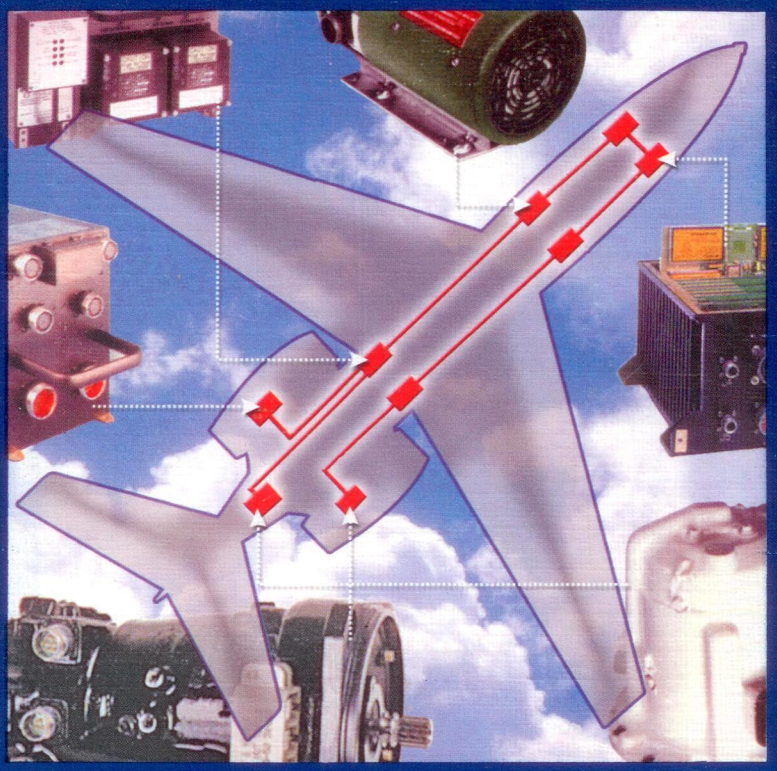
ISSN: 1993-6982

ВЕСТНИК МЭИ



В НОМЕРЕ:

Кафедра «Электротехнические комплексы автономных объектов»: становление, этапы развития, современное состояние

Состояние и проблемы развития синхронно-гистерезисного привода

Политехническая интернет-лаборатория МЭИ

Алгоритмы обнаружения аномальных наблюдений в коррелированных временных рядах

2007

ИЗДАТЕЛЬСКИЙ ДОМ МЭИ

3

Алгоритмы обнаружения аномальных наблюдений в коррелированных временных рядах*

А. Ш. Авшалумов, Г. Ф. Филаретов

Рассматривается задача обнаружения аномальных наблюдений в коррелированных случайных последовательностях. Под аномальными здесь понимаются значения, существенно нарушающие гладкость траектории наблюдаемой последовательности. Предлагаются два алгоритма выделения аномальных наблюдений, основанные на использовании первых и вторых разностей. Излагаются теоретические основы алгоритмов, приводится их подробное описание, достаточное для практической реализации. На модельном примере демонстрируется работоспособность алгоритмов и их эффективность для решения указанной задачи.

Введение

Задача обнаружения аномальных наблюдений является одной из стандартных задач обработки экспериментальных данных. При этом традиционно под аномальными понимаются наблюдения, резко выделяющиеся на фоне прочих наблюдений. Если наблюдаемые флюктуации носят случайный характер, то под подозрение попадают крайние члены вариационного ряда, в первую очередь минимальное $x_{\min}$ или максимальное $x_{\max}$ значения, присутствующие в выборке.

Известны статистические процедуры, позволяющие корректно установить, можно или нет отнести указанные значения к категории аномальных. Такая проверка осуществляется с помощью специальных статистических критериев и сводится к сравнению некоторого вычисляемого по экспериментальным данным показателя с критическим табличным значением при заданном уровне значимости и фиксированном числе наблюдений N [1]. В основе такого рода критериев лежат две предпосылки: наблюдаемая случайная величина подчиняется нормальному (гауссовскому) распределению, наблюдения статистически независимы. Кроме того, предполагается, что количество возможных аномальных наблюдений относительно невелико по сравнению с N.

При анализе дискретных случайных процессов (временных рядов) использование подобных критериев некорректно вследствие коррелированности ординат процесса и, следовательно, нарушения второй предпосылки. Более того, само понятие аномальности отдельных наблюдений в данном случае мало продуктивно. Далее рассматривается иной подход к обнаружению аномальных наблюдений, учитывающий специфические особенности временных рядов.

Исходная идея и предпосылки

Предлагаемый подход основан на свойствах траектории любой коррелированной случайной последовательности обладать определенной степенью гладкости, тем большей, чем более коррелированными являются соседние наблюдения. В этой ситуации под *аномальными* логично понимать наблюдения, *достаточно грубо нарушающие гладкость траектории* наблюдаемой дискретной последовательности. Иными словами, аномальными будут считаться точки, соответствующие положительному или отрицательному

*Статья поступила в редакцию 07.02.07.

выбросу процесса, существенно искажающему те или иные характеристики его гладкости.

Возможны различные способы формирования критериев, связанных с характеристиками гладкости процесса и позволяющих обнаружить их значимые отклонения от нормы, что и будет свидетельствовать о наличии аномальности. В этой связи можно предложить различные алгоритмы обнаружения аномальных в указанном смысле наблюдений. Далее рассматриваются два таких алгоритма. В их основе лежат следующие предпосылки:

а) наблюдаемый дискретный процесс $z(t)$ может быть представлен в виде суммы: $z(t) = x(t) + i(t)$;

б) процесс $x(t)$ является дискретным гауссовским процессом с дисперсией σ_X^2 и нормированной автокорреляционной функцией $\rho_{XX}(k)$, $k = 0, 1, 2 \ldots$;

в) процесс $i(t)$ состоит из импульсов, искажающих значения $x(t)$ и приводящих к появлению аномальности наблюдений; моменты появления импульсов случайны и неизвестны. Иными словами, выброс носит «точечный» характер;

г) предполагается, что импульсы разделены по крайней мере единичным интервалом;

д) количество выбросов относительно невелико по сравнению с общим числом наблюдений N;

е) амплитуды импульсов малы по сравнению с σ_X, что делает практически невозможным их визуальное выявление.

Алгоритм 1-го порядка

В данном алгоритме в качестве характеристик, связанных с гладкостью траекторий, используются, что вполне естественно, разности 1-го порядка $\nabla_t = z_t - z_{t-1}$, где $z_1, z_2, \ldots, z_N$ — наблюдаемые значения временного ряда, $t = \overline{2, N}$. Очевидно, что положительному выбросу будет соответствовать некоторая точка x_k, для который справедливы следующие соотношения:

$$\left.\begin{aligned} \nabla_k &= z_k - z_{k-1} > 0; \\ \nabla_{k+1} &= z_{k+1} - z_k < \nabla_k . \end{aligned}\right\} \qquad (1)$$

Аналогичным образом для отрицательного выброса в точке x_l имеем

$$\left.\begin{aligned} \nabla_l &= z_l - z_{l-1} < 0; \\ \nabla_{l+1} &= z_{l+1} - z_l > \nabla_l . \end{aligned}\right\} \qquad (2)$$

Характер изменения значений временного ряда первых разностей $\nabla_t = z_t - z_{t-1}$ при появлении импульсной составляющей i_t наглядно представлен на рис. 1.

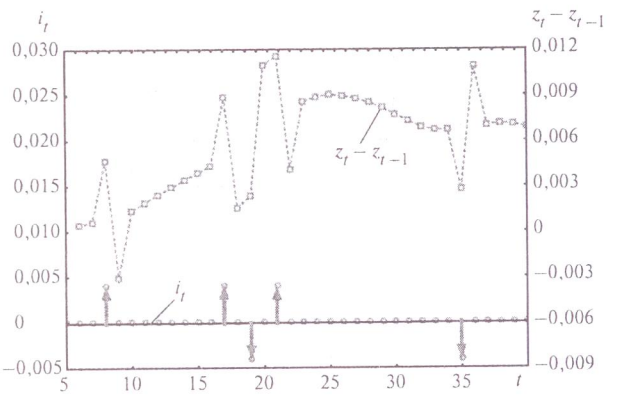


Рис. 1

Как следует из этого рисунка, в момент появления положительного импульса (график i_t) имеет место положительный выброс на графике первых разностей $\nabla_t = z_t - z_{t-1}$, вслед за которым наблюдается отклонение в обратную сторону. Аналогичным образом при отрицательном импульсе на графике первых разностей появляется сначала отрицательный выброс, а затем противоположное отклонение. Таким образом, каждый импульс i_t ведет к появлению двух точек, отклоняющихся от более или менее плавной траектории графика первых разностей, что в принципе может быть использовано в качестве критерия для обнаружения аномальных наблюдений (далее он будет именоваться вспомогательным). Однако на практике более эффективным является иной способ такого обнаружения. Он основан на следующих рассуждениях.

Очевидно, что при выполнении предпосылки б) о гауссовости процесса $x(t)$ приращения $\nabla_{t\,x}$ также подчиняются гауссовскому распределению со следующими параметрами:

математическое ожидание $m_{\nabla x} = 0$;

$$\text{дисперсия } \sigma_{\nabla x}^2 = 2\sigma_X^2[1 - \rho_{XX}(1)], \qquad (3)$$

где σ_X^2 — дисперсия гауссовской последовательности; $\rho_{XX}(k)$ — значение ее нормированной корреляционной функции для дискретного сдвига $k = 1$.

Несложно найти и значение нормированной корреляционной функции приращений $\rho_{\nabla x}(k)$, которое для сдвига $k = 1$ оказывается равным

$$\rho_{\nabla x}(1) = \frac{2\rho_{XX}(1) - 1 - \rho_{XX}(2)}{2[1 - \rho_{XX}(1)]}. \qquad (4)$$

Зная указанные параметры, можно построить на плоскости $(\nabla_{t\,x}, \nabla_{t+1,\,x})$ эллипс рассеяния приращений с заданной доверительной вероятностью P, содержащий экспериментальные точки, когда имеет

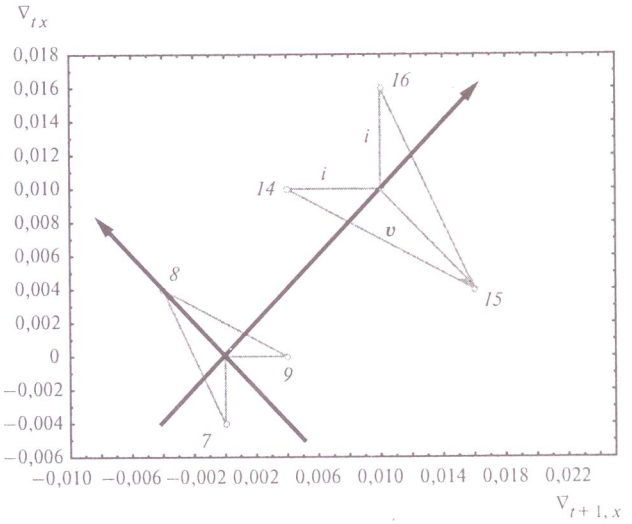


Рис. 2

место только процесс $x(t)$. Границы такого эллипса определяются соотношением

$$\frac{1}{\sigma^2_{\nabla x}\left[1-\rho^2_{\nabla x}(1)\right]}\times$$

$$\times\left[\nabla^2_{tx}-2\rho_{\nabla x}(1)\nabla_{tx}\nabla_{t+1,x}+\nabla^2_{t+1,x}\right]=\chi^2_P(2), \quad (5)$$

где $\chi^2_P(2)$ — P-квантиль χ^2-распределения с двумя степенями свободы [2].

Поскольку значение $\rho_{\nabla x}(1)$ практически всегда положительно, экспериментальные точки концентрируются вдоль главной оси эллипса, наклоненной под углом 45° относительно оси ординат.

Рассмотрим теперь отдельно поле рассеяния точек, соответствующих импульсам $i(t)$. На рис. 2 представлены данные чисто иллюстративного примера появления двух импульсов разной полярности и амплитуды в дискретные моменты времени $t = 8$ и $t = 15$. Первый из этих импульсов появляется на фоне постоянного значения процесса ∇_{tx} вблизи данного значения t, второй — на фоне линейного возрастания значений ∇_{tx}.

И в том, и в другом случае поле рассеяния точек в окрестностях выброса на плоскости $(\nabla_{tx}, \nabla_{t+1,x})$ имеет характерную стреловидную форму с основанием, находящимся либо в начале координат, либо на наклонной прямой, проходящей через начало координат под углом 45° относительно оси ординат; при этом направление отклонения точки основания и ее расстояние от начала координат зависят от наклона фоновой составляющей. Вектор $\boldsymbol{v}$, соответствующий импульсу i_t, ортогонален указанной прямой и имеет длину $|\boldsymbol{v}_t| = v_t = \sqrt{2}|i_t|$. На рис. 2 положительному выбросу соответствует точка *8*, а отрицательному — точка *15*.

Каждую из точек выбросов сопровождают две точки отклонений в противоположную сторону (точки *7* и *9* для выброса в точке *8*; точки *14* и *16* для выброса в точке *15*). В дальнейшем будем именовать их сателлитными точками. Их отклонение от указанной ранее прямой в ортогональном к ней направлении равно $v_{ct} = i_t/\sqrt{2}$.

Если в качестве фоновой составляющей рассмотреть кривые, описываемые полиномами второй, третьей и т.д. степени, то основания стреловидных фигур будут смещаться относительно выделенной прямой. Однако общая стреловидная конфигурация расположения точек, связанных с появлением импульсов i_t, при этом сохраняется.

Рассмотрим теперь общий случай появления импульсов на фоне случайного процесса $x(t)$. Общее поле рассеяния будет тогда представлять собой наложение двух рассмотренных выше полей, причем упомянутая ранее наклонная прямая совпадет с главной осью эллипса рассеяния. Отсюда следует *критерий обнаружения* аномальных наблюдений алгоритма 1: *аномальными считаются точки, выходящие за границу эллипса рассеяния* (5). Ясно, что при этом желательно использовать достаточно высокий уровень доверительной вероятности P (например, 0,990÷0,999) с тем, чтобы снизить риск отнесения к аномальным точек, не являющихся таковыми. При этом, конечно, существует и другая опасность: отнесение к аномальным и некоторых сателлитных точек.

В тех случаях, когда априори известно, что импульсы i_t могут иметь только один знак (односторонний вариант), эта информация может быть использована для отсева сателлитных точек. Если такой информации нет, то для прояснения ситуации приходится анализировать непосредственно реализацию процесса первых разностей с использованием вспомогательного критерия. Такой анализ следует проводить только в окрестностях точек, вызывающих сомнение. Поводом для такого сомнения является выход за границу эллипса рассеяния сразу нескольких следующих подряд точек, что недопустимо при выполнении предпосылки г). Выполнение этой предпосылки гарантирует однозначное разделение точек возможной аномальности и сателлитных точек.

В целом данный алгоритм выделения аномальных наблюдений включает в себя следующую последовательность операций:

а) вычисление приращений $\nabla_t = z_t - z_{t-1}$ временного ряда $z_1, z_2, \ldots, z_N$, $t = \overline{2, N}$;

б) оценивание дисперсии ряда приращений σ^2_∇ процесса z_t и его нормированной корреляционной функции для единичного сдвига $\rho_{\nabla\nabla}(1)$;

в) выбор доверительной вероятности P (обычно $P = 0{,}990 \div 0{,}999$);

г) выделение аномальных наблюдений: точка z_t относится к категории аномальных, если имеет место неравенство

$$\frac{1}{\sigma_\nabla^2\left[1-\rho_{\nabla\nabla}^2(1)\right]}\times$$

$$\times\left[\nabla_t^2-2\rho_{\nabla\nabla}(1)\nabla_t\nabla_{t+1}+\nabla_{t+1}^2\right]>\chi_P^2(2)\,; \qquad (6)$$

д) анализ выделенных точек на предмет исключения ложно выделенных точек (устранение сателлитных точек путем использования вспомогательного критерия или некоторых иных точек, противоречащих априорной информации о возможной полярности выбросов); в конечном итоге остаются только те точки, которые достаточно надежно идентифицируются как аномальные;

е) исключение обнаруженных аномальных наблюдений из временного ряда $z_1, z_2, \ldots, z_N$: выделенные точки заменяются на новые значения, которые определяются с помощью интерполяции по соседним точкам (как правило, достаточной является линейная интерполяция по двум соседним точкам).

Описанная процедура обнаружения аномальных наблюдений носит приближенный характер. Это обусловлено следующими причинами.

Во-первых, на практике вместо теоретических значений при построении эллипса рассеяния приходится использовать соответствующие оценки, найденные по выборке ограниченной длины.

Во-вторых, поскольку вычисление этих оценок производится по всем элементам выборки, появляется некоторое смещение, обусловленное аномальными наблюдениями, если таковые имеются. В этой связи можно ожидать, что предполагаемая процедура обнаружения аномальных наблюдений будет давать приемлемые результаты для сравнительно больших выборок и при весьма малом возможном числе аномальных наблюдений, когда их количество составляет единицы процентов объема выборки N. Именно в этом случае различие в характеристиках процессов z_t и x_t будет невелико.

Для улучшения качества метода предлагается придать ему итерационный характер.

Первая итерация реализуется с помощью указанной выше процедуры. Вторая итерация осуществляется аналогично первой, но с использованием скорректированных по п. е) значений временного ряда. При этом новый эллипс рассеяния будет более узким, а выделенные ранее точки уже не будут выходить за его границу. Зато могут выявиться новые аномальные наблюдения.

После каждой очередной итерации необходимо зафиксировать положение точек, выделенных на всех итерациях. При этом вновь может быть обнаружено появление сразу нескольких следующих подряд выделенных точек, что требует дополнительного исследования с помощью вспомогательного критерия в окрестностях такого рода нарушений.

Итерационный процесс может быть продолжен до тех пор, пока на очередной итерации не будет выявлено ни одного нового аномального наблюдения или их количество резко уменьшится по сравнению с предыдущими итерациями.

Представляет интерес вопрос о чувствительности предложенной процедуры. Очевидно, что ее практическая ценность невелика, если с ее помощью можно выявлять только такие аномальные точки, которые и без того легко обнаружить чисто визуально по осциллограмме процесса. В первом приближении ответить на данный вопрос можно с помощью следующих рассуждений.

Как несложно уяснить, наибольшая критическая длина вектора $\boldsymbol{v}$ соответствует его размещению на малой оси эллипса рассеяния. Известно, что длина этой полуоси:

$$B = \sigma_{\nabla x}\sqrt{\chi_P^2(2)[1-\rho_{\nabla x}(1)]} =$$

$$= \sigma_{\nabla x}\sqrt{\chi_P^2(2)\frac{3-4\rho_{XX}(1)+\rho_{XX}(2)}{2[1-\rho_{XX}(1)]}}\,. \qquad (7)$$

Отсюда с учетом (3) следует

$$B = \sigma_X\sqrt{2[1-\rho_{XX}(1)]}\times$$

$$\times\sqrt{\chi_P^2(2)\frac{3-4\rho_{XX}(1)+\rho_{XX}(2)}{2[1-\rho_{XX}(1)]}} =$$

$$= \sigma_X\sqrt{\chi_P^2(2)[3-4\rho_{XX}(1)+\rho_{XX}(2)]}\,,$$

и, следовательно, критическая длина вектора $\boldsymbol{v}$ равна

$$v_t = \sqrt{2}\left|i_t\right| = B =$$

$$= \sigma_X\sqrt{\chi_P^2(2)[3-4\rho_{XX}(1)+\rho_{XX}(2)]}\,. \qquad (8)$$

Тогда чувствительность первого алгоритма может быть охарактеризована относительным значением амплитуды обнаруживаемого в наихудшем случае импульса:

$$\delta^{(1)} = \frac{\left|i_t\right|}{\sigma_X} = \frac{B}{\sigma_X\sqrt{2}} =$$

$$= \sqrt{\frac{1}{2}\chi_P^2(2)[3-4\rho_{XX}(1)+\rho_{XX}(2)]}\,. \qquad (9)$$

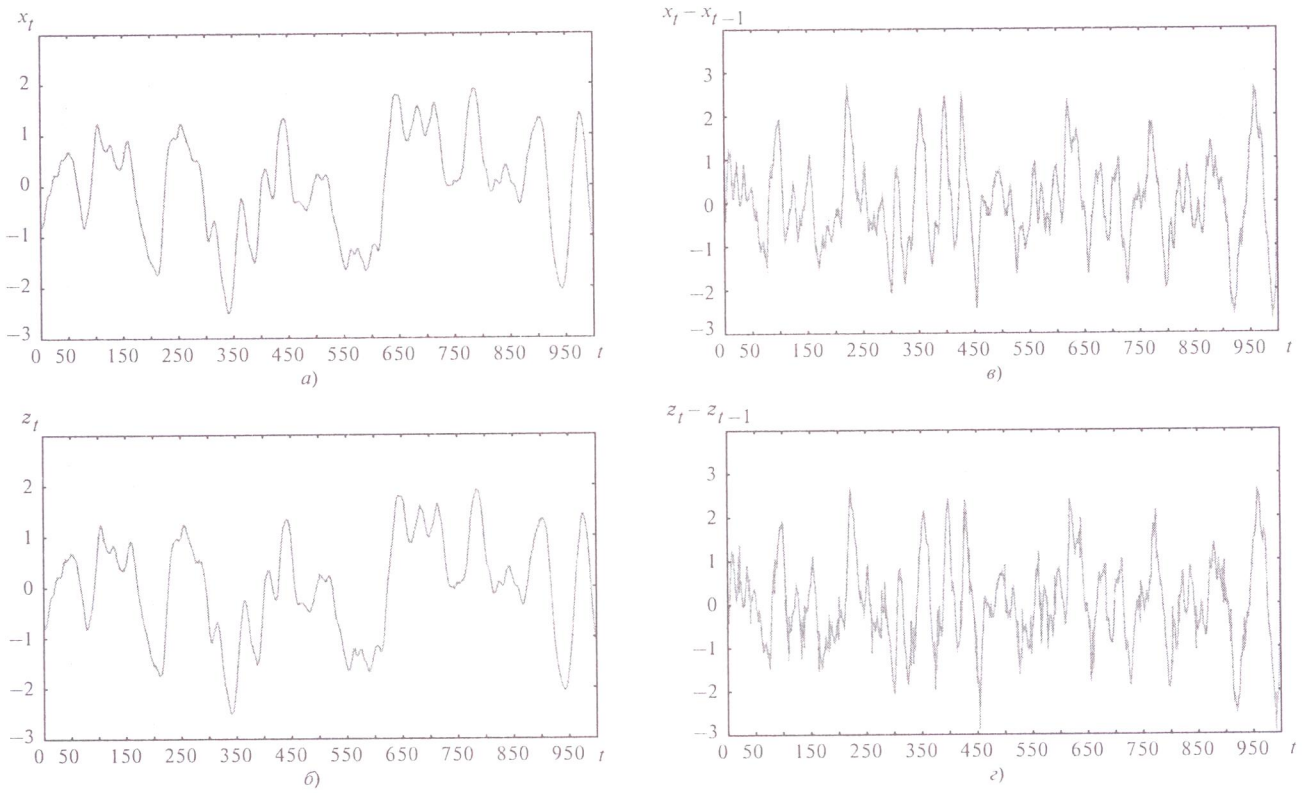


Рис. 3

Формула (9) после некоторых преобразований может быть приближенно представлена в иной, хорошо интерпретируемой форме:

$$\delta^{(1)} \approx \frac{\pi^2}{2}\sqrt{\chi_P^2(2)}\,\hat{\bar{N}}_0\hat{\bar{N}}_{\text{extr}}, \qquad (10)$$

где $\hat{\bar{N}}_0 = N_0/N$, $\hat{\bar{N}}_{\text{extr}} = N_{\text{extr}}/N$; N_0 и N_{extr} — соответственно число пересечений нулевого уровня центрированной реализацией наблюдаемого процесса длиной N и число его экстремумов (пересечение нулевого уровня реализацией процесса первых разностей). Ясно, что чем более гладкой является траектория наблюдаемого случайного процесса и, значит, чем меньше частота появления нулей и экстремумов, тем меньшие значения выбросов, нарушающих гладкость этой траектории, могут быть обнаружены.

Оценка показателя чувствительности $\delta^{(1)}$ может являться заключительным элементом процедуры выделения аномальных наблюдений. С этой целью используют реализацию процесса с исключенными аномальными точками, считая что в ней присутствует только компонента x_t. Именно для этой реализации находят оценки N_0 и N_{extr}, а затем с помощью соотношения (10) и значение $\delta^{(1)}$. Возможен и альтернативный вариант, основанный на формулах (7), (9), в которых вместо теоретических значений σ_X, σ_∇, $\rho_{\nabla\nabla}(1)$ используются их оценки $\hat{\sigma}_X$, $\hat{\sigma}_\nabla$, $\hat{\rho}_{\nabla\nabla}(1)$.

Определение показателя чувствительности может осуществляться и на начальном этапе (на первой итерации) с тем, чтобы получить хотя бы общее представление о том, насколько грубые отклонения могут быть обнаружены.

Проиллюстрируем сказанное модельным примером.

Исходный случайный процесс x_t представляет собой дискретный процесс, полученный путем пропускания гауссовского белого шума с нулевым математическим ожиданием через линейную динамическую систему 3-го порядка. Дисперсия процесса $\sigma_X^2 = 1$. Реализация данного процесса представлена на рис. 3, *а*. Аномальные наблюдения имитируются путем добавления к исходному процессу импульсов с амплитудой $i = \pm 0{,}035$, появляющихся в случайные моменты времени. В результате получаем наблюдаемый процесс $z_t = x_t + i_t$ (рис. 3, *б*). Общее число наблюдений $N = 1000$; количество аномальных наблюдений $N_a = 50$.

В силу малости введенных искажений их визуально обнаружить невозможно: реализации практически неотличимы. Мало что дает и сравнение кривых первых разностей, представленных (в нормированном виде) на рис. 3, *в* (для исходного процесса) и рис. 3, *г* (для процесса с аномальными наблюдениями).

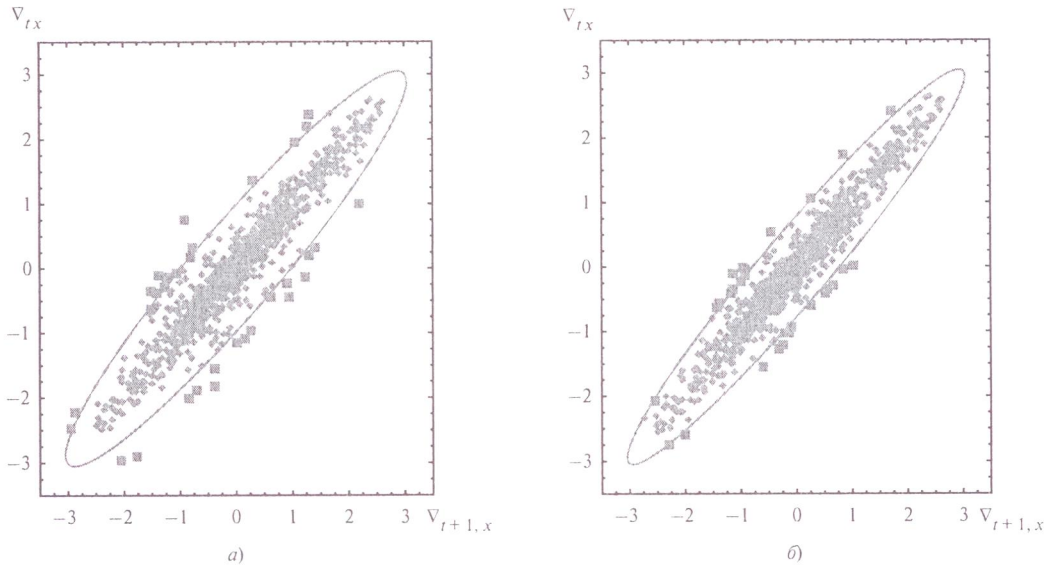


Рис. 4

Применение рассмотренного выше алгоритма дало следующие результаты. При использовании только основного критерия:

на первой итерации выделено 32 аномальных наблюдения (рис 4, *а*, — выделенные точки находятся вне эллипса рассеяния);

на второй итерации выделено 26 точек (рис. 4, *б*);

третьей итерации не потребовалось.

В результате выделено 58 точек; из них 44 выделены правильно, 6 аномальных точек не обнаружены; 12 точек ошибочно отнесены к разряду аномальных. Учитывая, что при построении эллипса рассеяния использовалось значение доверительной вероятности $P = 0{,}990$, последняя ошибка не представляется чрезмерной. Среди 58 выделенных точек обнаружено семь групп, содержащих точки, не разделенные единичным промежутком. Использование вспомогательного критерия позволило устранить это явление. В конечном итоге к числу аномальных была отнесена 51 точка с правильной локализацией 44 из них. Найденный после второй итерации показатель чувствительности $\delta^{(1)}$ оказался равным 0,027.

Алгоритм 2-го порядка

Данный алгоритм основан на идее выявления статистически значимого отклонения наблюдаемого значения z_t от точки z_t^*, получаемой путем линейной интерполяции по двум соседним точкам z_{t-1} и z_{t+1}.

Рассмотрим случайную переменную y_t:

$$y_t = z_t - z_t^* = z_t - \frac{z_{t+1} + z_{t-1}}{2}. \qquad (11)$$

Данная переменная при отсутствии аномальных наблюдений подчиняется гауссовскому распределению с нулевым математическим ожиданием и дисперсией

$$\sigma_y^2 = M\{y_t^2\} = \frac{1}{2}\sigma_X^2[3 - 4\rho_{XX}(1) + \rho_{XX}(2)]. \qquad (12)$$

Тогда наличие аномального наблюдения в точке z_t может быть установлено путем обнаружения значимого отклонения y_t от нуля, т.е. когда

$$|y_t| > u_{1-P}\sigma_y, \qquad (13)$$

здесь u_{1-P} — квантиль гауссовского распределения, соответствующий доверительной вероятности P.

Из (12), (13) непосредственно следует, что чувствительность данного алгоритма

$$\delta^{(2)} = u_{1-P}\left\{\frac{1}{2}[3 - 4\rho_{XX}(1) + \rho_{XX}(2)]\right\}^{1/2} \approx$$

$$\approx \frac{\pi^2}{2} u_{1-P}\hat{\bar{N}}_0\hat{\bar{N}}_{\text{extr}}. \qquad (14)$$

На практике нет необходимости вычислять значение σ_y^2 по формуле (12) и использовать затем соотношение (13). Существует более простой путь. Он становится очевидным, если представить y_t в другой форме:

$$y_t = z_t - \frac{z_{t+1} + z_{t-1}}{2} = \frac{(z_t - z_{t+1}) + (z_t - z_{t-1})}{2} =$$

$$= \frac{-\nabla_{t+1} + \nabla_t}{2} = \frac{-\nabla_{t+1}^{(2)}}{2}, \qquad (15)$$

где $\nabla_{t+1}^{(2)}$ — приращение 2-го порядка для момента времени $t + 1$. Тогда вместо (13) можно записать эквивалентную формулу, основанную на использовании приращений 2-го порядка (отсюда и название алгоритма):

$$\left|\nabla_{t+1}^{(2)}\right| > u_{1-P}\sigma_{\nabla^{(2)}}. \qquad (16)$$

В общем виде данный алгоритм выделения аномальных наблюдений включает в себя следующую последовательность операций:

а) вычисление приращений $\nabla_t = z_t - z_{t-1}$ временного ряда $z_1, z_2, \ldots, z_N$, $t = \overline{2, N}$;

б) вычисление приращений 2-го порядка $\nabla_t^{(2)} = \nabla_t - \nabla_{t-1}$, $t = \overline{3, N}$;

в) оценивание дисперсии ряда приращений 2-го порядка $\sigma_{\nabla^{(2)}}$;

г) выбор доверительной вероятности P (обычно $P = 0{,}950 \div 0{,}999$);

д) выделение аномальных наблюдений: точка z_t относится к категории аномальных, если имеет место неравенство (16);

е) исключение обнаруженных аномальных наблюдений из временного ряда $z_1, z_2, \ldots, z_N$: выделенные точки заменяются на новые значения, которые определяются, как и в предыдущем алгоритме, с помощью интерполяции по соседним точкам (например, путем линейной интерполяции по двум соседним точкам);

ж) выполнение следующей итерации для скорректированных по п. е) значений временного ряда и т.д. до тех пор, пока новые аномальные точки перестанут выделяться.

Данный алгоритм был использован для анализа того же модельного примера. Уже на первой итерации алгоритм обнаружил 53 аномальные точки, локализация 45 из них правильная; 8 точек ложные, 5 аномалий не выявлены. Среди 53 выделенных точек обнаружены 3 группы, содержащие точки, не разделенные единичным промежутком. Использование вспомогательного критерия позволило уменьшить количество ложно выделенных точек до пяти. Второй итерации в данном случае не потребовалось.

Доверительная вероятность P	0,990	0,995	0,999	0,9995
$\delta^{(2)}/\delta^{(1)}$	0,775	0,793	0,831	0,844

Очевидно, что второй алгоритм оказался более эффективным. Это может быть объяснено его большей чувствительностью. Действительно, из сопоставления формул (9), (14) следует:

$$\delta^{(2)}/\delta^{(1)} = u_{1-P}/\sqrt{\chi_P^2}. \qquad (17)$$

Это отношение всегда меньше единицы, что наглядно иллюстрируется данными таблицы.

Более детальный анализ показывает, что относительная эффективность алгоритма 2-го порядка повышается по мере увеличения гладкости траектории наблюдаемого случайного процесса и снижается с увеличением доли аномальных наблюдений в выборке.

Заключение

Отметим, что проблема выделения и устранения влияния аномальных наблюдений достаточно деликатна. По сути требуется сугубо индивидуальное рассмотрение каждого такого наблюдения с тем, чтобы исключая часть наблюдений как аномальные, не потерять ценную информацию об особенностях и тонких свойствах изучаемого процесса.

Возможна и такая ситуация, когда полезная информация содержится именно в выбросах. В качестве примера можно привести задачу анализа маломощных импульсных потоков на фоне относительно интенсивной помехи, представляющей собой коррелированный случайный процесс. Такая задача имеет место, например, при анализе нарушений ритма сердца.

Литература

1. **Введение** в теорию порядковых статистик: пер. англ. / под ред. А.Я. Боярского. М.: Статистика, 1970.

2. **Корн Г., Корн Т.** Справочник по математике для научных работников и инженеров. М.: Наука, 1968.