2013年春季统计学153课程（时间序列分析）：第一讲

阿迪蒂亚·冈图博伊纳

2013年1月22日

时间序列是一组数值观测值，每个观测值都记录在特定的时间点。 时间序列数据随处可见。 本课程的目标是教授如何分析这类数据。

我们将重点介绍两种时间序列分析方法：（1）时间域方法，和（2）频率域方法（也称为时间序列的谱分析或傅里叶分析）。 在时间域方法中，直接使用数据进行分析，而在频率域方法中，使用数据的离散傅里叶变换进行分析。

大致上，本课程将有60%的内容涉及时间域方法，40%的内容涉及频率域方法。

1 时间序列模型

时间序列分析的主要目标是开发数学模型，为样本时间序列数据提供合理的描述。

我们假设观测数据 $x_1, \dots\dots\dots, x_n$ 是一系列随机变量的实现 $X_1, \dots\dots\dots, X_n$. 建模的基本策略始终是从简单开始逐步建立。

我们将在本课程中研究时间序列模型。基本模型：

1. 白噪声
2. 确定性趋势 + 白噪声
3. 确定性季节性 + 白噪声
4. 确定性趋势 + 确定性季节性 + 白噪声
5. 平稳时间序列模型
6. 平稳ARMA模型
7. ARIMA模型
8. 季节性ARIMA模型
9. 建模和估计谱密度

1.1 白噪声

$X_1, \ldots\ldots\ldots\ldots, X_n$被称为白噪声，如果它们的均值为零，方差为 σ^2，并且不相关。

白噪声的一个重要特例是高斯白噪声，其中 $X_1, \ldots\ldots\ldots\ldots\ldots, X_n$独立同分布 $N(0, \sigma^2)$。

如何检查白噪声是否适用于给定的数据集？ 从预测的角度思考。 对于白噪声，给定的数据无法帮助预测 X_{n+1}。 对 X_{n+1}的最佳估计是 $E(X_{n+1}) = 0$。 特别地， X_1无法预测 X_2； X_2无法预测 X_3等等。 因此， $Y= (X_1, \ldots, X_{n-1})$ 和 $Z= (X_2, \ldots, X_n)$ 之间的相关系数必须接近于零。

计算 Y 和 Z之间的相关系数的公式为：

$$r := \frac{\sum_{t=1}^{n-1}(X_t - \bar{X}_{(1)})(X_{t+1} - \bar{X}_{(2)})}{\sqrt{\sum_{t=1}^{n-1}(X_t - \bar{X}_{(1)})^2 \sum_{t=1}^{n-1}(X_{t+1} - \bar{X}_{(2)})^2}}$$

其中

$$\bar{X}_{(1)} = \frac{\sum_{t=1}^{n-1} X_t}{n-1} \text{ and } \bar{X}_{(2)} = \frac{\sum_{t=1}^{n-1} X_{t+1}}{n-1}.$$

这个公式通常被简化为

$$r_1 = \frac{\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

其中$\bar{X} := \sum_{t=1}^{n} X_t/n$. 请注意，我们将其称为相关系数 r_1（注意下标1）。 这个量 r_1被称为样本自相关系数（样本ACF），是 $X_1, \ldots\ldots\ldots\ldots, X_n$在滞后一的相关性。滞后一是因为这个相关性是在 X_t和 X_{t+1}之间的。

当 $X_1, \ldots\ldots\ldots\ldots\ldots\ldots$，$X_n$是从白噪声中获得的， r_1接近于零，特别是当 n很大时。

类似地，可以在其他滞后位置定义样本自相关：

$$r_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \qquad \text{对于 } k= 1, 2, \ldots$$

这里有一个重要的数学事实：在某些附加条件下（对于高斯白噪声而言是满足的），如果 $X_1, \ldots$，X_n是白噪声，那么样本自相关 $r_1, r_2, \ldots$ 是独立地根据均值为零和方差为$1/n$的正态分布分布。

注意，方差随着 n的增加而减小，均值为零。 因此，对于大的 n，样本自相关应该非常接近于零。 还要注意，不同滞后的样本自相关是独立的。

因此，检查白噪声模型是否适合数据的一种方法是绘制样本自相关图。 这个图被称为自相关图。 使用 **R**中的函数 **acf**来获取自相关图。 自相关图中的蓝色带对应于 $\pm 1.96 n^{-1/2}$的水平。

如何解释自相关图？ 当 $X_1, \ldots$，X_n是白噪声时，固定的 r_k在蓝色带之外的概率等于0.05。 在蓝色带之外的 r_k值是显著的，即它提供了反对纯随机性的证据。 然而，至少有一个 r_k在带之外的概率随着绘制的系数数量的增加而增加。 例如，如果绘制了20个 r_k值，纯随机性下预计会得到一个显著值。

以下是判断自相关图是否显示白噪声的一些经验法则：

- 一个在边界之外的单个值可以忽略，但是两个或三个明显偏离边界的值表明不是纯随机性。

- 在具有某种物理解释的滞后期（如滞后一期或与季节变化对应的滞后期）上的一个显著值也表明存在非随机性的证据。

2013年春季统计学153课程（时间序列）：第二讲

阿迪蒂亚・冈图博伊纳

2013年1月24日

1 最后一节课

- 数据示例
- 白噪声模型
- 样本自相关函数和自相关图

2 趋势模型

许多时间序列数据集显示出递增或递减的趋势。 通过将时间的确定性趋势函数添加到白噪声中，可以得到这种数据集的简单模型：

$$X_t = m_t + Z_t$$

这里， m_t是一个确定性趋势函数， Z_t是白噪声。 有两种主要技术可以将这个模型拟合到数据中。

2.1 参数形式为 m_t和线性回归

假设 m_t的简单参数形式，比如线性或二次，并通过线性回归进行拟合。

2.2 平滑

在这里，我们估计 m_t而不对其形式做任何参数假设。

这个想法是从 $X_t = m_t + Z_t$中得到 m_t，我们需要消除 Z_t。 众所周知，通过平均可以消除噪声。考虑

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j}。 \qquad (1)$$

如果 m_t在线性区间$[t-q, t+q]$上是线性的，则检查

$$\frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j} = m_t。$$

因此，如果 m_t在$[t-q,\ t+q]$上近似线性，则

$$\hat{m}_t \approx m_t + \frac{1}{2q+1}\sum_{j=-q}^{q} Z_{t+j} \approx m_t.$$

计算接近端点的平均值时，$\hat{m}_t$的定义方程会有困难。 为了对抗这个问题，只需将 X_t定义为$t<1$ 时的 X_1 和$\mathrm{t}>\mathrm{n}$ 时的 X_n。

$\hat{m}_t$也被称为 X_t的简单移动平均。

关键问题：如何选择平滑参数 q？观察到：

$$\hat{m}_t = \frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j} + \frac{1}{2q+1}\sum_{j=-q}^{q} Z_{t+j}.$$

如果 q非常小，那么上述第二项并不是非常小，因此趋势估计也会包含一些噪声成分，从而使得$\hat{n}_t$非常嘈杂。 另一方面，如果 q很大，那么假设 m_t在 $[t-q,t+q]$ 上是线性的可能并不完全正确，因此，$\hat{m}_t$可能与 m_t不接近。这通常被称为偏差-方差权衡。 因此， q既不应该太小也不应该太大。

2.2.1 参数曲线拟合与平滑

假设有充分的理由相信存在一个潜在的线性或二次趋势函数。 在这种情况下，使用平滑仍然可以吗？

不，当真的存在一个潜在的线性趋势时，拟合一条直线可以给出更精确的趋势估计。 另一方面，通过平滑估计趋势只使用每个时间点的少数观测值，得到的估计结果不够精确。 这是放弃线性假设所付出的代价。 当没有理由相信存在一个潜在的线性趋势时，拟合一条直线可能毫无意义。 在这种情况下，平滑是可行的方法。

2.3 更一般的滤波

趋势函数 m_t的平滑估计（1）是线性滤波的特例。 线性滤波将观测时间序列 X_t转换为趋势估计值$\hat{m}_t$，通过线性运算实现：

$$\hat{m}_t = \sum_{j=-q}^{s} a_j X_{t+j}.$$

这些数字 $a_{-q}, a_{-q+1}, \ldots \ldots \ldots \ldots, a_{-1}, a_0, a_1, \ldots$ 称为滤波器的权重。 平滑方法显然是滤波器的特殊实例，其中 $s=q$和 $a_j = 1/(2q+1)$ 对于 $|j| \leq q$，其他情况下为0。

可以将滤波器看作是一个（线性）$\mathrm{sys}_t\mathrm{em}$，它将观测到的序列 X_t 作为输入，并产生趋势的估计值，即$\hat{n}_t$作为输出。

除了选择 $aj = 1/(2q+1)$ 对于 $|_j| \leq q$之外，人们通常还使用其他选择的滤波器。

(1)二项式权重：基于以下思想。 当我们估计趋势的值时
m_t at t，相比于 $X_{t\pm1}$，给予 X_t更高的权重，相比于 $X_{t\pm2}$等等。这样的权重示例为：

$$a_j = 2^{-q}\binom{q}{q/2+j} \qquad \text{对于 } j = -q/2, -q/2+1, \ldots, -1, 0, 1, \ldots, q/2\text{。}$$

与常规平滑一样，选择 q在这里也是一个问题。

(2)**Spencer的15点移动平均**：我们已经看到简单移动平均滤波器不会改变线性函数。 是否可能设计一个滤波器，使高阶多项式保持不变？ 例如，我们能否设计一个滤波器，使所有二次多项式保持不变。

是的！

对于一个具有权重的滤波器，使所有二次多项式保持不变，我们需要满足以下条件对于每个二次多项式m和$_t$：

$$\sum_j a_j m_{t+j} = m_t \qquad \text{对于所有}t$$

换句话说，如果m_t=α_t^^{2}+β_t+γ，我们需要

$$\sum_j a_j \left(\alpha(t+j)\text{\^{}}^2+\beta(t+j)+\gamma \qquad \right) = \alpha t\text{\^{}}^2+\beta t+\gamma \qquad \text{对于所有}t.$$

简化得到

$$\alpha t\text{\^{}}^2+\beta t+\gamma = (\alpha t\text{\^{}}^2+\beta t+\gamma) \qquad \sum_j a_j + (2\alpha t+\beta)\sum_j j a_j + \alpha \sum_j j\text{\^{}}^2 a j \qquad \text{对于所有}t.$$

如果满足这个条件，那么显然成立

$$\sum_j a_j = 1 \qquad \sum_j j a_j = 0 \qquad \sum_j j^2 a_j = 0. \tag{2}$$

这样的滤波器的一个例子是Spencer的15点移动平均值，定义为

$$a_0 = \frac{74}{320}, a_1 = \frac{67}{320}, a_2 = \frac{46}{320}, a_3 = \frac{21}{320}, a_4 = \frac{3}{320}, a_5 = \frac{-5}{320}, a_6 = \frac{-6}{320}, a_7 = \frac{-3}{320}$$

和 a_j= 0对于j >7. 此外，这个滤波器在对称性上是对称的，即 $a_{-1} = a_1$， a_{-2} =a$_2$等等。检查这个滤波器是否满足条件（2）。

因为这是一个对称的滤波器，可以检查它是否允许所有的三次多项式通过。

（3）指数平滑：一种非常流行的平滑方法（维基百科上有一个很大的页面关于这个）。 它也被用作一种预测技术。

在这种方法中，为了得到 $\hat{m}_t$，只使用了前面的观测值 X_{t-1}，X_{t-2}，X_{t-3}，. . . . 这些观测值被赋予的权重指数级地减小随着时间的推移。具体来说，

$$\hat{m}_t := \alpha X_{t-1} + \alpha(1-\alpha)X_{t-2} + \alpha(1-\alpha)^2 X_{t-3} + \cdots + \alpha(1-\alpha)^{t-2}X_1 + (1-\alpha)^{t-1}X_0.$$

检查权重是否相加为1。α是确定平滑程度的参数（ α在平滑中类似于 q）。 如果α接近1，平滑程度很小，反之亦然。

2013年春季统计学153课程（时间序列）：第三讲

阿迪蒂亚·冈图博伊纳

2013年1月29日

用于趋势消除的一阶差分

在上一节课中，我们学习了趋势模型： $X_t = m_t + Z_t$其中 m_t是确定性趋势函数，$\{Z_t\}$ 是白噪声。

在拟合模型 $X_t = m_t + Z_t$后得到的残差被研究，以确定它们是否是白噪声或者是否存在某种依赖结构可用于预测。

假设目标只是产生这种去趋势的残差。 差分是一种简单的技术，可以产生这种去趋势的残差。

只需看 $Y_t = X_t - X_{t-1}, t= 2, \dots , n$. 如果 $X_t = m_t + Z_t$的趋势 m_t是线性的，那么这个操作就是简单地去除它，因为如果 $m_t = \alpha t + b$, 那么 $m_t - m_{t-1} = \alpha$，所以 $Y_t = \alpha + Z_t - Z_{t-1}$.

假设第一阶差分序列 Y_t看起来像白噪声。 那么对于原始序列 X_{n+1}，什么样的预测是合理的？ 因为 Y_t像白噪声，我们通过样本均值 $\bar{Y} := (Y_2 + \cdots + Y_n)/(n - 1)$来预测 Yn_{+1}. 但是由于 $Y_{n+1} = X_{n+1} - X_n$，这导致了预测 $X_n + \bar{Y}\cdot$
X_{n+1}.

有时，即使经过差分，数据中仍然可以观察到趋势。 在这种情况下，只需再次进行差分。 使用符号 ∇表示差分是很有用的：

$$\nabla X_t = X_t - X_{t-1} \qquad \text{对于 } t = 2, \dots , n$$

而二次差分对应于

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2} \qquad \text{对于 } t = 3, \dots , n.$$

可以证明二次趋势在操作 ∇^2后会消失。 假设数据
$\nabla^2 X_t$看起来像白噪声，你将如何预测 X_{n+1}?

差分是产生去趋势残差的一种快速简便方法，也是ARIMA预测模型（稍后介绍）的关键组成部分。 然而，一个问题是它不会得出任何对趋势函数的估计。

2随机漫步模型（带漂移）

考虑以下模型 X_t：

$$R_t = \delta + R_{t-1} + W_t$$

对于 $t = 1, 2, \ldots,$初始条件为 $R_0 = 0$， W_t为白噪声。 这个模型也可以写成：

$$R_t = \delta t + \sum_{j=1}^{t} W_j$$

当 $\delta = 0$时，这个模型被称为随机漫步模型。 这经常用于建模趋势。 这将是一个随机模型，与之前的确定性模型不同。

考虑模型 $X_t = m_t + Z_t$ 其中 Z_t 是白噪声， m_t 是带漂移的随机漫步：
这是一个动态线性模型（DLM）的例子。W_t被称为演化误差
and Z_t被称为观测误差。

对于 X_t的差分序列是：

$$\nabla X_t = X_t - X_{t-1} = m_t - m_{t-1} + Z_t - Z_{t-1} = \delta + W_t + Z_t - Z_{t-1}.$$

因此， ∇X_t是一个去趋势的序列。 通过平稳模型对去趋势序列进行建模的方法将在以后研究。

季节性的3种模型

许多时间序列数据集表现出季节性。 模拟这种情况的最简单方法是： $X_t = s_t + Z_t$其中 s_t是一个已知周期 d的周期函数，即 $s_{t+d} = s_t$对于所有 t。 这样的函数 s模拟季节性。
这些模型适用于具有季节性模式的月度、季度或周度数据集。

然而，这个模型对于同时具有趋势和季节性的数据集不适用，这是更现实的情况。 稍后将对这些进行研究。

与趋势情况一样，处理季节性有三种不同的方法：拟合参数函数、平滑和差分。

3.0.1 拟合参数季节性函数

周期为 d的最简周期函数是： $a\cos(2\pi f\,t/d)$ 和 $a\sin(2\pi f\,t/d)$。 这里 f是一个正整数。 量 a被称为振幅，f/d被称为频率，它的倒数， d/f被称为周期。 较高的 f值意味着函数中的振荡更快。

更一般地说，

$$s_t = a_0 + \sum_{f=1}^{k} \left(a_j \cos(2\pi f\ t/d) + b_j \sin(2\pi f\ t/d)\right) \tag{1}$$

是一个周期函数。 选择一个值为 k（不要太大）并将其拟合到数据中。

对于 $d = 12$，不需要考虑大于6的 k值。 当 $k = 6$时，每个周期函数都可以用形式（1）来表示。 在我们研究时间序列的频域分析时，会更详细地讨论这个问题。

3.0.2 平滑

由于周期性，函数 s_t仅取决于 d值 s_1，s_2，$\ldots, s_d$。 显然， s_1可以通过 X_1，X_{1+d}，X_{1+2d}，$\ldots$的平均值来估计。 例如，对于月度数据，这对应于

通过对所有一月观测值求平均来估计一月的平均项。

$$\hat{s}_i := X_i, X_{i+d}, X_{i+2d}, \ldots \text{ 的平均值}$$

请注意，我们在这里从n个观测值中拟合12个参数（分别对应$s_1, \ldots, s_d$）。 如果n不是很大，拟合12个参数可能会导致过拟合。

3.0.3 差分

如何在不明确拟合季节性函数的情况下从数据中获得经过季节性调整的残差？ 回想一下，如果对于所有t，$s_{t+}\ d = s_t$，那么函数s是周期为d的周期函数。 我们在这里考虑的模型是：$X\ t = s_t + Z_t$。

因此，滞后-d的差分数据
$X_t - X_{t-d}$不显示任何季节性。 这种产生去季节化残差的方法被称为季节性差分。

4 数据转换

假设时间序列数据集具有趋势，并且随着趋势函数的增加，变异性也增加。 一个例子是R中的UKgas数据集。在这种情况下，使用对数或平方根转换数据，使得结果数据看起来具有合理的同方差性（方差相同）。

为什么要使用对数或平方根？ 了解一些方差稳定转换的知识会有所帮助。
假设 X是一个具有均值 m的随机变量。 一个非常启发式的计算给出了随机变量X的函数 $f(\ X)$的方差的近似值。 将 $f(X)$在其泰勒级数中展开，直到一阶项为止，围绕 m展开。

$$f(X) \approx f(m) + f'(m)(X - m)$$

因此,

$$\mathrm{var}(f(X)) \approx \mathrm{var}\ (f(m) + f'(m)(X - m)) = (f'(m))^2 \mathrm{var}(X).$$

因此如果

1. $\mathrm{var}(\ X) = Cm$ and $f(x) = \sqrt{x}$, 我们会得到 $\mathrm{var}(\quad X) \approx C/4.$
2. $\mathrm{var}(\ X) = Cm^2$ and $f(x) = \log x$, 我们会得到 $\mathrm{var}(\quad X) \approx C.$

关键是要注意，在上述两种情况下， $f(X)$的近似方差不再依赖于
m .

上述粗略计算揭示了对时间序列数据分析的以下见解。 一个模型形式为 $X_t = m_t + W_t$其中 m_t是一个确定性函数， W_t是纯随机或平稳的（下周）假设 X_t的方差不随 t变化。 然而，假设数据的时间图显示 X_t的方差随其均值 m_t增加，例如 $\mathrm{var}(X_t) \propto m_t$。 那么，粗略计算表明 $\mathrm{var}(\sqrt{X_t})$应该近似恒定（不依赖于 t），因此应该将模型 $m_t + W_t$拟合到变换后的数据 $\sqrt{X_t}$而不是原始数据 X_t。 同样地，如果 $\mathrm{var}(X_t) \propto m_t^2$，那么 $\mathrm{var}(\log X_t)$应该近似恒定。

因此，如果数据显示出趋势增加的变异性，那么可以应用对数或平方根等转换，具体取决于结果数据集在时间上的变异性是否恒定。

顺便说一下，计数数据通常通过泊松随机变量进行建模，泊松分布的方差等于其均值。 因此，在处理计数（泊松）数据时，通常使用平方根。

如果使用模型$X_t = m_t + W_t$，其中m_t是非确定性（随机）趋势函数，这将自动允许X_t的方差随$_t$变化。 在这种情况下，我们可能不需要对数据进行转换。 这些模型可以看作是状态空间模型的特例，我们稍后会简要介绍。

Box-Cox转换：平方根和对数是**Box-Cox**转换的特例，表示为：

$$\begin{aligned} Y_t &= \frac{X_t^\lambda - 1}{\lambda} \qquad \text{if } \lambda = 0 \\ &= \log X_t \text{ 如果 } \lambda = 0。 \end{aligned} \tag{2}$$

平方根基本上对应于 $\lambda = 1/2$。

2013年秋季统计学151课程（线性模型）：第四讲

阿迪蒂亚・冈图博伊纳

2013年9月10日

1 复习

1.1 回归问题

有一个响应变量 y和 p个解释变量 $x_1, \ldots, x_p$。目标是理解 y和 $x_1, \ldots, x_p$之间的关系。

有 n个主体，数据是从这些主体收集的变量。

响应变量的数据是 $y_1, \ldots$，y_n，并且用列向量 $Y = (y_1, \ldots, y_n)^T$（这里的 T表示转置）。

第 j个解释变量 x_j的数据是 $x_{1j}, x_{2j}, \ldots, x_{nj}$。这些数据由 $n \times p$矩阵 X表示，其中（i, j）的元素是 x_{ij}。换句话说， X的第 i行包含从第 i个主体收集的数据， X的第 j列包含第 j个变量的数据。

1.2 线性模型

1. $y_1, \ldots$，y_n 被假设为随机变量，但 x_{ij} 被假设为非随机的。
2. 假设

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i \quad \text{对于} \quad i = 1, \ldots, n$$

其中 $e_1, \ldots$，e_n 是不相关的随机变量，均值为零，方差为 σ^2。

用矩阵表示，上述第二个假设可以写成

$$Y = X\beta + e \qquad \text{其中} \quad \mathbb{E}e = 0, \quad Cov(e) = \sigma^2 I_n$$

其中 I_n 表示 $n \times n$ 的单位矩阵。 $\beta = (\beta_1, \ldots, \beta_p)^T$， $e = (e_1, \ldots, e_n)^T$。

如果 $Z_1, \ldots$, Z_n是随机变量，其中 $Z= (Z_1, \ldots, Z_n)^T$，那么 $Cov(Z)$表示的是 $n \times n$矩阵中的（i, j）项表示 Z_i和 Z_j之间的协方差。特别地，
$Cov(Z)$的$_i$th对角线项表示随机变量 Zi的方差。因此， $Cov(e) = \sigma^2 I_n$是一种简洁的
方式来表示当$_i =_j$时， $e\,i$和 $e\,j$之间的协方差等于0，当 $i = j$时，协方差等于 σ^2。

2 截距项

线性模型规定了

$$\mathbb{E}y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

对于 $i = 1, \ldots, n$。 这意味着当解释变量的值 $x_{i1}, \ldots, x_{ip}$都等于0时， $\mathbb{E}y_i = 0$。 当然，这并不总是一个合理的假设。 因此，稍微修改线性模型规定为

$$\mathbb{E}y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}. \tag{1}$$

现在当所有解释变量取零值时， $\mathbb{E}y_i$不必为零。该术语
上述β_0被称为截距项。 通常，在线性模型中，默认情况下，总是包括截距项。

如果我们将 x_0表示为始终取值为1的“变量”，那么（1）可以写成

$$\mathbb{E}y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}.$$

因此，带有截距项的模型与先前的线性模型相同，只是增加了这个附加变量以及解释变量。

有了截距项，可以将线性模型以矩阵形式表示为

$$Y = X\beta + e \quad \text{其中} \quad \mathbb{E}\ e = 0, \quad Cov\ (\ e\ ) = \ \sigma^2 I_n$$

其中，X表示一个$n \times (p+1)$的矩阵，其第一列由全为1的元素组成，其余列对应于p个解释变量的值，而$\beta = (\beta_0, ..., \beta p)^T$。

当$p = {}_1$时，即只有一个解释变量时，这个线性模型（带有截距项）被称为简单线性回归模型：$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$。

从现在开始，我们将始终考虑带有截距项的线性模型，这意味着X的第一列（一个$n \times (p+1)$的矩阵）始终是由1组成的向量，而β是长度为$p+1$的向量。

3 线性模型中的估计

量 $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ 和 $\sigma^2 > 0$ 是线性模型中的参数。 这些需要从数据中估计。 估计参数的过程也被称为将线性模型拟合到数据中。

让我们首先关注 β的估计。

线性模型的思想是通过线性组合 $\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$来解释响应值 y_i。 因此，通过最小化平方和来估计 β是有意义的。

$$S(\beta) := \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_p x_{ip})^2\text{。}$$

使用矩阵符号，可以写成

$$S(\beta) = ||Y - X\beta||^2.$$

向量$x=(x_1, \ldots\ldots\ldots\ldots\ldots, x_n)^T$ 的范数 $||x||$of 定义为 $||x||\sqrt{=}$ 注意等式 $||x||^2 = x^T x$. 利用这个等式，我们可以写成

$$S(\beta) = (Y - X\beta)^T(Y - X\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta.$$

通过微积分可以将其最小化。 对于$_i= 0, 1, \ldots\ldots$，对 β i进行偏导数，并令其等于0。很容易验证，p和将它们等于0。很容易验证

$$\nabla S(\beta) = 2X^T X\beta - 2X^T Y.$$

其中

$$\nabla S(\beta) = \left(\frac{\partial S(\beta)}{\partial \beta_1}, \ldots, \frac{\partial S(\beta)}{\partial \beta_p}\right).$$

表示对β的梯度 $S(\beta) = ($ $\beta_1, \ldots, \beta_p)^T$. 因此，最小化函数 $S(\beta)$ 的解满足以下等式

$$X^T X\beta = X^T Y. \tag{2}$$

这给出了 p个未知数的 p个线性方程 $\beta_1, \ldots, \beta_p$. 这个重要的方程组被称为正规方程. 它们的解，记为$\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ 给出了 β的最小二乘估计值. 如果一个主题的解释变量的值是 $\lambda_1, \ldots, \lambda$ p, 那么他的均值响应的估计值由$\hat{\beta}_0 + \lambda_1$ ^给出

$$\beta_1 + \cdots + \lambda_p \hat{\beta}_p.$$

出现了两个重要问题：（1）正规方程是否存在解
（2）如果存在解，那么解是否唯一？

第一个问题的答案是是的. 正规方程总是有解的。 原因是如下： $X^T Y$在 X^T的列空间中。 此外， X^T和 $X^T X$的列空间是相同的，因此 $X^T Y$总可以写成 $X^T Xu$的形式，其中 u是某个向量。

第二个问题的答案是如果 $X^T X$可逆，则是是的，如果 $X^T X$不可逆，则是否的.

正规方程（2）是否有唯一解？答案：一般来说是不的。如果 $X^T X$可逆，则是是的.

如果 $X^T X$可逆，则正规方程的解为$\hat{\beta} = (X^T X)^{-1} X^T Y$. 对于向量 $\lambda = (\lambda_1, \ldots, \lambda_p)^T$，线性函数$\lambda^T \beta$的估计值为 $\lambda^T (X^T X)^{-1} X^T Y$.

如果 $X^T X$不可逆，则正规方程有很多解。 在这种情况下，如何估计 β? 在这里，实际上向量 β无法估计。下面将进行解释。

4 当 $X^T X$不一定可逆时

在这种情况下，向量 β无法估计。

首先观察到 $X^T X$可逆等价于 X的秩等于 $p+1$。 因此，当 $X^T X$不可逆时， X的秩严格小于 $p+1$。 换句话说， X的某一列是其他列的线性组合，即至少一个解释变量是冗余的，可以写成其他解释变量的线性组合。

让我们在这里考虑一个例子。 假设 $p=2$，两个解释变量 x_1和 x_2实际上是相同的，即对于每个$_i= 1, \ldots, n$， $x_{i1} = x$ i_2。 显然， X的秩最多为2。 线性模型可以写成

$$y_i = \beta_0 + (\beta_1 + \beta_2)\, x_{i1} + \epsilon_i$$

对于 $i= 1, \ldots, n$. 从这些观察中可以清楚地看出，参数 β_1和 β_2无法被估计。 另一方面， $\beta_1 + \beta_2$可以被估计。

因此，当 X^TX不可逆时，参数向量 β无法被估计，而某些特殊的线性组合可以被估计。

可以证明，只有当 λ位于 X^T的列空间中时，线性组合 $\lambda^T\beta$才能被估计。 这等价于说 λ位于 X^TX的列空间中，因为 X^T和 X^TX的列空间总是相等的。

在刚才讨论的例子中，向量（$0,1,1$）^{T}位于 X^T的列空间中，这意味着 $\beta_1+\beta_2$是可估计的。 另一方面，向量（$0,1,0$）^{T}不在 X^T的列空间中，这意味着 β_1不可估计。

当 X^TX可逆时， X^T 的列空间包含所有（$p+1$）维向量，然后 β的每个线性组合都是可估计的。

5最小二乘估计

考虑正规方程 $X^TX\beta=X^TY$。 令$\hat{\beta}_{ls}$表示任意解（只有在 X^TX可逆时才唯一）。

令 $\lambda^T\beta$为可估计的（即 λ在 X^T的列空间中，或等价地在 X^TX的列空间中）。然后用 $\lambda^T\hat{\beta}_{ls}$估计$\lambda^T\beta$。这被称为 $\lambda^T\beta$的最小二乘估计。

结果5.1. 如果 $\lambda^T\beta$是可估计的，那么 $\lambda^T\hat{\beta}_{ls}$对于正规方程的每个解 $\hat{\beta}_{ls}$都是相同的。
换句话说， $\lambda^T\beta$的最小二乘估计是唯一的。

证明。 由于 $\lambda^T\beta$是可估计的，向量 λ位于 X^TX的列空间中，因此 $\lambda=X^TXu$，其中某个向量 u。因此，

$$\lambda^T\hat{\beta}_{ls}=u^TX^TX\hat{\beta}_{ls}=u^TX^TY$$

最后一个等式是因为$\hat{\beta}_{ls}$满足正规方程。 由于 u只取决于 λ，这证明了 $\lambda^T\hat{\beta}_{ls}$不依赖于正规方程解$\hat{\beta}_{ls}$的特定选择。

$\square$

因此，当 $\lambda^T\beta$是可估计的时候，它被最小二乘估计 $\lambda^T\hat{\beta}_{ls}$（这是唯一确定的）。 当 $\lambda^T\beta$不可估计时，当然没有意义去尝试估计它。

2013年秋季统计学151课程（线性模型）：第五讲

阿迪蒂亚・冈图博伊纳

2013年9月12日

1 最小二乘估计的线性模型中的 β

线性模型是

$$Y = X\beta + e \quad \text{其中} \quad \mathbb{E}\, e = 0, \quad Cov(e) = \sigma^2 I_n$$

其中 Y是一个 $n \times 1$向量，包含所有响应值， X是一个 $n \times (p+1)$矩阵，包含所有解释变量的值（ X的第一列是全为一的列）， $\beta = (\beta_0, \beta 1, \ldots, \beta_p)^T$ （β_0是截距）。

正如我们上次所见， β通过最小化 $S(\beta) = ||Y - X\beta||^2$来估计。 对 β求导并令其为零，可以得到正规方程

$$X^T X\beta = X^T Y.$$

如果 $X^T X$是可逆的（这等价于 X的秩等于 $p+1$），那么正规方程的解是唯一的，并且给出的解为

$$\hat{\beta} := (X^T X)^{-1} X^T Y$$

这是 β的最小二乘估计。

特殊情况2：简单线性回归

假设只有一个解释变量 x。 矩阵 X的大小为 $n \times 2$，其中 X的第一列由全为1的元素组成，第二列由解释变量 $x_1, \ldots, x_n$的值组成。因此

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}.$$

检查

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

其中 $\bar{x} = \sum_i x_i/n$。还有 $\bar{y} = \sum_i y_i/n$. 因为

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

我们得到

$$(X^TX)^{-1} = \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}.$$

另外

$$X^TY = \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

因此

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \sum_i x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.$$

简化得到

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \bar{y}\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \end{pmatrix}.$$

因此

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

和

$$\hat{\beta}_0 = \frac{\bar{y}\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \bar{y} - \hat{\beta}_1\bar{x}$$

如果我们得到一个新的主题，其解释变量值为 x，我们对其响应的预测为

$$y = \hat{\beta}_0 + \hat{\beta}_1 x. \qquad (1)$$

1）如果以上的预测结果在图表上绘制出来（以x轴为横轴），则会得到一条被称为回归线的直线。

回归线的表达式比（1）要好得多。要看到这一点，注意到

$$y = \beta\text{^}_0 + \beta\text{^}_1 x = y\text{平均值} - \bar{x}\beta\text{^}_1 + \beta\text{^}_1 x = y\text{平均值} + \beta\text{^}_1(x - \bar{x})$$

这可以写成

$$y - y\text{平均值} = \beta\text{^}_1(x - x\text{平均值}) = \frac{\sum_{i=1}^n (x\_i - x_{\text{平均值}})(y\_i - {}_{y\text{平均值}})}{\sum_{i=1}^n (x_i - \bar{x})^2}(x - x\text{平均值}) \qquad (2)$$

使用符号表示

$$r := \frac{\sum_{i=1}^n (x\_i - x_{\text{平均值}})(y\_i - y_{\text{平均值}})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad s_x := \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}, \quad s_y := \sqrt{\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2},$$

我们可以将预测方程（2）重写为

$$\frac{y - \bar{y}}{s_y} = r\,\frac{x - \bar{x}}{s_x}. \qquad (3)$$

r是 x和 y之间的相关性，始终在-1和1之间。

作为一个推论，注意如果（$x - \bar{x}$）/s_x= 1，即，如果受试者的解释变量值比样本均值高一个标准差，那么其响应变量的预测值只会比其均值高 r个标准差。 弗朗西斯・高尔顿将此称为回归到平庸，这就是回归的名字来源。

随机向量的3个基本均值和协方差公式

接下来，我们想要探索线性模型中$\hat{\beta} = (X^TX)^{-1}X^TY$作为 β的估计量的性质。 为了这个目的，我们需要一些关于均值和协方差的事实。

令 $Z= (Z_1, \ldots , Z_k)^T$为一个随机向量。 其期望 $\mathbb{E}Z$定义为一个向量，其第$_i$个元素是 $Z\ i$的期望，即 $\mathbb{E}Z= (\mathbb{E}Z_1, \mathbb{E}Z_2, \ldots , \mathbb{E}Z_k)^T$。

随机向量 Z的协方差矩阵，记作 $Cov(Z)$，是一个 $k\times k$的矩阵，其第(i, j)个元素是 Z_i和 Z_j的协方差。

如果另一个随机向量 $W= (W_1, \ldots , W_m)^T$，则 Z和 W的协方差矩阵，记作 $Cov(Z, W)$，是一个 $k \times m$的矩阵，其第(i, j)个元素是 Z_i和 W_j的协方差。 注意到， $Cov(Z, Z) = Cov(Z)$。

以下公式非常重要：

1.$\mathbb{E}(AZ + c) = A\ \ \mathbb{E}(Z) + c$对于任意常数矩阵 A和任意常数向量 c。

2. $Cov(AZ + c) = ACov(Z)A^T$对于任意常数矩阵 A和任意常数向量 c。

3.$Cov(AZ +c, BW + d) = ACov(Z, W)B$对于任意一对常数矩阵 A和 B以及任意一对常数向量 c和 d。

线性模型为

$$Y = X\beta + e \qquad \text{其中 } \mathbb{E}e= 0 \text{ 且 } Cov(e) = \sigma^2 I_n\text{。}$$

根据上述公式（记住 X和 β是固定的），

$$\mathbb{E}Y\ = X\beta \text{ 且 } Cov(Y) = \sigma^2 I_n\text{。}$$

最小二乘估计的4个性质

假设 X^TX是可逆的（等价地， $X\text{has rank } p+ 1$），并考虑最小二乘估计

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

作为β的估计量， $\hat{\beta}$有什么性质?

4.1 线性性

如果一个参数的估计量可以写成 AY的形式，那么它被称为线性的，其中 A是一个矩阵。 显然， $\hat{\beta} = (X^TX)^{-1}X^TY$具有这种形式，因此它是 β的线性估计量。

4.2 无偏性

如果一个参数的估计量的期望等于该参数（对于所有参数值），则称其为无偏的。

最小二乘估计的期望是（使用期望的公式： $\mathbb{E}AZ\ = A\mathbb{E}Z$）

$$\mathbb{E}\hat{\beta} = \mathbb{E}((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^T\mathbb{E}Y = (X^TX)^{-1}X^TX\beta = \beta$$

特别是，这意味着 $\mathbb{E}\hat{\beta}_i = \beta_i$对于每个 i，这意味着每个$\hat{\beta}_i$都是无偏估计 of β_i。 更一般地，对于每个向量 λ，数量 $\lambda^T\hat{\beta}$ 是 $\lambda^T\beta$的无偏估计。

4.3 协方差矩阵

估计器 $\hat{\beta}$ 的协方差矩阵可以使用公式轻松计算： $Cov(AZ) = ACov(Z)A^T$:

$$Cov(\hat{\beta}) = Cov((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^TCov(Y)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$$

特别是，β^的方差等于σ^^{2}乘以$(X$^$^TX)$^-^{1}的第$_i$个对角元素。 一旦我们学会如何估计σ，我们就可以用它来计算β^$_i$的标准误差。

4.4 最优性 - 高斯-马尔可夫定理

高斯-马尔可夫定理指出β^是BLUE（最佳线性无偏估计）。这意味着 β^是所有线性和无偏估计器中的“最佳”估计器。 这里，“最佳”是指方差最小。 这意味着β^$_i$在所有线性和无偏估计器中具有最小的方差。

β_i。

2013年秋季统计学151课程（线性模型）：第六讲

阿迪蒂亚·冈图博伊纳

2013年9月17日

我们再次考虑$Y = X\beta + e$，其中$\mathbb{E}[e] = 0$且$Cov(e) = \sigma$^$^2 I$_$_n$。通过解正态方程$X^T X\beta = X^T Y$来估计β。

1 回归平面

如果我们得到一个新的主题，其解释变量的值为 $x_1, \ldots, x_p$, 那么我们对其响应变量值的预测是

$$.\hat{\beta}_p x_p$$

这个方程表示一个平面，我们称之为回归平面。

2 拟合值

这些是线性模型对于N个主题的预测值。

解释变量的值为 $x_{i1}, \ldots, x_{ip}$对于第 i个主题。因此，线性模型对于第 i个主题的预测是

$$\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}.$$

因为第 i个主题的响应变量值为 y_i，所以上述预测有意义。因此

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} \qquad \text{对于 } i = 1, \ldots, n.$$

这些值$\hat{y}_1, \ldots, \hat{y}_n$被称为拟合值，向量$\hat{Y} = (\hat{y}_1, \ldots, \hat{y}_n)^T$被称为拟合值向量。 这个向量可以简洁地写成$\hat{Y}$ $= X\hat{\beta}\cdot$ 因为$\hat{\beta} = (X^T X)^{-1} X^T Y$, 我们可以写成

$$\hat{Y} = X(X^T X)^{-1} X^T Y.$$

拟合值向量 $\hat{Y}$ 是 Y 在列空间上的（正交）投影。
X.

因为乘以 H会将Y变成 $\hat{Y}$，所以这个矩阵
H被称为帽子矩阵。 在线性回归中非常重要。 它具有以下三个容易验证的性质：

1. 它是一个对称的 $n \times n$ 矩阵。
2. 它是幂等的，即 $H^2 = H$。
3. $HX = X$。

4. H和 X的秩是相同的。

这些可以从定义 $H = X(X^TX)^{-1}X^T$中轻松推导出来。 因为这些，我们得到

$$\mathbb{E}\hat{Y} = \mathbb{E}(HY) = H(\mathbb{E}Y) = HX\beta = X\beta = \mathbb{E}Y。 \tag{1}$$

因此$\hat{Y}$和 Y具有相同的期望值。同时

$$Cov(\hat{Y}) = Cov(HY) = HCov(Y)H^T = H(\sigma^2 I)H = \sigma^2 H.$$

3个残差

第$_i$个主题的 y_i和$\hat{y}_i$之间的差异被称为第$_i$个主题的残差。 $\hat{e}\ i := y_i - \hat{y}_i$. 向量 $\hat{e} = (\hat{e}_1, \ldots\ldots\ldots, \hat{e}_n)^T$被称为残差向量。显然

$$\hat{e} = Y - \hat{Y} = (I - H)Y.$$

残差向量$\hat{e}$充当了未观察到的误差向量 e的代理。

线性模型中残差最重要的事实是它们与 X的列空间正交。 这是因为 $HX = X$，所以

$$\hat{e}^T X = ((I - H)Y)^T X = Y^T(I - H)X = Y^T(X - HX) = 0。$$

因此，对于每个向量 u，$\hat{e}^T Xu = 0$，这意味着$\hat{e}$与 X的列空间正交。 X的第一列由1组成。 因为$\hat{e}$与X

的列空间中的所有内容正交
X，所以它必须与全1向量正交，这意味着

$$\sum_{i=1}^{n} \hat{e}_i = 0。$$

$\hat{e}$也与 X的每一列正交：

$$\sum_{i=1}^{n} \hat{e}_i x_{ij} = 0 \qquad \text{对于每个 } j$$

拟合值的向量属于 X的列空间，因为$\hat{Y} = X\hat{\beta}$。 因此，$\hat{e}$也与$\hat{Y}$正交。

因为 $X^T\hat{e} = 0$，残差满足 $rank(X) = p + 1$个线性等式。 因此，尽管有 n个，但它们实际上是有效的$n - p -_1$个。 因此， $n - p -_1$被称为残差的自由度。

$\hat{e}$的期望是

$$\mathbb{E}\hat{e} = \mathbb{E}((I - H)Y) = (I - H)(\mathbb{E}Y) = (I - H)X\beta = (X - HX)\beta = 0.$$

或者 $\mathbb{E}\hat{e} = \mathbb{E}(Y - \hat{Y}) = \mathbb{E}Y - \mathbb{E}\hat{Y} = 0$ 通过（1）.

$\hat{e}$的协方差矩阵是

$$Cov(\hat{e}) = Cov((I - H)Y) = (I - H)Cov(Y)(I - H) = \sigma^2(I - H). \tag{2}$$

注意残差具有不同的方差.

4 残差平方和

残差平方和的和称为RSS:

$$RSS=\sum_{i=1}^{n}\hat{e}_i^2=\hat{e}^T\hat{e}=Y^T(I-H)Y=e^T(I-H)e.$$

如果模型中没有解释变量（该模型只包含截距项），残差平方和是多少？ 这个数量被称为TSS（总平方和）。向量（$y_1-y,...,\bar{y}_n-\bar{y}$）有 $n-1$个自由度（因为这是一个大小为 n的向量，并且满足和为零的线性约束）.

简单线性回归中残差平方和是什么（当只有一个解释变量时）？ 在简单线性回归中，请检查以下内容：

$$RSS=\left(1-r^2\right)\sum_{i=1}^{n}(y_i-\bar{y})^2$$

其中 r是$(y_1,...,y_n)$和 $x=(x_1,...,x_n)$之间的样本相关性。 因为$1-r^2\le 1$，在简单线性回归中的残差平方和比没有解释变量的线性模型中的残差平方和要小。

一般来说，当我们向模型中添加更多的解释变量时，残差平方和会减少（或保持不变）。

5 决定系数

这更常被称为R平方。它的定义是

$$R^2:=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}.$$

这个定义的意义是什么？ 回归的一个目标是预测未来主题的响应变量的值。 为了达到这个目的，我们被给予数据 $y_1,...,y_n$和 x_{ij}对于 $i=1,...,n$和$j=1,...,p.$

假设我们被告知要预测未来主题的响应变量，而不使用任何关于解释变量的数据，即我们只能使用 $y_1,...,y_n$. 在这种情况下，很明显我们对下一个主题的预测将是$\bar{y}$. 这种预测方法在第 i个主题上的误差是$y_i-\bar{y}$，因此总误差为：

$$\text{总平方和}=\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2.$$

另一方面，如果我们被允许使用解释变量的数据，那么预测将会是

$$\hat{\beta}_0+x_1\hat{\beta}_1+\cdots+x_p\hat{\beta}_p.$$

对于第 i个主题的预测误差是残差$\hat{e}_i$，总误差是残差平方和：

$$RSS=\sum_{i=1}^{n}\hat{e}_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2.$$

因为使用解释变量总是比不使用它们更好，RSS总是小于或等于TSS（这个事实在于我们的模型中有一个截距）.

如果RSS与TSS相比非常小，这意味着解释变量在预测响应方面非常有用。 另一方面，如果RSS只比TSS稍微小一点，这意味着使用解释变量并没有带来太多的收益。 数量 R^2试图量化解释变量在预测响应方面的有用性。 它始终介于0和1之间。

1. 如果R^2很高，意味着残差平方和（R**SS**）相对于总平方和（**TSS**）要小得多，因此解释变量在预测响应方面非常有用。
2. 如果R^2很低，意味着残差平方和（R**SS**）仅比总平方和（**TSS**）稍微小一点，因此解释变量在预测响应方面不太有用。

必须注意的是，R^2是一个样本内预测准确度的度量。 换句话说，预测是在已经存在于样本中的对象上进行检查（而不是在新对象上进行检查）。 特别地，这些对象是模型拟合（或训练）的相同对象，因此通过拟合具有大量参数的模型，可以使R^2看起来非常好。

因为当向模型添加更多参数时，RSS会减小，所以当向模型添加更多参数时，R^2会增加。

2013年秋季统计学151课程（线性模型）：第七讲

阿迪蒂亚・冈图博伊纳

2013年9月19日

1 最后一节课

我们看了

1. 拟合值: $\hat{Y}=X\hat{\beta}=HY$其中 $H=X(X^TX)^{-1}X^TY$.. $\hat{Y}$是 Y在 X的列空间上的投影。

2.残差: $\hat{e}=Y-\hat{Y}=(I-H)Y.\hat{e}$与 X的列空间中的每个向量正交。
残差的自由度为 $n-p-1$。

3. 残差平方和: $RSS=\sum^n{}_{i=1}\hat{e}^2{}_i=\hat{e}^{Ti}\ \hat{e}_iY^T(I-H)Y$.. 当模型中添加更多解释变量时， RSS会减小。

4. 总平方和: $TSS=\sum^n{}_{i=1}(Y_i-\bar{Y})^2$.. 可以将其视为线性模型中的残差平方和（仅有截距项）

5. 决定系数或多重 R^2: 定义为 $1-(RSS/TSS)$.. 始终介于0和1之间. 较高的值意味着解释变量对响应变量有用较低的值意味着解释变量对响应变量无用. 当模型中添加更多解释变量时， R^2增加.

2 残差平方和的期望值

残差平方和的期望值是多少?

$$\mathbb{E}(RSS)=\mathbb{E}e^T(I-H)e=\mathbb{E}\left(\sum_{i,j}(I-H)(i,j)e_ie_j\right)=\sum_{i,j}(I-H)(i,j)(\mathbb{E}\,e_i\,e_j)$$

因为当$_i={}_j$时，$\mathbb{E}$（$e_i\,e_j$）等于0，而当$_i\neq{}_j$时，σ^2，我们得到

$$\mathbb{E}(RSS)=\sigma^2\sum_{i=1}^n(I-H)(i,i)=\sigma^2\left(n-\sum_{i=1}^n H(i,i)\right)$$

一个方阵的对角线元素之和被称为其迹，即tr（A）$=\sum{}_ia_{ii}$。因此，我们可以写成

$$\mathbb{E}(RSS)=\sigma^2(n-tr(H)).$$

关于迹的一个非常重要的事实是tr（AB）= tr（BA）。因此

$$tr(H)=tr(X(X^TX)^{-1}X^T)=tr((X^TX)^{-1}X^TX)=tr(I_{p+1})={}_{p+1}.$$

我们证明了

$$\mathbb{E}（RSS）=\sigma^2（n-p-1）。$$

因此，σ^2的无偏估计值为

$$\hat{\sigma^2} := \frac{RSS}{n-p-1}。$$

而 σ 的估计值为

$$\hat{\sigma} := \sqrt{\frac{RSS}{n-p-1}}。$$

这个$\hat{\sigma}$被称为残差标准误差。

$\hat{\beta}$的3个标准误差

我们已经看到 $\mathbb{E}\hat{\beta}=\beta$并且 $Cov（\hat{\beta}）=\sigma^2（X^TX）^{-1}$。因此，$\hat{\beta}_i$的标准误差被定义为$\hat{\sigma}$乘以（$X^TX$）$^{-1}$的第$_i$个对角线元素的平方根。标准误差给出了$\hat{\beta}\ i$作为 β_i估计值的准确性的一个概念。这些标准误差是线性模型摘要的R输出的一部分。

4个标准化或学生化残差

残差$\hat{e}_1$，...，$\hat{e}_n$具有不同的方差。实际上，因为 $Cov(\hat{e})=\sigma^2(I-H)$，我们有

$$var(\hat{e}_i)=\sigma^2(1-h_{ii})$$

其中 h_{ii} 表示 H 的第 i 个对角线元素。因为 h_{ii} 可以对不同的 i 不同，所以残差具有不同的方差。

如果我们将残差除以 $\sigma\sqrt{(1-h_{ii})}$，方差可以标准化为 1。但是因为 σ 是未知的，所以我们将其除以 $\hat{\sigma}\sqrt{(1-h_{ii})}$，并将结果称为标准化残差或学生化残差：

$$r_i=\frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}.$$

标准化残差 $r_1, ... ,r_n$在回归诊断中非常重要。关于未观察到的误差e1, . . . 的各种假设可以通过它们进行检验。, e_n可以通过它们进行检验。

5误差的正态性

到目前为止，我们所做的一切都只是在假设误差 $e_1, . . . , e_n$是不相关的，均值为零，方差为 σ^2的情况下。但是，如果我们想要对 β的线性组合进行假设检验或者构建置信区间，我们需要对误差进行分布假设。

例如，考虑测试零假设 $H_0:\beta_1=0$ 对备择假设 $H_1:\beta_1=0$的问题。如果 H_0成立，这意味着第一个解释变量在确定响应变量的期望值时（在其他解释变量存在的情况下）没有作用。

测试这个假设的一个明显方法是查看$\hat{\beta}_1$的值，然后拒绝 H_0如果 $|\hat{\beta}_1|$很大。
但是有多大？为了回答这个问题，我们需要了解在零假设 H_0下$\hat{\beta}_1$的分布。这样的研究需要对误差 $e_1, . . . , e_n$做一些分布假设。

误差的最常见假设是 $e_1, ..., e_n$独立地服从均值为零、方差为 σ^2的正态分布。 这在多元正态符号中写作 $e \sim N(0, \sigma^2 I_n)$。

6 多元正态分布

随机向量 $U=(U_1, \ldots, U_p)^T$如果U1, . . . , Up的联合密度函数为给定的参数μ和 Σ，则称为具有多元正态分布。

$$(2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(\frac{-1}{2}(u-\mu)^T\Sigma^{-1}(u-\mu)\right) \quad 对于u\in\mathbb{R}^d.$$

这里$|\Sigma|$表示Σ的行列式.

我们使用符号$U{\sim}N_p(\mu,\Sigma)$来表示U是多元正态分布，其参数为μ和Σ.

例子**6**.1. 多元正态分布的一个重要例子是当$U_1, ..., U_p$独立地服从均值为0，方差为σ^2的正态分布时. , U_p 独立地服从均值为0，方差为σ^2的正态分布. 在这种情况下，很容易证明$U=(U_1, ..., U_p)^T{\sim}N_p(0, \sigma^{2}I_p)$.

多元正态分布的最重要特性如下所述：

1. 当 $p=1$时，这就是通常的正态分布。
2. 均值和方差-协方差矩阵： $\mathbb{E}U=\mu$和 $Cov(U)=\Sigma$。
3. 通过矩阵相乘可以检查线性函数的独立性：如果两个线性函数 AU和 BU是独立的，当且仅当 $A\Sigma B^T=0$。 特别地，这意味着当且仅当Σ的（i, j）项等于0时， U 和 U_j是独立的。
4. 每个线性函数也是多元正态分布的： $a+AU \sim N(a+A\mu,\ A\Sigma A^T)$。
5. 假设 $U \sim N_p(\mu, I)$且 A是一个 $p\times p$symmetric和idempotent（symmetric意味着 $A^T=A$和idempotent意味着 $A^2=A$）矩阵。 那么（$U-\mu$）$^T A$（$U-\mu$）的卡方分布的自由度等于 A的秩。 这可以写作（$U-\mu$）$^T A$（$U-\mu$） $\sim \chi^2_{rank}(_A)$。

7 正态回归理论

我们假设 $e \sim N_n(0, \sigma^2 I_n)$。等价地， $e_1, ...$，e_n是独立的均值为0，方差为 σ^2的正态分布。

在这个假设下，我们可以计算到目前为止研究的许多数量的分布。

7.1 Y的分布

由于 $Y=X\beta+e$，我们有 $Y \sim N_n(X\beta, \sigma^2 I_n)$。

7.2 $\hat{\beta}$的分布

因为$\hat{\beta}=(X^TX)^{-1}X^TY$是$Y$的线性函数，所以它具有多元正态分布。 我们已经看到 $\mathbb{E}\hat{\beta}=\beta$和 $Cov(\hat{\beta})=\sigma^2(X^TX)^{-1}$。 因此$\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^TX)^{-1})$。

7.3 拟合值的分布

Y的帽子等于HY同时，Cov(Y的帽子)等于$Cov(HY)$等于σ的平方乘以H 因此，Y的帽子服从$N(X\beta, \sigma$的平方乘以$H)$

7.4 残差的分布

我们看到，估计的e的期望值等于0，而Cov(估计的e)等于σ的平方乘以$(I\text{-}H)$ 因此，估计的e服从$N(0, \sigma$的平方乘以$(I\text{-}H))$

7.5 残差和β的估计值的独立性

回想一下，如果U服从$N(\mu, \Sigma)$，那么AU和BU只有在$A\Sigma B^T$时才是独立的

这可以用来验证$\hat{\beta} = (X^TX)^{-1}X^TY$和 $\hat{e} = (I - H)Y$是独立的。 为了看到这一点，观察到它们都是$Y \sim N_n(X\beta, \sigma^2 I)$的线性函数。 因此，如果 $A = (X^TX)^{-1}X^TY$, $B = (I - H)$和 $\Sigma = \sigma^2 I$，那么

$$A\Sigma B^T = \sigma^2(X^TX)^{-1}X^T(I - H) = \sigma^2(X^TX)^{-1}(X^T - X^TH)$$

因为 $X^TH = (HX)^T = X^T$，我们得出结论 β 和 ˆ e 是独立的。

还要检查 Y和ˆ e 是否独立。

7.6 RSS的分布

回顾

$$RSS = \hat{e}^T\hat{e} = Y^T(I - H)Y = e^T(I - H)e。$$

所以

$$\frac{RSS}{\sigma^2} = \left(\frac{e}{\sigma}\right)^T (I - H) \left(\frac{e}{\sigma}\right).$$

因为 $e/\sigma \sim N_n(0, I)$且 $I - H$是对称且幂等的，秩为 $n - p - 1$，所以我们有

$$\frac{RSS}{\sigma^2} \sim \chi^2_{n-p-1}。$$

2013年秋季统计学151课程（线性模型）：第八讲

阿迪蒂亚・冈图博伊纳

2013年9月24日

1 正态回归理论

我们假设 $e \sim N_n(0, \sigma^2 I_n)$。等价地，$e_1, \ldots$，$e_n$是独立的正态分布，均值为0，方差为 σ^2。由于这个假设，我们可以计算以下内容：

1. Y的分布：由于 $Y = X\beta + e$，我们有 $Y \sim N_n(X\beta, \sigma^2 I_n)$。
2. $\hat{\beta}$的分布：$\hat{\beta} = (X^T X)^{-1} X^T Y \sim N_{p+1}(\beta, \sigma^2 (X^T X)^{-1})$。
3. 残差的分布：$\hat{e} = (I - H)Y$。. 我们看到 $\mathbb{E}\hat{e} = 0$，$Cov(\hat{e}) = \sigma^2 (I - H)$。
   因此，$\hat{e} \sim N_n(0, \sigma^2 (I - H))$。
4. 残差和 $\hat{\beta}$的独立性：回想一下，如果 $U \sim N_p(\mu, \Sigma)$，那么 AU和 BU是独立的，当且仅当 $A\Sigma B^T$。

   这可以用来验证$\hat{\beta} = (X^T X)^{-1} X^T Y$和 $\hat{e} = (I - H)Y$是独立的。为了看到这一点，观察到两者都是 $Y \sim N_n(X\beta, \sigma^2 I)$的线性函数。因此，如果 $A = (X^T X)^{-1} X^T Y$，$B = (I - H)$和$\Sigma = \sigma^2 I$，则

   $$A\Sigma B^T = \sigma^2 (X^T X)^{-1} X^T (I - H) = \sigma^2 (X^T X)^{-1} (X^T - X^T H)$$

   因为 $X^T H = (HX)^T = X^T$，我们得出结论 β 和 $\hat{}$ e 是独立的。
   还要检查 Y和$\hat{}$ e 是否独立。
5. **RSS的分布**: $RSS = \hat{e}^T \hat{e} = Y^T (I - H) Y = e^T (I - H) e$。所以

   $$\frac{RSS}{\sigma^2} = \left(\frac{e}{\sigma}\right)^T (I - H) \left(\frac{e}{\sigma}\right).$$

   因为 $e/\sigma \sim N_n(0, I)$且 $I - H$是对称且幂等的，秩为 $n - p - 1$，所以我们有

   $$\frac{RSS}{\sigma^2} \sim \chi^2_{n-p-1}。$$

2 如何测试 $H_0 : \beta_j = 0$

有两种等价的测试这个假设的方法。

2.1 第一次测试： t-测试

根据$\hat{\beta}_j$的值进行测试是很自然的，即如果 $|\hat{\beta}_j|$很大，则拒绝。那么多大才算很大呢？ 为了回答这个问题，我们需要看一下$\hat{\beta}_j$在零假设下的分布（称为零分布）。 在误差正态分布的情况下，我们已经知道$\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^TX)^{-1})$。换句话说，

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 v_j)$$

其中 v_j是$(X^TX)^{-1}$的第j个对角线元素。 在零假设下，当 $\beta_j = 0$，我们有

$$\frac{\hat{\beta}_j}{\sigma\sqrt{v_j}} \sim N(0, 1).$$

这可以用来构建一个测试，但问题是 σ是未知的。 因此，用估计值$\hat{\sigma}$替换它来构建测试统计量：

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}} = \frac{\hat{\beta}_j/\sigma\sqrt{v_j}}{\hat{\sigma}/\sigma} = \frac{\hat{\beta}_j/\sigma\sqrt{v_j}}{\sqrt{RSS/(n-p-1)\sigma^2}}$$

现在这里的分子是 $N(0,1)$。分母是 $\sqrt{}$此外，分子和分母是独立的。因此，我们得到

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} \sim t_{n-p-1}$$

其中 t_{n-p-1}表示自由度为 t的t-分布。

用于检验 $H_0 : \beta_j = 0$ 的p值可以得到

$$\mathbb{P}\left(|t_{n-p-1}| > \left|\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)}\right|\right).$$

请注意，当 $n-p-1$ 很大时， t-分布几乎与标准正态分布相同。

2.2 第二次测试： F检验

我们刚刚看到如何使用统计量 $\hat{\beta}_j/s.e(\hat{\beta}_j)$ 和 t-
分布来检验假设H0 :βj = 0。

这里有另一个自然的测试方法。 零假设 H_0 表示解释变量 x_j可以从线性模型中删除。 让我们称这个简化模型为 m。同时，让我们称原始模型为 M（这是完整模型： $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i$）。

以下提供了另一个对 H_0进行自然检验的方法。 在模型中，将残差平方和表示为
将模型中的残差平方和表示为 $RSS(m)$，将完全模型中的残差平方和表示为 $RSS(M)$。 始终成立的是$RSS(M) \leq RSS(m)$。 现在，如果 $RSS(M)$比 $RSS(m)$要小得多，这意味着解释变量 x_j对回归有很大的贡献，因此不能被剔除，即我们拒绝零假设 H_0。 另一方面，如果 $RSS(M)$只比 $RSS(m)$稍微小一点，那么 x_j在预测 y方面并没有真正做出很大的贡献，因此可以被剔除，即我们不拒绝 H_0。因此，可以通过检验统计量来测试 H_0：

$$RSS(m) - RSS(M)$$

如果这个值很大，我们将拒绝零假设。有多大？ 为了回答这个问题，我们需要看一下
零分布 RSS（m） $-RSS$（M） 我们将在下一堂课中展示

$$\frac{RSS(m)-RSS(M)}{\sigma^2} \sim \chi_1^2$$

在零假设下 由于我们不知道 σ^2，我们通过估计来得到

$$\hat{\sigma}^2 = \frac{RSS(M)}{n-p-1}$$

，得到检验统计量：

$$\frac{RSS(m)-RSS(M)}{RSS(M)/(n-p-1)}$$

分子和分母是独立的（将在下一堂课中展示） 如果分母是 RSS（m）/（$n-p$），则这种独立性不成立
因此，在零假设下

$$\frac{RSS(m)-RSS(M)}{RSS（M）/（n-p-1）} \sim F_{1,n-}{}^{p}$$

因此可以得到p-值

$$\mathbb{P}\left(F_{1,n-p-1} > \frac{RSS(m)-RSS(M)}{RSS(M)/(n-p-1)}\right).$$

这两个检验的等价性

事实证明，这两个检验用于测试 $H_0 : \beta_j= 0$ 在某种意义上是等价的，因为它们给出相同的p-值。这是因为

$$\left(\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)}\right)^2 = \frac{RSS(m)-RSS(M)}{RSS(M)/(n-p-1)}$$

这并不难证明，但我们将跳过它的证明。

2013年秋季统计学151课程（线性模型）：第九讲

阿迪蒂亚・冈图博伊纳

2013年9月26日

用于比较模型的假设检验

设 M表示完整的回归模型：

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i$$

其中有 p个解释变量。

让 m表示 M的子模型，该子模型通过对参数 $\beta =$ $(\beta_0, \ldots\ldots, \beta_p)$施加线性约束来获得。例如：

1. 对于约束 $\beta_1 = 0$，模型 m变为： $y_i = \beta_0 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$。
2. 对于 $\beta_1 = \beta_2 = 0$，模型 m变为： $y_i = \beta_0 + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + e_i$。
3. 对于 $\beta_1 = \cdots = \beta_p = 0$，模型 m变为： $y_i = \beta_0 + e_i$。
4. 对于$\beta_1 = \beta_2$，模型m变为$y_i = \beta_0 + \beta_1(x_{i1} + x_{i2}) + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + e_i$.
5. 对于$\beta_1 = 3$，模型m变为$y_i = \beta_0 + 3x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$.

我们如何测试假设H_0：m对立H_1：M？ 设q为解释变量的数量在m中。 这个测试可以通过注意到如果$RSS(m)$ - $RSS(M)$很大，那么m不是数据的一个好模型，因此我们拒绝H_0来进行。 另一方面，如果$RSS(m)$ - $RSS(M)$很小，我们不拒绝H_0。

因此，测试基于 RSS（m） $- RSS$（M）。 在零假设下，这个数量的分布是什么？我们将展示

$$\frac{RSS(m) - RSS(M)}{\sigma^2} \sim \chi^2_{p-q}$$

其中 q是 m中解释变量的数量， p是 M中解释变量的数量。
现在让我们证明这个事实。 让两个模型中的帽矩阵分别表示为 H（m）和 H（M）
写成

$$RSS(m) = Y^T(I - H(m))Y \quad \text{和} \quad RSS(M) = Y^T(I - H(M))Y$$

所以

$$RSS（m） - RSS（M） = Y^T（H（M） - H（m））Y。$$

我们需要 RSS（m） $- RSS$（M）的零分布。 因此，我们假设 $Y = X$（m）β（m）$+ e$（其中 X（m）是模型中的X矩阵（m）。重要的是要意识到H（m）X（m）$=X$（m）并且也 H（M）X（m）$=X$（m）。所以

$$RSS（m）-RSS（M）=e^T（H（M）-H（m））e。$$

H（M）-H（m）是一个对称的$n\times n$矩阵，秩为$p-q$。它也是幂等的，因为（H（M）-H（m））^^{2}=H（M）+H（m）-^2H（M）H（m）。 为了证明这一点，只需证明H（M）H（m）v=H（m）v对于每个向量v成立。现在回想一下H（m）v表示v在X（m）的列空间上的投影。并且
H（M）H（m）v项目 H（m）v项目投影到 X（M）的列空间上（等于原始 X矩阵）。但是因为 X（m）的列空间包含在 X（M）的列空间中，所以它遵循 H（m）v已经包含在 X（M）的列空间中。因此，它在 X（M）的列空间上的投影等于它本身。所以 H（M）H（m）$v=H$（m）v。因为 H（M）H（m）$=H$（m），根据（1）可知（H（M）$-H$（m））$^2=H$（M）$-H$（m）。因此，

在零假设下RSS（m）$-RSS$（M）

$$\frac{\qquad\qquad\qquad}{\sigma^2}=\left(\frac{e}{\sigma}\right)^T(H(M)-H(m))\frac{e}{\sigma}\sim\chi^2_{p-q}.$$

由于我们不知道 σ^2，我们通过估计来得到

$$\hat{\sigma}^2=\frac{RSS(M)}{n-p-1}$$

，得到检验统计量：

$$\frac{RSS(m)-RSS(M)}{RSS(M)/(n-p-1)}$$

分子和分母是独立的，因为 $RSS(m)-RSS(M)=\ Y^T(H(M)-H(m))Y$ 和 $RSS(M)=Y^T(I-H(M))Y$ 以及矩阵的乘积

$$(H(M)-H(m))\,(I-H(M))=H(M)-H(M)^2-H(m)+H(m)H(M)=H(M)-H(M)-H(m)+H(m)=0.$$

因此在零假设下

$$\frac{(\text{残差平方和}(m)-\text{残差平方和}(M))/(p-q)}{RSS(M)/(n-p-1)}\sim F_{p-q,n-p-1}.$$

因此可以得到p-值

$$\mathbb{P}\left(F_{p-q,n-p-1}>\frac{(\text{残差平方和}(m)-\text{残差平方和}(M))/(p-q)}{RSS(M)/(n-p-1)}\right).$$

如果零假设可以用一个关于 β的线性函数来表示，例如 $H_0:\beta_1+5\beta_3=5.$
那么也可以通过 t检验来进行测试；使用统计量：

$$\frac{\hat{\beta}_1+5\hat{\beta}_3-5}{s.e(\hat{\beta}_1+5\hat{\beta}_3)}$$

该统计量在 H_0下服从自由度为 $n-p-1$的 t分布。这个检验和相应的
F检验将具有相同的 p值。

1.1 检验所有解释变量

我们如何检验 $H_0:\beta_1=\cdots=\beta_p=0$ 的补集？只需将 m作为模型 $y_i=\beta_0+e_i$。
在这种情况下，$RSS(m)=TSS$和 $q=0$，$RSS(M)=RSS$。因此，p值为

$$\mathbb{P}\left\{F_{p,n-p-1}>\frac{(TSS-RSS)/p}{RSS/(n-p-1)}\right\}.$$

2013年春季统计学153课程（时间序列）：第十讲

阿迪蒂亚・冈图博伊纳

2013年2月21日

1 最佳线性预测

假设 Y 和 $W_1, \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots, W_m$ 是具有零均值和有限方差的随机变量。令 $\text{cov}(Y, W_i) = \zeta_i, i = 1, \ldots, m$ 和

$$\text{cov}(W_i, W_j) = \Delta(i, j) \qquad \text{对于} i, j = 1, \ldots, m.$$

在 $W_1, \ldots \ldots \ldots \ldots, W_m$方面，什么是最佳的线性预测器 Y?

最佳的线性预测器 $a_1, \ldots, a_m$的特点是 $Y - a_1W_1 - \cdots - a_mW_m$与 $W_1, \ldots, W_m$不相关。换句话说：$\text{cov}(Y - a_1W_1 - \cdots - a_mW_m, W_i) = 0$

$$\text{对于 } i = 1, \ldots, m.$$

注意，这给出了 m个未知数 $a_1, \ldots \ldots \ldots \ldots \ldots \ldots \ldots, a\ m$的$_m$个方程。第 i个方程可以重写为

$$\zeta_i - \Delta(i, 1)a_1 - \cdots - \Delta(i, m)a_m = 0.$$

换句话说，这意味着 $\zeta_i$equals the ith row of Δ multiplied by the vector $a = (a_1, \ldots \ldots \ldots \ldots, a_m)^T$ which is same as the ith element of the vector Δa. 因此，这些 m方程可以写成一行 $\Delta a = \zeta$.

得到最佳线性预测器系数的定义方程的另一种方法是找到 $a_1, \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots, a_m$的值，使其最小化

$$\begin{aligned}
F(\mathbf{a}) &:= \mathbb{E}\left(Y - a_1W_1 - \cdots - a_mW_m\right)^2 \\
&= \mathbb{E}\left(Y - a^TW\right)^2 \\
&= \mathbb{E}Y^2 - 2\mathbb{E}((a^TW)Y) + \mathbb{E}(a^TW\,W^Ta) \\
&= \mathbb{E}Y^2 - 2a^T\zeta + a^T\Delta a.
\end{aligned}$$

对 a进行微分并置零得到

$$-2\zeta + 2\Delta a = 0$$

因此，以 W^1为条件的 Y的最佳线性预测器为$\zeta^T\Delta^-{}_1W$。

这种情况在 $m = 1$（只有一个预测变量 W_1）时可能更为熟悉。当 $m = 1$时，我们有$\zeta 1 = \text{cov}(Y, W1)$ 和 $\Delta(1,1) = \text{var}(W1)$。因此，以$W_1$为条件的 Y的最佳预测器是

$$\frac{\text{cov}(Y, W_1)}{\text{var}(W_1)}W_1.$$

现在考虑一个平稳的均值为零的时间序列 $\{X_t\}$. 利用上述结果，取 $Y = X_n$和 $W_1 = X_{n-1}$，我们可以得到 X_n关于 X_{n-1}的最佳预测器

$$\frac{\text{cov}(X_n, X_{n-1})}{\text{var}(X_{n-1})} X_{n-1} = \frac{\gamma_X(1)}{\gamma_X(0)} X_{n-1} = \rho_X(1) X_{n-1}$$

关于 $X_{n-1}, X_{n-2}, \ldots\ldots\ldots\ldots\ldots\ldots,\ X_{n-k}$，什么是最佳预测器？ 在这里，我们取 $Y = X_n$ 和 $W_i = X_{n-i}$，对于 $i= 1, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots., k$. 因此

$$\Delta(i, j) = \text{cov}(W_i, W_j) = \text{cov}(X_{n-i}, X_{n-j}) = \gamma_X(i-j)$$

和

$$\zeta_i = \text{cov}(Y, W_i) = \text{cov}(X_n, X_{n-i}) = \gamma_X(i).$$

通过这些Δ和 ζ, 解出$\Delta a = \zeta$以获得 $X_{n-1}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$ 的系数 , X_{n-k}在最佳线性预测器中。

考虑AR(p)模型的特殊情况： $X_t - \phi_1 X_t$ $_{-}$根据定义方程和因果关系，可以直接得出$X_{n-}\ \phi_1\ X_n$ $_{-1} - \cdots\cdots\cdots - \phi_p X_{n-p}$与Xn不相关因此我们推断出，以$X_{n-1}$， X_{n-2}，. . .为基础的 X_n的最佳线性预测器等于 $\phi_1 X_{n-1} + \phi_2 X_{n-2} + \cdots + \phi_p X_{n-p}$.

2 偏自相关函数（pacf）

2.1 第一定义

设 $\{X_t\}$为均值为零的平稳过程。 在滞后 h处的偏自相关，表示为 $pacf(h)$，被定义为 X_t的最佳线性预测器中 X_{t-h}的系数，其中 X_{t-1}，. . . , X_{t-h}也包含在内。

检查一下自相关系数在滞后一期时是否与自相关函数$\text{pac}f(1)$相同。 但是对于$h > 1$的情况，$\text{pac}f(h)$可能与$\rho(h)$相差很大。

对于AR(p)模型：检查一下$\text{pac}f(p)$是否等于ϕp，并且对于$h > p$，$\text{pac}f(h)$是否等于0。

2.2 第二定义

从第一定义中，为什么这被称为相关性还不太清楚。 这将在第二定义中变得明显。

滞后h的偏自相关函数定义为X_t与X_t-h之间的相关性，其中中间变量X_t-$_1$，X_t-$_2$，...，X_t-h_{+1}的影响被移除。让$\beta_1 X_t$ $_{-1} + \cdots + \beta_{h-1} X_{t-h+1}$ 表示X_t 的最佳线性预测器，以$X_{t-1}, \ldots, X_{t-h+1}$ 为条件。 根据平稳性，这两个序列

$$X_t, X_{t-1}, \ldots, X_{t-h+1}$$

和

$$X_{t-h}, X_{t-h+1}, \ldots, X_{t-1}$$

具有相同的协方差矩阵。 当 $i \neq j$ 时，W_i 和 W_j 之间的协方差等于 $\gamma X(i-j)$，这与 $\tilde{W}_i$ 和 $\tilde{W}_j$ 之间的协方差相同。

因此，最佳线性预测 X_{t-h}的 X_t−h的线性预测 $_{-h+1}, \ldots, X_{t-1}$ 等于 $\beta_1 X_{t-h+1} + \cdots + \beta_{h-1} X_{t-1}$.

滞后 h的偏自相关函数定义为

$$pacf(h) = \operatorname{corr}\left(X_t - \beta_1 X_{t-1} - \cdots - \beta_{h-1} X_{t-h+1}, X_{t-h} - \beta_1 X_{t-h+1} - \cdots - \beta_{h-1} X_{t-1}\right).$$

换句话说， $pacf(h)$是 X_t
和 X_{t-h}之间最佳线性预测的误差的相关性，其中涉及的变量是 $X_{t-1}, \ldots\ldots\ldots\ldots, X_{t-h+1}$.

关键事实是对于AR(p)模型，当滞后$h > p$时，自相关函数pacf(h)等于零。 为了证明这一点：注意到当$h > p$时，以X_t-$_1, \ldots$为基础的X_t的最佳线性预测器， 即X_t-h_{+1}等于$\phi_1 X_t$-$_1$+$\phi_2 X_t$-$_2$+$\cdots$+$\phi p X_t$-p.
换句话说， $\beta_1 = \phi_1, \ldots\ldots\ldots$ $\beta_p = \phi_p$且 $\beta_i = 0$对于$_i > _p$.

因此对于h $> p$，我们有

$$_pacf(h) = \operatorname{corr}(X_t\text{-}\phi_1 X_t\text{-}_1\text{-}\cdots\text{-}\phi_p X_t\text{-}_p, X_t\text{-}h\text{-}\phi_1 X_t\text{-}h_{+1}\text{-}\cdots\text{-}\phi_p X_t\text{-}h_{+p}) = \operatorname{corr}(Z_t, X$$
$$_t\text{-}h\text{-}\phi_1 X_t\text{-}h_{+1}\text{-}\cdots\text{-}\phi_p X_t\text{-}h_{+p}) = 0,$$

由因果关系引起。

通过线性代数可以证明两个定义之间的等价性 $pacf(h)$。 我们将跳过这个推导。

3 从数据中估计 $pacf$

如何根据不同的滞后h从数据中估计 $pacf(\,h)$？ 系数 $a_1, \ldots\ldots\ldots\ldots\ldots\ldots, a_h$ of $X_{t-1}, \ldots$， X_{t-h}
在 X_t的最佳线性预测器中通过解一个形式为$\Delta a = \zeta$的方程获得。

现在，Δ和ζ的所有元素都是形如γ_X（i-j）的形式，其中i和j是某些值。 因此，一种自然的估计pacf（h）的方法是通过相应的样本自相关来估计Δ和ζ的条目，得到$\Delta\hat{}$和$\zeta\hat{}$，然后解方程$\Delta\hat{}\hat{}a$=$\zeta\hat{}$来求解$\hat{}a$。 请注意，pacf（h）恰好是a的h阶。

已经证明，当数据来自AR(p)模型时，滞后大于p的样本偏自相关在独立正态分布中近似地具有零均值和方差$1/n$。 因此，对于h $> p$，可以使用$\pm 1.96\sqrt{}(n\text{-}1)/^2$的区间来检验AR$(p)$模型是否适用。

4 总结

对于$MA(q)$模型，自相关函数$\rho_X(h)$在$h > q$时等于零。 同样对于$h > q$，样本自相关函数$r(h)$近似服从均值为0，方差为$w(hh)$/n的正态分布，其中$w(hh) := 1 + {}^2\rho^2(1) + \ldots + {}^2\rho^2(q)$。

对于一个$AR(p)$模型，偏自相关函数pacf(h)在$h > p$时等于零。 同样对于$h > p$，样本自相关函数$r(h)$近似服从均值为0，方差为1的正态分布。

如果一个数据集的样本自相关函数在某个滞后阶数截尾，我们使用MA模型。 如果样本偏自相关函数在某个滞后阶数截尾，我们使用AR模型。

如果以上两种情况都不发生怎么办？ 那么我们如何选择一个合适的ARMA模型？ 下面是一个一般的策略：

1. 尝试不同的p和q选择ARMA(p,q)模型。
2. 对于固定的p和q，将ARMA(p,q)模型拟合到数据中（我们很快会学习如何做到这一点）。

3. 检查拟合效果如何。选择p和q使得拟合效果好，同时确保没有过拟合。

如何检查模型是否很好地拟合了数据但没有过拟合？ 这是一个模型选择的问题。通常使用AIC、FPE、BIC等自动准则。 还应该运用判断力。

我们的计划如下：

1. 如何将ARMA模型拟合到数据中？
2. 如何评估拟合的好坏？
3. 通过自动模型选择技术选择 p和 q。

2013年秋季统计学151课程（线性模型）：第十一讲

阿迪蒂亚·冈图博伊纳

2013年10月03日

1 单因素方差分析

考虑模型

$$y_{ij} = \mu_i + e_{ij} \qquad \text{对于 } i = 1, \ldots, t \text{和 } j = 1, \ldots, n_i$$

其中 e_{ij}是均值为零方差为 σ^2的独立同分布正态随机变量，令 $\sum^t{}_{i=1} n_i = n.$

这个模型用于以下几种情况：

1. 有 t个处理和 n个对象。 每个对象只接受一个 j处理。
   $y_{i1}, \ldots, y_{in_i}$ 表示接受 $i$th处理的对象的分数。
2. 我们研究的是可以自然分为 t个组的t个对象的某种表现。 我们希望看到对象之间的表现差异是否可以通过这些不同的组来解释。 $y_{i1}, \ldots, y_{in}$ $_i$ 表示 $i$th组中对象的表现。

通常这个模型也可以写成

$$y_{ij} = \mu + \tau_i + e_{ij} \qquad \text{对于 } i = 1, \ldots, t \text{和 } j = 1, \ldots, n_i \qquad (_1)$$

其中，μ被称为基准分数，τ_i是第$_i$个处理的平均分数与基准分数之间的差异。 在这个模型中，μ和个体τ_i是不可估计的。 很容易证明，如果且仅如果$\lambda = \sum^{ti}{=}_1\lambda_i$，那么参数$\lambda\mu + \sum^{ti}{=}_1\lambda_i\tau_i$是可估计的。 由于缺乏可估计性，人们经常强加条件$\sum^{ti}{=}_1\tau_i{=}0$。这个条件确保了所有参数$\mu$和$\tau_1$，...，$\tau^t$都是可估计的。 此外，它提供了一个很好的解释。$\mu$表示基准响应值，$\tau_i$是响应值需要从基准$\mu$调整的值，用于组$_i$。因为$\sum$

当$_i = 0$时，一些调整将是正数，一些调整将是负数，但是所有组的整体调整平均值为零。

在这个模型中，如何测试假设 $H_0 : \mu_1 = \cdots = \mu_t$? 这只是一个线性模型，因此我们可以使用 F检验。 我们只需要找到完整模型（M）中的残差平方和（RSS），以及简化模型（m）中的残差平方和（RSS）。 完整模型中的残差平方和是多少？ 让 $\bar{y}_i = \sum^n{}_{j=1i}\ y_{ij}/n$和 $\bar{y} = \sum^t{}_{i=1}$ $\sum^{n_i}_{j=1} y_{ij}/n$。写

$$\begin{aligned}
\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 &= \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i+\bar{y}_i-\mu_i)^2 \\
&= \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 + 2\sum_{i=1}^{t}(\bar{y}_i-\mu_i)\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i) + \sum_{i=1}^{t}n_i(\bar{y}_i-\mu_i)^2 \\
&= \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 + \sum_{i=1}^{t}n_i(\bar{y}_i-\mu_i)^2.
\end{aligned}$$

因此， μ_i的最小二乘估计值为 $\hat{\mu}_i= \bar{y}_i$。 如果我们将 $\tau_i=\mu_i-\mu$写成，那么 μ的最小二乘估计值为 $\bar{y}$，其中$\sum$， τ_i的最小二乘估计值为 $\bar{y}_i - \bar{y}$。

全模型中的残差平方和为

$$RSS(M) = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2。$$

检查在简化模型中的残差平方和

$$残差平方和（m） = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{t} n_i (\bar{y}_i - \bar{y})^2 .$$

因此，用于检验 H_0的 F统计量为 $\mu_1 =\cdots = \mu_t$的

$$T = \frac{\sum_{i=1}^{t} n_i(\bar{y}_i - \bar{y})^2/(t-1)}{\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2/(n-t)}$$

该统计量在 H_0下服从 F分布，自由度为 $t-1$ 和 $n-t$。

2个排列检验

到目前为止，我们已经通过 F检验研究了线性模型中的假设检验。假设我们想要对完全线性模型（记为 M）中的 $\beta=(\beta_0,...,\beta_p)^T$进行线性假设检验。 我们首先构建一个包含完全模型 M中假设的简化模型。 将这个简化模型称为 m。 然后我们看一下这个量：$T:=$

$$\frac{(残差平方和(m) - 残差平方和(M))/(p-q)}{RSS(M)/(n-p-1)}.$$

如果 T很大，拒绝零假设是有意义的。 回答问题：大到什么程度才算大？因此，我们依赖于误差正态性的假设，即 $e \sim N(0,\sigma^2 I)$，来断定 $T \sim F_{p-q,n-p-1}$在 H_0下。 因此，可以通过计算 $\mathbb{P}\{F_{p-q,n-p-1} > T\}$来得到一个 p值。

假设我们不想假设误差正态性。 有没有办法得到一个 p值？ 在某些情况下，通过置换检验是可能的。 下面我们提供两个例子。

2.1 测试所有解释变量

我们想要测试的零假设是，所有解释变量可以在不假设 $e \sim N(0,\sigma^2 I)$的情况下被丢弃。 在零假设下，我们假设如果响应变量 y与解释变量没有关系。 因此，可以合理地假设在零假设下，响应变量 $y_1,...$，y_n在受试者之间是随机分布的，与预测变量无关。 这激发了以下测试：

1. 随机排列响应值： $y_1, \ldots\ldots\ldots, y_n$.
2. 计算数量
$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$
响应值是上一步中排列的值。
3. 重复上述两个步骤多次。

4. 这将产生大量的检验统计量值（每个响应值的排列都有一个）。 我们称它们为 $T_1, ..., T_N$. 计算 p 值作为 $T_1, ..., T_N$中超过原始检验统计量值 T （ T是使用实际未排列的响应值 $y_1, ..., y_n$计算的）。

这个测试的想法如下：从给定的数据中，我们计算出

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

我们需要知道在零假设下这个值有多极端。 在正态性的假设下，我们可以通过 F-分布来评估这个值。 但是我们需要在不假设正态性的情况下进行。 为此，我们尝试在零假设下生成这个数量的值。 这个想法是通过对响应值进行排列后计算统计量来实现的。 因为一旦响应值被排列，响应和解释变量之间的所有关联都会断开，使得

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

对于排列后的响应值，它们的值类似于在零假设下生成的值。 然后，P-值被计算为大于观察值的这些值的比例。

2.2 对单个解释变量进行测试

我们如何测试，比如说，第一个解释变量是否有用？ 我们计算 t-统计量：

$$\frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)}$$

并通过与 $t_{n-\ p-1}$分布进行比较来计算$_p$值（这需要正态性）。 如何在没有正态性的情况下进行?

我们可以通过对 x_1的值进行排列测试来进行跟踪。 对于每个排列，我们计算 t统计量，并且p值是这些t值中大于观察到的t值的比例的绝对值。

2013年春季统计学153课程（时间序列）：第十二讲

阿迪蒂亚・冈图博伊纳

2013年3月5日

1 计划

到目前为止：

1. 趋势和季节性
2. 稳定性
3. ARMA模型

时间域技术中即将出现的内容：

1. 如何将ARMA模型拟合到数据中。
2. ARIMA模型
3. SARIMA模型
4. 预测
5. 模型诊断和选择

2 复习：将AR模型拟合到数据中

假设已知顺序 p。通过调用R中的函数 $ar()$来执行。

1. **Yule Walker**或矩法：找到AR(p)模型，其自相关函数等于样本自相关函数在滞后0，1，...，p处。在R中使用 yw方法。
2. 条件最小二乘：最小化条件平方和： $\sum^n{}_{i=p+1}(x_i-\mu-\phi_1(x_{i-1}-\mu)-\cdots-\phi_p(x_{i-p}-\mu))^2$ over μ和 ϕ_1，...，ϕ_p。. 而 σ^2通过平均残差得到。 在R中使用**ols**方法。在该方法中，给定数据 x_1，...，x_n，R将数据拟合成模型 $x_t-\bar{x}=intercept+\phi(x_{t-1}-\bar{x})+residual$。 可以通过调用$\$x.intercept$获得拟合的截距值。 可以将其转换为形式为 $x_t=intercept+\phi x_{t-1}+residual$的模型。 查看R函数**ar.ols**的帮助页面。
3. 最大似然：最大化似然函数（写下来相对简单，但需要优化算法来最大化）。 在R中使用 mle方法。这个方法很复杂。

通常情况下，这三种方法得到的答案是相似的。 R中的默认方法是Yule-Walker。

3 自回归模型的Yule-Walker估计的渐近分布

对于大的n，$\sqrt{n}\left(\hat{\phi}-\phi\right)$的近似分布是均值为0，方差协方差矩阵为 $\sigma_Z^2\Gamma_p^{-1}$的正态分布，其中Γ_p是 $\gamma_X(i-j)$的(i,j)th元素。

例如，在AR(1)情况下：

$$\Gamma_p = \Gamma_1 = \gamma_X(0) = \sigma_Z^2/(1-\phi^2)。$$

因此$\hat{\phi}$近似服从均值为 ϕ和方差为$(1-\phi^2)/n$的正态分布。

对于AR(2)，使用

$$\gamma_X(0) = \frac{1-\phi_2}{1+\phi_2}\frac{\sigma_Z^2}{(1-\phi_2)^2-\phi_1^2} \quad 和 \quad \rho_X(1) = \frac{\phi_1}{1-\phi_2},$$

我们可以证明$(\hat{\phi}_1,\hat{\phi}_2)$的均值近似为$(\phi_1,\phi_2)$，方差-协方差矩阵为
$1/n$ 倍

$$\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}$$

2013年春季统计学153课程（时间序列）：第十三讲

阿迪蒂亚·冈图博伊纳

2013年3月7日

1 AR模型估计的渐近分布

对于Yule-Walker、条件最小二乘和最大似然估计，对于大的 n，$\sqrt{n}$的近似分布为

$-\left(\hat{\phi}-\phi\right)$ 的近似分布是均值为0，方差协方差矩阵为 σ

$^2_Z\Gamma_p^{-1}$的正态分布，其中Γ_p是 $\gamma_X(i-j)$的(i,j)th元素。

1.1 证明概要

为了简单起见，假设 $\mu=0$。 最容易使用条件最小二乘估计。AR(p)模型如下：

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t。$$

我们可以将这个模型写成矩阵表示形式：

$$X_t = \mathbb{X}_{t-1}^T\phi + Z_t$$

其中 $\mathbb{X}_{t-1}$是一个 $p\times 1$向量 $\mathbb{X}_{t-1}=(X_{t-1}, X_{t-2}, \ldots\ldots, X_{t-p})^T$， ϕ是一个 $p\times 1$向量 $(\phi_1, \ldots, \phi_p)^T$。
条件最小二乘法最小化平方和：

$$\sum_{t=p+1}^{n}\left(X_t - \phi^T\mathbb{X}_{t-1}\right)^2$$

关于 ϕ的解是：

$$\hat{\phi} = \left(\sum_{t=p+1}^{n}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^T\right)^{-1}\left(\sum_{t=p+1}^{n}\mathbb{X}_{t-1}X_t\right).$$

写成 $X_t = \mathbb{X}^{Tt}{}_{-1}\phi + Z_t$，我们得到

$$\hat{\phi} = \phi + \left(\sum_{t=p+1}^{n}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^T\right)^{-1}\left(\sum_{t=p+1}^{n}\mathbb{X}_{t-1}Z_t\right).$$

因此,

$$\sqrt{n}\left(\hat{\phi}-\phi\right) = \left(\frac{1}{n}\sum_{t=p+1}^{n}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^T\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\mathbb{X}_{t-1}Z_t\right). \quad (1)$$

以下断言是直观的（注意 $\mathbb{X}_{t-1}$和 Z_t在高斯假设下是不相关的，因此是独立的），并且可以严格证明：

$$\frac{1}{n}\sum_{t=p+1}^{n}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^{T}\to\Gamma_p \qquad \text{当 } n\to\infty\text{时，以概率收敛}$$

和

$$\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\mathbb{X}_{t-1}Z_t\to N\left(0,\sigma_Z^2\Gamma_p\right) \qquad \text{当 } n\to\infty\text{时，以分布收敛。}$$

这些结果可以与表达式（1）结合起来证明 $\sqrt{n}$ $^{-}\left(\hat{\phi}-\phi\right)$ 收敛于分布
to一个均值为0和方差协方差矩阵 $\sigma_Z^2\,\Gamma_p^{-1}$的正态分布。

1.2特殊实例

在AR(1)情况下：

$$\Gamma_p=\Gamma_1=\gamma_X(0)=\sigma_Z^2/(1-\phi^2)\text{。}$$

因此$\hat{\phi}$近似服从均值为 ϕ和方差为$(1-\phi^2)/n$的正态分布。

对于AR(2)，使用

$$\gamma_X(0)=\frac{1-\phi_2}{1+\phi_2}\frac{\sigma_Z^2}{(1-\phi_2)^2-\phi_1^2} \qquad \text{和} \qquad \rho_X(1)=\frac{\phi_1}{1-\phi_2},$$

我们可以证明$(\hat{\phi}_1,\hat{\phi}_2)$的均值近似为$(\phi_1,\phi_2)$，方差-协方差矩阵为
$1/n$ 倍

$$\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}$$

注意到$\hat{\phi}_1$和$\hat{\phi}_2$的近似方差是相同的。 观察到如果我们将AR(2)模型拟合到来自AR(1)的数据集中，那么 $\hat{\phi}_1$的估计可能不会改变太多，但标准误差会更高。我们失去了精度。 参见书中的例子3.34。

2更一般的ARMA模型拟合

2.1可逆性

考虑MA(1)模型的情况，其自相关函数给出 $\gamma_X(0)=\sigma_Z^2(1+\theta^2)$和 $\gamma_X(1)=\theta\sigma_Z^2$和 $\gamma_X(h)=0$对于所有 $h\geq^2$。 很容易看出，对于 $\theta=5$，$\sigma_Z{}^2=1$，我们得到与$\theta=1/5$，$\sigma_Z{}^2={}^25$相同的自相关函数。 换句话说，存在不同的参数值给出相同的自相关函数。 更一般地，参数对(θ,σ_Z^2)和$(1/\theta,\theta^2\sigma_Z^2)$对应于相同的自相关函数。

这意味着不能从数据中唯一地估计MA(1)模型的参数。 一个自然的修正是只考虑那些满足$|\theta|<1$的MA(1)。 这个条件被称为可逆性。 对于MA(1)模型，条件$|\theta|<1$等价于移动平均多项式 $\theta(z)$的所有根的模大于1。这给出了ARMA过程的可逆性的一般定义。

如果移动平均多项式 $\theta(z)$的所有根的模大于1，那么ARMA模型 $\phi(B)(X_t-\mu)=\theta(B)Z_t$被称为可逆的。可以证明（类似于因果性）

这个条件等价于 Z_t被写成 X_t的当前和过去值的线性组合。

从现在开始，我们只考虑平稳、因果和可逆的ARMA模型，即我们假设多项式 $\phi(z)$和 $\theta(z)$在单位圆盘内没有任何根。

现在让我们来研究将一个平稳、因果和可逆的ARMA模型拟合到数据的问题，假设阶数 p和 q已知。

对于AR模型拟合的三种方法都可以推广到（带有额外复杂性的）一般ARMA情况。 最简单的方法是先学习相关的R函数。 要使用的函数是$arima()$。 我们将在后面看到ARIMA是一类更一般的模型，包括ARMA模型作为一种特殊情况（实际上，ARIMA只是差分+ARMA）。这个函数 $arima()$可以用来拟合数据的ARMA模型。 它还有一个方法参数，有三个取值：CSS-ML、ML和CSS，默认为CSS-ML。

2.2 Yule-Walker或矩估计法

原则上，这通过解以下一组方程的某个子集来进行，以求解未知参数
$\theta_1, \ldots \ldots \ldots \ldots, \theta_q, \phi_1, \ldots$， ϕ_p和 σ_Z^2 （而 μ由样本均值估计）

$$\hat{\gamma}(k) - \phi_1\hat{\gamma}(k-1) - \cdots - \phi_p\hat{\gamma}(k-p) = (\psi_0\theta_k + \psi_1\theta_{k+1} + \cdots + \psi_{q-k}\theta_q)\sigma_Z^2$$

对于$0 \le k \le q$和

$$\hat{\gamma}(k) - \phi_1\hat{\gamma}(k-1) - \cdots - \phi_p\hat{\gamma}(k-p) = 0 \qquad \text{对于} k > q。$$

注意到 ψ_j上面是 $\theta_1, \ldots \ldots, \theta_q$和 $\phi_1, \ldots, \phi_p$的函数。

这种估计方法存在以下问题：

1. 这种方法很繁琐（除非我们处于纯AR情况下）：这些方程可能并不总是有解（例如，在MA($_1$)中，这种方法涉及求解 $r_1 = \theta/(1+\theta^2)$，当然在 $r_1 \notin [-0.5, 0.5]$时是没有解的）。 当这些方程无解时，参数的估计是任意的。

2. 得到的估计量效率低。 下面的其他技术给出了更好的估计结果（更小的标准误差）。

由于这些问题，没有人使用矩估计法来估计一般ARMA模型的参数。 甚至R都没有一个用于此目的的函数。 然而，需要注意的是，对于纯AR模型，这两个问题都消失了。

2.3 条件最小二乘法

首先考虑MA(1)模型的特殊情况： $X_t - \mu = Z_t + \theta Z_{t-1}$。 我们想要将这个模型拟合到数据 $x_1, \ldots, x_n$上。 如果数据确实是从这个模型生成的，那么

$$Z_1 = x_1 - \mu - \theta Z_0; Z_2 = x_2 - \mu - \theta Z_1; \ldots; Z_n = x_n - \mu - \theta Z_{n-1}。$$

如果我们将 Z_0设为其均值0，那么对于固定的 θ和 μ值，我们可以递归计算 $Z_1, \ldots, Z_n$。然后我们可以计算平方和 $\sum^n{}_{i=1} Z_i^2$。 这个值在不同的 θ值下会发生变化。 然后我们会选择使其最小的 θ值（通过优化过程实现）。

这也被称为条件最小二乘法，因为当我们试图最大化数据在条件概率下的条件似然时，这种最小化是得到的。

请注意，条件似然在AR($_1$)情况下与MA($_1$)情况下有所不同。 例如，在ARMA($_{1,\ 1}$)情况下，它以另一种不同的方式工作。 在这里，模型是 X_t - μ - $\phi(X\_\{t$-1} - μ) $= Z\_t + \theta Z\_\{t$-1}。 在这里，将 Z_1 设置为零是方便的。然后我们可以写成

$Z\_2 = x\_2$ - μ - $\phi(x\_1$ - μ); $Z\_3 = x\_3$ - μ - $\phi(x\_2$ - μ) - θZ_2; ...; $Z\_n = x\_n$ - μ - $\phi(x\_\{n$-1} - μ) - $\theta Z\_\{n$-1}。

在此之后，我们形成了平方和的总和 $\sum_{i=2}^{n} Z_i^2$，可以计算出每个固定值的θ、ϕ和μ。 然后，我们通过不同的未知参数值来最小化这些结果的平方和。

对于一般的ARMA(p, q)模型：

$$X_t - \mu - \phi_1(X_{t-1} - \mu) - \cdots - \phi_p(X_{t-p} - \mu) = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

我们设置 $Z_t = 0$，对于 $t \leq p$，并进行递归计算

$$Z_t = X_t - \mu - \phi_1(X_{t-1} - \mu) - \cdots - \phi_p(X_{t-p} - \mu) - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q}$$

对于$t = p +_1, \ldots, n$。 这等价于在$X_1, \ldots$上写出条件似然函数。 ，Xp和$Zt = 0$对于$t \leq p$。 如果$q = 0$（AR模型），最小化平方和等价于线性回归，不需要迭代技术。 如果$q > 0$，问题变为非线性回归，需要使用数值优化算法。

在R中，通过调用arima()函数并将method参数设置为CSS（CSS代表条件平方和）来执行此方法。

2.4 最大似然估计

这种方法原理上很简单。 假设误差$\{Z_t\}$服从高斯分布。 写出关于观测数据$x_1, \ldots$的似然函数。 ，x_{nt}以未知参数值$\mu, \theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p$和$\sigma Z^2$的形式。 对这些未知参数值进行最大化。

在R中，通过调用函数$arima()$并将方法参数设置为ML或CSS-ML来实现。ML当然代表最大似然。R使用优化算法来最大化似然函数。
这个算法是迭代的，并且需要合适的参数初始值。 在CSS-ML中，R通过CSS选择这些初始值。我不太清楚在ML中如何选择初始值。R中arima函数的默认方法是CSS-ML。CSS-ML和ML方法的R输出似乎是相同的。

2013年春季统计学153课程（时间序列）：第十四讲

阿迪蒂亚・冈图博伊纳

2013年3月12日

1 ARMA ML估计的渐近分布

设$\beta = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)$。$\beta$^的分布近似服从均值为$\beta$、方差协方差矩阵为$\sigma_Z{}^2\Gamma^-{}_{p,q}/$n的正态分布，其中$\Gamma_{p,q}$是一个形如的$(p+q)\times(p+q)$矩阵：

$$\begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix}$$

矩阵Γ的（i, j）th项等于γ_A（i） $-$其中$\{A_t\}$是AR(p)过程：ϕ（B）A_t=Z_t。类似地，矩阵$\Gamma\theta$的（i, j）$_t$h项等于γB（$i-j$）用于$AR(q)$过程θ（B）B_t=Z_t。Γ的（i, j）$_t$h项等于Ai和Bj之间的协方差。

1.1特殊情况

特别地，结果表明AR和MA模型的方差-协方差矩阵非常相似（唯一的区别在于符号）。在MA(1)情况下：

$$\Gamma_\theta = \Gamma_1 = \sigma_Z^2/(1-\theta^2).$$

因此，θ大致上服从均值为θ，方差为(1-θ^2)/n的正态分布。

对于MA(2)，(θ_1, θ_2)大致上服从均值为(θ_1, θ_2)的正态分布，方差协方差矩阵为1/n倍的单位矩阵。

$$\begin{pmatrix} 1-\theta_2^2 & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & 1-\theta_2^2 \end{pmatrix}$$

.

对于$ARMA(1, 1)$，要计算$\Gamma_{\phi\theta}$，我们必须找到A_1和B_1之间的协方差，其中A_1-$_\phi A_0$=Z_t，B_1+$_\theta B_0$=Z_t。写成

$$\Gamma_{\phi\theta} = \mathrm{cov}(A_1, B_1) = \mathrm{cov}(\phi A_0 + Z_1, -\theta B_0 + Z_t) = -\phi\theta\Gamma\phi\theta + \sigma Z^2$$

这给出了$\Gamma_{\phi\theta} = \sigma_Z^2/(1+\phi\theta)$。这表明（$\hat{\phi}$, $\hat{\theta}$）近似服从均值为（ϕ, θ）和方差-协方差矩阵为$1/n$乘以的逆

$$\begin{pmatrix} (1-\phi^2)^{-1} & (1+\phi\theta)^{-1} \\ (1+\phi\theta)^{-1} & (1-\theta^2)^{-1} \end{pmatrix}$$

.

2个ARIMA模型

ARIMA本质上是差分加上ARMA。我们之前已经看到，差分常用于去除时间序列数据的趋势和季节性。

例如，差分可以用于

1. 去除多项式趋势：假设数据来自模型 $Y_t = \mu_t + X_t$其中 μ_t是一个^{k}阶多项式， X_t是平稳的，那么^{k}阶差分 $\nabla kY_t = (I - B)\ kY_t$会得到平稳数据，可以拟合ARMA模型。

2. 随机游走模型：假设数据来自随机游走模型： $Y_t = Y_{t-1} + X_t$其中 X_t是平稳的。 显然 $\nabla Y_t = X_t$是平稳的，可以对这个差分数据拟合ARMA模型。

这种模型，在适当的差分之后，可以简化为ARMA模型，称为ARIMA模型。

定义**2.1** (ARIMA)**.**如果过程 Y_t是$ARIMA(p, d, q)$且均值为 μ，则 $X_t = (I - B)^d Y_t$是$ARMA(p, q)$且均值为 μ。换句话说：

$$\phi(B)(X_t - \mu) = \theta(B)Z_t,$$

其中 $\{Z_t\}$ 是白噪声。

3 拟合ARIMA模型

只需使用函数$arima(dataset, order = c(p, d, q))$。 我建议你总是使用这个函数。 如果你知道你想拟合一个纯AR模型，你可以考虑使用 $ar()$函数。

arima函数将给出 μ（在截距项下的估计值）， ϕ_1，...，ϕ_p和
θ_1，...，θ_q的估计值。 它还将给出估计的标准误差。 还提供了 σ^2的估计值。

2013年春季统计学153课程（时间序列）：第十五讲

阿迪蒂亚・冈图博伊纳

2013年3月14日

1 ARIMA预测

对于未来的观测值， x_{n+m}，使用 X_{n+m}的最佳线性预测器进行预测
X_1，...，X_n。 最佳线性预测器的系数涉及用于 x_1，...，x_n的ARIMA模型的参数。 这些参数是从数据中估计得到的。

我们已经看到了如何计算一个随机变量 Y在 $W_1, \dots, W_m$方面的最佳线性预测器。

假设所有的随机变量$Y, W_1, \dots, W_m$的均值为零。 那么最佳线性预测器就是 $a_1 W_1 + \cdots + a_m W_m$其中 $a_0, \dots$，a_m由以下方程组确定：

$$\mathrm{cov}(Y - a_1 W_1 - \cdots - a_m W_m, W_i) = 0 \qquad 对于\ i = 1, \dots, m.$$

上述方程给出了 m个未知数 $a_1, \dots, a\ m$的$_m$个方程。 这些方程可以以紧凑的形式写成 $\Delta a = \zeta$其中 $\Delta(i, j) = \mathrm{cov}(W_i, W_j)$ 且 $\zeta_i = \mathrm{cov}(Y, W_i)$。

如果随机变量$_Y$，W_1，...，W_m具有不同的均值： $\mathbb{E}_Y = \mu_Y$和 $\mathbb{E}W_i = \mu_i$，那么以 $W_1 - \mu_1$为条件的 $Y - \mu_Y$的最佳线性预测器是$W_1 - \mu_1$，...，$W_m - \mu_m$由 $a_1(W_1 - \mu_1) + \cdots + a_m(W_m - \mu_m)$给出，其中 $a1$，...，a_m由相同的方程式$\Delta a = \zeta$给出。 因此，在这些非零均值的情况下，以$W1$，...，W_m为条件的 Y的最佳线性预测器是

$$\mu_Y + a_1(W_1 - \mu_1) + \cdots + a_m(W_m - \mu_m)。$$

预测误差通过以下方式进行测量

$$\mathbb{E}\left(Y - \mu_Y - a_1(W_1 - \mu_1) - \cdots - a_m(W_m - \mu_m)\right)^2.$$

对于ARMA模型，存在迭代算法可以快速计算基于 $X_1, \dots$的最佳线性预测器of $X_{n+m}.$, X_n和相应的预测误差在 n和 m上进行递归计算，例如Durbin-Levinson和Innovations。这些方法不需要显式求逆矩阵Δ.

2时间序列数据分析

1. 探索性分析.
2. 确定是否有必要转换数据（无论是为了更好的解释还是为了稳定方差）.
3. 处理趋势或季节性. 可以通过拟合确定性模型或平滑或差分的方式进行.

4. 在去除趋势和季节性后，对残差拟合ARMA模型.

5. 检查拟合的ARMA模型是否合适（模型诊断）。

6. 预测。

3 模型诊断

在将ARIMA模型拟合到数据后，可以通过观察残差来形成残差序列： $x_i - \hat{x}^{i\ i-1}$，即观察值$x\ i$与基于前面的观察值$x_1, ..., x\ _{i-1}$的最佳线性预测之间的差异。 通常，通过将残差除以相应的预测误差的平方根来标准化残差。

如果模型拟合良好，标准化残差应该表现为均值为零、方差为一的独立同分布序列。可以通过观察残差图和自相关图来检查这一点。
还需要评估是否存在偏离正态分布的情况（可以通过查看Q-Q图来完成）。

设 $r_e(h)$为来自ARMA拟合的残差的样本自相关函数。 为了拟合良好，残差必须是均值为零、方差为一的独立同分布序列，这意味着对于$h= 1, 2, . . .$， $r_e(\ h)$。 必须是均值为0、方差为1的独立同分布序列/n。

除了绘制 r_e（h）之外，还有一个考虑到 r_e（h）
幅度的正式测试。 这是基于所谓的Q统计量的Ljung-Box-Pierce检验：

$$Q := n(n + 2) \sum_{h=1}^{H} \frac{r_e^2(h)}{n - h}.$$

在模型适应性的零假设下， Q的分布渐近地为 χ^2，自由度为 $H - p - q$。 最大滞后 H是任意选择的（通常为20）。 因此，如果观察到的 Q值超过 χ^2分布的（$1 - \alpha$）分位数，就会在水平α下拒绝零假设，自由度为 $H - p - q$。

4个季节性ARIMA模型

这些模型在小滞后（比如0和1）和某些季节性滞后（比如12）上具有非零自相关性，而在其他所有滞后上则为零自相关性。

考虑MA模型： $X_t\ = Z_t + \Theta Z_{t-12} = (1 + \Theta B^{12})Z_t$。 可以将其视为一个MA(12)
模型，其中 $\theta_1 = \cdots = \theta_{11} = 0$， $\theta_{12} = \Theta$。 这是一个平稳模型，其自协方差函数
仅在滞后0和12处非零。 因此，它被称为季节性MA(1)模型，季节周期为12。

一般化地，具有季节周期 s的季节性MA(Q)模型定义为

$$X_t\ = Z_t + \Theta_1 Z_{t-s} + \Theta_2 Z_{t-2s} + \cdots + \Theta_Q Z_{t-Qs}。$$

该平稳模型的自相关仅在滞后$0, s, 2s, . . . , Qs$处非零。 请注意，这只是一个MA(Qs)模型，其MA多项式为$1 + \Theta_1 z^s + \Theta_2 z^{2s} + \cdots + \Theta_Q z^{Qs}$。

对于co2数据集，我们需要一个具有非零自相关性的平稳模型，其滞后1、11、12和13具有自相关性（其他滞后具有零自相关性）。 这样一个模型的例子如下：

$$X_t = Z_t + \theta Z_{t-1} + \Theta Z_{t-12} + \theta\Theta Z_{t-13}.$$

更简洁地，这个模型可以写成

$$X_t = (1 + \theta B + \Theta B^{12} + \theta\Theta B^{13})Z_t = (1 + \theta B)(1 + \Theta B^{12})Z_t.$$

这只是一个具有$\mathrm{MA}(12)$多项式$(1 + \theta z)(1 + \Theta z^{12})$的$\mathrm{MA}(12)$模型。

很容易验证对于这个模型：

$$\gamma_X(h) = (1 + \theta^2)(1 + \Theta^2)\sigma_Z^2,$$

$$\rho_x(1) = \frac{\theta}{1 + \theta^2} \text{和} \quad \rho_X(12) = \frac{\Theta}{1 + \Theta^2}$$

和

$$\rho_X(11) = \rho_X(13) = \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)}.$$

在每个其他滞后期，自相关 $\rho_X(h)$ 等于零。

更一般地，我们可以考虑具有AR多项式 $\phi(z)\Phi(z)$ 和MA多项式 $\theta(z)\Theta(z)$ 的ARMA模型，其中

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p,$$

$$\Phi(z) = 1 - \Phi_1 z^s - \Phi_2 z^{2s} - \cdots - \Phi_p z^{Ps}$$

和

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q,$$

$$\Theta(z) = 1 + \Theta_1 z^s + \Theta_2 z^{2s} + \cdots + \Theta_Q z^{Qs}.$$

这被称为乘法季节性$\mathbf{ARMA}(p, q) \times (P, Q)_s$模型，具有季节周期 s。

在二氧化碳的例子中，我们想要使用这样的模型来处理第一和季节性差分数据。具体来说，我们想要使用模型$\mathrm{ARMA}(0, 1) \times (0, 1)_{12}$来处理季节性和第一阶差分数据：$\nabla\nabla_{12}X_t$。 如果差分序列 $\nabla^d\nabla^d{}_s Y_t$满足一个具有季节周期 s的$\mathrm{ARMA}(p, q) \times (P, Q)_s$模型，那么序列 $\{Y_t\}$被称为具有非季节性阶数p, d, q和季节性阶数P, D, Q以及季节周期 s的乘法季节性$\mathrm{AR}\mathbf{I}\mathrm{MA}$模型。

因此，我们希望将具有非季节性顺序0、1、1和季节性顺序0、1、1的乘法季节性ARIMA模型拟合到co2数据集上，季节周期为12。 可以使用带有seasonal参数的arima()函数将该模型拟合到数据上。

5 过度拟合作为诊断工具

在将适当的模型拟合到数据后，再拟合一个稍微更一般的模型。 例如，如果AR(2)模型看起来合适，可以使用AR(3)模型进行过度拟合。 在拟合AR(3)模型时，可以确认原始的AR(2)模型是否正确：

1. 额外的ϕ_3参数的估计值与零没有显著差异。
2. 共同参数ϕ_1和ϕ_2的估计值与其原始估计值没有显著变化。

如何选择这个通用模型以过拟合？ 在拟合更通用的模型时，不应同时增加AR和MA模型的阶数。 因为这会导致可辨识性问题。 例如：考虑MA(1)模型： $X_t = (1 + \theta B)Z_t$。 然后在两边乘以多项式$1 - \phi z$：我们可以看到 X_t也满足$\mathrm{ARMA}(1, 2)$模型： $X_t - \phi X_{t-1} = Z_t + (\theta - \phi)Z_{t-1} + \phi\theta Z_{t-2}$。

但请注意，参数 ϕ不是唯一的，因此如果我们对来自$\mathrm{MA}(1)$的数据集拟合$\mathrm{ARMA}(1,2)$模型，可能会得到一个任意的 ϕ估计。

一般来说，根据残差的分析找到过拟合模型是一个好主意。
例如，如果在拟合了$\mathrm{MA}(1)$模型之后，残差中仍然存在一个不太小的滞后2的相关性，则过拟合使用MA(2)而不是ARMA(1,1)模型。

2013年春季统计学153课程（时间序列）：第十六讲

阿迪蒂亚・冈图博伊纳

2013年3月19日

1 季节性ARMA模型

双向无限序列 $\{X_t\}$被称为具有周期 s的季节性$\mathrm{ARMA}(P, Q)$过程，如果它是平稳的，并且满足差分方程$\Phi(B^s)X_t = \Theta(B^s)Z_t$其中 $\{Z_t\}$是白噪声。

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

和

$$\Theta(B^s) = 1 + \Theta_1 B^2 + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}.$$

注意，这些也可以看作是$\mathrm{ARMA}(Ps, Qs)$模型。 然而请注意，这些模型有$P + Q + 1$（其中1是为了 σ^2 ）个参数，而一般的$\mathrm{ARMA}(Ps, Qs)$模型将有$Ps + Qs + 1$个参数。 因此，这些模型更加稀疏。

只有当$\Phi(z^s)$的每个根的幅度与1不同，才存在唯一的平稳解$\Phi(B^s)X_t = \Theta(B^s)Z_t$。只有当$\Phi(z^s)$的每个根的幅度严格大于1时，才存在因果平稳解。只有当$\Theta(z^s)$的每个根的幅度严格大于1时，才存在可逆平稳解。

这些模型的ACF和PACF仅在季节滞后$h = 0, s, 2s, 3s$等时存在非零值. 在这些季节滞后时，这些模型的ACF和PACF的行为与非季节性ARMA模型的情况完全相同：$\Phi(B)X_t = \Theta(B)Z_t$.

2个乘法季节性ARMA模型

在co2数据集中，对于第一个和季节性差分数据，我们需要拟合一个具有非零自相关的平稳模型，滞后为1、11、12和13（其他滞后的自相关为零）。 我们可以使用一个MA(13)模型来拟合这些数据，但是这将有14个参数，因此很可能过拟合数据。 通过将MA(1)模型与周期为12的季节性MA(1)模型相结合，我们可以得到一个更简洁的模型来拟合这个数据集。 具体来说，考虑以下模型

$$X_t = (1 + \Theta B^{12})(1 + \theta B)Z_t.$$

这个模型具有自相关函数：

$$\rho_x(1) = \frac{\theta}{1 + \theta^2} \text{和} \quad \rho_X(12) = \frac{\Theta}{1 + \Theta^2}$$

和

$$\rho_X(11) = \rho_X(13) = \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)}.$$

在每个其他滞后期，自相关 $\rho_X(h)$ 等于零。 因此，这是一个适合于co2数据集中的第一和季节性差分数据的模型。

更一般地，我们可以通过乘法将ARMA和季节性ARMA模型组合起来，以获得在季节滞后期方面具有特殊自相关性质的模型：

乘法季节性自回归滑动平均模型**ARMA**$(p, q) \times (P, Q)_s$
被定义为满足差分方程的平稳解：

$$\Phi(B^s)\phi(B)X_t = \Theta(B^s)\theta(B)Z_t.$$

我们上面看到的co2数据集的模型是ARMA(0, 1) $\times(0, 1)_{12}$.

另一个乘法季节性ARMA模型的例子是ARMA(0, 1) $\times(1, 0)_{12}$模型：

$$X_t - \Phi X_{t-12} = Z_t + \theta Z_{t-1}.$$

这个模型的自相关函数可以检查为 $\rho_X(h) = \Phi^h$对于 $h \geq 0$ 和

$$\rho_X(12h-1) = \rho_X(12h+1) = \frac{\theta}{1+\theta^2}\Phi^h \qquad \text{对于 } h = 0, 1, 2, \ldots$$

和 $\rho_X(h) = 0$ 在其他滞后值上.

当你得到一个平稳数据集，其自相关图在季节滞后上显示出有趣的模式时，考虑使用乘法季节性ARMA模型。 你可以使用R函数ARMAacf 来了解这些模型的自相关和偏自相关函数。

3个SARIMA模型

这些模型是通过将差分与乘法季节性ARMA模型相结合得到的。 这些模型用ARIMA$(p, d, q) \times (P, D, Q)_s$表示。 这意味着经过 d次差分和季节性差分 D次后，我们得到一个乘法季节性ARMA模型。 换句话说，如果$\{Y_t\}$满足差分方程，那么它就是ARIMA$(p, d, q) \times (P, D, Q)_s$。

$$\Phi(B^s)\phi(B)\nabla^{Ds}\ \nabla^d Y_t = \delta + \Theta(B^s)\theta(B)Z_t。$$

回想一下 $\nabla^d{}_s = (1-B^s)^d$和 $\nabla^d = (1-B)^d$表示差分运算符。

在co2的例子中，我们想要使用模型ARMA$(0, 1) \times (0, 1)_1 2$ 来处理季节性和第一阶差分的数据： $\nabla\nabla_{12}X_t$。 换句话说，我们想要拟合具有非季节性阶数$0, 1, _1$和季节性阶数$0, 1, _1$以及季节周期$_1$2的SARIMA模型到co2数据集中。 可以使用函数 $arima()$和 $seasonal$参数将该模型拟合到数据上。

4 AIC

AIC代表Akaike信息准则。它是一种模型选择准则，建议选择一个模型，其中：

$$AIC = -2\ \log(\text{最大似然}) + 2k$$

是最小的。 这里 k表示模型中的参数个数。 例如，在具有非零均值 μ的ARMA(p, q)模型的情况下，我们有 $k = p + q + 2$。

AIC定义中的第一项衡量了模型的拟合程度，即模型在给定数据集上的性能。项$2k$作为惩罚函数，惩罚具有过多参数的模型。

在比较给定数据集的一组模型时，可以使用AIC。还有其他标准。 例如，贝叶斯信息准则（BIC）考虑以下因素：

$$BIC = -2\log(\text{最大似然}) + k\log n。$$

请注意，上述惩罚项大于AIC的惩罚项。因此，BIC选择比AIC更简洁的模型。

5个时间序列交叉验证

阅读Rob Hyndman博客上的两篇文章：`http://robjhyndman.com/hyndsight/crossvalidation/` 了解交叉验证的简单介绍和`http://robjhyndman.com/hyndsight/tscvexample/`了解特定于时间序列的交叉验证。

有很多方法可以进行时间序列的交叉验证。 假设我们有连续 m年的月度数据 $x_1, ..., x_n$其中 $n = 12m$，目标是预测下一年的数据（这类似于期中考试问题，只是数据是每周而不是每月）。 假设我们有 ℓ竞争模型$M_1, ..., M_\ell$用于数据集。 我们可以使用交叉验证来选择这些模型中的一个，具体步骤如下：

1. 固定一个模型 M_i。固定$k < m$。
2. 将模型拟合到从第一个 k年的数据中。
3. 使用拟合的模型，预测第（k+ 1）年的数据。
4. 计算第（k+ 1）年的预测误差平方和。
5. 重复这些步骤，对于 $k = k_0, ... , m -1$，其中 k_0是您选择的任意值。
6. 对于 $k = k_0, ..., m - 1$，计算预测误差平方和的平均值。 将这个值表示为 CV_i，并称之为模型 M_i的交叉验证分数。
7. 计算每个 $_i= 1, ... , \ell$的 $CV\,i$，并选择具有最小交叉验证分数的模型。

2013年春季统计学153课程（时间序列）：第十七讲

阿迪蒂亚・冈图博伊纳

2013年3月21日

过拟合作为一种诊断工具

在将适当的模型拟合到数据后，再拟合一个稍微更一般的模型。 例如，如果AR(2)模型看起来合适，可以使用AR(3)模型进行过度拟合。 在拟合AR(3)模型时，可以确认原始的AR(2)模型是否正确：

1. 额外的ϕ_3参数的估计值与零没有显著差异。
2. 共同参数ϕ_1和ϕ_2的估计值与其原始估计值没有显著变化。

如何选择这个通用模型进行过拟合？ 在拟合更通用的模型时，不应同时增加AR和MA模型的阶数。 因为这会导致可识别性问题。 例如：考虑MA(1)模型： $X_t = (1+\theta B)Z_t$。 然后在两边乘以多项式$1-\phi z$：我们可以看到 X_t也满足ARMA(1, 2)模型： $X_t - \phi X_{t-1} = Z_t + (\theta - \phi)Z_{t-1} + \phi\theta Z_{t-2}$。但请注意，参数 ϕ不是唯一的，因此如果我们对来自MA(1)的数据集拟合ARMA(1, 2)模型，可能会得到任意的 ϕ估计。

一般来说，根据残差的分析找到过拟合模型是一个好主意。
例如，如果在拟合了MA(1)模型之后，残差中仍然存在一个不太小的滞后2的相关性，则过拟合使用MA(2)而不是ARMA(1,1)模型。

2个正弦和余弦

频域技术：使用正弦和余弦来研究时间序列数据。

正弦波：$R\cos(2\pi ft+\Phi)$。 以下术语是标准的。 R被称为振幅，f被称为频率，Φ被称为相位。. 数量$1/f$被称为周期，$2\pi f$被称为角频率。 请注意，正弦波的定义涉及三个参数R，f和Φ。

该函数也可以写成$A\cos 2\pi \mathrm{ft} + B\sin 2\pi \mathrm{ft}$，其中$A = R\cos\Phi$，$B = R\sin\Phi$。 这个参数化也有三个参数：f，$A$和$B$。

2013年春季统计学153课程（时间序列）：第十八讲

阿迪蒂亚・冈图博伊纳

2013年4月2日

1 正弦波

正弦波： $R\cos(2\pi f t + \Phi)$. 以下术语是标准的。R被称为振幅， f被称为频率，Φ被称为相位。 数量$1/f$被称为周期，$2\pi f$被称为角频率。 请注意，正弦波的定义涉及三个参数R、f和 Φ。

该函数也可以写成$A\cos 2\pi \mathrm{ft} + B\sin 2\pi \mathrm{ft}$，其中$A = R\cos\Phi$，$B = R\sin\Phi$。 这个参数化也有三个参数：f，$A$和$B$。

表示正弦波的另一种方式是使用复指数形式：

$$\exp(2\pi i f t) = \cos(2\pi f t) + i\sin(2\pi f t)。$$

因此

$$\cos(2\pi f t) = \frac{\exp(2\pi i f t) + \exp(-2\pi i f t)}{2} \text{ 和 } \sin(2\pi f t) = \frac{\exp(2\pi i f t) - \exp(-2\pi i f t)}{2i}.$$

因此 $A\cos 2\pi f t + B\sin 2\pi f t$也可以写成$\exp(2\pi i f t)$和$\exp(-2\pi i f t)$的线性组合。

在某些特定频率上的正弦波具有良好的正交性质。 考虑向量：

$$u = (1, \exp(2\pi i/n), \exp(2\pi i 2/n), \ldots, \exp(2\pi i(n-1)/n)).$$

这是频率为 $f = 1/n$的正弦波$\exp(2\pi i f t)$在时间点 $t = 0, 1, 2, \ldots\ldots\ldots\ldots\ldots\ldots, (n-1)$的值。

同时，定义与频率为f的正弦波$\exp(2\pi \mathrm{i} f t)$对应的向量，其中$f$=$j/n$，在$t$=0,1,...,($n$-1)时进行评估

$$u\_^j = (1,\ \exp(2\pi i^j/n),\ \exp(2\pi i 2^j/n),\ ...,\ \exp(2\pi i(n\text{-}1)^j/n))$$

这些向量$u_^0, u_^1, u_^2, \ldots u_^k$和$u_^l$之间的点积为零，如果^k=^{l} 因此，每个数据向量$x := (x_^1, ..., x_^n)$可以写成$u_^0, u_^1, ..., u_(^n$-$^1)$的线性组合

2 离散傅里叶变换

令 x 等于 $_0$，...，$x\_{_n}$-$_1$ 表示长度为 $_n$ 的数据（实数）。x

的离散傅里叶变换由 b_j 给出，其中 $_j = _{0,\,1,\,...,\,n}$-$_1$，

$$b\_j = \sum_{t=0}^{n^{-1}} x\_t \exp\left(\frac{-2\pi i j t}{n}\right) \qquad \text{对于}\ \ j = 0,\ ...,\ n\text{-}\ 1。$$

因此，$b\_0 = \sum x\_t$。对于 $1 \le j \le n$-1，

$$b\_{n-j} = \sum_t x\_t \exp\left(\frac{-2\pi i\,(n-j)\,t}{n}\right) = \sum_t x\_t \exp\left(\frac{2\pi ijt}{n}\right) \exp(-2\pi it) = \bar{b}\_j。$$

请注意，只有当 x为$_0, \ldots\ldots$时，这个关系才成立，x_{n-1}是实数。

因此，对于 $n=11$，DFT可以写成：

$$b_0, b_1, b_2, b_3, b_4, b_5, \bar{b}_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1。$$

而对于 $n = 12$，它是

$$b_0, b_1, b_2, b_3, b_4, b_5, b_6 = \bar{b}_6, \bar{b}_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1。$$

请注意，b_6必须是实数，因为 $b_6 = \bar{b}_6$。

DFT是通过R函数 `fft()` 计算的。

原始数据 $x_0, \ldots\ldots\ldots\ldots\ldots\ldots$，$x_{n-1}$可以通过DFT恢复：

$$x_t = \frac{1}{n}\sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi ijt}{n}\right) \qquad 对于\ t = 0, \ldots, n-1。$$

因此，对于 $n=11$，可以将数据视为11个实数 $x_0, x_1, \ldots$，x_{10}或者等价地，一个实数 b_0以及5个复数 $b_1, \ldots, b_5$。

对于 $n=12$，可以将数据视为12个实数 $x_0, \ldots$，x_{11}或者等价地，两个实数 b_0和 b_6，以及5个复数 $b_1, \ldots, b_5$。

以下恒等式始终成立：

$$n\sum_t x_t^2 = \sum_{j=0}^{n-1} |b_j|^2 .$$

因为 $b_{n-j} = \bar{b}_j$，它们的绝对值相等，我们可以将上述平方和恒等式写成以下形式。对于 $n=11$，

$$n\sum_t x_t^2 = b_0^2 + 2|b_1|^2 + 2|b_2|^2 + 2|b_3|^2 + 2|b_4|^2 + 2|b_5|^2$$

对于 $n = 12,$

$$n\sum_t x_t^2 = b_0^2 + 2|b_1|^2 + 2|b_2|^2 + 2|b_3|^2 + 2|b_4|^2 + 2|b_5|^2 + b_6^2.$$

请注意，对于$b=_6$，不需要取绝对值，因为它是实数。因为$b=_0=\sum_t x_t = n\bar{x}$，我们有

$$n\sum_t x_t^2 - b_0^2 = n\sum_t x_t^2 - n^2\bar{x}^2 = n\sum_t (x_t - \bar{x})^2.$$

因此，对于奇数n，比如n=11，平方和恒等式可以写为

$$\sum_t (x_t - \bar{x})^2 = \frac{2}{n}|b_1|^2 + \frac{2}{n}|b_2|^2 + \frac{2}{n}|b_3|^2 + \frac{2}{n}|b_4|^2 + \frac{2}{n}|b_5|^2$$

对于偶数n，比如n=12，平方和恒等式可以写为

$$\sum_t (x_t - \bar{x})^2 = \frac{2}{n}|b_1|^2 + \frac{2}{n}|b_2|^2 + \frac{2}{n}|b_3|^2 + \frac{2}{n}|b_4|^2 + \frac{2}{n}|b_5|^2 + \frac{1}{n}b_6^2.$$

2013年春季统计学153课程（时间序列）：第十九讲

阿迪蒂亚・冈图博伊纳

2013年4月4日

1 DFT复习

给定数据 $x_0, ..., x_{n-1}$, 它们的DFT由 $b_0, b_1, . . . , b_{n-1}$给出，其中

$$b_j := \sum_{t=0}^{n-1} x_t \exp\left(\frac{-2\pi ijt}{n}\right) \qquad 对于\ j = 0, 1, . . . , n-1.$$

记住的两个关键点是：

1. $b_{n-j} = \bar{b}_j$对于 $1 \leq j \leq n-1$

对于奇数值的 n，比如 $n = 11$，DFT由实数 b_0和$(n-1)/2$复数 $b_1, . . . , b_{(n-1)/2}$组成。

对于偶数值的n，比如$n=12$，DFT由两个实数b_0和$b_{n/2}$以及（n-2）/2个复数b_1，...，$b_{(n-2)/2}$组成

原始数据 x_0， x_1，...， $x_{(n-1)}$可以通过公式从DFT中恢复出来

$$x_t = \frac{1}{n}\sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi ijt}{n}\right) \qquad 对于\ t=0，1，...，n-1$$

这个公式可以简洁地写成

$$x = \frac{1}{n}\sum_{j=0}^{n-1} b_j u^j$$

其中$x=(x_0,...,x_{(n-1)})$表示数据向量，u^j表示在时间$t=0$，1，...，n-1处评估的正弦波$\exp(2\pi ijt/n)$ 我们在上一堂课中已经看到u^j是正交的，$u^j \cdot u^j = n$对于每个j 因此，我们有平方和恒等式

$$n\sum_t x_t^2 = \sum_{j=0}^{n-1} |b_j|^2 .$$

绝对值 b_j和 b_{n-j}相等，因为 $b_{n-j} = \bar{b}_j$，因此我们可以将上述平方和恒等式写成以下形式。对于 $n = 11$，

$$n\sum_t x_t^2 = b_0^2 + 2|b_1|^2 + 2|b_2|^2 + 2|b_3|^2 + 2|b_4|^2 + 2|b_5|^2$$

对于 $n=12$,

$$n\sum_t x_t^2 = b_0^2 + 2|b_1|^2 + 2|b_2|^2 + 2|b_3|^2 + 2|b_4|^2 + 2|b_5|^2 + b_6^2.$$

请注意，对于b_6，不需要取绝对值，因为它是实数。因为$b_0=\sum_t x_t = n\bar{x}$，我们有

$$n\sum_t x_t^2 - b_0^2 = n\sum_t x_t^2 - n^2\bar{x}^2 = n\sum_t (x_t-\bar{x})^2.$$

因此，对于奇数n，比如$n=11$，平方和恒等式可以写为

$$\sum_t (x_t-\bar{x})^2 = \frac{2}{n}|b_1|^2 + \frac{2}{n}|b_2|^2 + \frac{2}{n}|b_3|^2 + \frac{2}{n}|b_4|^2 + \frac{2}{n}|b_5|^2$$

对于偶数n，比如$n=12$，平方和恒等式可以写为

$$\sum_t (x_t-\bar{x})^2 = \frac{2}{n}|b_1|^2 + \frac{2}{n}|b_2|^2 + \frac{2}{n}|b_3|^2 + \frac{2}{n}|b_4|^2 + \frac{2}{n}|b_5|^2 + \frac{1}{n}b_6^2.$$

余弦波的2 DFT

令 $x_t = R\cos(2\pi f_0 t + \phi)$ 对于 $t=0,\dots,n-1$。我们在R中已经看到，当 f_0是傅里叶频率时（即，形式为 k/n的频率），DFT只有一个峰值，但当 f_0不是傅里叶频率时，会有泄漏。我们在这里证明这一点。

我们可以假设$0\le f_0\le 1/2$，而不失一般性，因为：

1。 如果 $f_0<0$，那么我们可以写成 $\cos(2\pi f_0 t+\phi) = \cos(2\pi(-f_0)t-\phi)$。显然，$-f_0\ge 0$。

2. 如果$f_0\ge 1$，则我们写

$$\cos(2\pi f_0 t+\phi) = \cos(2\pi[ft_0]t + 2\pi(ft-[ft_0])t+\phi) = \cos(2\pi(ft-[ft_0])t+\phi),$$

因为$\cos(\cdot)$是周期为2π的。 显然$0\le f-[ft_0]<1$.

3. 如果$f_0\in[1/2,1)$，则

$$\cos(2\pi f_0 t+\phi) = \cos(2\pi t - 2\pi(1-f_0)t+\phi) = \cos(2\pi(1-f_0)t-\phi)$$

因为$\cos(2\pi t-x)=\cos x$对于所有整数t. 显然$0<1-f_0\le 1/2$.

因此，给定一个余弦波$R\cos(2\pi ft+\phi)$，我们总是可以将其写成$R\cos(2\pi f_0 t+\phi')$，其中$0\le f_0\le 1/2$，而相位ϕ'可能与ϕ不同. 这个频率f_0被称为f的别名. 从现在开始，每当我们提到余弦波$R\cos(2\pi f_0 t+\phi)$，我们假设$0\le f_0\le 1/2$.

如果$\phi=0$，那么我们有$x_t=R\cos(2\pi f_0 t)$. 当$f_0=0$时，$x_t=R$，因此数据中没有振荡。当$f_0=1/2$时，$x_t=R\cos(\pi t)=R(-1)^t$，因此$f_0=1/2$对应于最大可能的振荡.

对于$0\le f_0\le 1/2$，$x(t)=R\cos(2\pi f_0 t+\phi)$的DFT是什么？公式为

$$b_j := \sum_{t=0}^{n-1} x_t \exp\left(\frac{-2\pi ijt}{n}\right).$$

假设$f = j/n$，我们将计算

$$b(f) = \sum_{t=0}^{n-1} x_t \exp(-2\pi i f_t)$$

计算这个DFT最简单的方法是用复指数形式表示余弦波：

$$x_t = \frac{R}{2}\left(e^{2\pi i f_0 t} e^{i\phi} + e^{-2\pi i f_0 t} e^{-i\phi}\right) \quad .$$

因此，首先计算复指数$e^{2\pi \mathrm{i} f_0 t}$的DFT是方便的。

$y(t) = e^{2\pi i f_0 t}$的DFT为

$y(t) = e^{2\pi i f_0 t}$的DFT给出

$$\sum_{t=0}^{n-1} y_t e^{-2\pi i f t} = \sum_{t=0}^{n-1} e^{2\pi i (f_0 - f)t}$$

其中 $f = j/n$. 我们用 $S_n(f_0 - f)$ 表示这个

$$S_n(f_0 - f) = \sum_{t=0}^{n-1} e^{2\pi i (f_0 - f)t}. \tag{1}$$

这可以明显地使用几何级数公式来计算

$$S_n(f_0 - f) = \frac{e^{2\pi i (f_0 - f)n} - 1}{e^{2\pi i (f_0 - f)} - 1}$$

很容易检查

$$e^{i\theta} - 1 = \cos\theta + i\sin\theta - 1 = 2e^{i\theta/2}\sin\theta/2.$$

因此

$$S_n(f_0 - f) = \frac{\sin \pi n (f_0 - f)}{\sin \pi (f_0 - f)} e^{i\pi(f_0 - f)(n-1)}$$

因此， y_t的DFT的绝对值 $= e^{2\pi i f_0 t}$的结果为

$$|S_n(f - f_0)| = \left|\frac{\sin \pi n (f_0 - f)}{\sin \pi (f_0 - f))}\right| \qquad \text{其中} \quad f = j/n$$

当 $f_0 = f$时，这个表达式变得没有意义。 但是当 $f_0 = f$时， $S_n(f_0 - f)$的值可以直接从(1)计算得到，等于 n.

通过绘制函数 $g \to (\sin \pi n g)/(\sin \pi g)$，可以最好地理解 $|S_n(f - f_0)|$的行为。
这解释了泄漏问题。

通过将余弦波的DFT写成复指数的形式，可以研究其行为。

如果 f_0不是 k/n的形式，对于任何 j，则形如 j/n的函数 $S_n(f - f_0)$对于所有形如j/n的 f都是非零的。当存在一个与频率 f_0不同于 j/n的正弦波时，观察到一个非零的DFT项 b_j的情况被称为泄漏。

由于频率 f_0不是 k/n的形式，泄漏存在于所有的DFT项 b_j中，但是其存在的幅度随着 j/n远离 f_0而衰减。 这是因为函数$S_n(f - f_0)$的形式。

泄漏存在两个问题：

1. 傅立叶分析通常用于分离不同频率的效应；因此泄漏是一种不希望出现的现象。

2. 由于频率为f_0的正弦波在j/n处的泄漏可能掩盖了频率为j/n的真正正弦波的存在。

如何消除泄漏？ 简单的答案是适当选择n（理想情况下，n应为所有振荡周期的倍数）。 例如，如果是月度数据，则最好有一整年的数据。 但这并不总是可能的。 我们将在以后学习一种减少泄漏的技术。

3 周期序列的离散傅立叶变换

假设数据x_0，x_1，...，x_n-1是周期为h的周期性数据，即对于所有整数$_t$和u，$x_{t+}h^\star u$=x_t。 让$_n$是h的整数倍，即$_n$=$k^\star h$。 例如，假设我们有在$_{10}$年内收集的月度数据，则：h=$_1$2，k=$_{10}$，$_n$=$_1$2$_0$。

假设数据的离散傅里叶变换为$x_0, x_1, ...$，$x(_n$-$_1)$是$b_0, b_1, ..., b_n$-$_1$ 假设数据在第一个周期的离散傅里叶变换为$x_0, x_1, ...$，$x(_h$-$_1)$是$\beta_0, \beta_1, ..., \beta_h$-$_1$

我们将用 $\beta_0, ...,$ β_h-$_1$来表示 b_j 为简单起见，令 $f =_j$ /n

根据定义

$$b\_j = \sum_{t=0}^{n-1} x_t = \exp(-2\pi i_t f)$$

将求和分解为

$$\sum_{t=0}^{h-1} + \sum_{t=h}^{2h-1} + \cdots + \sum_{t=(k-1)h}^{kh-1}$$

上面的第l个项可以计算为：

$$\begin{aligned}\sum_{t=(l-1)h}^{lh-1} x_t \exp\left(-2\pi i f\, t\right) &= \sum_{s=0}^{h-1} x_s \exp\left(-2\pi i f(s+(l-1)h)\right)\\ &= \exp\left(-2\pi i f(l-1)h\right)\sum_{s=0}^{h-1} x_s \exp(-2\pi i f\, s).\end{aligned}$$

因此

$$
\begin{aligned}
b_{-j} = & \sum_{l=1}^{k} \exp\ (-2\pi if(l-1)h) \sum_{s=0}^{h-1} x_s \exp(-2\pi if\ s) \\
&= \sum_{s=0}^{h-1} x_s \exp(-2\pi if\ s) \sum_{l=1}^{k} \exp\ (-2\pi if(l-1)h) \\
&= S_k(f\ h) \sum_{s=0}^{h-1} x_s \exp(-2\pi if\ s) \\
&= S_k(jh/n) \sum_{s=0}^{h-1} x_s \exp(-2\pi ijs/n) \\
&= S_k(j/k) \sum_{s=0}^{h-1} x_s \exp(-2\pi i(j/k)s/h)
\end{aligned}
$$

因此当$_j$不是 k的倍数时， $b\ j= 0$，当 j是 k的倍数时， $|b_j| = k|\beta_{j/k}|$。

因此，原始的DFT项 β_0，β_1等等。 现在出现为 $b_0 = k\beta_0$，$b_k = k\beta_1$，$b_{2k} = k\beta_2$等等，直到 $b_{(h-1)k} = k\beta_{h-1}$。 其他所有 b_j都为零。

4 DFT和样本自协方差函数

我们下面展示

$$
\frac{|b_j|^2}{n} = \sum_{|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi ijh}{n}\right) \qquad \text{对于 } j= 1, \ldots, n-1
$$

其中 $\hat{\gamma}(h)$ 是样本自协方差函数。 这给出了离散傅里叶变换和样本自协方差函数之间的重要联系。

为了看到这一点，首先观察到，根据等比级数求和公式，有

$$
\sum_{t=0}^{n-1} \exp\left(\frac{-2\pi ijt}{n}\right) = 0 \qquad \text{对于 } j= 1, \ldots, n-1\text{。}
$$

换句话说，如果数据是常数，即 $x_0 =\cdots = x_{n-1}$，则 b_0等于 nx_0，而对于其他所有的$_j$， $b\ j$等于0。 因此，我们可以写成：

$$
b_{-j} = \sum_{t=0}^{n-1} (x_t - \bar{x}) \exp\left(\frac{-2\pi ijt}{n}\right) \qquad \text{对于 } j= 1, \ldots, n-1\text{。}
$$

因此，对于 $j = 1, \ldots, n-1$，我们写成

$$
\begin{aligned}
|b_j|^2 = b_j \bar{b}_j &= {}^{n}\Sigma^{-1} \sum_{t=0}^{n-1}\sum_{s=0} (x - \bar{x}) \quad (x - \bar{x}) \ \exp\left(\frac{-2\pi i j t}{n}\right) \exp\left(\frac{2\pi i j s}{n}\right) \\
&= {}^{n}\Sigma^{-1} \sum_{t=0}^{n-1}\sum_{s=0} (x - \bar{x}) \quad (x - \bar{x}) \ \exp\left(\frac{-2\pi i j \ (t - s)}{n}\right) \\
&= \sum_{h=-(n-1)}^{n\,-1} \sum_{t,\, s:\, t-s=h} (x - \bar{x}) \quad (x -{}_{t} - \bar{x}) \ \exp\left(-\frac{2\pi i j h}{n}\right) \\
&= n \sum_{|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi i j h}{n}\right).
\end{aligned}
$$

2013年春季统计学153课程（时间序列）：第二十讲

阿迪蒂亚・冈图博伊纳

2013年4月9日

再次进行1 DFT

数据用 x_0，x_1, ...，x_{n-1}表示。

DFT用 b_0，b_1, ...，b_{n-1}表示。

DFT是从数据计算得出的。

$$b_j := \sum_{t=0}^{n-1} x_t \exp\left(\frac{-2\pi ijt}{n}\right) \qquad 对于\ j = 0, 1, \ldots, n-1. \tag{1}$$

数据是从DFT计算得出的。

$$x_t = \frac{1}{n} \sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi ijt}{n}\right) \qquad 对于\ t=0，1，...，n-1$$

在教科书中，DFT的公式为（1），额外乘以$(n^{-1})^{/2}$的因子。 我去掉了这个因子，以使定义与R函数fft兼容。

对于奇数值的 n，DFT由实数 b_0和$(n-1)/2$复数 $b_1, \ldots, b_{(n-1)/2}$组成。

对于偶数值的 n，DFT由两个实数 b_0和 $b_{n/2}$以及$(n-2)/2$复数 $b_1, \ldots, b_{(n-2)/2}$组成。

2 DFT是做什么的?

假设 $x_t = R\cos(2\pi f_0 t + \Phi)$，其中 $t = 0, 1, \ldots, n-1$。 我们在上一堂课中已经看到，我们只需要考虑范围在$0 \le f_0 \le 1/2$的频率（因为其他频率在区间$[0, 1/2]$中有一个别名）。

首先假设$f(_0)$的形式为k/n，其中$0 \le k/n \le 1/2$。然后DFT的表达式为

$$\begin{aligned} b_j &= \sum_{t=0}^{n-1} R\cos(2\pi(k/n)t + \Phi)\exp(-2\pi i(j/n)t) \\ &= \frac{Re^{(i\Phi)}}{2} \sum_{t=0}^{n-1} \exp\left(2\pi it\,\frac{j-k}{n}\right) + \frac{Re^{(-i\Phi)}}{2} \sum_{t=0}^{n-1} \exp\left(-2\pi it\ \frac{j+k}{n}\right). \end{aligned}$$

注意我们不需要考虑$_j$/n>1/2的DFT项。 因此我们假设0≤$_j$/n≤1/2。
由于原始余弦波的频率假设在[0,1/2]范围内，所以有0≤k/n≤
1/2. 当 j =k时，检查$0 < (k + j)/n <1$是否成立。 因此，当 $j = k$时，上述第二项始终为零，第一项为零，当 $j = k$时等于 $Re^{i\Phi}n/2$。 因此，对于频率为 k/n的余弦波的DFT，其中$0 \le k/n \le 1/2$，有 $b_k = n$ $Re^{i\Phi}/2$，对于$_j = k$，有 b j= 0，其中$0 \le j/n \le 1/2$。

现在考虑由多个频率线性组合的数据：

$$x_t = \sum_{l=1}^{m} R_l \cos(2\pi t(k_l/n) + \Phi_l) \tag{2}$$

其中每个 k_l都是满足$0 \le k_l/n \le 1/2$的整数。 因为DFT的定义在数据 $\{x_t\}$中是线性的，所以根据公式(2)，DFT的结果为

$$b_{-j} = \begin{cases} nR_l e^{i\Phi_l}/2 & \text{如果 } j = k_l \\ & \text{则为0，否则为1。} \end{cases}$$

对于$0 \le j/n \le 1/2$。

这表明DFT可以提取出数据中存在的频率。 DFT在某个频率上的强度（绝对值）与该频率上余弦波的振幅（R_l）成比例。

3 解释DFT

DFT将给定的数据表示为频率为 k/n的正弦波的组合。
频率为 k/n的形式被称为傅里叶频率。

假设我们有一个数据集 $x_0, \ldots, x_{n-1}$。 我们已经计算了它的DFT: $b_0, b_1, \ldots, b_{n-1}$，并且我们已经绘制了 $|b_j|$，其中 j= $1, \ldots, (n-1)/2$对于奇数 n，以及 j= $1, \ldots, n/2$对于偶数 n。

如果我们在这个图中看到一个单独的尖峰，比如在b_k处，我们可以确定数据是一个频率为k/n的正弦波。

如果我们看到两个尖峰，比如在 b_{k_1}处 和 b_{k_2}处，那么数据就稍微复杂一些：它是两个频率为k_1/n和k_2/n的正弦波的线性组合，这些正弦波的强度取决于尖峰的大小。

多个尖峰表明数据由许多傅里叶频率的正弦波组成，一般来说，这意味着数据更加复杂。

然而，有时候即使数据的结构并不是很复杂，也可以看到DFT中的多个尖峰。 一个典型的例子是由于存在一个非傅里叶频率的正弦波而导致的泄漏。

非傅里叶频率的正弦波的DFT计算如下：考虑信号$x^t = e^{2\pi f_0}{}_t$，其中$f_0 \in [0, 1/2]$不一定是k/n的形式。它的DFT由以下公式给出：

$$b_j := \sum_{t=0}^{n-1} x_t e^{-2\pi i t(j/n)} = \sum_{t=0}^{n-1} e^{2\pi i(f_0-(j/n))t}.$$

如果我们用函数表示

$$S_n(g) := \sum_{t=0}^{n-1} e^{2\pi i g t} \tag{3}$$

那么我们可以写成

$$b_j = S_n(f_0 - (j/n)).$$

函数 $S_n(g)$可以明显地使用几何级数公式计算得到

$$S_n(g) = \frac{e^{2\pi i g n} - 1}{e^{2\pi i g} - 1}$$

因为

$$e^{i\theta} - 1 = \cos\theta + i\sin\theta - 1 = 2e^{i\theta/2}\sin\theta/2,$$

所以我们得到

$$S_n(g) = \frac{\sin \pi n g}{\sin \pi g} e^{i\pi g(n-1)}$$

因此， y_t的DFT的绝对值 $= e^{2\pi i f_0 t}$的结果为

$$|b_j| = |S_n(f_0 - (j/n))| = \left|\frac{\sin \pi n(f_0 - (j/n))}{\sin \pi(f_0 - (j/n)))}\right|$$

当 $f_0 = j/n$时，这个表达式变得没有意义。但是当 $f_0 = f$时， $S_n(f_0 - j/n)$的值可以直接从（3）计算得到，等于 n。

通过绘制函数 $g \to (\sin \pi n g)/(\sin \pi g)$，可以最好地理解这个DFT的行为。

通过Hanning减少泄漏

Hanning是一种减少泄漏的技术，它说：将数据乘以窗口或淡出器：

$$w_t = 1 - \cos\ (2\pi t/n) \qquad \text{对于 } t = 0, 1, \ldots, n-1$$

然后进行DFT。

为什么它有效？ 以下是 $y_t = w_t e^{2\pi i f_0 t}$的DFT（下面的$f$代表$j$/n）

$$\begin{aligned}
b_y(f) &:= \sum_{t=0}^{n-1} w_t e^{2\pi i f_0 t} e^{-2\pi i f\, t} \\
&= \sum_{t=0}^{n-1} (1 - \cos(2\pi t/n))\, e^{2\pi i (f_0 - f)t} \\
&= \sum_{t=0}^{n-1} e^{2\pi i (f_0 - f)t} - \frac{1}{2}\sum_{t=0}^{n-1} e^{2\pi i (f_0 - f + 1/n)t} - \frac{1}{2}\sum_{t=0}^{n-1} e^{2\pi i (f_0 - f - 1/n)t} \\
&= S_n(f_0 - f) - \frac{1}{2} S_n(f_0 - f + 1/n) - \frac{1}{2} S_n \text{（时间序列）}.
\end{aligned}$$

显然，当f_0时， $S\ n$（$1/n$）$= {}_0$. 假设 $g = f_0 - f$. 为了消除泄漏，我们需要确保 f不等于f_0时， b_y（ f）接近于零。这几乎是不可能的，但我们可以启发性地展示，当 | $f - f_0$|相对于$1/n$来说足够大时， $b\ f$（$_f$）接近于零。

让 g 表示 $f - f_0$ ，那么

$$b_y(f) = S_n(g) - \frac{1}{2} S_n(g - 1/n) - \frac{1}{2} S_n\text{（}g + 1/n\text{）。} \tag{4}$$

我们在上一节中推导出

$$S_n(g) = \frac{\sin \pi n g}{\sin \pi g} e^{\pi i g\ (n-1)}\text{。}$$

因此，

$$S_n(g-1/n)=\frac{\sin\pi n(g-1/n)}{\sin\pi(g-1/n)}e^{\pi i(g-1/n)(n-1)}$$

现在$\sin\pi n$（$g-1/n$）$=$-$\sin\pi ng$，如果$1/n$相对于 g很小，那么$\sin\pi$（$g-1/n$）$\approx\sin\pi g$。还有当$1/n$相对于 g很小时，我们有

$$e^{\pi i(g-1/n)(n-1)}=e^{\pi ig(n-1)}e^{-i\pi(n-1)/n}\approx e^{\pi ig(n-1)}e^{-i\pi}=-e^{\pi ig(n-1)}。$$

因此，如果$1/n$相对于 g很小，我们有

$$S_n(g-1/n)\approx S_n(g)$$

同样地 $S_n(g+1/n)\approx S_n(g)$. 因此，根据（4），我们得到 $b_y(f)\approx 0$，只要 $f-f_0$相对于$1/n$不太小。同时 $b_y(f_0)=b(f_0)$。因此泄漏减少了。

然而，通常情况下，对于接近 f_0的 f， $b_y(f)$要比 $b(f)$大得多。因此，减少泄漏的代价是峰值在顶部稍微变圆，与没有使用汉宁窗的峰值相比。

5离散傅里叶变换和样本自协方差函数

我们下面展示

$$\frac{|b_j|^2}{n}=\sum_{|h|<n}\hat{\gamma}(h)\exp\left(-\frac{2\pi ijh}{n}\right)\qquad \text{对于 } j=1,\ldots,n-1$$

其中 $\hat{\gamma}(h)$ 是样本自协方差函数。这给出了离散傅里叶变换和样本自协方差函数之间的重要联系。

为了看到这一点，首先观察到，根据等比级数求和公式，有

$$\sum_{t=0}^{n-1}\exp\left(\frac{-2\pi ijt}{n}\right)=0\qquad \text{对于 } j=1,\ldots,n-1。$$

换句话说，如果数据是常数，即 $x_0=\cdots=x_{n-1}$，则 b_0等于 nx_0，而对于其他所有的$_j$， b_j等于0。因此，我们可以写成：

$$b_j=\sum_{t=0}^{n-1}(x_t-\bar{x})\exp\left(\frac{-2\pi ijt}{n}\right)\qquad \text{对于 } j=1,\ldots,n-1。$$

因此，对于 $j=1,\ldots,n-1$，我们写成

$$\begin{aligned}|b_j|^2=b_j\bar{b}_j&=\sum_{t=0}^{n-1}\sum_{s=0}^{n-1}(x_t-\bar{x})(x_s-\bar{x})\exp\left(\frac{-2\pi ijt}{n}\right)\exp\left(\frac{2\pi ijs}{n}\right)\\&=\sum_{t=0}^{n-1}\sum_{s=0}^{n-1}(x_t-\bar{x})(x_s-\bar{x})\exp\left(\frac{-2\pi ij(t-s)}{n}\right)\\&=\sum_{h=-(n-1)}^{n-1}\sum_{t,s:t-s=h}(x_t-\bar{x})(x_{t-h}-\bar{x})\exp\left(-\frac{2\pi ijh}{n}\right)\\&=n\sum_{|h|<n}\hat{\gamma}(h)\exp\left(-\frac{2\pi ijh}{n}\right).\end{aligned}$$

2013年春季统计学153课程（时间序列）：第二十一讲

阿迪蒂亚·冈图博伊纳

2013年4月11日

1 周期图

在上一堂课中，我们看到了DFT和样本自协方差函数之间的以下联系：

$$\frac{|b_j|^2}{n} = \sum_{h:|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi ijh}{n}\right) \qquad \text{对于 } j = 1, \ldots, [n/2].$$

这个函数

$$I(j/n) := \frac{|b_j|^2}{n} = \sum_{h:|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi ijh}{n}\right) \qquad \text{对于 } j = 1, \ldots, [n/2] \tag{1}$$

被称为数据 $x_0, x_1, \ldots, x_{n-1}$的周期图。 周期图给出了数据中各个频率的正弦波强度。

2 谱密度

假设 $\{X_t\}$是一个双重无限随机变量序列，它是平稳的。 让 $\{\gamma(h)\}$表示它们的自协方差函数。 类似于周期图的定义（1），我们定义

$$f(\lambda) := \sum_{h=-\infty}^{\infty} \gamma(h) \exp\left(-2\pi i\lambda h\right) \qquad \text{对于 } -1/2 \leq \lambda \leq 1/2 \tag{2}$$

并将此数量称为随机变量序列 $\{X_t\}$的谱密度。 因为复指数函数$^{-2\pi i\lambda h}$在 λ上都是周期性的，周期为1，所以我们只需要在长度为1的区间上定义 f，并且按照惯例，我们关注的是区间$[-1/2, 1/2]$。 实际上，注意到 f是对称的，我们只需要关注$[0, 1/2]$。

类似于周期图，谱密度将给出数据中不同频率的正弦波的强度。

我们已经根据自协方差函数定义了谱密度。 事实上，自协方差函数也可以从谱密度中获得：为了看到这一点，只需将（^{2}）两边都乘以e的$^{2\pi}$iλk，并从$\lambda =$ -$1/^2$到$\lambda = 1/^2$进行积分，得到：

$$\gamma(k) = \int_{1/2}^{1/2} {}^{e\text{的}2\pi i}\lambda^{k\text{乘以}}{}_{f}\,{}^{(}\lambda^{)}\;{}^{\text{乘以}}{}_{d\lambda} \tag{3}$$

换句话说，自协方差函数和谱密度提供了关于平稳过程$\{X_t\}$的等效信息。

然而，谱密度的定义存在一个问题。在（2）中的无穷和不一定总是有意义的。实际上，复指数exp（-2πiλh）的幅度始终为1，因此当$\{\gamma(h)\}$足够快地衰减时，和（2）才有意义。使（2）有意义的一个充分（但不必要）条件是$\sum_{h=-\infty}^{\infty}|\gamma(h)|<\infty$。

在（2）式右边的求和在一些情况下是没有意义的，但并不太难找到这样的例子。
例如，考虑过程 $X_t = A\cos 2\pi\lambda_1 t + B\sin 2\pi\lambda_1 t$其中 A和 B是不相关的随机变量，均值为0，方差为 σ^2，且$0<\lambda_1<1/2$是一个固定（非随机）频率。
这个过程显然是平稳的，其自协方差函数等于 $\gamma(h) = \sigma^2\cos 2\pi\lambda_1 h$。显然，
这个过程的衰减速度不够快，$\sum_h \gamma(h)\exp(-2\pi i\lambda h)$ 对于任何 λ都没有意义。

事实证明，并不是每个平稳过程 $\{X_t\}$都能定义一个谱密度。
但是人们总是可以定义一个谱分布函数。类比于随机变量（并非所有随机变量都有密度函数，但它们都有分布函数）。

在定义谱分布函数之前，让我们简要讨论一下基本期望。

2.1 期望的回顾

设X是一个随机变量。X的分布函数定义为$F(x)=\mathbb{P}\{X\leq x\}$。函数$F$非负、右连续、非减，并满足以下条件：

$$\text{极限}_{x\to-\infty} F(x)=0\text{，且}\qquad \text{极限}_{x\to\infty} F(x)=1\text{。}$$

对于X的函数$g(X)$的期望有时用以下符号表示：

$$\mathbb{E}[g(X)] = \int g(x)dF(x)\text{。}$$

在以下两种情况下，计算这个期望是相当容易的：

1. X是一个离散随机变量，取值为 $x_1<\cdots<x_k$，概率为 $p_1,\ldots,p_k$。. 在这种情况下，F在 x_i处有一个大小为 p_i的跳跃，并且在 x_i和 x_{i+1}之间是常数。而且，$\int g(x)dF(x) = \sum_i g(x_i)p_i$。
2. X具有密度 f。在这种情况下，$F(x) = \int_{-\infty}^{x} f(x)dx$ 和 $\mathbb{E}g(X) = \int g(x)f(x)dx$。

数量 $\int g(x)dF(x)$也可以用于非负、右连续、非递减且满足以下条件的 F定义：lim

$$\lim_{x\to-\infty} F(x)=0\text{，且}\qquad \text{极限}_{x\to\infty} F(x)=\sigma^2$$

对于一些$\sigma^2>0$。在这种情况下，F/σ^2是一个分布函数，因此积分 $\int g(x)dF(x)$ 可以被定义为

$$\int g(x)dF(x)=\sigma^2\int g(x)d\tilde{F}(x)\qquad \text{其中}\tilde{F}(x)=\frac{F(x)}{\sigma^2}.$$

2.2频谱分布函数和频谱密度

设$\{X_t\}$为一个平稳随机变量序列，$\gamma X(h)=\mathrm{cov}(X_t,X_{t+h})$表示自协方差函数。

由Herglotz（有时归因于Bochner）的定理表明，每个自协方差函数γ_X都可以写成：

$$\gamma_X(h)=\int_{-1/2}^{1/2} e^{2\pi ih\lambda}dF(\lambda),$$

其中 $F(\cdot)$ 是一个非负、右连续、非递减的函数，其定义域为$[-1/2,1/2]$，且 $F(-1/2)=0$ 和 $F(1/2)=\gamma_X(0)$. 此外， F是由 γ_X唯一确定的.

如果 F具有密度 f，则 f被称为 $\{X_t\}$的谱密度.

一个充分条件（但不是必要条件）是谱密度的存在条件 $\sum$在这种情况下，谱密度存在且由公式（2）给出.

2.3离散频谱示例

假设

$$X_t=\sum_{j=1}^{m}(A_j\cos(2\pi\lambda_jt)+B_j\sin(2\pi\lambda_jt)) \qquad \text{对于 } t=\ldots,-2,-1,0,1,2,\ldots,$$

其中频率$0<\lambda_1<\cdots<\lambda_m<1/2$是固定的， $A_1,B_1,A_2,B_2,\ldots,A_m,B_m$是不相关的随机变量，具有公共均值0和$\mathrm{var}(A_j)=\sigma_j^2=\mathrm{var}(B_j)$。 X_t和

X_{t+h}之间的协方差为：

$$\sum_j\sigma_j^2\left(\cos(2\pi\lambda_jt)\cos(2\pi\lambda_j(t+h))+\sin(2\pi\lambda_jt)\sin(2\pi\lambda_j(t+h))\right)=\sum_j\sigma_j^2\cos(2\pi\lambda_jh).$$

因为这个协方差不依赖于 t，所以过程 $\{X_t\}$是平稳的，具有自协方差函数
这个自协方差函数$\gamma_X(h)$ 可以写成：

$$\gamma_X(h)=\sum_{j=1}^{m}\sigma_j^2\left(\frac{e^{2\pi i\lambda_jh}+e^{-2\pi i\lambda_jh}}{2}\right)$$

因此 $\gamma_X(h)$ 等于 $\int_{-1/2}^{1/2}e^{2\pi ih\lambda}dF(\lambda)$ 其中 F对应于取值

$$-\lambda_m<\cdots<-\lambda_1<\lambda_1<\cdots<\lambda_m$$

带权重

$$\frac{\sigma_m^2}{2},\ldots,\frac{\sigma_1^2}{2},\frac{\sigma_1^2}{2},\ldots,\frac{\sigma_m^2}{2}.$$

请注意，这是一个对称分布。 因此，谱分布函数只在存在于 $\{X_t\}$的频率上有质量。 此外，特定频率 λ_j上的质量与该频率上的方差 σ_j^2成比例。 谱分布的总质量为：

$$\frac{\sigma_m^2}{2}+\cdots+\frac{\sigma_1^2}{2}+\frac{\sigma_1^2}{2}+\cdots+\frac{\sigma_m^2}{2}=\sigma_1{}^2+\cdots+\sigma_m^2=\gamma_X(0)\text{。}$$

2.4 白噪声

对于白噪声，当$h=0$时，$\gamma_X(h)=0$，且 $\gamma_X(0)=\sigma^2$。因此， $\sum|\gamma_X(h)|<\infty$，谱密度由以下公式给出：

$$f(\lambda)=\sum_{h=-\infty}^{\infty}\gamma_X(h)\exp(-2\pi i\lambda h)=\gamma_X(0)=\sigma^2 \qquad \text{对于所有的}-1/2\le\lambda\le1/2.$$

这个想法是所有频率在白噪声中以相等的数量存在。

2.5 ARMA过程的谱密度

定理**2.1**. 让$\{Y\}$是一个均值为零的平稳过程，其谱分布函数为FY。
定义$\{X\}$为

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \qquad \text{其中} \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

那么$\{X\}$是平稳的，其谱分布函数为：

$$F_X(\lambda) = \int_{-1/2}^{\lambda} \left| \sum_j \psi_j e^{2\pi i j \lambda} \right|^2 dF_Y(\lambda). \qquad \text{对于} -1/2 \leq \lambda \leq 1/2.$$

证明.X_t的自协方差（注意X_t的均值为零，因为$\{Y\}$被假设为零均值）：

$$\gamma_X(h) = \mathbb{E}X_t X_{t+h} = \mathbb{E}(\sum_j \psi_j Y_{t-j})(\sum_k \psi_k Y_{t+h-k}) = \sum_{j,k} \psi_j \psi_k \gamma_Y(h-k+j).$$

根据谱分布函数的定义，我们可以写成：

$$\gamma_Y(h-k+j) = \int_{-1/2}^{1/2} e^{2\pi i(h-k+j)\lambda} dF_Y(\lambda).$$

因此，

$$\begin{aligned}
\gamma_X(h) &= \sum_{j,k} \psi_j \psi_k \int_{-1/2}^{1/2} e^{2\pi i(h-k+j)\lambda} dF_Y(\lambda) \\
&= \int_{-1/2}^{1/2} e^{2\pi i h \lambda} \sum_{j,k} \psi_j \psi_k e^{-2\pi i k \lambda} e^{2\pi i j \lambda} dF_Y(\lambda) \\
&= \int_{-1/2}^{1/2} e^{2\pi i h \lambda} \left(\sum_j \psi_j e^{2\pi i j \lambda} \right) \left(\sum_k \psi_k e^{-2\pi i k \lambda} \right) dF_Y(\lambda) \\
&= \int_{-1/2}^{1/2} e^{2\pi i h \lambda} \left| \sum_j \psi_j e^{2\pi i j \lambda} \right|^2 dF_Y(\lambda)
\end{aligned}$$

这是一个形式：

$$\gamma_X(h) = \int_{-1/2}^{1/2} e^{2\pi i h \lambda} dF_X(\lambda)$$

与

$$dF_X(\lambda) = \left| \sum_j \psi_j e^{2\pi i j \lambda} \right|^2 dF_Y(\lambda).$$

证明完成。 □

由上述定理可知，如果 $\{Y_t\}$具有谱密度 f_Y，则 X_t也具有谱密度，其表示为

$$f_X(\lambda) = \left| \sum_{j=-\infty}^{\infty} \psi_j e^{2\pi i j \lambda} \right|^2 f_Y(\lambda) \qquad \text{对于} -1/2 \leq \lambda \leq 1/2.$$

如果我们使用符号 $\psi(z)=\sum_{j=-\infty}^{\infty}\psi_j z^j$，那么谱密度 $f_X(\lambda)$可以表示为：

$$f_X(\lambda)=|\psi(e^{2\pi i\lambda})|^2 f_Y(\lambda)$$

2013年春季统计学153课程（时间序列）：第二十二讲

阿迪蒂亚・冈图博伊纳

2013年4月16日

1 光谱分布函数

设$\{X_t\}$为一个平稳随机变量序列，γX（h）$=\mathrm{cov}(X_t, X_{t+}h)$表示自协方差函数。

由Herglotz（有时归因于Bochner）的定理表明，每个自协方差函数γ_X都可以写成：

$$\gamma_X(h) = \int_{-1/2}^{1/2} e^{2\pi i h\lambda} dF(\lambda),$$

其中 F（$\cdot$）是一个非负、右连续、非递减的函数，其定义域为$[-1/2, 1/2]$，且 F（$-1/2$）$=0$ 和 F（$1/2$）$=\gamma_X$（0）. 此外， F是由 γ_X唯一确定的.

这个函数 F 被称为 $\{X_t\}$的光谱分布函数。如果 F 有密度 f即 F can be written as

$$F(x) := \int_{-1/2}^{x} f(t)dt$$

那么 f被称为 $\{X_t\}$的光谱密度。

一个充分条件（但不是必要条件）是谱密度的存在条件 $\sum$ 在这种情况下，光谱密度存在并且由以下公式给出:

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-2\pi i\lambda h) \qquad \text{对于} -1/2 \le \lambda \le 1/2.$$

对于一个平稳过程来说，光谱分布函数和自协方差函数一样重要。

2 线性时不变滤波器

线性时不变滤波器使用一组指定的系数 $\{a_j\}$ for j =. . . to 根据以下公式将输入时间序列 $\{X_t\}$转换为输出时间序列 $\{Y_t\}$:

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}.$$

滤波器由系数 $\{a_j\}$决定，通常假设满足 $\sum_{j=-\infty}^{\infty} |a_j| < \infty$。

假设输入序列 $\{X_t\}$给定为

$$X_t = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}$$

这样的 $\{X_t\}$通常被称为冲激函数。 滤波器的输出 $\{Y_t\}$可以很容易地看出是 $Y_t = a_t$。 因此，滤波器系数 $\{a_j\}$通常被称为冲激响应函数。

到目前为止，我们已经看到了两个线性时不变滤波器的主要示例：（1）移动平均滤波器，其冲激响应函数为： $a_j= 1/(2q+ 1)$，其中 $|j| \leq q$，其他情况下 $a_j= 0$；（2）差分滤波器，对应于滤波器 $a_0= 1$，$a_1 = -1$，其他 a_j均为零。 我们已经看到这两个滤波器的作用非常不同；一个估计趋势，而另一个消除趋势。

假设输入时间序列 $\{X_t\}$是平稳的，自协方差函数为 γ_X。 那么 $\{Y_t\}$的自协方差函数是什么？观察到

$$\gamma_Y(h) := \text{cov}\left(\sum_j a_j X_{t-j}, \sum_k a_k X_{t+h-k}\right) = \sum_{j,k} a_j a_k \text{cov}(X_{t-j}, X_{t+h-k}) = \sum_{j,k} a_j a_k \gamma_X(h-k+j). \quad (1)$$

注意上述计算还表明 $\{Y_t\}$是平稳的。

现在假设输入平稳序列 $\{X_t\}$的谱密度是 f_X。 那么输出 $\{Y_t\}$的谱密度 fY是什么？

因为 $\{X_t\}$的谱密度等于 f_X，我们有

$$\gamma_X(h) = \int_{-1/2}^{1/2} e^{2\pi i h\lambda} f_X(\lambda) d\lambda.$$

因此，根据（1），我们有

$$\gamma_Y(h) = \sum_j \sum_k a_j a_k \int e^{2\pi i(h-k+j)\lambda} f_X(\lambda) d\lambda = \int e^{2\pi i h\lambda} f_X(\lambda) \left(\sum_j \sum_k a_j a_k e^{-2\pi i k\lambda} e^{2\pi i j\lambda}\right) d\lambda \quad (2)$$

现在让我们定义函数

$$A(\lambda) := \sum_j a_j e^{-2\pi i j\lambda} \qquad \text{对于} -1/2 \leq \lambda \leq 1/2.$$

注意，这个函数只依赖于滤波器系数 $\{a_j\}$. 根据（2），很明显有

$$\gamma_Y(h) = \int e^{2\pi i\lambda h} f_X(\lambda) A(\lambda) \overline{A(\lambda)} d\lambda,$$

在这里，当然，$\overline{A(\lambda)}$ 表示$A(\lambda)$的复共轭。因此，我们有

$$\gamma_Y(h) = \int e^{2\pi i\lambda h} f_X(\lambda) |A(\lambda)|^2 d\lambda$$

这显然是$\gamma_Y(h)$的形式= $\int e^{2\pi i\lambda h} f_Y(\lambda) d\lambda$。因此，我们有

$$f_Y(\lambda) = f_X(\lambda) |A(\lambda)|^2 \qquad \text{对于} -1/2 \leq \lambda \leq 1/2. \quad (3)$$

换句话说，滤波器对输入的频谱的作用非常容易解释。 它通过与函数$|A(\lambda)|^2$相乘来修改频谱。 根据$|A(\lambda)|^2$的值，输出中的某些频率可能会增强，而其他频率则会减弱。

这个函数 $\lambda \to |A(\lambda)|$^2 被称为滤波器的功率传递函数。 这个函数 $\lambda \to A(\lambda)$ 被称为滤波器的传递函数或频率响应函数。

频谱密度在研究滤波器的性质时非常有用。 虽然输出序列的自协方差函数 γ_Y 在很复杂的方式上依赖于输入序列的自协方差函数 γ_X，但两个频谱密度之间的依赖关系非常简单。

例子 **2.**1（差分滤波器的功率传递函数）。考虑滞后 s 的差分滤波器：$Y\_t = X\_t$ - $X\_\{t$-$s\}$，其中权重 $a\_0 = 1$，$a\_s =$ -1，对于其他所有的 j，$a\_j = 0$。 那么传递函数显然可以表示为

$$A(\lambda) = \sum_j a_j e^{-2\pi ij\lambda} = 1 - e^{-2\pi is\lambda} = 2i\sin(\pi s\lambda)e^{-\pi is\lambda},$$

在最后一个等式中，使用了公式 $1 - e^{i\theta} = -2i\sin(\theta/2)e^{i\theta/2}$。 因此，功率传输函数等于

$$|A(\lambda)|^2 = 4\sin^2(\pi s\lambda) \qquad \text{对于} -1/2 \leq \lambda \leq 1/2.$$

要理解这个函数，我们只需要考虑区间 $[0,1/2]$，因为它在
$[-1/2,1/2]$上对称。

当 $s = 1$时，函数 $\lambda \to |A(\lambda)|^2$在 $[0,1/2]$上是递增的。 这意味着一阶差分会增强数据中的高频率并减弱低频率。 因此，它会使数据更加波动。

对于较高的s值，函数$A(\lambda)$上下波动，并且在λ=$0, 1/s, 2/s$等取零值。 换句话说，它消除了所有周期为s的分量。

例**2.2.** 现在考虑移动平均滤波器，其对应的系数$a_j = 1/(\mathbf{2}q+1)$（$|j| \leq q$）。传递函数为

$$\frac{1}{2q+1}\sum_{j=-q}^{q} e\text{\^{}}(-2\pi ij\lambda) = \frac{S_{q+}1(\lambda) + S_{q+}1(-\lambda) - 1}{2q+1},$$

其中可以回忆起（第19讲）

$$S_n(g) := {}^{n}\Sigma^{-1}_{t=0} \exp(2\pi igt) = \frac{\sin(\pi ng)}{\sin(\pi g)} e^{i\pi g(n-1)}$$

。因此

$$S_n(g) + S_n(-g) = 2\frac{\sin(\pi ng)}{\sin(\pi g)}\cos(\pi g(n-1)),$$

这意味着传递函数由以下给出

$$A(\lambda) = \frac{1}{2q+1}\left(2\frac{\sin(\pi(q+1)\lambda)}{\sin(\pi\lambda)}\cos(\pi q\lambda) - 1\right),$$

这个函数只依赖于 q，并且可以绘制不同 q的值。 对于大的 q，它会迅速降至零。 解释是滤波器消除了输入过程中的高频成分。

3 ARMA过程的谱密度

假设 $\{X_t\}$是一个平稳的ARMA过程： ϕ（B）$X_t = \theta$（B）Z_t其中多项式 ϕ和 θ have在单位圆上没有公共零点。 由于平稳性，多项式 ϕ在单位圆上没有根。

首先，让我们写下 U_t的谱密度，用 $\{X_t\}$的谱密度来表示。 显然， U_t可以看作是应用于X_t的一个滤波器的输出。 该滤波器由以下形式给出：$a_0 = 1$，对于$1 \leq_j \leq p$， $a_j = -\phi j$，对于其他所有的$_j$， $a\ j = 0$。 让 $A\ \phi(\lambda)$表示该滤波器的传递函数。那么我们有

$$f_U(\lambda) = |A_\phi(\lambda)|^2 f_X(\lambda)。$$

同样地，利用 $U_t = \theta(B)Z_t$的事实，我们可以写成

$$f_U(\lambda) = |A_\theta(\lambda)|^2 f_Z(\lambda) = \sigma_Z^2 |A_\theta(\lambda)|^2 \tag{5}$$

其中 $A_\theta(\lambda)$ 是具有系数 a_0= 1 和 $a_j = \theta_j$（$1 \leq j \leq q$）的滤波器的传递函数，对于其他所有的j，a_j= 0。 将（4）和（5）等式化，我们得到

$$f_X(\lambda) = \frac{|A_\theta(\lambda)|^2}{|A_\phi(\lambda)|^2}\sigma_Z^2 \qquad 对于 -1/2 \leq \lambda \leq 1/2.$$

现在

$$A_\phi(\lambda) = 1 - \phi_1 e^{-2\pi i\lambda} - \phi_2 e^{-2\pi i(2\lambda)} - \cdots - \phi_p e^{-2\pi i(p\lambda)} = \phi(e^{-2\pi i\lambda})。$$

同样地 $A_\theta(\lambda) = \theta(e^{-2\pi i\lambda})$。因此，我们有

$$f_X(\lambda) = \sigma_Z^2 \frac{|\theta(e^{-2\pi i\lambda})|^2}{|\phi(e^{-2\pi i\lambda})|^2} \qquad 对于 -1/2 \leq \lambda \leq 1/2.$$

注意，由于平稳性，上述右侧的分母对于所有的 λ都不为零。

例子3.1 (MA(1))。对于*MA(***1***)*过程： $X_t = Z_t + \theta Z_{t-1}$，我们有 $\phi(z) = 1$和 $\theta(z) = 1 + \theta z$。因此

$$\begin{aligned} f_X(\lambda) &= \sigma_Z^2 \left|1 + \theta e^{2\pi i\lambda}\right|^2 \\ &= \sigma_Z^2 |1 + \theta\cos 2\pi\lambda + i\theta\sin 2\pi\lambda|^2 \\ &= \sigma_Z^2 \left[(1 + \theta\cos 2\pi\lambda)^2 + \theta^2\sin^2 2\pi\lambda\right] \\ &= \sigma_Z^2 \left[1 + \theta^2 + 2\theta\cos 2\pi\lambda\right] \qquad 对于 -1/2 \leq \lambda \leq 1/2. \end{aligned}$$

检查当 $\theta = -1$时，数量 $1 + \theta_2 + 2\theta\cos(2\pi\lambda)$等于第一个差分滤波器的功率传递函数。

例子3.2（AR(1)）。 对于AR(1)： $X_t - \phi X_{t-1} = Z_t$，我们有 $\phi(z) = 1 - \phi z$和 $\theta(z) = 1$。因此

$$f_X(\lambda) = \sigma_Z^2 \frac{1}{|1 - \phi e^{2\pi i\lambda}|^2} = \frac{\sigma_Z^2}{1 + \phi^2 - 2\phi\cos 2\pi\lambda} \qquad 对于 -1/2 \leq \lambda \leq 1/2.$$

例子3.3（AR(2)）.对于AR(2)模型： $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t$，我们有 $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ 和 $\theta(z) = 1$. 这里可以证明

$$f_X(\lambda) = \frac{\sigma_Z^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2)\cos 2\pi\lambda - 2\phi_2\cos 4\pi\lambda} \qquad 对于 -1/2 \leq \lambda \leq 1/2.$$

2013年春季统计学153课程（时间序列）：第二十三讲

阿迪蒂亚・冈图博伊纳

2013年4月18日

1非参数估计的谱密度

设 $\{X_t\}$为一个平稳过程，满足 $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$. 我们已经看到 $\{X_t\}$有一个由以下给出的谱密度

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-2\pi i \lambda h} \qquad 对于 -1/2 \leq \lambda \leq 1/2. \tag{1}$$

现在假设我们有数据 $x_1, \ldots$， x_n来自过程 $\{X_t\}$. 那么我们如何在不对底层过程做任何参数假设的情况下估计 f（λ）呢？这是我们下一个主题。

为什么我们要非参数地估计谱密度?

当我们对数据拟合ARMA模型时，我们首先观察样本自协方差函数或自相关函数，然后尝试找到与样本自协方差函数匹配的理论ARMA模型。现在样本自协方差函数是过程的非参数估计。 换句话说，我们首先通过非参数方法估计 γ（h）

$\hat{\gamma}$（h）然后找到一个与 γ（h）接近的ARMA模型的γ_{ARMA}（h）。

如果我们可以非参数地估计谱密度，我们可以类似地使用估计来选择参数模型。 我们简单地选择谱密度最接近非参数估计的ARMA模型。

估计谱密度的另一个原因来自于估计滤波器系数的问题。 假设我们知道两个过程$\{X_t\}$和$\{Y_t\}$通过一个线性时不变滤波器相互关联。 换句话说，当$\{X_t\}$是滤波器的输入时，$\{Y_t\}$是输出。 假设我们不知道滤波器系数，但我们有来自输入和输出过程的观测数据。 目标是估计滤波器。 在这种情况下，一个自然的策略是从数据中估计fX和fY的谱密度，然后使用$fY(\lambda) = fX(\lambda)|A(\lambda)|$^2来获得滤波器功率传递函数的估计（要获得传递函数本身的估计，需要使用交叉谱）。 这是谱分析的一个应用之一。 我们可能无法对$\{X_t\}$和$\{Y_t\}$进行参数化假设，因此非参数地估计谱密度是有意义的。

非参数估计谱密度比非参数估计自协方差函数更复杂。 主要原因是自然估计量效果不好。

由于公式（1）给出了谱密度与自协方差函数$\gamma_X(h)$的关系，估计$f(\lambda)$的一个自然想法是用其估计量$\hat{\gamma}$ (h)代替$\gamma_X(h)$，其中$|h|$<n（对于$|h|$>n不可能估计$\gamma(h)$）。这将得到估计量：

$\gamma\hat{}(h)$ = e^(-2πiλh)

$$I(\lambda) = \sum_{h:|h|<n} \text{当λ=j/n∈(0,1/2]时，上述量正好是周期图：} \qquad \text{对于}-1/2 \leq \lambda \leq 1/2.$$

当 $\lambda = j/n \in (0, 1/2]$时，上述数量就是周期图：

$$I(j/n) = \frac{|b_j|^2}{n} \qquad \text{在哪里 } b_j = \sum_t x\_t \exp\left(\frac{-2\pi ijt}{n}\right)$$

不幸的是，我（λ）不是 f_X的一个好估计。 这可以通过模拟很容易地看出。 只需从白噪声中生成数据并观察到周期图非常波动，而真实的谱密度是恒定的。 事实上，我（λ）是一个糟糕的估计量，这也可以通过以下方式在数学上验证。假设数据 x_t是由方差为 σ^2的高斯白噪声生成的（它们的均值为零，因为

它们是白噪声）。 对于 $j/n \in [0, 1/2]$， $|b_j|^2/n$的分布是什么？写出

$$\begin{aligned}
\frac{|b_j|^2}{n} &= \frac{1}{n}\left|\sum_{t=0}^{n-1} x_t \exp\left(\frac{-2\pi ijt}{n}\right)\right|^2 \\
&= \frac{1}{n}\left|\sum_t x_t \cos(2\pi jt/n) - i\sum x_t \sin(2\pi jt/n)\right|^2 \\
&= \frac{1}{n}\left(A_j^2 + B_j^2\right),
\end{aligned}$$

其中

$$A_j = \sum_t x_t \cos(2\pi jt/n)\text{和} \quad B_j = \sum_t x_t \sin(2\pi jt/n).$$

如果我们还假设 $x_1, \ldots\ldots\ldots\ldots, x_n$, 那么$(A_j, B_j)$ 是联合正态分布的

$$\text{var}A_j = \sigma^2 \sum_{t=0}^{n-1} \cos^2(2\pi jt/n) \text{ 和 var } \quad B_j = \sigma^2 \sum_{t=0}^{n-1} \sin^2(2\pi jt/n).$$

另外

$$\text{cov}(A_j, B_j) = \sigma^2 \sum_{t=0}^{n-1} \cos(2\pi jt/n)\sin(2\pi jt/n).$$

可以验证

$$\begin{aligned}
\sum_{t=0}^{n-1} \cos^2(2\pi jt/n) &= n \qquad \text{当 } j\text{为0或 } n/2\text{时} \\
&= n/2 \qquad \text{当 } j\text{既不是0也不是 } n/2\text{时}.
\end{aligned}$$

和

$$\begin{aligned}
\sum_{t=0}^{n-1} \sin^2(2\pi jt/n) &= 0 \qquad \text{当 } j\text{为0或 } n/2\text{时} \\
&= n/2 \qquad \text{当 } j\text{既不是0也不是 } n/2\text{时}.
\end{aligned}$$

和

$$\sum_{t=0}^{n-1} \cos(2\pi jt/n)\sin(2\pi jt/n) = 0.$$

因此，当j既不是0也不是n/2（回想一下0≤j/n≤1/2）时，我们有

$$\frac{\sqrt{2}A_j}{\sigma\sqrt{n}} \sim N(0,1)\text{和} \qquad \frac{\sqrt{2}B_j}{\sigma\sqrt{n}} \sim N(0,1)$$

这意味着

$$\frac{2}{n\sigma^2}A_j^2 \sim \chi_1^2 \quad \text{和} \quad \frac{2}{n\sigma^2}B_j^2 \sim \chi_1^2.$$

另外，因为它们是独立的，所以对于j/n$\in$(0,1/2)，我们有

$$\frac{2}{\sigma^2}I(j/n) = \frac{2|b_j|^2}{n\sigma^2} = \frac{2}{n\sigma^2}A_j^2 + \frac{2}{n\sigma^2}B_j^2 \sim \chi_2^2$$

或者$I(j/n) \sim (\sigma$^$^2/^2)\chi$^$^{2\ 2}$.

对于 j= 0或 $n/2$，我们有 B_j= 0和 $A_j \sim N(0, \sigma^2 n)$，这意味着 $|b_j|^2/n \sim \sigma^2\chi^2{}_1$。

重要的是要注意 $I(j/n)$的分布不依赖于 n。 人们还可以检查
对于$_j =_{j'}$，(A_j, B_j)与$(A\ j_{'}, B\ j')$是独立的。

因此，当数据x_1，...，x_n来自高斯白噪声模型时，周期图坐标$I(j/_n)$对于$0 < j \leq {}_n/^2$是独立的随机变量，其分布为$(\sigma$^$^2/^2)\chi$^$^{2\ 2}$对于$0 < j < {}_n/^2$和σ^$^2\chi$^$^2{}_1$对于$j = {}_n/^2$。 由于这种独立性和分布不依赖于$_n$，很明显$I(\lambda)$不是$f(\lambda)$的良好估计。 因为这种独立性和分布不依赖于$_n$，很明显$I(\lambda)$不是$f(\lambda)$的良好估计。

我们已经完成了对高斯白噪声数据的上述计算。 对于一般的ARMA过程，在一些正则条件下，可以证明当n趋于无穷大时，随机变量：

$$\frac{2I(j/n)}{f(j/n)}, \qquad \text{对于}0< j < n/2$$

近似独立地服从 $\chi^{2\ 2}$分布。

请注意，因为 $\chi^{2\ 2}$分布的均值为2，所以 $I(j/n)$的期望值近似为
$f(j/n)$。 换句话说，周期图近似无偏。 另一方面，
$I(j/n)$的方差近似为 $f^2(j/n)$。 因此，在高斯白噪声的情况下，例如，周期图坐标的方差为 σ^4，这并不随n的增大而减小。 这以及相邻周期图坐标的近似独立性使得周期图非常嘈杂，成为真实谱密度的一个糟糕的估计器。

2 修改周期图以获得谱密度的良好估计

2.1 方法一

近似分布结果使我们能够写成：

$$\frac{2I(j/n)}{f(j/n)} \approx 2 + 2U_j \qquad \text{对于 } 0< j < n/2,$$

其中 $U_1, U_2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots$ 是独立的，均值为零，方差为1。. 换句话说 $\{U_j\}$是白噪声。因此

$$I(j/n) = f(j/n) + U_j f(j/n) \qquad \text{对于 } 0< j < n/2.$$

因此，我们可以将 $I(j/n)$ 视为一个具有趋势 $f(j/n)$ 的不相关时间序列，我们希望估计。 我们以前对趋势估计的经验表明，我们可以通过使用一个滤波器（例如简单移动平均滤波器）对 $I(j/n)$ 进行平滑来实现这一点：

$$\frac{1}{2m+1}\sum_{k=-m}^{m} I\left(\frac{j+k}{n}\right).$$

更一般地，我们也可以考虑使用不等权重来得到以下形式的估计量：

$$\hat{f}(j/n) = \sum_{k=-m_n}^{m_n} W_n(k) I\left(\frac{j+k}{n}\right).$$

注意，如果我们取 $m_n = 0$，我们将得到周期图。 我们可以将$\hat{f}$的这个定义扩展到整个区间$[0, 1/2]$，具体方法如下：对于每个 $\lambda \in [0, 1/2]$，让 $g(\lambda, n)$表示最接近 λ的$1/\mathrm{n}$的倍数。定义

$$\hat{f}(\lambda) = \hat{f}(g(\lambda, n)).$$

可以证明这个估计量是一致的（即，随着n的增大，它越来越接近$f(j/n)$），只要：

$$\text{当}m\text{和}_n\text{趋向于无穷大时 当}_n\text{趋向于0时} \tag{2}$$

当n趋向于无穷大时 还需要权重$W(n)(k)$是对称的：$W(n)(k) = W(n)(\text{-}k)$，非负的$W(n)(k) \geq 0$，总和为1：$\sum^{m}(k) = n\text{-}^{m}_{n} W(_n)(k) = 1$，并且它们的平方和趋向于零：$\sum^{m}(k) = n\text{-}^{m}_{n}$ 当$_n$趋向于无穷大时，$W(_n)$^$^2(k)$趋向于零 当n趋向于无穷大时 请注意，所有这些条件都对简单移动平均滤波器成立，其中m选择如(2)中所示。

如果满足上述条件，则对于$0 < \lambda < 1/2$，我们有

$$\mathbb{E}\hat{f}(\lambda) \approx f(\lambda)\text{和}\mathrm{var}(\hat{f}(\lambda)) \approx \left(\sum_{k=-m_n}^{m_n} W_n^2(k)\right) f^2(\lambda).$$

当 λ等于0或$1/2$时，方差是上述方程给出的两倍。 期望值
仍然相同。

此外，$\hat{f}(\lambda_1)$和$\hat{f}(\lambda_2)$之间的协方差近似为零。

2.2 第二种方法

这是一种稍微不同的方法，用于得出与周期图不同的谱密度估计器。 周期图的定义为：

$$I(\lambda) = \sum_{h:|h|<n} \hat{\gamma}(h) \exp(-2\pi i \lambda h) \qquad \text{对于} -1/2 \leq \lambda \leq 1/2. \tag{3}$$

请注意，上述公式涉及到所有自协方差的估计 $\gamma_X(h)$ 对于 $|h| < n$. 现在我们知道，从一个大小为 n的样本中，很难得到对于h接近 n的 $\gamma_X(h)$ 的良好估计。 这经常被引用为周期图不是一个好的估计器的原因。 基于这个原因，获得更好的估计器的一个合理方法是通过在（3）的右侧截断求和，省略h接近 n的$\hat{\gamma}(h)$。 换句话说，我们考虑

$$\tilde{f}(\lambda) = \sum_{h:|h|\leq r} \hat{\gamma}(h) \exp(-2\pi i \lambda h).$$

如果我们假设$r = r(_n)$是$_n$的函数，使得$r(_n)$趋近于无穷大且$r(_n)/_n$趋近于0，那么f~是由$(2r+1)$个方差约为$1/_n$的项的和。 在这种情况下，在正则条件下，可以证明f~是f的一致估计量。

更一般地，我们可以取

$$\tilde{f}(\lambda) = \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \gamma\text{^}(h) \exp(\text{-}2\pi i \lambda h),$$

其中$w(x)$是一个对称函数，满足$w(0) = 1$，$|w(x)| \leq 1$且$w(x) = 0$，当$|x| > 1$。 这有时被称为滞后窗口谱密度估计器。

2.3 两种方法的等价性

我们将展示这两种改进周期图的方法：平滑和滞后窗口
谱密度估计器本质上是相同的。 为了看到这一点，我们首先需要一个I(λ) 和 ^γ(h) 之间的反关系
。 我们将 $I(\lambda)$ 定义为

$$I(\lambda) := \sum_{h:|h|<n} \text{当}\lambda\text{=j/n}\in(0,1/2]\text{时，上述量正好是周期图:} \qquad \text{对于}-1/2\leq\lambda\leq 1/2.$$

可以通过反转这个公式来用 $I(\lambda)$ 来表示 $\hat{\gamma}(k)$。 固定一个整数 k，使得 $|k|< n$，并将上述公式的两边都乘以 $e^{2\pi i\lambda k}$。 将得到的表达式对λ从 $-1/2$ 积分到 $1/2$，我们得到

$$\int_{-1/2}^{1/2} e^{2\pi i\lambda k} I(\lambda) d\lambda = \sum_{h:|h|<n} \hat{\gamma}(h) \int_{-1/2}^{1/2} e^{2\pi i\lambda(k-h)} d\lambda = \hat{\gamma}(k).$$

因此，这意味着

$$\hat{\gamma}(k) = \int_{-1/2}^{1/2} e^{2\pi i\lambda k} I(\lambda) d\lambda. \tag{4}$$

换句话说，函数 $I(\lambda)$ 是对应于样本自相关函数的谱密度。 使用公式（4），我们可以将滞后窗口谱密度估计器写为

$$\begin{aligned}
\tilde{f}(\lambda) &= \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \text{当}\lambda\text{=j/n}\in(0,1/2]\text{时，上述量正好是周期图:} \\
&= \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \int_{-1/2}^{1/2} e^{2\pi i\rho h} I(\rho) d\rho \; e^{-2\pi i\lambda h} \\
&= \int_{-1/2}^{1/2} I(\rho) \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi i(\rho-\lambda)h} d\rho.
\end{aligned}$$

通过变量变换 $\rho = \lambda + u$，我们得到

$$\tilde{f}(\lambda) = \int I(\lambda+u) \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi iuh} du.$$

让我们

$$W(u) = \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi iuh},$$

我们得到

$$\tilde{f}(\lambda) = \int I(\lambda+u) W(u) du.$$

因此，滞后窗口谱密度估计器$\tilde{f}$ 也可以看作是通过平滑周期图获得的。

2013年春季统计学153课程（时间序列）：第二十四讲

阿迪蒂亚・冈图博伊纳

2013年4月23日

在上一节课中，我们讨论了非参数估计谱密度的问题。 自然估计器是：

$$I(\lambda) := \sum_{h:|h|<n} \text{当}\lambda\text{=j/n}\in(0,1/2]\text{时，上述量正好是周期图：} \qquad \text{对于 } -1/2 \leq \lambda \leq 1/2$$

对于 $\lambda = j/n \in (0, 1/2]$，它与周期图相一致：

$$I(j/n) = \frac{|b_j|^2}{n} \qquad \text{在哪里 } b_j = \sum_t x_{\_t} \exp\left(\frac{\text{-}2\pi ijt}{n}\right)$$

周期图的关键结果是，在一些正则条件下，对于所有ARMA过程下的高斯噪声，当 n很大时，随机变量
：

$$\frac{2I(j/n)}{f(j/n)} \qquad \text{对于}0< j < n/2$$

近似独立地服从 χ^2 2分布。 因此， $I(\lambda)$ 不是 $f(\lambda)$ 的一个好的估计器。

我们研究了两种周期图的修改：

1.移动平均平滑：选择一个整数 $m \geq 1$，并通过以下方式估计 $f(j/n)$

$$\hat{f}(j/n) := \frac{1}{2m+1} \sum_{k=-m}^{m} I\left(\frac{j+k}{n}\right)$$

或者更一般地

$$\hat{f}(j/n) := \sum_{k=-m}^{m} W(k) I\left(\frac{j+k}{n}\right)$$

其中 $W(k)$ 是非负权重，总和为一。该估计基于近似表示 $I(j/n) \approx f(j/n) + U_j f(j/n)$，其中 $\{U_j\}$ 是白噪声，对于 $0< j < n/2$。

2.滞后窗口谱密度估计：选择一个整数 $r \geq 1$，并通过以下方式估计 $f(j/n)$

$$\hat{f}(j/n) := \sum_{h:|h|\leq r} \hat{\gamma}(h) \exp\left(-2\pi i\lambda h\right)$$

或者更一般地

$$\hat{f}(j/n) := \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \hat{\gamma}(h) \exp\left(-2\pi i\lambda h\right)$$

其中 $w(x)$ 是对称的，即 $w(x) = w(-x)$，满足 $w(0) = 1$， $|w(x)| \leq 1$，且 $w(x) = 0$
对于 $|x| > 1$。 该估计基于以下公式：

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\lambda h}$$

并且对于大的h，$\gamma(h)$趋近于零（因为$\sum_h |\gamma(h)| < \infty$)，而且它们也很难从数据中估计出来
。

这两个估计量的等价性

我们将展示这两种改进周期图的方法：平滑和滞后窗口
谱密度估计器本质上是相同的。 为了看到这一点，我们首先需要一个I(λ) 和 ˆγ(h) 之间的反关系
。 我们将 $I(\lambda)$ 定义为

$$I(\lambda) := \sum_{h:|h|<n} \text{当}\lambda=j/n\in(0,1/2]\text{时，上述量正好是周期图：} \qquad \text{对于}-1/2 \leq \lambda \leq 1/2.$$

可以通过反转这个公式来用 $I(\lambda)$ 来表示 $\hat{\gamma}(k)$。 固定一个整数 k，使得 $|k| < n$，并将上述公式的两边都乘以 $e^{2\pi i\lambda k}$。 将得到的表达式对λ从 $-1/2$ 积分到 $1/2$，我们得到

$$\int_{-1/2}^{1/2} e^{2\pi i\lambda k} I(\lambda) d\lambda = \sum_{h:|h|<n} \hat{\gamma}(h) \int_{-1/2}^{1/2} e^{2\pi i\lambda(k-h)} d\lambda = \hat{\gamma}(k).$$

因此，这意味着

$$\hat{\gamma}(k) = \int_{-1/2}^{1/2} e^{2\pi i\lambda k} I(\lambda) d\lambda. \tag{1}$$

换句话说，函数 $I(\lambda)$是对应于样本自相关函数的谱密度。 使用公式（1），我们可以将滞后窗口谱密度估计器写成

$$\begin{aligned}
\tilde{f}(\lambda) &= \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \text{当}\lambda=j/n\in(0,1/2]\text{时，上述量正好是周期图：} \\
&= \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \int_{-1/2}^{1/2} e^{2\pi i\rho h} I(\rho) d\rho \ e^{-2\pi i\lambda h} \\
&= \int_{-1/2}^{1/2} I(\rho) \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi i(\rho-\lambda)h} d\rho.
\end{aligned}$$

通过变量变换 $\rho = \lambda + u$，我们得到

$$\tilde{f}(\lambda) = \int I(\lambda + u) \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi i u h} du.$$

让我们

$$W(u) = \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi i u h},$$

我们得到

$$\tilde{f}(\lambda) = \int I(\lambda + u) W(u) du.$$

因此，滞后窗口谱密度估计器$\tilde{f}$ 也可以看作是通过平滑周期图获得的。

对于 $f(j/n)$的近似置信区间

回想一下，随机变量

$$\frac{2I(j/n)}{f(j/n)} \qquad 对于 0 < j < n/2$$

，近似独立地服从 $\chi^2{}_2$分布。

因此，近似地

$$\hat{f}(j/n) = \frac{1}{2m+1}\sum_{k=-m}^{m} I\left(\frac{j+k}{n}\right) \approx \frac{f(j/n)}{2(2m+1)}\sum_{k=-m}^{m}\frac{2I((j+k)/n)}{f((j+k)/n)}.$$

这将使我们能够以以下方式近似计算$\hat{f}(j/n)$ 的分布:

$$2(2m+1)\frac{\hat{f}(j/n)}{f(j/n)} \sim \chi^2_{2(2m+1)}.$$

如果 $\chi^2{}_{2(2m+1)}(\alpha/2)$ 和 $\chi^2{}_{2(2m+1)}(1-\alpha/2)$ 满足

$$\mathbb{P}\left\{\chi^2_{2(2m+1)}(\alpha/2) \le \chi^2_{2(2m+1)} \le \chi^2_{2(2m+1)}(1-\alpha/2)\right\} = 1-\alpha,$$

那么我们可以得出近似结论

$$\mathbb{P}\left\{\chi^2_{2(2m+1)}(\alpha/2) \le 2(2m+1)\frac{\hat{f}(j/n)}{f(j/n)} \le \chi^2_{2(2m+1)}(1-\alpha/2)\right\} \approx 1-\alpha.$$

这将导致对于 $f(j/n)$ 的置信区间，置信水平约为$1-\alpha$：

$$2(2m+1)\frac{\hat{f}(j/n)}{\chi^2_{2(2m+1)}(1-\alpha/2)} \le f(j/n) \le 2(2m+1)\frac{\hat{f}(j/n)}{\chi^2_{2(2m+1)}(\alpha/2)}.$$

2013年春季统计学153课程（时间序列）：第二十五讲

阿迪蒂亚・冈图博伊纳

2013年4月25日

假设 $x_1, ..., x_n$是我们假设来自均值为零方差为 σ^2的平稳过程 $\{X_t\}$的数据。

之前，我们研究了用于建模此类数据的平稳ARMA模型。 今天我们要问的是是否还有其他自然的建模方式。 特别是，我们要调查使用正弦和余弦函数来建模是否有意义，即 $x_1, ..., x_n$。

使用正弦和余弦函数的最简单的平稳模型是

$$X_t = A\cos(2\pi\lambda t) + B\sin(2\pi\lambda t)$$

其中$0 \leq \lambda \leq 1/^2$是一个固定常数，A和B是不相关的随机变量，均值为0，方差为σ^2。 我们在过去多次看到，$\{X_t\}$是平稳的，均值为0，方差为σ^2。

通过采用形式为线性组合的方式，可以构建更复杂的使用正弦和余弦函数的平稳模型。

$$X_t = \sum_{j=1}^{m} (A_j \cos(2\pi\lambda_j t) + B_j \sin(2\pi\lambda_j t)) \tag{1}$$

其中$0 \leq \lambda \leq 1/_2$是一个固定常数，$A1$，$B1$，A_2，B_2，...，A_m，B_m是所有不相关的随机变量，均值为零且

$$\mathrm{var}(A_j) = \mathrm{var}(B_j) = \sigma_j\text{^}2.$$

令$\sum_j=_1$^^{m} σ_j^$^2 = \sigma$^^{2}，使得过程$\{X_t\}$的方差等于σ^2.

事实证明，模型$(_1)$可以逼近任何平稳模型，只要m足够大且λ_1，...，λm和σ_1^2，...，σm^2选择合适。例如，选择

$$\lambda_j = 2\frac{j}{m}\text{和}\sigma_j\text{^}2 = \sigma\frac{2}{m} \qquad \text{对于} \;\; j = 1, ..., m$$

对于较大的 m，可以得到白噪声模型的很好近似。

为了用正弦和余弦来近似一个一般的平稳过程，我们可以使用其谱密度。
我们知道谱密度 $f(\lambda)$ 满足

$$\int_{-1/2}^{1/2} f(\lambda)d\lambda = \sigma\text{^}2.$$

对于实值平稳过程，谱密度关于零对称。因此，

$$\int_0^{1/2} f(\lambda)d\lambda = \frac{\sigma^2}{2}.$$

通过用黎曼和来近似上述左侧的积分（假设这样的近似是有效的），我们得到

$$\frac{\sigma^2}{2} = \sum_{j=1}^{m} \int_{\lambda_{j-1}}^{\lambda_j} f(\lambda) d\lambda \approx \frac{1}{2m} \sum_{j=1}^{m} f(\lambda_j) \qquad \text{其中 } \lambda_j = {}_j \ \frac{}{2m}.$$

现在我们取

$$\lambda_j = 2\frac{j}{m} \text{ 和 } \sigma_j\text{^}^2 = \frac{f(\lambda_j)}{m}$$

然后，对于大的 m，过程（1）将近似具有谱密度 f的平稳过程。 这种方法可以用来模拟ARMA过程，例如，不使用R中的 $arima.sim$函数。

当 m等于 ∞时，过程（1）可以被定义为有意义的，使得它具有完全相同的谱密度 f。 但这需要随机积分，超出了本课程的范围。

上述分析传达了谱密度在研究平稳过程中的关键作用。
它基本上告诉我们关于平稳过程的所有信息。

2013年春季统计学153课程（时间序列）：第二十六讲

阿迪蒂亚・冈图博伊纳

2013年4月30日

我们将继续讨论谱密度的估计。估计器由以下给出：

$$\hat{f}(j/n) := \sum_{k=-m}^{m} W_m(k) I\left(\frac{j+k}{n}\right)$$

权重集合 $\{W_m(k)\}$通常被称为核函数或谱窗口。

最简单的选择是 $W_m(k)$）

$$W_m(k) = \frac{1}{2m+1} \qquad \text{对于 } -m \leq k \leq m\text{。}$$

这个窗口被称为$Daniell$频谱窗口。

可以通过使用函数 $spec.pgram$和 $kernel$在R中直接获得这些估计值。

频谱窗口的带宽定义为加权分布的标准差。
实际上，正是这个标准差控制了估计量的偏差。 这可以通过二阶泰勒展开来证明。 期望值为$\hat{f}$ （j/n）

$$\mathbb{E}\hat{f}(j/n) = \sum_{k=-m}^{m} W_m(k) f\left(\frac{j+k}{n}\right)$$

为了简化表示，令 $\lambda = j/n$ 然后通过围绕 λ的二阶泰勒展开，我们得到

$$\mathbb{E}\hat{f}(\lambda) = \sum_{k=-m}^{m} W_m(k)\left(f(\lambda) + \frac{k}{n}f'(\lambda) + \frac{k^2}{2n^2}f''(\lambda)\right)$$

如果权重满足 $\sum_k W_m(k) = 1$并且 $\sum_k kW_m(k) = 0$（例如满足Daniell核函数），那么

$$\mathbb{E}\hat{f}(\lambda) - f(\lambda) = \frac{f''(\lambda)}{2} \sum_{k=-m}^{m} \left(\frac{k}{n}\right)^2 W_m(k)$$

核函数的带宽由以下公式给出

$$\sqrt{\sum_{k=-m}^{m} \left(\frac{k}{n}\right)^2 W_m(k)}.$$

对于Daniell核函数，带宽由均匀分布在$\{-m/n, -(m-1)/n, \ldots, (m-1)/n, m/n\}$上的标准差给出，该标准差非常接近于连续均匀分布在 $[-m/n, m/n]$ 上的标准差，即：

$$\sqrt{\frac{(2\text{米})^2}{12n^2}} \approx \sqrt{\frac{L^2}{12n^2}} = \frac{L}{n\sqrt{12}}$$

重复使用Daniell核函数会产生非均匀权重。例如，当 $m=1$时，Daniell核函数对应的三个权重为$(1/3, 1/3, 1/3)$。将其应用于一系列数字 $\{u_t\}$会得到更平滑的结果:

$$\hat{u}_t = \frac{u_{t-1} + u_t + u_{t+1}}{3}.$$

再次应用Daniell核函数到$\hat{u}_t$得到

$$\hat{\hat{u}}_t := \frac{\hat{u}_{t-1} + \hat{u}_t + \hat{u}_{t+1}}{3} = \frac{1}{9}u_{t-2} + \frac{2}{9}u_{t-1} + \frac{3}{9}u_t + \frac{2}{9}u_{t+1} + \frac{1}{9}u_{t+2}.$$

因此，应用Daniell核函数等同于将核函数（$1/9, 2/9, 3/9, 2/9, 1/9$）应用于数据。这是一个带有更高带宽的非均匀核函数。还要注意这些权重等于Daniell核函数的卷积。换句话说，如果 X_1和 X_2都具有概率质量函数（$1/3, 1/3, 1/3$），那么
如果我们不断应用Daniell核函数，我们会得到非常类似于高斯概率密度函数的频谱窗口。

另一个常见的核函数选择是修改的Daniell核函数，它在端点处放置了半权重。
这本书还讨论了Dirichlet核函数和Fejer核函数。