

ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЕ МОДЕЛИ

УДК 519.2

О МОДЕЛИРОВАНИИ УСТОЙЧИВЫХ СЛУЧАЙНЫХ ВЕЛИЧИН ПРИ α БЛИЗКИХ К ЕДИНИЦЕ

Архипов С.В., Багрова И.А.

Кафедра математической статистики и системного анализа

Поступила в редакцию 18.06.2010, после переработки 28.06.2010.

В работе уточняются формулы из [1] для параметров датчика, генерирующего положительные устойчивые числа при α близких к 1.

In this paper we clarify formulas for the parameters of the generator of positive stable numbers with values of α close to 1 provided in [1].

Ключевые слова: моделирование случайной величины, устойчивые распределения.

Keywords: modelling random variable, stable distributions.

Введение

В статье [1] был разработан датчик устойчивых случайных чисел, в основу моделирования которого была положена обобщенная центральная предельная теорема:

$$S_n = \frac{X_1 + \dots + X_n}{b \cdot n^{1/\alpha}} \xrightarrow{d} Y \text{ при } n \rightarrow \infty, \quad (1)$$

где $X_m, m = \overline{1, n}$ – случайные величины, имеющие распределение Парето с параметрами $\alpha \in (0, 1)$ и $\varepsilon > 0$, Y – устойчивая случайная величина. Для вычисления параметра b в [1] была предложена итерационная формула:

$$b_{k+1} = \left[\left(\Gamma(1 - \alpha) - \frac{b_k^{\alpha-1} \left(\frac{|t|\varepsilon}{n^{1/\alpha}} \right)^{1-\alpha}}{1 - \alpha} \right) \cdot \cos \frac{\pi\alpha}{2} \right]^{\frac{1}{\alpha}}, \quad (2)$$

полученная из требования сходимости характеристической функции суммы S_n к характеристической функции устойчивого распределения, взятого в форме (А) в предположении, что интеграл по дуге $I(\cdot)$ из [1], (2) является малой величиной при достаточно большом n .

Тестирование работы датчика при $\alpha \in [0.85; 1)$ показывает некоторое расхождение в районе моды графиков эмпирической и теоретической функций плотности, причем последняя имеет значения, полученные с помощью программы Stable.exe, разработанной Nolan'ом (см.[3]):

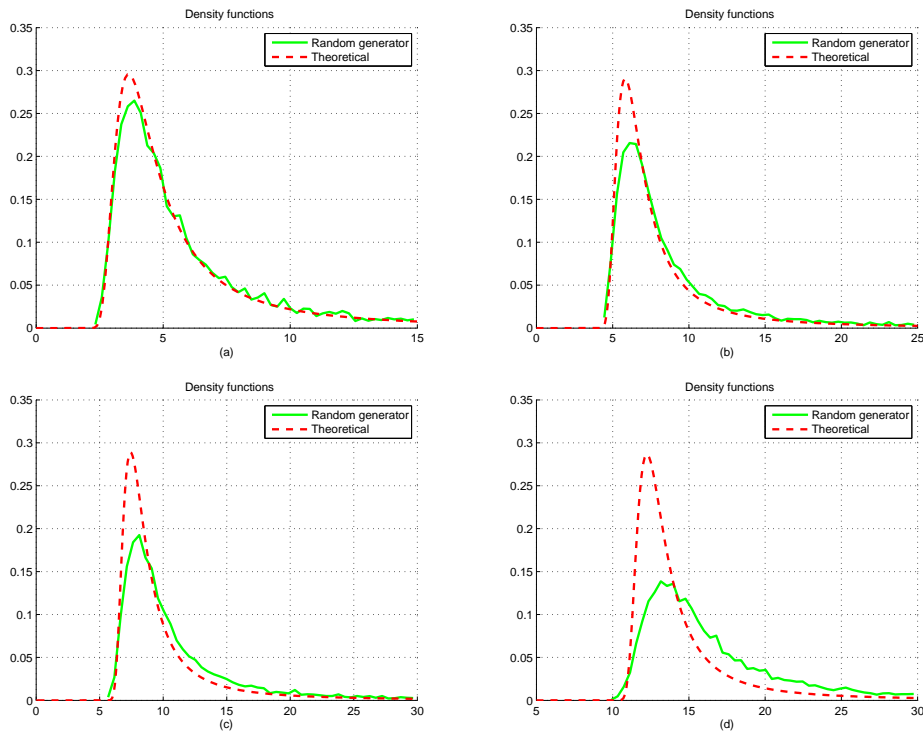


Рис. 1: График функций плотности сгенерированных чисел и устойчивой случайной величины в форме (A) при $n = 10^6$, $K = 10^4$,
 а) $\alpha = 0.85$, б) $\alpha = 0.9$, в) $\alpha = 0.92$, д) $\alpha = 0.95$

Для устранения этого недостатка можно увеличить количество слагаемых X_m . Однако эти действия не дают видимых улучшений. На рисунке 2 представлены функции плотности для α равных 0.9 и 0.95 при $n = 10^7$ (K – количество сгенерированных устойчивых чисел).

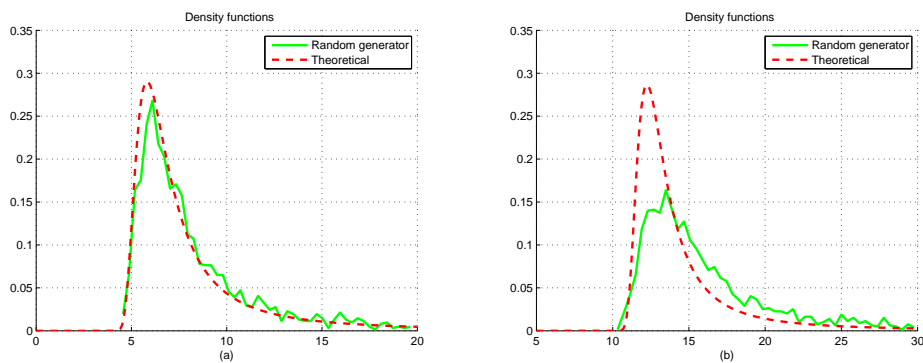


Рис. 2: График функций плотности сгенерированных чисел и устойчивой случайной величины в форме (A) при $n = 10^7$, $K = 2 \cdot 10^3$, а) $\alpha = 0.9$, б) $\alpha = 0.95$

Следовательно, возникает необходимость уточнения формулы (2).

1. О параметрах датчика устойчивых чисел в форме (А) при α близких к единице

Запишем асимптотическую формулу для характеристической функции суммы S_n , которая была получена в [1]:

$$f_{S_n}(t) = \exp \left\{ \frac{it\varepsilon}{bn^{1/\alpha-1}} + \frac{i|t|\varepsilon(-i \cdot \text{sign}(t))^\alpha}{bn^{1/\alpha-1}} \int_0^{-\frac{\pi}{2} \text{sign}(t)} e^{-\frac{|t|\varepsilon e^{i\varphi}}{bn^{1/\alpha}}} e^{i\varphi(1-\alpha)} d\varphi - \right. \\ \left. - \left(-\frac{it\varepsilon}{b} \right)^\alpha \Gamma\left(1-\alpha, \frac{|t|\varepsilon}{bn^{1/\alpha}}\right) \right\} + O\left(\frac{1}{n}\right).$$

При достаточно больших n и α близких к 1 подынтегральная функция имеет порядок $1 + O\left(\frac{1}{n^{1/\alpha}}\right) + O(1-\alpha)$, поэтому можно приближенно вычислить интеграл и выписать новое представление для характеристической функции:

$$f_{S_n}(t) = \exp \left\{ \frac{it\varepsilon}{bn^{1/\alpha-1}} + \frac{\pi}{2} (-i \cdot \text{sign}(t))^{\alpha+1} \frac{|t|\varepsilon}{bn^{1/\alpha-1}} - \right. \\ \left. - \left(-\frac{it\varepsilon}{b} \right)^\alpha \Gamma\left(1-\alpha, \frac{|t|\varepsilon}{bn^{1/\alpha}}\right) \right\} + O\left(\frac{1}{n}\right). \quad (3)$$

Обратим внимание на значения выражения $n^{1-1/\alpha}$. Если число слагаемых равно $n = 10^6$, то при $\alpha = 0.9, 0.95, 0.99$ они соответственно равны 0.2154, 0.4833, 0.8697. Получается, что при α достаточно близких к единице первым и вторым слагаемыми в показателе экспоненты (3) пренебречь нельзя.

Рассмотрим поведение характеристической функции устойчивого распределения в форме (А) при $\alpha \rightarrow 1$. Для этого разложим ее на два множителя:

$$f_Y(t) = e^{-|t|^\alpha} \cdot e^{i|t|^\alpha \text{sign}(t) \text{tg} \frac{\pi\alpha}{2}}.$$

Первый множитель стремится к $e^{-|t|}$ – характеристической функции распределения Коши с параметром $\lambda = 1$, а во втором множителе

$$|t|^\alpha \text{sign}(t) \rightarrow t, \quad \text{tg} \frac{\pi\alpha}{2} \rightarrow +\infty,$$

т.е. при $\alpha \rightarrow 1$ предельное распределение для S_n все больше похоже на стандартное распределение Коши, сдвинутое вправо на $\text{tg} \frac{\pi\alpha}{2}$.

Выпишем из показателя в (3) слагаемые, имеющие существенно ненулевую действительную часть при α близких к 1:

$$R_n(t) = \frac{\pi}{2} (-i \cdot \text{sign}(t))^{\alpha+1} \frac{|t|\varepsilon}{bn^{1/\alpha-1}} - \left(\frac{|t|\varepsilon}{b} \right)^\alpha \cos \frac{\pi\alpha}{2} \cdot \Gamma\left(1-\alpha, \frac{|t|\varepsilon}{bn^{1/\alpha}}\right), \quad (4)$$

где для степени комплексного числа было использовано равенство

$$\left(-\frac{it\varepsilon}{b} \right)^\alpha = \left(\frac{|t|\varepsilon}{b} \right)^\alpha \left(\cos \frac{\pi\alpha}{2} - i \sin \frac{\pi\alpha}{2} \cdot \text{sign}(t) \right).$$

Замечание 1. Присутствие параметра ε объясняется тем, что рассматривается общий случай с коэффициентом масштаба устойчивой плотности, равным ε .

Теперь в целях упрощения перейдем к соответствующим стандартным функциям плотности ($\lambda = \varepsilon = 1$). Модальное значение функции плотности распределения Коши равно $f(0) = 1/\pi \approx 0.3183$. Эмпирическая функция плотности нормированной суммы S_n имеет модальное значение, вычисленное исходя из следующих соображений. Параметр b , подобранный с помощью (2), делает в (4) множитель у $-(|t|\varepsilon)^\alpha$ близким к 1. Суммируя эту единицу с множителем перед $|t|(i \cdot \text{sign}(t))^{\alpha+1}$ в первом слагаемом, получаем приближенно коэффициент масштаба λ_1 распределения Коши

$$\lambda_1 \approx 1 + \frac{\pi}{2bn^{1/\alpha-1}},$$

а соответствующее модальное значение функции плотности S_n будет равно:

$$f_{S_n}(0) \approx \frac{1}{\pi\lambda_1}.$$

Подставляя, например, значение $\alpha = 0.9$, $n = 10^6$ и $b = 1.2013$, получаем $f_{S_n}(0) = 0.2483$, что и показывает график эмпирической плотности (см. Рис.1).

Вернемся к произвольному коэффициенту масштаба $\varepsilon > 0$. Для совпадения модальных значений функций плотности нужно добиться того, чтобы суммарный коэффициент масштаба был близок к ε . Для выполнения этого условия приравняем

$$\left(\frac{|t|\varepsilon}{b}\right)^\alpha \left(\frac{\frac{\pi}{2}(-i \cdot \text{sign}(t))^{\alpha+1}}{n^{1/\alpha-1}} \left(\frac{|t|\varepsilon}{b}\right)^{1-\alpha} - \cos \frac{\pi\alpha}{2} \cdot \Gamma\left(1 - \alpha, \frac{|t|\varepsilon}{bn^{1/\alpha}}\right)\right) = -(|t|\varepsilon)^\alpha. \quad (5)$$

Откуда получаем рекуррентную формулу для вычисления параметра \tilde{b}

$$\tilde{b}_{k+1} = \left[\cos \frac{\pi\alpha}{2} \left(\Gamma(1 - \alpha) - \frac{\tilde{b}_k^{\alpha-1} \left(\frac{|t|\varepsilon}{n^{1/\alpha}}\right)^{1-\alpha}}{1 - \alpha} \right) - \frac{\pi}{2} \frac{(-i \cdot \text{sign}(t))^{\alpha+1} \left(\frac{|t|\varepsilon}{\tilde{b}_k}\right)^{1-\alpha}}{n^{1/\alpha-1}} \right]^{1/\alpha}, \quad (6)$$

где была применена асимптотическая формула для верхней неполной гамма-функции (см. [1], с. 56).

Замечание 2. Если $n^{1-1/\alpha}$ достаточно мало, то вторым слагаемым можно пренебречь и мы приходим к формуле (2).

Отметим, что \tilde{b}_k является комплексным числом и зависит от n , ε и t . Первые два параметра задаются, а затем для каждого $t_j = -100 + j \cdot h$, $j = 0, 200$, $j \neq 100$ и $h = 1$ вычисляются $\tilde{b}_k(t_j)$ и нужно оставить такое, которое имеет минимальную сумму модулей разностей между экспонентами левой и правой частей в (5)

$$\sum_j \left| \exp\{\text{left}(t_j)\} - \exp\{\text{right}(t_j)\} \right|.$$

Поскольку нормирующий коэффициент должен быть вещественным, то положим $b = \text{Re } \tilde{b}$, при этом отбрасываемая мнимая часть по модулю имеет порядок 10^{-2} .

Например, для $\alpha = 0.9$ при различных n с помощью описанного выше алгоритма вычислены следующие значения параметра b :

Таблица 1: Параметры датчика (1) устойчивых чисел в форме (A) при $\alpha = 0.9$

n	10^3	10^4	10^5	10^6
b	1.5462	1.5481	1.5496	1.5509

Результат генерирования устойчивых чисел по схеме (1) с вычисленными коэффициентами b можно посмотреть на Рис.3.

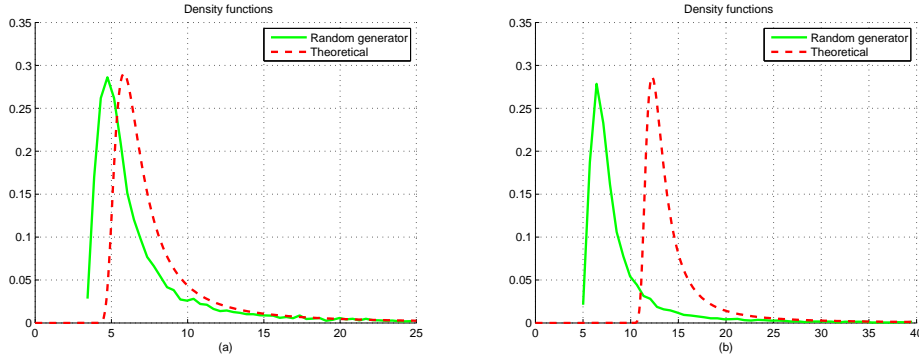


Рис. 3: График функций плотности сгенерированных чисел и устойчивой случайной величины в форме (A) при $n = 10^6$, $K = 10^4$, а) $\alpha = 0.9$, б) $\alpha = 0.95$

Откуда видно, что изменение b помимо увеличения модального значения естественным образом сдвинуло влево график эмпирической плотности. Поэтому необходимо добавить сдвиг a в (1).

Его значение определяется из рассмотрения выражений в (3), отвечающих за сдвиг. В пределе при $n \rightarrow \infty$ должно выполняться соотношение:

$$\frac{it\varepsilon}{\tilde{b}n^{1/\alpha-1}} + it \cdot \tilde{a} + i \left(\frac{|t|\varepsilon}{\tilde{b}} \right)^\alpha \sin \frac{\pi\alpha}{2} \Gamma \left(1 - \alpha, \frac{|t|\varepsilon}{\tilde{b}n^{1/\alpha}} \right) \cdot \text{sign}(t) \rightarrow i(|t|\varepsilon)^\alpha \text{tg} \frac{\pi\alpha}{2} \text{sign}(t).$$

Считая $|t|^\alpha \text{sign}(t) \approx t$ при α близких к 1, получаем выражение для коэффициента сдвига

$$\tilde{a} = \left(\frac{1}{\cos \frac{\pi\alpha}{2}} - \frac{\Gamma \left(1 - \alpha, \frac{|t|\varepsilon}{\tilde{b}n^{1/\alpha}} \right)}{\tilde{b}^\alpha} \right) \varepsilon^\alpha \sin \frac{\pi\alpha}{2} - \frac{\varepsilon}{\tilde{b}n^{1/\alpha-1}}. \quad (7)$$

Как уже было отмечено, на выходе итерационной формулы (5) мы имеем комплексное число \tilde{b} . Именно это значение подставляется в (7) для определения \tilde{a} . Оптимальные $\tilde{b}_{\text{опт}}$ и $\tilde{a}_{\text{опт}}$ будем выбирать следующим образом. Для каждого фиксированного t_j и заданных α, n, ε определим из (5) и (7) соответствующие \tilde{b} и \tilde{a} .

Затем для каждой пары \tilde{b} и \tilde{a} будем вычислять $MAE = \frac{1}{N} \sum_j |f_{S_n}(t_j) - f_Y(t_j)|$ и выберем такие $\tilde{b}_{\text{опт}}, \tilde{a}_{\text{опт}}$, которые минимизируют MAE .

Далее берем их действительные части:

$$a = \text{Re } \tilde{a}_{\text{опт}}, \quad b = \text{Re } \tilde{b}_{\text{опт}}.$$

Замечание 3. Теоретическое предположение $n \rightarrow \infty$ преобразует (5) и (7) в выражения для b и a , которые приведены в [4], 2.5, а именно

$$b = \pi / \left(2\Gamma(\alpha) \sin \frac{\pi\alpha}{2} \right) = \Gamma(1 - \alpha) \cos \frac{\pi\alpha}{2}, \quad a = 0.$$

Для практического применения вынесем параметры стандартного датчика ($\varepsilon = 1$)

$$S_n = \frac{X_1 + \dots + X_n}{b \cdot n^{1/\alpha}} + a \quad (8)$$

в таблицу

Таблица 2: Параметры датчика (8) устойчивых чисел в форме (A)

α	n	10^3	10^4	10^5	10^6	10^7	10^8
0.85	b	1.5355	1.5421	1.5466	1.548	1.5489	1.5496
	a	1.068	0.6953	0.4587	0.3044	0.2022	0.1345
0.875	b	1.5399	1.5431	1.5467	1.5498	1.5507	1.5513
	a	1.6757	1.1965	0.8509	0.6085	0.437	0.314
0.9	b	1.5462	1.5481	1.5496	1.5509	1.5524	1.5537
	a	2.6767	2.0673	1.5955	1.2313	0.9494	0.7337
0.925	b	1.5528	1.5537	1.5544	1.5551	1.5556	1.5561
	a	4.5063	3.7374	3.0983	2.5681	2.1284	1.7641
0.95	b	1.5593	1.5596	1.5598	1.5601	1.5603	1.5605
	a	8.4397	7.4771	6.6234	5.8667	5.1962	4.6021
0.99	b	1.5687	1.5687	1.5687	1.5688	1.5688	1.5688
	a	58.8431	57.4905	56.169	54.8778	53.6163	52.3837
0.999	b	1.5706	1.5706	1.5706	1.5706	1.5706	1.5706
	a	631.6791	630.2248	628.7739	627.3263	625.8821	624.4412

Отдельно для характеристического показателя $\alpha = 0.9$ вынесем в таблицу значения функционалов ошибок между теоретической и эмпирической функциями плотности на промежутке $[4.7, 11.1]$, на котором $\{x : f_Y(x) < \frac{1}{10} \max f_Y(x)\}$:

Таблица 3: Значения средних абсолютных ошибок, квадратов ошибок и абсолютных процентных ошибок при $\alpha = 0.9$, $K = 10^4$

n	10^3	10^4	10^5	10^6
MAE	0.010243	0.008152	0.010404	0.008523
MSE	0.000173	0.000112	0.000182	0.000125
MAPE	10.538143	8.461984	9.559375	8.470700

После использования уточненных значений для b и a датчик случайных чисел, основанный на формуле (8), дал следующие результаты:

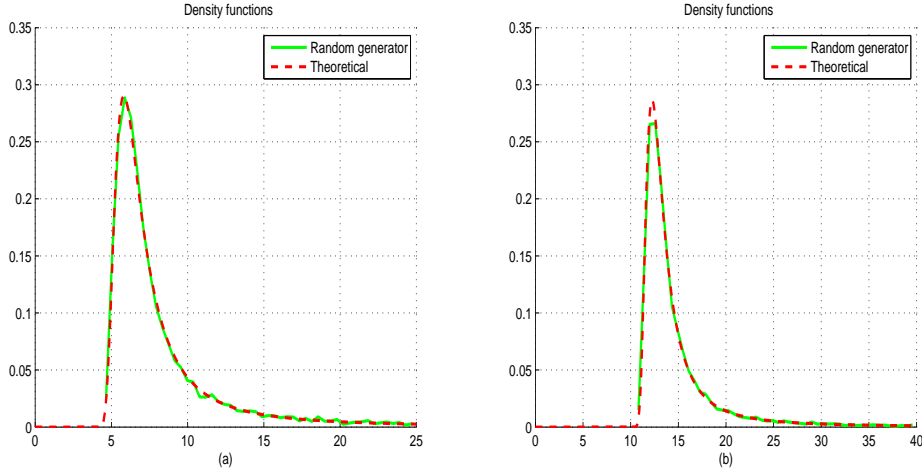


Рис.4: График функций плотности сгенерированных чисел и устойчивой случайной величины в форме (A) при $n = 10^6$, $K = 10^4$, а) $\alpha = 0.9$, б) $\alpha = 0.95$

2. О моделировании устойчивых чисел для других форм представлений

В работе [2] было отмечено, что при $\alpha \rightarrow 1$ наиболее естественным будет стремление устойчивой функции плотности с единичным коэффициентом масштаба к дельта-функции $\delta(1)$. Тогда плотность суммы паретовских случайных величин с $\varepsilon = 1$ также должна иметь пределом $\delta(1)$. В терминах характеристических функций это утверждение выглядит следующим образом (см. [1], (2), (3)):

$$f_{S_n}(t) \xrightarrow{n \rightarrow \infty} f_Y(t) = \exp\{-(-it)^\alpha\} \xrightarrow{\alpha \rightarrow 1} \exp(it \cdot 1).$$

Соответствующие формулы для вычисления $b = \operatorname{Re} \tilde{b}$, $a = \operatorname{Re} \tilde{a}$ имеют вид:

$$\tilde{b}_{k+1} = \left[\left(\Gamma(1 - \alpha) - \frac{\tilde{b}_k^{\alpha-1} \left(\frac{|t|\varepsilon}{n^{1/\alpha}}\right)^{1-\alpha}}{1 - \alpha} \right) - \frac{\pi (-i \cdot \operatorname{sign}(t))^{\alpha+1} \left(\frac{|t|\varepsilon}{b_k}\right)^{1-\alpha}}{n^{1/\alpha-1} \cos \frac{\pi\alpha}{2}} \right]^{1/\alpha}, \quad (9)$$

$$\tilde{a} = \left(1 - \frac{\Gamma\left(1 - \alpha, \frac{|t|\varepsilon}{bn^{1/\alpha}}\right)}{\tilde{b}^\alpha} \right) \varepsilon^\alpha \sin \frac{\pi\alpha}{2} - \frac{\varepsilon}{\tilde{b}n^{1/\alpha-1}}. \quad (10)$$

Замечание 4. Отсутствие множителя $\cos \frac{\pi\alpha}{2}$ (см.(6)) в формуле (9) значительно увеличивает b , что делает первое и второе слагаемые в показателе (3) достаточно маленькими.

Приведем некоторые значения параметров b и a , вычисленные с помощью (9) и (10) с последующей процедурой их оптимизации:

Таблица 4: Параметры датчика (8) устойчивых чисел, распределенных около 1 ($\varepsilon = 1$).

α	n	10^3	10^4	10^5	10^6	10^7	10^8
0.85	b	8.5332	8.5525	8.5662	8.5755	8.5797	8.5824
	a	0.1891	0.1247	0.0826	0.0549	0.0365	0.0243
0.875	b	9.9911	10.0074	10.0213	10.0322	10.0401	10.045
	a	0.2562	0.1831	0.1309	0.0938	0.0674	0.0485
0.9	b	12.1587	12.1712	12.1813	12.1895	12.1959	12.202
	a	0.3401	0.2625	0.2026	0.1564	0.1207	0.0933
0.925	b	15.7239	15.7316	15.7383	15.7439	15.7487	15.7528
	a	0.4451	0.369	0.3058	0.2535	0.2101	0.1741
0.95	b	22.7274	22.7312	22.7347	22.7378	22.7405	22.743
	a	0.5792	0.5131	0.4545	0.4025	0.3565	0.3157
0.99	b	104.1532	104.1533	104.1535	104.1537	104.1539	104.154
	a	0.8863	0.8659	0.846	0.8266	0.8076	0.789
0.999	b	1006.357	1006.357	1006.357	1006.357	1006.357	1006.357
	a	0.9858	0.9836	0.9813	0.9791	0.9768	0.9746
0.9999	b	10008.6377	10008.6377	10008.6377	10008.6377	10008.6377	10008.6377
	a	0.9983	0.9981	0.9979	0.9977	0.9974	0.9972

Продemonстрируем работу датчика при $\alpha = 0.95$ и $n = 10^5$ и $n = 10^7$:

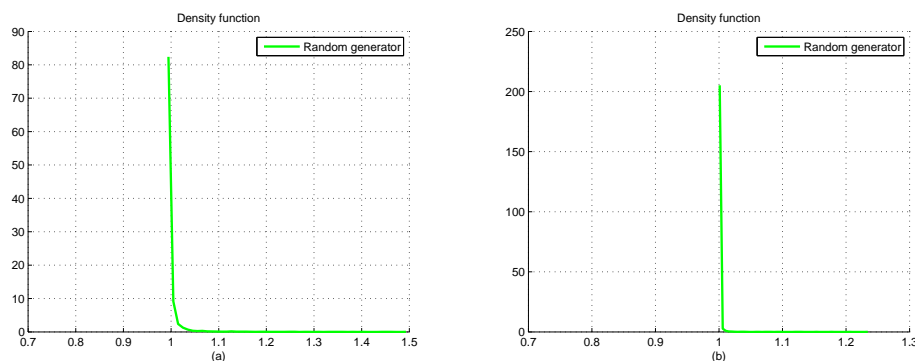


Рис.5: График функций плотности сгенерированных чисел при $n = 10^4$, $K = 10^4$,
a) $\alpha = 0.999$, b) $\alpha = 0.9999$

Отметим, что увеличение модального значения эмпирической функции плотности происходит с увеличением α . Тот факт, что ε – левая граница распределения Парето является параметром масштаба предельного устойчивого распределения, можно подтвердить с помощью графиков. Например, при $\varepsilon = 2$ и $\varepsilon = 10$ имеем:

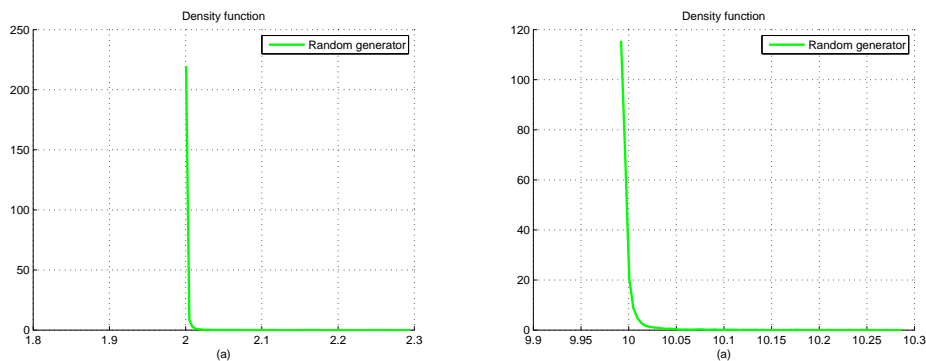


Рис. 6: График функций плотности сгенерированных чисел при $n = 10^4$, $K = 10^4$, $\alpha = 0.9999$, а) $\varepsilon = 2$, б) $\varepsilon = 10$

Формы представления (А) и (М) отличаются сдвигом: $-\operatorname{tg} \frac{\pi\alpha}{2}$. Поэтому, добавив это значение к выражению для a , получим параметр сдвига для формы (М):

$$a_M = a - \operatorname{tg} \frac{\pi\alpha}{2}.$$

Понятно, что значения параметра b из таблицы 2 остаются без изменения.

Приведем графики эмпирической и теоретической (со значениями из [3]) функций плотности:

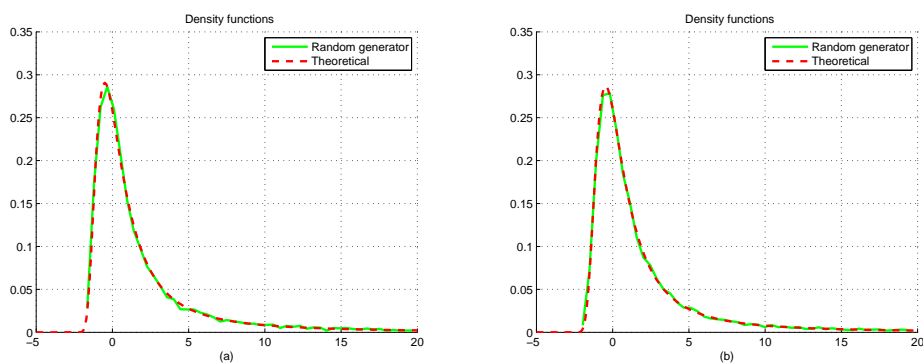


Рис. 7: График функций плотности сгенерированных чисел и устойчивой случайной величины в форме (М) при $n = 10^4$, $K = 10^4$, а) $\alpha = 0.9$, б) $\alpha = 0.95$

Замечание 5. В инструкции к программе Nolan'a stable.exe (см. [3]) говорится о неприменимости разработанного в ней датчика (форма (А)) при $|1 - \alpha| < 0.005$. Числа, генерируемые по формуле (8) с параметрами из (6) и (7), позволяют получить правильное расположение (график должен быть примерно в районе $\operatorname{tg} \frac{\pi\alpha}{2}$) эмпирической функции плотности:

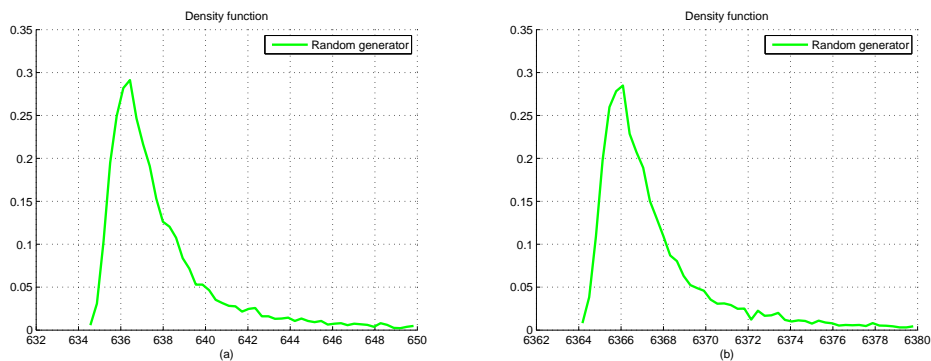


Рис. 8: График функций плотности сгенерированных чисел при $n = 10^4$, $K = 10^4$, а) $\alpha = 0.999$, б) $\alpha = 0.9999$

Заключение

В статье получены уточненные формулы для параметров датчика, генерирующего положительные устойчивые числа при α близких к 1. Приведены соответствующие таблицы для параметров α , a , b и n . Продемонстрирована работа датчика устойчивых чисел, имеющих различные формы представления.

Список литературы

- [1] Архипов С.В., Багрова И.А. О моделировании односторонних устойчивых случайных величин // Вестник Тверского государственного университета. Серия: Прикладная математика, выпуск 4(15). – Тверь: изд-во Тверского государственного университета, 2009. – с. 53-62.
- [2] Архипов С.В. Замечания к представлению характеристических функций устойчивых распределений. // Сб. «Сложные системы: обработка информации, моделирование и оптимизация». – Тверь: изд-во Тверского государственного университета, 2002. – с. 97-101.
- [3] J. P. Nolan. <http://academic2.american.edu/~jpnolan/stable/stable.html>
- [4] Uchaikin V.V., Zolotarev V.M. Chance and Stability. Stable Distributions and their Applications. – Utrecht: VSP, 1999. – 594 p.