12. Щукин, А. Н. Методика преподавания русского языка как иностранного: учебное пособие для вузов / А. Н. Щукин. – Москва : Высшая школа, 2003. – 334 с.

УДК 519.866.2:004.4

Кузьмина А. В.

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ПРОЦЕССОВ НА ЯЗЫКЕ R

Анна Викентьевна Кузьмина

кандидат физико-математических наук

kuzminaAV@bsu.by

Белорусский государственный университет, Минск, Республика Беларусь

RANDOM PROCESSES SIMULATION USING R LANGUAGE

Anna Vikentevna Kuzmina

Belarusian State University, Minsk, Republic of Belarus

Аннотация. *В статье рассматривается моделирование случайных процессов Леви — процессов, которые в настоящее время в наибольшей мере соответствуют природе эволюции движения цен финансовых активов, на языке R. В силу возможности прямого обращения из R к значениям акций, хранящимся в базе данных Oracle, моделирование процессов Леви на языке R является актуальной задачей. Процессы Леви используются как процессы, описывающие эволюцию логарифмических доходностей финансовых активов, в экспоненциальной модели Леви. В статье предлагаются алгоритмы моделирования некоторых процессов Леви, существенно сокращающие время моделирования указанных процессов по сравнению с ранее известными алгоритмами моделирования процессов Леви.*

Abstract. *The paper discusses random Levy processes simulation using the capabilities of the R language. Levy processes are processes that currently most closely* *correspond to the nature of the evolution of stock price movements. Due to the possibility of direct access from R to stock values stored in the Oracle database, modeling Levy processes in the R language is an urgent task. Levy processes are used as processes describing the evolution of the logarithmic returns of financial assets in the exponential Levy model. The article proposes modeling algorithms for some Levy processes that significantly reduce the modeling time of these processes compared to previously known Levy process modeling algorithms.*

Ключевые слова: *случайные процессы Леви, язык R.*

Keywords: *Levy processes, R language.*

Проблема моделирования процесса эволюции финансовых характеристик (цен акций, индексов) привлекает внимание многих ученых на протяжении более 100 лет. Первая попытка моделирования эволюции цен финансовых инструментов была предпринята Л. Башелье в 1900 году [1]. В настоящее время модели, построенные на основе процессов Леви, наиболее адекватно отражают процесс эволюции финансовых характеристик. К таким моделям относятся экспоненциальная модель Леви.

Язык статистического программирования R обладает функционалом для моделирования базовых вероятностных распределений таких, как нормальное распределение, распределение Пуассона, распределение Стьюдента, экспоненциальное распределение и др. Моделирование распределений, на которых основаны алгоритмы моделирования процессов Леви, является актуальной задачей программирования на языке R. Кроме того, возможно работа напрямую из R с базой данных Oracle. Построение соединения, так называемого «моста», между языком R и базой данных Oracle, в которой хранятся данные о финансовых активах, позволяет построить прогноз динами значений этих активов.

Процессы Леви используются как процессы, описывающие эволюцию логарифмических доходностей финансовых активов, в модели Леви [2]:

$$S_t = S_0 \exp(rt) \exp(X_t), \qquad (1)$$

где $S = \left(S_t\right)_{t\geq 0}$ — процесс цен акции, $S_0 > 0$, $X = \left(X_t\right)_{t\geq 0}$ — процесс Леви, $r > 0$ — процентная ставка.

Процесс CGMY — это один из процессов Леви, который используется для описания эволюции логарифмических доходностей акций в модели Леви (1). Процесс CGMY и его свойства изучены в работе авторов Carr P., Geman H., Madan D. P., Yor M. [3]. Приведем определения CGMY распределения и процесса CGMY.

Определение 1. Распределение F называется CGMY распределением с параметрами $C > 0$, G, $M > 0$, $Y < 2$, если его характеристическая функция имеет вид

$$\phi(u) = \exp(C\Gamma(-Y)((M - iu)^Y - M^Y + (G + iu)^Y - G^Y)),\ u \in (\Omega, \mathcal{F}, P), \tag{2}$$

где $\Gamma(\mathrm{x}) = \int\limits_0^\infty z^{x-1} e^{-z} dz$ — гамма-функция, $x > 0$. Случайную величину ξ с CGMY распределением будем обозначать следующим образом: $\xi \sim CGMY(C, G, M, Y)$.

Определение 2. Случайный процесс $X = \left(X_t\right)_{t\geq 0}$ с параметрами $C > 0$, $G > 0$, $M > 0$, $Y < 2$, заданный на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ со значениями в $\mathbb{R}$, такой, что $X_0 \overset{п.н.}{=} 0$, называется процессом Леви, если выполнены следующие условия:

1. X имеет независимые приращения;
2. X имеет стационарные приращения, которые подчиняются CGMY распределению: для любых $s \geq 0, t \geq 0$

$$X_{s+t} - X_s \sim CGMY(Ct, G, M, Y);$$

3. X обладает свойством стохастической непрерывности: для любых $t \geq 0$ и $\varepsilon > 0$

$$\lim_{s \to t} P(| X_s - X_t | > \varepsilon) = 0.$$

Мера Леви процесса CGMY имеет следующий вид:

$$\nu_{CGMY}(x)=\begin{cases} C\exp(Gx)(-x)^{-1-Y}\,dx, x<0, \\ C\exp(-Mx)x^{-1-Y}\,dx, x>0. \end{cases} \tag{3}$$

Эффективный алгоритм моделирования процесса CGMY основан на моделировании медленно растущего случайного процесса. Введем определения медленно растущего случайного распределения и медленно растущего случайного процесса.

Определение 3. Распределение F называется медленно растущим устойчивым распределением с параметрами $a>0$, $b>0$, $0<k<1$, если его характеристическая функция имеет вид:

$$\phi(u)=\exp(ab-a(b^{1/k}-2iu)^k)\,. \tag{4}$$

Случайную величину η с медленно растущим устойчивым распределением будем обозначать $\eta \sim TS(a,b,k)$.

Определение 4. Случайный процесс $Z=(Z_t)_{t\ge 0}$ с параметрами $a>0$, $b>0$, $0<k<1$, заданный на вероятностном пространстве $(\Omega,\mathcal{F},P)$ со значениями в $\mathbb{R}$, называется медленно растущим устойчивым процессом Леви, если выполнены следующие условия:

1) $Z_0 \overset{п.н.}{=} 0$;

2) Z имеет независимые и стационарные приращения;

3) приращения Z имеют медленно растущие устойчивое распределение, то есть для любых $s\ge 0, t\ge 0$

$$Z_{s+t}-Z_s \sim TS(at,b,k)\,.$$

Мера Леви медленно растущего устойчивого процесса имеет вид:

$$\nu_{TS}(x)=a2^k\frac{k}{\Gamma(1-k)}x^{-k-1}\exp\left(-\frac{1}{2}b^{1/k}x\right)1_{x>0}\,. \tag{5}$$

Теорема 1. [4] Пусть $X=(X_t)_{t\ge 0}-$ процесс CGMY с параметрами $C>0$, $G>0$, $M>0$, $Y<2$, заданный на вероятностном пространстве $(\Omega,\mathcal{F},P)$, с мерой

Леви (3), а $X^1 = \left(X_t^1\right)_{t\ge 0}$ и $X^2 = \left(X_t^2\right)_{t\ge 0}$ – медленно растущие устойчивые независимые процессы Леви с параметрами $C > 0$, $M > 0$, $Y < 2$ и $C > 0$, $G > 0$, $Y < 2$ соответственно такие, что их меры Леви имеют вид

$$\nu_1(x) = \frac{Ce^{-Mx}}{x^{1+Y}} 1_{x>0} \text{ и } \nu_2(x) = \frac{Ce^{-G|x|}}{|x|^{1+Y}} 1_{x<0}. \qquad (6)$$

Тогда CGMY процесс CGMY $X = \left(X_t\right)_{t\ge 0}$ представим как

$$X_t = X_t^1 - X_t^2. \qquad (7)$$

Доказательство теоремы проведено в работе Кузьминой А. В., Труша Н. Н. [4].

Дисперсионный гамма процесс [2] также является процессом Леви, который применяется для описания эволюции логарифмических доходностей акций в модели Леви (1). Приведем определения дисперсионного гамма распределения и дисперсионного гамма процесса.

Определение 5. Случайная величина X, заданная на вероятностном пространстве $(\Omega, \mathcal{F}, \mathrm{P})$ подчиняется дисперсионному гамма распределению с параметрами $\sigma > 0, \nu > 0, \theta \in \mathbb{R}$, если ее характеристическая функция имеет вид

$$\varphi_X^{VG}(u;\sigma,\nu,\theta) = \left(\frac{1}{1 - i\nu\theta u + (\sigma^2\nu/2)u^2}\right)^{\frac{1}{\nu}} \qquad (8)$$

где $\Gamma(x)$, $x \in \mathbb{R}_+$ — гамма-функция.

Случайную величину ϑ с дисперсионным гамма распределением с параметрами $\sigma > 0, \nu > 0, \theta \in \mathbb{R}$ будем обозначать как $\vartheta \sim V(\sigma, \nu, \theta)$.

Определение 6. Случайный процесс $V = \left(V_t\right)_{t\ge 0}$ с параметрами $\sigma > 0, \nu > 0, \theta \in \mathbb{R}$, заданный на вероятностном пространстве $(\Omega, \mathcal{F}, \mathrm{P})$ со значениями в $\mathbb{R}$ такой, что $V_0 \overset{п.н.}{=} 0$, называется дисперсионными гамма процессом, если выполнены следующие условия:

1) V имеет независимые приращения;

2) V имеет стационарные приращения с дисперсионным гамма распределением: для любых $s \geq 0, t \geq 0$

$$V_{t+s} - V_s \overset{d}{=} V_t - V_0 \sim V(\sigma\sqrt{t}, \nu / t, t\theta).$$

3) V обладает свойством стохастической непрерывности.

Предлагаемый алгоритм моделирования дисперсионного гамма процесса основан на моделировании гамма процессов. Приведем определение гамма распределения и гамма-процесса [2].

Определение 7. Случайная величина X подчинятся гамма распределению с параметром формы $a > 0$, и параметром масштаба $b > 0$, если ее характеристическая функция задается как

$$\varphi_X^{\Gamma}(u; a, b) = (1 - iu / b)^{-a} \quad x \in \mathbb{R}, \tag{9}$$

а $\Gamma(x) = \int_0^{\infty} z^{x-1} e^{-z} dz$, $x \in \mathbb{R}_+$ — гамма-функция. Случайную величину γ, подчиняющуюся гамма распределению, будем обозначать следующим образом $\gamma \sim \Gamma(a, b)$.

Определение 8. Случайный процесс $G = (G_t)_{t \geq 0}$ с параметрами $a > 0$, $b > 0$, заданный на вероятностном пространстве $(\Omega, \mathcal{F}, \mathrm{P})$ со значениями в $\mathbb{R}$ такой, что $G_0 \overset{п.н.}{=} 0$, называется гамма-процессом, если выполнены следующие условия:

1) G имеет независимые приращения;

2) G имеет стационарные приращения с гамма-распределением:

$$G_{s+t} - G_s \overset{d}{=} G_t - G_0 \sim \Gamma(at, b);$$

3) G обладает свойством стохастической непрерывности.

Дисперсионный процесс может быть смоделирован как разность двух гамма-процессов.

R — это язык программирования и свободная программная среда для статистической обработки данных и работы с графикой. Язык R содержит функции для моделирования некоторых законов распределения, например, нормального распределения, бета-распределения, гамма-распределения, равномерного распределения, распределения Пуассона и др. Однако, функции моделирования CGMY распределения и CGMY процесса не представлены в языке R. Поэтому проблема моделирования CGMY распределения и CGMY процесса на R является актуальной.

В первую очередь был запрограммирован алгоритм моделирования CGMY процесса, предложенный в работе Poirot J. и P. Tankov [5]. Этот алгоритм моделирует CGMY процесс с параметрами $C>0$, $G>0$, $M>0$, $Y<2$ используя винеровский процесс и случайную замену времени положительным невозрастающим $Y/2$-устойчивым процессом, который называют субординатором:

$$X_t = AS_t + W_{S_t},$$

где $W=(W_t)_{t\geq 0}$ — стандартный винеровский процесс в случайные моменты времени S_t, $S=(S_t)_{t\geq 0}$ — $Y/2$- устойчивый процесс Леви независимый от винеровского процесса, $A=\dfrac{G-M}{2}$.

Предлагаемый алгоритм позволяет смоделировать процесс CGMY как разность двух медленнорастущих независимых случайных процессов и основан на теореме1.

Алгоритм 1. [4] Моделирование процесса CGMY как разности медленнорастущих независимых случайных процессов.

1. Генерируем медленно растущий устойчивый процесс $X^1=(X_t^1)_{t\geq 0}$ с параметрами $C>0$, $M>0$, $0<Y<1$.

2. Генерируем медленно растущий устойчивый процесс $X^2=(X_t^2)_{t\geq 0}$ с параметрами $C>0$, $G>0$, $0<Y<1$.

3. CGMY процесс $X=(X_t)_{0\leq t<T}$ с параметрами $C,G,M>0$, $0<Y<1$ генерируем как

$$X_t = X_t^1 - X_t^2.$$

Алгоритм моделирования медленно растущего устойчивого процесса приведен в работе Poirot J. и P. Tankov [5].

При моделировании процесса CGMY с параметрами $C = 10$, $G = 7$, $M = 5$, $Y = 0.2$, $\varepsilon = 0.0001$ и временным шагом $\Delta t = 0.01$ как винеровского процесса с помощью случайной замены времени $Y/2$-устойчивым субординатором стандартное отклонение составляет 0.0241, а при моделировании этого процесса как разности двух медленно растущих устойчивых процессов при $K = 10000, \Delta t = 0.01$ стандартное отклонение составляет 0.0238. Преимущество способа моделирования процесса CGMY как разности двух медленно растущих устойчивых процессов относительно способа моделирования этого процесса как винеровского процесса с помощью случайной замены времени $Y/2$-устойчивым субординатором заключается в затрачиваемом на моделирование процесса CGMY времени. При моделировании процесса CGMY как разности двух медленно растущих устойчивых случайных процессов затрачивается значительно меньше времени, чем при моделировании этого процесса как винеровского процесса с помощью случайной замены времени. Например, при моделировании процесса CGMY как разности двух медленно растущих устойчивых случайных процессов $X = (X_t)_{0 \le t < T}$ с параметрами $C = 10$, $G = 7$, $M = 5$, $Y = 0.2$ и временным шагом $\Delta t = 0.01$ состоящего из $T = 1000$ значений потребуется 11 сек., а при моделировании такого же процесса CGMY как винеровского процесса с помощью случайной замены времени $Y/2$-устойчивым субординатором — 302 сек.

Язык R не содержит инструменты для моделирования дисперсионного гамма распределения и дисперсионного гамма процесса.

Алгоритм моделирования дисперсионного гамма процесса как разности двух независимых гамма процессов представлен в работе W. Schoutens [2]. Предлагаемый алгоритм моделирования основан на методе суперпозиции и позволяет проводить моделирование дисперсионного гамма процесса используя случайные величины с дисперсионными гамма распределением.

Алгоритм 2. **[6]** Моделирование дисперсионного гамма процесса $V = (V_t)_{t \ge 0}$ с параметрами σ, ν, θ, μ с помощью случайных величин с дисперсионным гамма распределением.

1. Генерируем независимые случайные величины с гамма распределением $\{\xi_k, k \ge 1\}$ с параметрами $a = \Delta t / \nu, b = 1/\nu$ используя функцию **dgamma** языка R

$$\xi_k \sim \Gamma(\Delta t / \nu, 1/\nu)\ \xi_k \sim \Gamma(\Delta t / \nu, 1/\nu),\ k \ge 1.$$

2. Генерируем независимые случайные величины с гамма распределением $\{\eta_k, k \ge 1\}$ как случайные величины с нормальным распределением и $E\eta_k = \theta \Delta t \xi_k$, $D\eta_k = \sigma^2 \Delta t \xi_k$ используя функции **dgamma и rnorm**:

$$\eta_k \sim N(\theta \Delta t \xi_k + \mu \Delta t, \sigma^2 \Delta t \xi_k),\ k \ge 1.$$

3. Генерируем дисперсионный гамма процесс $V = (V_t)_{t \ge 0}$ с параметрами σ, ν, θ, μ как

$$V_0 = 0,\ V_{k\Delta t} = V_{(k-1)\Delta t} + \eta_k,\ k \ge 1.$$

Преимущество предлагаемого способа моделирования дисперсионного гамма процесса по алгоритму 2 заключается в отсутствии необходимости моделирования каких-либо вспомогательных процессов, что позволяет существенно сократить время моделирования.

Список литературы

1. Bachelier, L. Theorie de la speculation / L. Bachelier // Annales de l'Ecole Normale Superieure. – 1900. – Vol. 17. – P. 21–86.
2. Schoutens, W. Levy processes in finance / W. Schoutens. – Bognor Regis: John Wiley & Sons Ltd, 2003. – 196 p.
3. Carr P., Geman H., Madan D. B., Yor M. The fine structure of asset returns: An empirical investigation // Journal of Business. 2002. Vol. 75, № 2. P. 305–332.
4. Труш, Н. Н. Моделирование процесса CGMY / Н. Н. Труш, А. В. Кузьмина // Вестник Белорусского государственного университета. Сер. 1, Физика, математика, информатика. – 2012. – № 2. – С. 102–105.

5. Poirot, J. Monte Carlo option pricing for tempered stable (CGMY) processes / J. Poirot, P. Tankov // Asia-Pacific Financial Market. – 2006. – Vol. 13, № 4. – P. 327.

6. Кузьмина, А. В. Моделироване дисперсионного гамма-процесса / А. В. Кузьмина // Вестник национальной академии наук Беларуси. Сер. физико-математических наук. – 2011. – № 1. – С. 70–74.

УДК [371.016:811.161.'243]:371.31

Линник Л. А., Петросян М. М.

ОБЛАКО СЛОВ КАК МЕТОД КОМПРЕССИИ ИНФОРМАЦИИ НАУЧНОГО ТЕКСТА

Любовь Александровна Линник
старший преподаватель
lyubovlinnik07@gmail.com
Мерри Мгеровна Петросян
старший преподаватель
merryhappy2013@yandex.ru
ФГБОУ ВО «Башкирский государственный медицинский университет», Россия, Уфа

TAG CLOUD AS A METHOD OF COMPRESSION OF INFORMATION OF SCIENTIFIC TEXT

Liubov Alexandrovna Linnik
Merri Mgerovna Petrosyan
Bashkir State Medical University, Russia, Ufa

Аннотация. *В статье рассматривается один из способов компрессии информации — создание облака слов, который позволяет определить наиболее общие субтемы текста. Даются сервисы, позволяющие составлять облако слов, сам метод компрессии сравнивается с другим методом сжатия научной информации — составлением аннотации. Определяется, что облако, в отличие от аннотации, представляет информацию более обобщенную, в визуальной форме,*