

**МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ПРОЦЕССОВ С  
НЕСТАЦИОНАРНЫМИ РЕАЛИЗАЦИЯМИ**  
**Кожевникова И.А.<sup>1</sup>, Швейкина В.И.<sup>2</sup>**

<sup>1</sup>Кожевникова Ирина Аркадьевна - кандидат физико-математических наук, старший научный сотрудник, ведущий научный сотрудник,

Лаборатория математической статистики,  
кафедра математической статистики и случайных процессов, механико-математический факультет,  
Московский государственный университет им. М.В. Ломоносова;

<sup>2</sup>Швейкина Валентина Ивановна - кандидат географических наук, старший научный сотрудник,

Лаборатория глобальной гидрологии,  
Институт водных проблем  
Российская академия наук,  
г. Москва

**Аннотация:** в статье исследована реализация, не являющаяся стационарной в широком смысле. Проанализированы способы аппроксимации рассматриваемого ряда с использованием нормального распределения и  $\Gamma$ -распределения. В обоих случаях с помощью метода максимального правдоподобия получены оценки параметров условно-гауссовской авторегрессии. Методом математического моделирования получены интервалы наиболее вероятных значений рассматриваемой характеристики.

**Ключевые слова:** нелинейная стохастическая условно-гауссовская авторегрессия, метод максимального правдоподобия, вероятностный прогноз.

УДК 551.461

**Введение. Определение 1.** Случайный процесс  $\xi(t)$  называется стационарным в широком смысле, если математическое ожидание  $E\xi(t)$  и дисперсия  $D\xi(t)$  не зависят от  $t$ , а ковариационная функция  $\text{COV}(\xi(t), \xi(s)) = K(s-t)$ .

**Определение 2.** Случайный процесс  $\xi(t)$  называется стационарным в узком смысле или просто стационарным, если  $n$ -мерные функции распределения вероятностей двух конечных групп переменных  $\xi(t_1), \xi(t_2), \dots, \xi(t_n)$  и  $\xi(t_1+u), \xi(t_2+u), \dots, \xi(t_n+u)$  совпадают и, следовательно, не зависят от  $u$ . Числа  $n$  и  $u$ , а также моменты времени  $t_1, t_2, \dots, t_n$  выбраны при этом совершенно произвольно. Обозначим

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P(\xi(t_1) \leq x_1, \xi(t_2) \leq x_2, \dots, \xi(t_n) \leq x_n), \quad \text{тогда}$$
$$F(x_1, x_2, \dots, x_n; t_1+u, t_2+u, \dots, t_n+u) = F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n).$$

Понятно, что знание только первых двух моментов не может всегда полностью определить случайный процесс  $\xi(t)$  и не в состоянии полностью заменить теорию случайных процессов, основанную на распределении вероятностей. Тем не менее, в некоторых случаях теория, основанная на рассмотрении только вторых моментов, оказывается вполне достаточной. Например, указанная теория справедлива в том случае, когда исследуемая реализация (данные) принадлежат выборке из нормального распределения, для которого момент произвольного порядка, даже смешанный, выражается только через первые два момента.

Исследование на стационарность обычно проводят для уточнения метода исследования. В заключение приведем пример случайной дискретной последовательности стационарной в широком смысле, но не стационарной в узком смысле.

1) Пример дискретной последовательности, стационарной в широком смысле, но не стационарной в узком смысле. Пусть  $\xi_1, \xi_2, \dots, \xi_n$  — последовательность независимых случайных величин, у которой  $\xi_{2t-1}$  имеют равномерное распределение вероятностей,  $R[-a, a]$ ,  $a > 0$ , а случайные величины  $\xi_{2t}$  имеют распределение вероятностей  $N(0, \sigma^2)$ ,  $\sigma^2 = a^2/3$ . Из задания параметров распределений следует, что  $E\xi_{2t-1} = E\xi_{2t} = 0$  и, кроме того, при равномерном

распределении  $D_{\xi_{2t-1}} = \frac{(a - (-a))^2}{12} = \frac{4a^2}{12} = \frac{a^2}{3}$ , а при гауссовском распределении аналогично

положим  $D_{\xi_{2t}} = \sigma^2 = \frac{a^3}{3}$ . Это означает, что независимые случайные величины, расположенные на

нечётных и чётных местах  $\xi_{2t-1} \sim R[-a, a]$ ,  $D_{\xi_{2t}} \sim N(0, a^2/3)$  имеют разные распределения, что обеспечивает отсутствие стационарности ости в широком смысле.

2) Пример последовательности стационарной в узком смысле, но не стационарной в широком смысле. Пусть случайная величина  $\xi$  имеет распределение Коши с плотностью

$$p(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

Моменты этой случайной величины не существуют, т.е.  $E\xi$ ,  $E\xi^2$ , ..., и так далее обращаются в бесконечность [Б.В. Гнеденко. Курс теории вероятностей. Физматгиз, 1965]. Это обстоятельство означает отсутствие стационарности в широком смысле. В узком смысле стационарность существует, это не сложно увидеть, используя последовательность независимых случайных величин, имеющих распределение Коши.

Ранее нами был разработан метод моделирования и прогнозирования гидрологических рядов, которые можно было представить в виде суммы детерминированной и случайной частей [1, 2]. При этом детерминированная часть модели представляла собой сумму полиномиального тренда и тригонометрического полинома, а случайная часть состояла из суммы независимых случайных величин, как правило, имеющих нормальное распределение. В этой статье анализируются последовательности наблюдений, которые не разделяются на детерминированную и случайную части и которые не являются стационарными в широком смысле. В 90-х годах 20 века А.Н.Ширяев [3] опубликовал метод обработки и прогнозирования не стационарных в широком смысле временных рядов на основании нелинейной условно-гауссовской стохастической модели ARCH(p)  $p \geq 1$ ,  $p$  – порядок авторегрессии, целое. В условной авторегрессии ARCH(p), входящая в неё последовательность независимых одинаково распределённых случайных величин имеет множитель, зависящий от предыдущих значений выборки. Значение имеет также вид распределения, которому подчиняется последовательность независимых одинаково распределённых случайных величин, входящая в ARCH(p). В статье рассматриваются два вида распределения последовательности независимых случайных величин: нормальное распределение и Г-распределение. В зависимости от вида распределения выведена формула оценки неизвестного параметра условной авторегрессии для порядка  $p = 1$ . Построены соответствующие моделированные траектории и рассмотрены методы вероятностного прогноза рассматриваемой последовательности.

**Исходные данные.** Исходными данными являются наблюдения за гидрохимической характеристикой, определяющей содержание железа в воде реки Исеть (рис. 1, 2, 3), которые наглядно показывают влияние города на состав воды после прохождения реки через город. Наблюдения проводились с 2000 года в течение 10 лет.

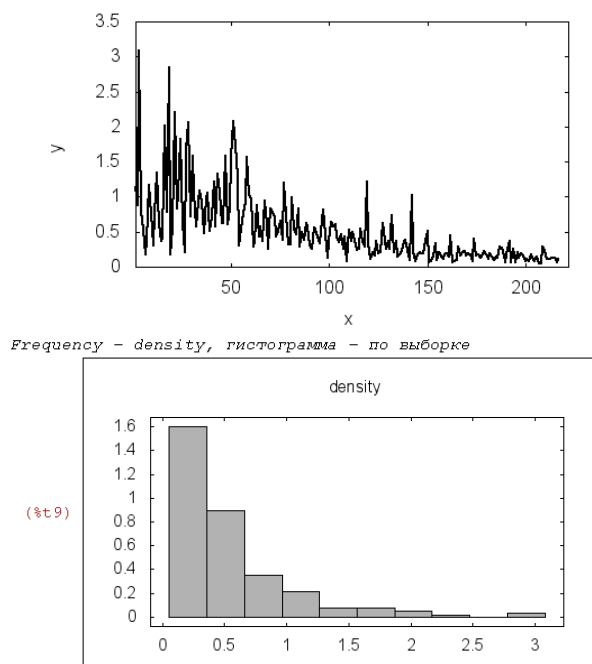


Рис. 1. Содержание железа в воде 7 км ниже г. Екатеринбурга, реализация и соответствующая реализации гистограмма

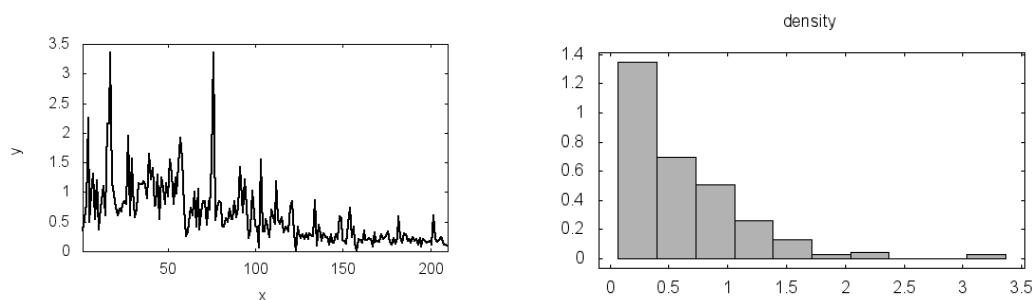


Рис. 2. Содержание железа в воде г. Екатеринбург, реализация и гистограмма

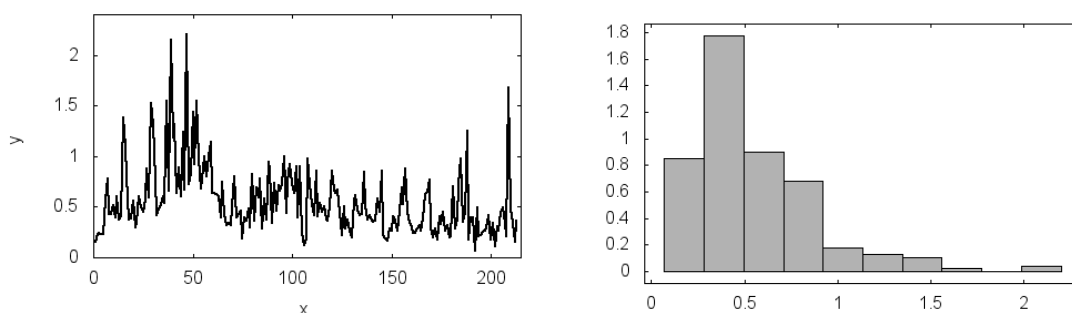


Рис. 3. Содержание железа в воде выше г. Екатеринбурга, реализация и соответствующая реализации гистограмма

Будем исследовать, например, реализацию наблюдений о содержании железа в воде в пункте выше г. Екатеринбурга, т.е. до протекания ее через город, длина реализации равна 213. Заметим также, что исследуемая последовательность наблюдений не является стационарной в широком смысле.

В дальнейшем в этой статье будет рассмотрено два случая:

- а) выборка, которой принадлежит реализация, имеет усеченное нормальное распределение;
- б) выборка, которой принадлежит реализация, имеет гамма-распределение.

В каждом из этих рассматриваемых случаев параметры плотности распределения были получены методом максимального правдоподобия на основании соответствующей реализации [4].

1. Рассмотрим плотности нормального распределения (рис. 4).

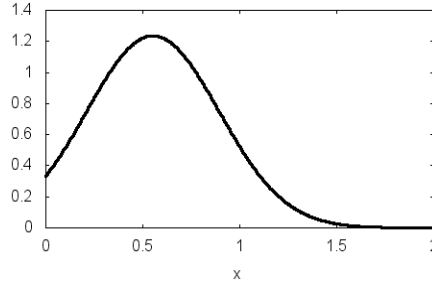


Рис. 4. Усеченное и нормированное нормальное распределение содержания железа в воде выше Екатеринбурга  $N(a, \sigma^2)$ ,  $a = 0.5520$ ,  $\sigma^2 = 0.1163$ , площадь под графиком изображенной плотности равна 1

**Условная авторегрессия.** Чтобы анализировать исследуемую последовательность наблюдений (рис 3), аппроксимируем её условной авторегрессией 1-го порядка ARCH(1) [3]

$$X_t = \beta_0 + \beta_1 X_{t-1} + \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2} \varepsilon_t, \quad (1)$$

где  $\varepsilon_t$  — последовательность независимых одинаково распределенных случайных величин с заданным распределением, а последовательность  $(X_t)$  удовлетворяет модели авторегрессии, со случайным ARCH(1)–шумом  $(\sqrt{\alpha_0 + \alpha_1 X_{t-1}^2} \varepsilon_t)_{t \geq 1}$ , причём  $\varepsilon_t = \frac{X_t - \beta_0 - \beta_1 X_{t-1}}{\sqrt{\alpha_0 + \alpha_1 X_{t-1}^2}}$ .

Параметры условной авторегрессии  $\alpha_0$ ,  $\alpha_1$  принимают одни и те же значения на протяжении всей статьи и равны:  $\alpha_0 = 0.9$ ,  $\alpha_1 = 0.2$ .

1. Пусть последовательность  $\varepsilon_t$  имеет нормальное распределение, многомерная плотность которого задаётся следующим образом:

$$p_\theta(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \prod_{k=1}^n (\alpha_0 + \alpha_1 x_{k-1}^2)^{1/2}} \cdot \exp\left(-\frac{1}{2} \sum_{k=1}^n \frac{(x_k - \beta_0 - \beta_1 x_{k-1})^2}{\alpha_0 + \alpha_1 x_{k-1}^2}\right). \quad (2)$$

Параметры  $\beta_0$  и  $\beta_1$ , входящие в (1) неизвестны. Получим их оценки методом максимального правдоподобия, дифференцируя логарифм плотности по соответствующему параметру.

$$\begin{cases} \frac{d \ln p_{(\alpha_0, \alpha_1, \beta_0, \beta_1)}}{d \beta_0}(x_1, \dots, x_n) = 0, \\ \frac{d \ln p_{(\alpha_0, \alpha_1, \beta_0, \beta_1)}}{d \beta_1}(x_1, \dots, x_n) = 0, \end{cases}$$

$$\ln p(x) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{k=1}^n \ln(\alpha_0 + \alpha_1 x_{k-1}^2) - \frac{1}{2} \sum_{k=1}^n \frac{(x_k - \beta_0 - \beta_1 x_{k-1})^2}{(\alpha_0 + \alpha_1 x_{k-1}^2)^2}.$$

Дифференцируем логарифм многомерной плотности (2) по неизвестным параметрам.

$$\begin{cases} \frac{d \ln p}{d \beta_0} = \frac{1}{2} \sum_{k=1}^n \frac{2(x_k - \beta_0 - \beta_1 x_{k-1})}{\alpha_0 + \alpha_1 x_{k-1}^2} = 0, \\ \frac{d \ln p}{d \beta_1} = \frac{1}{2} \sum_{k=1}^n \frac{2(x_k - \beta_0 - \beta_1 x_{k-1}) x_{k-1}}{\alpha_0 + \alpha_1 x_{k-1}^2} = 0, \end{cases}$$

и получаем систему уравнений

$$\begin{cases} \sum_{k=1}^n \frac{x_k}{\alpha_0 + \alpha_1 x_{k-1}^2} - \beta_0 \sum_{k=1}^n \frac{1}{\alpha_0 + \alpha_1 x_{k-1}^2} - \beta_1 \sum_{k=1}^n \frac{x_{k-1}}{\alpha_0 + \alpha_1 x_{k-1}^2} = 0, \\ \sum_{k=1}^n \frac{x_k x_{k-1}}{\alpha_0 + \alpha_1 x_{k-1}^2} - \beta_0 \sum_{k=1}^n \frac{x_{k-1}}{\alpha_0 + \alpha_1 x_{k-1}^2} - \beta_1 \sum_{k=1}^n \frac{x_k^2}{\alpha_0 + \alpha_1 x_{k-1}^2} = 0, \end{cases}$$

Записываем систему уравнений в традиционной форме и решаем её относительно параметров

$$\begin{cases} \beta_0 \sum_{k=1}^n \frac{1}{\alpha_0 + \alpha_1 x_{k-1}^2} + \beta_1 \sum_{k=1}^n \frac{x_{k-1}}{\alpha_0 + \alpha_1 x_{k-1}^2} = \sum_{k=1}^n \frac{x_k}{\alpha_0 + \alpha_1 x_{k-1}^2}, \\ \beta_0 \sum_{k=1}^n \frac{x_{k-1}}{\alpha_0 + \alpha_1 x_{k-1}^2} + \beta_1 \sum_{k=1}^n \frac{x_{k-1}^2}{\alpha_0 + \alpha_1 x_{k-1}^2} = \sum_{k=1}^n \frac{x_k x_{k-1}}{\alpha_0 + \alpha_1 x_{k-1}^2}. \end{cases} \quad (3)$$

или

$$\begin{cases} Ax + By = C, \\ Dx + Ey = F. \end{cases}$$

Решение этой системы записываем в виде

$$\beta_0 = \frac{CE - FB}{AE - BD}, \quad \beta_1 = \frac{AF - DC}{AE - BD}.$$

Для имеющейся реализации, в предположении нормальности, оценки параметров соответственно равны:

$$\hat{\beta}_0 = 0.33, \hat{\beta}_1 = 0.4$$

При этих значениях оценок параметров в для условной авторегрессии в случае нормального распределения  $N(0,1)$  получены следующие моделированные траектории (рис. 5).

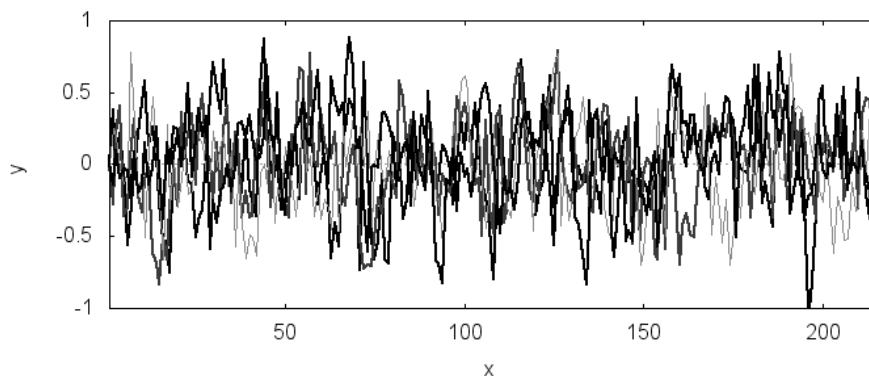


Рис. 5. Моделированные траектории для  $ARCH(1)$  гауссовского распределения  $\alpha_0 = 0.9$ ,  $\alpha_1 = 0.2$ ,  $\hat{\beta}_0 = 0.33, \hat{\beta}_1 = 0.4$ ,  $N(0; 0.3)$ , стандартное отклонение  $\sigma = 0.3$

Из рис. 5 видно, что моделированные траектории почти не имеют существенных положительных выбросов, характерных для исходной реализации, поэтому параметры оцененные в предположении нормального распределения, дают недостаточное представление о действительном распределении..

2. Рассмотрим  $\Gamma$ -распределение для последовательности  $\mathcal{E}_t$  и оценим его параметры по имеющейся реализации. Плотность вероятностей  $\Gamma$ -распределения имеет вид:

$$p(\theta, x) = \begin{cases} \frac{(x+a)^{k-1} e^{-(x+a)/\theta}}{\theta^k \Gamma(k)}, & x > a, \quad k > 0, \\ 0, & x < a. \end{cases}$$

Сдвиг  $a$  по оси  $Ox$  равен минимальному значению рассматриваемой реализации ( $a = -0.0822$ ) и его можно не оценивать с помощью функции правдоподобия, а определить по выборке. Пусть число степеней свободы  $\Gamma$ -распределения  $k = 2$ .

Логарифм функции правдоподобия для одного произвольного наблюдения  $1 \leq i \leq n$  выражается следующим образом:

$$\ln p(\theta, x_i) = (k-1) \ln(x_i) - \frac{x_i}{\theta} - k \ln \theta - \ln(\Gamma(k))$$

Логарифм функции правдоподобия для реализации  $x_1, \dots, x_n$  имеет вид:

$$\ln L(\theta, x_i) = \sum_{i=1}^n \ln(p(\theta, x_i)) = (k-1) \ln(x_i) - \frac{1}{\theta} \sum_{i=1}^n x_i - nk \ln(\theta) - n \ln(2\pi)$$

Дифференцируем функцию правдоподобия по параметру  $\theta$ , получим

$$\frac{d}{d\theta} \left( \sum_{i=1}^n \ln(p(\theta, x_i)) \right) = \frac{1}{\theta^2} \sum_{i=1}^n x_i - \frac{nk}{\theta} = 0.$$

Отсюда находим оценку неизвестного параметра  $\theta$ ,  $\hat{\theta}_n = \frac{1}{nk} \sum_{i=1}^n x_i$ .

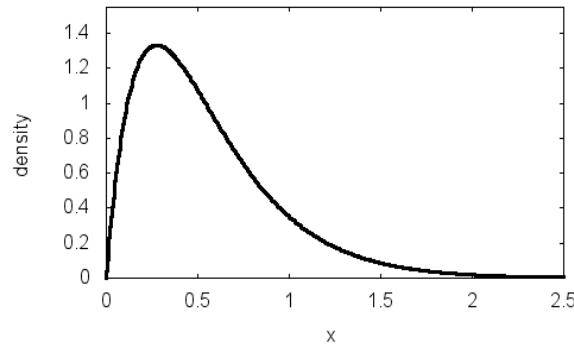


Рис. 6. Плотность  $\Gamma$ -распределения с параметром  $\hat{\theta}_n = 0.276$ , число степеней свободы  $k = 2$

Из сравнения рис. 3 (гистограмма, построенная по выборке) и рис. 6 (плотность  $\Gamma$ -распределения) можно сделать вывод в пользу  $\Gamma$ -распределения, но не усеченного гауссовского.

#### Оценки параметров условной авторегрессии в случае $\Gamma$ -распределения

Для условной авторегрессии ARCH(1)

$$X_t = \beta_0 + \beta_1 X_{t-1} + \sqrt{a_0 + \alpha_1 X_{t-1}^2} \varepsilon_t, \quad \alpha_0 = 0.9, \quad \alpha_1 = 0.2, \quad \beta_0 = 0$$

где  $\varepsilon_t = \frac{X_t - \beta_0 - \beta_1 X_{t-1}}{\sqrt{a_0 + \alpha_1 X_{t-1}^2}}$  имеет  $\Gamma$ -распределение с плотностью  $p(\theta, x) = \begin{cases} \frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)}, & x \geq 0, \quad k > 1, \quad \theta > 0, \\ 0, & x < 0, \end{cases}$

параметры которой оцениваются по реализации. Пусть  $k = 2$ ,  $\hat{\theta} = 0.276$ ,  $\beta_0 = 0$

Запишем многомерную плотность  $\Gamma$ -распределения для  $n$ -независимых одинаково распределенных случайных величин  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ , имеющих совместное распределение вероятностей

$$p_\theta(x_1, \dots, x_n) = \frac{1}{\theta^{kn} (\Gamma(k))^n} \prod_{j=1}^n \left( \frac{x_j - \beta_0 - \beta_1 x_{j-1}}{\sqrt{a_0 + \alpha_1 x_{j-1}^2}} \right)^{k-1} \cdot \exp \left( -\frac{x_j - \beta_0 - \beta_1 x_{j-1}}{\theta \cdot \sqrt{a_0 + \alpha_1 x_{j-1}^2}} \right),$$

где  $n$  — длина выборки. Будем считать  $\beta_0 = 0$ .

Чтобы оценить параметр  $\beta_1$ , логарифмируем выражение для  $n$ -мерной плотности. Распределения вероятностей

$$\ln(p_\theta(x_1, \dots, x_n)) = -kn \ln \theta - n \ln \Gamma(k) + \sum_{j=1}^n (k-1) \ln \underbrace{\left( \frac{x_j - \beta_0 - \beta_1 x_{j-1}}{\sqrt{a_0 + \alpha_1 x_{j-1}^2}} \right)}_{A_j} - \sum_{j=1}^n \frac{x_j - \beta_0 - \beta_1 x_{j-1}}{\theta \sqrt{a_0 + \alpha_1 x_{j-1}^2}}.$$

Дифференцируя по неизвестному параметру  $\beta_1$  и считая  $\beta_0 = 0$ , получаем уравнение для  $\beta_1$ :

$$\frac{1}{\theta} \sum_{j=1}^n \frac{x_{j-1}}{\sqrt{a_0 + \alpha_1 x_{j-1}^2}} - \sum_{j=1}^n \frac{x_{j-1}}{x_j - \beta_1 x_{j-1}} = 0.$$

Методом перебора получено два решения:  $\hat{\beta}_{11} = 0,52423$  и второе решение  $\hat{\beta}_{12} = 0,2271$ . Для этих оценок параметров получены следующие моделированные траектории при следующих значениях параметров:  $\beta_0 = 0$ ,  $\beta_{11} = 0.52423$ ,  $k = 2$ ,  $\theta = 0.276$ .

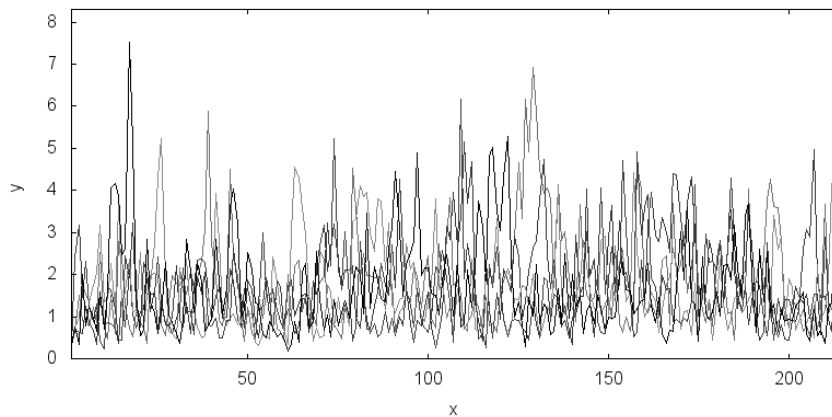


Рис. 7. Пять траекторий:  $k = 2$ ,  $\hat{\beta}_{11} = 0,52423$ ,  $\theta = 0,276$

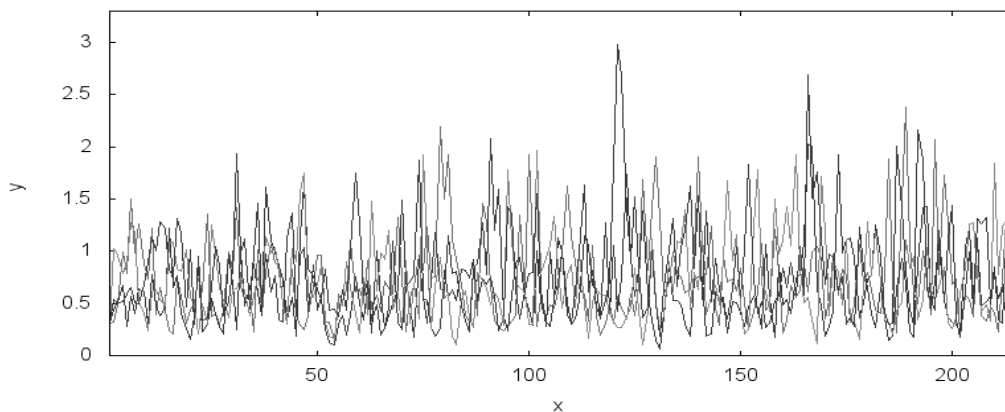


Рис. 8. Пять траекторий  $k = 2$ ,  $\theta = 0,276$   $\hat{\beta}_{12} = 0,227$ ,

Для дальнейшего анализа, основанного на моделировании траекторий, выбираем траектории, моделированные с параметрами  $k = 2$ ,  $\theta = 0.276$ ,  $\hat{\beta}_{12} = 0.227$ .

При  $\hat{\beta}_{11} = 0,52423$  моделированные траектории принимают слишком большие значения, не характерные для анализируемой реализации, изображённой на рис. 3.

**Вероятностное прогнозирование.** Рассмотрим вопрос о вероятностном прогнозировании исследуемой характеристики – содержание железа в воде реки Исеть (выше на 5 км от города Екатеринбурга). Будем использовать метод математического моделирования. Для этого моделируем 1000 траекторий, исходящих из последнего значения реализации, равного 0.312 (рис. 9).

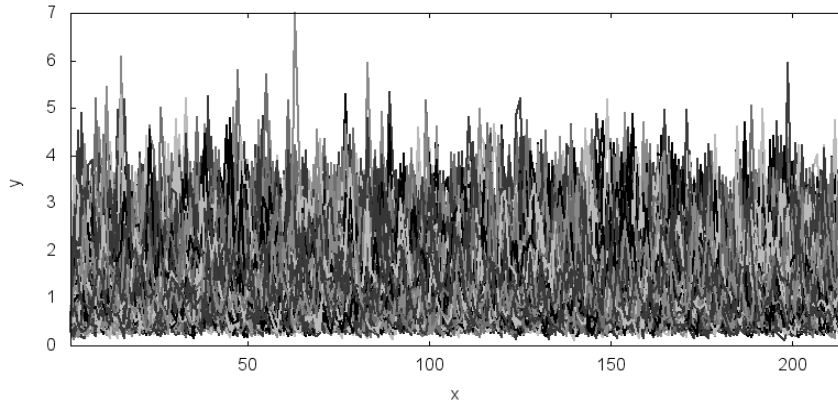


Рис. 9. 1000 моделированных траекторий длиной 213

Чтобы установить границы изменения наиболее вероятного ожидаемого значения содержания железа в воде, моделируем 1000 траекторий на 1 шаг и рассмотрим проекцию моделированных траекторий на вертикальное сечение, соответствующее первому шагу. Такие усеченные траектории, так называемый “факел” изображен на рис. 10.

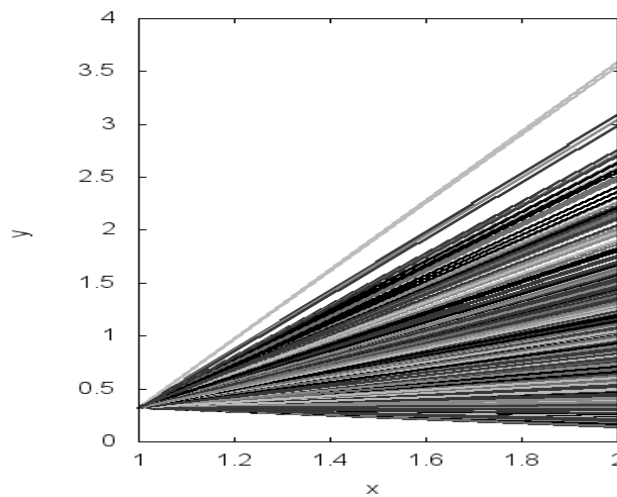


Рис. 10. Ожидаемые значения исследуемой характеристики на 1 шаге по 1000 траекториям

Траектория, соответствующая значениям в точках сечения при  $x = 2$  на рис. 10, приведена на рис. 11.

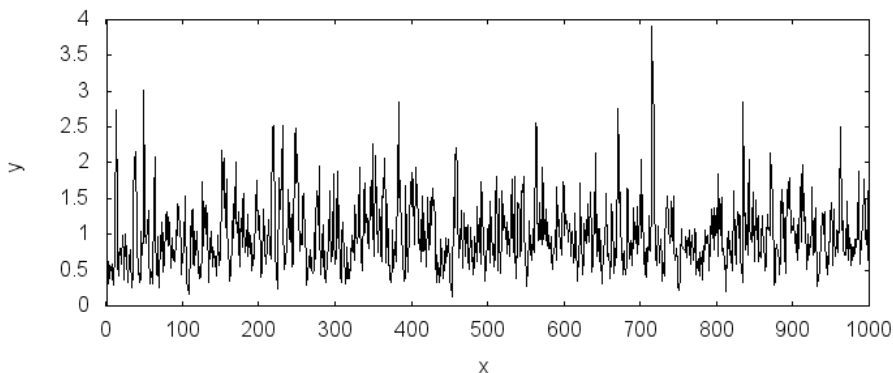


Рис. 11. График последовательности, построенный по 1000 точкам сечения 1-го шага

На рис. 12 приведена гистограмма – оценка плотности вероятностей всех ожидаемых значений при экстраполяции моделированных значений на 1 шаг.



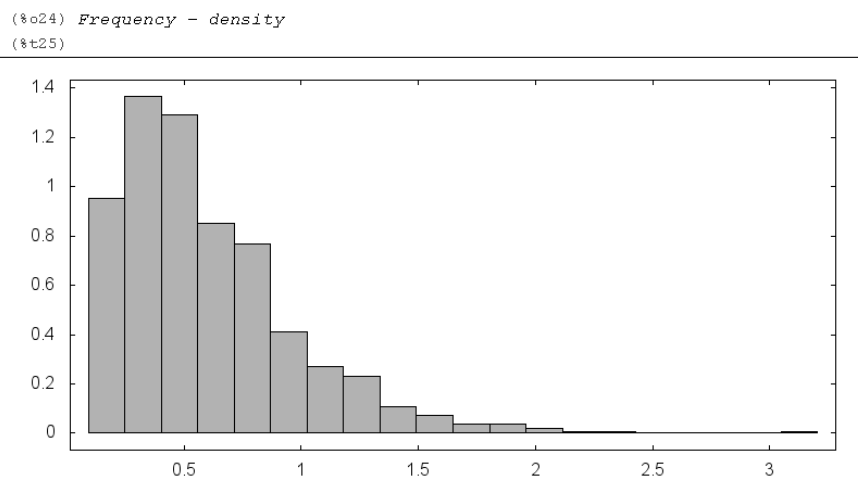


Рис. 12. Гистограмма, построенная по 1000 значений моделированной условной APCN(1) на один шаг вперед

Наиболее вероятными значениями содержания железа в воде реки Исеть при прогнозе на 1 шаг вперед будут значения, попадающие в интервал наибольшего скопления точек (максимум гистограммы), т.е. от 0.277 до 0.474.

**Заключение.** Статья носит методический характер, показано на примере гидрохимической характеристики речной воды, как моделировать и прогнозировать ряды, не являющиеся стационарными в широком смысле. Такие ряды наблюдений широко распространены не только в финансовых потоках [3], но и во многих других областях исследования. В том числе и в гидрологии. В результате проведенного анализа установлены следующие факты:

- 1) содержание железа в воде реки Исеть является случайной величиной, имеющей  $\Gamma$ -распределение, параметры которого оценены на основании имеющейся реализации.
- 2) рассматриваемая реализация не является реализацией стационарного в широком смысле случайного процесса.

Проведен сравнительный анализ аппроксимации рассматриваемого ряда с использованием нормального распределения и  $\Gamma$ -распределения. В обоих случаях с помощью метода максимального правдоподобия получены оценки параметров условно-гауссовской авторегрессии первого порядка. Методом математического моделирования получен интервал наиболее вероятных значений рассматриваемой характеристики (содержания железа в воде реки Исеть) на основании гистограмм.

#### Список литературы

1. Кожевникова И.А., Швейкина В.И. Использование экспоненциальной регрессии для построения модели колебаний уровня Каспийского моря // Известия РАН. Серия географическая, 2012. № 4. С. 62-68.
2. Кожевникова И.А., Швейкина В.И. Моделирование колебаний уровня озера Кинерет // Водные ресурсы, 2014. Том 41. № 6. С. 565-572.
3. Ширяев А.Н. Основы стохастической финансовой математики. Факты. Модели. М.: Т. 1. Фазис, 1998. 499 с.
4. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. М. Изд-во ЛКИ, 2010. 600 с.