二、时间序列的预处理

目录

平稳序列的定义

- 特征统计量
  - 概率分布族：$F_{t_1,t_2,\cdots,t_m}(x_1,x_2,\cdots,x_m), \forall m \in \mathbb{Z}^+, \forall t_i \in T$
  - 均值：$\mu_t = E(X_t) = \int_{-\infty}^{+\infty} x \,\mathrm{d}F_t(x)$
  - 方差：$\sigma_t^2 = D(X_t) = E[(X_t - \mu_t)^2] = \int_{-\infty}^{+\infty} (x-\mu_t)^2 \,\mathrm{d}F_t(x)$
  - 自协方差函数：$\gamma(t,s) = E[(X_t - \mu_t)(X_s - \mu_s)]$
  - 自相关系数 (ACF)：$\rho(t,s) = \dfrac{\gamma(t,s)}{\sqrt{D(X_t)\cdot D(X_s)}}$

平稳序列的定义

- 平稳时间序列的定义
  - 严平稳：$F_{t_1,t_2,\cdots,t_m}(x_1,x_2,\cdots,x_m) = F_{t_{1+\tau},t_{2+\tau},\cdots,t_{m+\tau}}(x_1,x_2,\cdots,x_m)$
  - 宽平稳：$E(X_t^2) < \infty, E(X_t) = \mu, \gamma(t,s) = \gamma(k,k+s-t)$
  - 通常情况下，严平稳能推出宽平稳成立，宽平稳序列不能反推严平稳成立
  - 服从柯西分布的严平稳序列就不是宽平稳序列
  - 当序列服从多元正态分布时，宽平稳可以推出严平稳
  - 多元正态分布：

$$f_{t_1,t_2,\cdots,t_n}(X) = (2\pi)^{-n/2}|\Gamma|^{1/2}\exp\left[-\frac{1}{2}(X-\mu)^{\mathrm{T}}\Gamma^{-1}(X-\mu)\right]$$

平稳序列的定义

- 平稳时间序列的统计性质
  - 常数均值：$E(X_t) = \mu$
  - 自协方差函数和自相关函数只依赖于时间的平移长度而与时间的起止点无关：$\gamma(t,s) = \gamma(k, k+s-t) := \gamma(s-t)$ ($\gamma(s-t)$ 是延迟 k 自协方差函数)
  - 常数方差：$D(X_t) = \gamma(0)$
  - 延迟 k 自相关系数：$\rho_k = \dfrac{\gamma(k)}{\gamma(0)}$
  - 规范性：$\rho_0 = 1, |\rho_k| \leq 1$
  - 对称性：$\rho_k = \rho_{-k}$
  - 相关阵为对称非负定阵

平稳性检验

- 图检验

时序图应该显示出该序列始终在一个常数值附近波动，而且波动的范围有界的特点。

- 单位根检验 (DF 检验)
  - 序列假定：$x_t = \phi_1 x_{t-1} + \xi_t$，$\xi_t \sim N(0, \sigma^2)$
  - $H_0 : |\phi_1| \geq 1$ (序列非平稳)，$H_1 : |\phi_1| < 1$ (序列平稳)
  - $|\phi_1| < 1$ 时的 t 统计量：$t(\hat{\phi}_1) = \dfrac{\hat{\phi}_1 - \phi_1}{S(\hat{\phi}_1)} \to N(0,1)$
  - $|\hat{\phi}_1| \geq 1$ 时的 DF 统计量：$\tau = \dfrac{|\hat{\phi}_1| - 1}{S(\hat{\phi}_1)} \to \dfrac{\int_0^1 W(r)\,\mathrm{d}W(r)}{\sqrt{\int_0^1 W^2(r)\,\mathrm{d}r}}$
  - $W(r)$ 是自由度为 r 的维纳过程：

    $W(0) = 0, W(1) \sim N(0,1), \sigma W(r) \sim N(0, r\sigma^2), W^2(r)/r \sim \chi^2(1)$

平稳性检验

- DF 检验的等价表达与三种类型
  - 序列假定：$\nabla x_t = \rho x_{t-1} + \xi_t$
  - $H_0 : \rho \geq 0$ (序列非平稳)，$H_1 : \rho < 0$ (序列平稳)
  - 统计量：$\tau = \dfrac{\hat{\rho}}{S(\hat{\rho})}$
  - 三种类型：

    无漂移项自回归结构：$x_t = \phi_1 x_{t-1} + \xi_t$；

    有漂移项自回归结构：$x_t = \phi_0 + \phi_1 x_{t-1} + \xi_t$；

    带趋势回归结构：$x_t = \alpha + \beta t + \phi_1 x_{t-1} + \xi_t$

平稳性检验

- 增广 DF 检验 (ADF 检验)
  - 序列假定：$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \xi_t$
  - $\rho = \phi_1 + \phi_2 + \cdots + \phi_p - 1$
  - $H_0 : \rho \geq 0$ (序列非平稳)，$H_1 : \rho < 0$ (序列平稳)
  - 统计量：$\tau = \dfrac{\hat{\rho}}{S(\hat{\rho})}$

纯随机性检验

- 纯随机序列 (白噪声序列)
  - 定义：$E(X_t)=\mu,\gamma(t,s)=\begin{cases}\sigma^2, & t=s\\ 0, & t\neq s\end{cases}$，记为 $X_t\sim \mathrm{WN}(0,\sigma^2)$
  - 性质：$\gamma(k)=0,D(X_t)=\gamma(0)=\sigma^2$
  - $H_0:\rho_1=\rho_2=\cdots=\rho_m=0$ (是纯随机序列)，$H_1:\exists\rho_k\neq 0$ (不是纯随机序列)
  - Q 统计量：$Q=n\sum\limits_{k=1}^{m}\hat{\rho}_k^2\to\chi^2(m)$
  - LB 统计量：$\mathrm{LB}=n(n+2)\sum\limits_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k}\to\chi^2(m)$

看程序读结论

增广 Dickey-Fuller 单位根检验							
类型	滞后	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
零均值	0	0.7840	0.8642	0.70	0.8614		
	1	0.9266	0.8915	1.77	0.9795		
	2	0.9120	0.8886	1.99	0.9873		
单均值	0	-9.4767	0.1269	-2.00	0.2858	2.74	0.3929
	1	-0.9933	0.8809	-0.37	0.9042	1.81	0.6184
	2	0.0918	0.9573	0.04	0.9559	1.99	0.5746
趋势	0	-33.0435	0.0004	-4.75	0.0025	11.36	0.0010
	1	-12.2151	0.2420	-1.96	0.6031	2.12	0.7594
	2	-10.6376	0.3337	-1.49	0.8140	1.41	0.8919

第一列是检验的模型类型，第二列是自相关延迟阶数。

第三列、第四列是 Rho 统计量的值及检验 P 值。

第五列、第六列是 Tau 统计量 (τ) 的值及检验 P 值。

第七列、第八列是回归模型显著性检验 F 统计量的值及检验 P 值。

看程序读结论

白噪声的自相关检查									
至滞后	卡方	自由度	Pr > 卡方	自相关					
6	47.81	6	<.0001	0.578	0.599	0.402	0.384	0.272	0.136

第一列是延迟阶数 k。

第二列是 LB 统计量的值。

第三列是该检验统计量的自由度。

第四列是 LB 检验统计量的 P 值。

第五列是延迟各阶自相关系数 $(\rho_1, \rho_2, \cdots, \rho_k)$。