蘑菇先生学习记

ARIMA时间序列模型(一)

2017-03-07 | 统计学 | 阅读量 65

基本概念

时间序列是什么?

定义：时间序列数据是按时间排序的观察序列，是目标在不同时间点下的一系列观察值。

所有的时间观察序列数据可以被标记为：$z_1, z_2, ..., z_T$,可以当作T个随机变量的一个实例：

$$(Z_1, Z_2, .., Z_T)$$

进一步定义：时间序列是一系列按照时间排序的随机变量。通常定义为双无穷随机变量序列。标记为：$Z_t, t \in \mathbb{Z}$,或者简记为：Z_t 。时间序列是离散时间下的随机过程。

回顾线性模型，响应变量Y和多个因变量X，线性模型表示为：

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

因变量X的信息是已知的，我们希望对响应变量Y做出推断。

在时间序列分析中，我们提出如下模型：

$$Y_t = \beta_o + \beta_1 Y_{t-1} + \varepsilon_t$$

在时间序列中，已知的信息包括：

- 时间下标t
- 过去的信息

两个典型的时间序列模型如下：

$$Z_t = a + bt + \varepsilon_t$$

and

$$Z_t = \theta_0 + \phi Z_{t-1} + \varepsilon_t$$

它们分别对应于确定性模型和随机模型，本文将讨论后者。

时间序列的均值，方差，协方差

- **均值函数（The mean function）**：对于一个时间序列$Z_t, t \in Z$,均值函数或平均序列被定义为：

  $$\mu_t = E(Z_t),\ t \in \mathbb{Z}$$

  μ_t是在t时刻的期望值，μ_t 在不同时刻可以是不同的值。

- **自协方差函数（The auto-covariance function）**：简记为ACVF，定义为：

  $$\gamma(t,s) = cov(Z_t, Z_s)\ t, s \in \mathbb{Z}$$

  其中，

  $$cov(Z_t, Z_s) = E[(Z_t - \mu_t)(Z_s - \mu_s)] = E(Z_t Z_s) - \mu_t \mu_s$$

- **方差函数（The variance function）**：特别是在s=t时，我们有：

  $$\gamma(t,t) = cov(Z_t, Z_t) = var(Z_t)$$

这就是Z_t的方差函数

- **自相关函数（The auto-correlation function）**：简记为ACF，定义为：

$$\rho(t,s) = corr(Z_t, Z_s),\ t,s \in \gamma(t,s) = cov(Z_t, Z_s)\ t,s \in \mathbb{Z}$$

其中，

$$corr(Z_t, Z_s) = \frac{cov(Z_t, Z_s)}{\sqrt{var(Z_t)var(Z_s)}} = \frac{\gamma(t,s)}{\sqrt{\gamma(t,t)\gamma(s,s)}}$$

ACVF和ACF有如下性质：

ACVF:

- $\gamma(t,t) = var(Z_t)$
- $\gamma(t,s) = \gamma(s,t)$
- $|\gamma(t,s)| \le \sqrt{\gamma(t,t)\gamma(s,s)}$

ACF:

- $\rho(t,t) = 1$
- $\rho(t,s) = \rho(s,t)$
- $|\rho(t,s)| \le 1$

一些重要的性质：

$$cov(aX, Y) = acov(X, Y)$$

$$cov(X, aY + bZ) = acov(X, Y) + bcov(X, Z)$$

$$cov(c_1Y_1 + c_2Y_2, d_1Z_1 + d_2Z_2) = c_1d_1cov(Y_1, Z_1) + c_2d_1cov(Y_2, Z_1) \\ + c_1d_2cov(Y_1, Z_2) + c_2d_2cov(Y_2, Z_2)$$

$$cov\left[\sum_{i=1}^{m} c_i Y_i, \sum_{j=1}^{n} d_j Z_j\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j cov(Y_i, Z_j)$$

最后一条性质经常用到。

随机游走

随机游走（The random walk）：令序列$a_t, t \in \mathbb{N}$是服从$i.i.d$独立同分布的随机变量。每个变量都是零均值，方差为σ_a^2,随机游走过程$Z_t, t \in \mathbb{N}$定义为:

$$Z_t = \sum_{j=1}^{t} a_j,\ t \in \mathbb{N}$$

另外，我们可以写作:

$$Z_t = Z_{t-1} + a_t,\ t \in \mathbb{N}, Z_0 = 0$$

- Z_t均值函数为:

$$\mu_t = E(Z_t) = E\left(\sum_{j=1}^{t} a_j\right) = \sum_{j=1}^{t} E(a_j) = 0$$

- Z_t方差函数为:

$$\gamma(t,t) = var(Z_t) = var\left(\sum_{j=1}^{t} a_j\right) = \sum_{j=1}^{t} var(a_j) = t \cdot \sigma_a^2$$

注意到，这一过程，方差会随着时间线性增长。

- ACVF自协方差函数：对于一切$t \le s$,

$$
\begin{aligned}
\gamma(t,s) &= cov(Z_t, Z_s) \\
&= cov\left(\sum_{j=1}^{t} a_j, \sum_{j=1}^{s} a_j\right) \\
&= cov\left(\sum_{j=1}^{t} a_j, \sum_{j=1}^{t} a_j + \sum_{j=t+1}^{s} a_j\right) \\
&= cov\left(\sum_{j=1}^{t} a_j, \sum_{j=1}^{t} a_j\right) \\
&= var\left(\sum_{j=1}^{t} a_j\right) = t \cdot \sigma_a^2
\end{aligned}
$$

- ACF自相关函数，根据定义有：

$$
\begin{aligned}
\rho(t,s) &= \frac{\gamma(t,s)}{\sqrt{\gamma(t,t)\gamma(s,s)}} \\
&= \frac{\sigma_a t}{\sqrt{\sigma_a^2 t \cdot \sigma_a^2 s}} \\
&= \sqrt{t/s},\ 1 \le t \le s
\end{aligned}
$$

当s=t+1时，

$$
\rho(t,t+1) = corr(Z_t, Z_{t+1}) = \sqrt{t/(t+1)} \approx 1,\ \text{当}t\text{无穷大}
$$

理解：随机游走可以看作，在时间轴上任意行走一步（大步或小步），是若干时刻的和。

移动平均

移动平均（a moving average）：假设$Z_t, t \in \mathbb{Z}$定义为：

$$
Z_t = a_t - 0.5a_{t-1},\ t \in \mathbb{Z}
$$

同样，a满足独立同分布，零均值，方差为σ_a^2

- Z_t均值函数为:

$$\mu_t = E(Z_t) = E(a_t) - 0.5E(a_{t-1}) = 0,\ t \in \mathbb{Z}$$

- Z_tf方差函数为:

$$var(Z_t) = var(a_t - 0.5a_{t-1}) = \sigma_a^2 + 0.5^2\sigma_a^2 = 1.25\sigma_a^2$$

- ACVF自协方差函数:

$$\begin{gathered} cov(Z_t, Z_{t-1}) = cov(a_t - 0.5a_{t-1}, a_{t-1} - 0.5a_{t-2}) = cov(a_t, a_{t-1}) \\ -0.5cov(a_t, a_{t-2}) - 0.5cov(a_{t-1}, a_{t-1}) - 0.5cov(a_{t-1}, a_{t-1}) \\ +0.5^2cov(a_{t-1}, a_{t-2}) = -0.5cov(a_{t-1}, a_{t-1}) \end{gathered}$$

  或者表示为:

$$\gamma(t, t-1) = -0.5\sigma_a^2, \forall t \in \mathbb{Z}$$

  对任意$k \geq 2$,

$$cov(Z_t, Z_{t-k}) = 0$$

  或者表示为,

$$\gamma(t, t-k) = 0,\ \forall k \geq 2, t \in \mathbb{Z}$$

- ACF自相关函数:

$$\begin{gathered} \rho(t, s) = -0.4, if\ |t-s| = 1 \\ \rho(t, s) = 0, if\ |t-s| \geq 2 \end{gathered}$$

  理解：移动平均可以看作，若干时刻的线性组合。

平稳性

强平稳性（strict stationarity）要求：时间序列Z_t为强平稳，只有当对任意的自然数n, 任意的时间点t_1,t_2,...,t_n以及任意的滞后k, 都满足Z_{t_1},Z_{t_2},...,Z_{t_n}的联合分布 和Z_{t_1-k},Z_{t_2-k},...,

Z_{t_n-k}相同。

弱平稳性(weak stationarity)要求：时间序列为弱平稳性，只有当均值函数μ_t不随时间变化，并且对于任意的时间t和任意的滞后k，都有$\gamma(t,t-k)=\gamma(0,k)$

对于弱平稳性，有如下标志：

$$\mu = E(Z_t)$$

$$\gamma_k = cov(Z_t, Z_{t-k}),\ (\gamma_{-k} = \gamma_k)$$

$$\rho_k = Corr(Z_t, Z_{t-k});\ (\rho_{-k} = \rho_k)$$

强平稳性和弱平稳性关系如下：

1. 强平稳性+有限的秒时刻 => 弱平稳性
2. 时间序列的联合分布为多元正太分布，那么这两种定义是一致的

白噪声

白噪声（White noise）：一个很重要的关于平稳性处理的例子就是所谓的白噪声处理。它被定义为满足独立同分布的随机变量a_t,零均值并且方差为$\sigma_a^2>0$,简记为：$WN(0,\sigma_a^2)$

显然，a_t满足强平稳性要求。

对于弱平稳性，注意到$\mu_t=E(a_t)=0$是一个常数，并且，

$$\gamma(t;t-k) = \begin{cases} \sigma_a^2, k=0 \\ 0, k\neq 0 \end{cases} := \gamma_k$$

,

$$\rho_k = \begin{cases} 1, k=0 \\ 0, k\neq 0 \end{cases}$$

有些书中定义白噪声为一系列不相关的随机变量。

前面我们提高的随机游走，由于Z_t的方差受时间影响线性变化$var(Z_t) = t\sigma_a^2$，并且协方差$\gamma(t,s) = t\sigma_a^2$,因此不仅仅受滞后k的影响，故不是平稳的时间序列。

令，

$$X_t = \nabla Z_t = Z_t - Z_{t-1}$$

则$X_t = a_t, \nabla Z_t$是平稳的。

前面我们还提到移动平均。是由白噪声构成的一个非平凡平稳时间序列。在前面那个例子里，我们有：

$$\rho_k = \begin{cases} 1, k = 0 \\ -0.4, k \pm 1 \\ 0, |k| \geq 2 \end{cases}$$

统计学 # 时间序列 # ARIMA # ARMA

神经网络(系列2)

ARIMA时间序列模型(二)

0 条评论，0 人参与。

0

我有话说...

使用社交帐号登录

发布前先点击左边的按钮登录

最新评论

还没有评论

友言?