时间序列分析第六章作业

Phlinsia

2024 年 4 月 30 日

Contents

0.1 练习 6.12

从包含 100 个观测值的时间序列中，我们计算得到以下自相关系数：

$$\begin{cases} r_1 = -0.49, \\ r_2 = 0.31, \\ r_3 = -0.21, \\ r_4 = 0.11 \end{cases}$$

且满足条件

$$r_k| < 0.09 \quad \text{当} \quad k > 4.$$

仅依据这些统计量，我们尝试初步确定适合该序列的 ARIMA 模型。考虑到样本相关性的标准差，通过 $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2$，如果某个自相关系数的绝对值显著大于这个值，就可能意味着序列不是完全随机的，存在某种结构。我们或许可以考虑 MA(2) 或 MA(3) 模型作为可能的选项。

如果我们暂时假定一个 MA(2) 结构，可以利用公式来评估 r_3 的方差。

$$\mathrm{Var}(r_k) = \frac{1}{n}\left[1 + 2\sum_{j=1}^{q}\rho_j^2\right] \text{ 当 } k > q$$

基于已知的 r_1 至 r_4，在假设 MA(2) 模型时，仅考虑 $q = 2$，计算 r_3 的方差。这里，ρ_1 和 ρ_2 分别对应 r_1 和 r_2 的值（注意，实际应用中 ρ 应通过偏自相关函数 PACF 准确计算，但此处直接用了 r 值作为近似处理），代入公式:

$$\mathrm{Var}(r_3) = \frac{1 + 3[(-0.49)^2 + (0.31)^2]}{100} = 0.016724$$

因此，r_3 的标准化残差是其值除以方差的平方根：

$$\frac{r_3}{\sqrt{\mathrm{Var}(r_3)}} = \frac{-0.21}{\sqrt{0.016724}} = -1.62.$$

鉴于该标准化残差的绝对值（$|-1.62|$）并未远大于 1，未超过统计显著性的典型临界值（通常对于 95% 的置信区间约为 1.96），我们根据这一证据得出结论，MA(2) 模型未被拒绝。因此，基于呈现的自相关结构，一个 MA(2) 模型似乎是该时间序列合理的初步模型设定。

0.2 练习 6.13

假设有一个长度为 121 的平稳时间序列，该序列产生了以下样本偏自相关系数：

$$\begin{cases} \hat{\phi}_{11} = 0.8, \\ \hat{\phi}_{22} = -0.6, \\ \hat{\phi}_{33} = 0.08, \\ \hat{\phi}_{44} = 0.00. \end{cases}$$

$$\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{121}} = 0.181$$

这个界限来源于正态分布的 95% 置信区间概念，即大约有 95% 的数据点会落在均值 ± 2 标准差的范围内。对于长度为 n 的时间序列，一个纯随机序列（即白噪声）的自相关系数的标准误大约为 $\frac{1}{\sqrt{n}}$，因此，两倍的标准误即为 $\frac{2}{\sqrt{n}}$。当 PACF 的绝对值超过这个界限时，我们倾向于认为该系数在统计上显著，意味着它可能对模型有所贡献。

提供的 PACF 值在前几阶显著不为零，特别是 $\hat{\phi}_{11} = 0.8$ 和 $\hat{\phi}_{22} = -0.6$，这两个值的绝对值远大于 0.181 这一显著性界限。由于 $\hat{\phi}_{33} = 0.08$ 和 $\hat{\phi}_{44} = 0.00$ 的值接近零，且从 $\hat{\phi}_{22}$ 到 $\hat{\phi}_{33}$ 有一个明显的下降，这表明序列的自回归特性主要体现在前两阶滞后上。因此，基于 PACF 的这种表现，我们倾向于认为一个 AR(2) 模型足够捕捉到序列的主要自相关特性，而更高阶的滞后系数在统计上并不显著，故不纳入模型考虑。

0.3 练习 6.15

下表给出了一序列及其一阶差分的样本自相关函数（ACF），其中 $n = 100$。

延迟 (lag)	Y_t 的 ACF	∇Y_t 的 ACF
1	0.97	-0.42
2	0.97	0.18
3	0.93	-0.02
4	0.85	0.07
5	0.80	-0.10

仅凭这些信息，我们会考虑使用哪个或哪些 ARIMA 模型对该序列进行分析？

一阶差分后，第 1 阶 ACF 为-0.42，之后的 ACF 值迅速减小并接近零，这提示序列经过一次差分后可能达到平稳。

样本自相关函数中缺乏衰减现象表明非平稳性。然而，在进行差分后，相关性显得更为合理。，

$$\because \rho_1 = -0.42 \therefore \mathrm{Var}(r_1) = \frac{1 + 2(-0.42)^2}{100} = 0.0135$$

$$\because \rho_2 = 0.18 \therefore \text{显著性为} \frac{\rho_2}{\sqrt{\mathrm{Var}(r_1)}} = \frac{0.18}{\sqrt{0.0135}} = 1.55$$

统计实践中，如果这个标准化值的绝对值大于约 1.96（对应于 95% 的置信水平），则认为该自相关系数在统计上显著不为零。此处，虽然 1.55 小于 1.96，但仍然较大，表明二阶自相关在统计上也是相对显著的，尽管可能不满足最严格的标准显著性水平。

差分后序列的 ACF 显示出了平稳性，并且至少有一阶自相关显著，这表明序列经过一次整合（I）后，可能还存在一定的自回归（AR）和滑动平均（MA）特性。IMA(1,1) 模型意味着序列经过一次差分（整合部分）后，其残差显示出一阶自回归和一阶滑动平均的影响。因此，根据差分后 ACF 的模式和显著性测试结果，IMA(1,1) 模型是进一步分析和检验的一个合理候选模型。

0.4 练习 6.16

针对长度为 64 的时间序列，样本偏自相关系数如下所示：

延迟 (Lag)	偏自相关系数 (PACF)
1	0.47
2	-0.34
3	0.20
4	0.02
5	-0.06

在这种情况下，我们应该考虑哪些模型呢？

在大样本情况下，一个纯随机序列（白噪声）的自相关系数大约会落在 $\pm\frac{2}{\sqrt{n}}$ 的范围内。因此，如果某个 PACF 的绝对值大于 $\frac{2}{\sqrt{n}}$，则认为它在统计上显著不为零，可能意味着该滞后阶在模型中是重要的。

因为 $\frac{2}{\sqrt{64}} = 0.25$，并且从滞后 3 开始的所有偏自相关系数的绝对值都小于 0.25。

PACF 在 Lag 1 和 Lag 2 处有显著的非零值（0.47 和-0.34），这意味着序列中存在明显的自回归特征，且在前两阶。而从 Lag 3 开始，PACF 的值（0.20, 0.02, -0.06）都小于显著性界限 0.25，这意味着在 Lag 3 及以后的自回归效应不显著。因此，基于 PACF 的这种“拖尾”（在高阶滞后逐渐减小至不显著）特性，我们可以推测原序列可能适合一个自回归模型，且其阶数为 2，即 AR(2) 模型。

0.5 练习 6.20

Simulate an AR(1) time series with $n = 48$ and with $\phi = 0.7$.

```
set.seed(241357); series=arima.sim(n=48,list(ar=0.7))
```

a

对于此模型，计算滞后 1 和滞后 5 的理论自相关系数。

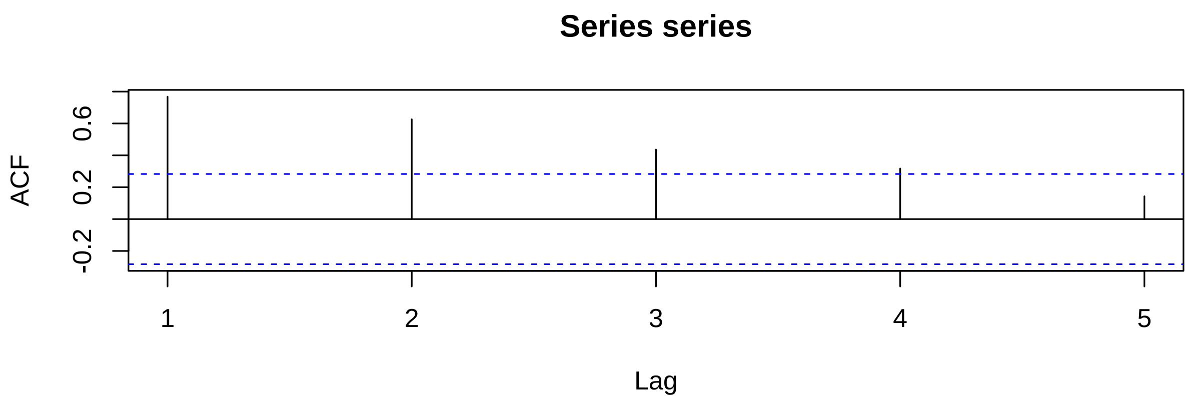
$$\rho_1 = \phi^1 = 0.7, \quad \rho_5 = \phi^5 = 0.16807$$

b

计算滞后 1 和滞后 5 的样本自相关系数，并将这些值与理论值进行比较。

$$\mathrm{Var}(r_1) \approx \frac{1-\phi^2}{n}, \quad \mathrm{Var}(r_k) \approx \frac{1}{n}\left(\frac{1+\phi^2}{1-\phi^2}\right) \quad \text{（当 } k \text{ 较大时）}$$

```
acf(series,lag.max=5)[1:5]
```




```
##
## Autocorrelations of series 'series', by lag
##
##     1     2     3     4     5
## 0.768 0.626 0.436 0.318 0.143
```

r_n 的标准误差计算为 $\sqrt{\frac{1-\phi^2}{n}} = \sqrt{\frac{1-(0.7)^2}{48}} = \sqrt{0.010625} \approx 0.10$。至于 r_5 的标准误差，则计算为

$$\sqrt{\frac{1}{n}\left[\frac{1+\phi^2}{1-\phi^2}\right]} = \sqrt{\frac{1}{48}\left[\frac{1+(0.7)^2}{1-(0.7)^2}\right]} \approx 0.25.$$

综合考虑这些标准误差，估计值 0.768 对于 ϕ 而言，以及 0.14 对于另一个系数，分别是真实值 0.7 和 0.16807 的极佳近似。

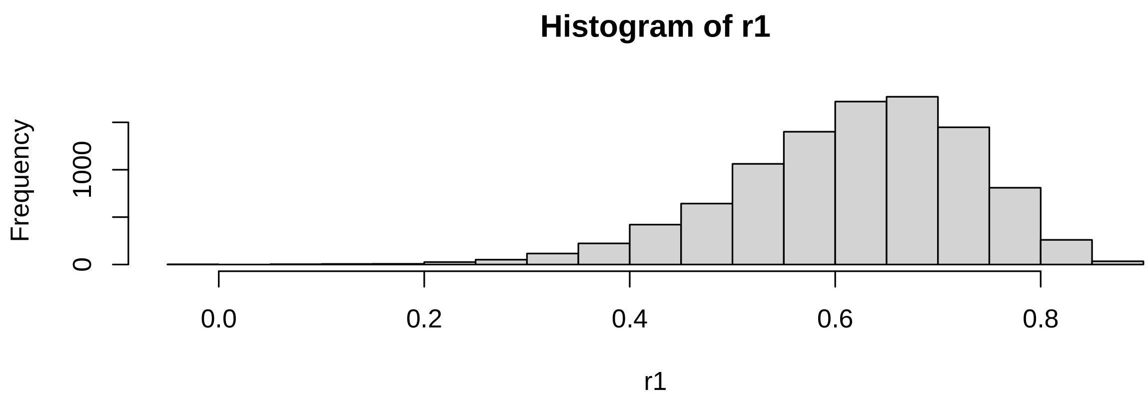
c

在相同的条件下使用新的模拟数据重复 (b) 部分的步骤。讨论在不同样本下估计的准确性如何变化。

d

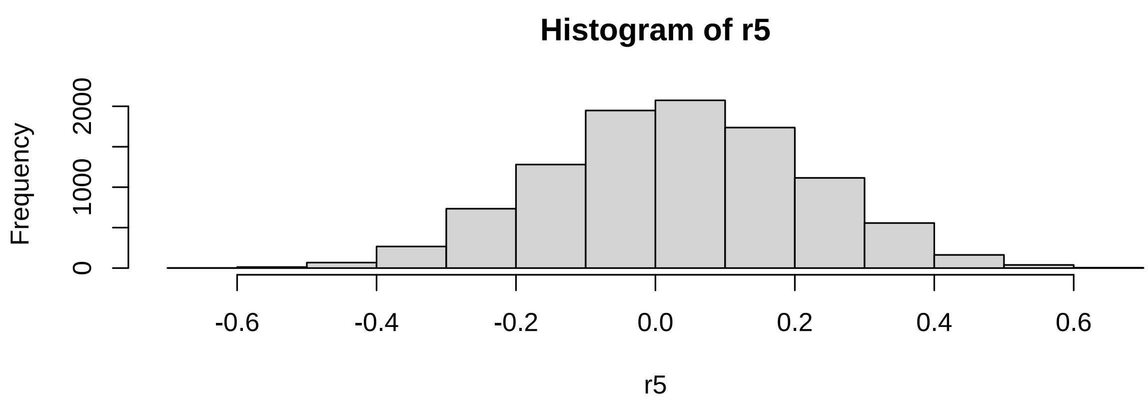
多次重复时间序列的模拟以及 r_1 和 r_5 的计算过程，构建 r_1 和 r_5 的抽样分布。描述在相同条件下选取不同样本时，这些估计的精度如何变化。

```
set.seed(132435); r1=rep(NA,10000); r5=r1
for (k in 1:10000) {series=arima.sim(n=48, list(ar=0.7));r1[k]=acf(series,lag.max=1,plot=F)$acf[1];r5[k
hist(r1); mean(r1); sd(r1); median(r1)
```




```
## [1] 0.6184299
## [1] 0.1145287
## [1] 0.6313287
```

```
hist(r5); mean(r5); sd(r5); median(r5)
```




```
## [1] 0.03277664
## [1] 0.1848059
## [1] 0.03243878
```

对于 r_1 的抽样分布，其均值为 0.618（$\rho_1 = 0.7$），中位数为 0.631，这与观察到的向较小值偏斜的情况一致。该分布的标准差为 0.11，与渐近理论预测的约 0.10 非常吻合。

至于 r_5 的抽样分布，其均值为 0.033（$\rho_5 = 0.168$），中位数为 0.032，这与该分布接近对称的特点相符。此分布的标准差为 0.18，与渐近理论值大约 0.25 较为相符。

本练习揭示了在样本量 $n = 48$ 如此之小时，准确估计简单 AR(1) 序列自相关函数的难度。你可以考虑使用更大的样本量，比如 $n = 96$ 或更大来重复此练习。此外，尝试不同的自回归参数 ϕ 值也是值得探索的方向。

0.6 练习 6.21

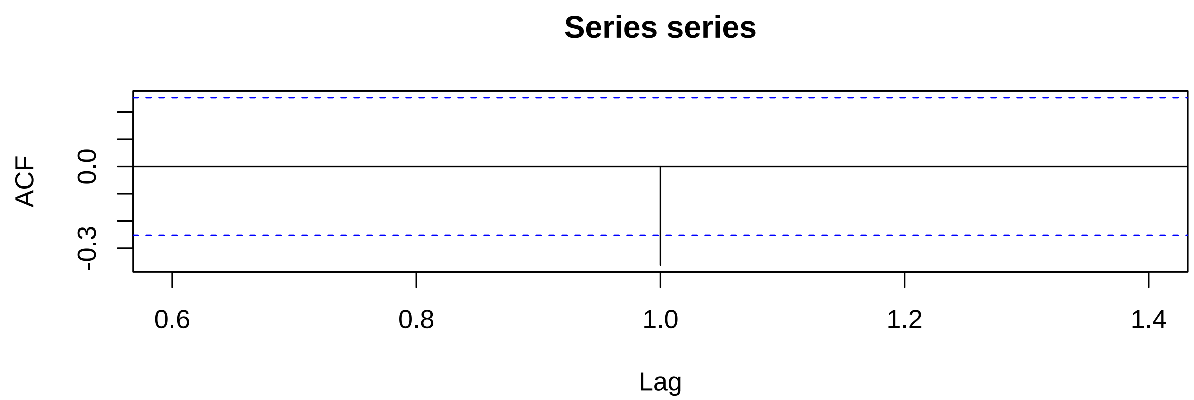
a

```r
set.seed(6453421); series=arima.sim(n=60,list(ma=-0.5))
```

$$\rho_1 = -\frac{\theta}{1+\theta^2} = -\frac{0.5}{1+(0.5)^2} = -0.4$$

b

```r
acf(series,lag.max=1)[1]
```



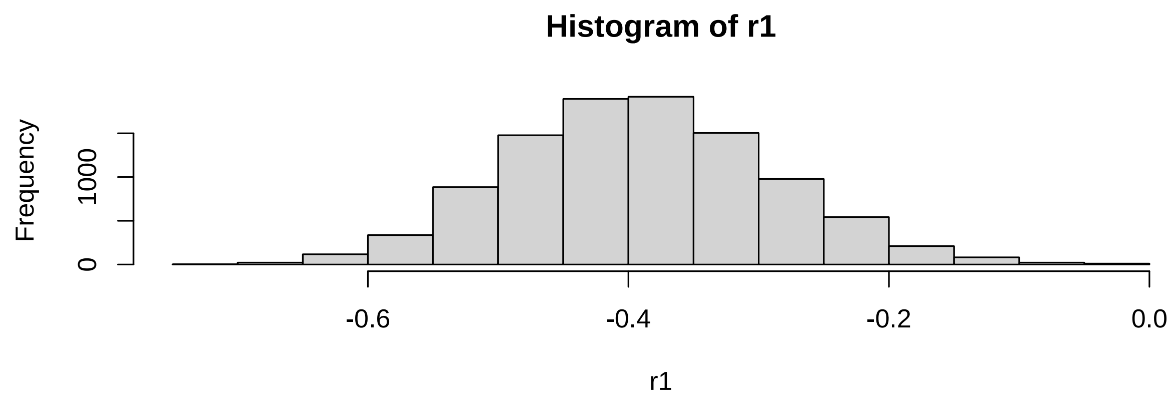

```
##
## Autocorrelations of series 'series', by lag
##
##      1
## -0.362
```

滞后 1 处的自相关系数估计值为 -0.362。r_1 的标准误差计算为 $\sqrt{c_{11}/n} = \sqrt{1-3\rho_1^2+4\rho_1^4/n} = \sqrt{1-3(-0.4)^2+4(-0.4)^4/60} \approx 0.10$。该估计值位于真值 -0.4 的两倍标准误差范围内，表明估计较为精确。

c

d

```r
set.seed(534261); r1=rep(NA,10000); r5=r1
for (k in 1:10000) {series=arima.sim(n=60, list(ma=-0.5));r1[k]=acf(series,lag.max=1,plot=F)$acf[1]}
hist(r1); mean(r1); sd(r1); median(r1)
```




```
## [1] -0.3901954
## [1] 0.1002125
## [1] -0.3931737
```

注意到 $\rho_1 = -0.4$。此处，抽样分布的均值为-0.390（中位数为-0.393），标准差为 0.100。给出的大样本标准差是 $\frac{0.79}{\sqrt{60}} = 0.102$，这与从抽样分布中获得的值非常接近，是一个极好的近似。

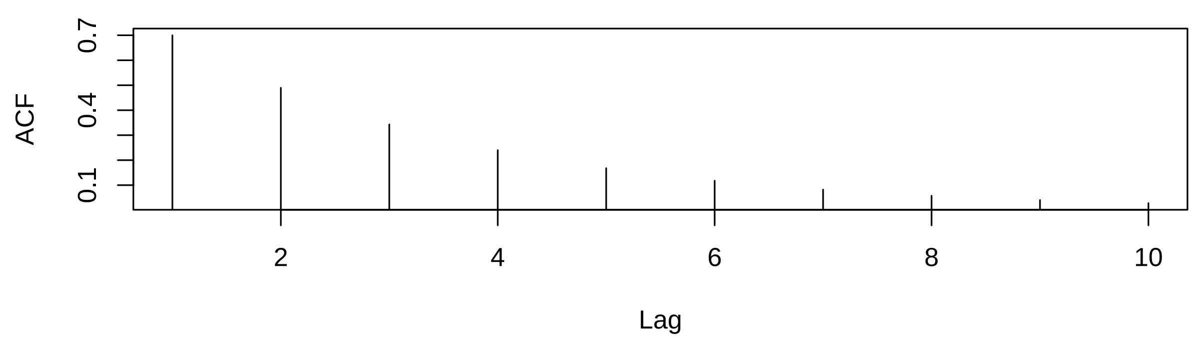
0.7 练习 6.25

a

```
round(ARMAacf(ar=0.7,lag.max=10),digits=3)
```

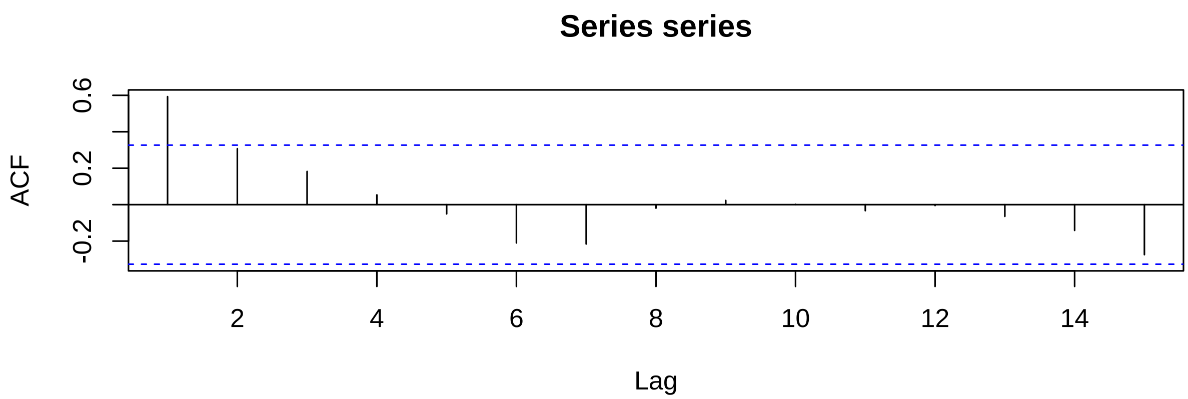
```
##     0     1     2     3     4     5     6     7     8     9    10
## 1.000 0.700 0.490 0.343 0.240 0.168 0.118 0.082 0.058 0.040 0.028
```

```
ACF=ARMAacf(ar=0.7,lag.max=10)
plot(y=ACF[-1],x=1:10,xlab='Lag',ylab='ACF',type='h'); abline(h=0)
```



b

```
set.seed(162534); series=arima.sim(n=36,list(ar=0.7)); acf(series)
```



模式匹配并不是很好，但请记住 $n = 36$。

c

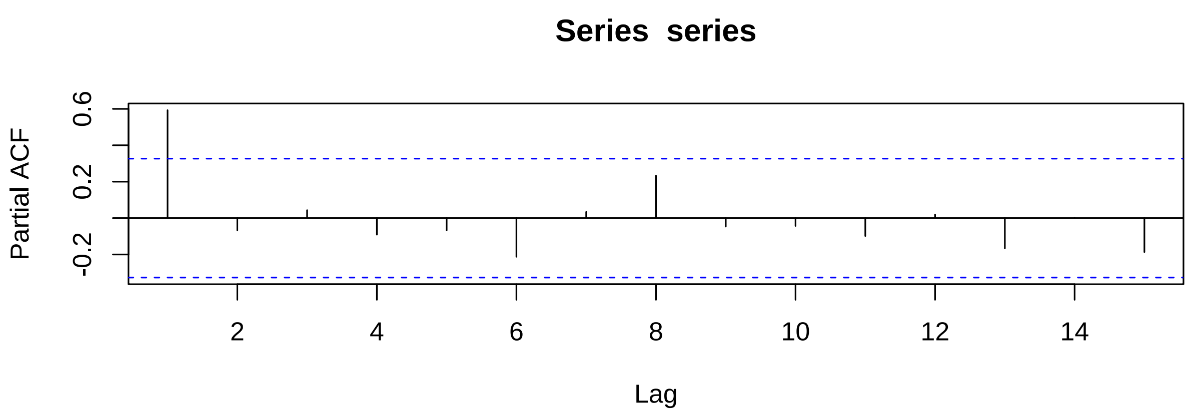
$\phi_{11} = 0$ 并且对于其它所有的 k，$\phi_{kk} = 0$。

d

参考 (b) 部分的答案，这里 r_1 的方差的平方根大约为 $\sqrt{Var(r_1)} \approx \sqrt{(1-\phi^2)/n} = \sqrt{(1-0.7^2)/36} = 0.12$，表明观测到的 r_1 值很好地落在了真实值的两倍标准误差之内。对于更高阶的滞后也一样。

e

```r
pacf(series)
```



利用大约的标准误差 $\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{36}} = 0.167$，样本偏自相关函数（PACF）与理论上的 PACF 匹配得相当好。

0.8 练习 6.26

a

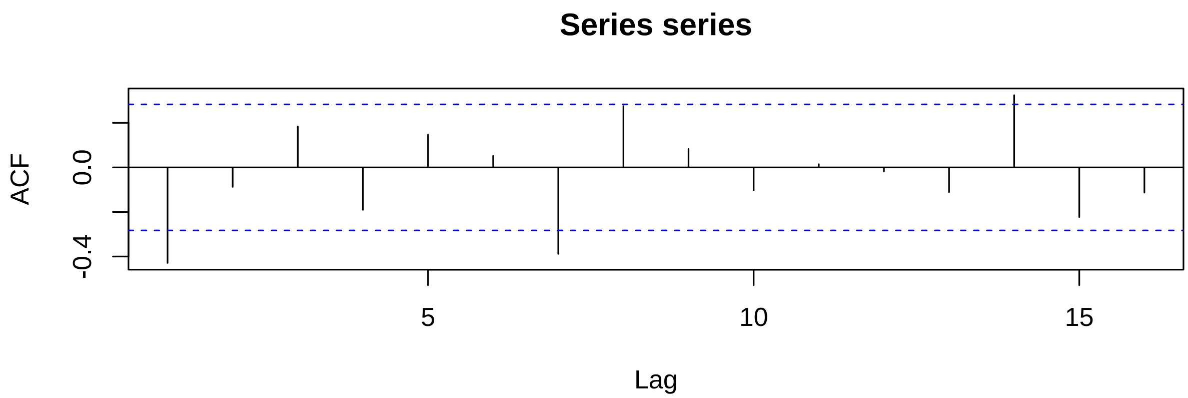
对于给定的模型，一阶偏自相关系数计算如下：

$$\rho_1 = -\frac{\theta}{1+\theta^2} = -\frac{0.5}{1+(0.5)^2} = -0.4$$

这是该模型中唯一非零的自相关系数。

b

```
set.seed(162534); series=arima.sim(n=48,list(ma=-0.5)); acf(series)
```



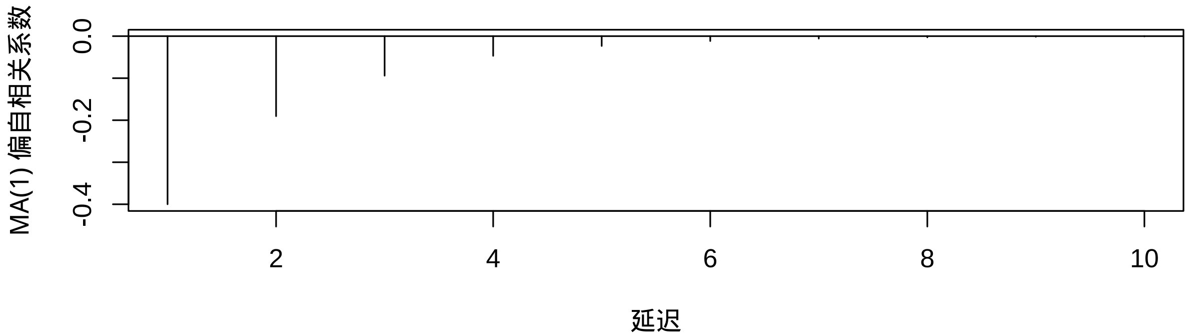
滞后 1 处的自相关系数看起来是合理的，但在滞后 7 和 14 处也出现了看似 “显著” 但实际上虚假的相关性。

c

对于 MA(1) 模型，偏自相关系数 (ϕ_{kk}) 的公式定义为：

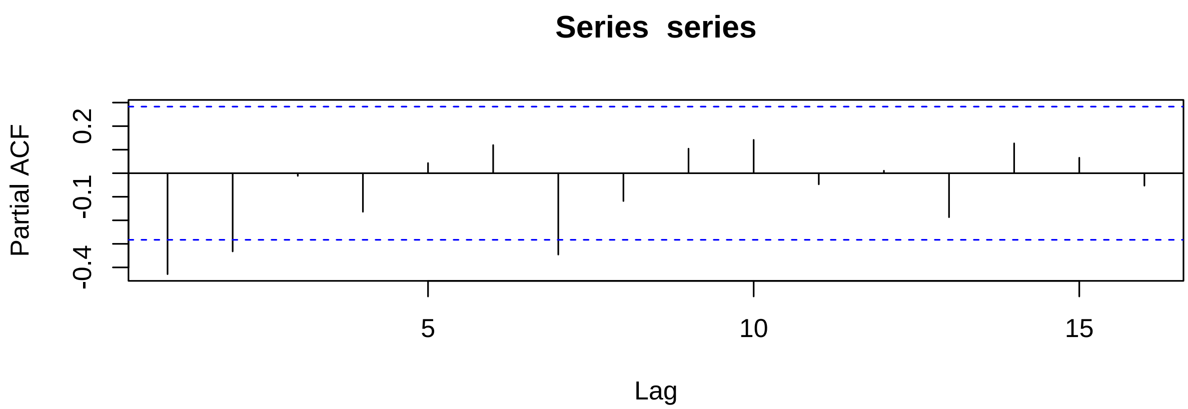
$$\phi_{kk} = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}} \text{ 当 } k \geq 1$$

```
theta=0.5; phikk=rep(NA,10)
for (k in 1:10) {phikk[k]=-(theta^k)*(1-theta^2)/(1-theta^(2*(k+1)))}
plot(phikk,type='h',ylab='MA(1) 偏自相关系数',xlab='延迟'); abline(h=0)
```



d

```
pacf(series)
```



只有前两阶的样本偏自相关系数与理论值匹配良好。然而，样本偏自相关系数的标准误大约为 $\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{48}} = 0.14$，表明在本例中仅滞后 7 的样本偏自相关系数超出了预期范围，其余的差异可视为抽样波动的结果。

0.9 练习 6.31

a

```
set.seed(15243); series=arima.sim(n=60,list(order=c(0,1,1),ma=-0.8))[-1]
library(urca); ur.df(series, type="none", selectlags="AIC")
```

```
## Warning: 程辑包'urca'是用R版本4.3.3 来建造的
```

```
##
## ###############################################################
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## ###############################################################
##
## The value of the test statistic is: -2.6684
```

b

```
ar(diff(series))
```

```
##
## Call:
## ar(x = diff(series))
##
## Coefficients:
##       1        2        3
## -0.7355  -0.5275  -0.3961
##
## Order selected 3  sigma^2 estimated as  0.8229
```

```
ur.df(diff(series), type="none", selectlags="AIC")
```

```
##
## ###############################################################
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## ###############################################################
##
```

```
## The value of the test statistic is: -8.356
```

c

```r
ur.df(diff(series), type="none", selectlags="AIC")
```

```
##
## ###############################################################
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## ###############################################################
##
## The value of the test statistic is: -8.356
```

```r
ar(diff(diff(series)))
```

```
##
## Call:
## ar(x = diff(diff(series)))
##
## Coefficients:
##       1        2        3        4        5
## -1.4156  -1.4512  -1.2708  -0.7384  -0.2884
##
## Order selected 5  sigma^2 estimated as  1.203
```

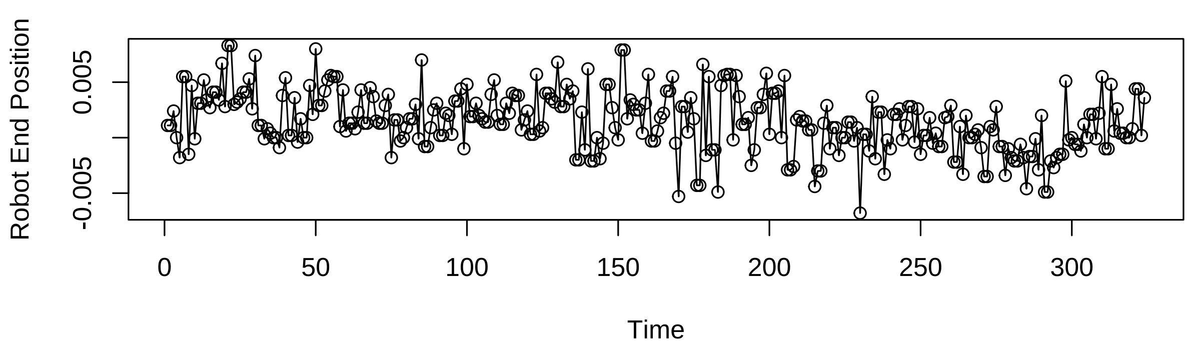
```r
ur.df(diff(series), type="none", selectlags="AIC")
```

```
##
## ###############################################################
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## ###############################################################
##
## The value of the test statistic is: -8.356
```

0.10 练习 6.36

a

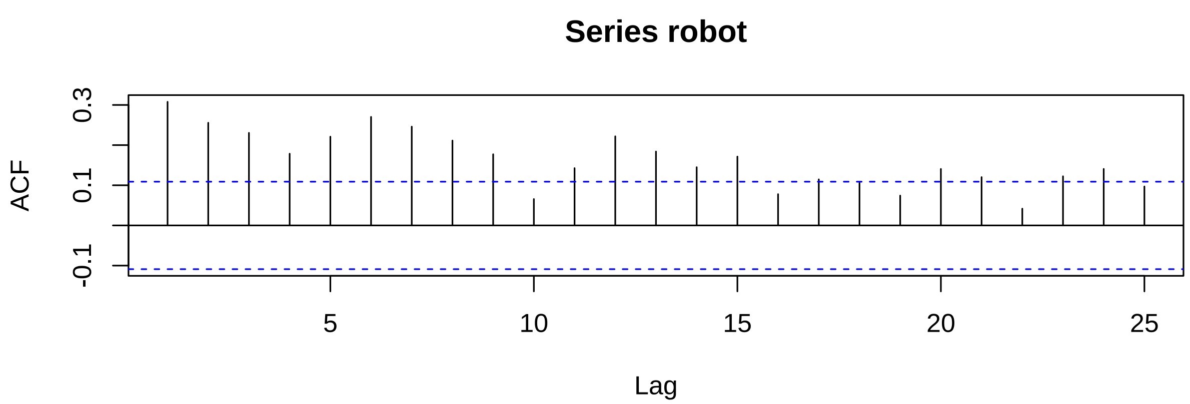
```r
data(robot)
plot(robot,type='o',ylab='Robot End Position')
```



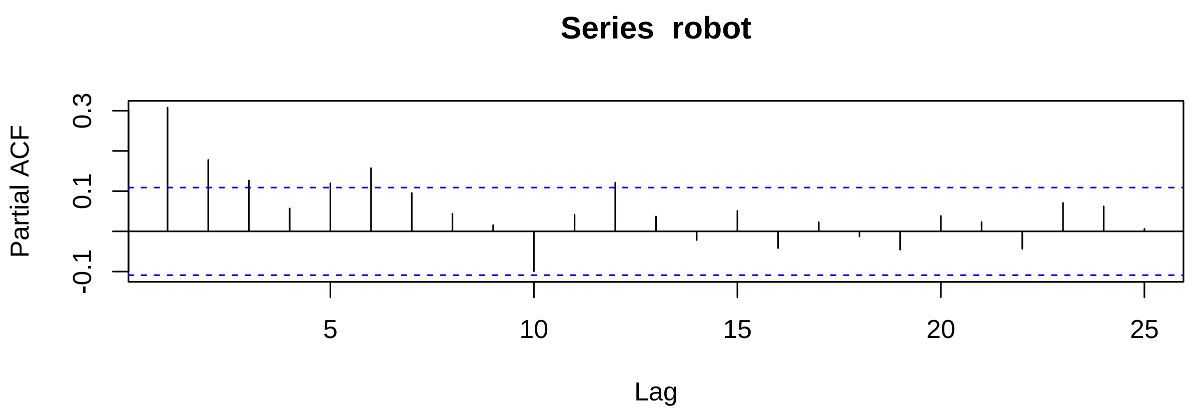
从这个图中，我们可能会尝试使用一个平稳模型，但同时也存在着足够的“漂移”现象，这使得我们怀疑序列可能存在非平稳性。

b

```r
acf(robot)
```




```r
pacf(robot)
```



这些图并不是特别明确，但偏自相关图（pacf）暗示了该序列可能适合一个 AR(3) 模型。

c

```r
eacf(robot)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x x x x x x x x o x  x  x  x
## 1 x o o o o o o o o o o  o  o  o
## 2 x x o o o o o o o o o  o  o  o
## 3 x x o o o o o o o o o  o  o  o
## 4 x x x x o o o o o o o  o  x  o
## 5 x x x o o o o o o o o  o  x  o
## 6 x o o o o x o o o o o  o  o  o
```

```
## 7 x o o x o x x o o o o  o  o  o
```

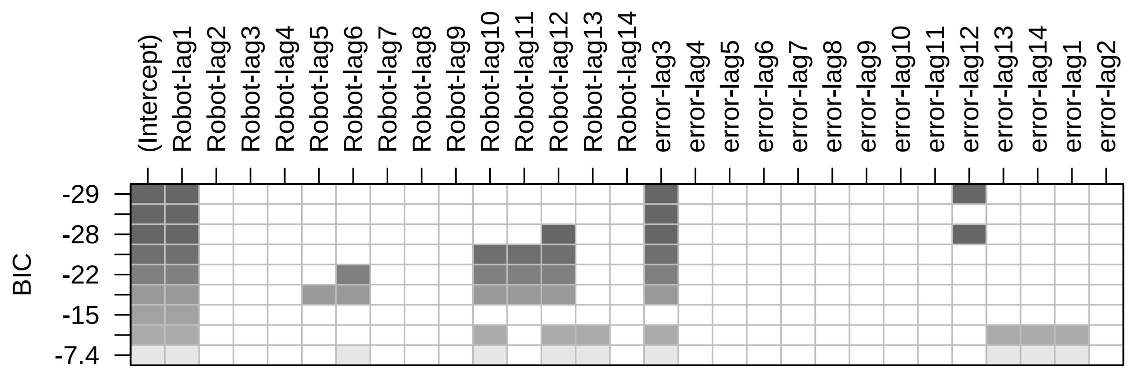
误差修正函数（EACF）则提示了一个 ARMA(1,1) 模型。

d

```r
plot(armasubsets(y=robot,nar=14,nma=14,y.name='Robot',ar.method='ols'))
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =
## force.in, : 2 linear dependencies found
```

```
## Reordering variables and trying again:
```



在这里，最佳模型包括了序列部分的一个 1 阶自回归项（AR term），以及移动平均部分（MA part）中的 3 阶和 12 阶滞后。