Лекция 5
Линейные гауссовские модели с дискретным временем. Модели авторегрессии-скользящего среднего

Артемов А. В.
бПМИ ФКН ВШЭ

17 февраля 2018

1 Введение

На абсолютно *эффективных рынках* наилучшим прогнозом будущего уровня цены финансового актива является текущая цена этого актива. Поэтому понятие мартингала стало одним из основных при исследовании динамики эволюции цен как стохастических последовательностей или процессов с определёнными свойствами их распределений. Однако при проведении конкретных расчетов одного лишь знания “мартингальности распределений” слишком мало — нужна более “тонкая” структура этих распределений, что приводит к необходимости детального рассмотрения самых разнообразных вероятностно-статистических моделей с целью выявления тех из них, свойства распределений которых лучше всего согласуются со свойствами эмпирических распределений, построенных по статистическим данным.

Предположение гауссовости распределений величин $h_1, \ldots, h_n$, конечно, выглядит привлекательным и с точки зрения теоретического анализа, и с точки зрения “статистики нормального распределения”. Но это предположение не всегда соответсвтует истинной картине поведения цен. Но какую альтернативу можно привести? Для этого вспомним разложение Дуба. Как известно, оно определяется с привлечением условных матожиданий вида $\mathsf{E}[h_n \mid \mathcal{F}_{n-1}]$. Тогда было бы разумно предположить, что не безусловные, а *условные* распределения являются гауссовскими:

$$\mathrm{Law}(h_n \mid \mathcal{F}_{n-1}) = \mathcal{N}(\mu_n, \sigma_n^2)$$

с некоторыми $\mathcal{F}_{n-1}$-измеримыми величинами $\mu_n = \mu_n(\omega)$ и $\sigma_n^2 = \sigma_n^2(\omega)$.

Оказывается, что $\mathsf{E}[h_n \mid \mathcal{F}_{n-1}] = \mu_n$ и $\mathsf{D}[h_n \mid \mathcal{F}_{n-1}] = \sigma_n^2$ (это следует из регулярности условного распределения — за доказательством обращайтесь к первому тому Ширяева). Тем самым видем смысл этих параметров — условное среднее и условная дисперсия распределения $\mathrm{Law}(h_n \mid \mathcal{F}_{n-1})$.

Само же распределение $\mathrm{Law}(h_n)$ является *взвесью* условных гауссовских распределений $\mathrm{Law}(h_n \mid \mathcal{F}_{n-1})$ с усреднением по распределению величин μ_n и σ_n^2.

Обычно наряду с $h = (h_n)$ вводится “стандартная” условно-гауссовская последовательность $\varepsilon = (\varepsilon_n)_{n \geqslant 1}$ $\mathcal{F}_n$-измеримых случайных величин таких, что

$$\mathrm{Law}(\varepsilon_n \mid \mathcal{F}_{n-1}) = \mathcal{N}(0, 1), \text{ где } \mathcal{F}_0 = \{\varnothing, \Omega\}.$$

Оказывается, что это будет последовательность *независимых* случайных величин с стандартным нормальным распределением, так как

$$\mathrm{Law}(\varepsilon_n \mid \varepsilon_1, \ldots, \varepsilon_{n-1}) = \mathcal{N}(0, 1).$$

Если $\omega_n^2 \neq 0$ поточечно для всех $n \geqslant 1$, то величины ε_n, задаваемые по правилу $\varepsilon_n \equiv (h_n - \mu_n)/\omega_n$, будут задавать стандартную гауссовскую последовательность. Тогда можно считать, что рассматриваемые условно-гауссовские последовательности представимы в виде

$$h_n = \mu_n + \sigma_n \varepsilon_n,$$

где $\varepsilon = (\varepsilon_n)$ — последовательность независимых $\mathcal{F}_n$-измеримых случайных величин с распределением $\mathcal{N}(0, 1)$.

Понятно, что более подробное изучение свойств последовательности $h = (h_n)$ зависит от структуры μ_n и σ_n^2. Именно это и делается в представляемых ниже моделях.

В теории временных рядов есть целый арсенал разнообразных *линейных* моделей, среди которых в первую очередь нужно назвать следующие:

- Модель *скользящего среднего* порядка q MA(q),
- Модель *авторегрессии* порядка p AR(p),
- Модель *авторегрессии и скользящего среднего* порядка (p, q) ARMA(p, q).

Эти модели широко исследуются в теории временных рядов, особенно в предположении *стационарности*. Вообще, для чего вводятся линейные модели? Они весьма просты, но при этом ими можно неплохо приближать весьма широкий класс стационарных последовательностей.

Вот только не все временные "эконометрические" ряды являются стационарными. Анализ показывает, что часто в данных вырисовываются три составляющие:

- Медленно меняющийся (например, "инфляционный") тренд x,
- Периодические или же апериодические циклы y,
- Нерегулярная, флуктуирующая ("стохастическая" или "хаотическая") компонента z.

В наблюдаемые данные h они могут входить весьма разнообразными способами. Образно это можно представить так: $h = x * y * z$, где вместо $*$ могут выступать сложение $+$, умножение $\times$ и так далее.

Ниже мы рассмотрим некоторые *линейные* (а затем и нелинейные) модели, преследуя цель дать представление об их структуре, особенностях, свойствах, применяемых в анализе данных.

Не стоит забывать, что конечной целью анализа статистических данных является *прогнозирование* дальнейшнего поведения. Качество этого прогнозирования зависит от удачного выбора модели, точности оценивания определяющих её параметров и качества экстраполяционного оценивания.

Во всех рассматриваемых далее моделях будем считать, что задана некоторая "базисная" последовательность $\varepsilon = (\varepsilon_n)$, которую в теории временных рядов обычно считают *белым шумом* и идентифицируют с источником случайности, определяющим стохастический характер исследуемых вероятностно-статистических объектов. При этом (в "L^2-теории") говорят, что последовательность является *белым шумом в широком смысле*.

Определение 1. Последовательность $\varepsilon = (\varepsilon_n)$ называется белым шумом в широком смысле, если
$$\forall m, n \in \mathbb{Z}, m \neq n : \mathsf{E}[\varepsilon_n] = 0, \mathsf{E}\left[\varepsilon_n^2\right] < \infty, \mathsf{E}[\varepsilon_m \varepsilon_n] = 0.$$

Другими словами, белый шум в широком смысле — это квадратично интегрируемая последовательность некоррелированных случайных величин с нулевыми средними.

Ещё вводят белый шум *в узком смысле*, который обычно называют просто *белым (гауссовским) шумом.*

Определение 2. Белый шум — это гауссовская последовательность, являющаяся белым шумом в широком смысле.

Другими словами, это последовательность независимых случайных величин с нормальными распределениями $\mathcal{N}(0, \sigma_n^2)$. Далее будем считать, что $\sigma_n^2 \equiv 1$. В таком случае обычно говорят, что ε есть стандартная гауссовская последовательность.

2 Модель скользящего среднего $\mathrm{MA}(q)$

В модели скользящего среднего порядка q, описывающей эволюцию последовательности $h = (h_n)$, предполагается следующие способ формирования значений h_n по белому шуму в широком смысле $\varepsilon = (\varepsilon_n)$:

$$h_n = (\mu + b_1 \varepsilon_{n-1} + \ldots + b_q \varepsilon_{n-q}) + b_0 \varepsilon_n,$$

где параметр q определяет порядок зависимости от "прошлого", а ε_n играет роль величин, "обновляющих" информацию, содержащуюся в $\mathcal{F}_{n-1} = \sigma(\varepsilon_{n-1}, \varepsilon_{n-2}, \ldots)$.

Далее, для компактности вводят *лаговый оператор* L, действующий по правилу $Lx_n = x_{n-1}$. Так как $L(Lx_n) = x_{n-2}$, то разумно ввести обозначение

$$L^2 x_n \equiv L(Lx_n) = x_{n-2},$$

и, в общем случае, $L^k x_n = x_{n-k}$.

Отметим следующие свойства лагового оператора:

$$\begin{aligned}
L(cx_n) &= cLx_n, \\
L(x_n + y_n) &= Lx_n + Ly_n, \\
(c_1 L + c_2 L^2) x_n &= c_1 L x_n + c_2 L^2 x_n = c_1 x_{n-1} + c_2 x_{n-2}, \\
(1 - \lambda_1 L)(1 - \lambda_2 L) x_n &= x_n - (\lambda_1 + \lambda_2) x_{n-1} + (\lambda_1 \lambda_2) x_{n-2}.
\end{aligned}$$

Пользуясь этими свойствами, модели $\mathrm{MA}(q)$ можно придать следующую форму: $h_n = \mu + \beta(L)\varepsilon_n$, где $\beta(L) = b_0 + b_1 L + \ldots + b_q L^q$.

Теперь положим $q = 1$. В таком случае

$$h_n = \mu + b_0 \varepsilon_n + b_1 \varepsilon_{n-1}.$$

Несложно проверить, что

$$\begin{gathered}
\mathsf{E}[h_n] = \mu, \mathsf{D}[h_n] = b_0^2 + b_1^2 \\
\mathsf{cov}(h_{n+1}, h_n) = b_0 b_1, \mathsf{cov}(h_{n+k}, h_n) = 0, k > 1.
\end{gathered}$$

Это означает, что $h = (h_n)$ — это последовательность с коррелированными соседними значениями (h_n и h_{n+1}), причём корреляция значений h_{n+k} и h_n при $k > 1$ равна нулю.

Из соотношений сверху следует, что у элементов последовательности $h = (h_n)$ матожидание, дисперсия и ковариация не зависят от n (впрочем, это определяется предположением стандартности последовательности ε и тем, что b_k не зависят от n). Отсюда следует, что последовательность $h = (h_n)$ является стационарной в широком смысле (просто по определению). Если же добавить то, что ε является гауссовской, то и h тоже будет гауссовской. Это означает, что все её параметры полностью задаются средним, дисперсией и ковариацией. Но тогда h будет стационарной и в узком смысле, так как для произвольных n, k и $i_1, \ldots, i_n$

$$\mathrm{Law}(h_{i_1}, \ldots, h_{i_n}) = \mathrm{Law}(h_{i_1+k}, \ldots, h_{i_n+k})$$

Теперь покажем одно интересное свойство модели MA(1). Пусть $(h_1, \ldots, h_n)$ — некоторая реализация, полученная в результате наблюдений величин h_k в моменты времени $k = 1, \ldots, n$. Далее, пусть $\overline{h}_n = (\sum_{k=1}^n h_k)/n$ — это временное среднее. Со статистической точки зрения обращение к "статистике" $\overline{h}_n$ представляет тот интерес, что $\overline{h}_n$ является естественным кандидатом для оценивания среднего μ.

Оказывается, что для стационарной в широком смысле последовательности h_n есть хороший критерий эргодичности, который похож на условие Слуцкого (на самом деле это оно и есть):

Теорема 1. *Пусть $h = (h_n)$ — стационарная в широком смысле последовательность, а $R(k) = \mathsf{cov}(h_{n+k}, h_n)$. Тогда*

$$\lim_{n\to\infty} \mathsf{E}\big[(\overline{h}_n - \mu)^2\big] = 0 \iff \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} R(k) = 0.$$

Доказательство. Без ограничения общности скажем, что $\mu = \mathsf{E}[h_n] = 0$.

Пусть $\mathsf{E}\big[(\overline{h}_n - \mu)^2\big] \to 0$. Тогда по неравенству Коши-Буняковского:

$$\left|\frac{1}{n}\sum_{k=1}^{n} R(k)\right|^2 = \left|\mathsf{E}\left[\frac{h_0}{n}\sum_{k=1}^{n} h_k\right]\right|^2 \leqslant \mathsf{E}\big[h_0^2\big]\,\mathsf{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n} h_k\right|^2\right] \xrightarrow[n\to\infty]{} 0.$$

Теперь докажем в другую сторону. Заметим, что

$$\mathsf{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n} h_k\right|^2\right] = \frac{1}{n^2}\,\mathsf{E}\left[\sum_{k=1}^{n} h_k^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} h_i h_j\right] = \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=0}^{i-1} R(j) - \frac{1}{n}R(0)$$

Зафиксируем произвольное $\delta > 0$ и найдём $n(\delta)$ такое, что для любого $l \geqslant n(\delta)$ выполнено, что

$$\left|\frac{1}{l}\sum_{k=0}^{l} h_k\right| \leqslant \delta.$$

Тогда для $n \geqslant n(\delta)$ имеем

$$\begin{aligned}\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=0}^{i-1} R(j)\right| &= \left|\frac{1}{n^2}\sum_{i=1}^{n(\delta)}\sum_{j=0}^{i-1} R(j) + \frac{1}{n^2}\sum_{i=n(\delta)+1}^{n}\sum_{j=0}^{i-1} R(j)\right| \leqslant \\ &\leqslant \left|\frac{1}{n^2}\sum_{i=1}^{n(\delta)}\sum_{j=0}^{i-1} R(j)\right| + \left|\frac{1}{n^2}\sum_{i=n(\delta)+1}^{n} i\cdot\frac{1}{i}\sum_{j=0}^{i-1} R(j)\right| \leqslant \\ &\leqslant \frac{1}{n^2}\left|\sum_{i=1}^{n(\delta)}\sum_{j=0}^{i-1} R(j)\right| + \delta\end{aligned}$$

Теперь вспомним, что $R(0) = const < \infty$. Тогда

$$\varlimsup_{n\to\infty} \mathsf{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n} h_k\right|^2\right] \leqslant \delta.$$

Устремляя δ к нулю, получаем желаемое. □

Тем самым мы получили весьма полезное свойство: MA(1) эргодична в среднеквадратичном, то есть среднее по времени стремится в среднеквадратичном смысле к среднему по ансамблю μ.

Теперь вспомним про корреляционную функцию. Для MA(1) она будет иметь вид

$$r(k) = \frac{\mathsf{cov}(h_{n+k}, h_n)}{\sqrt{\mathsf{D}[h_{n+k}]\,\mathsf{D}[h_n]}} = \frac{R(k)}{R(0)} = \begin{cases} 1, & k = 0 \\ \frac{b_0 b_1}{b_0^2 + b_1^2}, & k = 1 \\ 0, & k > 1 \end{cases}$$

Вернёмся к общему случаю MA(q). В качестве упражнения оставим вывод следующих тождеств:

$$\mathsf{E}[h_n] = \mu, \quad \mathsf{D}[h_n] = \sum_{k=0}^{q} b_k^2,$$

$$R(k) = \begin{cases} \sum_{i=0}^{q-k} b_i b_{k+i}, & k \leqslant q \\ 0, & k > q \end{cases}$$

Пользуясь этими формулами, моделью MA(q) можно пытаться моделировать поведение последовательностей $h = (h_n)$, у которых корреляция величин h_n и h_{n+k}, где $k > q$, нулевая. Но как это делать? Общий принцип подгонки следующий:

- Для начала, по выборке $(h_1, \ldots, h_n)$ строятся некоторые эмпирические характеристики: например,

$$\overline{h}_n = \frac{1}{n}\sum_{k=1}^{n} h_k \text{ — выборочное среднее} \tag{2.1}$$

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(h_k - \overline{h}_n) \text{ — выборочная дисперсия} \tag{2.2}$$

$$r_n(k) = \frac{1}{n\widehat{\sigma}_n^2}\sum_{i=k+1}^{n}(h_i - \overline{h}_n)(h_{i-k} - \overline{h}_n) \text{ — выборочное среднее} \tag{2.3}$$

- Далее, используя выражения для теоретических характеристик, производится варьирование параметров с целью подгонки теоретических значений под эмпирические.

- В конце провдодится *оценка кажества* подгонки, основываясь на знании эмпирических характеристик и их отклонений от теоретических распределений.

Ранее мы смотрели на модели с конечным q. Было бы разумно ввести обобщение, которое позволяет и бесконечное q: MA(∞). Устроено оно так:

$$h_n = \mu + \sum_{j=0}^{\infty} b_j \varepsilon_{n-j}.$$

Но, конечно, должны быть какие-то условия (как минимум, на сходимость). Если потребовать сходимость ряда $\sum b_j^2$, то ряд в формуле для h_n будет сходиться в среднеквадратичном смысле.

Для этой модели

$$\mathsf{E}[h_n] = \mu, \quad \mathsf{D}[h_n] = \sum_{k=0}^{\infty} b_k^2, \quad R(k) = \sum_{i=0}^{\infty} b_{k+i} b_i.$$

В теории стационарных случайных процессов принято говорить, что h_n есть "результат реакции физически осуществимого фильтра с импульсной переходной функцией $b = (b_n)$, когда на вход подается последовательность $\varepsilon = (\varepsilon_n)$".

Оказывается, что в определённом смысле "регулярная" стационарная (в широком смысле) последовательность может быть представлена моделью MA(∞). Результат, связанный с этим фактом, называется *разложением Вольда* стационарных последовательностей на «сингулярную» и «регулярную» части. За подробностями обращайтесь ко второму тому Ширяева.

3 Авторегрессионная модель AR(p)

Поехали дальше. Авторегрессионая модель порядка p определяется следующим образом:

$$h_n = \mu_n + \sigma\varepsilon_n, \quad \mu_n = a_0 + a_1 h_{n-1} + \ldots + a_p h_{n-p}.$$

Можно сказать, что модель AR(p) подчиняется *разностному уравнению* порядка p:

$$h_n = a_0 + a_1 h_{n-1} + \ldots + a_p h_{n-p} + \sigma\varepsilon_n.$$

Если совпользоваться лаговым оператором, то это уравнение можно записать в виде:

$$(1 - a_1 L - \ldots - a_p L^p) h_n = a_0 + \sigma\varepsilon_n.$$

Тогда, введя пару обознаений, уравнение приобретает вид

$$\alpha(L) h_n = \omega_n, \quad \omega_n = a_0 + \sigma\varepsilon_n, \quad \alpha(L) = 1 - a_1 L - \ldots - a_p L^p$$

В отличие от модели со скользящий средним, в этой модели нужно задавать *начальные* условия $(h_{1-p}, h_{2-p}, \ldots, h_0)$. Обычно их обнуляют, хотя можно считать их случайными и не зависящими от ε. В эргодических случаях асимптотическое поведение последовательности при $n \to \infty$ не зависит от начальных условий, и в этом смысле их конкретизация не столь существенна.

Рассмотрим модель AR(1) поподробнее. Выглядит она следующим образом:

$$h_n = a_0 + a_1 h_{n-1} + \sigma\varepsilon_n.$$

Этоа модель выделяется из общего класса AR(p) моделей тем, что из «прошлых» величин $h_{n-1}, \ldots, h_{n-p}$, h_n зависит только от ближайшего (по времени) значения h_{n-1}. Если добавить к этой модели независимость последовательности $\varepsilon = (\varepsilon_n)$ и независимость h_0 от неё, то получится *конструктивный* пример марковской цепи.

Несложно получить, что

$$h_n = a_0(1 + a_1 + a_1^2 + \ldots + a_1^{n-1}) + a_1^n h_0 + \sigma(\varepsilon_n + a_1\varepsilon_{n-1} + \ldots + a_1^{n-1}\varepsilon_1).$$

Отсюда видно, что свойства последовательности сильно зависят от a_1. Учитывая формулу, есть смысл различать три случая: $|a_1| < 1$, $|a_1| = 1$ и $|a_1| > 1$, причём второй случай является «пограничным». Смысл этого станет понятен позднее.

Из разложения понятно, что

$$\begin{aligned}
\mathsf{E}[h_n] &= a_0(1 + a_1 + \ldots + a_1^{n-1}) + a_1^n \,\mathsf{E}[h_0] = \frac{a_0(1 - a_1^n)}{1 - a_1} + a_1^n \,\mathsf{E}[h_0] \\
\mathsf{D}[h_n] &= a_1^{2n} \,\mathsf{D}[h_0] + \sigma^2(1 + a_1^2 + \ldots + a_1^{2(n-1)}) = a_1^{2n} \,\mathsf{D}[h_0] + \frac{\sigma^2(1 - a_1^{2n})}{1 - a_1^2} \\
\mathsf{cov}(h_n, h_{n-k}) &= a_1^{2n-k} \,\mathsf{D}[h_0] + \sigma^2 a_1^k (1 + a_2 + \ldots + a_1^{2(n-k-1)}) \\
&= a_1^{2n-k} \,\mathsf{D}[h_0] + \frac{\sigma^2 a_1^k (1 - a_1^{2(n-k)})}{1 - a_1^2}
\end{aligned}$$

Если $|a_1| < 1$ и $\mathsf{E}[h_0] < \infty]$, $\mathsf{D}[h_0] < \infty$, то при $n \to \infty$ последовательность «стационаризируется»:

$$\mathsf{E}[h_n] \to \frac{a_0}{1 - a_1}, \quad \mathsf{D}[h_n] \to \frac{\sigma^2}{1 - a_1^2}, \quad \mathsf{cov}(h_n, h_{n-k}) \to \frac{\sigma^2 a_1^k}{1 - a_1^2}$$

Теперь заметим, что если начальное распределение для h_0 является нормальным:

$$h_0 \sim \mathcal{N}\left(\frac{a_0}{1 - a_1}, \frac{\sigma^2}{1 - a_1^2}\right),$$

то последовательность $h = (h_n)$ является стационарной в узком смысле гауссовской последовательностью. Заметим, что для такой последовательности корреляция равна

$$r(k) = \frac{\mathsf{cov}(h_{n-k}, h_n)}{\sqrt{\mathsf{D}[h_{n-k}]\,\mathsf{D}[h_n]}} = a_1^k.$$

Зафиксируем n. В таком случае

$$h_n = a_0(1 + a_1 + a_1^2 + \ldots + a_1^{n-1}) + a_1^n h_0 + \sigma(\varepsilon_n + a_1\varepsilon_{n-1} + \ldots + a_1^{n-1}\varepsilon_1).$$

Если h_0 есть константа, то, введя обозначение $\mu = a_0(1 + a_1 + a_1^2 + \ldots + a_1^{n-1}) + a_1^n h_0$, получится модель MA($n-1$). В этом смысле иногда несколько вольно говорят, что «модель AR(1) может рассматриваться как модель MA(∞)».

Теперь скажем, что $|a_1| = 1$. Тогда

$$h_n = a_0 n + h_0 + \sigma(\varepsilon_1 + \ldots + \varepsilon_n).$$

Если ввести обозначение $\omega_n = a_0 + \sigma\varepsilon_n$, то модель будет иметь вид

$$h_n = h_0 + \omega_1 + \omega_2 + \ldots + \omega_n.$$

Это есть ни что иное, как случайное блуждание. Заметим, что

$$\mathsf{E}[h_n] = a_0 n + \mathsf{E}[h_0], \quad \mathsf{D}[h_n] = \sigma^2 n \xrightarrow[n \to \infty]{} \infty.$$

Случай с $|a_1| > 1$ называют *взрывающимся*, так как и среднее значение, и дисперсия растут с ростом n, причём экспоненциально быстро.

Теперь посмотрим на модель AR(2):

$$h_n = a_0 + a_1 h_{n-1} + a_2 h_{n-2} + \sigma\varepsilon_n \iff (1 - a_1 L - a_2 L^2)h_n = a_0 + \sigma\varepsilon_n.$$

Если $a_2 = 0$, то мы возвращаемся к модели AR(1). Введя обозначение $\omega_n = a_0 + \sigma\varepsilon_n$, получаем, что

$$(1 - a_1 L)h_n = \omega_n.$$

Вопрос: можно ли как-то «обратить» это равенство и сразу считать h_n только по $\omega = (\omega_n)$, не обращаясь к предыдущим значениям? Воспользуемся свойствами оператора L и заметим, что

$$(1 + a_1 L + a_1^2 L^2 + \ldots + a_1^k L^k)(1 - a_1 L) = 1 - a_1^{k+1} L^{k+1}.$$

Тогда

$$h_n = (1 + a_1 L + a_1^2 L^2 + \ldots + a_1^k L^k)\omega_n + a_1^{k+1} L^{k+1} h_n.$$

Если положить $k = n - 1$, то получим разложение, которое получали ранее:

$$\begin{aligned} h_n &= (1 + a_1 L + a_1^2 L^2 + \ldots + a_1^{n-1} L^{n-1})\omega_n + a_1^n h_0 = \\ &= (a_0 + \sigma\varepsilon_n) + a_1(a_0 + \sigma\varepsilon_{n-1}) + \ldots + a_1^{n-1}(a_0 + \sigma\varepsilon_1) + a_1^n h_0 = \\ &= a_0(1 + a_1 + \ldots + a_1^{n-1}) + a_1^n h_0 + \sigma(\varepsilon_n + a_1\varepsilon_{n-1} + \ldots + a_1^{n-1}\varepsilon_1). \end{aligned}$$

Если $|a_1| < 1$ и n достаточно велико, то неформально можно сказать, что

$$h_n \approx (1 + a_1 L + a_1^2 L^2 + \ldots + a_1^{n-1} L^{n-1})\omega_n = (1 + a_1 L + a_1^2 L^2 + \ldots + a_1^{n-1} L^{n-1})(1 - a_1 L)h_n.$$

Тем самым мы получаем, что «обратный» оператор $(1 - a_1 L)^{-1}$ разумно определить следующим образом:

$$(1 - a_1 L)^{-1} \equiv \sum_{k=0}^{\infty} a_1^k L^k = 1 + a_1 L + a_1^2 L^2 + \ldots + a_1^n L^n + \ldots$$

Оказывается, что если получаемый ряд для h_n сходится в среднеквадратичном смысле, то это разложение единственно.

Пользуясь этим рассуждением, можно найти похожее представление для AR(2). Так как

$$(1 - \lambda_1 L)(1 - \lambda_2 L) = 1 - (\lambda_1 + \lambda_2)L + \lambda_1\lambda_2 L^2,$$

то, определяя λ_1 и λ_2 из системы

$$\begin{cases} \lambda_1 + \lambda_2 = a_1 \\ \lambda_1\lambda_2 = -a_2 \end{cases}$$

получим, что

$$1 - a_1 L - a_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L).$$

Тогда

$$(1 - \lambda_1 L)(1 - \lambda_2)h_n = \omega_n \implies h_n = (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1}\omega_n$$

Воспользуемся методом неопределённых коэффициентов:

$$\frac{1}{(1 - \lambda_1 L)(1 - \lambda_2 L)} = \frac{A}{1 - \lambda_1 L} + \frac{B}{1 - \lambda_2 L} = \frac{A - A\lambda_2 L + B - B\lambda_1 L}{(1 - \lambda_1 L)(1 - \lambda_2 L)}$$

Тогда
$$\begin{cases} A+B=1 \\ A\lambda_2+B\lambda_1=0 \end{cases} \implies A=\frac{\lambda_1}{\lambda_1-\lambda_2}, B=-\frac{\lambda_2}{\lambda_1-\lambda_2}$$

Отсюда получаем, что
$$h_n=\frac{\lambda_1}{\lambda_1-\lambda_2}(1-\lambda_1 L)^{-1}\omega_n-\frac{\lambda_2}{\lambda_1-\lambda_2}(1-\lambda_2 L)^{-1}\omega_n.$$

Предположим, что все λ_i такие, что $|\lambda_i|<1$. Тогда
$$h_n=\sum_{k=0}^{\infty}(c_1\lambda_1^k+c_2\lambda_2^k)\omega_k,\ \text{где } c_1=\frac{\lambda_1}{\lambda_1-\lambda_2},\quad c_2=-\frac{\lambda_2}{\lambda_1-\lambda_2}.$$

Это рассуждение обобщается и на модель AR(p). Для неё
$$(1-a_1L-a_2L^2-\ldots-a_pL^p)h_n=\omega_n.$$

Опять же, разложим его на множители:
$$1-a_1L-a_2L^2-\ldots-a_pL^p=(1-\lambda_1L)\ldots(1-\lambda_pL)$$

Если все $|\lambda_i|<1$, то получится стационарное решение, которое будет единственным среди решений с конечным вторым моментом:
$$h_n=(1-\lambda_1L)^{-1}\ldots(1-\lambda_pL)^{-1}\omega_n.$$

Теперь сведём это в виду ряда. Для этого снова воспользуемся методом неопределённых коэффициентов:
$$\frac{1}{(1-\lambda_1z)\ldots(1-\lambda_pz)}=\frac{c_1}{1-\lambda_1z}+\ldots+\frac{c_p}{1-\lambda_pz}.$$

Умножая на $(1-\lambda_1z)\ldots(1-\lambda_pz)$, получаем уравнение, которое должно выполняться для всех z:
$$1=\sum_{k=1}^{p}c_k\prod_{i\neq k}(1-\lambda_iz).$$

Пподставляя значения $z=0,\lambda_k^{-1}$, $k\in\{1,2,\ldots,p\}$, получаем, что $c_1+\ldots+c_p=1$ и
$$c_k=\frac{\lambda_k^{p-1}}{\prod_{i\neq k}(\lambda_k-\lambda_i)}.$$

Следовательно, решение имеет вид
$$h_n=\sum_{k=0}^{\infty}(c_1\lambda_1^k+\ldots+c_p\lambda_p^k)\omega_{n-k}.$$

Это разложение помогает считать различные характеристики последовательности $h=(h_n)$: например, моменты $\mathsf{E}\left[h_n^k\right]$, ковариации, условные математические ожидания и так далее.

Что мы можем сказать про некоторые характеристики последовательности $h=(h_n)$, если она стационарна (в широком смысле)? Для этого воспользуемся определением модели. Тогда, если $\mu\equiv\mathsf{E}[h_k]$, то
$$\mu=a_0+a_1\mu+\ldots+a_p\mu \implies \mu=\frac{a_0}{1-(a_1+\ldots+a_p)}.$$

Ковариация $R(k) = \mathsf{cov}(h_{n+k}, h_n)$ при $k > 0$ же удовлетворяет следующему уравнению:

$$R(k) = \mathsf{cov}(h_{n+k}, h_n) = \mathsf{cov}(a_0 + a_1 h_{n+k-1} + \ldots + a_p h_{n+k-p} + \sigma\varepsilon_{n+k}, h_n) = \\ = a_1 R(k-1) + a_2 R(k-2) + \ldots + a_p R(k-p).$$

Если $k = 0$, то уравнение имеет вид

$$R(0) = a_1 R(1) + \ldots + a_p R(p) + \sigma^2.$$

Оказывается, что аналогичные самые уравнения верны и для корреляций $r(k)$. Их принято называть *уравнениями Юла-Уолкера.*

4 Модель авторегрессии и скользящего среднего $\mathrm{ARMA}(p, q)$ и интегральная модель $\mathrm{ARIMA}(p, d, q)$

Модель $\mathrm{ARMA}(p, q)$ совмещает в себе возможности и модели скользящего среднего, и авторегрессионной модели. Это отображено и в названии: $\mathrm{ARMA} = \mathrm{AR} + \mathrm{MA}$. Перейдём к определению.

Определение 3. Будем называть последовательность $h = (h_n)$ $\mathrm{ARMA}(p, q)$-*моделью*, если

$$h_n = \mu_n + \sigma\varepsilon_n, \text{ где } \mu_n = (a_0 + a_1 h_{n-1} + \ldots + a_p h_{n-p}) + (b_1\varepsilon_{n-1} + \ldots + b_q\varepsilon_{n-q}).$$

Без ограничения общности можно полагать, что $\sigma = 1$. Тогда

$$h_n - a_1 h_{n-1} - \ldots - a_p h_{n-p} = a_0 + \varepsilon_n + b_1\varepsilon_{n-1} + \ldots + b_q\varepsilon_{n-q}$$

Введём два оператора:

$$\alpha(L) = 1 - a_1 L - \ldots - a_p L^p, \quad \beta(L) = 1 + b_1 L + \ldots + b_q L^q.$$

Тогда модель можно записать следующим образом:

$$\alpha(L) h_n = a_0 + \beta(L)\varepsilon_n \iff h_n = \frac{a_0}{1 - (a_1 + \ldots + a_p)} + \frac{\beta(L)}{\alpha(L)}\varepsilon_n.$$

Опять же, поднимем вопрос о существовании стационарного решения этого уравнения. Из предыдущих рассуждений и вида уравнения понятно, что стационарность задаётся авторегрессионной компонентой модели $\mathrm{ARMA}(p, q)$. Следовательно, если все корни уравнения $1 - a_1 z - \ldots - a_p z^p = 0$ меньше единицы по модулю, то стационарное решение существует и единственно (в классе L^2). Для него

$$\mathsf{E}[h_n] = \frac{a_0}{1 - (a_1 + \ldots + a_p)}.$$

Со ковариацией же ситуация немного другая. Если $k > q$, то выполнено уравнение Юла-Уолкера:

$$R(k) = \mathsf{cov}(a_0 + a_1 h_{n+k-1} + \ldots + a_p h_{n+k-p} + \varepsilon_{n+k} + b_1\varepsilon_{n+k-1} + \ldots + b_q\varepsilon_{n+k-q}, h_n) = \\ = a_1 R(k-1) + \ldots + a_p R(k-p).$$

Если же $k \leqslant q$, то уже нужно учитывать корреляционную зависимость между h_n и ε_{n-l}, $l \geqslant 0$.

Рассмотрим модель ARMA$(1,1)$:

$$h_n - a_1 h_{n-1} = a_0 + \varepsilon_n + b_1 \varepsilon_{n-1}.$$

Предположим, что $|a_1| < 1$. Тогда

$$h_n = \frac{a_0}{1-a_1} + \frac{1+b_1L}{1-a_1L}\varepsilon_n = \frac{a_0}{1-a_1} + \left(\sum_{k=0}^{\infty} a_1^k L^k\right)(1+b_1L)\varepsilon_n =$$

$$= \frac{a_0}{1-a_1} + \sum_{k=0}^{\infty} a_1^k L^k \varepsilon_n + b_1 \sum_{k=0}^{\infty} a_1^k L^{k+1}\varepsilon_n = \frac{a_0}{1-a_1} + \varepsilon_n + (a_1+b_1)\sum_{k=1}^{\infty} a_1^{k-1}\varepsilon_{n-k}$$

Что мы можем сказать про ковариацию? На самом деле много. Пользуясь этим разложением, получаем, что при $k > 1$ $R(k) = a_1 R(k-1)$ и

$$R(1) = \mathsf{cov}(h_{n+1}, h_n) = \mathsf{cov}(a_0 + a_1 h_n + \varepsilon_{n+1} + b_1\varepsilon_n, h_n) = a_1 R(0) + b_1$$

Осталось решить:

$$R(0) = \mathsf{D}[h_n] = (a_1+b_1)^2 \sum_{k=0}^{\infty} a_1^{2k} + 1 = \frac{a_1^2 + 2a_1b_1 + b_1^2}{1-a_1^2} + 1 = \frac{1+2a_1b_1+b_1^2}{1-a_1^2},$$

$$R(1) = \frac{a_1 + 2a_1^2 b_1 + a_1 b_1^2}{1-a_1^2} + b_1 = \frac{a_1 + b_1 + a_1^2 b_1 + a_1 b_1^2}{1-a_1^2} = \frac{(a_1+b_1)(1+a_1b_1)}{1-a_1^2}$$

$$R(k) = \frac{(a_1+b_1)(1+a_1b_1)}{1-a_1^2} a_1^{k-1}, \quad k > 0.$$

Следовательно, корреляция равна

$$r(k) = \frac{(a_1+b_1)(1+a_1b_1)}{1+2a_1b_1+b_1^2} a_1^{k-1}, \quad k > 0.$$

Модели ARMA(p,q) достаточно хорошо изучены и успешно применяются при описании стационарных временных рядов. Однако стационарность есть не всегда. Но, переходя от временного ряда $x = (x_n)$ к ряду разностей $h = (h_n)$, где $h_n = \Delta x_n \equiv x_n - x_{n-1}$, или же разностей более высокого порядка: $h_n = \Delta^d x_n = (1-L)^d x_n$, получается получить стационарность. Именно из этих соображений и появилась модель ARIMA(p,d,q).

Определение 4. Будем говорить, что последоватеьность $x = (x_n)$ образует ARIMA(p,d,q)-модель, если последовательность $\Delta^d x = (\Delta^d x)$ образует ARMA(p,q)-модель.

Неформально это можно записать так:

$$\Delta^d \mathrm{ARIMA}(p,d,q) = \mathrm{ARMA}(p,q).$$

Проясним смысл модели на примере ARIMA$(0,1,1)$. Она устроена следующим образом: $h_n = \Delta x_n$, где $h = (h_n)$ является моделью ARMA$(0,1) =$ MA(1):

$$\Delta x_n = \mu + (b_0 + b_1 L)\varepsilon.$$

Если ввести оператор «интегрирования» S по правилу $S \equiv \Delta^{-1}$, или, что равносильно,

$$S = (1-L)^{-1} = 1 + L + L^2 + \dots$$

то формально можно записать, что $x_n = (Sh)_n$, где $h_n = \mu + b_0\varepsilon_n + b_1\varepsilon_{n-1}$.

Следовательно, $x = (x_n)$ можно рассматривать, как результат «интегрирования» последовательности $h = (h_n)$, подчиняющейся модели MA(1). Это объясняет происхождение названия: ARIMA = AR + I + MA, где I происходит от слова «Integrated».