

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МАТЕРИАЛЫ

**Первой Всероссийской молодежной
научной конференции
«МАТЕМАТИЧЕСКОЕ
И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ
ИНФОРМАЦИОННЫХ,
ТЕХНИЧЕСКИХ
И ЭКОНОМИЧЕСКИХ СИСТЕМ»**

Томск, 17–18 мая 2013 г.

*Под общей редакцией
кандидата технических наук И.С. Шмырина*

Томск
Издательский Дом Томского государственного университета
2013

рующий эффект централизации управления ритмом сердца, который обусловлен, в основном, степенью активации нервной системы.

• Вариационный размах (VAR или $MxDMn$) — соответствует разности между длительностью самого большого и самого маленького интервалов. Физиологический смысл $MxDMn$ обычно связывают с активностью парасимпатического отдела вегетативной нервной системы.

Как и говорилось выше, существует ряд других методов оценки ВСР. Остается много нерешенных проблем, которые требуют тщательной проработки и новых исследований в этой области. Хотелось бы разработать методы, в которых после анализа и обработки полученных данных можно было бы делать выводы о развивающихся болезнях, чтобы иметь возможность предотвратить их.

ЛИТЕРАТУРА

1. <http://www.kardi.ru/ru/index/Article?&ViewType=view&Id=37>
2. В. А. Снежицкий Методологические аспекты анализа вариабельности сердечного ритма в клинической практике
3. Баевский, Кириллов, Клецкин, 1984
4. Р. М. Баевский. Вариабельность сердечного ритма. Медико-физиологические аспекты (По материалам сайта Института Внедрения Новых Медицинских Технологий РАМЕНА www.ramena.ru)
5. Баевский Р. М., Иванов Г. Г., Чирейкин Л. В. Анализ вариабельности сердечного ритма при использовании различных электрокардиографических систем// Вестник аритмологии. — № 24. — 2001
6. Рисунки взяты с сайта <http://www.kardi.ru/>

АЛГОРИТМ ОБНАРУЖЕНИЯ СИНХРОННОСТИ ВРЕМЕННЫХ РЯДОВ

Е. С. Катаева

Томский государственный университет
E-mail: aleo@sibmail.com

Будем называть временные ряды синхронными, если их динамика развития сходна. Выявление синхронности поведения временных рядов различных метеорологических параметров является важной задачей — ее решение позволяет выдвигать гипотезу о наличии общих факторов и определять область их действия. Известные способы определения синхронности, такие как применение кластерного анализа в вычислительных пакетах или нахождение коэффициента корреляции, могут быть недоступны или не давать необходимой информации. В данной работе предлагается алгоритм выделения синхронности, основанный на анализе коэффициентов многочленов, аппроксимирующих временные ряды. Применялась аппроксимация двумя способами: разложением по полиномам Чебышева [1, 2] и тригонометрической регрессией Фурье.

Аппроксимация разложением по полиномам Чебышева

Случайные ошибки в значениях функции сильно искажают интерполяционные многочлены высоких степеней, а при интерполяции многочленами низких степеней теряется существенная информация. Поэтому при наличии случайных ошибок часто применяют «сглаживающую» аппроксимацию такими многочленами, которые минимизируют взвешенную среднеквадратическую ошибку аппроксимации.

Для функции $p(x)$, заданной на дискретном множестве $(m+1)$ точек $x_0, x_1, x_2, \dots, x_m$, требуется построить функцию $P(x)$ вида

$$P(x) = k_0 \varphi_0(x) + k_1 \varphi_1(x) + \dots + k_n \varphi_n(x) \quad (1)$$

так, чтобы минимизировать взвешенную среднеквадратическую ошибку вида

$$\sigma^2 = \sum_{k=0}^m \gamma_k [P(x_k) - p(x_k)]^2,$$

где γ_k — заданные положительные веса [1]. Это просто сделать в том случае, когда функции $\varphi_r(x)$ представляют собой многочлены степени r , попарно ортогональные с весами γ_k на заданном множестве точек, то есть когда

$$\sum_{i=0}^m \gamma_i \varphi_i(x_k) \varphi_j(x_k) = 0 \quad (i \neq j).$$

В данной работе в качестве ортогональных полиномов используются полиномы Чебышева 1-го рода. При числе слагаемых, равном числу значений ряда, полиномы Чебышева приводят аппроксимацию в интерполяционной формуле, при меньшем количестве членов можно говорить о среднеквадратическом приближении [3]. Поэтому дальше положим $n = m$. Коэффициенты k_i определяются следующим образом:

$$k_0 = \frac{\sum_{i=0}^m p(x_i) \varphi_0(x_i)}{\sum_{i=0}^m \varphi_0^2(x_i)}; k_1 = \frac{\sum_{i=0}^m p(x_i) \varphi_1(x_i)}{\sum_{i=0}^m \varphi_1^2(x_i)}; \dots; k_n = \frac{\sum_{i=0}^m p(x_i) \varphi_n(x_i)}{\sum_{i=0}^m \varphi_n^2(x_i)}.$$

Теперь можно представить функцию $k_0 \varphi_0(x) + k_1 \varphi_1(x) + \dots + k_n \varphi_n(x)$ в виде

$$p(x) = \frac{\sum_{i=0}^m p(x_i) \varphi_0(x_i)}{\sum_{i=0}^m \varphi_0^2(x_i)} \varphi_0(x) + \frac{\sum_{i=0}^m p(x_i) \varphi_1(x_i)}{\sum_{i=0}^m \varphi_1^2(x_i)} \varphi_1(x) + \dots + \frac{\sum_{i=0}^m p(x_i) \varphi_n(x_i)}{\sum_{i=0}^m \varphi_n^2(x_i)} \varphi_n(x).$$

Если $n < m$, то добавление нового члена $k_{n+1} \varphi_{n+1}(x)$ оставляет предыдущие члены без изменения.

Из тригонометрического тождества

$$\cos(n+1)\theta + \cos(n-1)\theta = 2 \cos \theta \cos n\theta, \quad (2)$$

полагая в нем $\theta = \arccos x$ для функций

$$T_n(x) = \cos(n \arccos x), \quad x \in [-1, 1], \quad n = 0, 1, \dots, \quad (3)$$

находим рекуррентное соотношение

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \quad (4)$$

Поскольку $T_0(x) = 1$, $T_1(x) = x$, то и при любом n функция $T_n(x)$ есть алгебраический многочлен степени n . Эти многочлены называются многочленами Чебышева первого рода [3]. С помощью рекуррентной формулы (4) их можно вычислять последовательно:

$$\begin{aligned}
 T_0(x) &= 1, \\
 T_1(x) &= x, \\
 T_2(x) &= 2x^2 - 1, \\
 T_3(x) &= 4x^3 - 3x, \\
 &\dots\dots\dots
 \end{aligned}$$

Можно рассматривать смещение сегмента $[-1, 1]$ на любой сегмент $[a, b]$. Так как линейное преобразование этих сегментов при $t \in [-1, 1]$ и $x \in [a, b]$ осуществляется по формуле $2x = (b - a)t + a + b$, то соответствующие смещенные многочлены Чебышева имеют вид: $T_n\left(\frac{2x - a - b}{b - a}\right)$.

Аппроксимация тригонометрической регрессией Фурье

Рассматриваемая в данном случае модель состоит в том, что наблюдаемая величина y_t представлена в виде суммы тренда $f(t)$, периодической составляющей $g(t)$ и ненаблюдаемой ошибки u_t , т.е. $y_t = f(t) + g(t) + u_t$, где все u_t некоррелированные, имеют средние значения $Mu_t = 0$ и дисперсии $Mu_t^2 = \sigma^2$.

На первом этапе методом наименьших квадратов выделяются и исключаются линейные тренды. При рассмотрении тренда только в наблюдаемые моменты времени его можно точно выразить через T линейно независимых тригонометрических функций. Если же период тренда равен $m < T$, то все его первые m значений затем повторяются еще $\left(\frac{T}{m} - 1\right)$ раз, т.е. всего h раз, поэтому в точное разложение функции в точках $t = 1, \dots, T$ достаточно включить m членов, которые дают точное представление функции в точках $t = 1, \dots, m$, а все остальные значения повторяют первые m значений.

Функции $\cos \frac{2\pi j}{m}t$, $\sin \frac{2\pi j}{m}t$ имеют период $\frac{m}{j}$, причем этот период укладывается в общей длине ряда $\frac{Tj}{m} = h_j$ раз, т.е. целое число раз, если j - целое. Подберем m таких функций с наименьшими периодами. В разложение необходимо включить константу, т.е. $\varphi_0(t) \equiv 1$.

Затем последовательно включаются пары тригонометрических функций $\cos \frac{2\pi j}{m}t$, $\sin \frac{2\pi j}{m}t$ при $j = 1, 2, \dots$, причем каждому j соответствует пара функций с периодом $\frac{m}{j}$; дополнительно включается функция $\varphi_{m-1}(t) = (-1)^t$ с $j = \frac{m}{2}$, имеющая период 2.

Окончательно получаем следующее представление периодического тренда:

$$f(t, \alpha) = \sum_{i=0}^{m-1} \alpha_i \varphi_i(t) = \alpha_0 + \sum_{j=1}^{\frac{m-1}{2}} \left(\alpha_{2j-1} \cos \frac{2\pi j}{m} t + \alpha_{2j} \sin \frac{2\pi j}{m} t \right) + \alpha_{m-1} (-1)^t.$$

Для оценки параметров α_i получившейся теоретико-вероятностной схемы используется метод наименьших квадратов. В [5] получены следующие оценки:

$$\begin{aligned} \hat{\alpha}_0 &= \bar{y}, \\ \hat{\alpha}_{2j-1} &= \frac{2}{T} \sum_{t=1}^T y_t \cos \left(\frac{2\pi j}{m} t \right), \quad j = 1, \dots, \frac{m-1}{2}, \\ \hat{\alpha}_{2j} &= \frac{2}{T} \sum_{t=1}^T y_t \sin \left(\frac{2\pi j}{m} t \right), \quad j = 1, \dots, \frac{m-1}{2}, \\ \hat{\alpha}_{m-1} &= \frac{1}{T} \sum_{t=1}^T y_t (-1)^t. \end{aligned}$$

Таким образом, точечные оценки тренда определяются выражением

$$\hat{y}_t = \bar{y} + \frac{2}{T} \sum_{j=1}^{\frac{m-1}{2}} \left(\cos \left(\frac{2\pi j}{m} t \right) \sum_{t=1}^T (y_t \cos \left(\frac{2\pi j}{m} t \right)) + \sin \left(\frac{2\pi j}{m} t \right) \sum_{t=1}^T (y_t \sin \left(\frac{2\pi j}{m} t \right)) \right) + \frac{(-1)^t}{T} \sum_{t=1}^T y_t (-1)^t.$$

Определение синхронности временных рядов

Исследуем спектр (коэффициенты) аппроксимирующих разложений для определения синхронно развивающихся рядов. Будем считать ряды *синхронными*, если многочлены, аппроксимирующие их, равны. 2 многочлена считаются *равными* в том и только в том случае, если равны коэффициенты при соответствующих степенях неизвестного. На практике будем считать коэффициенты *равными*, если их разность меньше определенного числа: $|A_{ij} - A_{kj}| < E$, где $A = \{A_{mj}\}$, m — номер временного ряда, j — номер коэффициента. Для каждой пары рядов рассчитывается число равных коэффициентов.

Были исследованы разложения рядов среднемесячных температур и уровней осадков по 35 метеостанциям на Алтае за 48 лет и рядов среднемесячных температур по 134 метеостанциям на Азиатской территории России. Результаты определения синхронности по разложениям Чебышева:

1) Для рядов значений температуры на алтайских станциях выделены 6 кластеров синхронности:

{Алейская, Благовещенка, Волчиха, Камень-на-Оби, Ключи, Кулунда, Рубцовск, Славгород, Шелаболиха};

{Бийск-Зональная, Тогул, Троицкое, Турочак};

{Змеиногоorsk, Краснощеково, Чарышское, Горно-Алтайск};

{Ак-Кем, Кара-Тюрек};

{Онгудай, Усть-Кокса};

{Чемал, Яйлю, Бея}.

2) Для рядов уровней осадков на алтайских станциях выделены 6 кластеров синхронности:

{Барнаул, Алейская, Бийск, Змеиногоorsk, Камень-на-Оби, Рубцовск, Славгород, Шелаболиха};

{Благовещенка, Волчиха, Ключи, Кулунда};

{Солонешное, Чарышское, Горно-Алтайск, Кара-Тюрек};
 {Тогул, Ак-Кем, Турочак, Усть-Кан, Усть-Кокса, Чемал, Бея};
 {Катанда, Кызыл-Озек, Яйлю};
 {Кош-Агач, Усть-Улаган, Бертек, Уландрык}.

3) Для рядов значений температуры на 134 станциях выделены 7 кластеров синхронности:

Результаты для значений температуры на Азиатской территории России проиллюстрированы на рисунках 1 и 2:

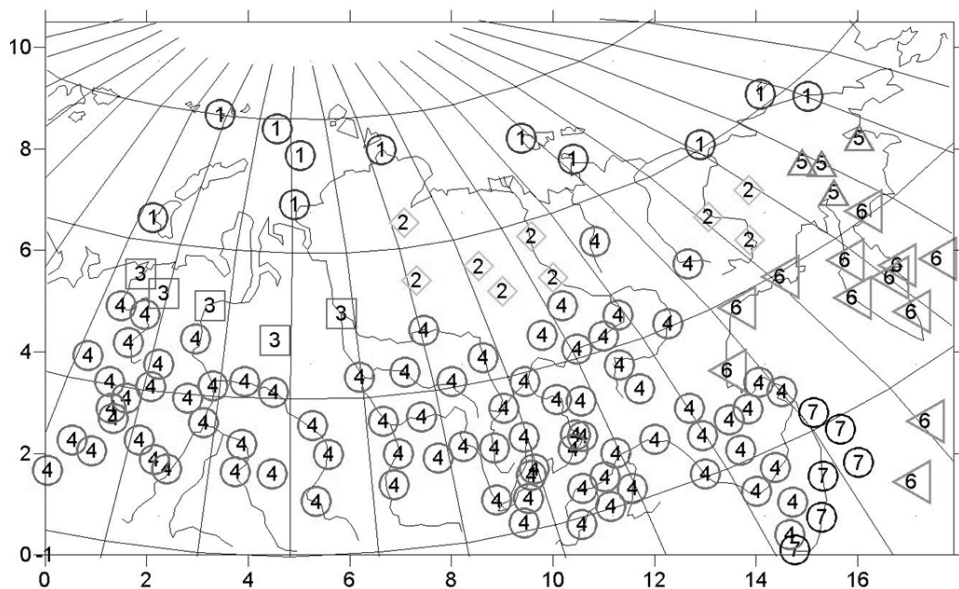


Рис. 1. Результат выделения станций с синхронной динамикой на АТР при аппроксимации с использованием полиномов Чебышева.

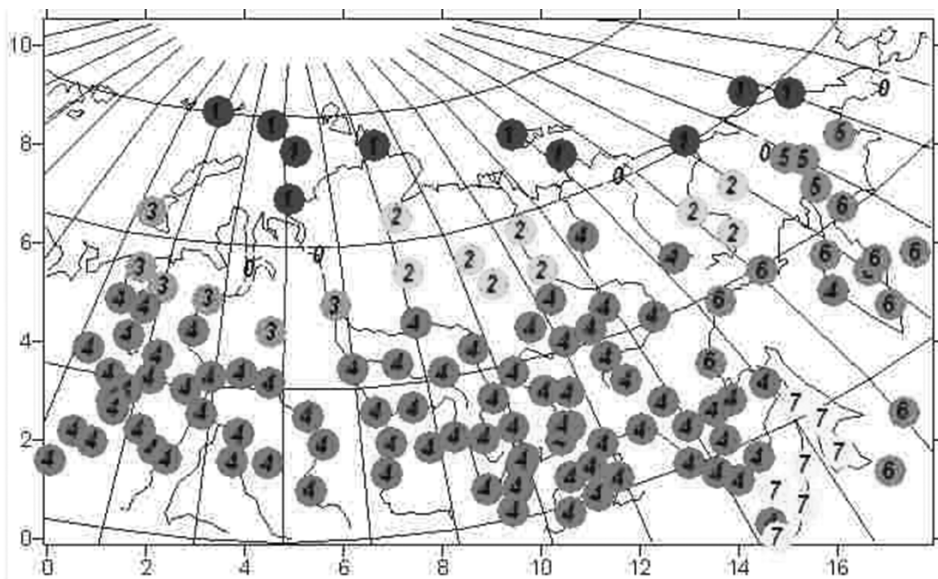


Рис. 2. Результат выделения станций с синхронной динамикой на АТР при аппроксимации тригонометрической регрессией Фурье.

Для рядов температуры на алтайских станциях такие же результаты получены и при исключении сезонной составляющей, из чего можно сделать вывод о сходной динамике сезонности.

ЛИТЕРАТУРА

1. Корн Г. Справочник по математике для научных работников и инженеров: определения, теоремы, формулы. / Г. Корн, Т. Корн. – М. : Главная редакция физико-математической литературы издательства «Наука», 1974. – 832 с.
2. Суетин П. К. Классические ортогональные полиномы / П. К. Суетин. – 2-е изд., доп. – М. : Главная редакция физико-математической литературы издательства «Наука», 1979. – 416 с.
3. Бухановский А. В. Вероятностный анализ и моделирование вертикально неоднородного океанологического поля / А.В. Бухановский, Е.Н. Захарченко, В.А. Рожков // Навигация и гидрография. – 1999. - № 9. – С.86-103.
4. Данилов Ю. А. Многочлены Чебышева / Ю. А. Данилов. – Минск : Вышэйшая школа, 1984. – 157 с.
5. Колемаев В. А. Теория вероятностей и математическая статистика / В.А. Колемаев, В.Н. Калинина. – М.: ИНФРА-М, 1997. – 302 с.

ИССЛЕДОВАНИЕ ЦЕНЫ ПОТРЕБИТЕЛЬСКОГО СПРОСА НА ТОВАР-НОВИНКУ ПО ДВУКРАТНО СПРАВА ЦЕНЗУРИРОВАННЫМ ДАННЫМ

М. А. Муравлева

Томский государственный университет
E-mail: marmyr90@rambler.ru

В процессе проведения маркетинговых исследований маркетологи постоянно сталкиваются с неполными наблюдениями, когда значение исследуемого показателя известно с точностью до некоторого интервала или даже пропущено. При этом зачастую обработка подобных данных проводится без учета их природы, к ним нередко применяются классические методы, ориентированные на полное знание каждого наблюдения. Полученные в итоге результаты являются некорректными и не позволяют сделать адекватные выводы. К неполным, цензурированным данным необходимо применять специально разработанные для этих целей модели [1], которые существенно повышают точность обработки [2]. Еще большее значение при этом придается знанию любой априорной информации об исследуемой случайной величине (с.в.), так как ее привлечение способствует повышению качества статистических процедур [2].

В данной работе рассматривалась случайная величина $\tau > 0$ — цена в рублях за единицу товара — с функцией распределения (ф.р.) $F(t) = P(\tau \leq t)$. Прогрессивно цензурированная выборка (ц.в.) объема N $(X, I) = \{(X_1, I_1), \dots, (X_N, I_N)\}$, где для $i = \overline{1, N}$

$$I_i = \begin{cases} 0, & X_i - \text{полное наблюдение;} \\ 1, & X_i - \text{наблюдение до цензурирования,} \end{cases}$$

строилась по следующей схеме: цензурирование интервалом I типа, моменты цензурирования T_1, T_2 — не случайны, количество неполных наработок в интервале $[T_1, T_2]$ — с.в., численно равная доле g , $0 < g < 1$, от суммарного числа наблю-