Предсказание временных рядов

Авторегрессионные модели

Сергей Горбачев

13 ноября 2017 г.

НИУ ВШЭ ФКН

Содержание

Обозначения

$X = \{X_t, t \in T\} = X_1, X_2 \ldots$ - Временной ряд (случайный процесс)

ϵ_t - Белый шум (в широком смысле)

$L^k(X_t) = X_{t-k}$ - Лаговый оператор

$R(k) = cov(X_{t+k}, X_t)$ - Автоковариационная функция

$r(k) = \dfrac{cov(X_{t+k}, X_t)}{\sqrt{D[X_{t+k}]D[X_t]}}$ - Автокоррелляционная функция

Стационарный временной ряд

Ряд является стационарным, если вероятностные характеристики не зависят от времени t.
То есть $P(X_1, ..., X_t \in B) = P(X_{1+\tau}, ..., y_{X+\tau} \in B)$
Ряд слабо стационарен, $E[X_t]$ постоянно и не зависит от t и $cov(X_t, X_{t+\tau})$ зависит только от τ.

Модель скользящего среднего

Обозначение MA(q)

$$X_t = \mu + \epsilon_t + \sum_{i=1}^{q} b_i \epsilon_{t-i} = \mu + (1 + \sum_{i=1}^{q} b_i L^i)\epsilon_t \tag{1}$$

Где $b_1, \ldots, b_q$ параметры модели, а μ математическое среднее X_t (обычно считается нулевым)

Свойства

Для модели MA(q) известны такие характеристики

$$E[X_t] = \mu$$

$$D[X_t] = \sum_{k=0}^{q} b_k^2$$

$$R(k) = cov(X_{t+k}, X_t) = \begin{cases} \sum_{i=0}^{q-k} b_i b_{i+k} & \text{k} \le q \\ 0 & \text{k > q} \end{cases}$$

Процесс моделирования

Есть выборка $X_1, \ldots, X_n$

1. Считаем эмпирические характеристики

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X_n})^2$$

$$S(k) = \frac{1}{nS^2} \sum_{i=k+1}^{n} (X_i - \overline{X_n})(X_{i-k} - \overline{X_n})$$

2. Подбираем параметры модели используя выражения для теоретических характеристик
3. Оцениваем качество по отклонении теоретических характеристик от эмпирических

Обобщение

Мы рассмотрели модель с конечным q. Также можно ввести обобщение MA(q):

$$X_t = \mu + \sum_{i=0}^{\infty} b_i \epsilon_{n-i} \tag{2}$$

$$E[X_t] = \mu$$

$$D[X_t] = \sum_{k=0}^{\infty} b_k^2$$

$$cov(X_{t+k}, X_t) = \sum_{i=0}^{\infty} b_i b_{i+k}$$

Авторегрессионная модель

Обозначение AR(p)

$$X_t = a_0 + \sum_{i=1}^{p} a_i X_{t-i} + \epsilon_t = a_0 + \sum_{i=1}^{p} a_i L^i X_t + \epsilon_t \tag{3}$$

Модель зависит от начальных условий $X_{t-p}, \ldots, X_0$, обычно их обнуляют

AR(1)

В простейшем случае для AR(1)

$$E[X_t] = a_0/(1 - a_1)$$

$$R(0) = D[X_t] = D[\epsilon]/(1 - a_1^2)$$

Свойства AR(p)

Автоковариационная и автокорреляционная функции удовлетворяют рекуррентным соотношениям:

$$R(k) = \sum_{i=1}^{p} a_i R(k-i)$$

$$r(k) = \sum_{i=1}^{p} a_i r(k-i)$$

$$\alpha(L) = (1 - \sum_{i=1}^{p} a_i L^i)$$

Для того, чтобы процесс был стационарным, необходимо, чтобы корни характеристического многочлена авторегрессионной части $alpha(z)$ лежали вне единичного круга в комплексной плоскости (были по модулю строго больше единицы)

ARMA = AR + MA

Обозначение ARMA(p, q)

$$X_t = a_0 + \epsilon_t + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=1}^{q} b_i \epsilon_{t-i} \tag{4}$$

$$(1 - \sum_{i=1}^{p} a_i L^i) X_t - a_0 = (1 + \sum_{i=1}^{q} b_i L^i) \epsilon_t \tag{5}$$

ARMA

$$\alpha(L) = (1 - \sum_{i=1}^{p} a_i L^i)$$

$$\beta(L) = (1 + \sum_{i=1}^{q} b_i L^i)$$

Стационарность достигается тогда же как и в AR

ARMA

Стационарный ARMA-процесс можно представить как бесконечный MA-процесс:

$$X_t = \alpha^{-1}(L)c + \alpha^{-1}(L)\beta(L)\varepsilon_t = c/a(1) + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

ARMA

Например ARMA(1,0)=AR(1) можно представить через MA(∞)

$$X_t = c/(1-a) + \sum_{i=0}^{\infty} a^i \varepsilon_{t-i}$$

Обозначение ARIMA(p,d,q)

$$(1 - \sum_{i=1}^{p} a_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^{q} b_i L^1)\epsilon_t \tag{6}$$

ARIMA

Существует несколько известных частных случаев модели ARIMA. Например, ARIMA(0,1,0) является моделью случайных блужданий.

$$X_t = X_{t-1} + \epsilon_t$$

Конец

1. Модель скользящего среднего (MA)
2. Авторегрессионная модель (AR)
3. ARMA
4. ARIMA

Источники

- http://statsoft.ru/home/textbook/modules/sttimser.htmlspectrum
- https://onlinecourses.science.psu.edu/stat510/node/48
- http://www.machinelearning.ru/
- https://en.wikipedia.org/wiki/Autoregressive%E2%80%93moving-average$_{m}odelcite_{r}ef - 1$