Стохастический дискретный процесс

Рассмотрим последовательность случайных переменных $X_1, X_2, ..., X_T, ...$, каждая из которых имеет одно единственное возможное значение соответственно $t = t_1, t = t_2, ..., t = t_T, ...$ Такая последовательность называется **дискретным стохастическим** процессом X_t .

Для такого процесса математическое ожидание может изменяться во времени $M(X_t) = \mu_t = \mu(t)$

Аналогично дисперсия также может изменяться во времени

$D(X_t) = \sigma^2{}_t = \sigma^2(t)$

Автоковариация $\mathbf{cov}(X_{t_1}, X_{t_2})$ в общем виде зависит от каждого t_1 и t_2

Определение: Конечная реализация $x_1, x_2, ..., x_T$ дискретного стохастического процесса $...X_1, X_2, ..., X_T, ...$ называется **временным рядом.**

Понятие стационарности.

Опр. 1 (**строгой** стационарности или стационарности в узком смысле) Случайный (стохастический) процесс является строго стационарным, если сдвиг во времени не меняет и одну из функций плотности распределения. То есть, если ко всем моментам времени прибавить некую целочисленную величину, то функция плотности при этом не изменится: $f_n(x_{t_1},\dots x_{t_n}) = f_n(x_{t_1+\Delta},\dots x_{t_n+\Delta})$ для всех n, моментов времени $x_{t_1},\dots x_{t_n}$ и целочисленном Δ.

Опр. 2 (**слабой** стационарности или стационарности в широком смысле) Случайный процесс называются стационарным в широком смысле, если он обладает постоянной средней и дисперсией (то есть дисперсия и математическое ожидание не зависят от времени), а ковариация зависит только от временного интервала между двумя отдельными наблюдениями.

То есть случайный процесс v_t называют *стационарным*, если для него выполняются следующие условия:

1) $M(v_t)= M(v_{t+k})=m=const;$

2) $D(\varepsilon_t)=\sigma^2=const;$

3) $cov(v_i,v_j)=cov(v_{i+l},v_{j+l})$ при любых $i\neq j$.

Соответственно временные ряды представленные строго стационарными и слабо стационарными случайными процессами называются стационарными временными рядами в узком и широком смысле.

Процесс «белого шума»

Опр. Процесс ε_t называется «белым шумом», если он удовлетворяет следующим трем требованиям:

1) $M(\varepsilon_t)=0;$

2) $D(\varepsilon_t)=\sigma^2=const;$

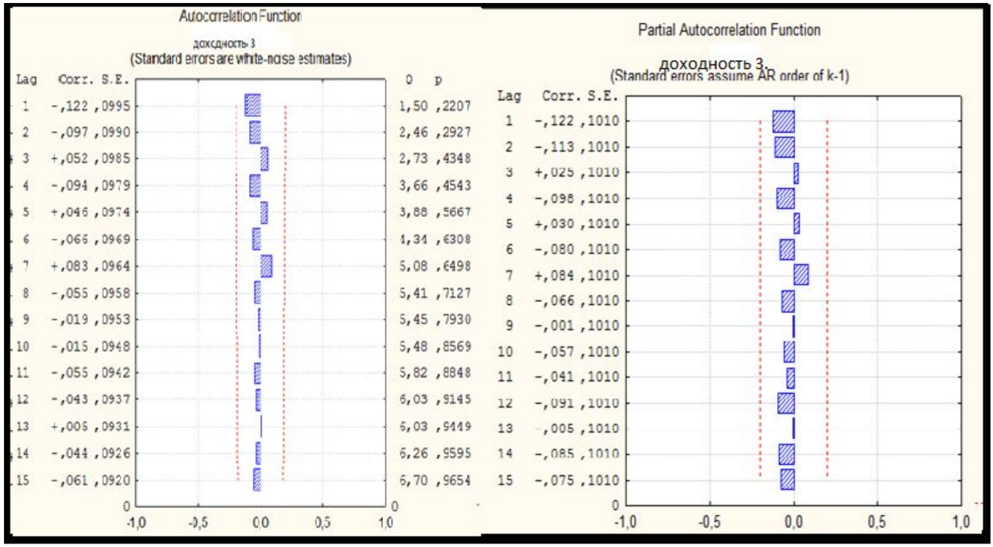
3) $cov(\varepsilon_t,\varepsilon_{t+k})=0$ при любых t и $k\neq 0$;

В случае, если процесс также удовлетворяет условию:

4) $\varepsilon_t \sim N(0;\,\sigma)$

То его называют Гауссовым белым шумом.

Для проверки на соответствие процесса «белому шум», достаточно проанализировать АКФ и ЧАКФ процесса и /или провести спектральный анализ. Теоретически белый шум имеет вид горизонтальной прямой с ординатой $\sigma^2/2\pi$



Модели скользящей средней

Опр.4. Модель скользящей средней (Moving Average) – это модель, где моделируемая величина задается линейной комбинацией от процессов белого шума, следующими друг за другом

$$Y_t = \varepsilon_t - b_1 \cdot \varepsilon_{t-1} - b_2 \cdot \varepsilon_{t-2} - \ldots - b_q \cdot \varepsilon_{t-q} \qquad \textbf{(3)}$$

Где q – количество лагов запаздывания называется порядком МА-модели.

Термин «скользящая средняя», используемый здесь не стоит путать с соответствующим термином, относящимся к непараметрическим методом поиска тренда.

Процессы скользящей средней – всегда стационарны в слабом смысле.

Введем понятие оператора сдвига:

$$L(X_t) = X_{t-1}$$

$$L^2(X_t) = X_{t-2}$$

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

Тогда процесс МА(q) можно записать как $Y_t = (1 - b_1 L - b_2 L^2 - \ldots - b_q L^q)\varepsilon_t$

Авторегрессионные процессы

<u>Опр. 3.</u> Авторегресонным (Auto-Regressive) называется процесс, при котором значение ряда находится в линейной зависимости от предыдущих значений.

Например, если текущее наблюдаемое значение является функцией всего лишь одного значения, непосредственно предшествующего наблюдению, то есть процесс зависит всего лишь от одного значения рассматриваемой переменной, то процесс называется авторегрссионным процессом первого порядка и обозначается AR(1). Это можно обобщить следующим образом: если анализируемый динамический процесс зависит от 1 до *p* временных лагов назад, то это авторегрессионный процесс порядка *p*, т. е. AR(*p*):

$$Y_t = \alpha_0 + \alpha_1 \cdot Y_{t-1} + \alpha_2 \cdot Y_{t-2} + \ldots + \alpha_p \cdot Y_p + \varepsilon_t \tag{1}$$

Здесь текущее значение Y – функция от p наиболее недавних предыдущих значений. p– порядок авторегрессии.

ε_t- процесс белого шума

Авторегрессионные процессы

Введем функцию **оператора сдвига**:

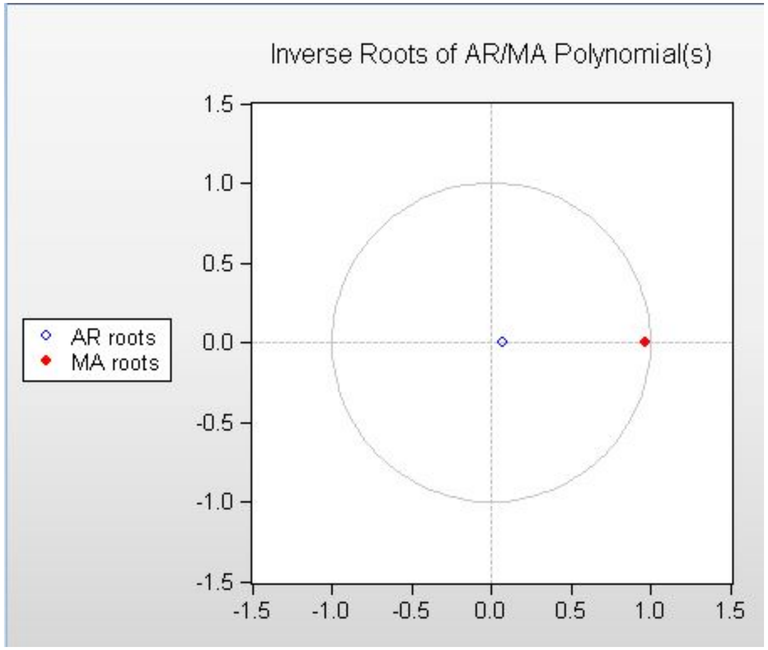
$$f_p(L) = 1 - \alpha_0 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_n L^p$$

Тогда процесс AR(p) можно записать в виде: $f_p(L) Y_t = \varepsilon_t$

Теорема: AR-процесс является стационарным тогда и только тогда, когда корни соответствующего ему характеристического уравнения

$$1 - \alpha_1 \cdot z - \alpha_2 \cdot z^2 - \ldots - \alpha_p \cdot z^p = 0$$ лежат вне единичного круга, то есть $|z_i| < 1$.

Тесты на наличие единичных корней: для АР-процесса выписывается характеристическое уравнение, разрешая которое определяют, входят ли его корни (комлекснозначные) в единичный круг или нет. Если все корни уравнения лежат вне круга, то процесс представленный АР-процессом является стационарным.



Определение: Процесс вида $Y_t = Y_{t-1} + \varepsilon_t$ называется процессом случайного блуждания или процессом единичного корня $|z| = 1$

Авторегрессионные модели скользящей средней (АРСС)

ОПР. Модели временных рядов, которые сочетают авторегрессионный процесс с моделью скользящей средней называются авторегрессионными моделями скользящей средней (АРСС, ARMA). Модель АРСС(p,q) имеет p временных лагов в авторегрессионном процессе и q интервалов в модели скользящей средней.

$$\hat{Y}_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \dots - b_q \varepsilon_{t-q} + \varepsilon_t \quad (4)$$

где ε_t остаточный член ошибки – белый шум.

Так АРСС(3,2) имеет вид:

$$\hat{Y}_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + a_3 Y_{t-3} - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} + \varepsilon_t$$

ARMA процесс моет быть представлен при самых общих допущениях как бесконечный AR-процесс или как бесконечный MA-процесс.

При анализе временных рядов рекомендуется выбирать конечные АРСС-процессы с наименьшим числом параметров, подлежащих оценке Как правило $p+q \le 4$.

Идентификация модели АРСС

Опр. Пусть дана модель АРСС(*p*, *q*) для временного ряда Y_t . Идентификацией этой модели называется процедура определения неизвестных значений *p* и *q*.

Существует несколько подходов к идентификации моделей АРСС.

I подход:

Идентификацию модели АРСС проводят на визуальном анализе коррелограммы АКФ и частичной коррелограммы ЧАКФ.

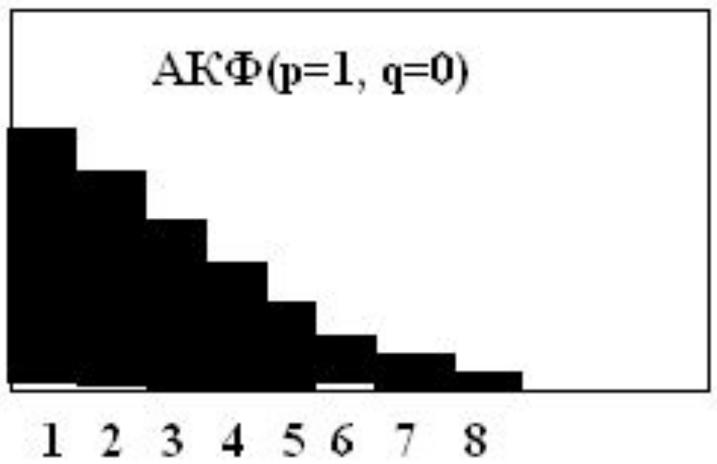
Здесь используются следующие особенности АКФ:

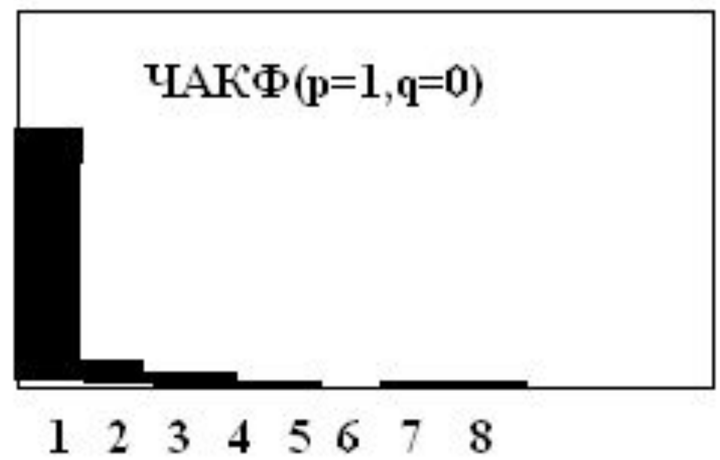
- В случае АР моделей ее модуль убывает по экспоненте, осциллируя около нуля;
- В случае СС(*q)* модели только первые q значений отличны от нуля;
- В модели АРСС(*p, q*), после *q-p* значений, АКФ имеет вид, такой же, как АР модель.

Свойства поведения ЧАКФ:

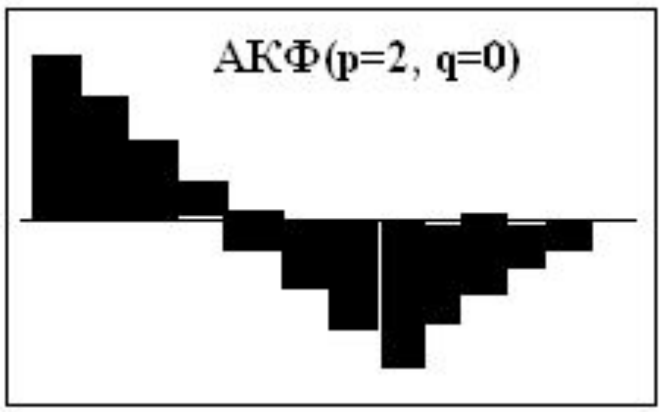
- Для АР(*p*) моделей, она равна нулю после первых *p* значений.
- Для моделей СС она экспоненциально убывает по модулю.
- Для АРСС(p,q) моделей, после первых p-q значений,она ведет себя также как и для модели СС.

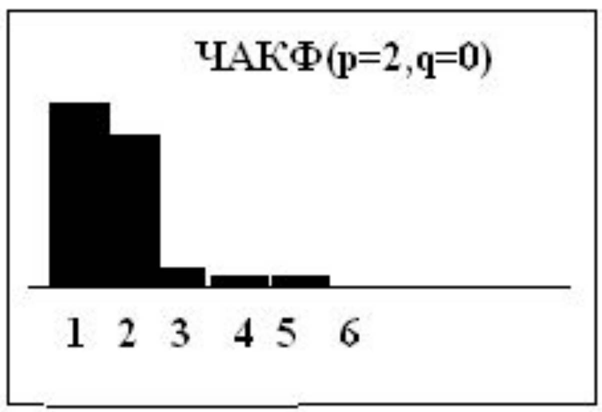
Идентификация модели АРСС (I подход)



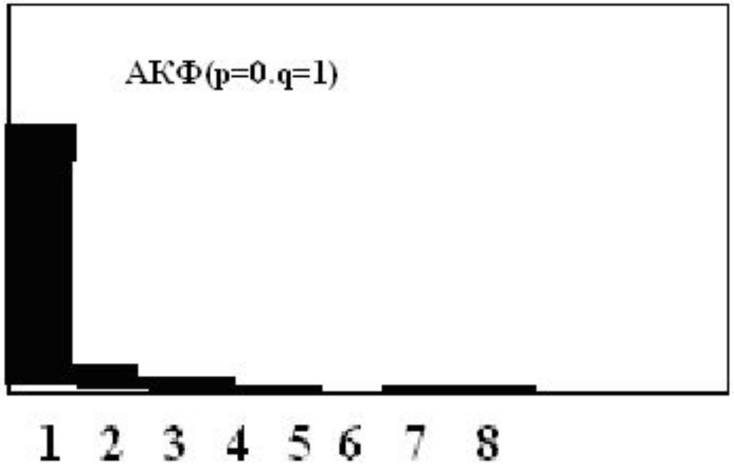


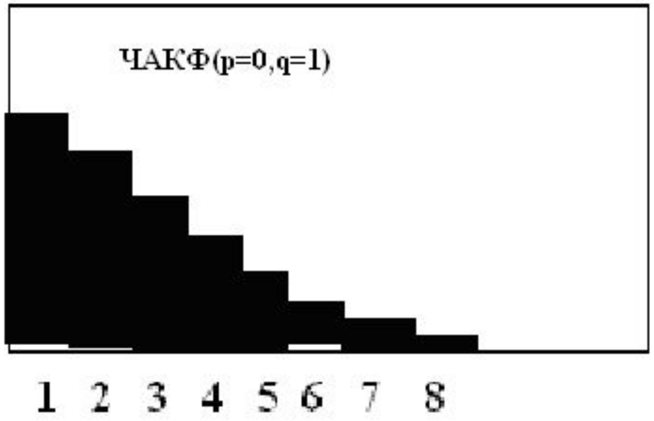
АРСС(1, 0)



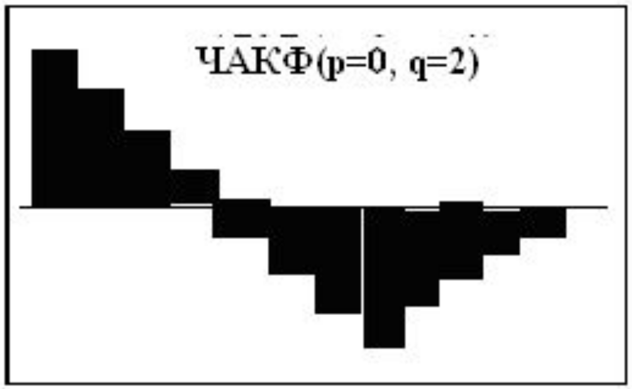


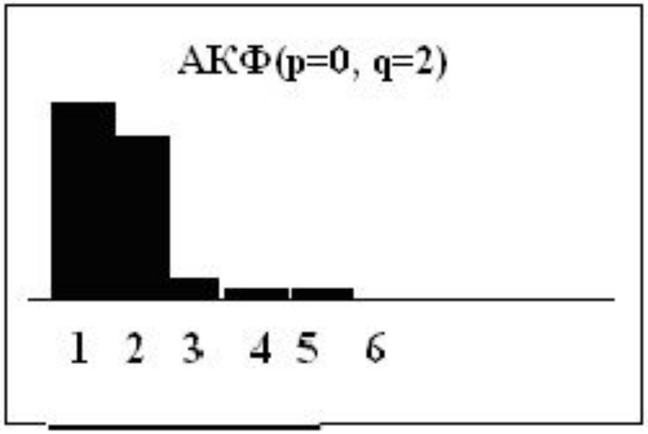
Идентификация модели АРСС (I подход)





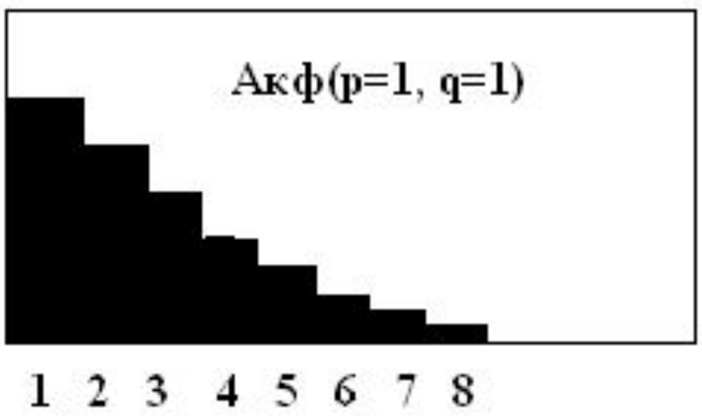
АРСС(0, 1)

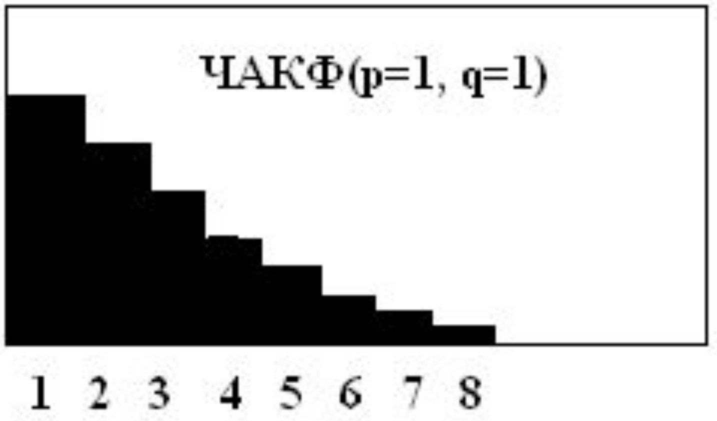




АРСС(0, 2)

Идентификация модели АРСС (I подход)





АРСС(1, 1)

Идентификация модели АРСС (II подход)

Для проверки автокорреляции в рядах, где присутствуют элементы авторегрессии и скользящей средней, используется критерий Люнга-Бокса.

Значение $LB_{\text{расч}}$ определяется по формуле:

$$LB_{расч} = n(n+2)\sum_{k=1}^{m}\left(\frac{1}{n-k}\right)r_k^2 \quad (5)$$

где m – максимальное число временных лагов, рассматриваемых в модели, p – порядок авторегрессии, q – порядок процесса скользящей средней, n – число наблюдений во временном ряду, r^2_k – коэффициент автокорреляции.

Из таблицы χ^2 критических значений находится $\chi^2_{табл}(m-p-q)$ для m-p-q степеней свободы. Затем сравнивают $LB_{\text{расч}}$ и $\chi^2_{табл}(m-p-q)$, если $LB_{\text{расч}} > \chi^2_{табл}(m-p-q)$, то порядок авторегрессионного процесса равен p, а порядок процесса скользящей средней - q.

Идентификация модели АРСС (III подход)

Также для идентификации модели АРСС подходит метод спектрального анализа.

Авторегрессионный процесс первого порядка $Y_t = \alpha \cdot Y_{t-1} + \varepsilon_t$ (при α>0)

Процесс скользящей средней первого порядка $Y_t = \varepsilon_t - b \cdot \varepsilon_{t-1}$ (при b<0)

Селекция моделей АРСС

Иногда один и тот же стохастический процесс, представленный временным рядом, может быть описан с помощью нескольких моделей АРСС различных порядков.

Возникает задача селекции (отбора) модели.

Для выбора наиболее подходящей модели используют информационные критерии Акайке (*AC*), Шварца (*SC*) и Ханна-Квина (*HQ*), определяемые по соответствующим формулам:

;

здесь n – общее число наблюдений временного ряда, k – число степеней свободы модели (для АРСС $k=p+q$), – остаточная или объясненная моделью дисперсия.

Выбирают АРСС-модель, для которой значения *AC*, *SC* и *HQ* критериев имеют наименьшее значение.

Проверка адекватности модели АРСС

Для проверки адекватности модели АРСС, заключающейся в тестировании оценок на достоверность (эффективность, состоятельность и несмещенность) необходимо проанализировать остатки $\acute{\varepsilon}$. Если модель АРСС (p; q)достаточно хорошо описана, то остатки являются независимыми нормально распределенными случайными величинами с нулевым математическим ожиданием, с постоянной дисперсией и нулевой атокорреляцией.

Поэтому для проверки адекватности модели (достоверности оцененных параметров) достаточно проверить остаточную компоненту на соответствии гауссову «белому шуму».

Для этого проводят спектральный анализ или анализ АКФ, ЧАКФ + тест на нормальность (Пирсона, Колмогорова-Смирнова, Бера-Жарка).

Тест Бера-Жарка

Соответствие распределения остатков ARMA-модели нормальному закону можно проверить с помощью теста Бера-Жарка, для которого определяется *JB*-статистика по формуле:

где иент асимметрии распределения

остатков,

– коэффициент эксцесса, n – объем выборки, – среднее значение остатков, p и q – порядки ARMA-модели. Нулевая гипотеза о «ненормальности» распределения остатков отклоняется на выбранном уровне значимости, если $JB>\chi^2_{табл}$, определённого для степеней свободы $n–p–q$ из таблицы критических значений χ^2-распределения.