



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

ARIMA как инструмент для предсказания будущего поведения временного ряда

Шульман Александра
ФКН ПИ, 3 курс

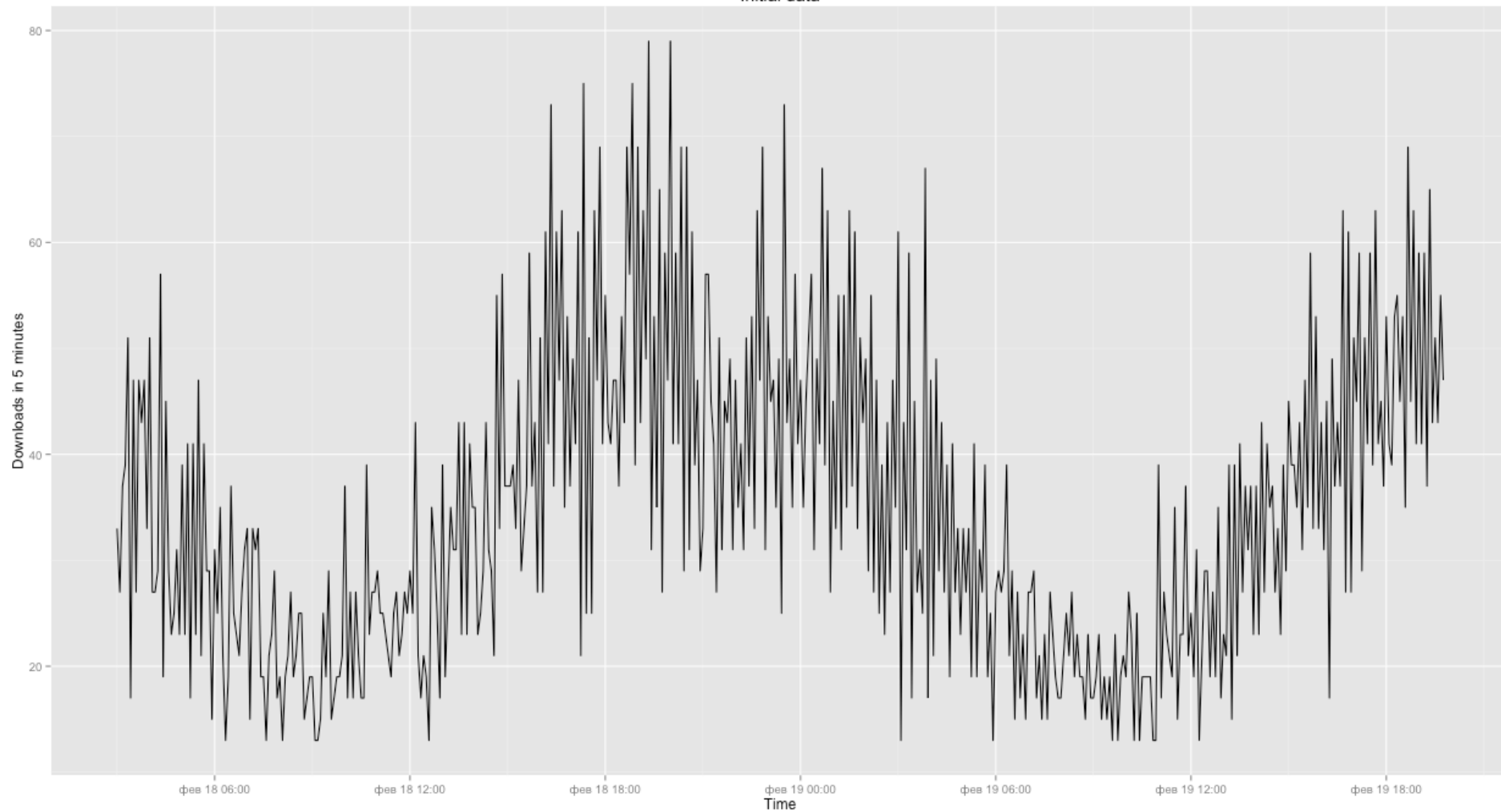
Временными рядами называют дискретные последовательности чисел, характеризующих состояние объекта наблюдения в отдельные моменты времени.

Ограничения и обозначения:

- Δt (интервал выборки) должен быть одинаковым на протяжении всего временного ряда;
- v_i – точка временного ряда, значение наблюдаемой величины в дискретный момент времени;
- i – порядковый номер точки или дискретное время;
- N (общее количество точек) – длина временного ряда;
- $\{v_i\} = v_1, v_2, \dots, v_N$ – длина временного ряда.

Пример временного ряда

Initial data



- 1. Мы будем считать наши наблюдаемые процессы *случайными*.**
Отсюда следует, что мы не можем использовать динамические модели. Вместо этого мы будем использовать *статистические* модели и аппараты статистики и теории вероятностей.
- 2. ARIMA — это пример эмпирической модели, конструируемой на экспериментальных данных.**
- 3. ARIMA использует структурные методы анализа временного ряда.**
Эти методы, в отличие от неструктурных, ориентированы на оценку по небольшому количеству параметров при дополнительных предположениях о свойствах наблюдаемых величин.

Постановка задачи

Пусть имеется скалярный временной ряд $\{v_i\}_{i=1}^N$. Необходимо найти (предсказать) значения величины v в моменты времени $t > t_N$.

Экстраполяция, а точнее, регрессия, не подходит для описания большинства временных зависимостей, так как при этом предполагают явный вид функциональной зависимости v от t .

Выход: моделирование ряда в виде случайного процесса $v = v(t_i, \omega)$, где ω — случайное событие (отсюда v_i — случайная величина).

Предположим, что базовой моделью нашего временного ряда является нормальный белый шум $a(t)$.

Значения в такой модели в различные моменты времени независимы, а значение в любой фиксированный момент времени распределено по нормальному закону:

$$p(a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{a^2}{2\sigma^2}}$$

Для такой модели среднее значение принято считать равным нулю, значение автокорреляционной функции при любой задержке (ненулевой) равно нулю. Остается оценить только дисперсию.

Однако чаще всего получается, что мы не можем применить модель нормального белого шума к выборке (например, значения автокорреляционной функции отличны от нуля). Тогда будем говорить о процессе как о нормальном белом шуме, преобразованном линейным фильтром:

$$v_n = a_n + \sum_{i=1}^{\infty} \psi_i a_{n-i}, \quad \sum_{i=1}^{\infty} \psi_i^2 \leq \text{const}$$

Модель скользящего среднего (Moving Average) Модель авторегрессии (AutoRegressive Model)

Несложно понять, что с увеличением i значения ψ_i быстро убывают, следовательно можно ограничиться количеством весов q .

$$v_n = a_n - \sum_{i=1}^q \theta_i a_{n-i}$$

Переписывая выражение нормального белого шума, преобразованного линейным фильтром, приходим к формуле:

$$v_n = a_n + \sum_{i=1}^{\infty} \pi_i v_{n-i}$$

Снова понимаем, что бесконечное количество весов нам не нужно, ограничиваемся количеством весов p :

$$v_n = a_n + \sum_{i=1}^p \phi_i v_{n-i}$$

Соединив модель скользящего среднего и модель авторегрессии, получим модель ARMA (p, q):

$$v_n = a_n + \sum_{i=1}^p \phi_i v_{n-i} - \sum_{i=1}^q \theta_i a_{n-i}$$

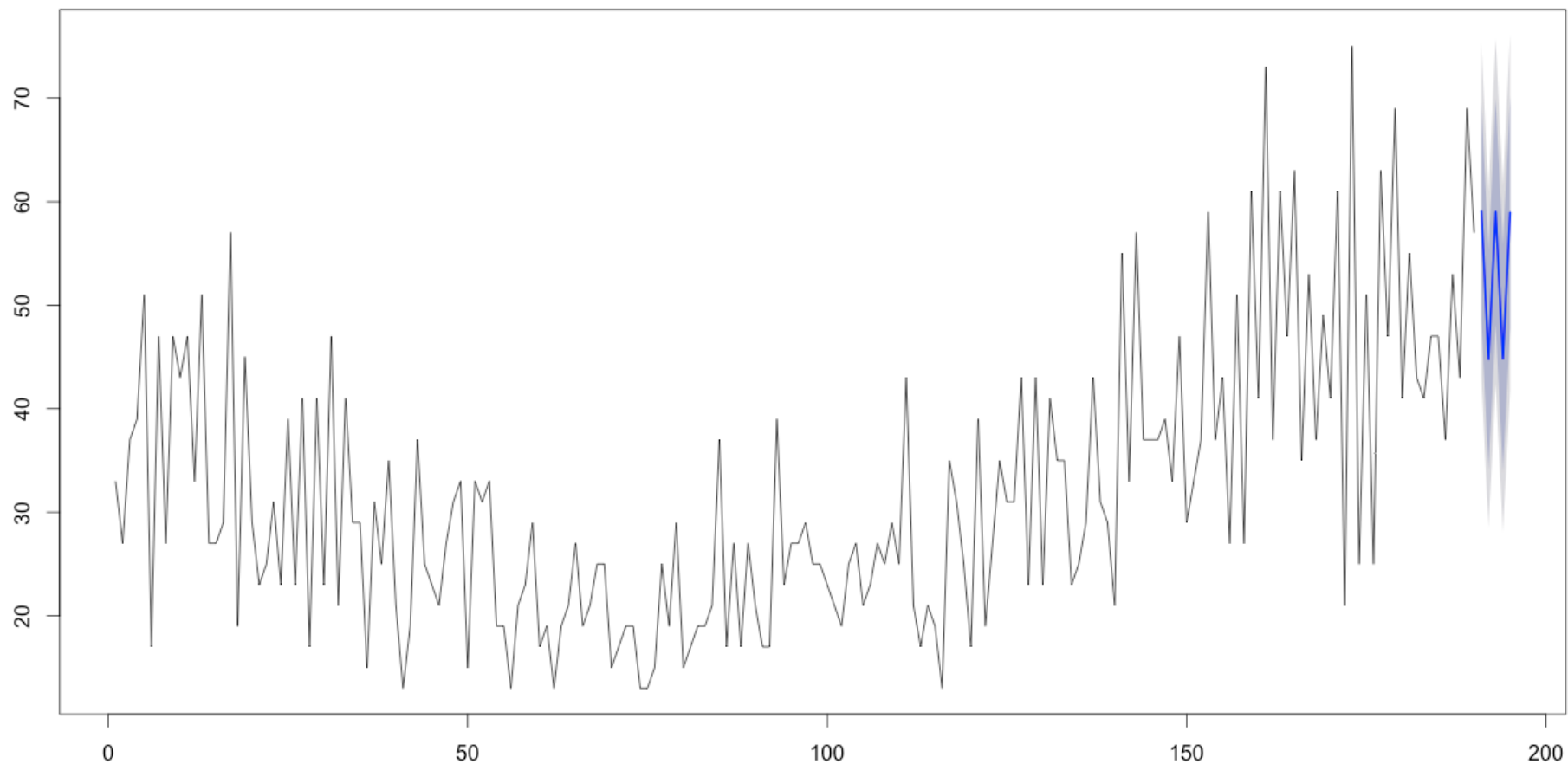
Проблема заключается в том, что если наш ряд имеет признаки нестационарности, то наша модель не может быть адекватной. Но стационарной может быть некоторая разность наблюдаемого процесса порядка d :

$$w_n = \nabla^d v_n, \quad \nabla v_n = v_n - v_{n-1}$$

Переходим к ARIMA (p, q, d) (AutoRegressive Integrated Moving Average):

$$w_n = a_n + \sum_{i=1}^p \phi_i w_{n-i} - \sum_{i=1}^q \theta_i a_{n-i} + \mu$$

Forecasts from ARIMA(1,1,2)





НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Go to file/function

Project: (None)

Environment History

Import Dataset Clear

Global Environment

values

fit

List of 16

coef : Named num [1:3] -0.9952 0.0871 -0.7571
..- attr(*, "names")= chr [1:3] "ar1" "ma1" "ma2"
sigma2 : num 67.2
var.coef : num [1:3, 1:3] 5.01e-05 -1.61e-04 1.50e-04 -1.61e-04 2.50e-03 ...
..- attr(*, "dimnames")=List of 2
.. ..\$: chr [1:3] "ar1" "ma1" "ma2"
.. ..\$: chr [1:3] "ar1" "ma1" "ma2"
mask : logi [1:3] TRUE TRUE TRUE
loglik : num -667
aic : num 1343
arma : int [1:7] 1 2 0 0 1 1 0
residuals: Time-Series [1:190] from 1 to 190: 0.033 -2.966 4.312 8.596 14.717 ...
call : language auto.arima(x = structure(list(x = structure(c(33, 27, 37, 39, 51, 17, 47, 27, 47, 43, 47, 33, 51, 27, 27, 29, 57, 19, 45, 29, 23, 25, 31, 23,...
series : chr "timeseries_date\$V2"
code : int 0
n.cond : int 0
model :List of 10
..\$ phi : num -0.995
..\$ theta: num [1:2] 0.0871 -0.7571
..\$ Delta: num 1
..\$ Z : num [1:4] 1 0 0 1
..\$ a : num [1:4] -12 -9.87 -12.25 69
..\$ P : num [1:4, 1:4] 0.00 0.00 0.00 -8.18e-17 0.00 ...
..\$ T : num [1:4, 1:4] -0.995 0 0 1 1 ...
..\$ V : num [1:4, 1:4] 1 0.0871 -0.7571 0 0.0871 ...
..\$ h : num 0
..\$ Pn : num [1:4, 1:4] 1.00 8.71e-02 -7.57e-01 -7.71e-17 8.71e-02 ...
bic : num 1356
aicc : num 1343
x : Time-Series [1:190] from 1 to 190: 33 27 37 39 51 17 47 27 47 43 ...
attr(*, "class")= chr [1:2] "ARIMA" "Arima"

mask : logi [1:3] TRUE TRUE TRUE

Console

timeseries_date\$V2
> plot(forecast(fit, 5))
> plot(forecast(fit, 20))
> plot(forecast(fit, 15))
> plot(forecast(fit, 3))
> plot(forecast(fit, 4))
> plot(forecast(fit, 5))

Files Plots Packages Help Viewer



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Спасибо за внимание!

101000, Россия, Москва, Мясницкая ул., д. 20

Тел.: (495) 621-7983, факс: (495) 628-7931

www.hse.ru