Course: Computer statistics

Lecture 12: ARIMA models.

Lecturer: Oleksandr Dykhovychnyi

Лекція 11. Моделі ARIMA

I. Модель
II. Побудова моделі x2
III. Приклад побудови моделі й прогнозу

Розглянуті у попередній лекції підходи до дослідження й прогнозування часового ряду на підставі згладжування не спирались на дослідження ймовірнісних механізмів, якими спричинена поведінка ряду. На дослідженні статистичних зав'язків між членами ряду побудовані лінійні моделей, до яких у першу чергу відносяться моделі **ARIMA** або моделі **Дженкінса-Бокса**.

I. Модель

Будемо говорити, що часовий ряд $X(t)$ задовольняє рівняння **авторегресії та ковзного середнього** або відповідає моделі **ARMA** (autoregressive moving average), якщо він задовольняє рівняння:

$$X(t) = \sum_{i=1}^{p} a_i X(t-i) + \sum_{i=1}^{q} b_i \varepsilon(t-i) + \varepsilon(t) + a_0 ,$$

де

- p - порядок авторегресії;
- $a_1, a_2, ..., a_p$ -коефіцієнти авторегресії;
- q -порядок ковзного середнього;
- $b_1, b_2, ..., b_q$ - коефіцієнти ковзного середнього;
- a_0 - деяка стала;
- $\varepsilon(t)$ - «білий шум», послідовність однаково розподілених

незалежних випадкових величин з нульовим середнім і дисперсією $\sigma_\varepsilon^{\ 2}$.

Часовий ряд називають **стаціонарним**, якщо середнє ряду є сталим

$$EX(t) = Const \ ,$$

а кореляційна функція

$$R(i) = Cov_{X(t)X(t+i)}$$

залежить лише від відстані між елементами ряду $X(t)$ та $X(t+i)$. Умова стаціонарності для моделі АРСС полягає у наступному:
Корені характеристичного рівняння
x2

$$1 - a_1 x - a_2 x^2 - \ldots - a_p x^p = 0$$

лежать за межами одиничного кола.

У випадку, коли часовий ряд не є стаціонарним, іноді вдається замінити його рядом з різниць двох сусідніх значень. Якщо ця процедура не допомагає, то її треба повторити. Тоді, одержаний в результаті такої процедури часовий ряд, описуються моделлю **ARIMA** (**autoregressive integrated moving average**), авторегресії та проінтегрованого ковзного середнього. Таким чином, побудова моделі **ARIMA** зводиться до побудови моделі **ARMA**.

II. Побудова моделі

Задача побудови моделі **ARMA** передбачає наступні кроки:

- підбір оптимальних порядків p і q ;
- оцінювання коефіцієнтів $a_1, a_2, \ldots, a_p$, $b_1, b_2, \ldots, b_q$;
- оцінювання $\sigma_\varepsilon^{\ 2}$.

Зазвичай розв’язання цих задач передбачає наступну процедуру. Для кожної можливої пари порядків (p, q) складають систему Юла-Уокера:

$$R(q+i) = \sum_{j=1}^{p} a_j R(|q+i-j|), \ i = \overline{1, p} \ .$$

Для знаходження коефіцієнтів ковзного середнього, також складають нелінійну систему, розв'язуючи яку, знаходять $b_1, b_2, ..., b_q$, $\sigma_\varepsilon{}^2$. Далі вибирається та пара порядків (p^*, q^*), для якої певний інформаційний критерій, приміром, **AIC** (**Akaike Information Critera**) набуває мінімального значення:

$$AIC = \ln(\sigma_\varepsilon^2) + \frac{2(p+q)}{T} \ ,$$

або

$$BIC = \ln(\sigma_\varepsilon^2) + \frac{(p+q)\ln T}{T} \ .$$

де T- довжина інтервалу спостережень за часовим рядом.

x2

III. Приклад побудови моделі й прогнозу

Розглянемо на прикладі процедуру побудови моделі **ARIMA** та прогнозування на прикладі часового ряду тривалості життя англійських королів.

Приклади:

```
> # зчитуємо часовий ряд
> kings <- scan("http://robjhyndman.com/tsdldata/misc/kings.dat"
                        ,skip=3   )
Read 42 items
> kingsts <- ts(kings) # формуємо часовий ряд
> kingsts
Time Series:
Start = 1
End = 42
Frequency = 1
 [1] 60 43 67 50 56 42 50 65 68 43 65 34 47 34 49 41 13 35 53 56 16 4
3 69 59 48
[26] 59 86 55 68 51 33 49 67 77 81 67 71 81 68 70 77 56
> plot.ts(kingsts) # будуємо графік
```

Будуємо графік ряду kingsts (Рис. 1)

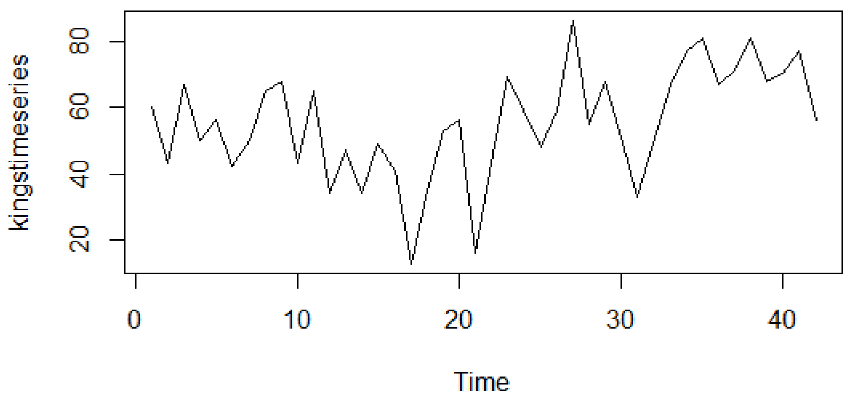

Рис.1(Рисунок згенеровано RStudio).

За виглядом графіка (Рис.1) видно, що цей ряд «несхожий» на стаціонарний. Спробуємо його «покращити» шляхом взяття різниць. Це реалізує функція **diff().**

Приклади:

> kingstsdiff1 <- diff(kingsts, differences=1) # бере
#мо різниці першого порядку
> plot.ts(kingstsdiff1) #будуємо графік

Будуємо графік ряду kingstsdiff1 (Рис. 2)

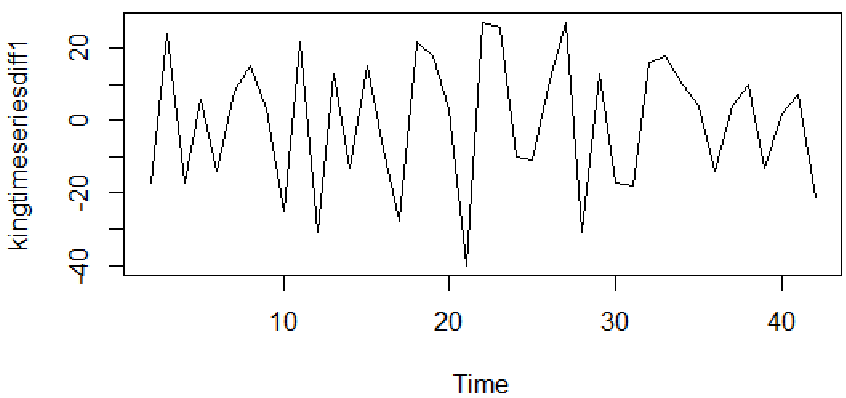

Рис.2 (Рисунок згенеровано RStudio).

Очевидно, що ряд став більш схожим на стаціонарний. Його характеристики:

Приклади:

```
>summary(kingstsdiff1)
Min.       1st Qu.     Median     Mean    3rd Qu.     Max.
-40.00000 -14.00000   4.00000  -0.09756  15.00000  27.00000
```

Далі оцінимо кореляційну функцію ряду та побудуємо її графік (Рис. 3).

Приклади:

```
> acf(kingstsdiff1, lag.max=20)
> acf(kingstsdiff1, lag.max=20, plot=FALSE)

Autocorrelations of series ‘kingstsdiff1’, by lag

    0      1      2      3      4      5      6      7      8      9     10
1.000 -0.360 -0.162 -0.050  0.227 -0.042 -0.181  0.095  0.064 -0.116 -0.071
   11     12     13     14     15     16     17     18     19     20
0.206 -0.017 -0.212  0.130  0.114 -0.009 -0.192  0.072  0.113 -0.093
```

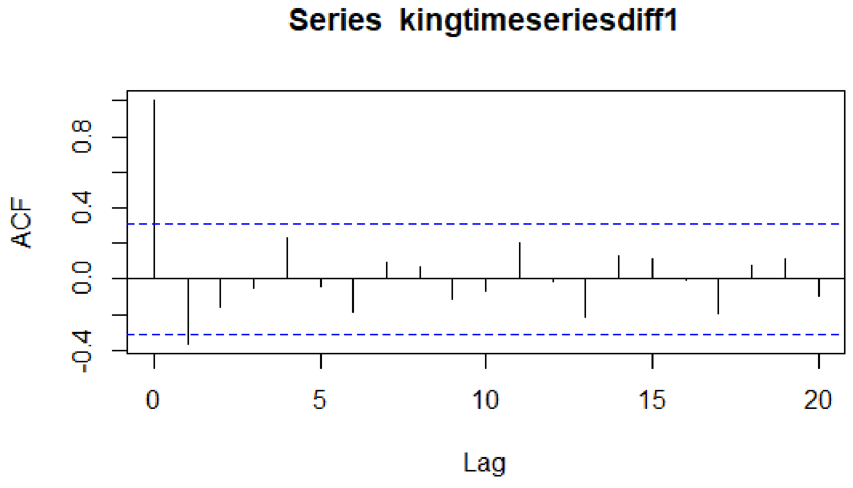

Рис.3 (Рисунок згенеровано RStudio).

З графіка кореляційної функції (Рис.3) видно, що кореляція із затримкою 1 (-0,360) перевищує межі значущості, а всі інші кореляції між лагами 1-20 не перевищують границі значущості. Тобто, ймовірний порядок авторегресії – 1.

Зробимо автоматичний підбір моделі **ARIMA(p,d,q)** (p- порядок авторегресії, d - порядок різниці, q-порядок ковзного середнього) за допомогою інформаційного критерію. Для цього потрібно підключити пакет **forecast**, у якому це реалізовано функцією **auto.arima ()**.

Приклади:

```
> library("forecast")
> install.packages("forecast")
> auto.arima (kingsts)
Series: kingsts
ARIMA(0,1,1)

Coefficients:
         ma1
     -0.7218
s.e.  0.1208

sigma^2 estimated as 236.2:  log likelihood= -170.06
AIC = 344.13   AICc=344.44   BIC=347.56
```

Як бачимо з резюме, оптимальною моделлю ряду є **ARIMA(0,1,1)**, тобто, порядок авторегресії p=0, порядок різниці d=1, порядок ковзного середнього q=1. Коефіцієнт ковзного середнього $b_1 = -0.7218$, $\sigma_\varepsilon = 0.1208$. Отже, модель має вигляд:

$$X(t) - X(t-1) = \varepsilon(t) - 0.7218\varepsilon(t-1) \ .$$

Далі на підставі побудованої моделі будуємо прогноз.

Приклади:

```
> # прив’язуємо модель до ряду
> kingstsarima <- arima(kingsts, order=c(0,1,1))
> kingstsforecasts <- forecast(kingstsarima, h=5)
> kingstsforecasts # друкуємо прогноз у п’яти точках
Point  Forecast     Lo 80     Hi 80     Lo 95      Hi 95
43     67.75063 48.29647 87.20479 37.99806  97.50319
44     67.75063 47.55748 87.94377 36.86788  98.63338
45     67.75063 46.84460 88.65665 35.77762  99.72363
46     67.75063 46.15524 89.34601 34.72333 100.77792
47     67.75063 45.48722 90.01404 33.70168 101.79958
> plot(kingstsforecasts) # будуємо графік прогнозу
```

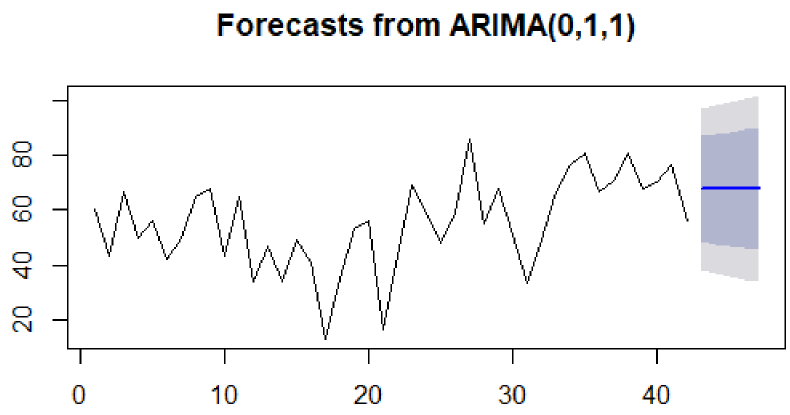

Рис.4 (Рисунок згенеровано RStudio).

x2

Прогноз (Рис.4) зображено у вигляді синьої лінії, інтервал прогнозування 80% - заштрихована область фіолетового кольору, а інтервал прогнозування 95% - як заштрихована сіра область.