

时间序列分析及应用

4.6 AR 模型的预测

刘洋

四川师范大学数学科学学院

2022 年 4 月 2 日



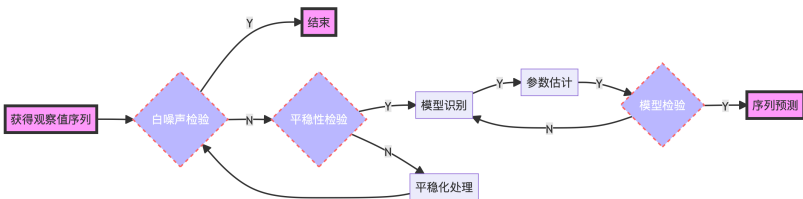


图 1: 时间序列建模流程图

- ① 4.6.1 最佳预测
- ② 4.6.2 最佳线性预测
- ③ 4.6.3 AR 模型的预测
- ④ 4.6.4 修正预测
- ⑤ 参考资料

给定时间序列 X_1, \dots, X_t , 如何预测下一时刻序列值 X_{t+1} 或 k 步以后的序列值 X_{t+k} 是时间序列分析的主要目的之一.

人们希望预测得越“准确”越好. 这里的准确性是在某种准则意义下的, 常见的准则有均方误差准则. 由于平稳序列可以中心化, 故本节仅考虑零均值且方差有限的平稳序列.

本节讨论平稳序列的最佳预测和最佳线性预测, 并将其应用于 AR 模型的预测问题.

- ① 4.6.1 最佳预测
- ② 4.6.2 最佳线性预测
- ③ 4.6.3 AR 模型的预测
- ④ 4.6.4 修正预测
- ⑤ 参考资料

最佳预测

设 $\hat{X}_t(k)$ 是由时间序列 X_1, \dots, X_t 对于 X_{t+k} 的预测值. 为了衡量预测值 $\hat{X}_t(k)$ 的精确程度, 常用的衡量准则为二次损失函数

$$E \left(X_{t+k} - \hat{X}_t(k) \right)^2. \quad (4.42)$$

因此, 我们希望预测值 $\hat{X}_t(k)$ 能使得二次损失函数 (4.42) 达到最小. 若能找到某一 $\hat{X}_t(k)$ 使得 (4.42) 达到最小, 则称 $\hat{X}_t(k)$ 为 X_{t+k} 的最佳预测. 记 $\mathbf{X}_t = (X_1, \dots, X_t)^\top$, 我们有下面的结论.

最佳预测

定理 (4.4)

给定时间序列 X_1, \dots, X_t , 则使得二次损失函数 (4.42) 达到最小的预测值为条件期望

$$\hat{X}_t(k) = E(X_{t+k} | \mathbf{X}_t)$$

最佳预测

证 对于某一预测函数 $f(\mathbf{X}_t)$, 其二次损失函数

$$\begin{aligned}
 & E(X_{t+k} - f(\mathbf{X}_t))^2 \\
 &= E(X_{t+k} - E(X_{t+k} | \mathbf{X}_t) + E(X_{t+k} | \mathbf{X}_t) - f(\mathbf{X}_t))^2 \\
 &= E(X_{t+k} - E(X_{t+k} | \mathbf{X}_t))^2 + 2E(\xi_{t+k}) \\
 &\quad + E(E(X_{t+k} | \mathbf{X}_t) - f(\mathbf{X}_t))^2,
 \end{aligned} \tag{4.43}$$

式中 $\xi_{t+k} = (X_{t+k} - E(X_{t+k} | \mathbf{X}_t))(E(X_{t+k} | \mathbf{X}_t) - f(\mathbf{X}_t))$. 易知, $E(X_{t+k} | \mathbf{X}_t)$ 和 $f(\mathbf{X}_t)$ 是 \mathbf{X}_t 的函数.

最佳预测

因此, 给定数据 $\mathbf{X}_t = (X_1, \dots, X_t)^\top$, 我们有

$$\begin{aligned}
 & E(\xi_{t+k} \mid \mathbf{X}_t) \\
 &= (E(X_{t+k} \mid \mathbf{X}_t) - f(\mathbf{X}_t)) \times E(X_{t+k} - E(X_{t+k} \mid \mathbf{X}_t)) \\
 &= (E(X_{t+k} \mid \mathbf{X}_t) - f(\mathbf{X}_t)) \times (EX_{t+k} - E(E(X_{t+k} \mid \mathbf{X}_t))) \\
 &= (E(X_{t+k} \mid \mathbf{X}_t) - f(\mathbf{X}_t)) \times 0 \\
 &= 0
 \end{aligned}
 \tag{4.44}$$

上面倒数第二个等式成立的原因是条件期望有一个特殊的性质, 即条件期望的期望等于无条件期望.

最佳预测

从而, 由 (4.44) 式可得

$$E(\xi_{t+k}) = E(E(\xi_{t+k} | \mathbf{X}_t)) = 0 \quad (4.45)$$

把 (4.45) 式带入 (4.43) 可得,

$$\begin{aligned} & E(X_{t+k} - f(\mathbf{X}_t))^2 \\ &= E(X_{t+k} - E(X_{t+k} | \mathbf{X}_t))^2 + E(E(X_{t+k} | \mathbf{X}_t) - f(\mathbf{X}_t))^2 \end{aligned} \quad (4.46)$$

显然, (4.46) 式中等式右边的第一项与预测函数 $f(\mathbf{X}_t)$ 无关, 而第二项是非负的.

最佳预测

欲使 (4.46) 式达到最小, 只需让预测函数

$$f(\mathbf{X}_t) = E(X_{t+k} | \mathbf{X}_t)$$

即可. 从而完成了定理 4.4 的证明.

由定理 4.4 可知, 给定数据 \mathbf{X}_t , 其 k 步最佳预测为其条件期望. 然而, 在实际应用中, 条件期望的具体表达式往往未知, 故难以具体实现. 因此, 人们往往考虑简单的线性预测模型.

最佳线性预测

给定某一中心化的平稳序列 $\mathbf{X}_t = (X_1, \dots, X_t)^T$ ，考虑用 X_1, \dots, X_t 的某一线性组合

$$\mathbf{a}^T \mathbf{X}_t = \sum_{i=1}^t a_i X_i$$

来预测 k 步之后的序列值 X_{t+k} ，其中 $\mathbf{a} = (a_1, \dots, a_t)^T$ 。人们希望寻找某一向量 \mathbf{a} 使得 $\mathbf{a}^T \mathbf{X}_t$ 尽可能逼近 X_{t+k} 。

一般地，其逼近程度用二次损失函数 (4.42) 来衡量。

最佳线性预测

因此, 人们寻找某一向量 a 使得

$$E (X_{t+k} - \mathbf{a}^T \mathbf{X}_t)^2 \quad (4.47)$$

达到最小. 设 \mathbf{a}^* 使得 (4.47) 式达到最小, 我们称 $(\mathbf{a}^*)^T \mathbf{X}_t$ 为 X_{t+k} 的最佳线性预测, 并记为 $\hat{X}_t(k)$ 或 $L(X_{t+k} | \mathbf{X}_t)$. 由于 $X_{t+k} - \mathbf{a}^T \mathbf{X}_t$ 为预测误差, 则最佳线性预测是所有线性预测中, 使得均方误差达到最小的预测.

最佳线性预测

比较 (4.42) 式和 (4.47) 式可知, 最佳线性预测的预测效果往往比定理 4.4 中给出的最佳预测的预测效果差, 因为最佳线性预测只是在线性类里面寻找最佳的向量, 而最佳预测在均方误差意义下是最好的预测.

最佳线性预测

对于均值 μ 不为零的平稳序列 Y_1, \dots, Y_t , 其中 $Y_i = X_i + \mu$, 则其 k 步最佳线性预测可表示为

$$\hat{Y}_t(k) = L(X_{t+k} | \mathbf{X}_t) + \mu,$$

即先对序列 Y_1, \dots, Y_t 都减去均值 μ 得到 X_1, \dots, X_t , 再由其最佳线性预测加上均值 μ 即可. 显然, $E\hat{Y}_t(k) = \mu = E(Y_{t+k})$.

最佳线性预测的求解

对于任意 t 维向量 \mathbf{b} , 我们有

$$\begin{aligned}
 & E (X_{t+k} - \mathbf{b}^T \mathbf{X}_t)^2 \\
 &= E (X_{t+k} - \mathbf{a}^T \mathbf{X}_t + \mathbf{a}^T \mathbf{X}_t - \mathbf{b}^T \mathbf{X}_t)^2 \\
 &= E (X_{t+k} - \mathbf{a}^T \mathbf{X}_t)^2 + E ((\mathbf{a} - \mathbf{b})^T \mathbf{X}_t)^2 + \\
 & \quad 2(\mathbf{a} - \mathbf{b})^T (E (\mathbf{X}_t X_{t+k}) - E (\mathbf{X}_t \mathbf{X}_t^T) \mathbf{a})
 \end{aligned} \tag{4.48}$$

最佳线性预测的求解

显然, 假如如下的关系成立

$$E(\mathbf{X}_t X_{t+k}) - E(\mathbf{X}_t \mathbf{X}_t^T) \mathbf{a} = \mathbf{0}_t \quad (4.49)$$

其中 $\mathbf{0}_t$ 为 t 维零向量, 则由 (4.48) 式易知,

$$E(X_{t+k} - \mathbf{b}^T \mathbf{X}_t)^2 \geq E(X_{t+k} - \mathbf{a}^T \mathbf{X}_t)^2$$

最佳线性预测的求解

此时, $\mathbf{a}^T \mathbf{X}_t$ 为最佳线性预测, 且预测均方误差

$$\begin{aligned} E(X_{t+k} - \mathbf{a}^T \mathbf{X}_t)^2 &= EX_{t+k}^2 + \mathbf{a}^T E(\mathbf{X}_t \mathbf{X}_t^T) \mathbf{a} - 2\mathbf{a}^T E(\mathbf{X}_t X_{t+k}) \\ &= EX_{t+k}^2 + \mathbf{a}^T E(\mathbf{X}_t \mathbf{X}_t^T) \mathbf{a} - 2\mathbf{a}^T E(\mathbf{X}_t \mathbf{X}_t^T) \mathbf{a} \\ &= EX_{t+k}^2 - \mathbf{a}^T E(\mathbf{X}_t \mathbf{X}_t^T) \mathbf{a} \end{aligned}$$

由上面的讨论可知, 假如向量 \mathbf{a} 满足 (4.49) 式, 则相应的线性组合 $\mathbf{a}^T \mathbf{X}_t$ 即为 X_{t+k} 的最佳线性预测.

因此, 我们称 (4.49) 式为预测方程, 即可以通过解方程 (4.49) 来获得最佳线性预测的向量 \mathbf{a} .

最佳线性预测的求解

记 $\mathbf{\Gamma} = E(\mathbf{X}_t \mathbf{X}_t^T)$, $\boldsymbol{\eta}_k = E(\mathbf{X}_t \mathbf{X}_{t+k})$, 则预测方程 (4.49) 即为一个非齐次线性方程组

$$\mathbf{\Gamma} \mathbf{a} = \boldsymbol{\eta}_k, \quad (4.50)$$

其中 $\mathbf{\Gamma} \in \mathbb{R}^{t \times t}$, $\boldsymbol{\eta}_k \in \mathbb{R}^t$. 易知 $\mathbf{\Gamma}$ 为半正定矩阵, 因为对于任意的 t 维向量 \mathbf{c} , $\mathbf{c}^T \mathbf{\Gamma} \mathbf{c} = E(\mathbf{c}^T \mathbf{X}_t)^2 \geq 0$. 假设矩阵 $\mathbf{\Gamma}$ 可逆, 则由预测方程 (4.50) 式, 易知解向量为

$$\hat{\mathbf{a}} = \mathbf{\Gamma}^{-1} \boldsymbol{\eta}_k \quad (4.51)$$

假如 $\mathbf{\Gamma}$ 不可逆, 即其行列式 $\det(\mathbf{\Gamma}) = 0$, 则由线性代数的知识可知预测方程 (4.50) 的解向量仍然存在. 下面给出其一个解向量.

最佳线性预测的求解

对于半正定矩阵 Γ , 总存在一个正交矩阵 Q 使 Γ 对角化, 即

$$Q\Gamma Q^T = E(QX_t X_t^T Q^T) = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$$

其中 $\lambda_1 \geq \dots \geq \lambda_r > 0$ 为 Γ 的 r 个大于零的特征值. 记

$Y = QX_t = (Y_1, \dots, Y_t)^T$, 并设 ξ 为 Y 的前 r 个随机变量, 即 $\xi = (Y_1, \dots, Y_r)^T$, $\mathbf{0}_{t-r}$ 为 $t-r$ 维零向量, 则

$$E(\xi\xi^T) = \text{diag}(\lambda_1, \dots, \lambda_r), Y = \begin{bmatrix} \xi \\ \mathbf{0}_{t-r} \end{bmatrix}, \text{ a.s.}$$

最佳线性预测的求解

若记 $\zeta = [E(\xi\xi^T)]^{-1} E(\xi X_{t+k})$, 则可以验证

$$\hat{\mathbf{a}} = \mathbf{Q}^T \begin{bmatrix} \zeta \\ \mathbf{0}_{t-r} \end{bmatrix} \quad (4.52)$$

满足预测方程 (4.50),

最佳线性预测的求解

因为

$$\begin{aligned}
 \mathbf{Q}\Gamma\hat{\mathbf{a}} &= \mathbf{Q}\Gamma\mathbf{Q}^T \begin{bmatrix} \zeta \\ \mathbf{0}_{t-r} \end{bmatrix} \\
 &= \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) \begin{bmatrix} \zeta \\ \mathbf{0}_{t-r} \end{bmatrix} \\
 &= \begin{bmatrix} E(\boldsymbol{\xi}X_{t+k}) \\ \mathbf{0}_{t-r} \end{bmatrix} = E \left[\begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{0}_{t-r} \end{pmatrix} X_{t+k} \right] \\
 &= E[\mathbf{Y}X_{t+k}] = E[\mathbf{Q}\mathbf{X}_tX_{t+k}] \\
 &= \mathbf{Q}\eta_k
 \end{aligned}$$

上面等式分别左乘矩阵 \mathbf{Q}^T ，即得 (4.50) 式。因此 (4.52) 式是预测方程 (4.50) 式的一个解。

最佳线性预测的求解

当 Γ 不可逆时, 根据矩阵论的知识, 解向量不是唯一的. 设 $\hat{\mathbf{a}}$ 如 (4.52) 所示, 则 $\hat{\mathbf{a}}^T \mathbf{X}_t$ 为 X_{t+k} 的最佳线性预测. 对于任意其他向量 $\mathbf{b} \in \mathbb{R}^T$, 根据 (4.48) 式, 我们有

$$E(X_{t+k} - \mathbf{b}^T \mathbf{X}_t)^2 = E(X_{t+k} - \hat{\mathbf{a}}^T \mathbf{X}_t)^2 + E((\hat{\mathbf{a}} - \mathbf{b})^T \mathbf{X}_t)^2$$

因此, 若 $\mathbf{b}^T \mathbf{X}_t$ 也是最佳线性预测, 则有 $E((\hat{\mathbf{a}} - \mathbf{b})^T \mathbf{X}_t)^2 = 0$, 即

$$\hat{\mathbf{a}}^T \mathbf{X}_t = \mathbf{b}^T \mathbf{X}_t, \quad \text{a.s.}$$

换句话说, 最佳线性预测是几乎处处唯一的.

最佳线性预测的求解

综上所述, 我们有如下的结论. 定理 4.5 对于预测方程 (4.49) 式或 (4.50) 式, 当 Γ 可逆时, 最佳线性预测的解向量 $\hat{\mathbf{a}}$ 如 (4.51) 式所示, 当 Γ 不可逆时, 解向量 $\hat{\mathbf{a}}$ 可采用 (4.52) 式.

需要指出的是, 定理 4.5 中讨论的是 $k \geq 1$ 的情形. 当 $k \leq 0$ 时, 即需预测的时刻只是当前或历史时刻, 则易证其最佳预测和最佳线性预测都是其本身, 即

$$L(X_{t+k} | \mathbf{X}_t) = X_{t+k}, \text{ 当 } k \leq 0.$$

最佳线性预测的求解

由 (4.49) 式可知,

$$E\left(\left(X_{t+k} - \mathbf{a}^T \mathbf{X}_t\right) \mathbf{X}_t^T\right) = \mathbf{0}_t^T, \quad (4.53)$$

即预测误差 $X_{t+k} - \mathbf{a}^T \mathbf{X}_t$ 和 \mathbf{X}_t 线性无关. 我们称预测 $\mathbf{a}^T \mathbf{X}_t$ 为 X_{t+k} 关于 \mathbf{X}_t 的线性投影. 因此, 在线性预测类中, 线性投影是均方误差最小的线性预测.

对于任意的 t 维列向量 \mathbf{b} ,

$$E\left(\left(X_{t+k} - \mathbf{a}^T \mathbf{X}_t\right) \left(\mathbf{b}^T \mathbf{X}_t\right)\right) = 0,$$

因此预测误差 $X_{t+k} - \mathbf{a}^T \mathbf{X}_t$ 垂直于由 \mathbf{X}_t 张成的线性空间 $H = \{\mathbf{b}^T \mathbf{X}_t, \mathbf{b} \in \mathbb{R}^t\}$. 因此, $\mathbf{a}^T \mathbf{X}_t$ 是 X_{t+k} 在空间 H 上的投影.

最佳线性预测的求解

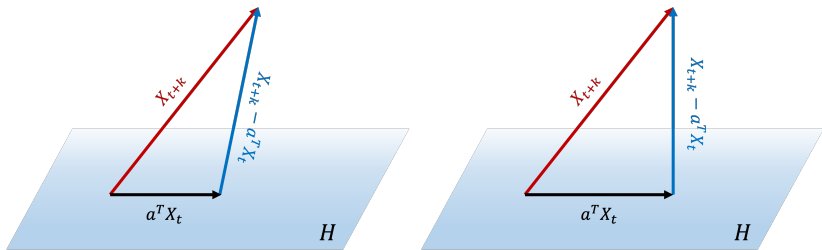


图 2: 投影演示

预测误差 $X_{t+k} - \mathbf{a}^T \mathbf{X}_t$ 垂直于由 \mathbf{X}_t 张成的线性空间 $H = \{\mathbf{b}^T \mathbf{X}_t, \mathbf{b} \in \mathbb{R}^t\}$. 因此, $\mathbf{a}^T \mathbf{X}_t$ 是 X_{t+k} 在空间 H 上的投影.

最佳线性预测与最小二乘回归的关系

线性预测是由已有的序列 \mathbf{X}_t 的某个线性组合对将来的序列值进行预测, 因此 X_{t+k} 可表示为

$$X_{t+k} = \mathbf{a}^T \mathbf{X}_t + \varepsilon_{t+k},$$

其中没有常数项的原因是该序列的均值为零. 上式与线性回归的形式非常相似.

最佳线性预测与最小二乘回归的关系

假设观测值序列 $\mathbf{x}_t = (x_{t-n+1}, x_{t-n+2}, \dots, x_t)^\top$, 给定 T 个数据对 $\{x_{t+k}, \mathbf{x}_t\}$, $t = 1, \dots, T$, 则 x_{t+k} 的线性预测为

$$x_{t+k} = b^T \mathbf{x}_t + \varepsilon_{t+k}.$$

最佳线性预测与最小二乘回归的关系

则使残差平方和

$$\sum_{t=1}^T (x_{t+k} - \mathbf{b}^T \mathbf{x}_t)^2$$

达到最小的参数 \mathbf{b} 的最小二乘解为

$$\begin{aligned} \mathbf{b} &= \left[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^T \right]^{-1} \left[\sum_{t=1}^T \mathbf{x}_t x_{t+k} \right] \\ &= \left[\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^T \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t x_{t+k} \right] \end{aligned} \quad (4.54)$$

最佳线性预测与最小二乘回归的关系

比较 (4.54) 式和最佳线性预测的参数估计 (4.51) 式可知, (4.54) 式中的参数 \mathbf{b} 是由样本矩 $\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^T$ 和 $\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t x_{t+k}$ 估计的, 而 (4.51) 式中的参数 \mathbf{a} 是由总体矩 $E(\mathbf{X}_t \mathbf{X}_t^T)$ 和 $E(\mathbf{X}_t X_{t+k})$ 估计的.

因此, 线性回归是对观测序列的处理, 而最佳线性预测是对于总体信息的处理. 不过对于平稳序列 \mathbf{X}_t , 当该序列具有各态历经 (遍历) 性时, 可以证明, 随着数据量 T 的增大, 样本矩将依概率收敛于总体矩. 因此, \mathbf{b} 也依概率收敛于 \mathbf{a} . 所以, 线性回归得到的系数是最佳线性预测的参数的一致估计.

最佳线性预测和最佳预测相等的情形

一般而言，在均方误差意义下，最佳预测比最佳线性预测效果好。然而，在某些情形下，这两者是相等的。下面给出最佳预测和最佳线性预测相等的两种特殊情形。需要指出的是，这里的相等都是在几乎处处 (a.s.) 意义下相等的。

定理 (4.6)

对于零均值的时间序列 X_t ，若 $(X_1, \dots, X_t, X_{t+k})^T$ 服从多元正态分布 $N(\mathbf{0}, \Sigma)$ ，则最佳预测和最佳线性预测相等，即

$$E(X_{t+k} | \mathbf{X}_t) = L(X_{t+k} | \mathbf{X}_t)$$

最佳线性预测和最佳预测相等的情形

证设 $\hat{X}_t(k) = L(X_{t+k} | \mathbf{X}_t)$, 则根据 (4.53) 式可知, 预测误差 $X_{t+k} - \hat{X}_t(k)$ 和 \mathbf{X}_t 线性无关, 根据多元正态分布的性质可知, $X_{t+k} - \hat{X}_t(k)$ 和 \mathbf{X}_t 相互独立. 从而 $X_{t+k} - \hat{X}_t(k)$ 与任意可测函数 $g(\cdot)$ 作用于 \mathbf{X}_t 而得到的随机变量 $g(\mathbf{X}_t)$ 都相互独立.

最佳线性预测和最佳预测相等的情形

从而对于任意的预测函数 $g(\mathbf{X}_t)$, 我们有

$$\begin{aligned}
 E(X_{t+k} - g(\mathbf{X}_t))^2 &= E\left(X_{t+k} - \hat{X}_t(k) + \hat{X}_t(k) - g(\mathbf{X}_t)\right)^2 \\
 &= E\left(X_{t+k} - \hat{X}_t(k)\right)^2 + E\left(\hat{X}_t(k) - g(\mathbf{X}_t)\right)^2 + \\
 &\quad 2E\left(\left(X_{t+k} - \hat{X}_t(k)\right)\left(\hat{X}_t(k) - g(\mathbf{X}_t)\right)\right) \\
 &= E\left(X_{t+k} - \hat{X}_t(k)\right)^2 + E\left(\hat{X}_t(k) - g(\mathbf{X}_t)\right)^2 \\
 &\geq E\left(X_{t+k} - \hat{X}_t(k)\right)^2
 \end{aligned}
 \tag{4.55}$$

最佳线性预测和最佳预测相等的情形

上面倒数第一个等式成立的原因是在于, $\hat{X}_t(k)$ 和 $g(\mathbf{X}_t)$ 都是由 \mathbf{X}_t 变化而来的, 从而 $X_{t+k} - \hat{X}_t(k)$ 与 $\hat{X}_t(k) - g(\mathbf{X}_t)$ 相互独立, 故根据序列的零均值可知,

$$E \left[\left(X_{t+k} - \hat{X}_t(k) \right) \left(\hat{X}_t(k) - g(\mathbf{X}_t) \right) \right] = 0.$$

则由最佳预测的定义及 (4.55) 式可知, $\hat{X}_t(k)$ 即为最佳预测.

定理 4.6 说明当时间序列为零均值的正态过程时, 最佳预测与最佳线性预测是相等的. 需要注意的是, 该结论并没有要求序列必须满足平稳性. 进一步的, 若取消零均值的限制时, 该结论仍然成立. 具体证明过程留给读者.

最佳线性预测和最佳预测相等的情形

定理 (4.7)

若平稳序列 X_t 可以表示为

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad (4.56)$$

其中 ε_t 为白噪声序列, 则最佳预测和最佳线性预测相等的充要条件为

$$E(\varepsilon_{t+1} \mid \varepsilon_t, \varepsilon_{t-1}, \dots) = 0 \quad (4.57)$$

最佳线性预测和最佳预测相等的情形

定理 4.7 的证明过程可参考何书元 (2003). 特别地, 当 $\{\varepsilon_t\}$ 为高斯白噪声时, (4.57) 式显然成立. 由于 $AR(p)$ 序列 X_t 可以表示为 (4.56) 式, 因此, 由定理 4.7 可知, 当 ε_t 为独立白噪声时, 最佳预测与最佳线性预测是相等的.

特别地, ε_t 为高斯白噪声时, 任意两个时刻的噪声都相互独立, 故 (4.57) 式成立, 此时最佳预测与最佳线性预测是相等的.

- ① 4.6.1 最佳预测
- ② 4.6.2 最佳线性预测
- ③ 4.6.3 AR 模型的预测**
- ④ 4.6.4 修正预测
- ⑤ 参考资料

AR 模型的预测

假设平稳序列 X_t 服从 $AR(p)$ 模型

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t$$

其中 $\{\varepsilon_t\}$ 为零均值白噪声. 考虑用 $\mathbf{X}_t = (X_1, \dots, X_t)^T$ 预测 k 步之后的序列值 X_{t+k} .

AR(p) 序列的最佳预测

考虑 $\{\varepsilon_t\}$ 为零均值高斯白噪声情形下的 AR(p) 序列 X_t 的最佳预测. 根据定理 4.4, X_{t+k} 的最佳预测 $\hat{X}_t(k)$ 为条件期望 $E(X_{t+k} | \mathbf{X}_t)$. 当 $k=1$ 时,

$$\begin{aligned}\hat{X}_t(1) &= E(X_{t+1} | \mathbf{X}_t) \\ &= E((a_1 X_t + a_2 X_{t-1} + \cdots + a_p X_{t+1-p} + \varepsilon_{t+1}) | \mathbf{X}_t) \\ &= E((a_1 X_t + a_2 X_{t-1} + \cdots + a_p X_{t+1-p}) | \mathbf{X}_t) + E(\varepsilon_{t+1} | \mathbf{X}_t) \\ &= \sum_{j=1}^p a_j X_{t+1-j},\end{aligned}$$

AR(p) 序列的最佳预测

上面最后一个等式成立的原因如下:

已知序列 $\mathbf{X}_t = (X_1, \dots, X_t)^T$, 则条件期望

$E(X_{t+1-j} | \mathbf{X}_t) = X_{t+1-j}, j = 1, \dots, p$; 且 ε_{t+1} 为将来时刻的噪声, 与已有数据 \mathbf{X}_t 独立, 故 $E(\varepsilon_{t+1} | \mathbf{X}_t) = E(\varepsilon_{t+1}) = 0$.

AR(p) 序列的最佳预测

当 $k = 2$ 时, 最佳预测 $\hat{X}_t(2)$ 为条件期望 $E(X_{t+2} | \mathbf{X}_t)$, 因此,

$$\begin{aligned}
 \hat{X}_t(2) &= E(X_{t+2} | \mathbf{X}_t) \\
 &= E((a_1 X_{t+1} + a_2 X_t + \cdots + a_p X_{t+2-p} + \varepsilon_{t+2}) | \mathbf{X}_t) \\
 &= a_1 E(X_{t+1} | \mathbf{X}_t) + E((a_2 X_t + a_3 X_{t-1} + \cdots + a_p X_{t+2-p}) | \mathbf{X}_t) \\
 &\quad + E(\varepsilon_{t+2} | \mathbf{X}_t) \\
 &= a_1 \hat{X}_t(1) + a_2 X_t + a_3 X_{t-1} + \cdots + a_p X_{t+2-p}
 \end{aligned}$$

AR(p) 序列的最佳预测

当 $k = 3$ 时, 最佳预测 $\hat{X}_t(3)$ 可表示为,

$$\begin{aligned}
 \hat{X}_t(3) &= E(X_{t+3} | \mathbf{X}_t) \\
 &= E((a_1 X_{t+2} + a_2 X_{t+1} + a_3 X_t + \cdots + a_p X_{t+3-p} + \varepsilon_{t+3}) | \mathbf{X}_t) \\
 &= a_1 E(X_{t+2} | \mathbf{X}_t) + a_2 E(X_{t+1} | \mathbf{X}_t) \\
 &\quad + E((a_3 X_t + a_4 X_{t-1} + \cdots + a_p X_{t+3-p}) | \mathbf{X}_t) + E(\varepsilon_{t+3} | \mathbf{X}_t) \\
 &= a_1 \hat{X}_t(2) + a_2 \hat{X}_t(1) + a_3 X_t + a_4 X_{t-1} + \cdots + a_p X_{t+3-p}
 \end{aligned}$$

AR(p) 序列的最佳预测

由数学归纳法易知, 当 $k > 1$ 时, $\hat{X}_t(k)$ 的最佳预测为

$$\hat{X}_t(k) = \begin{cases} \sum_{j=1}^{k-1} a_j \hat{X}_t(k-j) + \sum_{j=k}^p a_j X_{t+k-j}, & k \leq p \\ \sum_{j=1}^p a_j \hat{X}_t(k-j), & k > p \end{cases}$$

AR(p) 序列的最佳预测

综上所述, 当 $\{\varepsilon_t\}$ 为零均值高斯白噪声时, AR(p) 序列 X_t 的 k 步最佳预测 $\hat{X}_t(k)$ 可表示为

$$\hat{X}_t(k) = \begin{cases} \sum_{j=1}^p a_j X_{t+1-j}, & k = 1 \\ \sum_{j=1}^{k-1} a_j \hat{X}_t(k-j) + \sum_{j=k}^p a_j X_{t+k-j}, & 1 < k \leq p \\ \sum_{j=1}^p a_j \hat{X}_t(k-j), & k > p \end{cases} \quad (4.58)$$

从 (4.58) 式可知, 若已知 AR(p) 序列的系数 a_1, \dots, a_p , 我们只用数据 $\{X_t, X_{t-1}, \dots, X_{t+1-p}\}$ 即可预测 X_{t+k} , 而不需要其他数据. 因此, 该预测式子非常简单, 这也是实践中常用 AR 模型的原因之一. 然而, 当 ε_t 不是独立序列时, (4.58) 式往往不再成立, 即此时最佳预测难以有显式表达式.

AR(p) 序列的最佳线性预测

由前一小节的讨论可知, 当 $\{\varepsilon_t\}$ 为高斯白噪声时, 最佳线性预测与最佳预测是一致的. 然而, 当 $\{\varepsilon_t\}$ 不是独立白噪声时, 最佳线性预测与最佳预测一般不相等. 然而当历史数据足够多时, 最佳线性预测往往近似于最佳预测. 并且由于最佳线性预测有显式表达式, 易于操作, 人们往往考虑 AR(p) 序列的最佳线性预测.

AR(p) 序列的最佳线性预测

假设 $\{\varepsilon_t\}$ 为白噪声. 当 $1 \leq t \leq p-1$ 时, 自协方差矩阵 $\mathbf{\Gamma} = E(\mathbf{X}_t \mathbf{X}_t^T)$ 往往可逆. 此时, 根据 (4.51) 式, X_{t+k} 的最佳线性预测的预测向量 $\mathbf{a} = \mathbf{\Gamma}^{-1} \boldsymbol{\eta}_k$, 其中 $\boldsymbol{\eta}_k = E(\mathbf{X}_t X_{t+k})$. 则最佳线性预测为

$$\hat{X}_t(k) = L(X_{t+k} | \mathbf{X}_t) = \boldsymbol{\eta}_k^T \mathbf{\Gamma}^{-1} \mathbf{X}_t$$

预测的均方误差为

$$E\left(X_{t+k} - \hat{X}_t(k)\right)^2 = EX_{t+k}^2 - \mathbf{a}^T E(\mathbf{X}_t \mathbf{X}_t^T) \mathbf{a} = \gamma_0 - \boldsymbol{\eta}_k^T \mathbf{\Gamma}^{-1} \boldsymbol{\eta}_k \quad (4.59)$$

AR(p) 序列的最佳线性预测

由于 AR(p) 序列 X_t 可表示为

$$X_t = \sum_{j=0}^{\infty} g_j \varepsilon_{t-j}$$

则根据白噪声的性质可知, X_t 与 ε_{t+k} 线性无关. 从而当 $t \geq p$ 时,

$$\begin{aligned} \hat{X}_t(k) &= L \left(\sum_{j=1}^p a_j X_{t+k-j} + \varepsilon_{t+k} \mid \mathbf{X}_t \right) = L \left(\sum_{j=1}^p a_j X_{t+k-j} \mid \mathbf{X}_t \right) \\ &= \sum_{j=1}^p a_j L (X_{t+k-j} \mid \mathbf{X}_t). \end{aligned}$$

AR(p) 序列的最佳线性预测

当 $j \geq k$ 时, $L(X_{t+k-j} | \mathbf{X}_t) = X_{t+k-j}$. 因此, X_{t+k} 的最佳线性预测可表示为

$$\hat{X}_t(k) = L(X_{t+k} | \mathbf{X}_t) = \begin{cases} \sum_{j=1}^p a_j X_{t+1-j}, & k = 1, \\ \sum_{j=1}^{k-1} a_j \hat{X}_t(k-j) + \sum_{j=k}^p a_j X_{t+k-j}, & 1 < k \leq p, \\ \sum_{j=1}^p a_j \hat{X}_t(k-j), & k > p \end{cases} \quad (4.60)$$

上式和 (4.58) 式非常类似, (4.60) 式只不过把 (4.58) 式中最佳预测替换为最佳线性预测而已. 然而这两个式子的条件是有所不同的. (4.58) 式中要求噪声为高斯白噪声, 而 (4.60) 式中仅要求噪声为白噪声即可.

AR(p) 序列的最佳线性预测

例 (4.12)

AR(1) 模型的预测. 设

$$X_t = a_1 X_{t-1} + \varepsilon_t.$$

其中 ε_t 为白噪声.

记 $\mathbf{X}_t = (X_1, \dots, X_t)^\top$, 则根据 (4.60) 式, X_{t+k} 的最佳线性预测为

$$\hat{X}_t(k) = L(X_{t+k} | \mathbf{X}_t) = a_1 \hat{X}_t(k-1) = a_1^2 \hat{X}_t(k-2) = \dots = a_1^k X_t \quad (4.61)$$

则预测均方误差为 $E(X_{t+k} - \hat{X}_t(k))^2 = E(X_{t+k} - a_1^k X_t)^2$.

AR(p) 序列的最佳线性预测

根据 AR(1) 序列的平稳性, 可知 $|a_1| < 1$. 则当 $k \rightarrow \infty$ 时, $a_1^k \rightarrow 0$. 故预测均方误差

$$E \left(X_{t+k} - \hat{X}_t(k) \right)^2 \rightarrow EX_{t+k}^2 = \gamma_0, \quad k \rightarrow \infty$$

上式说明, 当预测的步长越大, 预测的精度越低. 当 $\{\varepsilon_t\}$ 为高斯白噪声时, (4.61) 式也是 AR(1) 模型的最佳预测.

AR(p) 序列的最佳线性预测

例 (4.13)

AR(2) 模型的预测. 设

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t$$

其中 $\{\varepsilon_t\}$ 为白噪声.

AR(p) 序列的最佳线性预测

根据 (4.60) 式, X_{t+k} 的最佳线性预测为

$$\hat{X}_t(1) = L(X_{t+1} | \mathbf{X}_t) = a_1 X_t + a_2 X_{t-1}$$

$$\hat{X}_t(2) = L(X_{t+2} | \mathbf{X}_t) = a_1 \hat{X}_t(1) + a_2 X_t$$

$$\hat{X}_t(3) = L(X_{t+3} | \mathbf{X}_t) = a_1 \hat{X}_t(2) + a_2 \hat{X}_t(1)$$

...

$$\hat{X}_t(k) = L(X_{t+k} | \mathbf{X}_t) = a_1 \hat{X}_t(k-1) + a_2 \hat{X}_t(k-2)$$

即 AR(2) 模型的预测可以采用递推的形式得到. 从中可见, 最佳线性预测 $\hat{X}_t(k)$ 最终可以由 X_t 和 X_{t-1} 的线性组合表出.

基于无限数据的 $AR(p)$ 序列预测

前面讨论的是有限历史数据的情形. 若历史数据为无穷多, 即已知数据 $\{X_t, X_{t-1}, \dots\}$, t 为当前时刻, 则 $AR(p)$ 序列的最佳预测有另外一种表达形式, 而且最佳预测的预测误差有简单表达式. 此时, 根据定理 4.7 可知, $AR(p)$ 序列的最佳预测与最佳线性预测相等.

基于无限数据的 $AR(p)$ 序列预测

根据 Green 公式, $AR(p)$ 序列 X_t 可表示为

$$X_t = \sum_{j=0}^{\infty} g_j \varepsilon_{t-j} \quad (4.62)$$

其中 g_j 为 Green 函数的系数. 则

$$X_{t+k} = (\varepsilon_{t+k} + g_1 \varepsilon_{t+k-1} + \cdots + g_{k-1} \varepsilon_{t+1}) + (g_k \varepsilon_t + g_{k+1} \varepsilon_{t-1} + \cdots).$$

上式中, $\varepsilon_t, \varepsilon_{t-1}, \cdots$ 为当前及过去的噪声序列.

基于无限数据的 $AR(p)$ 序列预测

记

$$\begin{aligned} e_t(k) &= \varepsilon_{t+k} + g_1\varepsilon_{t+k-1} + \cdots + g_{k-1}\varepsilon_{t+1} \\ \hat{X}_t(k) &= g_k\varepsilon_t + g_{k+1}\varepsilon_{t-1} + \cdots \end{aligned} \quad (4.63)$$

则 $X_{t+k} = e_t(k) + \hat{X}_t(k)$. 由 AR 模型的传递形式 (4.10) 可知, $\hat{X}_t(k)$ 可表示为 $\{X_t, X_{t-1}, \cdots\}$ 的线性组合, 即

$$\hat{X}_t(k) = \sum_{j=0}^{\infty} d_j X_{t-j}$$

若 ε_t 是均值为 0, 方差为 σ^2 的高斯白噪声, 则 $\{\varepsilon_{t+k}, \cdots, \varepsilon_{t+1}\}$ 与 $\{\varepsilon_t, \varepsilon_{t-1}, \cdots\}$ 相互独立.

基于无限数据的 $AR(p)$ 序列预测

故由无限多的历史数据 $\mathbf{X}_t = \{X_t, X_{t-1}, \dots\}$ 可得
 $E(\hat{X}_t(k) | \mathbf{X}_t) = \hat{X}_t(k), \text{Var}(\hat{X}_t(k) | \mathbf{X}_t) = 0.$

根据定理 4.4, X_{t+k} 的最佳预测为

$$\begin{aligned} E(X_{t+k} | \mathbf{X}_t) &= E\left(\left(e_t(k) + \hat{X}_t(k)\right) | \mathbf{X}_t\right) \\ &= E(e_t(k) | \mathbf{X}_t) + E(\hat{X}_t(k) | \mathbf{X}_t) \\ &= E(e_t(k)) + \hat{X}_t(k) = \hat{X}_t(k). \end{aligned} \quad (4.64)$$

基于无限数据的 $AR(p)$ 序列预测

即最佳预测为 $\hat{X}_t(k)$, 且预测的方差为

$$\begin{aligned}
 \text{Var}(X_{t+k} | \mathbf{X}_t) &= \text{Var}(e_t(k) | \mathbf{X}_t) + \text{Var}(\hat{X}_t(k) | \mathbf{X}_t) \\
 &= \text{Var}(e_t(k)) \\
 &= (1 + g_1^2 + \cdots + g_{k-1}^2) \sigma^2
 \end{aligned} \tag{4.65}$$

基于无限数据的 $AR(p)$ 序列预测

由于噪声服从正态分布, 故 $\hat{X}_t(k)$ 也服从正态分布. 由 (4.64) - (4.65) 式可知, 当噪声为高斯白噪声且知无限多的历史样本数据 $\mathbf{x}_t = \{x_t, x_{t-1}, \dots\}$ 时, 对于 X_{t+k} 的最佳预测将服从正态分布 $N(\hat{x}_t(k), \text{Var}(e_t(k)))$, 从而得到预测的 $(1 - \alpha)$ 的置信区间为

$$\left(\hat{x}_t(k) \pm z_{\frac{\alpha}{2}} (1 + g_1^2 + \dots + g_{k-1}^2)^{\frac{1}{2}} \sigma \right), \quad (4.66)$$

式中 $z_{\frac{\alpha}{2}}$ 表示标准正态分布的上 $\frac{\alpha}{2}$ 分位数.

基于无限数据的 $AR(p)$ 序列预测

例 (4.14)

设某城市每年的降雨量 (单位: mm) 近似服从 $AR(2)$ 模型

$$X_t = 200 + 0.5X_{t-1} + 0.3X_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 100)$$

已知 2009 – 2011 这三年的降雨量分别为: 1100 1000 1120. 请预测该市今后四年的降雨量及其 95% 的置信区间.

基于无限数据的 $AR(p)$ 序列预测

例 (4.14)

设某城市每年的降雨量 (单位: mm) 近似服从 $AR(2)$ 模型

$$X_t = 200 + 0.5X_{t-1} + 0.3X_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 100)$$

已知 2009 – 2011 这三年的降雨量分别为: 1100 1000 1120. 请预测该市今后四年的降雨量及其 95% 的置信区间.

由于噪声为高斯白噪声, 而且模型给定, 因此该模型可看为根据无限多的历史数据而得到的, 故最佳预测与最佳线性预测相等. 因此我们只需考虑最佳线性预测.

基于无限数据的 $AR(p)$ 序列预测

(1) 预测值的计算. 根据 (4.60) 式, 今后四年降雨量的预测值如下所示:

$$2012 \text{ 年: } \hat{x}_{2011}(1) = 200 + 0.5 \times x_{2011} + 0.3 \times x_{2010} = 1060,$$

$$2013 \text{ 年: } \hat{x}_{2011}(2) = 200 + 0.5 \times \hat{x}_{2011}(1) + 0.3 \times x_{2011} = 1066,$$

$$2014 \text{ 年: } \hat{x}_{2011}(3) = 200 + 0.5 \times \hat{x}_{2011}(2) + 0.3 \times \hat{x}_{2011}(1) = 1051,$$

$$2015 \text{ 年: } \hat{x}_{2011}(4) = 200 + 0.5 \times \hat{x}_{2011}(3) + 0.3 \times \hat{x}_{2011}(2) = 1062.1.$$

基于无限数据的 $AR(p)$ 序列预测

(2) 预测方差的计算. 根据 Green 函数的递推公式, 可得该 $AR(2)$ 模型的 Green 系数为

$$g_0 = 1$$

$$g_1 = a_1 g_0 = 0.5$$

$$g_2 = a_1 g_1 + a_2 g_0 = 0.5 \times 0.5 + 0.3 \times 1 = 0.55$$

$$g_3 = a_1 g_2 + a_2 g_1 = 0.5 \times 0.55 + 0.3 \times 0.5 = 0.425$$

从而可得预测方差为

$$\text{Var}(\hat{x}_{2011}(1)) = g_0^2 \sigma^2 = 100$$

$$\text{Var}(\hat{x}_{2011}(2)) = (g_0^2 + g_1^2) \sigma^2 = 125$$

$$\text{Var}(\hat{x}_{2011}(3)) = (g_0^2 + g_1^2 + g_2^2) \sigma^2 = 155.25$$

$$\text{Var}(\hat{x}_{2011}(4)) = (g_0^2 + g_1^2 + g_2^2 + g_3^2) \sigma^2 = 173.3125$$

基于无限数据的 $AR(p)$ 序列预测

(3) 置信区间的计算. 由于标准正态分布的 0.975 分位点 $z_{0.975} = 1.96$, 则根据 (4.66) 式, 我们可得 2012 - 2015 年预测的置信区间如下所示.

2012 年: (1040.40, 1079.60),

2013 年: (1044.09, 1087.91),

2014 年: (1026.58, 1075.42),

2015 年: (1036.30, 1087.90).

从中可知, 预测的置信区间随着预测步长的增加而增大.

- ① 4.6.1 最佳预测
- ② 4.6.2 最佳线性预测
- ③ 4.6.3 AR 模型的预测
- ④ 4.6.4 修正预测**
- ⑤ 参考资料

修正预测

预测的步长越长, 未知信息就越多, 从而估计的精度就越差. 然而, 随着时间的发展, 我们在原有的观测值 $\{\cdots, x_{t-1}, x_t\}$ 的基础上, 不断获得新的观测值 $\{x_{t+1}, x_{t+2}, \cdots\}$. 这些新观测值带来更多的信息, 从而预测未来时刻的未知信息逐渐减少.

因此, 利用新观测值的信息, 我们可以更好地预测未来的序列值 x_{t+k} , 预测精度将提高. 这就是所谓的修正预测.

修正预测

修正预测有两种处理方式.

一种处理方法是把新的观测值和原数据合并, 重新拟合模型, 然后再利用拟合后的模型预测 x_{t+k} ;

另一种处理方法是利用原来的拟合模型, 然后利用新观测值修正原来的拟合模型, 从而得到新的拟合模型. 当新的观测序列很多时或易于操作时, 可采用第一种方法. 然而, 当新的观测并不多时, 第一种方法不是最佳选择.

此时, 第二种方法将更加简便. 下面介绍第二种处理方法.

修正预测

假设已知观测值 $\{\cdots, x_{t-1}, x_t\}$, 则其 k 步最佳预测如 (4.63) 式所示. 若获得时刻 $t+1$ 的观测值 x_{t+1} , 即现有数据为 $\{\cdots, x_{t-1}, x_t, x_{t+1}\}$, 则根据 (4.63) 式, x_{t+k} 的预测值为

$$\begin{aligned}\hat{x}_{t+1}(k-1) &= g_{k-1}\varepsilon_{t+1} + g_k\varepsilon_t + g_{k+1}\varepsilon_{t-1} + \cdots \\ &= g_{k-1}\varepsilon_{t+1} + \hat{x}_t(k)\end{aligned}\quad (4.67)$$

其中 $\varepsilon_{t+1} = x_{t+1} - \hat{x}_t(1)$ 表示 x_{t+1} 的一步预测误差, $\hat{x}_{t+1}(k-1)$ 表示根据数据 $\{\cdots, x_{t-1}, x_t, x_{t+1}\}$ 对 $k-1$ 步后的 x_{t+k} 的预测值, 而 $\hat{x}_t(k)$ 表示根据数据 $\{\cdots, x_{t-1}, x_t\}$ 对 k 步后的 x_{t+k} 的预测值. 实际上, 修正后的预测值 $\hat{x}_{t+1}(k-1)$ 和原来的预测值 $\hat{x}_t(k)$ 都是对序列值 x_{t+k} 的预测.

修正预测

此时, 修正后的预测误差为

$$e_{t+1}(k-1) = x_{t+k} - \hat{x}_t(k) = g_0 \varepsilon_{t+k} + \cdots + g_{k-2} \varepsilon_{t+2}.$$

因此, 其预测误差的方差为

$$\text{Var}(e_{t+1}(k-1)) = (1 + g_1^2 + \cdots + g_{k-2}^2) \sigma^2,$$

其中 σ^2 为白噪声的方差. 与原预测的误差方差 (4.65) 相比, 修正后的预测方差减少了 $g_{k-1}^2 \sigma^2$, 因此, 修正后的预测精度有所提高.

修正预测

一般地, 假如获得新观测值 $x_{t+1}, \dots, x_{t+l} (1 \leq l < k)$, 则 x_{t+k} 的修正预测值为

$$\begin{aligned}\hat{x}_{t+1}(k-1) &= g_{k-l}\varepsilon_{t+l} + \dots + g_{k-1}\varepsilon_{t+1} + g_k\varepsilon_t + g_{k+1}\varepsilon_{t-1} + \dots \\ &= g_{k-l}\varepsilon_{t+l} + \dots + g_{k-1}\varepsilon_{t+1} + \hat{x}_t(k)\end{aligned}$$

其中 $\varepsilon_{t+j} = x_{t+j} - \hat{x}_{(t+j-1)+1} (1 \leq j \leq l)$ 为 x_{t+j} 的一步预测误差. 此时, 修正后的预测方差为

$$\text{Var}(e_{t+l}(k-l)) = (1 + g_1^2 + \dots + g_{k-l-1}^2) \sigma^2.$$

从上面的分析可知, 当我们获得新的观测值时, 修正后的预测方差将减少, 从而提高了预测精度. 而且这种修正方式简单, 易于操作.

修正预测

例 (4.15)

(例 4.14 续) 假如已知 2012 年该市的降雨量为 1100 mm, 预测后三年该市的降雨量及其 95% 的置信区间.

修正预测

例 (4.15)

(例 4.14 续) 假如已知 2012 年该市的降雨量为 1100 mm, 预测后三年该市的降雨量及其 95% 的置信区间.

(1) 计算 2012 年的一步预测误差:

$$\varepsilon_{2012} = x_{2012} - \hat{x}_{2011}(1) = 1100 - 1060 = 40.$$

修正预测

(2) 计算修正预测值:

$$\hat{x}_{2012}(1) = g_1 \varepsilon_{2012} + \hat{x}_{2011}(2) = 1086$$

$$\hat{x}_{2012}(2) = g_2 \varepsilon_{2012} + \hat{x}_{2011}(3) = 1073$$

$$\hat{x}_{2012}(3) = g_3 \varepsilon_{2012} + \hat{x}_{2011}(4) = 1079.1.$$

修正预测

(2) 计算修正预测值:

$$\hat{x}_{2012}(1) = g_1 \varepsilon_{2012} + \hat{x}_{2011}(2) = 1086$$

$$\hat{x}_{2012}(2) = g_2 \varepsilon_{2012} + \hat{x}_{2011}(3) = 1073$$

$$\hat{x}_{2012}(3) = g_3 \varepsilon_{2012} + \hat{x}_{2011}(4) = 1079.1.$$

(3) 计算修正预测方差:

$$\text{Var}(e_{2012}(1)) = g_0^2 \sigma^2 = 100$$

$$\text{Var}(e_{2012}(2)) = (g_0^2 + g_1^2) \sigma^2 = 125$$

$$\text{Var}(e_{2012}(3)) = (g_0^2 + g_1^2 + g_2^2) \sigma^2 = 155.25$$

修正预测

(4) 该市降雨量 $l(1 \leq l \leq k-1)$ 步预测值的 95% 的置信区间为

$$\left(\hat{x}_{2012}(l) - 1.96\sqrt{\text{Var}(e_{2012}(l))}, \hat{x}_{2012}(l) + 1.96\sqrt{\text{Var}(e_{2012}(l))} \right).$$

相应结果见表 4.7, 从中可知, 修正预测的置信区间比原预测的置信区间有所缩小, 即预测精度有所提高.

表 4.7 修正预测与原预测的比较

	原预测		修正预测	
年份	置信区间	长度	置信区间	长度
2012	(1040.40, 1079.60)	39.2		
2013	(1044.09, 1087.91)	43.82	(1066.40, 1105.60)	39.2
2014	(1026.58, 1075.42)	48.84	(1051.09, 1094.91)	43.82
2015	(1036.30, 1087.90)	51.60	(1054.68, 1103.52)	48.84

- ① 4.6.1 最佳预测
- ② 4.6.2 最佳线性预测
- ③ 4.6.3 AR 模型的预测
- ④ 4.6.4 修正预测
- ⑤ 参考资料

参考资料

- [1] 周永道, 王会琦, 吕王勇. 时间序列分析及应用. 北京: 高等教育出版社, 2015.



图 3: 时间序列分析及应用

感谢

Thanks

本人非统计专业，若有不妥之处，恳请批评指正。

作者：图灵的猫 邮箱：turingcat@126.com 主页：turingcat.club