非平稳时间序列  $y_n$ ，考虑系数随时间

变化的 AR 模型：

$$y_n = \sum_{j=1}^m a_{nj} y_{n-j} + w_n \quad (13.9)$$

其中,  $w_n$  是均值为 0、方差为  $\sigma^2$  的正态白噪声。该模型称为  $m$  阶时变系数 AR 模型 (time-varying coefficients AR model), 而  $a_{nj}$  是时刻  $n$  处滞后为  $j$  的自回归系数, 称为时变 AR 系数。

对于给定的时间序列  $y_1, \dots, y_N$ , 即使仅考虑自回归系数, 时变系数 AR 模型也包含  $mN$  个未知数, 简单地对模型 (13.9) 应用最大似然法或最小二乘法无法获得有意义的结果。因此, 与趋势模型和季节调整模型类似, 我们为参数的时间变化引入随机趋势成分模型。在趋势模型中, 我们将趋势成分  $t_n$  视为未知数并为其时间变化引入模型; 而在时变系数 AR 模型中, 由于 AR 系数  $a_{nj}$  随时刻  $n$  变化, 我们引入以下约束：

$$\Delta^k a_{nj} = v_{nj}, \quad j = 1, \dots, m \quad (13.10)$$

其中,  $\Delta$  是关于时刻  $n$  的差分算子, 满足  $\Delta a_{nj} = a_{nj} - a_{n-1,j}$ 。  $k$  通常取 1 或 2。  $v_n = (v_{n1}, \dots, v_{nm})^T$  是均值向量为 0、方差协方差矩阵为  $Q$  的  $m$  维正态白噪声。通常假设当  $i \neq j$  时  $v_{ni}$  与  $v_{nj}$  相互独立, 因此  $Q$  是以  $\tau_{11}^2, \dots, \tau_{mm}^2$  为对角元的对角矩阵, 可写为  $Q = \text{diag}\{\tau_{11}^2, \dots, \tau_{mm}^2\}$ 。这里进一步假设所有方差相等, 即  $\tau_{11}^2 = \dots = \tau_{mm}^2 = \tau^2$ 。关于该假设的含义和合理性将在第 13.4 节探讨。

接下来, 为了在约束模型 (13.10) 下估计时变系数 AR 模型 (13.9), 我们将这两个模型表示为状态空间模型。首先, 当  $k=1$  时定义  $x_{nj} = a_{nj}$ ; 当  $k=2$  时定义  $x_{nj} = (a_{nj}, a_{n-1,j})^T$  为  $k$  维向量。此时, 式 (13.10) 可以表示为系统模型形式：

$$x_{nj} = F^{(k)} x_{n-1,j} + G^{(k)} v_{nj} \quad (13.11)$$

其中,

$$F^{(1)} = G^{(1)} = 1$$

$$F^{(2)} = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad G^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (13.12)$$

如果将式 (13.9) 右侧的第  $j$  项视为该模型的第  $j$  个成分, 则可表示为：

$$a_{nj} y_{n-j} = H_n^{(kj)} x_{nj} \quad (13.13)$$

其中,  $H_n^{(1j)} = y_{n-j}$ ,  $H_n^{(2j)} = (y_{n-j}, 0)$ 。此时, 时变系数 AR 模型的第  $j$  个成分利用  $F^{(k)}$ 、 $G^{(k)}$  和  $H_n^{(kj)}$  可表示为状态空间模型:

$$\begin{aligned} x_{nj} &= F^{(k)}x_{n-1,j} + G^{(k)}v_{nj} \\ a_{nj}y_{n-j} &= H_n^{(kj)}x_{nj} \end{aligned} \quad (13.14)$$

此外, 利用  $H^{(1)} = 1$ 、 $H^{(2)} = (1, 0)$ , 可以写成:

$$H_n^{(kj)} = y_{n-j}H^{(k)} \quad (13.15)$$

利用上述成分模型, 可得到时变系数 AR 模型的状态空间表示:

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \\ y_n &= H_nx_n + w_n \end{aligned} \quad (13.16)$$

其中,  $F$  和  $G$  分别是如下  $km \times km$  和  $km \times m$  矩阵,  $H_n$  和  $x_n$  则是如下  $km$  维向量:

$$\begin{aligned} F &= \begin{bmatrix} F^{(k)} & & \\ & \ddots & \\ & & F^{(k)} \end{bmatrix} = I_m \otimes F^{(k)} \\ G &= \begin{bmatrix} G^{(k)} & & \\ & \ddots & \\ & & G^{(k)} \end{bmatrix} = I_m \otimes G^{(k)} \\ H_n &= [H_n^{(k1)}, \dots, H_n^{(km)}] = (y_{n-1}, \dots, y_{n-m}) \otimes H^{(k)} \\ x_n &= \begin{cases} (a_{n1}, \dots, a_{nm})^T, & k=1 \text{ 时} \\ (a_{n1}, a_{n-1,1}, \dots, a_{nm}, a_{n-1,m})^T, & k=2 \text{ 时} \end{cases} \\ Q &= \begin{bmatrix} \tau^2 & & \\ & \ddots & \\ & & \tau^2 \end{bmatrix}, \quad R = \sigma^2 \end{aligned} \quad (13.17)$$

在这里,  $I_m$  是  $m \times m$  单位矩阵,  $\otimes$  表示矩阵的 Kronecker 积。若  $A$  是  $k \times \ell$  矩阵,  $B$  是  $m \times n$  矩阵, 则  $A \otimes B$  是  $km \times \ell n$  矩阵。

综上所述, 时变系数 AR 模型 (13.9) 及其关于系数时间变化的约束模型 (13.10)

可以表示为状态空间模型。例如当  $m=2$ 、 $k=2$  时, 状态空间模型如下:

$$\begin{aligned} \begin{bmatrix} a_{n1} \\ a_{n-1,1} \\ a_{n2} \\ a_{n-1,2} \end{bmatrix} &= \begin{bmatrix} 2 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-1,1} \\ a_{n-2,1} \\ a_{n-1,2} \\ a_{n-2,2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_{n1} \\ v_{n2} \end{bmatrix} \\ y_n &= (y_{n-1}, 0, y_{n-2}, 0) \begin{bmatrix} a_{n1} \\ a_{n-1,1} \\ a_{n2} \\ a_{n-1,2} \end{bmatrix} + w_n \\ \begin{bmatrix} v_{n,1} \\ v_{n,2} \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2 & 0 \\ 0 & \tau^2 \end{bmatrix}\right), \quad w_n \sim N(0, \sigma^2) \end{aligned} \quad (13.18)$$

上述状态空间模型中包含若干待估参数。其中，观测噪声  $w_n$  的方差  $\sigma^2$  和系统噪声  $v_n$  的方差  $\tau^2$  可以利用第 9 章的方法计算对数似然，从而通过最大似然法进行估计。此外，AR 阶数  $m$  和约束模型的阶数  $k$  可选取使 AIC 最小的值。一旦估计出模型参数  $m$ 、 $k$ 、 $\sigma^2$  和  $\tau^2$ ，即可应用平滑算法获得状态向量的平滑值  $x_{n|N}$ 。此时，该向量的第  $(j-1)k+1$  个分量即为  $a_{n|N}$ ，即  $a_{nj}$  的估计值。

时变系数 AR 模型的平滑计算需要  $km \times km \times N$  的数组。因此，当 AR 阶数  $m$  或数据量  $N$  非常大时，可能会出现内存不足的情况。最简单的解决方法是假设系数每隔预设的  $r$  步变化一次。若取  $r$  为 20，所需内存将减至  $\frac{1}{20}$ 。如果系数变化平滑且缓慢，这种假设对结果影响不大。即使没有内存问题，为了在下一节说明的时变谱图中便于展示，也最好使用较大的  $r$ 。要在  $r > 1$  的情况下执行卡尔曼滤波，只需每计算一次预测步骤，重复执行  $r$  次滤波步骤即可。

#### 例：地震数据的时变系数 AR 模型

R 语言的 TSSS 包中的函数 `tvar` 用于估计时变系数 AR 模型。该函数的输入包括：

<code>y</code> :	单变量时间序列
<code>trend.order</code> :	趋势阶数 (1 或 2)
<code>ar.order</code> :	AR 阶数
<code>span</code> :	局部平稳跨度 ( $r$ )
<code>outlier</code> :	异常值发生的时间点
<code>tau2.ini</code> :	系统噪声 $\tau^2$ 的初始估计值。当 <code>tau2.ini=NULL</code> 时，会在 $\tau^2 = 2^{-k}, k = 1, 2, \dots$ 范围内寻找最优值
<code>delta</code> :	搜索步长
<code>plot</code> :	逻辑变量。为 TRUE (默认) 时绘制 PARCOR 图

函数 `tvar` 的输出如下：

<code>arcoef</code> :	时变 AR 系数
<code>sigma2</code> :	观测噪声方差 $\sigma^2$
<code>tau2</code> :	系统噪声方差 $\tau^2$
<code>llkhood</code> :	模型的对数似然
<code>aic</code> :	模型的 AIC
<code>parcor</code> :	时变偏自相关系数 (时变 PARCOR)

```

> data(MYE1F)
> # 阶数 k=1
> z <- tvar( MYE1F, trend.order = 1, ar.order = 8, span = 20,
tau2.ini = 1.0e-03, delta = 1.0e-04 )
> # 阶数 k=2
> z <- tvar( MYE1F, trend.order = 2, ar.order = 8, span = 20,
tau2.ini = 6.6e-06, delta = 1.0e-06 )

```

表 13.1 地震波数据的时变系数 AR 模型的 AIC

$m$	$k=1$	$k=2$	$m$	$k=1$	$k=2$
1	6536.6	6520.4	6	4884.1	4873.8
2	5568.6	5643.2	7	4874.8	4878.7
3	5113.7	5134.5	8	4859.5	4866.9
4	4866.1	4853.9	9	4868.3	4884.9
5	4894.2	4886.0	10	4881.2	4911.9

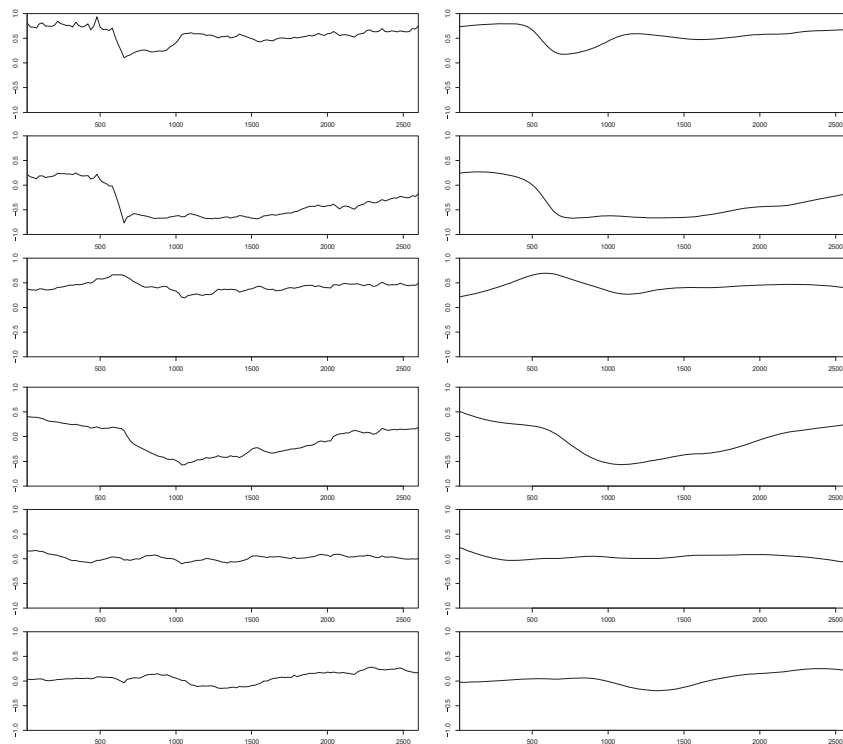


图 13.2 标准化地震波数据的时变 PARCOR 估计值。仅显示前 6 个。左图： $k=1$ ，右图： $k=2$ 。横轴：时刻，纵轴：时变 PARCOR

表 13.1 汇总了将  $r=20$  的各种阶数时变系数 AR 模型拟合至图 13.1 标准化地震波数据时的 AIC 值。由于时变系数 AR 模型假设方差  $\sigma^2$  是恒定的，因此当方差

显著变化 (如图 1.1(f) 所示) 时, 最好使用如图 13.1 所示变换为方差近似恒定的数据。\$m\$ 是 AR 阶数, \$k\$ 是差分阶数。无论 \$k=1\$ 还是 \$k=2\$, 都在 \$m=8\$ 时 AIC 最小。

图 13.2 展示了在 \$k=1\$ 和 \$k=2\$ 两种情况下, 采用 AIC 取得最小值时对应阶数的时变系数 AR 模型, 对各系数随时间变化的估计结果。需要指出的是, 这里并未直接给出 AR 系数本身, 而是仅显示了时变 PARCOR \$b\_{ni}\$ 中 \$i=1\sim 6\$ 的变化情况。图的左侧对应 \$k=1\$ 的情形, 右侧对应 \$k=2\$ 的情形。基于局部平稳 AR 模型的分析结果表明, 在 \$n=630\$ 时刻 P 波到达, 在 \$n=1026\$ 时刻 S 波到达, 可以看到与此相对应, 各系数也发生了变化。由 \$k=1\$ 得到的估计值波动较为明显。另一方面, 由 \$k=2\$ 得到的估计值较为平滑,乍看之下似乎更为理想, 但实际上并未能充分反映地震波到达所引起的急剧变化。根据 AIC 的判别结果, 可以认为 \$k=1\$ 的模型更为合适。关于既能够刻画系数的急剧变化, 又能获得变动较小的估计值的方法, 将在 13.5 节中加以说明。

### 13.3 时变谱的估计

对于平稳 AR 模型, 其功率谱由下式给出:

$$p(f) = \frac{\sigma^2}{|1 - \sum_{j=1}^m a_j e^{-2\pi i j f}|^2}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2} \quad (13.19)$$

因此, 对于 AR 系数作为时刻 \$n\$ 的函数 \$a\_{nj}\$ 给出的时变系数 AR 模型 (13.9), 可以将时刻 \$n\$ 的瞬时谱定义为:

$$p_n(f) = \frac{\sigma^2}{|1 - \sum_{j=1}^m a_{nj} e^{-2\pi i j f}|^2}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2} \quad (13.20)$$

利用前述方法, 可以将 AR 系数 \$a\_{nj}\$ 估计为时间的函数, 因此功率谱 \$\{p\_n(f)\}\$ (\$n=1, \dots, N\$) 也可以视为时间的函数。这被称为**时变谱** (time-varying spectrum)。

然而, 在实际现象中, 方差的变化和波形的变化往往同时发生。因此, 在时变谱估计中, 首先估计时变方差 \$\hat{\sigma}\_n^2\$, 并通过

$$\tilde{y}_n = \frac{y_n}{\sqrt{\hat{\sigma}_n^2}} \quad (13.21)$$

将数据转换为标准化数据, 再估计时变系数 AR 模型。此时, 原始时间序列的时变谱计算公式为:

$$p_n(f) = \frac{\hat{\sigma}_n^2 \hat{\sigma}_n^2}{|1 - \sum_{j=1}^m \hat{a}_{nj} e^{-2\pi i j f}|^2}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2} \quad (13.22)$$

**例: 地震数据的时变谱**

R 语言包 TSSS 中的函数 `tvspec` 利用 `tvar` 的输出计算时变谱。该函数的输入包括：

`arcoef`: 时变 AR 系数  
`sigma2`: 观测噪声方差  
`span`: 局部平稳跨度  
`nf`: 计算频谱的频率点数

```
> data( MYE1F )
> par( mar=c(2,2,1,1) )
> # k=1 的情况
> z <- tvar( vnordata,trend.order = 1,ar.order = 8,span = 20,
tau2.ini = 5.0e-03,delta = 1.0e-04,plot = FALSE)
> spec <- -tvspec(zarcoef, zsigma2,span = 20,nf = 400)
> plot(spec,tvv = vtvv, dx = 2, dy = 0.10 )
>
> # k=2 的情况
> z <- tvar(vnordata,trend.order = 2,ar.order = 8,span = 20,
tau2.ini = 4.4e-05,delta = 1.0e-07,plot = FALSE)
> spec <- -tvspec(zarcoef, zsigma2,span = 20,nf = 400)
> plot(spec,tvv = vtvv, dx = 2, dy = 0.10)
```

图 13.3 给出了根据式 (13.22) 计算得到的时变谱。左侧为  $k=1$  模型的估计结果，右侧为  $k=2$  模型的估计结果。横轴表示频率，纵轴表示谱的对数，斜向表示时间的推移。可以看到，随着 P 波的到达， $f=0.25$  附近的功率逐渐增强；随后 S 波到达时， $f=0.1$  附近的功率进一步增强。此后，谱峰位置在逐渐向右移动的同时，逐步回到原先常微动的谱形态。

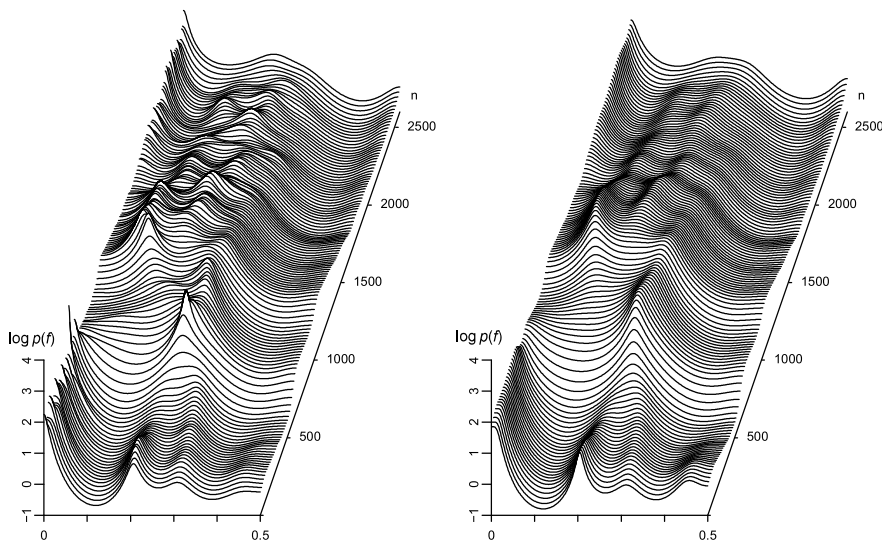


图 13.3 地震数据的时变谱。左图： $k=1$ ，右图： $k=2$

### 13.4 时变系数 AR 模型的系统噪声假设

在第 13.2 节中, 我们展示了通过趋势成分模型表达 AR 系数的时间变化, 并利用状态空间模型估计时变系数 AR 模型。在式 (13.17) 的状态空间模型中, 假设系统噪声的方差协方差矩阵为  $Q = \text{diag}\{\tau^2, \dots, \tau^2\}$  形式的对角矩阵。虽然这看起来是非常强的假设, 但本节将通过考虑 AR 算子的频率响应函数的平滑性, 来说明该假设可以自然地导出。首先, 考虑时变 AR 系数的傅里叶变换:

$$A(f, n) = 1 - \sum_{j=1}^m a_{nj} e^{-2\pi i j f}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2} \quad (13.23)$$

这相当于将 AR 模型视为白化滤波器时的频率响应函数。此时, 式 (13.20) 的时变谱可表示为:

$$p_n(f) = \frac{\sigma^2}{|A(f, n)|^2} \quad (13.24)$$

频谱的特性由  $A(f, n)$  决定, 因此通过适当控制  $A(f, n)$  的平滑度, 可以获得良好的时变谱估计值。在此, 通过考虑  $A(f, n)$  随时刻  $n$  变化的平滑性, 可以引入如下约束模型。为了评价  $A(f, n)$  随时间的阶跃变化, 取关于时刻  $n$  的  $k$  阶差分:

$$\Delta^k A(f, n) = \sum_{j=1}^m \Delta^k a_{nj} e^{-2\pi i j f} \quad (13.25)$$

考虑平方积分:

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} |\Delta^k A(f, n)|^2 df = \sum_{j=1}^m (\Delta^k a_{jn})^2 \quad (13.26)$$

因此, 通过减小 AR 系数  $k$  阶差分的平方和, 可以减小频谱随时间的变化。由于式 (13.26) 对各项赋予了相等的权重, 这相当于假设了使用共同的  $\tau^2$  的模型:

$$\Delta_n^k a_{jn} = v_{nj}, \quad v_{nj} \sim N(0, \tau^2), \quad j = 1, \dots, m \quad (13.27)$$

即在式 (13.10) 中取  $\tau_{11}^2 = \dots = \tau_{mm}^2 = \tau^2$  的假设是自然导出的。

### 13.5 关于系数的剧烈变化

如图 13.1 所示的经过标准化处理的地震数据那样, 当在某一时刻有新的信号加入并导致波形发生急剧变化时, 如图 13.2 所示, 所估计得到的时变系数往往会呈现较大的离散, 或者无法充分对应这种急剧变化。在这种情况下, 若采用第 8 章介

绍的局部平稳 AR 模型, 即可得到变化点的估计值, 利用这些信息有时能够获得更为理想的时变系数估计结果。

假设在时刻  $n = n_1, \dots, n_p$  处存在急剧变化。此时, 式 (13.10) 中的误差项  $v_{nj}$  的绝对值将取非常大的值, 因此有必要仅在  $n = n_1, \dots, n_p$  这些时刻增大系统噪声的方差  $\tau^2$ 。严格来说, 通过增大  $\tau^2$ , 会使得  $a_{nj}$  的  $k$  阶差分变化幅度增大。因此, 在  $k = 1$  的情况下,  $a_{nj}$  将呈现阶梯状变化, 并形成不连续点; 而在  $k = 2$  的情况下, 则表现为斜率发生急剧变化的折点。若要在  $k \geq 2$  的情形下产生不连续点, 则需要对整个状态向量引入噪声。为此, 可以对  $x_{n|n-1}$  与  $V_{n|n-1}$  进行初始化, 或在  $V_{n|n-1}$  的所有对角元上加上较大的值。该方法不仅适用于时变系数的估计, 同样也可以应用于趋势等的估计问题。

#### 例: 伴随突发结构变化的时变谱估计

在函数 `tvspec` 中处理突发结构变化时, 需要利用参数 `outlier` 指定变点位置:

```
> data(MYE1F)
> z <- tvar( vnordata, trend.order = 2, ar.order = 8, span = 20,
  outlier = c(630, 1026), tau2.ini = 6.6e - 06, delta = 1.0e - 06,
  plot = FALSE)
> spec <- tvspec(zarcoef, zsigma2, span = 20, nf = 400)
> plot(spec, tvv = vtvv, dx = 2, dy = 0.10 )
```

图 13.4 给出了在  $k = 2$ 、 $m = 8$  的设定下, 并假定在  $n = 630$  和  $n = 1026$  处存在急剧变化时, 所估计得到的时变系数及其对应的时变谱。可以看到, 系数的急剧变化仅在与 P 波和 S 波到达时刻相对应的这两个位置发生显著变化, 而在其他时刻则呈现出平滑的变化趋势。此外, 在 S 波对应的区间内, 一方面主导频率随时间逐渐减弱, 另一方面直流分量 ( $f = 0$ ) 逐步增强, 清楚地刻画了谱形逐渐向常时微动谱过渡的过程。

#### ● 章末习题

1. 请思考除地震波外, 方差随时间变化的序列示例。
2. 请思考估计系数随时间变化的回归模型的方法。

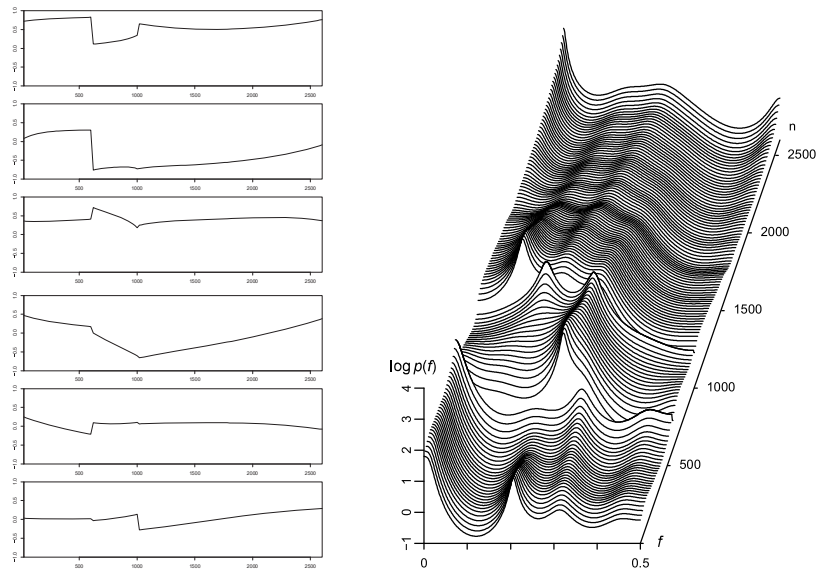


图 13.4 在参数变化中加入不连续点估计出的时变 PARCOR (左) 与时变谱 (右)

## 第 14 章

# 非高斯型模型

本章将状态空间模型扩展到系统噪声或观测噪声为非高斯型的情形。非高斯型模型适用于处理由于系统结构变化而引起的参数急剧变化，以及异常值的处理等问题。在非高斯型模型的情况下，卡尔曼滤波和平滑算法往往不能得到好的结果。然而，对于非高斯型状态空间模型，也可以像卡尔曼滤波那样导出递推公式，并借助数值计算实现滤波和平滑算法。本章还将介绍非高斯型状态空间模型的各种应用，以及向非线性状态空间模型的扩展。

### 14.1 非高斯型模型的必要性

正如前章中的例子所示，许多时间序列模型都可以利用状态向量  $x_n$ ，表示为线性高斯型状态空间模型

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \\ y_n &= Hx_n + w_n \end{aligned} \quad (14.0)$$

的形式。该状态空间模型作为时间序列模型非常自然，而且可以利用卡尔曼滤波和平滑算法等高效的计算方法，因此可以灵活地构造时间序列模型。不过，尽管基于状态空间表示的时间序列建模极为方便，但并不是所有时间序列都能用线性高斯型状态空间模型来建模。例如，在概率结构随时间变化的非平稳时间序列中，这种变化方式往往同时包含缓慢平滑的变化和急剧的变化。在这种情况下，通常的线性高斯型状态空间模型虽然能够较好地表示缓慢平滑的变化，但若恰当地表示急剧变化，则必须构造更加复杂的模型。此外，为了消除数据中经常出现的异常值的影响，还需要引入异常值的自动检测方法或稳健估计方法。再者，包含非线性的系统或离散过程，也都是通常状态空间模型难以妥善处理的例子。

下面考虑如何解决这些问题。在状态空间模型中，概率结构的变化往往对应于

状态的变化。如果对系统噪声  $v_n$  假设如柯西分布那样的重尾分布, 就可以用一个分布同时表示高概率发生的平滑变化以及极小概率发生的急剧变化。同样地, 为了处理异常值, 也可以对观测噪声  $w_n$  使用重尾分布。此外, 当系统存在非线性, 或者观测值服从离散分布时, 状态的分布必然成为非高斯型。像这样, 为了解决传统状态空间模型难以处理的各种问题, 就有必要引入非高斯性。以下各节将说明针对非高斯型状态空间模型的滤波和平滑方法及其应用。

## 14.2 非高斯型状态空间模型与状态估计

考虑如下的状态空间模型:

$$x_n = Fx_{n-1} + Gv_n, \quad (\text{系统模型}) \quad (14.1)$$

$$y_n = Hx_n + w_n, \quad (\text{观测模型}) \quad (14.2)$$

系统噪声  $v_n$  和观测噪声  $w_n$  分别是服从密度函数  $q(v)$  与  $r(w)$  的白噪声, 但与前章中的状态空间模型不同, 它们不一定服从正态分布。在这种情况下, 状态向量  $x_n$  一般也将成为非高斯型。像这样, 不再假定高斯分布时, 模型 (14.1)、(14.2) 称为**非高斯型状态空间模型** (non-Gaussian state-space model)。显然, 这种非高斯型模型是通常状态空间模型的扩展。

与第 9 章相同, 记截至时刻  $j$  为止所观测到的时间序列信息为  $Y_j \equiv \{y_1, \dots, y_j\}$ 。此外, 将截至时刻  $j$  的状态  $x_n$  的实现值集合记为  $X_j \equiv \{x_1, \dots, x_j\}$ 。再进一步, 设初始向量  $x_0$  服从密度函数  $p(x_0|Y_0)$ 。此时, 状态估计问题就是求  $p(x_n|Y_m)$ , 也就是在信息  $Y_m$  给定时状态  $x_n$  的条件分布。特别地, 当  $n > m$ 、 $n = m$ 、 $n < m$  时, 分别称为预测、滤波和平滑化。

卡尔曼滤波给出的是状态的均值和方差协方差矩阵, 但对于非高斯型状态空间模型, 则需要直接计算分布。然而, 即使在这种情况下, 对于状态空间模型 (14.1)、(14.2), 利用  $p(x_n|x_{n-1}, Y_{n-1}) = p(x_n|x_{n-1})$ ,  $p(y_n|x_n, Y_{n-1}) = p(y_n|x_n)$  成立这一事实, 仍可按如下方式递推地求得预测分布与滤波分布。

[一步预测]

$$\begin{aligned} p(x_n|Y_{n-1}) &= \int_{-\infty}^{\infty} p(x_n, x_{n-1}|Y_{n-1}) dx_{n-1} \\ &= \int_{-\infty}^{\infty} p(x_n|x_{n-1}, Y_{n-1}) p(x_{n-1}|Y_{n-1}) dx_{n-1} \\ &= \int_{-\infty}^{\infty} p(x_n|x_{n-1}) p(x_{n-1}|Y_{n-1}) dx_{n-1} \end{aligned} \quad (14.1)$$

[滤波]

$$\begin{aligned}
 p(x_n|Y_n) &= p(x_n|y_n, Y_{n-1}) \\
 &= \frac{p(y_n|x_n, Y_{n-1})p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})} \\
 &= \frac{p(y_n|x_n)p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})} \quad (14.0)
 \end{aligned}$$

其中,  $p(y_n|Y_{n-1})$  由  $\int p(y_n|x_n)p(x_n|Y_{n-1})dx_n$  求得。一步预测公式 (14.2) 是卡尔曼滤波中一步预测的扩展。 $p(x_n|x_{n-1})$  是在状态  $x_{n-1}$  给定时  $x_n$  的密度函数, 它由系统模型 (14.1) 决定。因此, 只要给定  $x_{n-1}$  的滤波分布  $p(x_{n-1}|Y_{n-1})$ , 就可以计算  $p(x_n|Y_{n-1})$ 。另一方面, 滤波公式 (14.0) 是卡尔曼滤波中滤波步骤的扩展。 $p(y_n|x_n)$  是在状态  $x_n$  给定时  $y_n$  的密度函数, 由观测模型 (14.2) 决定。因此, 只要给定  $x_n$  的预测分布  $p(x_n|Y_{n-1})$ , 就可以计算  $p(x_n|Y_n)$ 。

接下来考虑平滑化问题。对于状态空间模型 (14.1)、(14.2), 利用  $p(x_n|x_{n+1}, Y_N) = p(x_n|x_{n+1}, Y_n)$  成立这一事实, 可得

$$\begin{aligned}
 p(x_n, x_{n+1}|Y_N) &= p(x_{n+1}|Y_N)p(x_n|x_{n+1}, Y_N) \\
 &= p(x_{n+1}|Y_N)p(x_n|x_{n+1}, Y_n) \\
 &= p(x_{n+1}|Y_N) \frac{p(x_n|Y_n)p(x_{n+1}|x_n, Y_n)}{p(x_{n+1}|Y_n)} \\
 &= p(x_{n+1}|Y_N) \frac{p(x_n|Y_n)p(x_{n+1}|x_n)}{p(x_{n+1}|Y_n)} \quad (14.-2)
 \end{aligned}$$

由此, 关于平滑化问题也可以得到如下的递推公式。

[平滑化公式]

$$\begin{aligned}
 p(x_n|Y_N) &= \int_{-\infty}^{\infty} p(x_n, x_{n+1}|Y_N)dx_{n+1} \\
 &= p(x_n|Y_n) \int_{-\infty}^{\infty} \frac{p(x_{n+1}|Y_N)p(x_{n+1}|x_n)}{p(x_{n+1}|Y_n)} dx_{n+1} \quad (14.-2)
 \end{aligned}$$

在 (14.-2) 的右边中,  $p(x_{n+1}|x_n)$  由系统模型 (14.1) 决定。另外,  $p(x_n|Y_n)$  和  $p(x_{n+1}|Y_n)$  则由 (14.2)、(14.0) 求得。因此, (14.-2) 表明: 只要给定  $p(x_{n+1}|Y_N)$ , 就可以计算  $p(x_n|Y_N)$ 。又考虑到  $p(x_N|Y_N)$  可以通过 (14.0) 计算, 因此与第 9 章中的平滑化相同, 通过依次应用 (14.-2), 就能求得平滑分布  $p(x_{N-1}|Y_N), \dots, p(x_1|Y_N)$ 。

### 14.3 状态估计公式的数值实现

如上所述, 即使对于非高斯型状态空间模型, 也可以导出扩展卡尔曼滤波形式的递推估计公式。利用这些公式, 就能够统一处理广泛的时间序列模型。但问题在于, 滤波和平滑化的公式在实际中应如何计算。

对于线性高斯型状态空间模型, 条件分布  $p(x_n|Y_{n-1})$ 、 $p(x_n|Y_n)$  以及  $p(x_n|Y_N)$  全都是正态分布, 因此只需评价其均值和方差协方差矩阵即可, 而 (14.2)、(14.0) 以及 (14.-2) 就与通常的卡尔曼滤波和平滑算法等价。然而, 一般情况下, 状态的条件分布  $p(x_n|Y_j)$  是非高斯型分布, 因此仅靠均值和方差协方差矩阵无法确定该分布。另一方面, 将这些分布用适当调节均值向量和方差协方差矩阵的高斯分布来近似的方法, 如扩展卡尔曼滤波和二次滤波等, 虽已为人所知, 但并不总是能得到好的结果。

本节将说明一种通过直接对非高斯型分布进行数值近似, 从而导出非高斯型滤波与非高斯型平滑算法的方法。近似的方法可以考虑阶梯函数近似、折线近似、样条近似等。此时, (14.2)~(14.-2) 就可以借助数值计算来实现。这类数值方法过去被认为由于需要大规模数值计算而不切实际, 但随着计算机速度的提高, 如今至少对于低维问题而言, 它已成为一种极其有力的方法。

本节中, 将采用最简单的阶梯函数来近似出现在 (14.2)、(14.0)、(14.-2) 中的密度函数。具体地, 设待近似的密度函数为  $f(t)$  ( $-\infty < t < \infty$ )。为了用阶梯函数来近似该函数, 首先将定义域划分为  $d$  个小区间:  $t_0 < t_1 < \dots < t_d$ 。其中,  $t_0$  取为足够小的数,  $t_d$  取为足够大的数; 并且为简便起见, 假设分点等间隔, 即有  $t_i = t_0 + i\Delta t$ 。另外, 在实际计算程序中, 通常会随着时刻  $n$  的变化, 对  $t_0$  和  $t_i$  的值进行自适应平移, 但为了简便, 这里假设所有分点都是固定不变的。记如此定义的  $d$  个区间  $(t_{i-1}, t_i]$  上  $f(t)$  的值为  $f_i$ 。当  $f(t)$  是阶梯函数时,  $f_i = f(t_i)$ 。一般情况下,  $f_i$  可由

$$f_i = \Delta t \int_{t_{i-1}}^{t_i} f(t) dt \quad (14.-2)$$

求得。利用这些值,  $f(t)$  的近似函数可由  $\{d; t_0, \dots, t_d; f_1, \dots, f_d\}$  来表示。以下将这样得到的近似函数记为  $\tilde{f}(t)$ 。

为了数值实现非高斯型滤波与平滑公式, 如表 14.1 所示, 需要对  $p(x_n|Y_{n-1})$ 、 $p(x_n|Y_n)$ 、 $p(x_n|Y_N)$  以及系统噪声的密度函数  $q(v)$  进行近似。观测噪声的密度函数  $r(v)$  则最好不离散化, 而是直接用于计算。

以下为简便起见, 以一维趋势模型

$$\begin{aligned} x_n &= x_{n-1} + v_n \\ y_n &= x_n + w_n \end{aligned} \quad (14.-2)$$

为例说明数值计算方法。

表 14.1 密度函数的近似

密度函数	近似	记号
$p(x_n Y_{n-1})$	$\{d; t_0, \dots, t_d; p_1, \dots, p_d\}$	$\tilde{p}(t)$
$p(x_n Y_n)$	$\{d; t_0, \dots, t_d; f_1, \dots, f_d\}$	$\tilde{f}(t)$
$p(x_n Y_N)$	$\{d; t_0, \dots, t_d; s_1, \dots, s_d\}$	$\tilde{s}(t)$
$q(v)$	$\{2d+1; t_{-d}, \dots, t_d; q_{-d}, \dots, q_d\}$	$\tilde{q}(v)$

[一步预测] 对  $i = 1, \dots, d$

$$\begin{aligned}
 p_i &= \tilde{p}(t_i) = \int_{t_0}^{t_d} \tilde{q}(t_i - s) \tilde{f}(s) ds \\
 &= \sum_{j=1}^d \int_{t_{j-1}}^{t_j} \tilde{q}(t_i - s) \tilde{f}(s) ds \\
 &= \Delta t \sum_{j=1}^d q_{i-j} f_j
 \end{aligned} \tag{14.-3}$$

[滤波] 对  $i = 1, \dots, d$

$$f_i = \tilde{f}(t_i) = \frac{r(y_n - t_i) \tilde{p}(t_i)}{C} = \frac{r(y_n - t_i) p_i}{C} \tag{14.-2}$$

其中,  $C$  可由

$$\begin{aligned}
 C &= \int_{t_0}^{t_d} r(y_n - t) \tilde{p}(t) dt = \sum_{j=1}^d \int_{t_{j-1}}^{t_j} r(y_n - t) \tilde{p}(t) dt \\
 &= \Delta t \sum_{j=1}^d r(y_n - t_j) p_j
 \end{aligned} \tag{14.-2}$$

计算得到。

[平滑化] 对  $i = 1, \dots, d$

$$\begin{aligned}
 s_i &= \tilde{s}(t_i) = \tilde{f}(t_i) \int_{t_0}^{t_d} \frac{\tilde{q}(t_i - u) \tilde{s}(u)}{\tilde{p}(u)} du \\
 &= \tilde{f}(t_i) \sum_{j=1}^d \int_{t_{j-1}}^{t_j} \frac{\tilde{q}(t_i - u) \tilde{s}(u)}{\tilde{p}(u)} du \\
 &= \Delta t \cdot f(t_i) \sum_{j=1}^d \frac{q_{i-j} s_j}{p_j}
 \end{aligned} \tag{14.-3}$$

不过, 在实际计算中, 预测和平滑化步骤之后, 还需要对近似得到的密度函数进行归一化, 使其在整个区间上的积分值为 1。例如, 对于  $f_i$ , 需要计算

$$I(f) = \int_{t_0}^{t_d} f(t) dt = \Delta t (f_1 + \dots + f_d) \tag{14.-3}$$

并将其修正为  $f_i \equiv \frac{f_i}{I(f)}$ 。

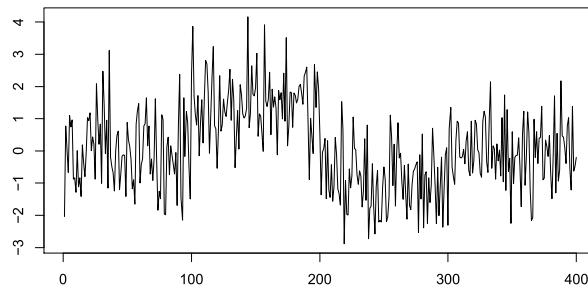


图 14.1 由 (14.15) 生成的数据

## 14.4 非高斯型趋势模型

在第 11 章的趋势模型中, 如果假定系统噪声  $v_n$  或观测噪声  $w_n$  服从非高斯型分布, 就可以得到非高斯型趋势模型 (non-Gaussian trend model)。本节将结合数值例子说明非高斯型模型的特点。

### 例：非高斯型趋势估计

R 包 TSSS 中的函数 `ngsmth`, 针对具有非高斯型噪声分布的一维状态空间模型, 执行非高斯型滤波和平滑化。该函数的输入为

**y:** 一元时间序列。  
**noisev:** 系统噪声密度函数的类型  
     1: 高斯 (正态) 分布  
     2: Pearson VII 型分布族  
     3: 双边指数分布  
**tau2** 系统分布的散布 (方差) 参数  
**bv:** 系统噪声分布的形状参数  
**noisew:** 观测噪声的分布类型  
     1: 高斯 (正态) 分布  
     2: Pearson VII 型分布族  
     3: 双边指数分布  
     4: 双重指数分布  
**sigma2:** 观测噪声的散布 (方差) 参数  
**bw:** 观测噪声的形状参数  
**initd:** 初始状态的分布类型  
     1: 高斯 (正态) 分布  
     2: 均匀分布  
     3: 双边指数分布  
**k:** 数值积分的分点数  
**plot:** 逻辑变量。若为 TRUE (默认值), 则将 **trend** 和 **smt** 作图显示

该函数的输出为:

**trend:** 趋势平滑后验分布的 7 个百分位点  
**smt:** 时刻  $n = 1, \dots, N$  的平滑密度函数

```

> y <- data( PfilterSample )
> par(mar=c(3,3,1,1)+0.1)
> #
> # 系统噪声: 高斯型 (正态分布)
> s1 <- ngsmth( y, noisev = 1, tau2 = 1.4e-02, noisew = 1, sigma2 = 1.048 )
> # 3D plot
> plot( s1, "smt", theta = 25, phi = 30, expand = 0.25, col="white" )
> #
> # 系统噪声: Pearson 分布族 (b=0.6)
> s2 <- ngsmth( y, noisev = 2, tau2 = 2.11e-10, bv = 0.6,
noisew = 1, sigma2 = 1.042 )
> plot( s2, "smt", theta = 25, phi = 30, expand = 0.25, col="white" )
> #
> # 系统噪声: 柯西分布 (Pearson 分布族 b=1)
> s2 <- ngsmth( y, noisev = 2, tau2 = 3.53e-5, bv = 1.0,
noisew = 1, sigma2 = 1.045 )

```

图 14.1 给出了按如下模型生成的数据。

$$y_n \sim N(\mu_n, 1), \quad \mu_n = \begin{cases} 0, & 1 \leq n \leq 100 \\ 1, & 101 \leq n \leq 200 \\ -1, & 201 \leq n \leq 300 \\ 0, & 301 \leq n \leq 400 \end{cases} \quad (14.-3)$$

在中途 3 处平均值发生了变化。为了估计这种伴随结构变化而发生变化的数据均值函数，这里考虑如下 1 阶趋势模型。

$$\begin{aligned} t_n &= t_{n-1} + v_n \\ y_n &= t_n + w_n \end{aligned} \quad (14.-3)$$

其中，对观测噪声  $w_n$  假定为正态分布，而对系统噪声  $v_n$ ，作为非高斯分布的例子，考虑 Pearson VII 型分布族

$$q(v_n) = \frac{C}{(\tau^2 + v_n^2)^b}, \quad \frac{1}{2} < b < \infty \quad (14.-3)$$

这里  $C$  是使  $q(v)$  在全区间上的积分为 1 的归一化常数，其值为  $C = \tau^{2b-1} \Gamma(b) / \Gamma(\frac{1}{2}) \Gamma(b - \frac{1}{2})$ 。Pearson 分布族在  $b = 1$  时为柯西分布，在  $b = (k+1)/2$  时为自由度为  $k$  的  $t$  分布，而在  $b \rightarrow \infty$  时则趋于正态分布。因此，通过改变  $b$  的取值，可以表示各种左右对称且比正态分布尾部更重的密度函数。

表 14.2 汇总了当  $b = \frac{3}{k}$  ( $k = 0, \dots, 5$ ) 所表示的 6 个  $b$  的取值对应的 Pearson 分布族下，参数的极大似然估计值、AIC 值等结果。当  $b = 0.60$  时，AIC 最小，而正态分布模型 ( $b = \infty$ ) 时 AIC 最大。

表 14.2 非高斯型模型

$b$	$\tau^2$	$\sigma^2$	对数似然	AIC
0.60	$0.211 \times 10^{-9}$	1.042	-597.19	1198.38
0.75	$0.299 \times 10^{-7}$	1.043	-597.39	1198.78
1.00	$0.353 \times 10^{-4}$	1.045	-597.99	1198.98
1.50	$0.303 \times 10^{-2}$	1.045	-599.13	1202.26
3.00	$0.406 \times 10^{-1}$	1.046	-600.40	1204.80
$\infty$	$0.140 \times 10^{-1}$	1.048	-600.69	1205.38

图 14.2 画出了高斯型模型 ( $b = \infty$ ) 和非高斯型模型 ( $b = 0.6$ ) 得到的趋势  $t_n$  的平滑分布  $p(t_n|Y_N)$  随时间变化的情形。对于高斯型模型，随着时刻  $n$  的推移，分布缓慢地向左右变动；而对于非高斯型模型，分布极其稳定，仅在 3 处发生了位置变化。图 14.3 左侧给出了在各时刻采用高斯型模型所估计的分布均值以及  $\pm 1, 2, 3 \times$  (标准差)。可以看到趋势出现了较大的起伏，并且无法很好地表示图 14.1 中明显可见

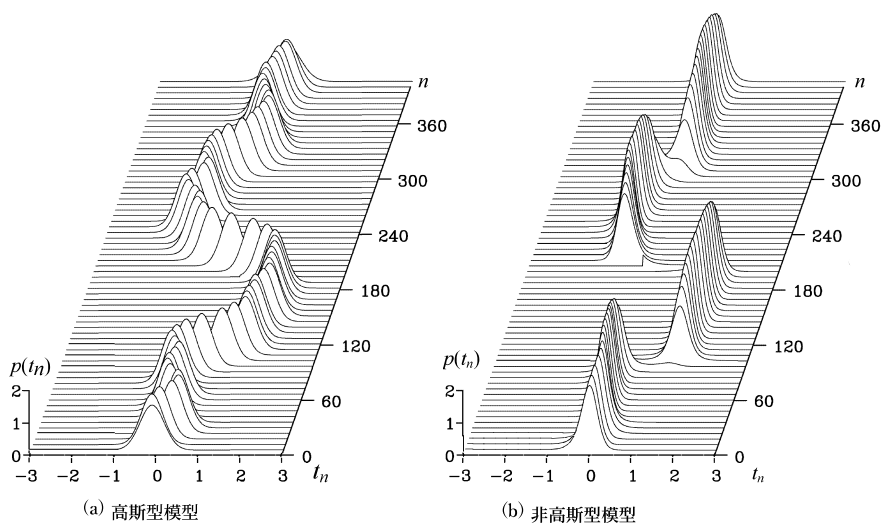


图 14.2 趋势平滑后验分布随时间的变化

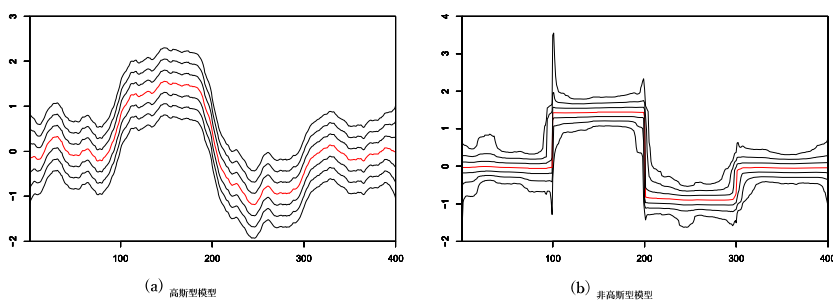


图 14.3 利用高斯型模型与非高斯型模型进行趋势估计

的跳变。另一方面，图 14.3 右侧的 7 条曲线是在非高斯型模型下，对应于高斯型分布中位数以及  $\pm 1 \sim \pm 3 \times (\text{标准差})$  的 0.13, 2.27, 15.87, 50.0, 84.13, 97.73, 99.87 百分位点。比较左右两图可以看出，采用非高斯型模型得到的估计值要平滑得多，而且对跳变的检测效果也很好。

## 14.5 非对称分布——时变方差模型

现实中得到的数据中，有些由于其生成机制的特性或某种非线性变换的结果，会呈现非对称分布。对于这种数据，如果假定通常的模型并应用最小二乘法等，可能会得到严重偏倚的结果。当观测噪声的分布已知时，直接使用其分布形式  $r(w)$  进行平滑化，就能得到较好的结果。

本节考虑 13.1 节中用于估计时变方差的模型

$$\begin{aligned} t_m &= t_{m-1} + v_m \\ y_m &= t_m + w_m \end{aligned} \quad (14.-3)$$

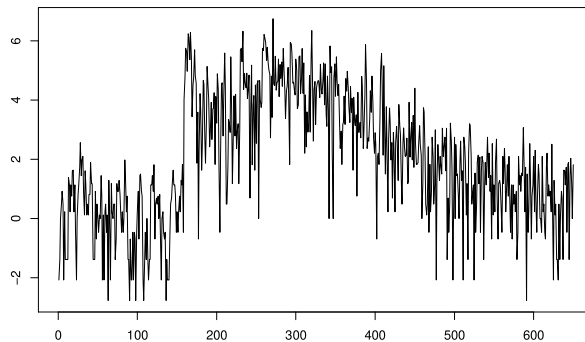


图 14.4 变换后的地震波数据

在第 13 章中，曾用正态分布  $N(-\zeta, \frac{\pi^2}{6})$  近似观测噪声  $w_n$  的双重指数分布；本节将比较两种情形：一种是用正态分布近似，另一种是利用精确密度函数进行数值计算。

图 14.4 是对图 1.1(f) 中的地震波数据做了滤波变换  $\tilde{y}_n = y_n - 0.5y_{n-1}$  之后，每隔一点取一个值，并进一步按照 13.1 节的方法进行变换所得的数据。进行这样的预处理，是因为本节的方法假定白噪声，并且是为了减弱数据前半部分所见到的强自相关。

对该数据，将  $\sigma^2 = \frac{\pi^2}{6}$  固定，并用极大似然法估计高斯型模型，可得  $\hat{\tau}^2 = 0.04908$ ， $AIC = 2541.36$ 。图 14.5 中给出了用这一模型估计的趋势，可以看到其上下起伏波动。这是因为图 14.4 中所见到的向下较大的

```

> ## 地震波数据
> data(MYE1F)
> # 数据变换
> n <- length( MYE1F )
> yy <- rep(0, n)
> for (i in 2:n) yy[i] <- MYE1F[i] - 0.5 * MYE1F[i-1]
> m <- seq(1, n, by = 2)
> y <- yy[m]
> z <- tvvar( y, trend.order = 2, tau2.ini = 4.909e-02, delta = 1.0e-06)
>
> # 系统噪声：正态分布，观测噪声：正态分布
> s3 <- ngsmth( z$sm, noisew = 1, tau2 = z$tau2, noisew = 2, sigma2 = pi*pi/6, k =
190 )
>
> # 系统噪声：柯西分布，观测噪声：双重指数分布
> s3 <- ngsmth( z$sm, noisew = 2, tau2 = 0.00026, noisew = 2, noizew=4,
sigma2 = pi*pi/6, k = 190 )

```

偏差作为异常值起作用，并且趋势对其敏感地作出了响应，从而产生了波动。

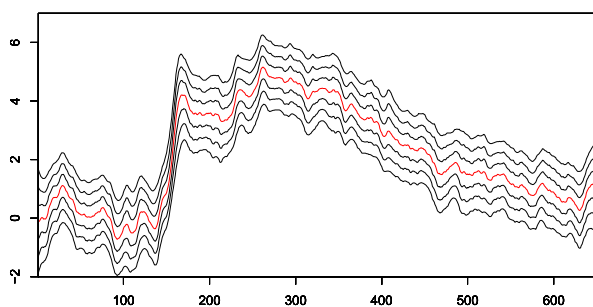


图 14.5 利用高斯型模型对趋势（方差的对数）的估计

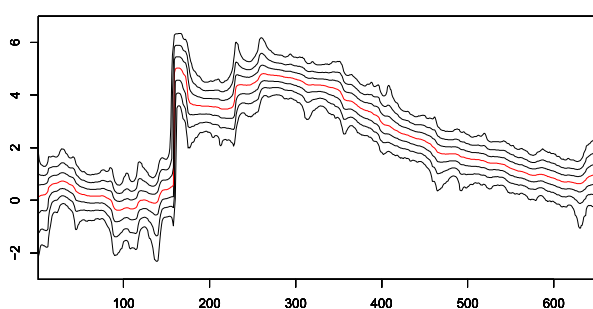


图 14.6 利用非高斯型模型对趋势（方差的对数）的估计

另一方面，当把状态空间模型 (14.-3) 的系统噪声与观测噪声的密度函数分别设为柯西分布和双重指数分布

$$\begin{aligned} q(x) &= \frac{1}{\pi} \frac{\tau}{\tau^2 + x^2} \\ r(x) &= e^{x - e^x} \end{aligned} \quad (14.-2)$$

时，非高斯型模型的结果为  $\hat{\tau}^2 = 0.00017$ ， $AIC=2517.00$ 。由 AIC 的差异可以看出，非高斯型模型比高斯型模型拟合得更好。实际上，图 14.6 中的趋势比高斯型模型下平滑得多，并且很好地表示了伴随 P 波与 S 波到达而产生的方差急剧增大。另外，在该模型中，即使得到明显向下偏离的数据，也只是最下方的 0.13 百分位曲线发生波动，而 50 百分位点并未受到很大影响。由此可见，如果采用适当的模型进行非高斯型平滑化，就能够自动检测参数的跳变，并且得到一种不受分布偏斜影响的估计方法。

**例：日经 225 平均股价数据**

```

> data( Nikkei225 )
> y <- diff( log( Nikkei225 ) )
> # 数据变换与参数估计
> z <- tvvar( y, trend.order = 2, tau2.ini = 4.909e-02, delta = 1.0e-06)
>
> # 非高斯型平滑化
> s3 <- ngsmth( z$sm, noisew = 1, tau2 = 0.07, noisew = 4, sigma2 = pi*pi/2, k =
190)
> tt <- s3$trend
>
> # 结果显示
> par( mfrow=c(3,1), mar=c(2,2,1,1) )
> plot( z$sm,type="l", ylim=c(-14,-4) )
> plot( tt[,4],type="l", ylim=c(-14,-4),lwd=2 )
> for (j in 1:7){
> lines( tt[,j],type="l", ylim=c(-14,-4) )
> }
> plot( exp(tt[,4]/2),type="l" )

```

时变方差的估计方法可用于金融时间序列中随机波动率的估计。图 14.7(a) 是对日经 225 平均股价数据的对数差分序列，按照 13.1 节的方法进行变换后的结果。对于该数据，使用非高斯型模型 (14.18)，其趋势所对应的  $\log \sigma_n^2$  的后验分布如 (b) 所示。进一步地，由 (b) 中间的曲线 (50% 点) 取一半再取指数，即

$$\exp(\log \hat{\sigma}_n^2 / 2) = \sqrt{\hat{\sigma}_n^2} = \hat{\sigma}_n$$

便可得到波动率的估计值。

这种时变方差的估计方法也可以立即应用于周期图的平滑。由于已知周期图服从自由度为 2 的  $\chi^2$  分布，因此取其对数后得到的序列就服从与 (14.19) 相同的分布。因此，本节的方法也适用于对数周期图的平滑化。

另外，利用非高斯型模型，还可以从经过

$$z_n = \log y_n^2 \quad (14.-3)$$

变换后的数据中估计时变方差。在这种情况下，只需将由自由度为 1 的  $\chi^2$  分布取对数变换后得到的密度函数

$$r(w) = \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{w - e^w}{2} \right\} \quad (14.-3)$$

作为观测噪声的密度函数即可。在这种情况下，无需作出  $\log \sigma_{2m-1}^2 = \log \sigma_{2m}^2$  这一假设，也能估计时变方差。

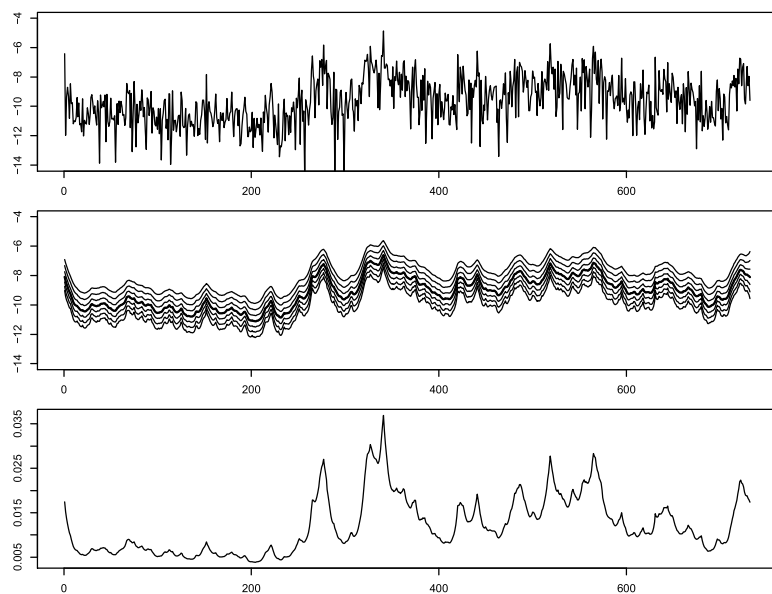


图 14.7 (a) 日经 225 数据的变换值, (b)  $\log \sigma_n^2$  的后验分布, (c) 波动率的估计值。

## 14.6 非高斯型状态空间模型的应用

本章的非高斯型状态空间模型具有极其广泛的应用范围。本节将针对若干情形说明非高斯型状态空间模型的使用方法。

### 14.6.1 利用混合高斯分布处理异常值

可以认为异常值服从与通常观测噪声分布不同的另一种分布。为了处理时间序列中的异常值,可以假定异常值以概率  $\alpha \ll 1$  出现,并将观测噪声  $w_n$  的分布设为混合分布

$$r(w) = (1 - \alpha)\varphi(w; \mu, \sigma^2) + \alpha\varphi(w; \xi, \delta^2) \quad (14.-2)$$

其中,  $\varphi(x; \mu, \sigma^2)$  表示均值为  $\mu$ 、方差为  $\sigma^2$  的正态分布的密度函数。右边第 1 项对应正常观测值的分布,第 2 项对应异常值的分布。特别地,当  $\xi = \mu = 0$ 、 $\delta^2 \gg \sigma^2$  时,异常值可以看作均值为 0 但具有很大方差噪声的观测值。柯西分布等可以视为对此类混合分布的近似。因此,也可以通过将柯西分布或 Pearson 分布族等用作观测噪声的密度函数来处理异常值。

另一方面,当像图 1.1(h) 中的地下水位数据那样,异常总是向上出现时,可以令  $\xi \gg \mu$ 、 $\delta^2 = \sigma^2$ ,从而在忽略这些异常值的同时估计趋势。

### 14.6.2 非平稳离散序列

非高斯型模型也可以应用于非平稳二项分布和泊松分布等离散过程的平滑化。下面考虑取整数值的时间序列  $\{m_1, \dots, m_N\}$ 。设该时间序列在第  $n$  期取值  $m_n$  的概率服从泊松分布

$$P(m_n|p_n) = \frac{e^{-p_n} p_n^{m_n}}{m_n!} \quad (14.-2)$$

其中,  $p_n$  表示某一事件在单位时间内发生的平均次数。这里, 泊松分布的均值并不是常数, 而是随时间逐渐变化。这样的模型称为**非平稳泊松过程模型**(nonstationary Poisson process model)。如果将趋势成分模型用作  $p_n$  的时间变化模型, 就可以得到  $p_n = p_{n-1} + v_n$  这样的模型。不过, 由于  $p_n$  始终必须满足  $0 < p_n < \infty$ , 因此可以通过  $\theta_n = \log p_n$  定义新的变量, 并把它作为状态来引入趋势成分模型。因此, 用于平滑非平稳泊松过程的模型为

$$\theta_n = \theta_{n-1} + v_n \quad (14.-1)$$

$$P(m_n|\theta_n) = \frac{e^{-p_n} p_n^{m_n}}{m_n!} \quad (14.0)$$

其中  $p_n = e^{\theta_n}$ 。这里, 第一式对应系统模型, 第二式对应观测模型, 它们各自给出了条件分布。因此, 对于这种离散序列, 也可以应用本章的方法, 并从数据中估计随时间变化的泊松概率。

完全类似地, 非高斯型状态空间模型也可以应用于非平稳二项分布的平滑化。设在第  $n$  期的  $\ell_n$  次观测中, 某一事件发生  $m_n$  次的概率服从二项分布

$$P(m_n|p_n, \ell_n) = \binom{\ell_n}{m_n} p_n^{m_n} (1-p_n)^{\ell_n - m_n} \quad (14.0)$$

这里,  $p_n$  称为二项概率, 表示某一事件发生的平均概率。在这种情况下,  $p_n$  随时间变化的模型称为**非平稳二项过程模型**(nonstationary binary process model), 并可利用如下模型来估计  $p_n$ :

$$\theta_n = \theta_{n-1} + v_n$$

$$P(m_n|\theta_n, \ell_n) = \binom{\ell_n}{m_n} p_n^{m_n} (1-p_n)^{\ell_n - m_n} \quad (14.0)$$

其中,  $p_n = e^{\theta_n} / (1 + e^{\theta_n})$ 。

### 14.6.3 直接估计时变方差的方法

在 14.5 节中, 我们通过对时间序列平方值的对数进行平滑化来估计时变方差; 而利用如下模型, 也可以进行更为直接的估计:

$$\begin{aligned} t_n &= t_{n-1} + v_n \\ y_n &\sim N(0, e^{t_n}) \end{aligned} \quad (14.0)$$

在该模型中, 方差始终为正, 因此将其对数记为  $t_n$ , 并把它作为状态。此时,  $e^{t_n}$  就是时刻  $n$  的方差。

如果利用这一方法从日经 225 平均股价数据的对数差分序列中估计  $t_n$ , 则可得与图 4.8 类似的结果。估计时得到的对数似然值在两种方法之间完全不同, 这是因为用于模型估计的数据不同。若对 4.6 节中的数据变换进行应用, 并对 AIC 作修正, 则可以确认这两个模型实际上是相同的。

### 14.6.4 非线性状态空间模型

下面考虑由以下非线性状态空间模型表示的系统:

$$\begin{aligned} x_n &= f(x_{n-1}) + w_n \\ y_n &= h(x_n) + \varepsilon_n, \end{aligned} \quad (14.0)$$

其中,  $y_n$  和  $x_n$  分别是 1 维时间序列和  $m$  维状态向量,  $f(x)$  和  $h(x)$  是状态  $x$  的非线性函数。另外,  $w_n$  与  $\varepsilon_n$  分别是具有密度函数  $q(w)$ 、 $r(\varepsilon)$  的  $m$  维和 1 维白噪声, 并与  $x_n$  和  $y_n$  的过去独立。初始状态向量  $x_0$  服从密度函数  $p(x_0)$ 。

对于非线性状态空间模型 (14.0), 条件分布  $p(x_n|x_{n-1})$  和  $p(y_n|x_n)$  为

$$\begin{aligned} p(x_n|x_{n-1}) &= q(x_n - f(x_{n-1})) \\ p(y_n|x_n) &= r(y_n - h(x_n)). \end{aligned} \quad (14.0)$$

此外, 还可验证以下关系成立:

$$\begin{aligned} p(x_n|x_{n-1}, Y_{n-1}) &= p(x_n|x_{n-1}) \\ p(y_n|x_n, Y_{n-1}) &= p(y_n|x_n) \\ p(x_n|x_{n+1}, Y_N) &= p(x_n|x_{n+1}, Y_n). \end{aligned} \quad (14.0)$$

由于一步预测、滤波和平滑化公式仅依赖于 (14.1) 中的关系推导而来, 因此, 非高斯型滤波与平滑化公式同样可以应用于非线性状态空间模型。关于这一方法的详细内容, 请参见 [23] 和 [?].

## ● 章末习题

1. 说明柯西分布和双边指数分布在建模中的作用。
2. 当  $y_n$  服从均值为 0、方差为 1 的正态分布时, 验证  $v_n = \log y_n^2$  的分布由 (14.21) 给出。
3. 考虑如下模型: 趋势以概率  $\alpha$  按正态分布发生变化, 其余情况下保持不变。

## 第 15 章

# 粒子滤波与平滑化

本章介绍作为高维非线性、非高斯型状态空间模型的滤波与平滑方法而开发的粒子滤波<sup>\*1</sup>（也称为蒙特卡罗滤波、Bootstrap 滤波）。这种方法不同于传统的近似法，它用大量粒子来近似一般的非高斯型分布。借此，只需通过各粒子的时间演化和重采样（有放回抽样）这两个简单操作，就能够实现复杂系统的滤波与平滑化。

### 15.1 非线性·非高斯型状态空间模型与分布的近似

本章考虑如下形式的非线性、非高斯型状态空间模型

$$x_n = F(x_{n-1}, v_n) \quad (15.1)$$

$$y_n = H(x_n, w_n), \quad (15.2)$$

其中， $y_n$  是时间序列， $x_n$  是  $k$  维状态向量，(15.1) 称为系统模型或状态模型，(15.2) 称为观测模型。它是对前章所使用的状态空间模型的推广。系统噪声  $v_n$  与观测噪声  $w_n$  分别服从密度函数  $q(v)$  与  $r(w)$ ，是  $\ell$  维与 1 维白噪声。另外，初始状态向量  $x_0$  服从密度函数  $p_0(x)$ 。函数  $F$  与  $H$  一般为非线性函数。不过，对于  $y = H(x, w)$ ，假定在给定  $x$  与  $y$  时， $w$  可唯一确定，并且可用关于  $y$  可微的函数  $G$  表示为  $w = G(y, x)$ 。满足这一条件的例子包括  $H(x, w) = H_1(x) + w$  和  $H(x, w) = e^x w$ ，此时分别有  $G(y, x) = y - H_1(x)$ 、 $G(y, x) = e^{-x} y$ 。为简便起见，本章中假定  $y_n$  与  $w_n$  为一维，但也可以扩展到多维情形。

本章将针对这一非线性、非高斯型状态空间模型，讨论状态估计问题。与前章相同，记到时刻  $t$  为止得到的观测值为  $Y_t \equiv \{y_1, \dots, y_t\}$ ，并说明如何通过粒子近似来计算预测分布  $p(x_n | Y_{n-1})$ 、滤波分布  $p(x_n | Y_n)$  以及平滑化分布  $p(x_n | Y_N)$ 。

<sup>\*1</sup>（译注）原文为“粒子フィルタ”（particle filter）。该方法也称为蒙特卡罗滤波（Monte Carlo filter）或 Bootstrap 滤波，在中文文献中也常称为“粒子滤波器”。

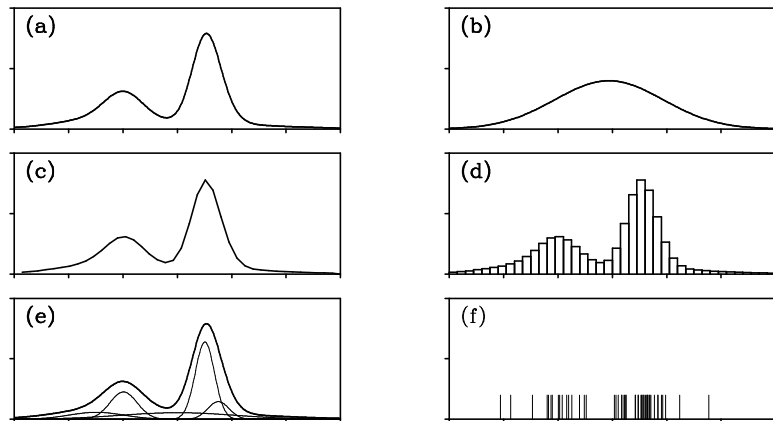


图 15.1 非高斯型分布的近似方法：(a) 假定的真实分布，(b) 正态近似，(c) 折线近似，(d) 阶梯函数近似，(e) 混合正态近似，(f) 用实现值进行近似

图 15.1 展示了近似非高斯型分布的方法。假定 (a) 中双峰分布为真实分布。(b) 表示用单个正态分布对其进行近似的情况，相当于扩展卡尔曼滤波或将模型作线性/正态近似的情形。当真实分布为双峰以上或左右不对称时，这种方法无法得到好的近似。(c) 表示用 40 个分点的折线进行近似，(d) 表示用 40 个分点的阶梯函数进行近似。前章中的非高斯型滤波与平滑，在数值计算时使用的就是这些近似；实际计算中，分点数往往取更大，通常可达数百。这些方法能够以相当高的精度近似任意分布。(e) 表示用 5 个正态分布的混合进行近似的情形。细线表示各个正态分布的贡献部分，粗线表示它们相加后得到的混合正态分布。高斯和滤波/平滑化所使用的正是这种近似。

本章介绍的粒子滤波方法不同于这些近似方法，它用大量粒子来表示分布。每个粒子都是从待表示分布中独立生成的一个实现值。图 15.1(f) 表示用 100 个实现值对 (a) 中的分布进行近似的情形。横轴上的短竖线表示实现值的位置。在 (a) 中密度函数较高的区域，粒子分布得更多。图 15.2 比较了由这些实现值构成的经验分布函数与真实分布函数。(a)、(b)、(c) 分别对应  $m=10, 100, 1000$  的情形。粗线表示真实分布函数，细线表示经验分布函数。当  $m=10$  时，与真实分布函数偏差较大；但随着  $m=100, 1000$  的增加，可以看出能够得到非常好的近似。

在粒子滤波中，预测、滤波、平滑化的分布都用  $m$  个“粒子”按表 15.1 所示进行近似。粒子数  $m$  会根据模型复杂度和所需精度，通常取 1000 个到 10 万个左右。这相当于用由大量“粒子”所定义的经验分布函数，来近似各个分布函数。

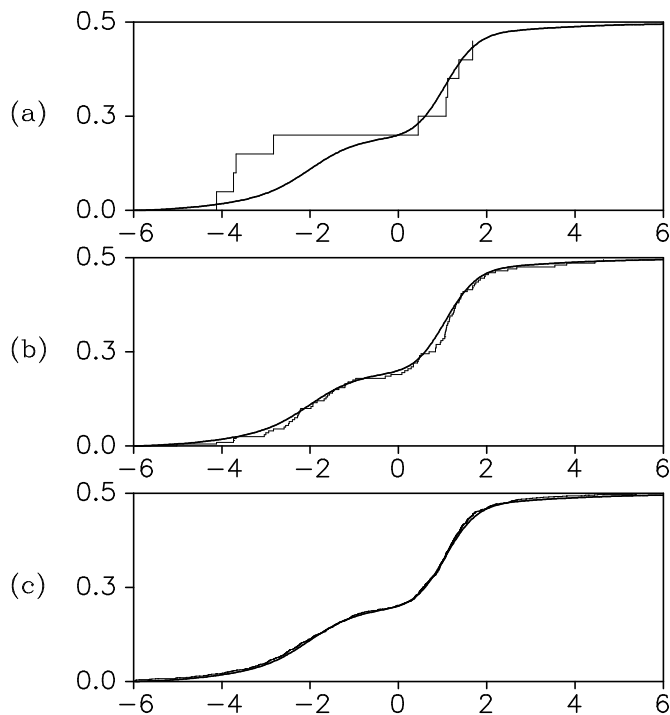


图 15.2 改变粒子数时经验分布函数与真实分布的比较：(a)  $m = 10$ , (b)  $m = 100$ , (c)  $m = 1000$

## 15.2 粒子滤波

本节给出一种粒子滤波算法，用于逐次生成近似预测分布与滤波分布的粒子。根据粒子滤波，服从预测分布的粒子  $\{p_n^{(1)}, \dots, p_n^{(m)}\}$  可以由前一期滤波实现值  $\{f_{n-1}^{(1)}, \dots, f_{n-1}^{(m)}\}$  生成，而滤波实现值  $\{f_n^{(1)}, \dots, f_n^{(m)}\}$  则可以通过对预测分布的实现值  $\{p_n^{(1)}, \dots, p_n^{(m)}\}$  进行重采样来生成。

### 15.2.1 一步预测

在预测步骤中，假定已经给定可视为状态  $x_{n-1}$  的条件分布  $p(x_{n-1}|Y_{n-1})$  的实现值的  $m$  个粒子  $\{f_{n-1}^{(1)}, \dots, f_{n-1}^{(m)}\}$ ，以及可视为系统噪声  $v_n$  的独立实现值的粒子

表 15.1 粒子滤波中的近似

分布	密度函数	由粒子进行的近似
预测分布	$p(x_n Y_{n-1})$	$\{p_n^{(1)}, \dots, p_n^{(m)}\}$
滤波分布	$p(x_n Y_n)$	$\{f_n^{(1)}, \dots, f_n^{(m)}\}$
平滑化分布	$p(x_n Y_N)$	$\{s_{n N}^{(1)}, \dots, s_{n N}^{(m)}\}$
系统噪声的分布	$p(v_n)$	$\{v_n^{(1)}, \dots, v_n^{(m)}\}$

$\{v_n^{(1)}, \dots, v_n^{(m)}\}$ 。也就是说, 对  $i = 1, \dots, m$  有

$$f_{n-1}^{(i)} \sim p(x_{n-1}|Y_{n-1}), \quad v_n^{(i)} \sim q(v) \quad (15.2)$$

成立。此时,

$$p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)}) \quad (15.2)$$

所定义的  $p_n^{(j)}$ , 可视为从  $x_n$  的一步预测分布中生成的粒子。(参见附录 D)

### 15.2.2 滤波

接下来, 在滤波步骤中, 首先计算基于观测值  $y_n$  得到的粒子  $p_n^{(j)}$  的“似然”(或者贝叶斯系数)  $\alpha_n^{(j)}$ 。即, 对  $j = 1, \dots, m$  定义

$$\alpha_n^{(j)} = p(y_n | p_n^{(j)}) = r(G(y_n, p_n^{(j)})) \left| \frac{\partial G}{\partial y_n} \right|$$

其中,  $G$  是函数  $H$  的逆函数,  $r$  是观测噪声  $w$  的密度函数。 $\alpha_n^{(j)}$  可以看作表示粒子  $p_n^{(j)}$  重要程度的权重系数。

接着, 对  $p_n^{(1)}, \dots, p_n^{(m)}$  按其“似然”  $\alpha_n^{(1)}, \dots, \alpha_n^{(m)}$  成比例的概率进行重采样, 从而得到  $m$  个粒子  $f_n^{(1)}, \dots, f_n^{(m)}$ 。也就是说, 新的粒子  $f_n^{(j)}$ , ( $j = 1, \dots, m$ ) 由

$$f_n^{(j)} = \begin{cases} p_n^{(1)} & \text{概率 } \alpha_n^{(1)} / (\alpha_n^{(1)} + \dots + \alpha_n^{(m)}) \\ \vdots & \vdots \\ p_n^{(m)} & \text{概率 } \alpha_n^{(m)} / (\alpha_n^{(1)} + \dots + \alpha_n^{(m)}) \end{cases} \quad (15.2)$$

进行有放回抽样得到。此时  $\{f_n^{(1)}, \dots, f_n^{(m)}\}$  可视为来自滤波分布  $p(x_n | Y_n)$  的实现值。

(附录 D)

### 15.2.3 粒子滤波算法

将以上结果归纳起来, 可得如下粒子滤波算法。

1. 对  $j = 1, \dots, m$ , 生成  $k$  维随机数  $f_0^{(j)} \sim p_0(x)$ 。
2. 对  $n = 1, \dots, N$  执行以下步骤。
  - (i) 对  $j = 1, \dots, m$ :
    - 生成  $\ell$  维随机数  $v_n^{(j)} \sim q(v)$ 。
    - 计算  $p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)})$ 。
    - 计算  $\alpha_n^{(j)} = r(G(y_n, p_n^{(j)})) \left| \frac{\partial G}{\partial y_n} \right|$ 。
  - (ii) 从  $\{p_n^{(1)}, \dots, p_n^{(m)}\}$  中按  $\{\alpha_n^{(1)}, \dots, \alpha_n^{(m)}\}$  成比例的概率进行  $m$  次重采样(有放回抽样), 生成  $\{f_n^{(1)}, \dots, f_n^{(m)}\}$ 。

### 15.2.4 模型的似然

状态空间模型 (15.1)、(15.2) 通常包含噪声方差以及  $F$ 、 $H$  的系数等未知参数  $\theta$ 。

给定观测值  $y_1, \dots, y_N$  时, 由 (9.21) 可知参数  $\theta$  的对数似然为

$$\ell(\theta) = \sum_{n=1}^N \log p(y_n | Y_{n-1}), \quad (15.3)$$

其中,  $p(y_1 | Y_0) = p_0(y_1)$ 。这里, 利用预测分布的粒子近似

$$\begin{aligned} p(y_n | Y_{n-1}) &= \int p(y_n | x_n) p(x_n | Y_{n-1}) dx_n \\ &\cong \frac{1}{m} \sum_{j=1}^m p(y_n | p_n^{(j)}) = \frac{1}{m} \sum_{j=1}^m \alpha_n^{(j)}, \end{aligned} \quad (15.3)$$

则对数似然可近似为

$$\ell(\theta) = \sum_{n=1}^N \log p(y_n | Y_{n-1}) \cong \sum_{n=1}^N \log \left( \sum_{j=1}^m \alpha_n^{(j)} \right) - N \log m \quad (15.3)$$

参数的极大似然估计值  $\hat{\theta}$  可以通过数值方式极大化这一对数似然来求得。

不过, 需要注意的是, 利用粒子滤波得到的对数似然中存在由蒙特卡罗方法带来的近似误差。解决这一问题的一种方法是, 将参数  $\theta$  也视为状态向量的一部分, 从而进行联合估计的自组织建模。

### 15.2.5 关于重采样方法

本节讨论滤波步骤中所需的重采样方法。基于随机抽样的基本算法如下。

- 对  $j = 1, \dots, m$ , 重复执行以下 (a)~(c)。

(a) 生成均匀随机数  $u_n^{(j)} \in U[0, 1]$ 。

(b) 寻找满足  $\frac{1}{C} \sum_{\ell=1}^{i-1} \alpha_n^{(\ell)} < u_n^{(j)} \leq \frac{1}{C} \sum_{\ell=1}^i \alpha_n^{(\ell)}$  的  $i$ 。

其中,  $C = \sum_{\ell=1}^m \alpha_n^{(\ell)}$ 。

(c) 令滤波粒子为  $f_n^{(j)} = p_n^{(i)}$ 。

不过, 上述重采样的目的, 是将由带权重  $\{\alpha_n^{(1)}, \dots, \alpha_n^{(m)}\}$  的粒子  $\{p_n^{(1)}, \dots, p_n^{(m)}\}$  所定义的概率密度函数, 重新表示为等权经验分布函数的形式, 因此并不一定必须进行严格的随机抽样。考虑到这一点, 在重采样方法中, 关于排序和随机数的生成方式可以有各种变形。一种改进方法如下, 即采用分层抽样。也就是说, 将区间  $(0, 1]$

划分为  $m$  个等长小区间, 并在每个区间内求一个  $u_n^{(j)}$ 。在这种情况下, 步骤 (a) 可由以下两种方法之一替代:

$$(a-S) \text{ 生成均匀随机数 } u_n^{(j)} \sim U\left(\frac{j-1}{m}, \frac{j}{m}\right]。$$

$$(a-D) \text{ 对固定的 } \alpha \in (0, 1], \text{ 令 } u_n^{(j)} = \frac{j-\alpha}{m}。$$

对于  $\alpha$ , 可以取  $1/2$ , 也可以用随机数来确定。严格地说, 要进行分层抽样, 首先需要将  $p_n^{(1)}, \dots, p_n^{(m)}$  按大小顺序重新排列, 并相应地重新排列  $\alpha_n^{(1)}, \dots, \alpha_n^{(m)}$ , 然后再进行上述抽样。然而, 排序在粒子滤波中, 尤其当粒子数  $m$  很大时, 会成为最耗时的运算; 并且, 如果将分层抽样的目的理解为: 从权重总和为  $1/m$  的粒子组中恰好抽取 1 个粒子, 那么实际上并不一定需要按顺序重新排序。

根据比较实验的结果, 与严格随机抽样相比, 采用分层的方法能够得到好得多的近似; 而且在两种分层方法中, (a-D) 的确定性方法比 (a-S) 的分层抽样法还要略好一些。

### 15.2.6 数值例

为了说明一步预测分布和滤波分布是如何用“粒子”近似的, 这里给出仅一步的粒子滤波结果 ( $n = 1$ )。考虑如下 1 阶线性、非高斯型模型。

$$\begin{aligned} x_n &= x_{n-1} + v_n \\ y_n &= x_n + w_n \end{aligned} \quad (15.3)$$

其中,  $v_n$  与  $w_n$  分别服从密度函数为  $q(v) = \sqrt{0.1}\pi^{-1}(v^2 + 0.1)^{-1}$  的柯西分布以及标准正态分布  $N(0, 1)$  的白噪声。另外, 假定初始分布  $p_0(x_0)$  也服从标准正态分布。不用说, 在粒子滤波中, 正态性的假设并不是本质性的。在上述条件下, 求得了预测分布  $p(x_1|Y_0)$  以及滤波分布  $p(x_1|Y_1)$ 。粒子数取  $m = 100$ 。之所以采用较少的粒子数  $m = 100$ , 是为了使图更易于观察; 在通常计算中, 会使用更多的粒子。

图 15.3(a) 中的实线表示所假定的初始分布  $p_0(x_0)$ 。竖线表示从该分布生成的 100 个实现值的位置, 条形函数表示它们的直方图。另一方面, (b) 中的粗线表示真实分布函数, 细线表示由 (a) 中的实现值得到的经验分布函数。同样地, (c)、(d) 给出了系统噪声  $v_1$  的密度函数、实现值等。

(e)、(f) 与预测分布  $p(x_1|Y_0)$  有关。(e) 中的曲线是由 (a) 与 (c) 的密度函数通过数值积分求得的“真实”密度函数。(f) 中的实线则是对其进一步积分得到的“真实”

分布函数。另一方面, (e) 中的竖线表示将 (a)、(c) 中的实现值代入式 (D) 后得到的  $p_1^{(j)}$  的位置。其直方图构成了对真实函数的近似。经验分布函数与真实分布函数之间的关系则示于 (f)。

(g) 中的曲线是在假定得到观测值  $y_1 = 2$  时的滤波分布, 它由式 (14.5) 的非高斯型滤波求得。(g) 中的粒子位置与 (e) 相同, 但竖线高度与似然  $\alpha_n^{(j)}$  成正比。因此, (h) 中的累积分布函数其阶梯宽度与  $\alpha_n^{(j)}$  成正比, 虽然阶梯位置与 (f) 相同, 但由于阶梯高度不同, 因此与等步长的 (f) 完全不同, 并近似了滤波分布。(i) 表示重采样后的直方图、粒子位置以及精确的滤波分布。可以看出, (i) 中的密度函数与 (j) 中的经验分布函数, 都分别很好地近似了 (g)、(h)。

### 15.3 粒子平滑化

上一节中的粒子滤波方法, 只需保存过去的粒子, 就可以扩展到平滑化的情形。以下约定,  $(s_{1|n}^{(j)}, \dots, s_{n|n}^{(j)})^T$  表示联合分布  $p(x_1, \dots, x_n | Y_n)$  的第  $j$  个实现值。此时, 为了进行平滑化, 只需将 15.2.3 节算法步骤 2 中的 (ii) 改为如下形式即可。

(ii-S) 对于  $j = 1, \dots, m$ , 通过对  $(s_{1|n-1}^{(j)}, \dots, s_{n-1|n-1}^{(j)}, p_n^{(j)})^T$  进行重采样, 生成  $(s_{1|n}^{(j)}, \dots, s_{n-1|n}^{(j)}, s_{n|n}^{(j)})^T$ 。

在这一修改中, 只需将  $\{(s_{1|n-1}^{(j)}, \dots, s_{n-1|n-1}^{(j)}, p_n^{(j)})^T, j = 1, \dots, m\}$  按照 15.2.3 节步骤 2(ii) 中的权重进行重采样, 就能够实现非线性、非高斯型模型的固定区间平滑化。但是, 如下所示, 由于实际计算是基于有限个 ( $m$  个) 样本进行的, 实际上不同粒子的数目会单调不减; 随着不断重复重采样, 粒子会集中到少数几个粒子上, 最终将损害分布的形状。

因此, 在平滑化算法中, 可以考虑将 (ii-S) 修改为如下形式。

(ii-L) 对于  $j = 1, \dots, m$ , 通过对  $(s_{n-L|n-1}^{(j)}, \dots, s_{n-1|n-1}^{(j)}, p_n^{(j)})$  进行重采样, 生成  $(s_{n-L|n}^{(j)}, \dots, s_{n-1|n}^{(j)}, s_{n|n}^{(j)})$ 。其中, 令  $f_n^{(j)} = s_{n|n}^{(j)}$ , 并将  $L$  取为固定值。

有趣的是, 这一修正后的算法, 等价于进行  $L$  滞后的固定滞后平滑化。当  $L$  取较大值时,  $p(x_n | Y_{n+L})$  能够较好地近似  $p(x_n | Y_N)$ , 但是,  $\{x_{n+L}^{(1)}, \dots, x_{n+L}^{(m)}\}$  的分布反而会偏离  $p(x_n | Y_{n+L})$ 。因此,  $L$  不宜取得过大。通常取 20 以下较为合适。

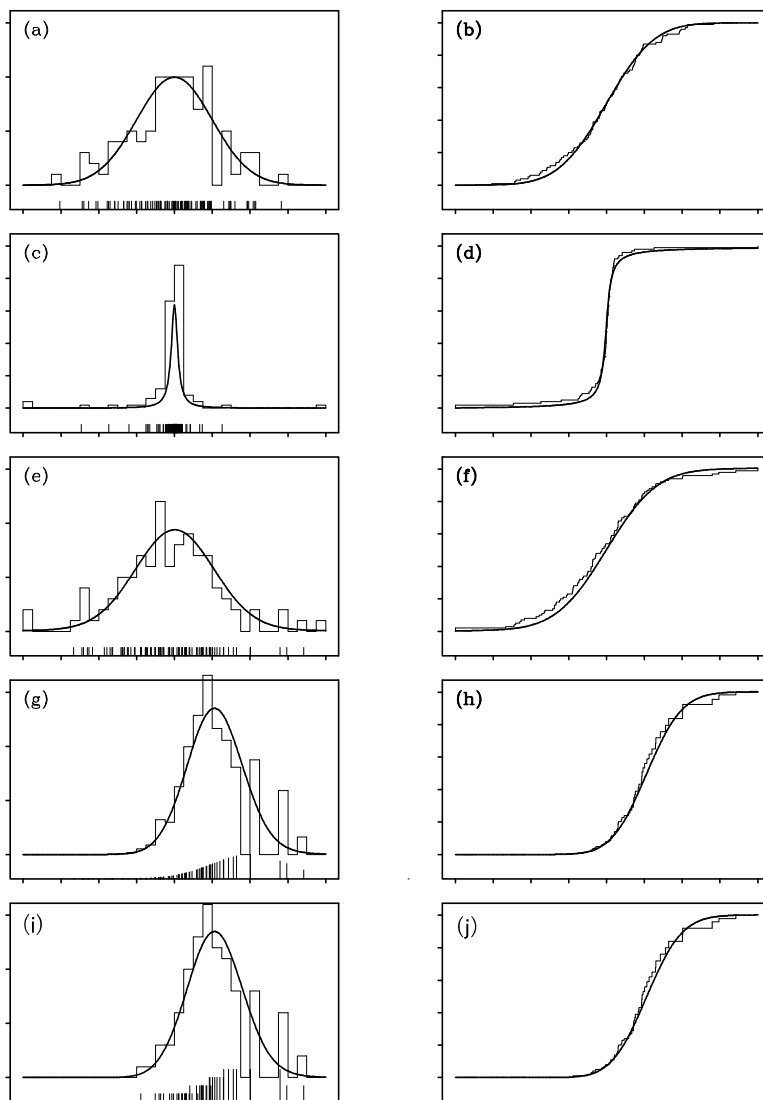


图 15.3 一步粒子滤波：(a) 初始分布的密度函数、实现值与直方图，(b) 初始分布的分布函数与经验分布函数，(c) 系统噪声的密度函数、实现值与直方图，(d) 系统噪声的分布函数与经验分布函数，(e) 预测分布的密度函数、实现值与直方图，(f) 预测分布的分布函数与经验分布函数，(g) 滤波分布的密度函数、带权实现值与直方图，(h) 滤波分布的分布函数与经验分布函数，(i) 重采样后的实现值与直方图，(j) 重采样后的经验分布函数

下面的例子是针对图 14.1 中人工生成的数据，假定使用如下的一阶趋势模型

$$\begin{aligned} x_n &= x_{n-1} + v_n \\ y_n &= x_n + w_n \end{aligned} \quad (15.3)$$

并用粒子滤波进行状态估计时得到的结果。这里假定  $w_n$  服从均值为 0、方差为  $\sigma^2$  的正态分布，而对系统噪声则考虑正态分布和柯西分布两种情形。

TSSS 包中的函数 `pfilter` 针对具有非高斯型系统噪声和观测噪声的线性状态空间模型，执行粒子滤波与平滑化。该函数需要如下输入：

**y:** 一元时间序列  
**m:** 粒子数  
**model:** 系统噪声的模型:  
     = 0: 正态分布  
     = 1: 柯西分布  
     = 2: 混合高斯分布  
**lag:** 固定滞后平滑化的滞后数  
**initd:** 状态的初始分布:  
     = 0:  $N(0,1)$   
     = 1:  $U(-4,4)$   
     = 2:  $\text{Cauchy}(0,1)$   
     = 3: 以  $x=0$  为中心的 Delta 函数  
**sigma2:** 观测噪声的方差  $\sigma^2$   
**tau2:** 系统噪声的方差  $\tau^2$   
**plot:** 逻辑变量。若为 TRUE (默认值), 则显示后验分布图

该函数的输出为：

**post.dist:** 后验分布  
**llkhood:** 模型的对数似然

图 15.4(a) 给出了系统噪声为正态分布时, 通过卡尔曼滤波得到的精确滤波分布。中央的粗线表示分布的均值, 三个灰色区域分别表示  $\pm 1\sigma$ 、 $\pm 2\sigma$ 、 $\pm 3\sigma$  的范围。另一方面, (b)、(c)、(d) 中取  $L=20$ , 并分别对应  $m=100$ 、 $1,000$ 、 $10,000$  的粒子滤波结果。图中的粗线表示中位数, 深灰色区域表示  $\pm 1\sigma$  范围, 浅灰色区域表示  $\pm 2\sigma$  范围。在 (d) 中, 还进一步给出了  $\pm 3\sigma$  范围。(b) 中  $m=100$  时, 虽然抓住了 (a) 的大致特征, 但波动较大, 不能算是很好的近似。然而, 随着粒子数的增加, 可以得到更精确的近似; 在 (c) 的  $m=1,000$  情况下, 虽然  $\pm 3\sigma$  点的波动仍稍大, 但其余范围已经得到了较好的近似。在 (d) 的  $m=10,000$  情况下, 则成为对 (a) 的非常好的近似。

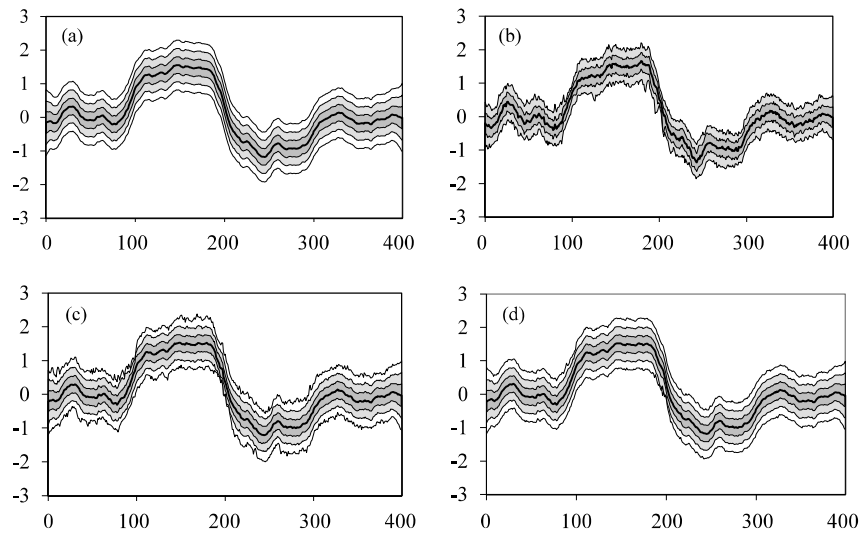


图 15.4 粒子滤波的结果：(a) 卡尔曼滤波得到的精确值，(b)~(d) 固定滞后 ( $L=20$ ) 粒子平滑化的结果 ((b)  $m=1,000$ , (c)  $m=10,000$ , (d)  $m=100,000$ )

```

> # 输入测试数据：在 N=101, 201, 301 处有水平跳变
> y <- data( pfilter_sample )
>
> # 卡尔曼滤波
> # trend( y, trend.order = 1, tau2.ini = 1.4e-2, delta = 0.0 )
>
> # 粒子滤波
> # 系统噪声：正态分布，观测噪声：正态分布
> # 粒子数 m = 1,000:
> pfilter(y, m = 1000 , model = 0, lag = 20 , initd = 0, sigma2 = 1.048, tau2 =
1.40e-2)
>
> # 粒子数 m = 10,000:
> pfilter(y, m = 10000, model = 1, lag = 20 , initd = 0, sigma2 = 1.048, tau2 =
1.40e-2)
>
> # 粒子数 m = 100,000:
> pfilter(y, m = 100000, model = 2, lag = 20 , initd = 0, sigma2 = 1.048, tau2 =
1.40e-2)

```

```

> # 非高斯型平滑化
> # 系统噪声：柯西分布
> ngsmth( y, noisev = 2, bv = 1.0, tau2 = 1.40e-2, noisew = 1, sigma2 = 1.048 )
> #
> # 粒子滤波·平滑化
> # 系统噪声：柯西分布
> pfilter( y, m = 100000, model = 1, lag = 20 , initd = 0, sigma2 = 1.045, tau2 =
3.53e-5)

```

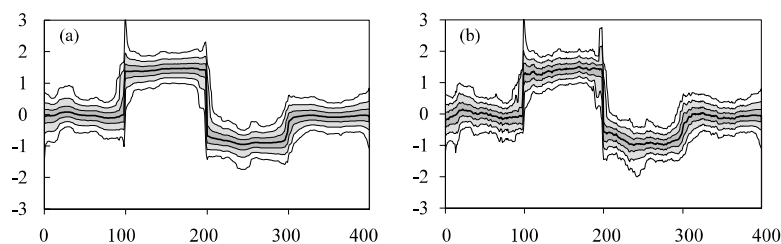


图 15.5 利用柯西分布进行平滑化: (a) 用非高斯型平滑化算法计算得到的精确分布. (b) 粒子平滑化结果 ( $m = 100,000$ )。

图 15.5 给出了系统噪声服从柯西分布时的结果。左图是第 14 章中的非高斯型平滑化得到的精确结果, 右图是取  $m = 10000$ 、 $L = 20$  时粒子滤波得到的平滑化结果。与左图相比, 虽然波动稍大一些, 但在  $n = 100$ 、 $200$ 、 $300$  处的变化仍被很好地捕捉到了。另外,  $\pm 3\sigma$  的误差范围也被非常好地近似了。

### 15.3.1 非线性平滑化

利用粒子滤波, 可以对非线性状态空间模型执行滤波与平滑化。图 15.6 的 (a) 和 (b) 给出了由如下非线性模型

$$\begin{aligned} x_n &= \frac{1}{2}x_{n-1} + \frac{25x_{n-1}}{x_{n-1}^2 + 1} + 8\cos(1.2n) + v_n \\ y_n &= \frac{x_n^2}{10} + w_n \end{aligned} \quad (15.3)$$

生成的  $x_n$  与  $y_n$ 。其中,  $v_n \sim N(0, 1)$ ,  $w_n \sim N(0, 10)$ ,  $v_0 \sim N(0, 5)$ 。这里考虑基于 100 个观测值  $y_n$ , 估计未知状态  $x_n$  的问题。 $x_n$  会不时在正负两个区域之间切换, 但在观测模型中, 观测到的是平方后的  $x_n^2$ , 因此几乎无法判断其符号。已知扩展卡尔曼滤波在这种情况下常常会发散。图 15.6(c) 给出了用粒子滤波求得的  $x_n$  的后验分布  $p(x_n|Y_n)$ 。可以看出, 即使在非线性模型中, 也能够得到未知状态  $x_n$  的良好估计。

TSSS 包中的函数 `pfilterNL` 用于对上述非线性状态空间模型执行粒子滤波和平滑化。该函数需要如下输入:

- y: 一元时间序列
- m: 粒子数
- lag: 固定滞后平滑化的滞后数
- sigma2: 观测噪声的方差  $\sigma^2$
- tau2: 系统噪声的方差  $\tau^2$
- plot: 逻辑变量。若为 TRUE (默认值), 则显示后验分布图

该函数的输出为:

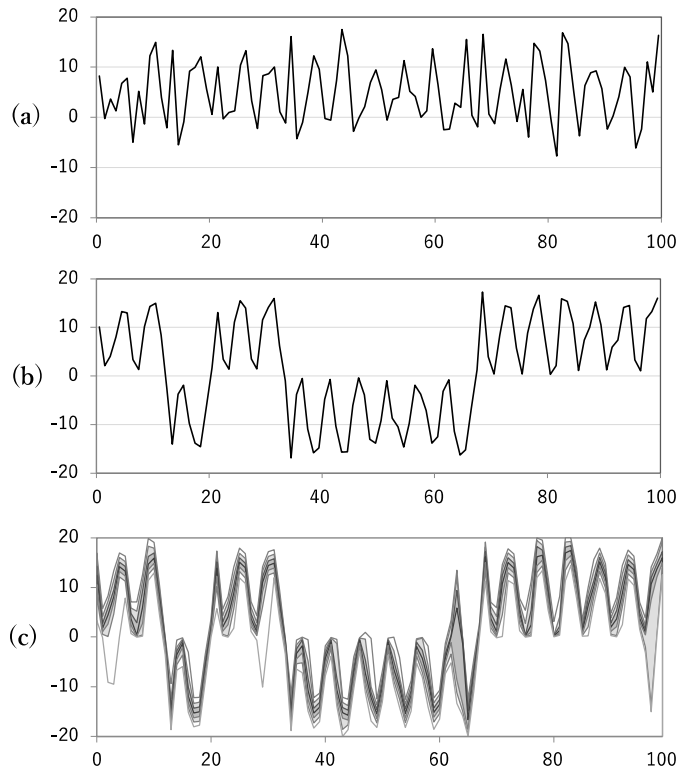


图 15.6 非线性平滑化: (a) 数据  $y_n$ , (b) 未知状态  $x_n$ , (c) 通过粒子平滑化得到的状态  $x_n$  的平滑分布

`post.dist`: 后验分布。但其中, `post.dist[,4]` 表示中位数,  
`post.dist[,j]`,  $j=1,2,3,5,6,7$ : 表示  $-3, -2, -1, 1, 2, 3$  个标准差  
`llkhood`: 模型的对数似然

```
> data(NLmodel)
> x <- NLmodel[, 1]
> pfilterNL(x, m = 100000, lag = 20, sigma2 = 10.0, tau2 = 1.0, xrange = c(-20,
20), seed = 2019071117)
```

对于这种非线性模型, 已知扩展卡尔曼滤波有时会发散 ([23])。图 15.6(c) 给出了通过粒子滤波和平滑化所得到的平滑化后验分布。可以看出, 利用这种方法能够得到状态  $x_n$  非常好的估计值。

为了使粒子滤波和平滑化成功, 关键在于能够恰当地表示状态的非高斯型分布。图 15.7(a) 与 (b) 分别给出了在  $t=30$  和 48 时的状态边缘后验分布  $p(x_t|y_{1:t+j})$ ,  $j=-1, 0, 1, 2$ 。无论是在  $t=30$  还是在 48 的情形中, 预测分布  $p(x_t|y_{1:t-1})$  和滤波分布  $p(x_t|y_{1:t})$  都是双峰型的。但是, 在  $t=30$  时的 2-滞后平滑化  $p(x_{30}|y_{1:32})$  中, 分布左半部分消失了, 分布变成了单峰型。在  $t=48$  时也能看到同样的现象。这种情况

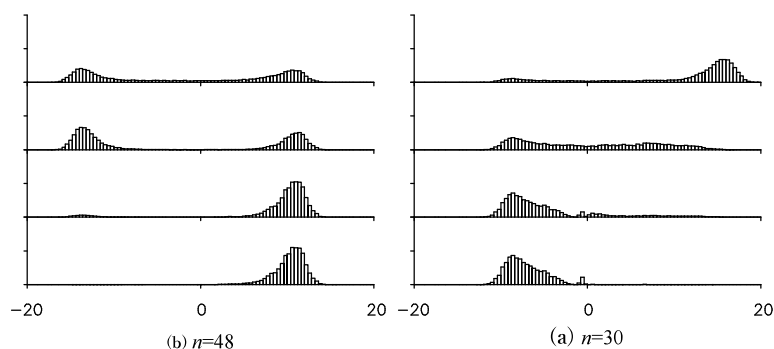


图 15.7  $t = 30$  (左) 和  $t = 48$  (右) 时的固定滞后平滑化。从上到下依次为：预测分布、滤波分布、1-滞后以及 2-滞后平滑化。

下，在预测分布中右侧峰值低于左侧峰值。然而，在 1-滞后和 2-滞后平滑化的结果中，右侧区域的峰值消失了。在这种情况下，扩展卡尔曼滤波很可能得到符号相反的结果。

### ● 章末习题

1. 针对图 1(g) 中的日经 225 数据, 考虑一个同时描述趋势与波动率变化的模型。
2. 在蒙特卡罗滤波中, 若希望将估计精度提高一个数量级, 需要将粒子数增加到什么程度?
3. 在蒙特卡罗滤波中, 如果不将粒子限制为等概率, 而是直接利用其权重, 是否就可以不进行重采样?
4. 比较固定区间平滑化和固定滞后平滑化所需保存数据量的差异。

## 第 16 章

# 仿真

通过生成服从各种分布的随机数,可以获得白噪声的实现值。时间序列模型将时间序列表示为以白噪声为输入的系统的输出。因此,给定时间序列模型后,可以利用随机数生成白噪声的实现值并投入模型,从而生成时间序列。本章将介绍服从各种分布的随机数的生成方法、服从 ARMA 模型的时间序列的生成方法,以及利用状态空间模型的统一仿真方法。

### 16.1 均匀随机数的生成

服从某种分布独立生成的值的序列称为**随机数**(random number)。换言之,随机数就是白噪声的实现值。时间序列模型将时间序列表示为以白噪声为输入的系统的输出。因此,以随机数为输入,可以生成服从各种模型的时间序列。这种时间序列的生成方法称为**仿真**(simulation)。

在计算机中使用随机数时,通常使用按适当算法生成的伪随机数。进行时间序列仿真所需的服从正态分布等分布的随机数,是先生成服从均匀分布的**均匀随机数**(uniform random number),再通过适当变换求得。

长期以来广泛用于生成均匀随机数的方法是**同余法**(congruence method)。该方法以适当的整数  $I_0$  为初始值,通过

$$I_k \equiv aI_{k-1} + c \pmod{m} \quad (16.0)$$

依次生成  $I_1, I_2, \dots$ 。当  $c = 0$  时,特别称为乘法型同余法。此时,  $I_k$  取  $[0, m)$  上的值,因此  $\frac{I_k}{m}$  为  $[0, 1)$  上的分布。

但是,该方法生成的数列的周期最多为  $m$ ,因此无法生成长度超过  $m$  的随机数。

$a$ 、 $c$ 、 $m$  选择适当的整数即可,例如

$$\begin{aligned} a=1229 & \quad c=351750 & m=1664501 \\ a=1103515245 & \quad c=12345 & m=2^{31} \end{aligned}$$

等组合据说可以生成较好的均匀随机数 [27]。后者的组合在 1990 年代之前被 C 语言所使用。当  $m=2^p$ 、 $a=5 \pmod{m}$  且  $c$  为奇数时，该算法生成的序列可以达到最大可能周期  $m$ ，已知 0 到  $m-1$  的所有数恰好各出现一次。

在带滞后的 Fibonacci 法中，整数序列  $I_n$  由以下递推关系生成。

$$I_{n+p} = I_{n+q} + I_n \pmod{m} \quad (16.1)$$

其中  $m$  通常设为  $2^k$  ( $k$  为比特长度)。此时，生成的序列最大周期为  $2^{k-1}(2^p-1)$ 。当前 C 语言中，为实现约  $2^{60}$  的周期，使用  $p=31$ 、 $q=3$ 、 $m=2^{31}$ 。

GFSR (generalized feedback shift register) 法 [28] 中， $k$  维二进制数列由以下递推关系生成。

$$I_{n+p} = I_{n+q} \oplus I_n \quad (16.1)$$

其中  $p > q$ ， $\oplus$  表示异或 (exclusive OR) 运算。整数  $p$  和  $q$  的选取要使多项式  $X^p + X^q + 1$  不可约。此时，生成序列的最大周期为  $2^p - 1$ 。满足此条件的  $p$  和  $q$  的例子有以下组合。

$$\begin{aligned} p=521 & \quad q=158 & \text{周期 } 2^{521}-1 \\ p=521 & \quad q=32 & \text{周期 } 2^{521}-1 \\ p=607 & \quad q=273 & \text{周期 } 2^{607}-1 \end{aligned}$$

Twisted GFSR 法 [?] 中， $k$  维二进制数列由以下递推关系生成。

$$I_{n+p} = I_{n+q} \oplus I_n A \quad (16.1)$$

其中  $A$  为伴随矩阵等  $k \times k$  正则矩阵。常用的 TT800 算法使用  $p=25$ 、 $q=7$ ，周期为  $2^{pk}-1$ 。进一步，Mersenne Twister 由

$$I_{n+p} = I_{n+q} \oplus I_{n+1} B \oplus I_n A \quad (16.1)$$

生成，当  $p=624$  时，适当选择  $A$  和  $B$ ，生成的序列周期为  $2^{19937}-1$ ，已知在 623 维空间中均匀分布 [29]。

与基于某种算法由软件生成的这些伪随机数不同，也有由硬件生成的物理随机数。物理随机数原理上可以生成无周期、无相关的序列，因此在仿真中需要更严格的随机数时被使用。

## R 中均匀随机数的生成

```

> # 生成均匀随机数
> y <- runif( 100 )
>
> y <- runif( 100, min = -1, max = 1)

```

在 R 环境中, 可以通过函数 `runif` 生成基于 Mersenne Twister 的均匀随机数。调用该函数时需要以下输入:

`n`: 随机数的长度  
`min, max`: 均匀分布的上限和下限

当未指定 `min` 或 `max` 时, 使用默认值 0 和 1。

## 16.2 白噪声的生成

为了生成正态白噪声的实现值, 只需生成服从正态分布的随机数序列即**正态随机数** (normal random number)。正态随机数的生成方法中, Box-Muller 法广为人知。该方法是, 设 2 个独立的  $[0, 1)$  上的均匀随机数为  $U_1$ 、 $U_2$ ,

$$\begin{aligned} X_1 &= \sqrt{-2\log U_1} \cos 2\pi U_2 \\ X_2 &= \sqrt{-2\log U_1} \sin 2\pi U_2 \end{aligned} \quad (16.1)$$

由此定义的  $X_1$  和  $X_2$  互相独立地服从标准正态分布  $N(0, 1)$ 。但是, 实际上使用如下改进的不含 `cos`、`sin` 计算的 Marsaglia 算法, 可以更快速地生成正态随机数。

1. 生成均匀随机数  $U_1$ 、 $U_2$ 。
2. 令  $V_1 = 2U_1 - 1$ 、 $V_2 = 2U_2 - 1$ 。
3. 令  $S^2 = V_1^2 + V_2^2$ 。
4. 若  $S^2 \geq 1$  则返回第 1 步。

$$5. X_1 = V_1 \sqrt{\frac{-2\log S}{S}}, X_2 = V_2 \sqrt{\frac{-2\log S}{S}}$$

### R 中正态随机数的生成

```

> # 生成标准正态随机数
> y <- rnorm( 100 )
>
> # 生成均值 0、标准差 2 的正态随机数
> y <- rnorm( 100, mean = 1, sd = 2)

```

在 R 环境中，正态随机数可以用函数 `rnorm` 生成。所需输入如下：

`n`: 随机数的长度  
`mean`: 正态分布的均值  
`sd`: 正态分布的标准差

当未指定 `mean` 或 `sd` 时，分别使用默认值 0 和 1。

```
> # 生成均匀随机数
> plot( runif(200), type = "l", ylim = c(0,1) )
>
> # 标准正态随机数
> y <- rnorm( 200 )
> plot( y, type = "l", ylim = c(-3,3) )
>
> hist( y ) # 直方图
> acf( y ) # 自相关函数
```

图 16.1 的 (a) 是用 R 的函数 `runif` 生成的 200 个均匀随机数。相对地，(b) 是用函数 `rnorm` 生成的正态随机数（白噪声）。(c) 和 (d) 分别展示了这 200 个正态随机数的直方图和样本自相关函数。同样，(e) 和 (f) 是将数据数增加到 2000 个时的直方图和样本自相关函数。两个直方图都比较接近正态分布的密度函数，样本自相关函数在  $\pm 1/\sqrt{n}$  的误差范围内，随着数据数的增加趋向于 0。

为了进行非高斯型状态空间模型的仿真，需要求服从任意密度函数  $p(x)$  的白噪声实现值的方法。当非高斯型分布为柯西分布、 $\chi^2$  分布、双指数分布等特殊分布时，有时可以通过适当变换均匀随机数或正态随机数来求得。首先介绍一些这样的例子。

### 16.2.1 $\chi^2$ 分布

自由度  $j$  的  $\chi^2$  分布是用独立的  $j$  个正态随机变量  $X_1, \dots, X_j$

$$\chi_j^2 = X_1^2 + \dots + X_j^2 \quad (16.1)$$

表示的随机变量的分布。因此，要生成服从  $\chi^2$  分布的随机数，只需先生成  $j$  个正态随机数，通过 (16.2.1) 求  $\chi_j^2$  即可。特别地，令  $j = 2$  可得到服从指数分布的随机数。进一步令  $Z = \log \chi_2^2$  可得到服从双指数分布的随机数。但是，设  $u$  为均匀随机数，指数随机数可由

$$v = -\log u \quad (16.1)$$

求得，双指数随机数可由

$$v = \log(-\log u) \quad (16.1)$$

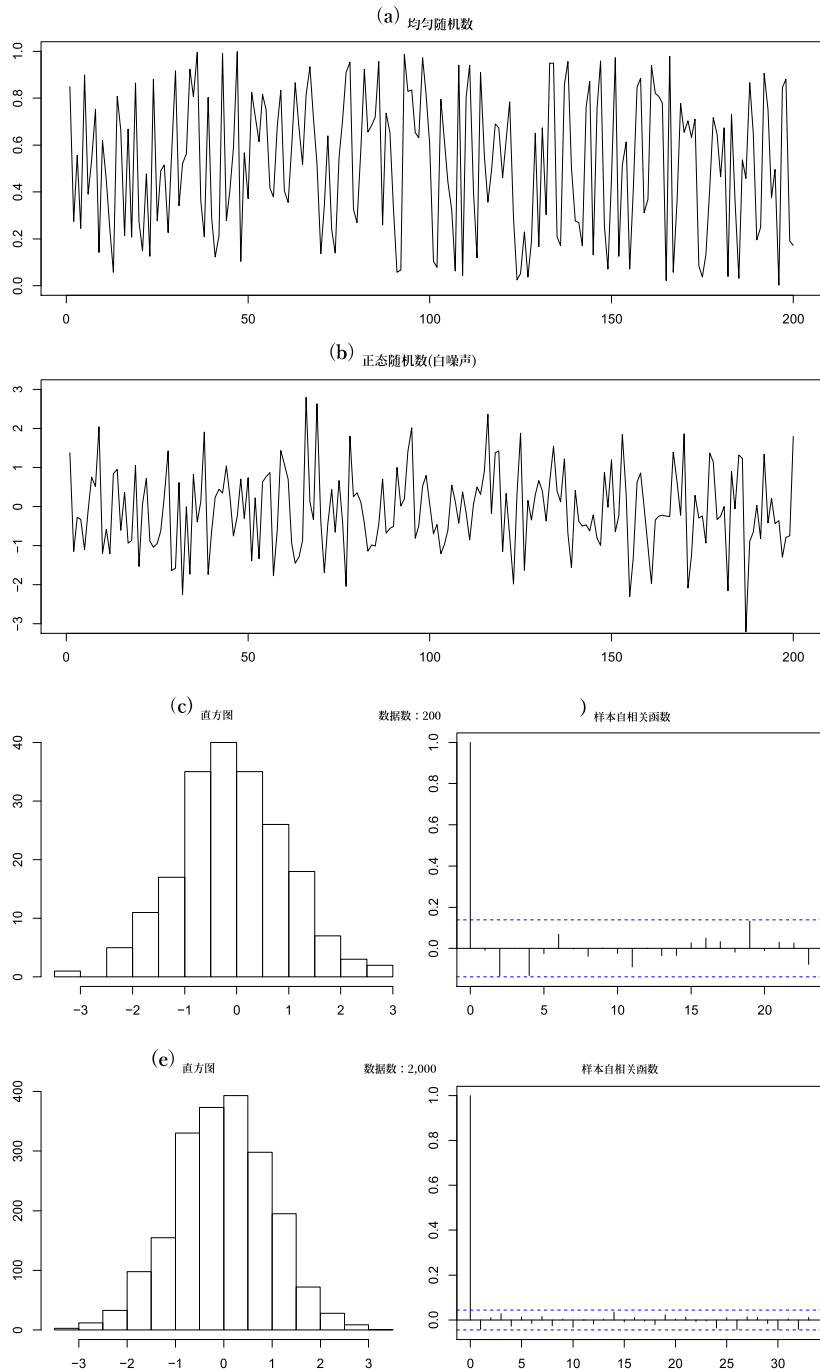


图 16.1 均匀随机数和白噪声的实现值、直方图和自相关函数。

直接从均匀随机数求得。

### 16.2.2 柯西分布

第 14 章非高斯型模型中使用的柯西分布 (Cauchy distribution) 的密度函数为

$$p(x) = \frac{1}{\pi(x^2 + 1)} \quad (16.1)$$

要生成服从柯西分布的随机数, 首先生成  $[0, 1)$  上的均匀随机数  $u$ , 当  $u \neq 0.5$  时令

$$v = \tan \pi u \quad (16.1)$$

即可。

### 16.2.3 服从任意分布的随机数的生成

一般来说, 要生成服从任意密度函数  $f(x)$  的随机数, 可以使用分布函数的逆函数方法。分布函数由

$$F(x) = \int_{-\infty}^x f(t) dt \quad (16.1)$$

定义的单调递增函数, 满足  $0 \leq F(x) \leq 1$ 、 $F(-\infty) = 0$ 、 $F(\infty) = 1$ 。求满足  $F(G(y)) = y$  的逆函数  $G(y) \equiv F^{-1}(y)$  ( $0 \leq y \leq 1$ ), 则  $G$  为从  $(0, 1)$  到  $(-\infty, \infty)$  的函数。此时, 令  $u$  为  $(0, 1)$  上的均匀随机数, 则  $v = G(u)$  为服从密度函数  $f(v)$  的随机数。利用该方法和第 4 章定义的密度函数, 可以生成服从各种分布的随机数。

在 R 环境中, 可以用 `rXXX` 形式的函数生成服从以下各种连续分布的随机数。实际计算中将 `xxx` 替换为适当的名称即可:

`rchisq`:  $\chi^2$  分布  
`rt`: 自由度  $df$  的  $t$  分布  
`rcauchy`: 柯西分布  
`rexp`: 指数分布

```
> # 生成服从各种分布的随机数
> y <- rnorm( 100 ) # 正态分布
>
> y <- rcauchy( 100, location = 0, scale = 1 ) # 柯西分布
>
> y <- rt( 100, df = 3, ncp = 0 ) # t 分布
>
> y <- rchisq( 100, df = 2, ncp = 0 ) #  $\chi^2$  分布
```

### 16.3 ARMA 模型的仿真

假设已给定正态白噪声  $v_1, \dots, v_N$ 。此时 ARMA 模型

$$y_n = \sum_{j=1}^m a_j y_{n-j} + v_n - \sum_{j=1}^{\ell} b_j v_{n-j}, \quad (16.2)$$

的仿真在获得过去观测值  $y_1, \dots, y_m$  的情况下，以它们为初始值，噪声  $v_n$  利用生成的随机数即可进行仿真。

另一方面，当未给定初始值时，在  $k = \max(m, \ell)$  时，可以对  $j = 1, \dots, k$  使用  $y_j = 0$  作为初始值。此时，对  $n = k+1, \dots, N$ ，将  $y_{n-1}, \dots, y_{n-m}$  和  $v_n, v_{n-1}, \dots, v_{n-\ell}$  代入 (16.2) 即可生成  $y_n$ 。

当使用 0 作为初始值时，为了消除任意设定的初始值的影响，丢弃最初的  $L$  个数据，使用剩余的  $n = N - L$  个是很重要的。为此， $L$  需要取足够大的值，使得第 6 章定义的脉冲响应函数可以视为已衰减到 0。更简单的方法是使用  $L = 1000$  这样足够大的值。

#### 例：AR 模型的仿真

以下 R 代码的示例基于给定的初始值执行 AR 模型的仿真。

```
> n <- 200
> m <- 2
> ii <- m+1:n
> mm <- 1:m
> a <- c( 1.5, -0.8 )
> par( mfrow=c(2,1), mar=c(2,2,1,1) 0)
>
> # 生成正态随机数
> y <- rnorm( n )
> plot( y, type="l" )
>
> # 生成 AR 过程
> y[1] <- 0
> y[2] <- 0.5
>
> for( i in ii ){
>   y[i] <- v[i]
>   for( j in mm ) {
>     y[i] <- y[i] + a[j]*y[i-j]
>   }
> }
>
> plot( y, type="l" )
```

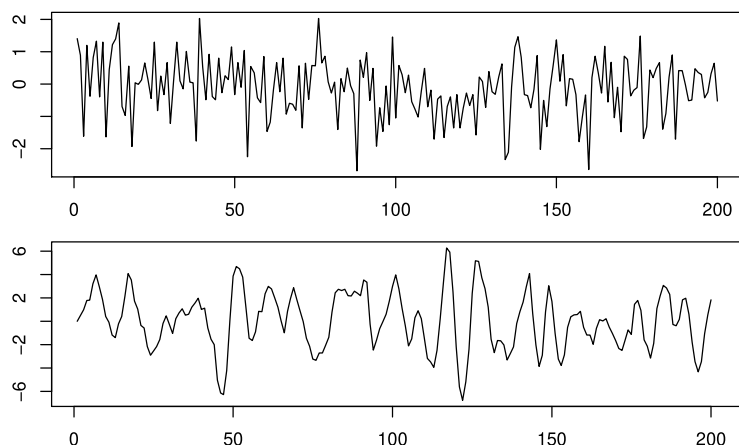


图 16.2 正态白噪声（上图）与 AR 模型的仿真（下图）。

图 16.3 展示了 2 阶 AR 模型的仿真结果。初始状态为  $y_1 = 0$ 、 $y_2 = 0.5$ ，使用正态白噪声  $v_n$  生成了包含 2 个初始值在内的 200 个观测值。上图展示了生成的正态白噪声  $v_n$ 。而下图展示了使用该白噪声生成的 AR 过程。

## 16.4 利用状态空间模型的仿真

利用前节方法生成的白噪声实现值，可以生成服从给定模型的时间序列。这称为时间序列的**仿真** (simulation)。假设时间序列模型以状态空间模型

$$x_n = Fx_{n-1} + Gv_n \quad (16.3)$$

$$y_n = Hx_n + w_n \quad (16.4)$$

的形式表示。这里，给定状态向量的初始值  $x_0$  和均值 0、方差协方差矩阵  $Q$  的  $k$  维白噪声实现值  $v_1, \dots, v_N$ ，通过反复应用 (16.3)，可以简单地求得状态向量的实现值  $x_1, \dots, x_N$ 。进一步，给定均值 0、方差协方差矩阵  $R$  的  $l$  维白噪声实现值  $w_1, \dots, w_N$ ，通过 (16.4) 即可生成时间序列  $y_1, \dots, y_N$ 。

服从方差协方差矩阵  $Q$  的  $k$  维正态白噪声，可以用前节方法生成的一维白噪声来生成。将  $k$  个一维正态白噪声实现值排列为  $u_n = (u_n^{(1)}, \dots, u_n^{(k)})^T$ ，则  $u_n$  显然为均值向量 0、方差协方差矩阵为单位矩阵  $I$  的  $k$  维正态白噪声。这里，通过正定对称矩阵  $Q$  的 **Cholesky 分解** (Cholesky decomposition) 求得满足  $Q = LL^T$  的  $k \times k$  下三角矩阵  $L$ 。此时，

$$v_n = Lu_n \quad (16.4)$$

由此定义  $v_n$ ，则

$$E[v_n v_n^T] = E[Lu_n u_n^T L^T] = LE[u_n u_n^T]L^T = LI_k L^T = Q \quad (16.5)$$

因此,  $v_n$  为服从均值向量 0、方差协方差矩阵  $Q$  的  $k$  维正态白噪声。

### 利用 R 进行高斯型状态空间模型的仿真

```
> # 利用状态空间模型进行仿真
> simssm(n = n1, trend, seasonal.order = m2, seasonal, arcoef, ar, tau1, tau2,
tau3, sigma2 = z$sigma2, seed = 333)
```

TSSS 程序包的函数 `simssm` 用于进行线性状态空间模型的仿真。需要如下设定状态空间模型的参数:

<code>n</code> :	仿真生成的数据数
<code>trend.order</code> :	趋势分量模型的阶数 (0 或 1)
<code>trend</code> :	<code>trend.order</code> >0 时, 趋势分量的初始值:
<code>seasonal.order</code> :	季节分量模型的阶数 (0,1,2)
<code>seasonal</code> :	<code>seasonal.order</code> >0 时, 季节分量长度 $p-1$ 的初始值:
<code>p</code> :	<code>seasonal.order</code> >0 时, 一个季节的长度 (通常为 12)
<code>ar.order</code> :	AR 分量模型的阶数 (0 或正整数)
<code>ar</code> :	<code>ar.order</code> >0 时, 长度 <code>ar.order</code> 的 AR 分量初始估计值:
<code>sigma2</code> :	观测噪声的方差 $\sigma^2$
<code>tau1</code> :	<code>trend.order</code> >0 时, 趋势分量的方差 $\tau_1^2$
<code>tau2</code> :	<code>seasonal.order</code> >0 时, 季节分量模型的方差 $\tau_2^2$
<code>tau3</code> :	<code>ar.order</code> >0 时, AR 分量模型的方差 $\tau_3^2$
<code>plot</code> :	逻辑变量。为 TRUE (默认值) 时, 显示仿真生成的数据

函数 `simssm` 的输出对象 "simulate" 给出高斯型状态空间模型的仿真数据。

### 例：随机游走模型和 AR 模型的仿真

函数 `simssm` 可以按如下方式进行带观测噪声的随机游走模型和 AR 模型的仿真:

```
> # 利用状态空间模型进行仿真
> # 带观测噪声的随机游走模型
> simssm(n = 200, seasonal.order = 0, arcoef=0.5, ar=0, tau3=pi*pi/6, sigma2 = 1,
seed = 333)
>
> # 2 阶 AR 模型
> simssm(n = 200, seasonal.order = 0, arcoef=c(1.5,-0.8), ar=c(0,0.1), tau3=1,
sigma2 = 0, seed = 567)
```

图 16.3 展示了随机游走模型

$$\begin{aligned}x_n &= x_{n-1} + v_n \\ y_n &= x_n + w_n\end{aligned}\quad (16.5)$$

的仿真结果。其中，系统噪声  $v_n$  为  $N(0, \frac{\pi^2}{6})$ ，观测噪声  $w_n$  为  $N(0, 1)$  的正态白噪声，初始值为  $x_0 = 0$ 。

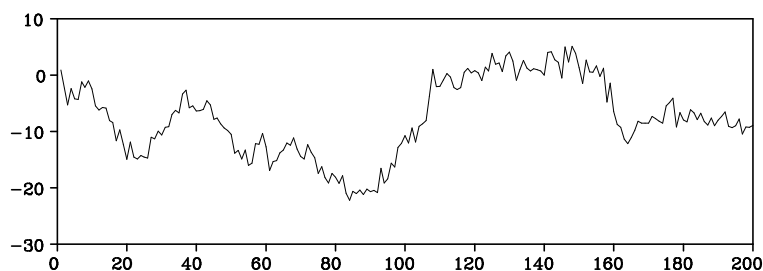


图 16.3 随机游走模型的仿真

#### 例：季节调整模型的仿真

函数 `simssm` 可以按如下方式进行季节调整模型的仿真：

```
> # BLSALLFOOD 数据的仿真
> data(BLSALLFOOD)
> # 季节调整模型的估计> m1 <- 2; m2 <- 1; m3 <- 2
> z <- season(BLSALLFOOD, trend.order = m1, seasonal.order = m2, ar.order = m3)
>
> # 状态空间模型参数的设定
> n1 <- length(BLSALLFOOD)
> trend <- z$trend[m1:1]
> arcoef <- z$arcoef
> period <- 12
> seasonal <- z$seasonal[(period-1):1]
> ar <- z$ar[m3:1]
> tau1 <- z$tau2[1]
> tau2 <- z$tau2[2]
> tau3 <- z$tau2[3]
>
> # 利用状态空间模型进行仿真
> simssm(n = n1, trend, seasonal.order = m2, seasonal, arcoef, ar, tau1, tau2,
tau3, sigma2 = z$sigma2, seed = 333)
```

图 16.4 是利用图 12.6 所示的包含平稳 AR 分量的季节调整模型 ( $m_1 = 2$ 、 $m_2 = 1$ 、 $m_3 = 2$ 、 $p = 12$ ) 的季节调整结果进行仿真的结果。其中，初始值  $x_0$  使用了从数据得到的估计值  $x_{0|N}$ 。图 1.1 (d) 的时间序列与季节变动很相似，但趋势的变动完全不同。

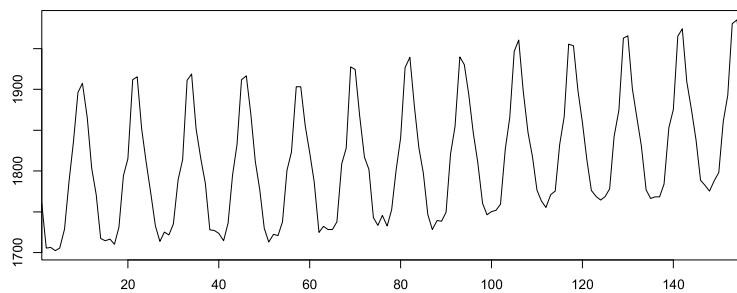


图 16.4 季节调整模型的仿真

## 16.5 利用非高斯型状态空间模型的仿真

前节的方法可以简单地扩展到非高斯型时间序列模型的仿真。时间序列的非高斯型状态空间模型

$$x_n = Fx_{n-1} + Gv_n \quad (16.6)$$

$$y_n = Hx_n + w_n \quad (16.7)$$

中,  $v_n$  和  $w_n$  分别为服从由密度函数  $q(v)$  和  $r(w)$  定义的一般分布的白噪声。为简单起见, 以下假设  $v_n$  的  $k$  个分量和  $w_n$  的  $\ell$  个分量互相独立。为了进行上述非高斯型状态空间模型的仿真, 需要生成服从给定非高斯型分布  $p(x)$  的随机数。

程序包 TSSS 的函数 `ngsim` 用于进行非高斯型状态空间模型的仿真。需要以下输入:

n:	仿真生成的数据数
trend.order:	趋势分量模型的阶数 (0 或 1)
trend:	trend.order > 0 时, 趋势分量模型的初始值
seasonal.order:	季节分量模型的阶数 (0,1,2)
seasonal:	seasonal.order > 0 时, 季节分量模型长度 $p-1$ 的初始值
p:	seasonal.order > 0 时, 季节分量一个周期的长度 (通常为 12)
ar.order:	AR 分量模型的阶数 (0 或正整数)
ar:	ar.order > 0 时, AR 分量模型长度 ar.order 的初始值
noisew:	观测噪声的分布类型 -1: 柯西分布 -2: 指数分布 -3: 双指数分布 0: 双指数分布 + 欧拉常数 1: 正态分布 2: Pearson 分布族 3: 双指数分布
wminmax:	观测噪声的上限值和下限值
paramw:	观测噪声分布的参数 noisew = 1 时, 方差 $\sigma^2$ noisew = 2 时, 散布参数 ( $\sigma^2$ ) 和形状参数 ( $bw$ )
noisev:	系统噪声的分布类型 -1: 柯西分布 -2: 指数分布 -3: 双指数分布 0: 双指数分布 + 欧拉常数 1: 正态分布 2: Pearson 分布族 3: 双指数分布
vminmax:	系统噪声的上限值和下限值
paramv:	系统噪声分布的参数 noisev = 1 时, 方差 $\tau^2$ noisev = 2 时, 散布参数 ( $\tau^2$ ) 和形状参数 ( $bv$ )
plot:	逻辑变量。TRUE (默认值) 时, 绘制仿真数据

函数 `ngsim` 的输出对象 "simulate" 给出非高斯型状态空间模型的仿真数据。

**例：具有双指数分布噪声的趋势模型**

```
> # 以双指数分布为系统噪声的趋势模型仿真
> ar3 <- ngsim(n = 400, arcoef = 1.00, noisew = 1, paramw = 1, noisev = 3, paramv
= 1, seed = 555)
> plot(ar3, use = c(201, 400))
```

图 16.5 是在图 16.3 的随机游走模型中, 将系统噪声  $v_n$  改为双指数随机数  $+\zeta$  ( $\zeta$  为欧拉常数, 0.57722) 进行仿真的结果。为了比较, 图 16.5 的 (a) 和 (b) 展示了图 16.2 中使用的正态分布和双指数分布的密度函数。两个密度函数的均值和方差相同, 但仿真结果有很大差异。在与 (a) 对应的图 16.2 的情况下, 趋势周围的变动上下对称, 而在与 (b) 对应的 (c) 的情况下, 变动方式因上下方向的不同而有很大差异, 呈现出在股价变动等中常见的特征。

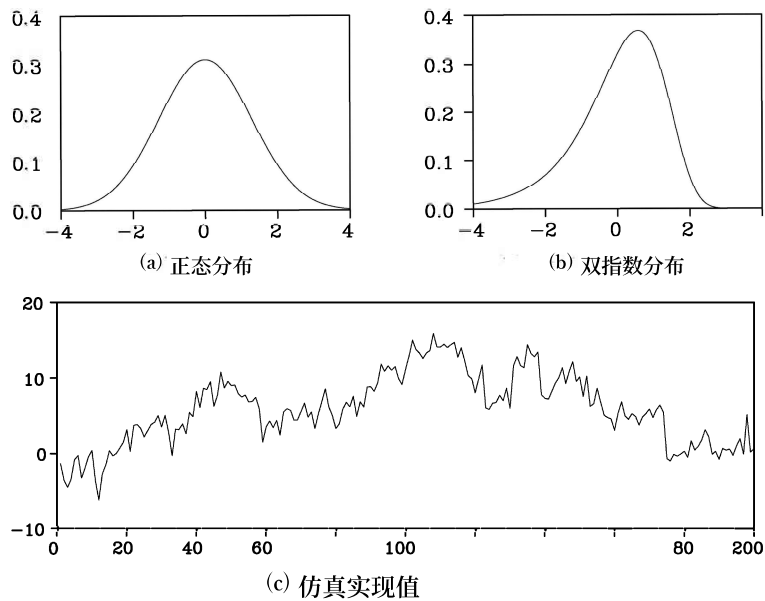


图 16.5 系统噪声的密度函数和仿真结果

**例：AR 模型：高斯型噪声和非高斯型噪声**

图 16.6 是对 1 阶 AR 模型

$$y_n = 0.9y_{n-1} + v_n \quad (16.7)$$

进行仿真的结果。其中，(a) 是  $v_n$  使用正态分布  $N(0,1)$  的情况，(b) 是使用  $t$  分布的情况。即使阶数和系数相同的 AR 模型，当噪声性质不同时，生成的时间序列看起来有很大差异。特别是在 (b) 中偶尔出现大的跳跃，整体的变动幅度超过 2 倍。

```
> # AR 模型的仿真
> # 正态噪声
> ar1 <- ngsim(n = 400, arcoef = 0.90, noisew = 1, paramw = 1, noisev = 1, paramv
= 1, seed = 555)
> plot(ar1, use = c(201, 400))
>
> # 非高斯型噪声 (t 分布)
> ar1 <- ngsim(n = 400, arcoef = 0.95, noisew = 1, paramw = 1, noisev = 2, paramv
= c(1,1), seed = 555)
> plot(ar1, use = c(201, 400))
```

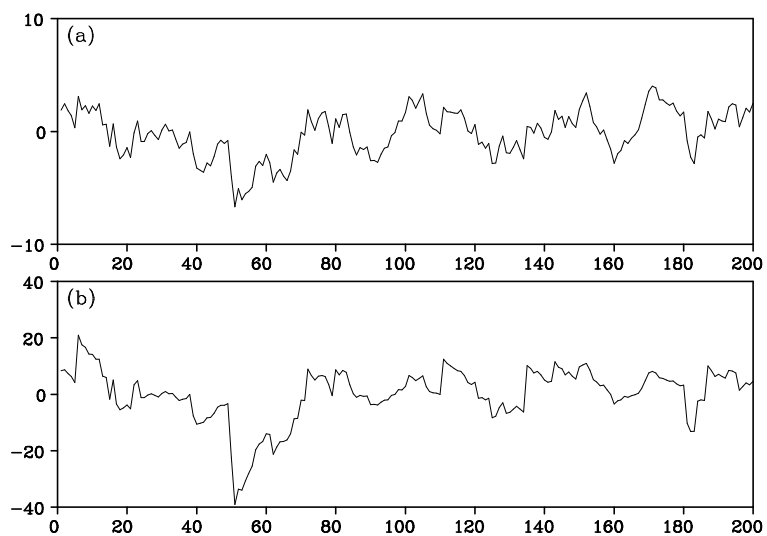


图 16.6 使用不同噪声分布的 AR 模型仿真。(a) 正态分布, (b)  $t$  分布

### ● 章末习题

1. 生成  $[0,1]$  上的均匀随机数  $U$  和  $V$  共 2 个, 令  $W = U + V$  时,  $W$  是什么样的随机数?
2. 在 AR 模型  $y_n = ay_{n-1} + v_n$ ,  $v_n \sim N(0,1)$  的仿真中, 通常生成  $m+n$  个, 仅使用后半部分的  $n$  个。  $m$  应设定为多大的值?
3. 给出基于当前已有的观测值, 对未来值进行仿真的方法。

## 附录 A

# 非线性最优化的算法

为简单起见, 考虑目标函数  $f(x)$  可以通过二阶 Taylor 展开表示为

$$f(x) = f(x^*) + (x - x^*)^T g(x^*) + \frac{1}{2}(x - x^*)^T H(x - x^*) \quad (\text{A.1})$$

的情况。其中,  $x = (x_1, \dots, x_m)^T$ ,

$$g(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_m} \end{bmatrix}, \quad H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_m \partial x_m} \end{bmatrix} \quad (\text{A.1})$$

由 (A.1) 可得

$$g(x) = g(x^*) + H(x - x^*) \quad (\text{A.1})$$

设  $x^*$  为使目标函数  $f(x)$  最小的点, 则  $g(x^*) = 0$  成立, 因此

$$g(x) = H(x - x^*) \quad (\text{A.1})$$

因此, 设当前近似值为  $x$ , 若  $H$  和  $g(x)$  已知, 则

$$x^* = x - H^{-1}g(x) \quad (\text{A.1})$$

即可求得解  $x^*$ 。

在实际问题中,  $f(x)$  能像 (A.1) 那样以二次形式表示的情况较少, 仅计算一次 (A) 往往无法得到解  $x^*$ 。此外, 矩阵  $H$  通常不是常数而是随  $x$  变化, 因此从适当的初始值  $x_0$  出发, 通过

$$x_k = x_{k-1} - H_{k-1}^{-1}g(x_{k-1}) \quad (\text{A.1})$$

反复迭代。这就是**牛顿法** (Newton-Raphson method)。

但是, 时间序列模型的似然函数是参数的复杂函数, 解析地求  $H_k$  往往很困难, 因此使用称为**拟牛顿法**的如下算法。

1. 令  $k = 0$ , 适当确定初始值  $x_0$  和  $H_0^{-1}$ 。通常取  $H_0 = I$ 。

2. 计算  $g(x_{k-1})$ 。
3. 通过  $h_k = H_{k-1}^{-1}g(x_{k-1})$  求方向向量。
4. 令  $x_k = x_{k-1} - \lambda_k h_k$ , 通过线性搜索确定使  $f(x_k)$  最小的  $\lambda_k$ 。
5. 通过下列公式之一求  $H_k^{-1}$  的估计值。(DFP 公式, BFGS 公式)

$$H_k^{-1} = H_{k-1}^{-1} + \frac{\Delta x_k \Delta x_k^T}{\Delta x_k^T \Delta g_k} - \frac{H_{k-1}^{-1} \Delta g_k \Delta g_k^T H_{k-1}^{-1}}{\Delta g_k^T H_{k-1}^{-1} \Delta g_k} \quad (\text{A.2})$$

$$H_k^{-1} = H_{k-1}^{-1} - \frac{\Delta x_k \Delta g_k^T H_{k-1}^{-1}}{\Delta x_k^T \Delta g_k} - \frac{H_{k-1}^{-1} \Delta g_k \Delta x_k^T}{\Delta x_k^T \Delta g_k} + \left\{ 1 + \frac{\Delta g_k^T H_{k-1}^{-1} \Delta g_k}{\Delta x_k^T \Delta g_k} \right\} \frac{\Delta x_k \Delta x_k^T}{\Delta x_k^T \Delta g_k} \quad (\text{A.2})$$

其中,  $\Delta x_k = x_k - x_{k-1}$ ,  $\Delta g_k = g(x_k) - g(x_{k-1})$ 。

6. 若未满足收敛条件, 则令  $k = k + 1$  返回第 (2) 步。

在这种拟牛顿法中, 无需计算  $H_k$  或  $H_k^{-1}$ , 而且通过线性搜索自动调节步长  $\lambda_k$ , 因此即使实际的目标函数在当前点  $x_k$  附近与二次曲面有较大偏离时, 也能在一定程度上应对。

## 付録 B

# Levinson 算法的推导

假设已给定自协方差函数  $C_k$ , ( $k = 0, 1, \dots$ ), 且  $m-1$  阶 AR 模型的 Yule-Walker 解已经求得。为了明确表示这是  $m-1$  阶模型的系数, 将其记为  $\hat{a}_1^{m-1}, \dots, \hat{a}_{m-1}^{m-1}$ , 预测误差记为  $v_n^{m-1}$ , 预测误差方差记为  $\hat{\sigma}_{m-1}^2$ 。接下来考虑利用这些估计值, 高效计算  $m$  阶 AR 模型参数的 Yule-Walker 估计值  $\hat{a}_1^m, \dots, \hat{a}_m^m, \hat{\sigma}_m^2$  的方法。

系数  $\hat{a}_1^{m-1}, \dots, \hat{a}_{m-1}^{m-1}$  满足  $m-1$  阶的 Yule-Walker 方程, 因此

$$C_k = \sum_{j=1}^{m-1} \hat{a}_j^{m-1} C_{j-k} \quad (k = 1, \dots, m-1) \quad (\text{B.0})$$

成立。这意味着关于  $v_n^{m-1}$ ,

$$\text{E}\{v_n^{m-1} y_{n-k}\} = \text{E}\left\{\left(y_n - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} y_{n-j}\right) y_{n-k}\right\} = 0 \quad (\text{B.0})$$

在  $k = 1, \dots, m-1$  的范围内成立。

这里, 将时间方向反转, 考虑用未来数据表示当前值的  $m-1$  阶后向 AR 模型 (backward AR model)

$$y_n = \sum_{i=1}^{m-1} a_i y_{n+i} + w_n^{m-1} \quad (\text{B.0})$$

由于  $C_{-k} = C_k$  成立, 该后向 AR 模型也满足相同的 Yule-Walker 方程 (B)。因此, 关于后向预测误差  $w_n^{m-1}$  也有

$$\text{E}\{w_{n-m}^{m-1} y_{n-k}\} = \text{E}\left\{\left(y_{n-m} - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} y_{n-m+j}\right) y_{n-k}\right\} = 0 \quad (\text{B.0})$$

对  $k = 1, \dots, m-1$  成立。

这里, 定义前向预测误差  $v_n^{m-1}$  和后向预测误差  $w_{n-m}^{m-1}$  的加权和为

$$\begin{aligned} z_n &\equiv v_n^{m-1} - \beta w_{n-m}^{m-1} \\ &\equiv y_n - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} y_{n-j} - \beta \left( y_{n-m} - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} y_{n-m+j} \right) \\ &= y_n - \sum_{j=1}^{m-1} \left( \hat{a}_j^{m-1} - \beta \hat{a}_{m-j}^{m-1} \right) y_{n-j} - \beta y_{n-m} \end{aligned} \quad (\text{B.-1})$$

$z_n$  是  $y_n, y_{n-1}, \dots, y_{n-m}$  的线性组合, 而且由 (B)、(B) 可知, 对  $k = 1, \dots, m-1$ ,

$$E\{z_n y_{n-k}\} = 0 \quad (\text{B.-1})$$

始终成立。因此, 若确定  $\beta$  的值使得对  $k = m$  也有  $E(z_n y_{n-m}) = 0$  成立, 则  $m$  阶 Yule-Walker 方程就得到满足。为此,

$$\begin{aligned} E(z_n y_{n-m}) &= E\left\{y_n - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} y_{n-j} - \beta \left(y_{n-m} - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} y_{n-m+j}\right)\right\} y_{n-m} \\ &= C_m - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} C_{m-j} - \beta \left(C_0 - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} C_j\right) = 0 \end{aligned} \quad (\text{B.-1})$$

由此可知, 令

$$\begin{aligned} \beta &= \left(C_0 - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} C_j\right)^{-1} \left(C_m - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} C_{m-j}\right) \\ &= \left(\hat{\sigma}_{m-1}^2\right)^{-1} \left(C_m - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} C_{m-j}\right) \end{aligned} \quad (\text{B.-1})$$

即可。因此, 将该  $\beta$  记为  $\hat{a}_m^m$ , 令

$$\hat{a}_j^m \equiv \hat{a}_j^{m-1} - \hat{a}_m^m \hat{a}_{m-j}^{m-1} \quad (\text{B.-1})$$

则由 (B.0)~(B.-1) 可知

$$E\left\{\left(y_n - \sum_{j=1}^m \hat{a}_j^m y_{n-j}\right) y_{n-k}\right\} = C_k - \sum_{j=1}^m \hat{a}_j^m C_{k-j} = 0 \quad (\text{B.-1})$$

对  $k = 1, \dots, m$  成立, 即  $\hat{a}_1^m, \dots, \hat{a}_m^m$  为  $m$  阶 Yule-Walker 方程的解。此外

$$\begin{aligned} \hat{\sigma}_m^2 &= C_0 - \sum_{j=1}^m \hat{a}_j^m C_j \\ &= C_0 - \sum_{j=1}^{m-1} (\hat{a}_j^{m-1} - \hat{a}_m^m \hat{a}_{m-j}^{m-1}) C_j - \hat{a}_m^m C_m \\ &= C_0 - \sum_{j=1}^{m-1} \hat{a}_j^{m-1} C_j - \hat{a}_m^m \left(C_m - \sum_{j=1}^{m-1} \hat{a}_{m-j}^{m-1} C_j\right) \\ &= \hat{\sigma}_{m-1}^2 - \hat{a}_m^m \left(C_m - \sum_{j=1}^{m-1} \hat{a}_{m-j}^{m-1} C_j\right) \end{aligned} \quad (\text{B.-3})$$

由 (B.-1) 可知  $C_m - \sum_{j=1}^{m-1} \hat{a}_{m-j}^{m-1} C_j = \hat{a}_m^m \hat{\sigma}_{m-1}^2$ , 利用此关系可得

$$\hat{\sigma}_m^2 = \hat{\sigma}_{m-1}^2 (1 - (\hat{a}_m^m)^2) \quad (\text{B.-2})$$

由此即可计算  $\hat{\sigma}_m^2$ 。

## 付録 C

# 卡尔曼滤波和平滑算法的推导

### C.1 卡尔曼滤波

[预测] 由  $x_n = F_n x_{n-1} + G_n v_n$  可得

$$\begin{aligned} x_{n|n-1} &= \mathbf{E}(x_n | Y_{n-1}) \\ &= \mathbf{E}(F_n x_{n-1} + G_n v_n | Y_{n-1}) \\ &= F_n \mathbf{E}(x_{n-1} | Y_{n-1}) \\ &= F_n x_{n-1|n-1} \end{aligned} \quad (\text{C.-2})$$

$$\begin{aligned} V_{n|n-1} &= \mathbf{E}(x_n - x_{n|n-1})^2 \\ &= \mathbf{E}(F_n(x_{n-1} - x_{n-1|n-1}) + G_n v_n)^2 \\ &= F_n \mathbf{E}(x_{n-1} - x_{n-1|n-1})^2 F_n^T + G_n \mathbf{E}(v_n^2) G_n^T \\ &= F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T \end{aligned} \quad (\text{C.-4})$$

其中, 为简单起见将  $xx^T$  简记为  $x^2$ 。

[滤波] 将  $y_n$  的预测误差记为  $\varepsilon_n$ , 由  $y_n = H_n x_n + w_n$

可得

$$\begin{aligned} \varepsilon_n &\equiv y_n - \mathbf{E}(y_n | Y_{n-1}) \\ &= H_n x_n + w_n - \mathbf{E}(H_n x_n + w_n | Y_{n-1}) \\ &= H_n x_n + w_n - H_n \mathbf{E}(x_n | Y_{n-1}) \\ &= H_n(x_n - x_{n|n-1}) + w_n \end{aligned} \quad (\text{C.-6})$$

因此

$$\text{Var}(\varepsilon_n) = H_n V_{n|n-1} H_n^T + R_n \quad (\text{C.-5})$$

$$\begin{aligned} \text{Cov}(x_n, \varepsilon_n) &= \text{Cov}(x_n, H_n(x_n - x_{n|n-1}) + w_n) \\ &= \text{Var}(x_n - x_{n|n-1}) H_n^T \\ &= V_{n|n-1} H_n^T \end{aligned} \quad (\text{C.-6})$$

此时, 利用  $Y_n = \{Y_{n-1}, y_n\} = Y_{n-1} \oplus \varepsilon_n$  以及在正态分布情况下  $x_n$  关于  $Y_n$  的条件期望

可以用正交投影表示这一性质, 可得

$$\begin{aligned} x_{n|n} &= E(x_n|Y_n) = \text{Proj}(x_n|Y_n) \\ &= \text{Proj}(x_n|Y_{n-1}, \varepsilon_n) \\ &= \text{Proj}(x_n|Y_{n-1}) + \text{Proj}(x_n|\varepsilon_n) \end{aligned} \quad (\text{C.-7})$$

$\text{Proj}(x_n|\varepsilon_n)$  可通过将  $x_n$  对  $\varepsilon_n$  进行回归得到, 由 (C.4)、(C.5) 可得

$$\begin{aligned} \text{Proj}(x_n|\varepsilon_n) &= \text{Cov}(x_n, \varepsilon_n) \text{Var}(\varepsilon_n)^{-1} \varepsilon_n \\ &= V_{n|n-1} H_n^T (H_n V_{n|n-1} H_n^T + R_n)^{-1} \varepsilon_n \\ &\equiv K_n \varepsilon_n \end{aligned} \quad (\text{C.-8})$$

因此

$$x_{n|n} = x_{n|n-1} + K_n \varepsilon_n \quad (\text{C.-8})$$

进一步

$$\begin{aligned} V_{n|n-1} &= E(x_n - x_{n|n-1})^2 \\ &= E(x_n - x_{n|n} + K_n \varepsilon_n)^2 \\ &= V_{n|n} + K_n \text{Var}(\varepsilon_n) K_n^T \end{aligned} \quad (\text{C.-9})$$

由此可得

$$\begin{aligned} V_{n|n} &= V_{n|n-1} - K_n H_n V_{n|n-1} \\ &= (I - K_n H_n) V_{n|n-1} \end{aligned} \quad (\text{C.-9})$$

## C.2 平滑

设  $\delta_{n+1} \equiv x_{n+1} - x_{n+1|n}$  为  $x_{n+1}$  的预测误差, 令

$$Z_n \equiv Y_n \oplus \delta_{n+1} \oplus \{v_{n+1}, \dots, v_N, w_{n+1}, \dots, w_N\}$$

此时

$$\begin{aligned} z_n &\equiv \text{Proj}(x_n|Z_n) \\ &= \text{Proj}(x_n|Y_n) + \text{Proj}(x_n|\delta_{n+1}) \\ &\quad + \text{Proj}(x_n|v_{n+1}, \dots, v_N, w_{n+1}, \dots, w_N) \end{aligned} \quad (\text{C.-10})$$

可以分解为上式, 且

$$\begin{aligned} \text{Proj}(x_n|Y_n) &= x_{n|n} \\ \text{Proj}(x_n|\delta_{n+1}) &= \text{Cov}(x_n, \delta_{n+1}) \text{Var}(\delta_{n+1})^{-1} \delta_{n+1} \\ \text{Proj}(x_n|v_{n+1}, \dots, v_N, w_{n+1}, \dots, w_N) &= 0 \end{aligned} \quad (\text{C.-10})$$

成立。进一步

$$\begin{aligned} \text{Var}(\delta_{n+1}) &= V_{n+1|n} \\ \text{Cov}(x_n, \delta_{n+1}) &= \text{Cov}(x_n, F_{n+1}(x_n - x_{n|n}) + G_{n+1}v_{n+1}) \\ &= E(x_n - x_{n|n})^2 F_{n+1}^T \\ &= V_{n|n} F_{n+1}^T \end{aligned} \quad (\text{C.-11})$$

因此, 令  $A_n = V_{n|n}F_{n+1}^T V_{n+1|n}^{-1}$ , 则

$$z_n = x_{n|n} + A_n(x_{n+1} - x_{n+1|n}) \quad (\text{C.-10})$$

这里, 考虑到  $Z_N$  生成  $Y_N$ , 可得

$$\begin{aligned} x_{n|N} &= \text{Proj}(x_n|Y_N) \\ &= \text{Proj}(\text{Proj}(x_n|Z_N)|Y_N) \\ &= \text{Proj}(z_n|Y_N) \\ &= x_{n|n} + A_n(x_{n+1|N} - x_{n+1|n}) \end{aligned} \quad (\text{C.-12})$$

进一步, 由

$$x_n - x_{n|N} + A_n x_{n+1|N} = x_n - x_{n|n} + A_n x_{n+1|n} \quad (\text{C.-12})$$

利用  $E\{(x_n - x_{n|N})x_{n+1|N}^T\} = E\{(x_n - x_{n|n})x_{n+1|n}^T\} = 0$ , 可得

$$V_{n|N} + A_n E\{x_{n+1|N}x_{n+1|N}^T\}A_n^T = V_{n|n} + A_n E\{x_{n+1|n}x_{n+1|n}^T\}A_n^T \quad (\text{C.-12})$$

这里, 进一步利用

$$\begin{aligned} E\{(x_{n+1} - x_{n+1|N})x_{n+1|N}^T\} &= 0 \\ E\{(x_n - x_{n|n})x_{n+1|n}^T\} &= 0 \end{aligned}$$

可得

$$\begin{aligned} E\{x_{n+1|N}x_{n+1|N}^T\} &= E\{(x_{n+1|N} - x_{n+1} + x_{n+1})(x_{n+1|N} - x_{n+1} + x_{n+1})^T\} \\ &= V_{n+1|N} + E\{x_{n+1}x_{n+1}^T\} + 2E\{(x_{n+1|N} - x_{n+1})x_{n+1}^T\} \\ &= V_{n+1|N} + E\{x_{n+1}x_{n+1}^T\} - 2E\{(x_{n+1|N} - x_{n+1})(x_{n+1|N} - x_{n+1})^T\} \\ &= E\{x_{n+1}x_{n+1}^T\} - V_{n+1|N} \end{aligned} \quad (\text{C.-17})$$

$$E\{x_{n+1|n}x_{n+1|n}^T\} = E\{x_{n+1}x_{n+1}^T\} - V_{n+1|n} \quad (\text{C.-16})$$

将此代入 (C.2) 可得

$$V_{n|N} = V_{n|n} + A_n(V_{n+1|N} - V_{n+1|n})A_n^T \quad (\text{C.-16})$$

## 附录 D

# 粒子滤波的算法

### (a) 一步预测

假设已给定可视为状态  $x_{n-1}$  的条件分布  $p(x_{n-1}|Y_{n-1})$  实现值的  $m$  个粒子  $\{f_{n-1}^{(1)}, \dots, f_{n-1}^{(m)}\}$ , 以及可视为系统噪声  $v_n$  独立实现值的粒子  $\{v_n^{(1)}, \dots, v_n^{(m)}\}$ 。即假设对  $i = 1, \dots, m$ ,

$$f_{n-1}^{(i)} \sim p(x_{n-1}|Y_{n-1}), \quad v_n^{(i)} \sim q(v) \quad (\text{D.0})$$

成立。此时,  $x_n$  的一步预测分布为

$$\begin{aligned} p(x_n|Y_{n-1}) &= \int \int p(x_n, x_{n-1}, v_n|Y_{n-1}) dx_{n-1} dv_n \\ &= \int \int p(x_n|x_{n-1}, v_n, Y_{n-1}) p(v_n|x_{n-1}, Y_{n-1}) p(x_{n-1}|Y_{n-1}) dx_{n-1} dv_n \end{aligned} \quad (\text{D.0})$$

由于系统噪声与之前的状态和观测值独立, 有  $p(v_n|x_{n-1}, Y_{n-1}) = p(v_n)$ , 另外, 由于  $x_n$  仅依赖于  $x_{n-1}$  和  $v_n$ , 有  $p(x_n|x_{n-1}, v_n, Y_{n-1}) = p(x_n|x_{n-1}, v_n) = \delta(x_n - F(x_{n-1}, v_n))$ , 因此

$$p(x_n|Y_{n-1}) = \int \int \delta(x_n - F(x_{n-1}, v_n)) p(v_n) p(x_{n-1}|Y_{n-1}) dx_{n-1} dv_n \quad (\text{D.-1})$$

因此, 当  $p(v_n)$  和  $p(x_{n-1}|Y_{n-1})$  的实现值  $\{v_n^{(j)}\}$  和  $\{f_{n-1}^{(j)}\}$  已知时,  $p(x_n|Y_{n-1})$  的实现值  $\{p_n^{(j)}\}$  可由

$$p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)}) \quad (\text{D.-1})$$

求得。

### (b) 滤波

当分布  $p(x_n|Y_{n-1})$  的  $m$  个独立实现值  $p_n^{(1)}, \dots, p_n^{(m)}$  已知时, 相当于分布  $p(x_n|Y_{n-1})$  由经验分布函数

$$P_n(x) = \frac{1}{m} \sum_{i=1}^m I(x, p_n^{(i)}), \quad (\text{D.-1})$$

近似。其中,  $I(x, a)$  是当  $x < a$  时  $I(x, a) = 0$ , 其他情况  $I(x, a) = 1$  的函数。这意味着预测分布  $p(x_n|Y_{n-1})$  由概率函数

$$\Pr(x_n = p_n^{(j)}|Y_{n-1}) = \frac{1}{m}, \quad j = 1, \dots, m \quad (\text{D.-1})$$

近似。此时, 给定观测值  $y_n$  后, 后验分布可由

$$\begin{aligned} \Pr(x_n = p_n^{(j)}|Y_n) &= \Pr(x_n = p_n^{(j)}|Y_{n-1}, y_n) \\ &= \lim_{\Delta y \rightarrow 0} \frac{\Pr(x_n = p_n^{(j)}, y_n \leq y \leq y_n + \Delta y | Y_{n-1})}{\Pr(y_n \leq y \leq y_n + \Delta y | Y_{n-1})} \\ &= \frac{p(y_n | p_n^{(j)}) \Pr(x_n = p_n^{(j)} | Y_{n-1})}{\sum_{i=1}^m p(y_n | p_n^{(i)}) \Pr(x_n = p_n^{(i)} | Y_{n-1})} \\ &= \frac{\alpha_n^{(j)} \cdot \frac{1}{m}}{\sum_{i=1}^m \alpha_n^{(i)} \cdot \frac{1}{m}} = \frac{\alpha_n^{(j)}}{\sum_{i=1}^m \alpha_n^{(i)}}. \end{aligned} \quad (\text{D.-3})$$

求得。与  $\Pr(x_n = p_n^{(j)}|Y_n)$  对应的累积分布函数

$$\frac{1}{\sum_{i=1}^m \alpha_n^{(i)}} \sum_{i=1}^m \alpha_n^{(i)} I(x, p_n^{(i)}) \quad (\text{D.-3})$$

仅在  $m$  个点  $p_n^{(1)}, \dots, p_n^{(m)}$  处有与  $\alpha_n^{(1)}, \dots, \alpha_n^{(m)}$  成比例的量 (即 (15.2.4) 式右边) 的跳跃。

由 (D) 式, 已经得到了滤波分布的近似, 但为了在下一时点进行预测步骤的计算 (D), 需要用  $m$  个粒子  $f_n^{(1)}, \dots, f_n^{(m)}$  重新表示。这相当于将 (D) 的分布用经验分布函数

$$\frac{1}{m} \sum_{i=1}^m I(x, f_n^{(i)}). \quad (\text{D.-2})$$

表示, 通过以概率

$$\Pr(f_n^{(j)} = p_n^{(i)}|Y_n) = \frac{\alpha_n^{(i)}}{\alpha_n^{(1)} + \dots + \alpha_n^{(m)}}, \quad j = 1, \dots, m \quad (\text{D.-1})$$

对  $m$  个实现值  $\{p_n^{(1)}, \dots, p_n^{(m)}\}$  进行重采样, 即可求得  $\{f_n^{(1)}, \dots, f_n^{(m)}\}$ 。

假设  $\Pr(x_1 = s_{1|n-1}^{(j)}, \dots, x_{n-1} = s_{n-1|n-1}^{(j)} | Y_{n-1}) = 1/m$  以及  $v_n^{(j)} \sim q(v)$  成立, 定义  $(p_{1|n-1}^{(j)}, \dots, p_{n|n-1}^{(j)})$  如下。

$$p_{i|n-1}^{(j)} = \begin{cases} s_{i|n-1}^{(j)}, & i = 1, \dots, n-1 \\ F(s_{n-1|n-1}^{(j)}, v_n^{(j)}), & i = n \end{cases} \quad (\text{D.0})$$

此时,  $(p_{1|n-1}^{(j)}, \dots, p_{n|n-1}^{(j)})$  可视为给定观测值  $Y_{n-1}$  时  $(x_1, \dots, x_n)$  的联合分布的实现值。

接下来, 当观测值  $y_n$  给定时, 分布  $\Pr(x_1 \leq p_{1|n-1}^{(j)}, \dots, x_n \leq p_{n|n-1}^{(j)} | Y_{n-1})$  按以下方式更新。

$$\begin{aligned}
 & \Pr(x_1 = p_{1|n-1}^{(j)}, \dots, x_n = p_{n|n-1}^{(j)} | Y_n) \\
 &= \Pr(x_1 = p_{1|n-1}^{(j)}, \dots, x_n = p_{n|n-1}^{(j)} | Y_{n-1}, y_n) \\
 &= \frac{p(y_n | x_1 = p_{1|n-1}^{(j)}, \dots, x_n = p_{n|n-1}^{(j)})}{p(y_n | Y_{n-1})} \\
 &\quad \times \Pr(x_1 = p_{1|n-1}^{(j)}, \dots, x_n = p_{n|n-1}^{(j)} | Y_{n-1}) \\
 &= \frac{p(y_n | p_{n|n-1}^{(j)}) \Pr(x_1 = p_{1|n-1}^{(j)}, \dots, x_n = p_{n|n-1}^{(j)} | Y_{n-1})}{p(y_n | Y_{n-1})}. \tag{D.-3}
 \end{aligned}$$

由于  $p_{n|n-1}^{(j)}$  与滤波算法 (D) 中的  $p_n^{(j)}$  相同, 平滑分布  $p(x_1, \dots, x_n | Y_n)$  只需对  $m$  个  $n$  维向量  $(p_{1|n-1}^{(j)}, \dots, p_{n|n-1}^{(j)})^T$ ,  $j = 1, \dots, m$ , 以与滤波相同的权重进行重采样即可。

## 参考文献

- [1] 赤池弘次, 北川源四郎編, 時系列解析の実際 I, II, 朝倉書店, 1994, 1995.
- [2] 赤池弘次, 中川東一郎, ダイナミックシステムの統計的解析と制御, サイエンス社, 1972.
- [3] Akaike, H., “A Bayesian extension of the minimum AIC procedure of autoregressive model fitting”. *Biometrika*, **66**, 237-24, 1979.
- [4] Akaike, H. ほか, TIMSAC-78, Computer Science Monographs, No.11, 1979.
- [5] Bartlett, M.S., On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series, *Supplement to the Journal of the Royal Statistical Society*, **8**, 1946.
- [6] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.
- [7] Brockwell, P. J., Davis, R. A., and Fienberg, S. E. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media.
- [8] Doucet, A., de Freitas, and Gordon, N., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001.
- [9] 尾崎統, 北川源四郎編, 時系列解析の方法, 朝倉書店, 1990,
- [10] 片山徹, 応用カルマンフィルタ, 朝倉書店, 1983.
- [11] Kitagawa, G. and Gersch, W., *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, No.106, Springer Verlag, 1996.
- [12] 坂元慶行, 石黒真木夫, 北川源四郎, 情報量統計学, 共立出版, 1983.
- [13] Sakamoto, Y., Ishiguro, M., Kitagawa, G., *Akaike Information Criterion Statistics*, D. Reidel Publishing Company, 1986.
- [14] 日野幹雄, スペクトル解析, 朝倉書店, 1977.
- [15] 広松毅, 浪花貞夫, 経済時系列分析, 朝倉書店, 1990.

- [16] 得丸英勝, 添田喬, 中溝高好, 秋月影雄, 計数・測定=ランダムデータ処理の理論と応用, 培風館, 1982.
- [17] Harvey, A.C., Time Series Models, Philip Allan Publishers, 1981.  
(邦訳) 国友直人, 山本拓, 時系列モデル入門, 東京大学出版会, 1985.
- [18] 今野浩, 山下浩, 非線形計画法, 日科技連, 1978.
- [19] 伏見正則, 乱数, 東京大学出版会, 1989.
- [20] 小西貞則, 北川源四郎, 情報量規準, 朝倉書店, 2004.
- [21] Anderson, B.D.O. and Moore, J., B., *Optimal Filtering*, Prentice-Hall, Inc., 1979.
- [22] Kitagawa, G., “Non-Gaussian state-space modeling of nonstationary time series (with discussion),” *J. Amer. Statist. Assoc.*, 82, 1987, 1032–1063.
- [23] Kitagawa, G., “A nonlinear smoothing method for time series analysis”, *Statistica Sinica*, 1, 371–388, 1991.
- [24] Kitagawa, G., “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models,” *Journal of Computational and Graphical Statistics*, Vol.5, No.1, 1996 1–25.
- [25] Kitagawa, G. and Gersch, W., “A smoothness priors time varying AR coefficient modeling of nonstationary time series”, *IEEE Trans. on Auto. Control*, **AC-30**, 48–56, 1985.
- [26] Kitagawa, G. and Gersch, W., *Smoothness Priors Analysis of Time Series*, (Vol. 116, Springer, New York, 1996.
- [27] Knuth, D.E, *The Art of Computer Programming* (日本語版), ASCII Addison Wesley Programming Series, 2004.
- [28] Lewis, T.G. and Payne, W.H, “Generalized feedback shift register pseudorandom number algorithm,” *JACM*, **20** 456–468. 1973.
- [29] Matsumoto, M. and Nishimura, T., “Mersenne twister, A 623-dimensionally equidistributed uniform pseudorandom number generator,” *ACM Trans. on Modeling and Computer Simulations*, 1998.
- [30] Shumway, R. H. and Stoffer, D. S. *Time Series Analysis and its Applications: with R Examples*. Springer, 2017.
- [31] West, M. and Harrison, J., *Bayesian Forecasting and Dynamic Models*, Springer,

1996.

## 章末习题解答

## 第 1 章

- 适当选择采样间隔是很重要的。间隔过大则无法捕捉时间序列的特征，而过小则用于建模的参数可能被消耗在非本质的部分，反而无法表达关键特征。主要周期分量的  $1/5 \sim 1/10$  左右是一个参考标准。
- (解答示例) 记录某商店每天销售的速溶咖啡个数的时间序列。由于是正值时间序列，分布呈上下不对称。此外，由于打折或非价格促销策略，销量有时会显著增加。
- (1) 将 (1.1) 式关于  $y$  求解得  $y = e^z / (1 + e^z)$ 。  
(2)  $z = \log\{(y - a)/(b - y)\}$ , 逆变换为  $y = (a + be^z)/(1 + e^z)$ 。
- 当原始时间序列包含噪声并可表示为  $y_n = a + bn + \varepsilon_n$  时，差分后  $\Delta y_n = b + \varepsilon_n - \varepsilon_{n-1}$ ，虽然趋势被去除，但噪声的结构反而变得更复杂。
- 前一年的趋势变化会影响今年的修正值。例如，如果上一年度中途趋势的斜率增加，即使今年度趋势斜率保持不变，修正后看起来斜率也会从中途开始减小。
- (1) 当  $y_n = T_n + w_n$ ,  $T_n = a + bn$ ,  $w_n \sim N(0, \sigma^2)$  为白噪声时

$$\begin{aligned}\hat{T}_n &= \frac{1}{3}(y_{n-1} + y_n + y_{n+1}) \\ &= \frac{1}{3}(T_{n-1} + T_n + T_{n+1}) + \frac{1}{3}(w_{n-1} + w_n + w_{n+1})\end{aligned}$$

其中，右边第一项等于  $T_n$ 。而右边第 2 项的均值为 0，由  $E(w_{n-1} + w_n + w_{n+1})^2 = 3\sigma^2$  可知方差为  $\sigma^2/3$ 。

- 移动平均可以通过增大项数获得平滑的估计值，但具有无法检测急剧结构变化以及对异常值敏感的缺点。移动中位数具有相反的性质。

## 第 2 章

- 高斯（正态）分布的形状完全由均值和方差决定。因此，如果弱平稳性保证均值和方差不变，则分布本身也不变，强平稳性成立。
- 将每个时刻服从标准柯西分布的序列视为时间序列时，其概率分布恒定且满足 (2.4)，因此是强平稳的。但由于柯西分布不存在均值和方差，不能说弱平稳性成立。

3.  $C_k = E(y_n - \mu)(y_{n+k} - \mu) = E(y_{n+k} - \mu)(y_n - \mu) = E(y_n - \mu)(y_{n-k} - \mu) = C_k$ . 第

三个等式仅在  $y_n$  平稳时成立。

4.  $C_0 = Ey_n^2 = Ev_n^2 - 2cEv_nv_{n-1} + c^2E_{v_{n-1}}^2 = 1 + c^2$ ,  $C_1 = E(v_n - cv_{n-1})(v_{n-1} - cv_{n-2}) = -c$ ,  $E_k = 0 (k \leq 2)$ 。

5. 设  $\alpha = (\alpha_1, \dots, \alpha_k)$  为不全为 0 的任意向量, 则

$$\begin{aligned} \alpha C \alpha^T &= \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j C_{i-j} = \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j E(y_{n-i} y_{n-j}) \\ &= E \left( \sum_{i=1}^k \alpha_i y_{n-i} \right)^2 \geq 0 \end{aligned}$$

6. (1) 对 (2.21) 式两边取期望得

$$E[\hat{C}_k(i, j)] = \frac{1}{N} \sum_{n=k+1}^N C_k(i, j) = \frac{N-k}{N} C_k(i, j)$$

(2) 由 (2.21) 式定义的样本互协方差是非负定的, 因此在 AR 模型的估计等中能保证平稳性, 这是其优点。将分母替换为  $N$  可以得到  $C_k(i-j)$  的无偏估计量, 但会失去这个优点。

7. (1) 当数据数为  $n$  时  $\hat{C}_0 \sim N(0, 2\sigma^2/n)$ ,  $\hat{C}_k \sim N(0, \sigma^2/n)$ ,  $\hat{R}_k \sim N(0, n^{-1})$

(2) 检验对所有  $k$  是否成立  $|\hat{R}_k| < \frac{\alpha}{\sqrt{n}}$ 。  $\alpha$  例如取 2。

### 第 3 章

1. 由  $e^{-2\pi i k f} = \cos(2\pi k f) - i \sin(2\pi k f)$  可得  $p(f) = \sum_{k=-\infty}^{\infty} C_k \cos(2\pi k f) -$

$i \sum_{k=-\infty}^{\infty} C_k \sin(2\pi k f)$ , 但由于  $C_k$  是偶函数,  $\sin(2\pi k f)$  是奇函数,  $C_k \sin(2\pi k f)$  为奇函数, 上式右边第二项为 0。

2.  $p(f) = 1 + C^2 - 2C \cos(2\pi f)$

3.

$$\begin{aligned} p(f) &= \sum_{k=-\infty}^{\infty} C_k e^{-2\pi i k f} = \frac{\sigma^2}{1-a^2} \sum_{k=-\infty}^{\infty} a^{|k|} e^{-2\pi i k f} \\ &= \frac{\sigma^2}{1-a^2} \left( \frac{1}{1-ae^{-2\pi i f}} + \frac{1}{1-ae^{2\pi i f}} - 1 \right) \\ &= \frac{\sigma^2}{1-a^2} \frac{1-a^2}{|1-ae^{-2\pi i f}|^2} = \frac{\sigma^2}{|1-ae^{-2\pi i f}|^2} \end{aligned}$$

4.

$$\begin{aligned}
& \frac{1}{N} \left| \sum_{n=1}^N y_n \exp(-2\pi i(n-1)j/N) \right|^2 \\
&= \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^N y_n y_m \exp(-2\pi i(n-1)j/N) \exp(2\pi i(m-1)j/N) \\
&= \frac{1}{N} \left\{ \sum_{n=1}^N y_n^2 + 2 \sum_{k=1}^{N-1} \sum_{n=k+1}^N y_n y_{n-k} \exp(-2\pi i k j/N) \right\} \\
&= \hat{C}_0 + 2 \sum_{k=1}^{N-1} \hat{C}_k \exp(-2\pi i k j/N)
\end{aligned}$$

5. 对 (3.9) 式两边取期望得

$$\begin{aligned}
E[\hat{p}(f)] &= \sum_{k=1-N}^{N-1} E[\hat{C}_k] e^{-2\pi i k f} = \sum_{k=1-N}^{N-1} \frac{N-|k|}{N} C_k e^{-2\pi i k f} \\
&= \sum_{k=1-N}^{N-1} C_k e^{-2\pi i k f} - \frac{1}{N} \sum_{k=1-N}^N |k| C_k e^{-2\pi i k f}
\end{aligned}$$

## 第 4 章

1. 对数似然为  $\ell = \sum_{i=1}^n \log f(m_i|\lambda) = \sum_{i=1}^n \{-\lambda + m_i \log \lambda - \log m_i!\}$ 。因此

$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^n \left\{ -1 + \frac{m_i}{\lambda} \right\} = -n + \frac{1}{\lambda} \sum_{i=1}^n m_i = 0$$

可得  $\hat{\lambda} = n^{-1} \sum_{i=1}^n m_i$ 。2. (1)  $\hat{\mu}_0 = \frac{1}{n+m} \left( \sum_{i=1}^n x_i + \sum_{i=1}^m y_i \right)$ ,  $\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\hat{\mu}_2 = \frac{1}{m} \sum_{i=1}^m y_i$ ,  $\hat{\sigma}_0^2 =$ 

$$\frac{1}{n+m} \left( \sum_{i=1}^n (x_i - \hat{\mu}_0)^2 + \sum_{i=1}^m (y_i - \hat{\mu}_0)^2 \right),$$

$$\hat{\tau}_0^2 = \frac{1}{n+m} \left( \sum_{i=1}^n (x_i - \hat{\mu}_1)^2 + \sum_{i=1}^m (y_i - \hat{\mu}_2)^2 \right) \text{ 此时}$$

$$\text{AIC}_0 = (n+m)(\log 2\pi \hat{\sigma}_0^2 + 1) + 2 \times 2$$

$$\text{AIC}_1 = (n+m)(\log 2\pi \hat{\tau}_0^2 + 1) + 2 \times 3$$

进行比较, 若  $\text{AIC}_0 < \text{AIC}_1$ , 则视为均值相等更好。(2)  $\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_0)^2$ ,  $\hat{\sigma}_2^2 = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{\mu}_0)^2$ ,  $\text{AIC}_2 = n(\log 2\pi \hat{\sigma}_1^2 + 1) + m(\log 2\pi \hat{\sigma}_2^2 + 1) + 2 \times 3$ 。若  $\text{AIC}_0 < \text{AIC}_2$ , 则视为方差相等更好。这里,由  $\sum_{k=-\infty}^{\infty} |C_k| < \infty$  可知, 对任意  $\varepsilon > 0$ , 存在满足  $\sum_{k=m+1}^{\infty} |C_k| < \varepsilon/2$  的  $m$ 。

此时

$$\begin{aligned} \left| \frac{1}{N} \sum_{k=-\infty}^{\infty} |k| C_k e^{-2\pi i k f} \right| &\leq \frac{2}{N} \sum_{k=1}^{\infty} k |C_k| \leq \frac{2}{N} \sum_{k=1}^m k C_0 + 2 \sum_{k=m+1}^{\infty} \frac{k}{N} |C_k| \\ &\leq \frac{2}{N} \frac{m(m+1)}{2} C_0 + \varepsilon \end{aligned}$$

当  $N$  增大时右边第 1 项可以任意小, 因此  $E[\hat{p}(f)] \rightarrow p(f)$ 。

3. 正态分布模型的 AIC 为: 令  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i$ ,  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2$ , 则  $AIC_0 = n \log 2\pi \hat{\sigma}^2 + n + 2 \times 2$ 。

柯西分布模型的 AIC 为: 将 (4.36) 式最大化求最大似然估计值  $\hat{\mu}$ 、 $\hat{\tau}^2$  时,  $AIC_1 = n \log \hat{\tau}^2 - 2n \log \pi - \sum_{i=1}^n \log \{(y_i - \hat{\mu})^2 + \hat{\tau}^2\} + 2 \times 2$ 。若  $AIC_0 < AIC_1$ , 则判断正态分布模型更好。

4. (1) 由二项分布模型  $f(m|n, p) = {}_n C_m p^m (1-p)^{n-m}$  可得对数似然  $\ell(p) = \log {}_n C_m + m \log p + (n-m) \log(1-p)$ 。因此由  $\frac{d\ell}{dp} = 0$  得  $\frac{m}{p} - \frac{n-m}{1-p} = 0$ , 解得  $\hat{p} = \frac{m}{n}$ 。

(2) (1) 的模型 AIC 为  $AIC_1 = -2 \log {}_n C_m - 2m \log \hat{p} - 2(n-m) \log(1-\hat{p}) + 2$ 。另一方面, 假设等概率时  $p = \frac{1}{2}$ , 故  $AIC_0 = -2 \log {}_n C_m + 2n \log 2$ 。若  $AIC_0 < AIC_1$ , 则视为等概率更好。

5. 由  $\frac{d}{d\theta} \log f(y|\theta) = f(y|\theta)^{-1} \frac{df(y|\theta)}{d\theta}$  可得

$$\begin{aligned} \frac{d^2}{d\theta^2} \log f(y|\theta) &= \frac{d}{d\theta} \left( \frac{d}{d\theta} \log f(y|\theta) \right) \\ &= \frac{d}{d\theta} \left( f(y|\theta)^{-1} \frac{df}{d\theta} \right) = -f(y|\theta)^{-2} \left( \frac{df}{d\theta} \right)^2 + f(y|\theta)^{-1} \frac{d^2 f}{d\theta^2} \\ &= - \left( \frac{d}{d\theta} \log f(y|\theta) \right)^2 + f(y|\theta)^{-1} \frac{d^2 f}{d\theta^2} \end{aligned}$$

因此,

$$\begin{aligned} J &= -E \left[ \frac{d^2}{d\theta^2} \log f(y|\theta) \right] \\ &= E \left( \frac{d}{d\theta} \log f(y|\theta) \right)^2 - E \left[ f(y|\theta)^{-1} \frac{d^2 f}{d\theta^2} \right] \\ &= I - \int_{-\infty}^{\infty} \frac{d^2 f}{d\theta^2} dy = I - \frac{d^2}{d\theta^2} \int_{-\infty}^{\infty} f(y|\theta) dy = I \end{aligned}$$

6. 由  $z = k(y) = \log y$ ,  $f(z) = (2\pi\sigma^2)^{-1} \exp\{-(z-\mu)^2/2\sigma^2\}$  可得  $\frac{dk}{dy} = y^{-1}$ 。因此

$$g(y) = \left| \frac{dk}{dy} \right| f(k(y)) = \frac{1}{y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log y - \mu)^2}{2\sigma^2}}$$

1. 令  $\mu$  的最大似然估计值为  $\hat{\mu}$ , 参数个数为 1, 因此

$$\text{AIC} = n \log 2\pi\sigma^2 - \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu})^2 + 2$$

2. (1) 由  $S = \sum_{n=1}^N (y_n - ax_n^2 - bx_n)^2$  可得

$$\begin{aligned} \frac{dS}{da} &= -2 \sum_{n=1}^N x_n^2 (y_n - ax_n^2 - bx_n) \\ &= -2 \left( \sum_{n=1}^N x_n^2 y_n - a \sum_{n=1}^N x_n^4 - b \sum_{n=1}^N x_n^3 \right) = 0 \\ \frac{dS}{db} &= -2 \sum_{n=1}^N x_n (y_n - ax_n^2 - bx_n) \\ &= -2 \left( \sum_{n=1}^N x_n y_n - a \sum_{n=1}^N x_n^3 - b \sum_{n=1}^N x_n^2 \right) = 0 \end{aligned}$$

解之得

$$\begin{aligned} \hat{a} &= \frac{\sum x_n^2 y_n \sum x_n^2 - \sum x_n y_n \sum x_n^3}{\sum x_n^4 \sum x_n^2 - (\sum x_n^3)^2} \\ \hat{b} &= \frac{\sum x_n y_n \sum x_n^4 - \sum x_n^2 y_n \sum x_n^3}{\sum x_n^4 \sum x_n^2 - (\sum x_n^3)^2} \end{aligned}$$

(2)  $y_n = ax_n(x_n - b) + \varepsilon_n$ . 通过数值最优化, 对两个变量  $a, b$  最小化  $S = \sum_{n=1}^N (y_n - ax_n^2 - abx_n)^2$  可求得最小二乘估计值。

## 第 6 章

1. (1) 在 (6.21) 式中令  $m = 1$ ,  $1 - a_1 B = 0$  的根在单位圆外为条件。因此  $|B| = |a_1^{-1}| > 1$  即  $|a_1| < 1$ .

(2) 由 (6.21) 式,  $m = 2$  时根为  $B = (a_1 \pm \sqrt{a_1^2 + 4a_2})/2$ . 因此平稳性条件为: 当  $a_1^2 + 4a_2 \geq 0$  时  $a_2 < 1 - a_1$  且  $a_2 < 1 + a_1$ , 当  $a_1^2 + 4a_2 < 0$  时  $a_2 > -1$ . 综上, 满足平稳性的区域为由三条直线  $a_2 < 1 - a_1$ 、 $a_2 < 1 + a_1$ 、 $a_2 > -1$  围成的三角形。

2. (1)  $\sigma^2$

(2) 由  $y_{n+2} = ay_{n+1} + v_{n+2} = a^2 y_n + av_{n+1} + v_{n+2}$  可得  $y_{n+2|n} = a^2 y_n$ . 因此  $E(y_{n+2} - y_{n+2|n})^2 = a^2 E v_{n+1}^2 + E v_{n+2}^2 = (1 + a^2)\sigma^2$

(3) 可以表示为  $y_{n+k} = v_{n+k} + g_1 v_{n+k-1} + g_2 v_{n+k-2} + \dots$ . 其中, 由 AR(1) 的脉

冲响应函数得  $g_i = a^i$ 。此时  $y_{n+k|n} = g_k v_n + g_{k+1} v_{n+1} + \dots$ ，因此

$$\begin{aligned} E(y_{n+k} - y_{n+k|n})^2 &= E(v_{n+k} + g_1 v_{n+k-1} + \dots + g_{k-1} v_{n+1})^2 \\ &= (1 + g_1^2 + \dots + g_{k-1}^2) \sigma^2 \\ &= (1 + a^2 + \dots + a^{2(k-1)}) \sigma^2 = \frac{1 - a^{2k}}{1 - a^2} \sigma^2 \end{aligned}$$

3. (1) 由  $C_0 = -0.9C_1 + 1$ ,  $C_1 = -0.9C_0$  得  $C_0 = 0.81C_0 + 1$ ,  $C_0 = 1/0.19 = 5.26$ ,

$$C_k = 5.26 \times (-0.9)^k$$

(2)  $C_0 = 25/7$ ,  $C_1 = 75/28$ ,  $C_3 = 15/14$ ,  $C_4 = -9/28$

(3)  $C_0 = 1 + b^2$ ,  $C_1 = -b$ ,  $C_k = 0$  ( $k > 1$ )

(4) 令  $E y_n v_{n-k} = g_k$ , 则  $g_0 = 1$ ,  $g_1 = a - b$ 。因此  $C_0 = aC_1 + g_0 - bg_1$ ,  $C_1 = aC_0 - bg_0$

可得

$$C_0 = \frac{1 - 2ab + b^2}{1 - a^2}, \quad C_k = \frac{(1 - ab)(a - b)}{1 - a^2} a^{k-1}$$

4. (1) 将  $v_n = y_n - ay_{n-1}$  代入  $v_n = bv_{n-1} + w_n$  的两边得

$$y_n = (a + b)y_{n-1} - aby_{n-2} + w_n$$

即得到 2 阶 AR 模型。

(2)  $y_n$  的自协方差函数为  $a^k(1 - a^2)^{-1}$  ( $k = 0, 1, 2, 3$ )。因此  $C_0 = (1 - a^2)^{-1} + 0.01$ ,  $C_k = a^k(1 - a^2)^{-1}$

5. (1) 将  $C_0 = 1 + b^2$ ,  $C_1 = -b$ ,  $C_k = 0$  ( $k > 1$ ) 代入  $p(f) = \sum_{k=-\infty}^{\infty} C_k e^{-2\pi i k f}$  得

$$p(f) = (1 + b^2) - be^{-2\pi i f} - be^{2\pi i f} = |1 - be^{-2\pi i f}|^2$$

(2) 由  $dp/df = -4\pi a \sin(2\pi f) / \{1 - 2a \cos(2\pi f) + a^2\}^2 = 0$  得  $\sin(2\pi f) = 0$ , 即  $f = 0$  或  $1/2$  时  $p(f)$  取极值。当  $a > 0$  时在  $f = 0$  取最大值, 在  $f = 1/2$  取最小值。 $a < 0$  时相反。

6. 在  $y_n = v_n - bv_{n-1} = (1 - bB)v_n$  中, 由  $(1 - bB)^{-1} = 1 + bB + b^2B^2 + \dots$  得  $y_n = -by_{n-1} - b^2y_{n-2} - \dots + v_n$ , 即为无穷阶 AR 模型。

7. (1)  $y_{n+1|n} = -bv_n$ ,  $y_{n+k|n} = 0$ , 因此  $E\varepsilon_{n+1|n}^2 = 1$ ,  $E(\varepsilon_{n+k|n}^2) = 1 + b^2$  ( $k \geq 1$ )

(2)  $y_n = v_n + av_{n-1} + a^2v_{n-2} + \dots$ ,  $E(\varepsilon_{n+k|n}^2) = 1 + a^2 + \dots + a^{2(k-1)}$

(3) 由  $y_n = v_n + v_{n-1} + v_{n-2} + \dots$  得  $E(\varepsilon_{n+k|n}^2) = k$ 。

## 第 7 章

1. (1)  $AIC_m = N(\log 2\pi \hat{\sigma}_{n-1}^2 + \log(1 - (a_m^m)^2) + 2(m + 1)) = AIC_{m-1} + N \log(1 - (a_m^m)^2) + 2$ 。因此, 若  $N \log(1 - (a_m^m)^2) + 2 < 0$ , 即  $(a_m^m)^2 > 1 - e^{-2/N}$ , 则判断

AR( $m$ ) 比 AR( $m-1$ ) 更好。

$$(2) \sigma_1^2 = (1 - 0.9^2) \times 1 = 0.19, \sigma_2^2 = 0.1216, \sigma_3^2 = 0.1107, \sigma_4^2 = 0.1095, \sigma_5^2 = 0.1071 \text{ (显示到小数点后 4 位)}$$

$$(3) \text{AIC}_0 = 100(\log 2\pi \times 1) + 2 \times 1 = 285.78$$

$$\text{AIC}_1 = 100(\log 2\pi \times 0.19 + 1) + 2 \times 2 = 121.71$$

$\text{AIC}_2 = 79.09, \text{AIC}_3 = 71.65, \text{AIC}_4 = 72.65, \text{AIC}_5 = 72.37$ 。因此 3 阶为最优。

- Yule-Walker 法的优点是始终保证平稳性, 但特别在数据较少时偏差较大。这是因为相当于假设观测数据区间前后为 0。最小二乘法的优点是可以获得偏差较小的估计值, 且数据的合并等操作比较容易, 但不能保证平稳性, 且最初几个点用于条件化而不能直接用于参数估计。PARCOR 法如果限于 AR 模型的估计, 可以非常高效地获得接近最大似然估计值的估计值。
- 分别对时间序列  $x_n$  和  $y_n$  拟合一元 AR 模型, 其 AIC 值分别记为  $\text{AIC}_x$ 、 $\text{AIC}_y$ 。另一方面, 对  $(x_n, y_n)^T$  拟合二元 AR 模型, 求  $\text{AIC}_0$ 。若  $\text{AIC}_x + \text{AIC}_y < \text{AIC}_0$ , 则视  $x_n$  和  $y_n$  为独立更好。

## 第 8 章

- 考虑  $y_n = \mu + a_1 y_{n-1} + \cdots + a_m y_{n-m} + v_n$  所表示的增加 1 个参数的局部平稳模型即可。
- 将  $y_n$  在  $n = 1, \cdots, k$  时设为  $N(0, \sigma_1^2)$ ,  $n = k+1, \cdots, N$  时设为  $N(0, \sigma_2^2)$  的模型与  $y_n \sim N(0, \sigma_0^2)$ ,  $n = 1, \cdots, N$  进行比较即可。AIC 分别为

$$\begin{aligned} \text{AIC}_{12} &= N \log 2\pi + k \log \hat{\sigma}^2 + (N-k) \log \hat{\sigma}_2^2 + N + 2 \times 2 \\ \text{AIC}_0 &= N \log 2\pi + N \log \hat{\sigma}_0^2 + N + 2 \end{aligned}$$

若  $\text{AIC}_{12} < \text{AIC}_0$ , 则认为方差发生了变化。

- 当  $n = 1, \cdots, k$  时  $y_n = a_0 + a_1 x_n + \cdots + a_p x_n^p + \varepsilon_n$ ,  $n = k+1, \cdots, N$  时  $y_n = b_0 + b_1 x_n + \cdots + a_q x_n^q + \delta_n$ 。其中  $\varepsilon_n \sim N(0, \sigma_1^2)$ ,  $\delta_n \sim N(0, \sigma_2^2)$ 。

$$\text{AIC} = N \log 2\pi + k \log \hat{\sigma}_1^2 + (N-k) \log \hat{\sigma}_2^2 + N + 2 \times (p+q+4)$$

## 第 9 章

- 对 (9.1) 两边左乘  $T$  得  $Tx_n = TF_n x_n + TG_n v_n = TF_n T^{-1} Tx_n + TG_n v_n$ 。这里令  $z_n = Tx_n$ , 利用  $y_n = H_n x_n + w_n = H_n T^{-1} Tx_n + w_n$ , 可以验证得

到等价的表示。

2. (1) 初始值:  $x_{0|0} = 0, V_{0|0} = 100$

$$\text{预测: } x_{n|n-1} = x_{n-1|n-1}, V_{n|n-1} = V_{n-1|n-1} + \tau^2$$

$$\text{滤波: } G_n = V_{n|n-1} / (V_{n|n-1} + 1),$$

$$x_{n|n} = x_{n|n-1} + G_n(y_n - x_{n|n-1}),$$

$$V_{n|n} = (1 - G_n)V_{n|n-1} = G_n$$

(2)  $V_{n+1|n} = V_{n|n-1} / (V_{n|n-1} + 1) + \tau^2$

(3) 在上式中令  $V_{n+1|n} = V_{n|n-1} = V$ 。

(4) 设第(3)题的解为  $V$ , 则滤波的协方差为  $V' = V / (V + 1)$ ,

$$x_{n+1|n} = x_{n|n} = x_{n|n-1} + \frac{V}{V+1}(y_n - x_{n|n-1})$$

3.  $V_P = V, V_F = G = V / (V + 1), A = V / (V + 1) / V = 1 / (V + 1)$ , 由  $V_S = V_F + A^2(V_S - V_P)$  得  $V_S = V / (V + 2)$

$\tau^2$	1	0.1	0.01	0.001
$V_P$	1.1618	0.370	0.105	0.032
$V_F$	0.618	0.270	0.095	0.031
$V_S$	0.447	0.156	0.050	0.016

## 第 10 章

1. 由  $x_n = (y_n, y_{n-1}, \dots, y_{n-m+1})^T$  可知  $E x_n x_n^T$  的  $(i, j)$  分量为  $E y_{n-i+1} y_{n-j+1} = C_{i-j}$ 。因此, 计算  $m$  阶 AR 模型的自协方差函数  $C_k (k = 0, 1, \dots)$ , 按如下方式确定即可。

$$V_{0|0} = \begin{bmatrix} C_0 & C_1 & \cdots & C_{m-1} \\ C_1 & C_0 & \cdots & C_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m-1} & C_{m-2} & \cdots & C_0 \end{bmatrix}$$

2. 在(10.9)式的变换后变量中, 绝对值较大的区域内对数似然的变化很小, 使用数值差分的最优化可能无法起作用。因此, 需要将变换后参数的绝对值限制在例如 20 左右以下。

## 第 11 章

1. (1) 在  $y_{n+k} = y_{n+k-1} + v_{n+k} = \dots = y_n + v_{n+1} + \dots + v_{n+k}$  中, 由  $E v_{n+i} = 0, E v_{n+i}^2 = \sigma^2$  可得  $y_{n+k|n} = y_n, E \varepsilon_{n+k|n}^2 = k \sigma^2$

(2) 由  $k\sigma^2 = 4 \times 40,000 = 160,000$  得  $\sqrt{k}\sigma = 400$ 。正态分布中大于  $1\sigma$  的概率为 0.16。

(3) 即使作为随机游走, 其方差  $\sigma^2$  也会随时间变动。因此, 可以考虑考虑波动率变化的模型

$$y_n = y_{n-1} + \sigma_n w_n, \\ \log \sigma_n^2 = a + b \log \sigma_{n-1}^2 + v_n \quad (\text{或 } \log \sigma_n^2 = \log \sigma_{n-1}^2 + v_n)$$

2. (示例解答) 可以考虑以三角函数模型为趋势的模型。

3. 省略

## 第 12 章

1. AR 模型的特征根接近 1 时, 与趋势分量的区分变得困难。因此, 在 AR 系数的估计中, 最好加以限制使特征根的绝对值不接近 1。

2. 令  $S_n = \sum_{j=1}^6 (a_j \cos \frac{\pi j n}{6} + b_j \sin \frac{\pi j n}{6})$ , 即可得到用三角函数表示的季节分量。

3. (1) 令  $\beta_{n1} = \beta_{n7}$ ,  $\beta_{n2} = \dots = \beta_{n6}$ , 可得到具有两个参数的星期效应分量。此时

$$td_n = \beta_{n1}(d'_{n1} + d'_{n7}) + \beta_{n2}(d'_{n2} + \dots + d'_{n6})。$$

(2) 假设  $\beta_{n2} = \dots = \beta_{n6}$ , 同样可得到具有三个参数的模型。

4. 黄金周等假期因年份不同而休假模式有所变化, 需要考虑其影响。此外, 闰年的影响仅靠星期调整可能无法充分修正。

5. 省略

## 第 13 章

1. (示例解答) 股价数据

2. 在时变系数回归模型  $y_n = \beta_{0n} + \beta_{1n}x_{n1} + \dots + \beta_{mn}x_{nm} + \varepsilon_n$  中, 对系数  $\beta_{0n}, \beta_{1n}, \dots, \beta_{mn}$ , 例如考虑随机游走模型  $\beta_{jn} = \beta_{j,n-1} + v_n$  即可。该模型可以用状态空间模型表示, 因此可以像时变系数 AR 模型一样用卡尔曼滤波进行估计。

## 第 14 章

1. 与正态分布相比, 分布集中在中心的同时尾部较重。因此适合于表达偶尔发生的结构变化和异常值。

2. 当  $y \sim N(0,1)$  时令  $z = y^2$ ,  $x = \log z$ 。此时  $z$  小于  $t$  的概率为

$$F(t) = \text{Prob}(y^2 < t) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{t}} e^{-\frac{y^2}{2}} dy$$

对  $t$  求导得

$$f(t) = \frac{dF}{dt} = \frac{1}{\sqrt{2\pi}} t^{-\frac{1}{2}} e^{-\frac{t}{2}}$$

(自由度 1 的  $\chi^2$  分布)。进一步利用 (4.51) 的逆求对数变换  $x = h(z)$  的分布, 可得如下  $g(x)$ 。

$$g(x) = \left| \frac{dh^{-1}}{dx} \right| f(e^x) = e^x \frac{1}{\sqrt{2\pi}} (e^x)^{-\frac{1}{2}} e^{-\frac{e^x}{2}} = \frac{1}{\sqrt{2\pi}} e^{\frac{x}{2} - \frac{e^x}{2}}.$$

3. 在趋势模型中将系统噪声的分布设为

$$p_v(x) = \begin{cases} \delta(0) & \text{以概率 } 1 - \alpha \\ N(0, \tau^2) & \text{以概率 } \alpha \end{cases}$$

即可。其中  $\delta(0)$  是集中在原点的分布函数。

## 第 15 章

1. 设  $y_n$  为时间序列,  $t_n$  为趋势, 则

$$\begin{aligned} t_n &= t_{n-1} + v_n, \\ \log \sigma_n^2 &= \log \sigma_{n-1}^2 + u_n, \\ y_n &= t_n + \sigma_n w_n, \end{aligned}$$

假设  $u_n \sim N(0, \tau_1^2)$ ,  $v_n \sim N(0, \tau_2^2)$ ,  $w_n \sim N(0, 1)$  的模型。该模型的观测模型是非线性的。

2. 一般来说, 在蒙特卡罗法中, 精度提高一个数量级需要 100 倍的粒子数。

3. 如果完全不进行重采样, 许多粒子的权重将接近 0, 精度变差。

4. 当粒子数为  $m$ , 数据数为  $N$ , 状态维度为  $k$ , 滞后为  $L$  时, 固定区间平滑需要保存  $m \times k \times N$  个, 固定滞后平滑需要保存  $m \times k \times L$  个左右的数据。注意一般  $N \gg L$ 。

## 第 16 章

1. 设  $f(x)$  为均匀分布的密度函数,  $f(x) = 1$  ( $0 \leq x \leq 1$ ),  $= 0$  (其他)。此时

$$g(x) = \int_{-\infty}^{\infty} f(y)f(x-y)dy = \int_0^1 f(x-y)dy = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2 \\ 0 & \text{其他} \end{cases}$$

因此  $W$  的密度函数为三角形。

2. 令  $y_0 = 1$ ,  $v_n \equiv 0$ , 则  $y_n = a^n$ 。因此, 取  $m$  使得  $a^m < c$  ( $c$  例如为  $10^{-3}$ ) 以上即可忽略初始值的影响。当  $a = 0.9, 0.5, 0.1$  时,  $m$  分别为 66、10、5。
3. 利用截至当前的观测值  $y_1, \dots, y_n$ , 构造状态向量  $x_n$ , 用随机数生成  $x_{n+1}, x_{n+2}, \dots$  即可。利用 (9.5) 式的状态进行 AR 模型的仿真时, 令  $x_n = (y_n, y_{n-1}, \dots, y_{n-m+1})^T$  即可。

# 索引

- AIC, 68, 104, 121, 127, 156  
 ARMA 模型, 86, 146  
 ARMA 模型的仿真, 228  
 AR 模型, 103  
 AR 模型, 229  
 autocorrelation function, 33  
  
 BFGS 公式, 238  
 BLSALLFOOD 数据, 19, 48, 172, 175  
 Box-Cox 变换, 73  
 Burg 算法, 109  
  
 DFP 公式, 238  
  
 FFT, 54  
  
 GFSR 法, 223  
 Görtzel 法, 53  
  
 K-L 情报量, 62  
  
 maximum temperature data, 27, 162  
 Mersenne Twister, 223  
  
 PARCOR, 89, 188  
 PARCOR 法, 107  
  
 sunspot number data, 150  
 sunspot number data continued, 152  
  
 twisted GFSR 法, 223  
  
 WHARD 数据, 19, 48, 155, 157, 170, 178  
  
 赤池信息量准则, 68  
 异常值, 21  
 异常值, 205  
 相位谱, 97  
 一变量时系列, 20  
 一样分布, 59  
 均匀随机数, 222  
 1 期先预测, 132, 134, 194  
 移动平均系数, 86  
 移动平均模型, 87  
 脉冲响应函数, 149  
 脉冲响应函数, 87  
 后向 AR 模型, 239  
 熵最大化原理, 62  
 回归模型, 77  
 $\chi^2$  分布, 59, 225  
 高斯型时间序列, 20  
 高斯分布, 59  
 可逆性, 150  
 Kullback-Leibler 信息量, 62  
 卡尔曼滤波, 241  
 观测噪声, 131  
 观测模型, 131  
 气温数据, 164  
 季节分量, 167  
 季节分量模型, 167  
 季节调整模型, 169, 231  
 局部平稳 AR 模型, 120  
 互谱密度函数, 96  
 欠测值, 21, 142  
 结构变化, 123  
 同余法, 222  
 柯西分布, 59, 203, 205, 227  
 固定空间平滑化, 135  
 相干性, 97  
 Cholesky 分解, 229  
 混合分布, 205  
 最高气温数据, 17, 47, 81, 156  
 最小二乘法, 106, 115  
 最小二乘法, 77  
 最大对数似然, 80  
 最大似然估计, 149  
 最尤推定值, 104, 141  
 最大似然法, 66  
 残差, 77  
 散点图, 30, 35  
 时间差分算子, 159  
 时间移位算子, 87  
 时系列, 15  
 自回归移动平均模型, 86  
 自回归系数, 86  
 自回归模型, 86  
 自协方差函数, 32, 88, 149  
 自相关函数, 32  
 地震波数据, 19, 48, 124, 183, 187, 202  
 阶数, 77, 86  
 阶数选择, 80  
 指数分布, 59  
 指数随机数, 225  
 系统噪声, 131  
 系统模型, 131  
 实现值, 60  
 自动划分, 121  
 季度数据, 167  
 时变 AR 系数, 184  
 时变系数 AR 模型, 184  
 时变谱, 188  
 时变方差模型, 181, 201  
 仿真, 222, 229  
 仿真, 232  
 弱平稳, 31  
 频率响应函数, 91, 190  
 拟牛顿法, 237  
 状态, 131  
 状态空间表示, 146  
 状态空间模型, 131, 229  
 初始状态, 149  
 初始值, 152  
 幅度谱, 97  
 数值的实现, 195  
 数值最优化, 67  
 谱, 42  
 谱窗, 52  
 正规分布, 59

- 正态随机数, 224  
 線形時系列, 21  
 船舶数据, 17, 36, 47, 55, 114  
 互协方差函数, 37  
 互相关函数, 37  
 相对功率贡献率, 98  
 对数似然, 65  
 对数尤度, 140  
 对数似然関数, 66  
 太阳黑子数数据, 17, 111, 151  
 多项式回归模型, 155  
 多元 AR 模型, 96, 112, 115  
 多元时间序列, 35  
 多变量時系列, 20  
 多元自回归模型, 96  
 長期予測, 135  
 长期预测, 172  
 月度数据, 167  
 平稳 AR 分量, 172, 174  
 定常時系列, 20  
 平稳序列, 32  
 平稳性, 150  
 $t$  分布, 235  
 数据的追加, 83  
 数据的变换, 72  
 特性方程, 93  
 折中参数, 164  
 趋势, 155  
 趋势分量模型, 159  
 趋势模型, 162  
 二項分布, 206  
 2 重指数分布, 59, 203  
 双指数分布, 181, 234  
 双指数随机数, 225  
 Householder 变换, 79, 126  
 豪斯霍尔德变换, 122  
 Householder 法, 78, 107, 116  
 白噪声, 32, 43  
 参数估计, 139  
 功率谱密度函数, 42  
 功率谱, 90, 92  
 可逆, 93  
 Pearson 分布族, 59, 200  
 非高斯型时间序列, 20  
 非高斯型状态空间模型, 194  
 非高斯型状态空间模型, 232  
 非高斯型趋势模型, 198  
 非高斯型模型, 193  
 直方图, 29  
 非线性最优化, 237  
 非線形時系列, 21  
 非定常時系列, 20  
 非平稳二项过程模型, 206  
 非平稳 Poisson 过程模型, 206  
 非定常離散系列, 206  
 标准季节调整模型, 169  
 样本自协方差函数, 33  
 样本自相关函数, 33  
 样本谱, 46  
 样本互协方差函数, 38  
 样本互相关函数, 38  
 样本平均, 33  
 样本平均向量, 38  
 周期图, 46  
 滤波, 133, 134, 195  
 傅里叶变换, 42, 54  
 子集回归模型, 84  
 分布函数, 227  
 分布関数, 58  
 平滑, 241  
 平滑化, 133, 134  
 平滑化公式, 195  
 平均对数似然, 65  
 平均值函数, 31  
 平均向量, 37  
 变化时点, 126  
 偏自相关系数, 89  
 变量选择, 84  
 Poisson 分布, 206  
 補間, 142  
 白噪声, 32  
 密度関数, 58  
 Jacobian, 73  
 似然, 66  
 尤度, 140  
 Yule-Walker 估计值, 105  
 Yule-Walker 法, 112  
 Yule-Walker 方程, 89, 96, 240  
 星期效应, 175  
 星期调整, 175  
 予測, 133, 136  
 预测误差, 239  
 滞后, 32  
 随机数, 222  
 随机游走模型, 160, 234  
 離散時間時系列, 20  
 Levinson 算法, 104, 239