УДК 519.2:530.1:600.1

МЕТОДИКА ОЦЕНКИ НАРУШЕНИЙ СТАТИСТИЧЕСКОЙ УСТОЙЧИВОСТИ ПРОЦЕССОВ, НЕ ИМЕЮЩИХ ДИСПЕРСИИ

И.И. Горбань
Институт проблем математических машин и систем НАН Украины
e-mail: igor.gorban@yahoo.com

1. Введение

Одним из удивительных физических феноменов является *статистическая устойчивость массовых явлений*, проявляющаяся в стабильности статистик.

Современная теория вероятностей (включающая в широком понимании и математическую статистику) описывает массовые явления с помощью случайных (стохастических) математических моделей, характеризуемых вероятностной мерой. В основе построения таких моделей лежит физическая гипотеза идеальной статистической устойчивости, предполагающая наличие сходимости частоты реальных событий и средних значений физических величин.

Многие годы гипотеза идеальной статистической устойчивости не вызывала сомнений. Однако, последние экспериментальные исследования различных физических величин и процессов на больших интервалах наблюдения показали, что она не находит экспериментального подтверждения.

На относительно небольших временных, пространственных или пространственно-временных интервалах наблюдения увеличение объема данных приводит к уменьшению уровня флуктуаций статистик. Однако при больших объемах эта тенденция не прослеживается: достигнув определенного значения, уровень флуктуаций практически не меняется или возрастает. Это указывает на отсутствие сходимости реальных статистик (их несостоятельность).

Статистическая устойчивость – свойство, характеризующее *как статистику, так и сам процесс.* Под статистической устойчивостью обычно подразумевают статистическую устойчивость определенной статистики или определенного класса статистик. В теории под *статистической устойчивостью процесса в узком смысле* понимается сходимость *эмпирической функции распределения*, а под *статистической устойчивостью процесса в широком смысле* – сходимость *выборочного среднего* и *выборочного среднеквадратического отклонения (СКО).*

При решении практических задач обычно не важна специфика поведения процесса на бесконечно большом интервале наблюдения, хотя именно она заложена в основу формального определения статистически устойчивого процесса. Более существенны особенности поведения процесса *на рассматриваемых интервалах наблюдения*: отсутствие или наличие тенденции стабилизации параметров, характеризующих нарушение статистической устойчивости. Если на интервале наблюдения прослеживается тенденция стабилизации этих параметров, процесс, считается статистически устойчивым, в противном же случае – статистически неустойчивым.

Разные статистики и разные процессы, как правило, имеют разные интервалы статистической устойчивости.

Если интервалы наблюдения не превышают интервал статистической устойчивости, возможно корректное использование классических вероятностно-статистических методов обработки; если же интервалы наблюдения превышают интервал статистической устойчивости, приходится использовать другие методы, в частности методы новой *физико-математической теории гиперслучайных явлений* [1], специально разработанной в интересах учета нарушений статистической устойчивости.

Нарушения статистической устойчивости вызываются различными причинами. Поэтому связать их с определенными особенностями процессов непросто.

В современном понимании феномен статистической устойчивости формально не связан с существованием вероятностной меры [1]. Поэтому не только случайные процессы, характеризуемые вероятностной мерой, могут быть статистически неустойчивыми. Исследования показали, что некоторые *детерминированные процессы* – статистически неустойчивые. Среди нестационарных процессов встречаются как *статистически устойчивые*, так и *неустойчивые процессы*. *Статистически неустойчивые случайные процессы не обязательно нестационарные*. Существуют *стационарные в узком смысле статистически неустойчивые случайные процессы.*

Методика фиксации нарушений статистической устойчивости и измерения степени нарушений базируется на расчете различных параметров. Некоторые из них (γ_N, μ_N, h_N, l_N) характеризуют нарушения статистической устойчивости по отношению к среднему, другие (Γ_N, M_N, H_N, L_N) – нарушения устойчивости по отношению к СКО [1]. В формулах, описывающих указанные параметры, фигурирует *дисперсия процесса* или *ее оценка*. Поэтому *применение методики возможно лишь тогда, когда процесс имеет определенную дисперсию*.

Не все распределения имеют дисперсии. Естественно возникает вопрос: как поступать в случае отсутствия дисперсии? *Целью работы* является *адаптация методики оценки нарушений статистической устойчивости для процессов, не имеющих дисперсии*.

2. Распределения, не имеющие моментов

Известно множество распределений случайной величины, у которых либо вообще нет моментов, либо отсутствуют моменты выше первого. К ним относятся распределение Коши (распределение Стьюдента первого порядка), а также при определенном соотношении параметров распределения Парето, Фишера-Снедекора (F распределение), Фреше и др.

При отсутствии *математического ожидания* или *дисперсии*, как правило, плотность распределения и функция распределения имеют *«тяжелые хвосты» (heavy (fat) tails)*. Такие распределения описывают реальные физические явления, в которых *наблюдаются редкие, но существенно влияющие на статистику, события*. Примеры двух распределений с «тяжелыми хвостами» приведены на рис. 1.

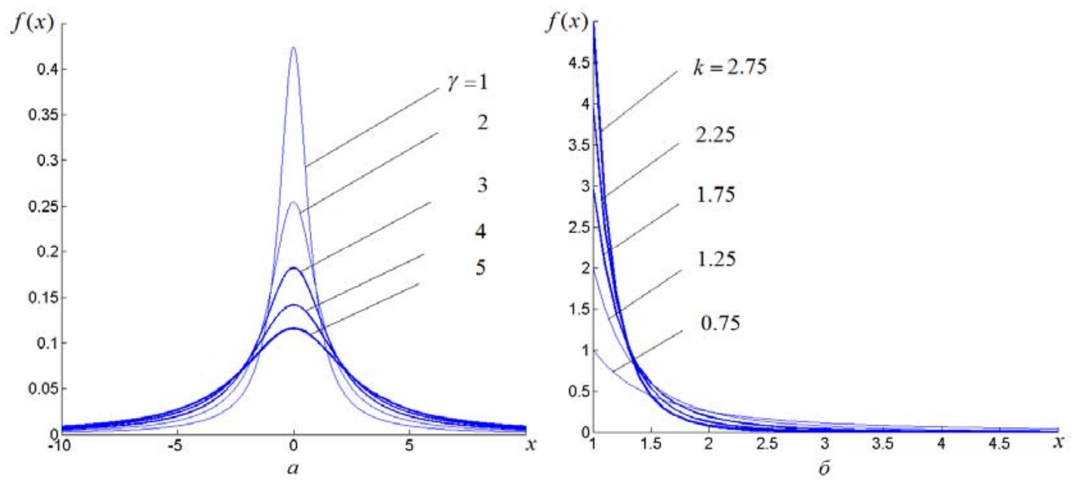

Рис. 1. Функции распределения Коши (*а*) и Парето (*б*), $x_0 = 0$

Распределение Коши и *распределение Парето* описываются соответственно плотностями распределения

$$f(x) = \frac{1}{\pi}\left(\frac{\alpha}{(x-x_0)^2+\alpha^2}\right) \text{ и } f(x) = \frac{k}{(x-x_0)^{k+1}},$$

где x_0 – параметр сдвига, $\alpha > 0$ – параметр масштаба, параметр $k > 0$.

В распределении Коши моменты всех порядков отсутствуют (хотя существует интеграл в смысле главного значения, описывающий первый момент). В распределении Парето при $k \le 1$ отсутствуют моменты всех порядков, а при $k \le 2$ – отсутствуют моменты, начиная со второго порядка.

3. Параметры статистической неустойчивости

Для исследования статистической устойчивости процессов с определенными дисперсиями часто используют *параметры статистической неустойчивости по отношению к среднему* $h_N = \gamma_N / \gamma_{0N}$ и *СКО* $H_N = \Gamma_N / \gamma_{0N}$. Их оценки описываются выражениями $h_N^* = \gamma_N^* / \gamma_{0N}$, $H_N^* = \Gamma_N^* / \gamma_{0N}$, где γ_N^* и Γ_N^* – оценки параметров статистической неустойчивости γ_N и Γ_N:

$$\gamma_N^* = \mathrm{M}^*[\bar{D}_{Y_N}] \big/ \mathrm{M}^*[\bar{D}_{X_N}],\ \Gamma_N^* = \mathrm{M}^*[\bar{D}_{Z_N}] \big/ \mathrm{M}^*[\bar{D}_{X_N}], \quad (1)$$

$\mathrm{M}^*[\cdot]$ – оператор усреднения по реализациям, $\bar{D}_{Y_N} = \frac{1}{N-1}\sum_{n=1}^{N}(Y_n - \bar{m}_{Y_N})^2$ – выборочная дисперсия флуктуации выборочного среднего $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (n = \overline{1,N})$, $\bar{m}_{Y_N} = \frac{1}{N}\sum_{n=1}^{N} Y_n$ – выборочное среднее флуктуации среднего, $\bar{D}_{X_N}$ – выборочная дисперсия процесса X_n $(n = \overline{1,N})$, γ_{0N} – параметр статистической неустойчивости шума по отношению к среднему γ_N, рассчитанный теоретически [1] для эталонной статистически устойчивой последовательности N некоррелированных отсчетов с постоянной дисперсией и нулевым математическим ожиданием, $\bar{D}_{Z_N} = \frac{1}{N-2}\sum_{n=2}^{N}(Z_n - \bar{m}_{Z_N})^2$ – выборочная дисперсия флуктуации выборочного СКО $Z_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - Y_n)^2} \quad (n = \overline{2,N})$, $\bar{m}_{Z_N} = \frac{1}{N-1}\sum_{n=2}^{N} Z_n$ – среднее флуктуации выборочного СКО.

Процесс считается статистически устойчивым по отношению к среднему и СКО на рассматриваемом интервале наблюдения, если на этом интервале оценки h_N^* и H_N^* не выходят за верхнюю границу $h_{0N}^+ = \gamma_{0N}^+ / \gamma_{0N}$ коридора устойчивости, где $\gamma_{0N}^+ = \gamma_{0N} + \varepsilon\sigma_{\tilde{\gamma}_{0N}}$ – верхняя граница коридора устойчивости параметра γ_N, ε – константа, характеризующая ширину коридора, $\sigma_{\tilde{\gamma}_{0N}}$ – СКО эталонной статистически устойчивой последовательности.

Если процесс не имеет дисперсии, использовать оценки h_N^* и H_N^* нельзя. Обойти указанную трудность можно, заменив в выражении (1) выборочную дисперсию $\bar{D}_{X_N}$ на *специально подобранную робастную статистику.* Для распределений Коши и Парето хороший результат дает величина $N s_{X_N}^{*2}$, где $s_{X_N}^* = \mathrm{med}\left(\left|X_n - m_{X_N}^*\right|,\quad n = \overline{1,N}\right)$ – абсолютное медианное отклонение, $m_X^* = \mathrm{med}\left(X_n \quad n = \overline{1,N}\right)$ – медианное смещение.

Результаты расчетов оценок h_N^*, H_N^* с использованием такой статистики представлены на рис. 2. Пунктирными линиями на рисунке изображены верхние границы коридора устойчивости, соответствующие $\varepsilon = 3$. Для моделирования использовалось 100 реализаций.

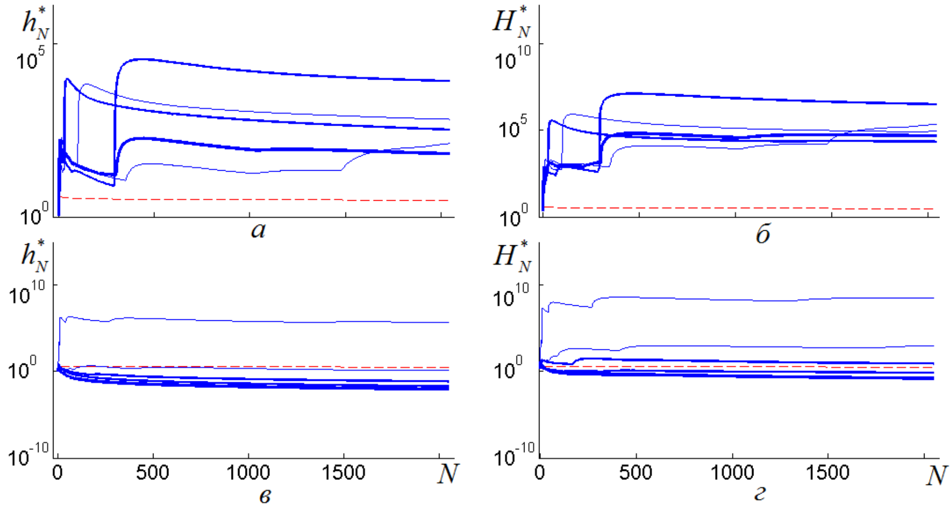

Рис. 2. Параметры статистической неустойчивости h_N^*, H_N^* для соответственно распределения Коши (*а*, *б*) и Парето (*в, г*). Параметры распределений указаны на рис. 1 (толщина кривых на обоих рисунках согласованы)

Согласно результатам моделирования процесс, описываемый распределением Коши, статистически неустойчив по отношению к среднему и СКО при любых соотношениях параметров. Процесс же, описываемый распределением Парето, статистически неустойчив по отношению к среднему при $k = 0.75$ и устойчив при $k > 1$. По отношению к СКО он статистически неустойчив при $k = 0.75, 1.25, 1.75$ и статистически устойчив при $k > 2$. Эти результаты полностью согласуются с теорией.

Обратим внимание, что *предлагаемая модификация методики применима лишь для процессов, не имеющих дисперсии*. Моделирование показывает, что хотя в частном случае некоррелированного гауссовского процесса получаемые оценки h_N^* и H_N^* не выходят за верхнюю границу коридора устойчивости (т.е. верно фиксируют наличие статистической устойчивости), однако оказываются сильно заниженными по сравнению с теоретическими. Поэтому при небольших нарушениях статистической устойчивости зафиксировать эти нарушения не представляется возможным.

Отсюда следует, что *область практического применения новой методики ограничена процессами с редкими, но существенно влияющими на статистику, отсчетами*.

4. Выводы

1. Известная методика оценки нарушений статистической устойчивости, разработанная для случайных процессов с определенными дисперсиями, адаптирована для случайных процессов, не имеющих дисперсии.

2. Область практического применения новой модификации методики ограничена процессами с редкими, но существенно влияющими на статистику, отсчетами.

Список литературы

1. Горбань И.И. Феномен статистической устойчивости [Электронный ресурс] / Горбань И.И. – К.: Наукова думка, 2014. – 444 с. – Режим доступа: http://www.immsp.kiev.ua/perspages/gorban_i_i/index.html.