Анализ данных

Хашин С.И.
http://math.ivanovo.ac.ru/dalgebra/Khashin/index.html

Ивановский университет

.

ARIMA, SARIMA, ...

.

Иваново-2023

План

Сглаживание

Полиномиальное сглаживание

Стационарность

ARIMA

Сглаживание

Двойное экспоненциальное сглаживание

Пусть $X = (x_i)$ — временной ряд и пусть Y — его экспоненциальной сглаживание с параметром α:

$$Y_i = \alpha X_i + (1 - \alpha) Y_{i-1}$$

На Питоне:

```
# Экспоненцильное сглаживание вектора x с параметром alpha
def exp_smoothing(x, alpha):
  x_smoth = np.zeros(len(x))
  x_smoth[0] = x[0]*alpha
  for i in range(1, len(x)):
    x_smoth[i] = alpha*x[i] + (1-alpha)*x_smoth[i-1]
  return x_smoth

alpha = 1/30
Y = exp_smoothing(X, alpha)
```

Двойное экспоненциальное сглаживание

Пусть Z — экспоненциальное сглаживание ряда из разностей $Y_i - Y_{i-1}$ с параметром β

```
beta = 1/100
Z = exp_smoothing(Y[1:]-Y[:-1], beta)
```

Это, в некотором смысле, даёт разбиение ряда на две составляющие — уровень Y_i и тренд Z_i.

Полиномиальное сглаживание

Будем аппроксимировать временной ряд $X = (x_n)$ многочленом степени d:

$$x_n \approx w_0 + w_1 n + w_1 n^2 + \dots w_d n^d.$$

$$w_0 \approx x_0,$$

$$w_0 + w_1 + w_1 + \dots w_d \approx x_1,$$

$$w_0 + w_1 2 + w_1 2^2 + \dots w_d 2^d \approx x_2,$$

$$w_0 + w_1 3 + w_1 3^2 + \dots w_d 3^d \approx x_2,$$

$$\dots$$

Полиномиальное сглаживание

Или, в матричном вида: $Aw = b$, где

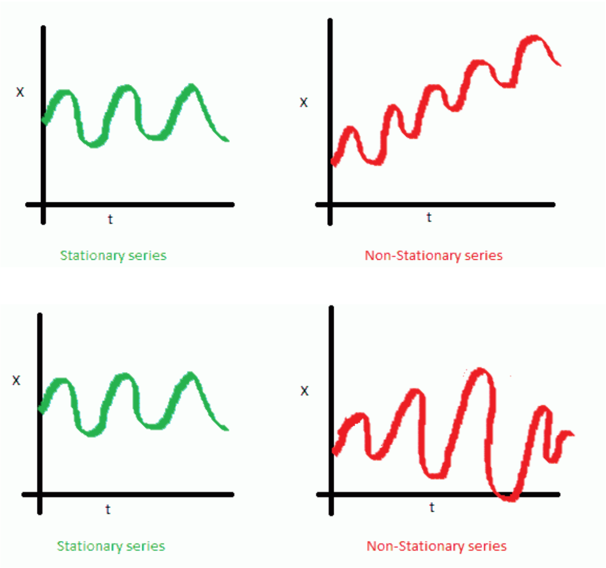
$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2^2 & \dots & 2^d \\ & & \dots & & \end{pmatrix}$$

```
def matr_pow(d,N):
  A = np.zeros((N, d+1))
  A[:,0] = 1
  for i in range(1, N):
    A[i] = np.array([i**j for j in range(d+1)])
  return A

print(matr_pow(2,5))
```

Полиномиальное сглаживание

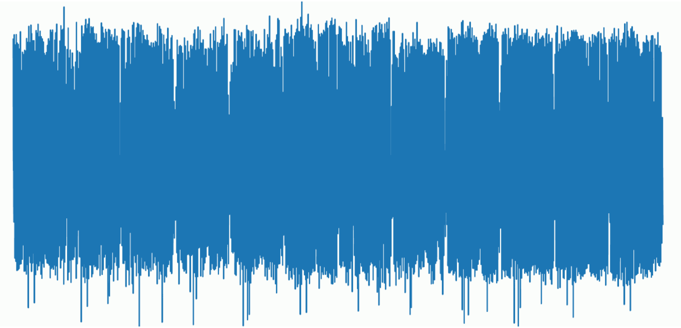
```
# приближение ряда Y многочленом степени d>=0
def poly_approx(Y, d):
   N = len(Y)
   A = matr_pow(d, N)
   # система уравнений: A*w ~= Y
   AtA = A.T.dot(A)
   AtY = A.T.dot(Y)
   w = np.linalg.solve(AtA, AtY)
   return w

Y = [2+2*n+n*n for n in range(10)]
print(poly_approx(Y,2))
```

Стационарность



Стационарность разностей

```
X = np.load("time_series\\GermanE.npz")['data'][:, 0]
Y = X[1:]-X[:-1]
plt.plot(np.arange(len(Y)), Y)
plt.show()
```



Белый шум

Белый шум — это последовательность независимых нормальных СВ $(\xi_1, \xi_2 \dots)$ с МО=0 и дисперсией σ^2. В каких пределах должно лежать среднее значение белого шума за N шагов? Согласно центральной предельной теореме сумма $\xi_1 + \dots + \xi_N$ имеет МО ноль и дисперсию $N\sigma^2$. Среднее значение

$$\frac{\xi_1 + \dots + \xi_N}{N}$$

имеет МО $= 0$ и дисперсию σ^2/N, то есть ср.кв.отклонение равно $\sigma/\sqrt{N}$. По правилу 3σ, среднее арифметическое N значений белого шума должно лежать в пределах $\pm 3\sigma/\sqrt{N}$. Например, при $N = 100$ получаем интервал $\pm\sigma/3$, при $N = 1000$ около $\pm\sigma/10$.
Значение ковариационной функции $covar(k)$ должно быть около 0 при $k > 0$.

ARIMA

```
import statsmodels.api as sm
...
sm.tsa.ARIMA(src_data_model, order=(p,d,q),
   freq='W').fit(full_output=False, disp=0)
```

p — порядок компоненты AR
d — порядок интегрированного ряда
q — порядок компоненты MA

порядок интегрированного ряда

Если исходный ряд стационарен, полагаем $d = 0$. Иначе переходим к ряду из разностей. Если он стационарен, полагаем $d = 1$. Иначе переходим к ряду из вторых разностей и так далее.

ARIMA

Линейная рекуррентная последовательность:

$$x_n = a_{k-1}x_{n-1} + a_{k-2}x_{n-2} + \dots + a_0 x_{n-k}.$$

$$\lambda^k - a_{k-1}\lambda^{k-1} - a_{k-2}\lambda^{k-2} - \dots - a_0 = 0.$$

— характеристическое уравнение. Пусть $\{\lambda_1, \dots, \lambda_k\}$ — его корни. Тогда для любых a_j последовательность

$$x_n = a_1\lambda_1^n + \dots + a_k\lambda_k^n \ .$$

удовлетворяет рекуррентному соотношению.

ARIMA

Если хотя бы одни из корней характеристического уравнения (вещественный или комплексный) по модулю больше 1, то последовательность x_n будет быстро (экспоненциально) расти. Поэтому мы должны подобрать такие коэффициенты a_i, чтобы

$$x_n \approx a_{k-1}x_{n-1} + a_{k-2}x_{n-2} + \cdots + a_0x_{n-k}.$$

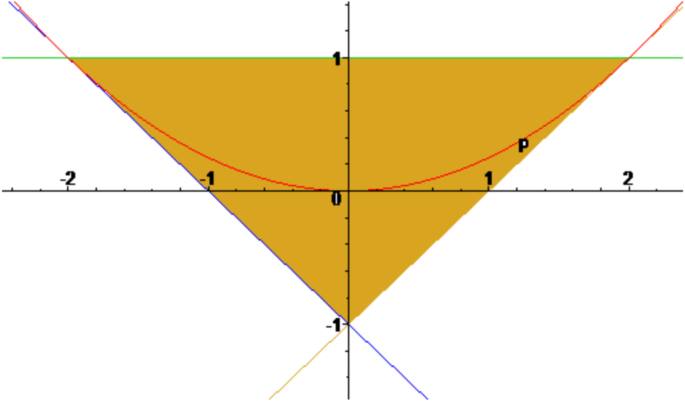
и при этом все корни характеристического многочлена

$$\lambda^k - a_{k-1}\lambda^{k-1} - a_{k-2}\lambda^{k-2} - \cdots - a_0 = 0.$$

по модулю ≤ 1.

ARIMA

Например, при $k = 2$, характеристический многочлен $f(\lambda) = \lambda^2 + p\lambda + q$. Допустимые (p, q):



ee

ee