Авторегрессионные модели

Р.В. Балуев — Анализ временных рядов

СПбГУ

СПб, 8 мая 2020 г.

Откуда они берутся

Можно записать следующее тригонометрическое тождество:

$$\sin n\alpha = 2\cos\alpha \sin(n-1)\alpha - \sin(n-2)\alpha. \tag{1}$$

Отсюда следует, что $x_n = \sin n\alpha$ можно представить такой моделью:

$$x_n = a x_{n-1} - x_{n-2}, \quad a = 2\cos\alpha. \tag{2}$$

Таким образом, периодический сигнал можно выразить как линейную комбинацию его предыдущих значений, взятых с некоторым лагом.

Определение

Авторегрессионной моделью временного ряда, $AR(p)$, будем называть следующую последовательность:

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} + u_n, \tag{3}$$

где u_n — шумовая добавка (случайная величина). Мы будем предполагать $u_n \in \mathcal{N}(0, \sigma^2)$, т.е. нормальным. При этом разные u_n статистически независимы.

Шумовые добавки u_n делают модель не строго периодической, а лишь приближённо, т.е. модель похожа на узкополосный шум. По этой причине такие модели удобно использовать для описания квазипериодических явлений, как например солнечная активность (она носит циклический характер, но при этом не строго периодична).

Числа $a_1, a_2, \ldots, a_p$ и σ^2 являются параметрами модели, которые определяют ее основных характеристики (корреляционная функция, спектр мощности).

Непрерывная AR-модель

Для дальнейшего удобства обращения с формулами, введём интегральный аналог AR-модели (3):

$$x(t) = -\int\limits_{-\infty}^{+\infty} a(\tau)x(t-\tau)d\tau + u(t), \qquad (4)$$

где $u(t)$ — шумовой случайный процесс. Теперь применим к (4) преобразование Фурье:

$$X(\nu) = -A(\nu)X(\nu) + U(\nu), \quad F(\nu) = \int\limits_{-\infty}^{+\infty} f(t)e^{-2\pi i\nu t}dt. \qquad (5)$$

Спектр мощности

Таким образом,

$$X(\nu) = \frac{U(\nu)}{1 + A(\nu)}, \implies |X(\nu)|^2 = \frac{|U(\nu)|^2}{|1 + A(\nu)|^2}. \tag{6}$$

Отсюда следует соотношение между спектрами мощности:

$$G_x(\nu) = \frac{G_u(\nu)}{\left|1 + \int_{-\infty}^{+\infty} a(t) e^{-2\pi i \nu t} dt\right|^2}. \tag{7}$$

Если $u(t)$ — белый шум, то $G_u \propto \sigma^2$.

Спектр мощности: обратно к дискретному случаю

Для первоначального определения (3) функцию $a(t)$ можно записать так:

$$a(t) = \sum_{k=1}^{p} a_k \delta(t - t_k), \quad t_k = k\Delta t. \tag{8}$$

И аналогично для $x(t)$ и $u(t)$. В этом случае непрерывное преобразование Фурье автоматически заменяется на дискретное, а его квадрат модуля представляет собой периодограмму, $P_x(\nu)$ или $P_u(\nu)$. Отсюда:

$$P_x(\nu) = \frac{P_u(\nu)}{\left|1 + \sum_{k=1}^{p} a_k e^{-2\pi i \nu t_k}\right|^2}. \tag{9}$$

При этом $\mathbb{E}P_u \equiv \sigma^2$, соответственно следует формула для $\mathbb{E}P_x$.

Корреляционная функция

Домножая (5) на $X^*(\nu)$, запишем:

$$|X(\nu)|^2 = -A(\nu)|X(nu)|^2 + U(\nu)X^*(\nu). \tag{10}$$

Применим теперь обратное преобразование Фурье и воспользуемся теоремой Винера–Хинчина и теоремой о свёртке:

$$k_x(\tau) = -\int\limits_{-\infty}^{+\infty} a(t)k_x(\tau - t)dt + R_{xu}(\tau), \tag{11}$$

где R_{xu} — кросс-корреляционная функция $x(t)$ и $u(t)$. Аналогично, применим комплексное сопряжение к (5), домножим на $U(\nu)$, применим обратное преобразование Фурье:

$$R_{xu}(\tau) = -\int\limits_{-\infty}^{+\infty} a^*(-t)R_{xu}(\tau - t)dt + k_u(\tau). \tag{12}$$

Корреляционная функция

Теперь воспользуемся тем фактом, что связь между x и u носит характер причинно-следственной, то есть $x(t)$ не может зависеть от $u(t')$, если $t' > t$. Математически это следует из того, что $a(t) = 0$ при $t < 0$ (это естественное требование на a). Физически, настоящее может зависеть от прошлого, но не от будущего. Отсюда следует, что $R_{xu}(t) = 0$, если $t > 0$.
Тогда в (11) можно опустить член $R_{xu}(\tau)$, если $\tau > 0$. А чтобы получить $R_{xu}(0)$, подставим $\tau = 0$ в (12):

$$R_{xu}(0) = k_u(0) = \sigma^2. \tag{13}$$

Интеграл исчез, потому что $a^*(-t)R_{xu}(-t) = 0$ (либо первый сомножитель нулевой, либо второй).

Корреляционная функция

Таким образом, корреляционная функция должна удовлетворять следующим интегральным уравнениям:

$$k_x(\tau) = -\int\limits_0^{+\infty} a(t) k_x(\tau - t) dt, \quad \tau > 0,$$

$$k_x(0) = \sigma^2 - \int\limits_0^{+\infty} a(t) k_x(t) dt, \quad \tau = 0. \qquad (14)$$

Здесь мы учли чётность $k_x(\tau)$.

Корреляционная функция — дискретный случай

Вернёмся к нашей основной модели $AR(p)$:

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} + u_n. \tag{15}$$

Нам нужно определить дискретную коррелляционную функцию $\varkappa_m = \mathbb{E}(x_n x_{n\pm m})$. Домножим (15) на x_{n-m} и пременим оператор мат. ожидания:

$$\mathbb{E}(x_n x_{n-m}) = -\sum_{k=1}^{p} a_k \mathbb{E}(x_{n-k} x_{n-m}) + \mathbb{E}(u_n x_{n-m}). \tag{16}$$

Отсюда следует

$$\varkappa_m = -\sum_{k=1}^{p} a_k \varkappa_{|m-k|} + \mathbb{E}(u_n x_{n-m}). \tag{17}$$

Уравнения Юля–Уолкера

Аналогично непрерывному случаю, x_{n-m} не может зависеть от будущего u_n. Значит $\mathbb{E}(u_n x_{n-m}) = 0$, если $m > 0$. Корреляционный коэффициент $\mathbb{E}(x_n u_n)$ можно получить, домножая (15) на u_n:

$$\mathbb{E}(u_n x_n) = \mathbb{E}u_n^2 = \sigma^2. \tag{18}$$

Тогда получаем такую линейную систему уравнений:

$$\varkappa_m = \begin{cases} \sigma^2 - \sum_{k=1}^{p} a_k \varkappa_k, & m = 0, \\ -\sum_{k=1}^{p} a_k \varkappa_{|m-k|}, & m = 1, 2, \ldots, p. \end{cases} \tag{19}$$

Эти уравнения можно записать в матричной форме:

$$\begin{bmatrix} \varkappa_0 & \varkappa_1 & \varkappa_2 & \ldots & \varkappa_p \\ \varkappa_1 & \varkappa_0 & \varkappa_1 & \ldots & \varkappa_{p-1} \\ \varkappa_2 & \varkappa_1 & \varkappa_0 & \ldots & \varkappa_{p-2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \varkappa_p & \varkappa_{p-1} & \varkappa_{p-2} & \ldots & \varkappa_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \ldots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix} \tag{20}$$

Эта система и называется уравнениями Юля–Уолкера (Yule–Walker).

Уравнения Юля–Уолкера

$$\begin{bmatrix} \varkappa_0 & \varkappa_1 & \varkappa_2 & \dots & \varkappa_p \\ \varkappa_1 & \varkappa_0 & \varkappa_1 & \dots & \varkappa_{p-1} \\ \varkappa_2 & \varkappa_1 & \varkappa_0 & \dots & \varkappa_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \varkappa_p & \varkappa_{p-1} & \varkappa_{p-2} & \dots & \varkappa_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \tag{21}$$

Входные величины здесь — корреляционные коэффициенты $\varkappa_m$. Они могут определяться на основе смещённых или несмещённых оценок

$$\tilde{\varkappa}_m = \frac{1}{N} \sum_{k=1}^{N-m} x_k x_{k+m}, \quad \tilde{\varkappa}'_m = \frac{1}{N-m} \sum_{k=1}^{N-m} x_k x_{k+m}, \quad m = 0, 1, 2, \dots, p, \tag{22}$$

при этом p должно быть меньше N.
Неизвестные здесь — это параметры σ^2 и $a_1, a_2, \dots, a_p$.

Уравнения Юля–Уолкера

$$\begin{bmatrix} \varkappa_0 & \varkappa_1 & \varkappa_2 & \dots & \varkappa_p \\ \varkappa_1 & \varkappa_0 & \varkappa_1 & \dots & \varkappa_{p-1} \\ \varkappa_2 & \varkappa_1 & \varkappa_0 & \dots & \varkappa_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \varkappa_p & \varkappa_{p-1} & \varkappa_{p-2} & \dots & \varkappa_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \qquad (23)$$

Матрица в левой части имеет здесь специфический вид, и относится к классу матриц Тёплица (диагонально-постоянная матрица). Поэтому для решения этой линейной системы разработаны специальные численные алгоритмы, учитывающие особенности задачи:

1. Алгоритм Левинсона.
2. Алгоритм Бэрга–Андерсена (maximum entropy method — MEM).

Метод максимальной энтропии Бэрга (Burg) (основы)

Рассмотрим стационарный центрированный гауссовский случайный процесс $x(t)$. Мы наблюдаем набор его значений $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ в дискретные моменты (равномерный временной ряд). При этом каждое x_k распределено нормально с нулевым средним и одинаковой дисперсией σ^2. Плотность их совместного распределения есть также многомерное нормальное распределение $p_N(x_1, x_2, \ldots, x_N)$, или $p_N(\mathbf{x})$. Введём понятие энтропии, или информации по Шеннону:

$$H = -\int\limits_{\mathbb{R}^N} p_N(\mathbf{x}) \log p_N(\mathbf{x}) d\mathbf{x}. \tag{24}$$

Энтропия дает меру размазанности распределения $p_N(\mathbf{x})$.

Метод максимальной энтропии Бэрга (Burg) (основы)

Если все x_k статистически взаимонезависимы (ДБШ — дискретный белый шум), то $p_N(\mathbf{x})$ распадается на произведение одномерных плотностей p_1 одинакового вида $\mathcal{N}(0, \sigma^2)$. Тогда

$$
\begin{aligned}
H &= -\int\limits_{\mathbb{R}^N} p_1(x_1)\dots p_1(x_N)\log(p_1(x_1)\dots p_1(x_N))d\mathbf{x} = \\
&= -N\int\limits_{\mathbb{R}^N} p_1(x)\log p_1(x)dx. \qquad (25)
\end{aligned}
$$

В общем случае это не так ($x(t)$ не обязательно ДБШ), однако энтропия всё равно растёт пропорционально N, т.е. физический смысл имеет удельная энтропия $h = \lim_{N\to\infty} H/N$.
Для ДБШ величина характеристика h просто соответствует нормальному распределению $p_1(x)$, однако в общем случае она определяется корреляционной функцией или (эквивалентно) спектром мощности.

Метод максимальной энтропии Бэрга (Burg) (основы)

Для гауссовского случайного процесса удельную энтропию можно выразить через спектр мощности $G(\nu)$ так:

$$h = \frac{1}{2}\log(2\nu_{\mathrm{Ny}}) + \frac{1}{4\nu_{\mathrm{Ny}}}\int\limits_{-\nu_{\mathrm{Ny}}}^{+\nu_{\mathrm{Ny}}} \log G(\nu)d\nu, \tag{26}$$

где ν_{Ny} — частота Найквиста.
Идея метода MEM в том, чтобы решить задачу оптимизации с ограниченинями:

$$\int\limits_{-\nu_{\mathrm{Ny}}}^{+\nu_{\mathrm{Ny}}} \log G(\nu)d\nu \longmapsto \max,$$

$$\int\limits_{-\nu_{\mathrm{Ny}}}^{+\nu_{\mathrm{Ny}}} G(\nu)e^{2\pi i\nu\Delta tm}d\nu = \varkappa_m, \tag{27}$$

относительно коэффициентов модели $AR(p)$ и σ^2, которые и определяют вид $G(\nu)$ (см. выше).

Метод максимальной энтропии Бэрга (Burg) (основы)

Таким образом, метод MEM ищет такое решение для параметров модели, чтобы при заданных корреляционных коэффициентах получить наиболее «размазанные», статистически неопределенные значения случайного процесса. Принцип максимума h выражает стремление избежать, по возможности, необоснованно информативных предположений относительно случайного процесса (т.е. каких-либо дополнительных неявных предположений помимо оцененных значений $\varkappa_m$). Заметим, минимизации h (в отличие от максимизации) можно было бы достичь тривиально, положив $\sigma \to 0$ (тогда $h \to -\infty$).