Сибирский журнал
индустриальной математики
Октябрь–декабрь, 2010. Том XIII, № 4(44)

УДК 519.233

МОДЕЛИРОВАНИЕ И ИДЕНТИФИКАЦИЯ ПОСЛЕДОВАТЕЛЬНОСТИ ЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН С СИММЕТРИЧНЫМ УСТОЙЧИВЫМ РАСПРЕДЕЛЕНИЕМ*)

А. П. Ковалевский, В. С. Костин, В. Е. Хиценко

Моделируется стационарная случайная последовательность, элементы которой имеют симметричное абсолютно непрерывное устойчивое распределение. Совместное распределение элементов последовательности определяется тремя параметрами: параметром Херста, параметром устойчивого закона, масштабным параметром. Обоснованы и реализованы сильно состоятельные методы оценивания этих параметров. Проведено сравнение различных методов оценивания параметров.

Ключевые слова: устойчивое распределение, фрактальный гауссовский шум.

Введение. Для временных рядов, встречающихся, в частности, в экономических исследованиях, характерна долговременная зависимость элементов. Стандартным средством моделирования этой зависимости служит фрактальное броуновское движение — модель, введенная почти одновременно Колмогоровым и Винером. Название «фрактальное броуновское движение» восходит к работе [1].

Фрактальное броуновское движение является автомодельным процессом в смысле следующего определения. Автомодельными называют процессы $Y(t)$, обладающие свойством масштабной инвариантности: существует $\alpha > 0$ такое, что для любого $\lambda > 0$ выполнено равенство $Y(\lambda t) \stackrel{d}{=} \lambda^{1/\alpha} Y(t)$. Здесь $\stackrel{d}{=}$ означает совпадение распределений.

Однако фрактальное броуновское движение не является удовлетворительной моделью экономических временных рядов, приращения которых, как правило, отличаются от нормальных более «тяжелыми хвостами», на это указывается в работе [2, гл. 3].

Другой автомодельный процесс — процесс Леви, т. е. процесс с независимыми приращениями, распределенными по устойчивому закону, введенный в работах Леви и Хинчина (см. [3]). В этой модели нет зависимости приращений, проявляющей себя в экономических данных.

Обобщением обеих этих моделей в дискретном времени является линейный автомодельный процесс, введенный Такку. Его свойства систематически изучены в монографии [4]. Приращения этого процесса задаются как скользящие средние от независимых случайных величин, распределенных по устойчивому закону.

В данной работе предложен способ генерации последовательности негауссовских случайных величин, имеющих устойчивый абсолютно непрерывный симметричный закон распределения. При этом случайные величины являются

*) Работа выполнена при частичной финансовой поддержке Министерства образования и науки РФ (проект РНП.2.1.1.1379) и Совета по грантам Президента РФ для поддержки ведущих научных школ (проект НШ–1172.2003.1).

© 2010 Ковалевский А. П., Костин В. С., Хиценко В. Е.

зависимыми и имеется возможность регулировать характер зависимости приращений по аналогии с показателем Херста фрактального броуновского движения. Частными случаями таких последовательностей являются фрактальный гауссовский шум и последовательность приращений процесса Леви.

Последовательность частичных сумм элементов таких последовательностей не является автомодельной (за исключением двух упомянутых выше частных случаев), но является полезной моделью временных рядов, встречающихся при экономическом анализе. Модель характеризуется тремя параметрами, для которых обоснованы и реализованы состоятельные процедуры оценивания. Наряду с оценками по методу максимального правдоподобия, трудоемкость вычисления которых очень велика, предложены и обоснованы сильно состоятельные процедуры оценивания параметров, имеющие малую трудоемкость вычисления, линейно зависящую от объема исследуемых данных. Для параметра автомодельного закона используется метод моментов с логарифмической функцией, а для параметра Херста — бинарный знаковый метод. Моделирование показывает, что эти способы дают погрешность оценивания, близкую в среднеквадратическом смысле к погрешности метода максимального правдоподобия. В то же время трудоемкость их гораздо ниже.

1. Моделирование. Изложим алгоритм моделирования последовательности с зависимыми приращениями, распределенными по устойчивому негауссовскому симметричному закону с абсолютно непрерывным распределением. Алгоритм состоит в том, что сначала генерируется фрактальный гауссовский шум, т. е. последовательность приращений фрактального броуновского движения с заданной зависимостью, определяемой показателем Херста. Затем эта нормальная последовательность преобразуется в последовательность с устойчивым симметричным законом распределения, но с более тяжелыми хвостами и с сохранением требуемой зависимости. В ряде работ (начиная с [1]; подробное описание алгоритма и результаты моделирования приводятся в [6]) используется метод скользящего среднего для моделирования фрактального броуновского движения. Этот метод является асимптотическим, т. е. фрактальное броуновское движение появляется лишь в пределе. Оценки скорости сходимости найдены в [7]. Мы используем точный метод моделирования, основанный на разложении ковариационной матрицы на нижнюю и верхнюю треугольные.

Алгоритм состоит из четырех этапов.

1. Генерирование вектора $\mathbf{W} = (W_1, \dots, W_n)^T$ независимых случайных величин со стандартным нормальным распределением.

2. Преобразование вектора $\mathbf{W}$ во фрактальный гауссовский шум с параметром H, $0 < H < 1$, на основе разложения корреляционной матрицы [10]: матрицу

$$R = (r(i-j))_{i,j=1}^n, \tag{1}$$

где $r(k) = \operatorname{corr}(X_i; X_{i+k}) = \frac{1}{2}(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H})$, подвергаем декомпозиции по Холецкому к виду $R = AA^T$, где A — нижняя треугольная матрица. Далее формируем вектор фрактального гауссовского шума по схеме $\mathbf{X} = A\mathbf{W}$. Действительно, корреляционная матрица этого вектора совпадает с R:

$$\mathbf{E}(\mathbf{X}\mathbf{X}^T) = \mathbf{E}(A\mathbf{W}\mathbf{W}^T A^T) = R.$$

3. Преобразование приращений к автомодельному закону по формуле $Y_i = {F_\alpha}^{-1}(\Phi(X_i))$, где Φ — функция распределения стандартного нормального закона, F_α — функция распределения симметричного автомодельного закона.

Вычисление функции распределения F_α симметричного абсолютно непрерывного автомодельного закона рассматривается в п. 3. Параметр α может принимать значения из полуинтервала $(1; 2]$.

Отметим, что Y_i распределены как приращения автомодельного процесса с независимыми приращениями, однако частичные суммы случайных величин Y_i не образуют автомодельного процесса (при одновременном выполнении

неравенств $H \neq 1/2$ и $\alpha \neq 2$). Отсутствие автомодельности можно проверить непосредственно вычислением распределения суммы $Y_1 + Y_2$. Параметрами процесса являются показатель Херста $H \in (0;1)$ и коэффициент автомодельности приращений $\alpha \in (1;2]$. При $H = 1/2$ случайные величины Y_i являются приращениями процесса Леви (в дискретном времени). При $\alpha = 2$ случайные величины Y_i образуют фрактальный гауссовский шум (в дискретном времени).

2. Вычисление функции распределения симметричного устойчивого закона. Следуя [5], будем использовать параметризацию (S) из [4], которая в случае симметричного распределения совпадает с параметризацией (A) Золотарева [8]. В соответствии с этой параметризацией характеристическая функция симметричного абсолютно непрерывного автомодельного закона имеет вид $\varphi(\lambda) = \exp(-\sigma^\alpha|\lambda|^\alpha)$. В частности, при $\alpha = 2$ получаем центрированное нормальное распределение с дисперсией $2\sigma^2$.

Пусть F_α — функция распределения симметричного абсолютно непрерывного устойчивого закона. Соответствующая плотность распределения равна $f_\alpha(x) = \frac{1}{\pi}\int\limits_0^\infty \cos(tx)\exp(-t^\alpha)\,dt$.

Раскладывая косинус в ряд и интегрируя, можно прийти к формуле

$$F_\alpha(x) = \frac{1}{2} + \frac{1}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{\alpha(2n+1)!}\Gamma\left(\frac{2n+1}{\alpha}\right). \tag{2}$$

Здесь Γ — гамма-функция. Ряд сходится для всех $x \in \mathbf{R}$ при $\alpha \in (1;+\infty)$. В области сходимости ряд имеет смысл функции распределения устойчивого закона при $\alpha \leq 2$. Скорость сходимости ряда зависит от значений параметра α и переменной x. Чем меньше α и больше $|x|$, тем больше требуется слагаемых. Так, при $\alpha = 1,1$, $x = 10$ и заданной точности 10^{-6} требуется порядка 10^{10} слагаемых. При $\alpha = 1$ ряд сходится только для $|x| < 1$. Таким образом, практические вычисления по формуле (2) возможны для $\alpha > 1$, не слишком близких к единице. Поэтому есть потребность в алгоритме, позволяющем вычислять F_α с высокой точностью для любых значений $\alpha \in (1;2)$.

Мы будем использовать алгоритм Нолана, изложенный в [5]. Алгоритм основан на том, что интеграл, вычисляющий функцию распределения, приводится с помощью специальной замены к интегралу от ограниченной функции по конечному промежутку. Принимаем для нашей задачи $\beta = 0$ (распределение не содержит дискретной компоненты), $\sigma = 1$ (параметр масштаба равен единице), $\mu = 0$ (смещение нулевое).

Введенную в [5] функцию $V(\theta;\alpha,\beta)$ при $\beta = 0$ обозначим $V_\alpha(\theta)$. Получаем

$$V_\alpha(\theta) = \left(\frac{\cos\theta}{\sin(\alpha\theta)}\right)^{\alpha/(\alpha-1)}\frac{\cos((\alpha-1)\theta)}{\cos\theta}. \tag{3}$$

Переформулируем теорему 1 из [5]: для $x > 0$ имеем

$$\begin{gathered} f_\alpha(x) = \frac{\alpha x^{1/(\alpha-1)}}{\pi(\alpha-1)}\int\limits_0^{\pi/2} V_\alpha(\theta)\exp(-x^{\alpha/(\alpha-1)}V_\alpha(\theta))\,d\theta; \\ F_\alpha(x) = 1 - \frac{1}{\pi}\int\limits_0^{\pi/2}\exp(-x^{\alpha/(\alpha-1)}V_\alpha(\theta))\,d\theta; \\ f_\alpha(-x) = f_\alpha(x); \quad F_\alpha(-x) = 1 - F_\alpha(x). \end{gathered} \tag{4}$$

Кроме того, $f_\alpha(0) = \Gamma(1+1/\alpha)/\pi$, $F_\alpha(0) = 1/2$.

В [5] отмечено, что подынтегральная функция $\exp(-x^{\alpha/(\alpha-1)}V_\alpha(\theta))$ ограничена, непрерывна и монотонно возрастает по θ на отрезке интегрирования. Отметим, что она равна нулю в точке 0 и единице в точке $\pi/2$. Вычисление значений $F_\alpha(x)$ будем проводить численным интегрированием по формуле (4). Исключением являются случаи больших или малых значений x. При больших x подынтегральная функция очень близка к нулю всюду на отрезке, за исключением малой окрестности точки $\pi/2$. При близких к нулю x подынтегральная функция очень близка к единице всюду на отрезке, за исключением малой окрестности точки 0.

В зависимости от требуемой точности $\Delta > 0$ и значения переменной x найдем нижнюю границу интервала интегрирования θ_- для больших значений x. Пусть $\theta_- = \pi/2 - \varepsilon$. При $\varepsilon \to 0$ функция $V_\alpha(\pi/2 - \varepsilon)$ одного порядка малости с $(\sin\varepsilon)^{\alpha/(\alpha-1)-1} \sim \varepsilon^{1/(\alpha-1)}$. Из равенства $\exp(-x^{\alpha/(\alpha-1)}\varepsilon^{1/(\alpha-1)}) = \Delta$ получаем

$$\varepsilon = (-\ln\Delta)^{\alpha-1}x^{-\alpha}, \quad \theta_- = \pi/2 - (-\ln\Delta)^{\alpha-1}x^{-\alpha}.$$

Этот интервал будем использовать при $\varepsilon < (-\ln\Delta)^{-1}$, т. е. при $x > -\ln\Delta$.

Найдем верхнюю границу интервала интегрирования θ_+ для малых значений x в зависимости от $\Delta > 0$ и x. Пусть $\theta_- = \pi/2 - \varepsilon$. При $\theta \to 0$ функция $V_\alpha(\theta)$ эквивалентна $(\alpha\theta)^{-\alpha/(\alpha-1)} \sim \varepsilon^{1/(\alpha-1)}$. Из эквивалентности $\exp(-x^{\alpha/(\alpha-1)}(\alpha\theta)^{-\alpha/(\alpha-1)}) \sim 1 - \Delta$ получаем $\theta \sim x/(\alpha\Delta^{(\alpha-1)/\alpha})$.

Положим θ_+ равным правой части: $\theta_+ = x/(\alpha\Delta^{(\alpha-1)/\alpha})$. Этот интервал будем использовать при $\theta_+ < 0,1$, т. е. при $x < 0,1\alpha\Delta^{(\alpha-1)}/\alpha$.

Итак, получаем следующий алгоритм. Для данных $\alpha \in (1;2)$, $\Delta > 0$ определяем интервал интегрирования:

если $x > -\ln\Delta$, то $\theta_- = \pi/2 - (-\ln\Delta)^{\alpha-1}x^{-\alpha}$, иначе $\theta_- = 0$;

если $x < 0,1\alpha\Delta^{(\alpha-1)/\alpha}$, то $\theta_+ = x/(\alpha\Delta^{(\alpha-1)/\alpha})$, иначе $\theta_+ = \pi/2$.

На интервале $(0;\theta_-)$ принимаем подынтегральную функцию равной нулю;

на интервале $(\theta_+;\pi/2)$ принимаем подынтегральную функцию равной единице;

на интервале $(\theta_-;\theta_+)$ вычисляем интеграл на основе квадратурной формулы Гаусса — Кронрода с 61 точкой [9, § 12.9]; для отрицательных значений x используем формулу $F_\alpha(-x) = 1 - F_\alpha(x)$.

Сравнение результатов вычислений с результатами работы алгоритма Нолана показывает, что для $\alpha \in [1,1;2]$ и $x \in [-10;10]$ погрешность не превосходит 10^{-14}.

3. Оценивание параметров. Оценивание параметров методом максимального правдоподобия по выборке $\mathbf{Y} = (Y_1, \ldots, Y_n)$ состоит в максимизации плотности совместного распределения

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{n/2}\sigma^n\sqrt{\det R}}\exp\left(-\frac{1}{2\sigma^2}\mathbf{X}^T R^{-1}\mathbf{X}\right), \quad \mathbf{X} = (X_1, \ldots, X_n), \qquad (5)$$
$$X_i = \Phi^{-1}(F_\alpha(Y_i)).$$

Максимизацию нужно проводить по совокупности параметров α, H, σ. Однако такая процедура оказывается вычислительно сложной, и для выборок большого объема ее реализовать не удается. Поэтому будем использовать приближенный алгоритм: сначала найдем оценку параметра α в начальном приближении, уточним ее, затем найдем оценки параметров σ и H.

В качестве начального приближения оценки параметра α будем использовать оценку по «хвосту» распределения. Эта оценка основана на том, что левый и правый «хвосты» симметричного устойчивого распределения с параметром α асимптотически пропорциональны $x^{-\alpha}$. Следовательно, для достаточно больших h плотность распределения случайной величины $Y_i\mathbf{I}\{|Y_i| \geq h\}$ приближенно равна $f(x) = K|x|^{-\alpha-1}\mathbf{I}\{|x| \geq h\}$.

Найдем константу K из условий нормировки

$$2\int\limits_h^\infty f(x)\,dx = \frac{2Kh^{-\alpha}}{\alpha}, \quad K = \alpha h^\alpha/2,$$

и оценку параметра α

$$\ln f(x) = \ln\alpha + \alpha\ln h - \ln 2 - (\alpha+1)\mathbf{I}\{|x|\ge h\}\ln|x|,$$
$$\frac{\partial}{\partial\alpha}\ln f(x) = \frac{1}{\alpha} + \ln h - \mathbf{I}\{|x|\ge h\}\ln|x|,$$
$$\sum_{i=1}^n\left(\frac{1}{\alpha} + \ln h - \mathbf{I}\{|Y_i|\ge h\}\ln|Y_i|\right) = 0, \quad \hat\alpha = \frac{1}{\overline{\mathbf{I}\{|Y|\ge h\}\ln|Y|} - \ln h}.$$

Проведенные расчеты показывают, что к наиболее точным результатам приводит выбор в качестве h 15-процентной точки, т. е. числа, которое по модулю превосходят 15 % элементов выборки.

Теорема 1. *Оценка $\tilde\alpha$, максимизирующая*

$$\sum_{i=1}^n \ln f_\alpha(Y_i), \tag{6}$$

является сильно состоятельной оценкой параметра α.

Доказательство. Докажем существование математического ожидания $\mathbf{E}\ln f_\alpha(Y_1) = 2\int\limits_0^\infty f_\alpha(x)\ln f_\alpha(x)\,dx$.

Так как $f_\alpha(+0) = f_\alpha(0) = \frac{1}{\pi}\Gamma(1+1/\alpha) < \infty$, то $\int\limits_0^1 f_\alpha(x)\ln f_\alpha(x)\,dx$ сходится. Так как $f_\alpha(x) \sim Cx^{-\alpha-1}$ при $x\to+\infty$, то $\int\limits_1^\infty f_\alpha(x)\ln f_\alpha(x)\,dx$ сходится. Таким образом, $\mathbf{E}\ln f_\alpha(Y_1)$ существует.

Случайные величины $\ln f_\alpha(Y_i)$ — это функции от Z_i, компонент фрактального гауссовского шума. Так как последовательность $\{Z_i\}$ стационарна и ее корреляционная функция стремится к нулю с ростом лага, то для последовательности $\{\ln f_\alpha(Y_i)\}$ выполнен усиленный закон больших чисел:

$$\frac{1}{n}\sum_{i=1}^n \ln f_\alpha(Y_i) \to \mathbf{E}\ln f_\alpha(Y_1) \tag{7}$$

с вероятностью 1 согласно эргодической теореме Биркгофа (см., например, [12, с. 308]).

В случае $H = 1/2$ оценка $\tilde\alpha$ является оценкой максимального правдоподобия и ее сильная состоятельность следует из общих теорем об оценках максимального правдоподобия в силу существования $\mathbf{E}\ln f_\alpha(Y_1)$, а также ограниченности $f_\alpha(x)$ и ее непрерывности по параметру α при $\alpha\in[1;2]$ (см. [13, с. 122, следствие 3]).

В случае $H \ne 1/2$ оценка $\tilde\alpha$ не является оценкой максимального правдоподобия, но она отыскивается максимизацией той же суммы (6), что и в случае независимых компонент, и в силу сходимости (7) является сильно состоятельной. Доказательство завершено.

Для оценивания параметров σ^2 и H преобразуем данные к нормальному закону по формуле (5). В работе [10] проведено исследование методов оценивания параметра H фрактального броуновского движения и показано преимущество оценки максимального правдоподобия.

Совместная плотность распределения n последовательных приращений ФБД, представленных вектором значений $\mathbf{X} = (X_1, \ldots, X_n)$, равна

$$f(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\sigma^n(\det R)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{x}^T R^{-1}\mathbf{x}\right).$$

В качестве логарифмической функции правдоподобия здесь выступает логарифм этой плотности, взятой в точке $\mathbf{X}$:

$$L(\mathbf{X}; H, \sigma) = -\ln(2\pi)^{n/2} - n\ln\sigma - \frac{1}{2}\ln\det R - \frac{1}{2\sigma^2}\mathbf{X}^T R^{-1}\mathbf{X}.$$

Из условия $\dfrac{\partial L}{\partial \sigma} = 0$ находим оценку максимального правдоподобия $\hat{\sigma}^2 = \dfrac{1}{n}\mathbf{X}^T R^{-1}\mathbf{X}$ и, подставляя ее в предыдущее равенство, получаем

$$L(\mathbf{X};\ H) = -\ln(2\pi)^{n/2} - \frac{n}{2}\ln\mathbf{X}^T R^{-1}\mathbf{X} - \frac{1}{2}\ln\det R - \frac{n}{2}.$$

Итак, нам достаточно найти минимум функции $l(\mathbf{X}; H) = n\ln\mathbf{X}^T R^{-1}\mathbf{X} + \ln\det R$, т. е. оценка максимального правдоподобия параметра H равна

$$\widehat{H} = \arg\max_{0<H<1} l(\mathbf{X}; H).$$

Эту оценку отыскиваем методом золотого сечения (см., например, [16, п. 5.5]). В качестве значений на первом шаге выбираем точки $u_-^{(1)} = (3-\sqrt{5})/2$, $u_+^{(1)} = (\sqrt{5}-1)/2$, образующие золотые сечения интервала $(0;1)$. Значения на втором шаге — золотые сечения интервала, выбираемого в зависимости от знака разности $l\big(\mathbf{X}; u_-^{(1)}\big) - l\big(\mathbf{X}; u_+^{(1)}\big)$.

При поиске оценки параметра H обращение теплицевой симметричной матрицы R реализовывалось по экономичному алгоритму, предложенному в [11, с. 142–160], в ходе реализации которого одновременно вычисляется и детерминант.

4. Бинарный знаковый метод оценивания параметра H. В качестве альтернативы трудоемкому методу оценивания, изложенному в п. 3, рассмотрим следующий бинарный знаковый метод оценивания параметра Херста. Вариант его модификации, приведенный в [14], менее удачен в силу больших положительных корреляций слагаемых.

Знаковый метод [15] оценивания корреляционной функции гауссовской стационарной последовательности с нулевым математическим ожиданием основан на подсчете частоты перемены знака элементами последовательности. Вычислительная сложность этого алгоритма имеет порядок n в отличие от модифицированного алгоритма максимального правдоподобия, имеющего сложность порядка n^2.

Нам потребуется следующая

Лемма 1 [15, с. 236]. *Если (ξ, η) — двумерная нормальная случайная величина с нулевым вектором математического ожидания, то*

$$\mathbf{P}\{\xi\eta < 0\} = \frac{1}{\pi}\arccos\rho,$$

где ρ — коэффициент корреляции случайных величин ξ и η.

Для фрактального гауссовского шума коэффициент корреляции между соседними случайными величинами равен $r(1) = 2^{2H-1} - 1$. Поэтому вероятность того, что последовательные случайные величины имеют противоположные знаки, согласно лемме 1 равна

$$\mathbf{P}\{X_i X_{i+1} < 0\} = \frac{1}{\pi} \arccos(2^{2H-1} - 1). \tag{8}$$

Но и последовательные суммы k случайных величин обладают тем же свойством. Обозначим через $S_{k,i}$ сумму k случайных величин начиная с номера $i+1$:

$$S_{k,i} = \sum_{j=i+1}^{i+k} X_j = S_{k+i} - S_i.$$

Тогда $\mathbf{D} \sum\limits_{j=1}^{n} S_{k,k(j-1)} = \mathbf{D} S_{kn} = \sigma^2 k^{2H} n^{2H}$ и

$$\mathbf{E} S_{k,0} S_{k,kj} = \frac{\sigma^2 k^{2H}}{2} (|j+1|^{2H} + |j-1|^{2H} - 2|j|^{2H}).$$

Поэтому коэффициент корреляции и вероятность иметь противоположные знаки для $S_{k,0}$ и $S_{k,k}$ такие же, как для X_i и X_{i+1}. В силу стационарности то же верно для $S_{k,i}$ и $S_{k,k+i}$ при произвольном $i > 0$.

Таким образом, появляется возможность улучшить оценку параметра H, используя агрерирование случайной последовательности, т. е. суммируя индикаторы перемены знака блоками равной длины, полученными суммированием соседних элементов последовательности.

Введем в рассмотрение бинарный знаковый алгоритм оценивания параметра H, при котором исходные значения суммируются блоками длины 2^k (см. рис. 1).

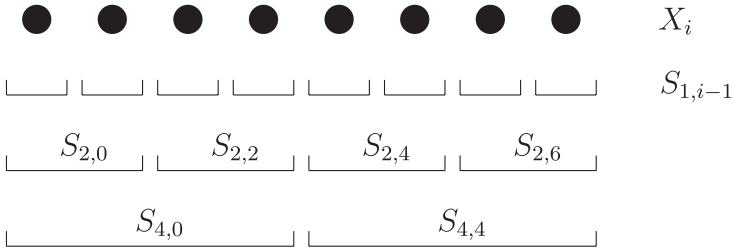

Рис. 1

Для упрощения изложения будем предполагать, что объем данных n является целой степенью числа 2, т. е. $\log_2 n$ — целое число.

Рассмотрим статистику

$$U_n = \sum_{k=0}^{\log_2 n - 1} \sum_{j=0}^{n2^{-k}-2} \mathbf{I}\{S_{2^k, 2^k j} S_{2^k, 2^k(j+1)} < 0\}.$$

В качестве оценки вероятности перемены знака используется частота перемены знака, подсчитанная на основании статистики U_n:

$$p_n^* = \frac{U_n}{\sum\limits_{k=0}^{\log_2 n - 1} (n2^{-k} - 1)} = \frac{U_n}{2n - 2 - \log_2 n}.$$

Оценка H_n^*, вычисляемая на основании статистики p_n^*, имеет вид $H_n^* = \frac{1}{2}\big(1 + \log_2\big(1 + \cos\big(\pi p_n^*\big)\big)\big)$.

5. Вычисление ковариаций при нулевой гипотезе. Здесь предполагается, что выполнена нулевая гипотеза $H = 1/2$, т. е. исходные случайные величины независимы: образуют выборку из нормального распределения с нулевым математическим ожиданием.

Введем следующее обозначение: $G_{k,k'}$ — ковариация индикаторов перемены знака парами блоков из k и k' слагаемых, середины которых совпадают. Введя обозначение $j' = k + j - k'$, определим $G_{k,k'}$ формально как

$$\begin{aligned} G_{k,k'} = \operatorname{cov}(J_{k,j}, J_{k',j'}) &= \mathbf{P}\{S_{k,j}S_{k,j+k} < 0, S_{k',j'}S_{k',j'+k'} < 0\} \\ &\quad - \mathbf{P}\{S_{k,j}S_{k,j+k} < 0\}\mathbf{P}\{S_{k',j'}S_{k',j'+k'} < 0\} \\ &= \mathbf{P}\{S_{k,j}S_{k,j+k} < 0, S_{k',j'}S_{k',j'+k'} < 0\} - 1/4. \end{aligned}$$

Здесь $J_{k,j} = \mathbf{I}\{S_{k,j}S_{k,j+k} < 0\}$ — индикатор перемены знака последовательными суммами k слагаемых, первое из которых имеет номер $j+1$. Рис. 2 поясняет введенное обозначение.

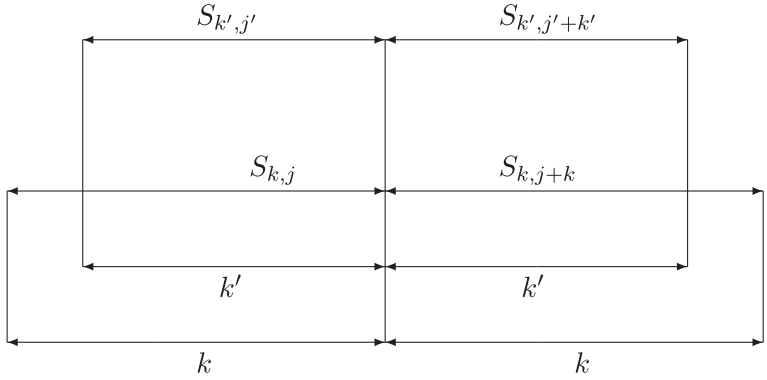

Рис. 2

Отметим, что по определению $G_{k,k'} = G_{k',k}$. Вычислим коэффициенты $G_{k,k'}$ в случае, когда $k' \le k$.

Теорема 2. *Пусть* $0 < k' \le k$. *Тогда* $G_{k,k'} = \left(\frac{1}{\pi}\arccos\sqrt{k'/k} - 1/2\right)^2$.

Доказательство. Будем использовать обозначение $j' = j + k - k'$. С учетом симметрии

$$\begin{aligned} G_{k,k'} &= 2\mathbf{P}\{S_{k,j} < 0, S_{k,j+k} > 0, S_{k',j'} < 0, S_{k',j'+k'} > 0\} \\ &\quad + 2\mathbf{P}\{S_{k,j} > 0, S_{k,j+k} < 0, S_{k',j'} < 0, S_{k',j'+k'} > 0\} - 1/4. \end{aligned}$$

В силу независимости левой и правой частей

$$\begin{aligned} G_{k,k'} &= 2\mathbf{P}\{S_{k,j} < 0, S_{k',j'} < 0\}\mathbf{P}\{S_{k,j+k} > 0, S_{k',j'+k'} > 0\} \\ &\quad + 2\mathbf{P}\{S_{k,j} > 0, S_{k',j'} < 0\}\mathbf{P}\{S_{k,j+k} < 0, S_{k',j'+k'} > 0\} - 1/4 \end{aligned}$$

и в силу симметрии

$$\begin{aligned} G_{k,k'} &= 2\mathbf{P}\{S_{k,j} < 0\}\mathbf{P}\{S_{k,j}S_{k',j'} > 0 \mid S_{k,j} < 0\}\mathbf{P}\{S_{k,j+k} > 0\} \\ &\times \mathbf{P}\{S_{k,j+k}S_{k',j'+k'} > 0 \mid S_{k,j+k} > 0\} + 2\mathbf{P}\{S_{k,j} > 0\}\mathbf{P}\{S_{k,j}S_{k',j'} < 0 \mid S_{k,j} > 0\} \\ &\times \mathbf{P}\{S_{k,j+k} < 0\}\mathbf{P}\{S_{k,j+k}S_{k',j'+k'} < 0 \mid S_{k,j+k} < 0\} - 1/4 \\ &= \frac{1}{2}(1 - \mathbf{P}\{S_{k,j}S_{k',j'} < 0\})^2 + \frac{1}{2}(\mathbf{P}\{S_{k,j}S_{k',j'} < 0\})^2 - 1/4. \end{aligned}$$

Обозначим $p = \mathbf{P}\{S_{k,j}S_{k',j'} < 0\}$. Тогда $G_{k,k'} = \frac{1}{2}((1-p)^2 + p^2) - 1/4 = (p - 1/2)^2$. Так как суммы $S_{k,j}$ и $S_{k',j'}$ имеют k' общих слагаемых, то их коэффициент корреляции $r = \sqrt{k'/k}$. Используя лемму 1, получаем утверждение теоремы. Теорема доказана.

Теорема 3. *Если $H = 1/2$, то при $n \to \infty$ имеет место слабая сходимость последовательности случайных величин $\sqrt{n}\left(p_n^* - 1/2\right)$ к нормальному закону с нулевым математическим ожиданием и дисперсией*:

$$\sigma^2 = \frac{1}{8} + \sum_{s=1}^{\infty} 2^{-s} \left(\frac{1}{\pi} \arccos 2^{-s/2} - \frac{1}{2}\right)^2 \approx 0,1654.$$

Доказательству теоремы предпошлем следующую лемму. Она формализует утверждение о том, что оценка, являющаяся в некотором смысле предельной для последовательности асимптотически нормальных оценок, обладает свойством асимптотической нормальности.

Лемма 2. *Пусть выполнены следующие условия*:

1) *для любого $K < \infty$ оценка $\theta_n^*(K)$ является асимптотически нормальной оценкой параметра θ с коэффициентом σ_K*;

2) *имеет место сходимость $\sigma_K \to \sigma$ при $K \to \infty$, $0 < \sigma < \infty$*;

3) *для любого $\varepsilon > 0$ выполнено $\lim\limits_{K\to\infty} \lim\limits_{n\to\infty} \mathbf{P}\{\sqrt{n}|\theta_n^*(K) - \theta_n^*| \geq \varepsilon\} = 0$. Тогда θ^* — асимптотически нормальная оценка параметра θ с коэффициентом σ.*

Доказательство. Для любого $\varepsilon > 0$ выполнено

$$\mathbf{P}\{\sqrt{n}(\theta_n^* - \theta) < t\} \geq \mathbf{P}\left\{\sqrt{n}(\theta_n^*(K) - \theta) < t - \varepsilon, \sqrt{n}\left|\theta_n^* - \theta_n^*(K)\right| < \varepsilon\right\}.$$

Так как для любых событий A и B выполнено $\mathbf{P}(AB) \geq \mathbf{P}(A) - \mathbf{P}(\bar{B})$, то, переходя к пределу, получаем

$$\lim_{n\to\infty} \mathbf{P}\left\{\sqrt{n}\left(\theta_n^* - \theta\right) < t\right\}$$
$$\geq \lim_{n\to\infty} \mathbf{P}\left\{\sqrt{n}\left(\theta_n^*(K) - \theta\right) < t - \varepsilon)\right\} - \lim_{n\to\infty} \mathbf{P}\left\{\sqrt{n}\left|\theta_n^* - \theta_n^*(K)\right| \geq \varepsilon\right\}.$$

Согласно условию 1) леммы 2 первый из пределов правой части равен $\Phi(\sigma_K(t - \varepsilon))$, где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2}\, dt$ — функция Лапласа.

Согласно условию 3) леммы 2 второй из пределов правой части равен $\delta_K > 0$, где δ_K можно выбором K сделать сколь угодно малым. Итак,

$$\lim_{n\to\infty} \mathbf{P}\left\{\sqrt{n}\left(\theta_n^* - \theta\right) < t\right\} \geq \Phi(\sigma_K(t - \varepsilon)) - \delta_K.$$

Аналогично, $\lim\limits_{n\to\infty} \mathbf{P}\left\{\sqrt{n}\left(\theta_n^* - \theta\right) < t\right\} \leq \Phi(\sigma_K(t + \varepsilon)) + \delta_K$.

Так как $\sigma_K \to \sigma$, $\delta_K \to 0$ при $K \to \infty$, функция Φ непрерывна, а константы $\varepsilon > 0$ и $K < \infty$ выбираются произвольно, то $\lim\limits_{n\to\infty} \mathbf{P}\left\{\sqrt{n}\left(\theta_n^* - \theta\right) < t\right\} = \Phi(\sigma t)$. Лемма доказана.

Следствие 1. *Если в условиях леммы 2 заменить условие 3) на условие $\lim\limits_{K\to\infty} \lim\limits_{n\to\infty} n\mathbf{E}\left(\theta_n^*(K) - \theta_n^*\right)^2 = 0$, то утверждение леммы сохранится.*

Доказательство очевидно из неравенства Чебышева.

Доказательство теоремы. 1) Вычислим $\mathbf{D}U_n$. Из определения получаем

$$\mathbf{D}U_n = \sum_{m=0}^{\log_2 n - 1} \mathbf{D}J_{2^m,0}(n2^{-m} - 1) + 2 \sum_{m=1}^{\log_2 n - 1} \sum_{m'=0}^{m-1} G_{2^{m'},2^m} \cdot (n2^{-m} - 1)$$
$$= \frac{1}{4}(2n - 2 - \log_2 n) + 2 \sum_{m=1}^{\log_2 n - 1} \sum_{m'=0}^{m-1} \left(\frac{1}{\pi} \arccos 2^{(m'-m)/2} - \frac{1}{2}\right)^2 (n2^{-m} - 1).$$

При $n \to \infty$ имеет место асимптотика

$$\mathbf{D}U_n \sim \frac{n}{2} + 2n \sum_{m=1}^{\infty} 2^{-m} \sum_{m'=0}^{m-1} \left(\frac{1}{\pi} \arccos 2^{(m'-m)/2} - \frac{1}{2} \right)^2 .$$

Сделаем замену $s = m - m'$ и поменяем порядок суммирования:

$$\mathbf{D}U_n \sim \frac{n}{2} + 2n \sum_{s=1}^{\infty} \sum_{m=s}^{\infty} 2^{-m} \left(\frac{1}{\pi} \arccos 2^{-s/2} - \frac{1}{2} \right)^2$$
$$= \frac{n}{2} + 4n \sum_{s=1}^{\infty} 2^{-s} \left(\frac{1}{\pi} \arccos 2^{-s/2} - \frac{1}{2} \right)^2 .$$

Следовательно,

$$\mathbf{D}p_n^* \sim \frac{1}{4n^2} \left(\frac{n}{2} + 4n \sum_{s=1}^{\infty} 2^{-s} \left(\frac{1}{\pi} \arccos 2^{-s/2} - \frac{1}{2} \right)^2 \right)$$
$$= \frac{1}{n} \left(\frac{1}{8} + \sum_{s=1}^{\infty} 2^{-s} \left(\frac{1}{\pi} \arccos 2^{-s/2} - \frac{1}{2} \right)^2 \right).$$

Докажем сходимость по распределению. Обозначим

$$U_n(K) = \sum_{m=0}^{K-1} \sum_{j=0}^{n2^{-m}-2} \mathbf{I}\{S_{2^m,2^m j} S_{2^m,2^m(j+1)} < 0\}.$$

Тогда

$$\mathbf{D}U_n(K) = \sum_{m=0}^{K-1} \mathbf{D}J_{2^m,0}(n2^{-m} - 1) + 2 \sum_{m=1}^{K-1} \sum_{m'=0}^{m-1} G_{2^{m'},2^m}(n2^{-m} - 1)$$
$$= \frac{1}{4} \left(\frac{n(1-2^{-K})}{1-1/2} - K \right) + 2 \sum_{m=1}^{K-1} \sum_{m'=0}^{m-1} (n2^{-m} - 1) \left(\frac{1}{\pi} \arccos 2^{(m'-m)/2} - \frac{1}{2} \right)^2 .$$

При $n \to \infty$ имеет место $\mathbf{D}U_n(K) \sim Cn$, $C > 0$.

Обозначим

$$p_n^*(K) = \frac{U_n(K)}{\sum\limits_{m=0}^{K-1} (n2^{-m} - 1)} = \frac{U_n(K)}{2n(1-2^{-K}) - K}.$$

Проверим условия леммы 2. Центральная предельная теорема для $U_n(K)$ имеет место в силу того, что $U_n(K)$ представима в виде суммы элементов стационарной последовательности 1-зависимых случайных величин (числа перемен знаков в последовательных блоках длины 2^{K+1}), при этом $\mathbf{D}U_n(K) \to \infty$. Следовательно, $\lim\limits_{n\to\infty} \mathbf{P}\{\sqrt{n}(p_n^* - 1/2) < t\} = \Phi(\sigma_K t)$.

Теперь вычислим предельную дисперсию

$$\sigma_K^2 = \lim_{n\to\infty} n\mathbf{D}p_n^*(K)$$
$$= \frac{1}{4(1-2^{-K})^2} \left(\frac{1}{2}(1-2^{-K}) + 2 \sum_{m=1}^{K} \sum_{m'=0}^{m-1} 2^{-m} \left(\frac{1}{\pi} \arccos 2^{(m'-m)/2} - \frac{1}{2} \right)^2 \right).$$

Сделаем замену $s = m - m'$ и поменяем порядок суммирования:

$$\sigma_K^2 = \frac{1}{4(1-2^{-K})^2}\left(\frac{1}{2}(1-2^{-K}) + 2\sum_{s=1}^{K}\sum_{m=s}^{K} 2^{-m}\left(\frac{1}{\pi}\arccos 2^{s/2} - \frac{1}{2}\right)^2\right)$$
$$= \frac{1}{4(1-2^{-K})^2}\left(\frac{1}{2}(1-2^{-K}) + 2\sum_{s=1}^{K}(2^{-s+1}-2^{-K})\left(\frac{1}{\pi}\arccos 2^{s/2} - \frac{1}{2}\right)^2\right).$$

Переходя к пределу при $K \to \infty$, устанавливаем сходимость $\sigma_K \to \sigma$.

Разность

$$p_n^* - p_n^*(K) = \frac{U_n}{2n-2-\log_2 n} - \frac{U_n(K)}{2n(1-2^{-K})-K}$$

имеет нулевое математическое ожидание.

Обозначим $\Delta_n(K) = U_n - U_n(K)$. Тогда

$$p_n^* - p_n^*(K)$$
$$= \frac{\Delta_n(K)}{2n-2-\log_2 n} - \frac{U_n(K)(2n2^{-K}+K-\log_2 n)}{(2n(1-2^{-K})-K)(2n-2-\log_2 n)} \stackrel{\text{def}}{=} A_n + B_n.$$

Заметим, что

$$\mathbf{E}\big(p_n^* - p_n^*(K)\big)^2 = \mathbf{D}\big(p_n^* - p_n^*(K)\big) \le \mathbf{D}A_n + \mathbf{D}B_n + 2\sqrt{\mathbf{D}A_n\mathbf{D}B_n},$$
$$\mathbf{D}A_n \sim \frac{\mathbf{D}U_n(K)2^{-2K}}{4n^2}, \quad \mathbf{D}B_n \sim \frac{\mathbf{D}\Delta_n(K)}{4n^2}.$$

Так как

$$\Delta_n(K) = \sum_{m=K}^{\log_2 n-1}\sum_{j=0}^{n2^{-m}-2} \mathbf{I}\{S_{2^m,2^m j}S_{2^m,2^m(j+1)} < 0\},$$

то в силу сходимости ряда $\sum\limits_{s=0}^{\infty} G_{1,2^s}$

$$\mathbf{D}\Delta_n(K) = \sum_{m=K}^{\log_2 n-1}\mathbf{D}J_{2^m,0}(n2^{-m}-1) + 2\sum_{m=K}^{\log_2 n-1}\sum_{m'=0}^{m-1} G_{2^{m'},2^m}(n2^{-m}-1)$$
$$= \sum_{m=K}^{\log_2 n-1}(n2^{-m}-1)\left(\frac{1}{4} + 2\sum_{m'=0}^{m-1} G_{2^{m'},2^m}\right) \le C_0\sum_{m=K}^{\log_2 n-1}(n2^{-m}-1) \le 2^{1-K}C_0 n.$$

Переходя к пределу при $n \to \infty$, а затем при $K \to \infty$, обосновываем выполнение условия 3) леммы 2. Доказательство завершено.

Так как согласно (8) вероятность перемены знака p и параметр H связаны соотношением $H = \frac{1}{2}(1 + \log_2(1 + \cos(\pi p)))$, то оценка $\widetilde{H}_n$, вычисляемая на основании статистики p_n^*, имеет вид $\widetilde{H}_n = \frac{1}{2}(1 + \log_2(1 + \cos(\pi p_n^*))$.

Согласно теореме об асимптотической нормальности [13] эта оценка является асимптотически нормальной с коэффициентом $B(H) = |H'(p)|\sigma(H)$. При $p = H = 1/2$ получаем $B = B(1/2) = \frac{\pi}{2\ln 2}\sigma$.

Итак, приходим к следующему утверждению.

Следствие 1. *Если $H = 1/2$, то при $n \to \infty$ имеет место слабая сходимость последовательности случайных величин $\sqrt{n}(\widetilde{H}_n - 1/2)$ к нормальному закону с нулевым математическим ожиданием и дисперсией*

$$B^2 = \frac{\pi^2}{4\ln^2 2}\left(\frac{1}{8} + \sum_{s=1}^{\infty} 2^{-s}\left(\frac{1}{\pi}\arccos 2^{-s/2} - \frac{1}{2}\right)^2\right) \approx 0,8494.$$

При $n = 1024$ получаем стандартное отклонение $B/32 \approx 0,02880$.

Моделирование подтверждает этот результат, а также показывает, что при изменении H в диапазоне от 0,1 до 0,7 дисперсия оценки по бинарному знаковому методу меняется несущественно.

6. Результаты для модельных данных. Алгоритм моделирования и алгоритмы оценивания параметров были реализованы для каждого значения H от 0,1 до 0,9 с шагом 0,1. Шаг по α также равен 0,1, значения от 1,1 до 1,9. Результаты получены усреднением 7218 вычислений для каждого значения α и H. Объем выборки n во всех случаях равнялся 1024.

В табл. 1–3 приведены средние выборочные значения и выборочные среднеквадратические отклонения оценок параметров α и H.

Среднеквадратическое отклонение оценки $\widehat{H}$ меньше при малых H, а оценки $\widetilde{H}$ почти не зависит от параметра. Оценка бинарным знаковым методом имеет отрицательное смещение, возрастающее по модулю с ростом H.

Таблица 1

Результаты оценивания параметра α модифицированным методом максимального правдоподобия

α	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9
$\widehat{\alpha}$	1,101	1,201	1,303	1,402	1,503	1,601	1,700	1,803	1,903
$\widehat{\sigma}_\alpha$	0,023	0,026	0,028	0,029	0,031	0,032	0,032	0,032	0,029

Таблица 2

Результаты оценивания параметра H модифицированным методом максимального правдоподобия

H	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$\widehat{H}$	0,100	0,200	0,300	0,399	0,499	0,599	0,699	0,799	0,898
$\widehat{B}_H$	0,011	0,015	0,017	0,018	0,019	0,020	0,020	0,021	0,021

Таблица 3

Результаты оценивания параметра H бинарным знаковым методом

H	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$\widetilde{H}$	0,100	0,200	0,299	0,399	0,499	0,600	0,699	0,798	0,893
$\widetilde{B}_H$	0,032	0,031	0,031	0,029	0,029	0,029	0,030	0,037	0,045

Сравнивая результаты вычислений, приведенные в табл. 2 и 3, следует отметить, что отношение $\widetilde{B}_H/\widehat{B}_H$ не превосходит 1,5 для $0,5 \le H \le 0,7$. Отметим, что такие значения параметра H типичны для ряда приложений. При этих значениях параметра H для достижения такого же среднеквадратического отклонения оценки модифицированным знаковым методом требуется примерно в 2,25 раза больше наблюдений, чем для оценки модифицированным методом максимального правдоподобия. Использование модифицированного знакового

метода представляется оправданным ввиду его малой вычислительной сложности в ситуации, когда объем наблюдаемых данных велик.

Авторы благодарят рецензента за ряд замечаний, позволивших улучшить изложение.

ЛИТЕРАТУРА

1. *Mandelbrot B. B., Van Ness J. W.* Fractional brownian motions, fractional noises and applications // SIAM Rev. 1968. V. 10. P. 422–437.
2. *Ширяев А. Н.* Основы стохастической финансовой математики. Т. 1. Факты. Модели. М.: Фазис, 1998.
3. *Ибрагимов И. А., Линник Ю. В.* Независимые и стационарно связанные величины. М.: Наука, 1965.
4. *Samorodnitsky G., Taqqu M. S.* Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. New York; London: Chapman & Hall, 1994.
5. *Nolan J. P.* An algorithm for evaluating stable densities in Zolotarev's (M) parameterization // Math. Comput. Model. 1999. V. 29. P. 229–233.
6. *Закревская Н. С., Ковалевский А. П.* Алгоритм идентификации фрактального броуновского движения по разности оценок // Сб. науч. трудов НГТУ. 2004. № 2(36). С. 29–36.
7. *Аркашов Н. С., Борисов И. С.* Гауссовская аппроксимация процессов частных сумм скользящих средних // Сиб. мат. журнал. 2004. Т. 45, № 6. С. 1221–1255.
8. *Золотарев В. М.* Современная теория суммирования независимых случайных величин. М.: Наука, 1986.
9. *Рыжиков Ю.* Вычислительные методы. СПб.: БХВ-Петербург, 2007.
10. *Jennane R., Harba R., Jacquet G.* Analysis methods for fractional brownian motion: theory and comparative results // Traitement du Signal. 1996. V. 13, N 4. P. 289–302.
11. *Воеводин В. В., Тыртышников Е. Е.* Вычислительные процессы с теплицевыми матрицами. М.: Наука, 1987.
12. *Скороход А. В. Элементы теории вероятностей* и случайных процессов. Киев: Вища школа, 1980.
13. *Боровков А. А.* Математическая статистика. Новосибирск: Наука, 1997.
14. *Ковалевский А. П.* Модифицированный знаковый метод тестирования фрактальности гауссовского шума // Пробл. передачи информации. 2008. Т. 44, вып. 1. С. 45–58.
15. *Яглом А. М.* Корреляционная теория стационарных случайных функций. Л.: Гидрометеоиздат, 1981.
16. *Федунец Н. И., Черников Ю. Г.* Методы оптимизации. М.: Горная книга, 2009.

Ковалевский Артем Павлович
Новосибирский государственный технический университет
пр. Карла Маркса, 20 630092 г. Новосибирск
Новосибирский госуниверситет
ул. Пирогова, 2 630090 г. Новосибирск
Костин Виталий Сергеевич
Институт экономики и организации промышленного производства СО РАН
пр. Лаврентьева, 17 630090 г. Новосибирск
Хиценко Владимир Евгеньевич
Новосибирский государственный технический университет
E-mail: pandorra@ngs.ru

Статья поступила 10 декабря 2009 г.
Окончательный вариант 25 апреля 2010 г.