

УДК 531.19+53.088+519.25

## СТАТИСТИКА ДРОБНЫХ МОМЕНТОВ: НОВЫЙ МЕТОД КОЛИЧЕСТВЕННОГО «ПРОЧТЕНИЯ» ПРОИЗВОЛЬНОЙ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

*Р.Р. Нигматуллин*

### Аннотация

Найден статистический смысл моментов *целого*  $\Delta_N^{(p)}$  ( $p = 1, 2, \dots$ ) и *дробного*  $\Delta_N^{(p)}$  ( $0 < p < \infty$ ) порядков, рассчитанных для некоторой случайной последовательности, содержащей  $N$  произвольных точек. Моменты высших порядков позволяют свести анализируемую случайную последовательность к некоторому конечному набору  $k$  статистически устойчивых целых моментов  $\Delta_k^{(p)}$  ( $p = 1, 2, \dots, k$ ), принадлежащих исходной последовательности. Найденные условия статистической устойчивости и близости, выраженные в терминах высших моментов  $\Delta_N^{(p)} = \Delta_{N+k}^{(p)}$  ( $p = 1, 2, \dots, k$ ), позволяют найти  $k$  неизвестных устойчивых точек и предсказать возможное будущее (устойчивое по отношению к прошлому времени) поведение анализируемой случайной последовательности. Функция обобщенного среднего (ФОС), определяемая как  $G_N^{(p)} = (\Delta_N^{(p)})^{1/p}$ , может быть эффективно использована при анализе статистически близких случайных последовательностей, содержащих большое число измеренных точек ( $N \gg 1$ ). Найденны приближенные аналитические выражения для ФОС  $G_N^{(p)}$  для произвольных значений  $p$  из интервала  $(-\infty < p < \infty)$ . Они дают возможность подогнать *произвольную* случайную последовательность, преобразованную в пространстве моментов в детерминированную ФОС, и выразить *количественно* исходную случайную последовательность в терминах некоего «универсального» набора редуцированных (подгоночных) параметров, входящих в приближенное аналитическое выражение для ФОС. Эти подгоночные параметры могут быть использованы для построения так называемой *калибровочной* кривой, когда возникает необходимость сравнения одной случайной последовательности с другой по отношению к изменениям некоторого внешнего доминантного фактора (малого сигнала). Целочисленные моменты легко обобщаются на *дробные* и даже *комплексные* моменты, которые позволяют ввести определение нецелых моментов и ФОС, содержащие комплексные величины. ФОС может быть также рассмотрена как функция двух и более переменных для анализа многомерных случайных последовательностей, содержащих два, три и более независимых индекса. Следует особо подчеркнуть тот факт, что статистика дробных моментов (СДМ), предлагаемая в этой работе, совершенно *свободна* от каких-либо модельных (априорных) представлений и предположений о природе случайности и поэтому может быть эффективно использована для количественного сравнения произвольных случайных последовательностей, используя для этой цели набор подгоночных параметров, получаемых из сравнения соответствующих ФОС. Найдено соотношение между величиной дробного момента и параметром неэкстенсивности  $q$ , входящим в обобщенное определение энтропии, предложенное К. Цаллисом. Для доказательства сверхчувствительности метода, основанного на статистике дробных моментов (СДМ), рассмотрена важная проблема по защите пластиковых карточек, товарных знаков и других ценных документов от подделок. Некоторые поучительные примеры детектирования сверхслабых ( $S/N = 10^{-2}$ ,  $10^{-3}$ ) сигналов, полученных на модельных данных и буквально «растворенных» в исходной последовательности, показывают высокую эффективность СДМ и могут быть использованы в качестве исходной базы для дальнейших приложений при сравнении

реальных шумов, модифицированных внешним малым фактором. Анализ реальных данных по диэлектрической спектроскопии, реализованный в рамках СДМ, предоставляет уникальную возможность для количественного сравнения каждого отдельного экспериментального измерения и позволяет выразить влияние нейтральной аддитивной добавки с помощью калибровочной кривой без детального знания подгоночной функции, которую невозможно получить для сложных систем в рамках упрощенных моделей.

### Введение

Общая цель любого метода развитого или предлагаемого в современной прикладной статистике заключается в том, чтобы улучшить чувствительность и селективность отношения сигнал/шум и уменьшить возможные систематические ошибки, связанные с временными флуктуациями возможных неконтролируемых параметров измеряемого прибора. Другой важной проблемой, существующей в прикладной математической статистике, является выработка некоего «универсального» языка для сравнения двух случайных последовательностей, имеющих в принципе различное количество  $N_1 \neq N_2$  случайных точек.

В настоящее время существует ряд методов, которые могут быть использованы для решения первой проблемы. Можно отметить анализ параметра Херста для случайных последовательностей, имеющих изначально фрактальное происхождение [1]. Мощный метод, основанный на идее вейвлетов, обобщающий традиционный Фурье-анализ [2–5], широко используется в настоящее время для извлечения слабых сигналов из шумов. Недавно для анализа различных шумовых последовательностей и их качественного сравнения был предложен метод, развитый Р.М. Юльметьевым и его учениками [6–8], основанный на записи уравнений Цванцига – Мори для дискретных случайных последовательностей. Необходимо упомянуть также феноменологический метод флуктуационно-шумовой спектроскопии (ФШС), предложенный С.Ф. Тимашевым с соавторами [9–11], который оказался весьма эффективным для детектирования необычного (аномального) поведения различных случайных последовательностей, таких, как перемежаемость, самоорганизованная критичность, и выявления всплесков и разрывов различного рода.

Однако внимательный анализ современных методов по «прочтению» различных случайных последовательностей показывает, что в настоящее время не существует «универсального» языка для количественного сравнения одной произвольной случайной последовательности с другой, обладающими различными статистическими характеристиками. Иными словами, в настоящее время не существует универсального способа сокращенного описания (редукции) и количественного сравнения для двух и более случайных последовательностей произвольной природы, которые обладают различным или близким статистическим «рисунком». Более того, все перечисленные выше методы обладают малой чувствительностью к детектированию малого сигнала с амплитудой ( $S$ ) по отношению к величине уровня шума ( $N$ ), особенно в случаях, когда  $S/N \ll 1$ .

Совсем недавно автором данной работы были предложены новые статистические методы, основанные на идее квантованных гистограмм [12, 13] и построении упорядоченных флуктуационных вилок (разбросов) [14, 15]. Эти новые методы доказали свою высокую эффективность при детектировании слабых сигналов, которые полностью скрыты внутри анализируемой случайной последовательности. Единственным недостатком этих новых методов является тот факт, что они являются неуниверсальными и «работают» достаточно избирательно, причем причины такой избирательности зависят от скрытого в случайной последовательности тренда, который сильно меняет статистический «рисунок» исходной последовательно-

сти и является неконтролируемым фактором, препятствующим детектированию слабого сигнала. Чтобы выработать более универсальные методы по детектированию сверхслабых сигналов ( $S/N \ll 1$ ), необходимо вернуться к «азам» исходной математической статистики и постараться выработать методы, которые не «чувствуют» наличие тренда и могут быть применены в ситуациях, когда этот тренд присутствует или полностью отсутствует.

На первый взгляд может показаться, что такая задача неразрешима, так как каждая случайная последовательность имеет свой неповторимый статистический рисунок, и поэтому любые попытки найти универсальный количественный и редуцированный язык для её описания бессмысленны и заведомо обречены на провал. Однако внимательный анализ существующих методов позволяет сформулировать следующий парадоксальный вопрос: *можно ли произвольную случайную последовательность преобразовать в детерминированную гладкую кривую, выбрав в качестве независимой переменной такой контролируемый параметр, по отношению к которому исходная случайность становится детерминированной?* При этом предполагается, что детерминированная кривая может быть подогнана под некоторую аналитическую функцию и представлена количественно неким набором подгоночных параметров.

Чтобы получить аргументированный и положительный ответ на этот вопрос, необходимо внимательно проанализировать простейшие методы, принятые для анализа случайных последовательностей в традиционной математической статистике. Если быть достаточно строгим по отношению к поиску «универсального» языка, пригодного для описания произвольной случайной последовательности без каких-либо априорных предположений о её статистической природе, то можно сказать, что зачатки такого «универсального» языка в статистике существуют с самого начала её становления. Автор имеет в виду, прежде всего, традиционное использование среднего арифметического и дисперсии (т. е. первого и второго моментов). Но когда такое сравнение в терминах среднего и дисперсии оказывается явно недостаточным, логично было пойти дальше и использовать целочисленные моменты более высоких порядков. Но в силу каких-то необъяснимых причин набор целочисленных моментов более высоких порядков (строго говоря, начиная с пятого) в современной статистике оказался невостребованным.

В этой статье предлагаем потенциальному читателю познакомиться с основами статистики дробных моментов (СДМ), включающей в себя весь набор целочисленных моментов более высоких порядков в качестве частного случая. СДМ, используемая в качестве независимой переменной индекс момента, оказывается весьма эффективной при детектировании сверхслабых сигналов, полностью скрытых в исходном шуме, но каким-то образом модифицирующих исходный шумовой тренд. СДМ является универсальной и совершенно не зависит от априорных предположений, накладываемых обычно на исходный шум, включает в себя традиционную статистику, основанную на 4-х целочисленных моментах, и предоставляет уникальную возможность по унифицированному и редуцированному количественному сравнению произвольной пары случайных последовательностей, имеющих в принципе различное число случайных точек ( $N_1 \neq N_2$ ). Случайные отличия или искажения могут быть выражены количественно в терминах подгоночных параметров для детерминированной функции обобщенного среднего (ФОС), вычисленной в пространстве дробных моментов. ФОС определяется соотношением  $G_N^{(p)} = \left(\Delta_N^{(p)}\right)^{1/p}$ , где  $\Delta_N^{(p)}$  определяет абсолютное значение момента порядка  $p$  для последовательности, образованной из  $N$  точек. Подгоночные параметры ФОС, найденные с помощью метода собственных координат (СК), могут быть использованы в качестве

универсального и редуцированного языка для сравнения произвольных случайных последовательностей различного объёма друг с другом. Статистика дробных моментов совместно с ФОС помогает найти универсальную калибровочную кривую, когда некоторые подгоночные параметры имеют монотонную зависимость по отношению к некоторому доминантному внешнему фактору. Этот фактор слегка возмущает исходную последовательность, которая предполагается свободной от присутствия внешнего «сигнала». Внешний фактор или «сигнал» может быть ассоциирован с концентрацией растворенной малой примеси, температурой, влажностью и любым другим внешним контролируемым фактором.

Практическое применение СДМ и ФОС было недавно реализовано в работе [16], где калибровочная кривая, построенная с помощью нового метода, позволила повысить чувствительность стандартного лабораторного ИК спектрометра более чем на порядок величины по отношению к детектированию малой концентрации растворенной примеси. Как уже упоминалось ранее, в традиционной статистике высшие моменты практически не используются, хотя для определенных приложений моменты третьего и четвертого порядков, выраженные в терминах асимметрии и эксцесса, оказываются востребованными [17, 18].

В этой статье хотим показать новые потенциальные возможности СДМ для «количественного» прочтения произвольных случайных последовательностей. Эффективность СДМ основана на применении ФОС, которая, в свою очередь, может быть аппроксимирована с высокой точностью некоторой аналитической функцией, содержащей конечное число подгоночных параметров. Следует отметить, что выбор этой функции *не* является произвольным; её вид следует из самой структуры построения ФОС. Эти подгоночные параметры можно рассматривать как своеобразные «отпечатки пальцев», которые позволяют количественно сравнить в терминах этих параметров произвольные последовательности друг с другом. СДМ свободна от каких-либо модельных (априорных) представлений о природе шума и поэтому методы, основанные на СДМ, могут быть с успехом использованы для извлечения слабых сигналов и построения искомым калибровочных кривых для выборок различного объёма. Увеличивая количество подгоночных параметров, всегда можно (если в этом есть необходимость) найти количественные отличия между двумя случайностями. Если эти различия несущественны, то всегда можно определить доверительный интервал, попадая в который две сравниваемые последовательности становятся уже «неразличимыми». Ниже будет показано, как «работает» СДМ для малых выборок, различающихся между собой всего на одну(!) точку. Для таких последовательностей традиционные методы статистики становятся уже неприменимыми. Интересно отметить также, что СДМ тесно связана с параметром неэкстенсивности  $q$ , входящим в определение обобщенной энтропии по Цаллису. В этом подходе обобщенная энтропия получает новое истолкование и может быть также использована для сравнения различных последовательностей.

Содержание статьи разбито на следующие разделы. В разд. 1 рассматриваются вопросы предсказуемости и стабильности случайных последовательностей, выраженные в терминах целочисленных моментов высоких порядков. В разд. 2 показано, как найти подгоночные параметры ФОС с помощью метода СК. В том же разделе даны определения *комплексных моментов* и ФОС на случай многомерных (имеющих более одного индекса) случайных последовательностей. В разд. 3 показана связь между ФОС и обобщенной энтропией по Цаллису. Это соотношение расширяет возможности неэкстенсивной статистической механики по истолкованию стохастических свойств различных последовательностей. Важное практическое приложение по статистической защите пластиковых карточек и других ценных документов от подделок рассмотрено схематически в разд. 4. Некоторые

модельные эксперименты, демонстрирующие мощь СДМ в извлечении сверхслабых сигналов ( $S/N = 10^{-2}, 10^{-3}$ ) и в детектировании статистической близости различных распределений, рассмотрены в разд. 5. Некоторые проблемы, основанные на анализе реальных диэлектрических данных и построении искомым калибровочных кривых без конкретных знаний о природе физической модели, рассмотрены в разд. 6. Основные результаты и выводы статьи приводятся в разд. 7.

### 1. Оценка статистической устойчивости случайных последовательностей, основанная на целочисленных моментах произвольного порядка

Рассмотрим некий отрезок случайной последовательности образованной системой  $j = 1, 2, \dots, N$  точек, локализованной в полосе  $y_{\min} \leq \{y_j\} \leq y_{\max}$ . Для некоторого множества  $N$  случайных точек абсолютный момент порядка  $p$  определяется выражением:

$$\Delta_N^{(p)} = \frac{1}{N} \sum_{j=1}^N y_j^p. \quad (1)$$

Основная проблема может быть поставлена следующим образом. Пусть основная случайная последовательность содержит  $N$  точек. К этому набору точек добавляется другой случайный набор  $k$  точек. Спрашивается, какому условию должен удовлетворять случайный набор  $k$  вновь добавленных точек, чтобы сохранить неизменными величины первых  $k$  моментов, изначально принадлежащих исходной последовательности?

Математически это условие можно записать в виде

$$\Delta_k^{(p)} \equiv \frac{1}{k} \sum_{s=1}^k y_{N+s}^p = \frac{1}{N} \sum_{j=1}^N y_j^p \equiv \Delta_N^{(p)}, \quad (2)$$

где  $p = 1, 2, \dots, k$ .

Решение этой проблемы может помочь в *предсказании* статистической устойчивости исходной последовательности. В Приложении условие (2) и соответствующее решение рассмотрены более детально. Условие (2) эквивалентно решению соответствующей нелинейной системы уравнений для отыскания искомого набора статистически устойчивых  $k$  точек

$$\frac{y_{N+1}^p + y_{N+2}^p + \dots + y_{N+k}^p}{k} = \Delta_N^{(p)}. \quad (3)$$

Здесь  $p = 1, 2, \dots, k$ . Используя рекуррентное соотношение Ньютона [19]

$$p a_p + a_{p-1} S_1 + a_{p-2} S_2 + \dots + a_0 S_p = 0, \quad a_0 = 1, \quad (4)$$

где  $S_p \equiv \sum_{j=1}^k y_{N+j}^p$ .

Система уравнений (3) может быть сведена к нахождению корней *следующего* характеристического полинома

$$y^k + a_1 y^{k-1} + \dots + a_{k-1} y + a_k = 0 \quad (5)$$

для произвольных  $k \leq N$ . Для  $k \leq 4$  эта система уравнений может быть решена аналитически [20], для  $5 \leq k \leq 100$  – численными методами [21] с помощью стабильных алгоритмов, для  $k > 100$  необходимо разработать специальные методы

с целью увеличения стабильности и точности получаемых корней полинома (5). Поучительно привести решения для одной и двух добавляемых точек.

Для  $k = 1$

$$y_{N+1} = \Delta_N^{(1)}, \quad (6)$$

добавляемая точка должна совпадать со средним арифметическим исходной последовательности.

Для  $k = 2$

$$y_{N+1} = \Delta_N^{(1)} + \sqrt{\Delta_N^{(2)} - \left(\Delta_N^{(1)}\right)^2}, \quad y_{N+2} = \Delta_N^{(1)} - \sqrt{\Delta_N^{(2)} - \left(\Delta_N^{(1)}\right)^2}, \quad (7)$$

две добавляемые точки должны совпадать со значениями среднеквадратичных отклонений, отсчитанных от среднего значения. Поэтому выражения (6) и (7) могут быть истолкованы как условие нахождения *предсказуемых* точек, сохраняющих статистическую устойчивость исходной случайной последовательности.

Для аналитического анализа выражений, соответствующим трем и четырёх добавляемым точкам, удобно использовать выражения для *относительных* моментов, отсчитанных от первого момента  $\Delta_N^{(1)}$ :

$$M_p \equiv \frac{1}{N} \sum_{j=1}^N x_j^p \equiv \frac{1}{N} \sum_{j=1}^N \left(y_j - \Delta_N^{(1)}\right)^p. \quad (8)$$

Основываясь на результатах, полученных в Приложении, легко прийти к выводу, что система уравнений (3) в этом случае упрощается и принимает вид

$$\frac{x_{N+1}^p + x_{N+2}^p + \dots + x_{N+k}^p}{k} = M_p. \quad (9)$$

Учитывая, что  $M_1 = 0$ , легко получить упрощенные выражения для отыскания искомых корней.

Случай трех корней  $r = 1, 2, 3$  ( $k = 3$ )

$$y_{N+r} = \lambda u_r + \Delta_N^{(1)}, \quad \lambda = \frac{2M_3}{3M_2}, \quad (10)$$

$$u^3 - bu - b = 0, \quad b = \frac{27}{8\beta}, \quad \beta = \frac{M_3^2}{M_2^3}.$$

Случай четырех корней:  $r = 1, 2, 3, 4$  ( $k = 4$ )

$$y_{N+r} = \lambda u_r + \Delta_N^{(1)}, \quad \lambda = \frac{2M_3}{3M_2},$$

$$u^4 - bu^2 - bu - c = 0, \quad (11)$$

$$b = \frac{9}{2\beta}, \quad c = \frac{b^2}{4}(\gamma - 2), \quad \beta = \frac{M_3^2}{M_2^3}, \quad \gamma = \frac{M_4}{M_2^2}.$$

Эти выражения допускают дальнейшее аналитическое исследование, так как соответствующие уравнения для безразмерной переменной  $u$  содержат только один ( $b$ ) ( $k = 3$ ) и два ( $b$  и  $c$ ) параметра ( $k = 4$ ).

Для  $k > 4$  величины  $y_{N+k}$  находятся численно. Поэтому поведение  $k$  точек ( $y_{N+1}, y_{N+2}, \dots, y_{N+k}$ ) становится *полностью предсказуемым*, если эти точки

удовлетворяют системе (3). Для того чтобы выразить все моменты в одних и тех же размерных единицах, полезно определить следующее выражение

$$G_N^{(p)} = \left[ \frac{1}{N} \sum_{j=1}^N y_j^p \right]^{1/p} \equiv \left( \Delta_N^{(p)} \right)^{1/p}. \quad (12)$$

Это выражение служит определением функции обобщенного среднего (ФОС) [19], зависящей от величины текущего момента  $p$ . Полезно отметить основные свойства выражения (12). Для  $p = -1$  оно определяет *гармоническое* среднее

$$G_N^{(-1)} = \left[ \frac{1}{N} \sum_{j=1}^N y_j^{-1} \right]^{-1} \equiv \frac{1}{\left( \Delta_N^{(1)} \right)}. \quad (13)$$

Для  $p = 0$  выражение (12) совпадает с *геометрическим* средним

$$\lim_{p \rightarrow 0} G_N^{(p)} = \exp \left[ \frac{1}{N} \sum_{j=1}^N \ln(y_j) \right] = \left[ \prod_{j=1}^N y_j \right]^{1/N} \quad (14)$$

и для  $p = 1$  оно, естественно, совпадает с *арифметическим* средним. Предельные значения определяются из соотношений

$$\begin{aligned} \lim_{p \rightarrow \infty} G_N^{(p)} &= \max(y_1, y_2, \dots, y_N) \equiv \max(y), \\ \lim_{p \rightarrow -\infty} G_N^{(p)} &= \min(y_1, y_2, \dots, y_N) \equiv \min(y). \end{aligned} \quad (15)$$

ФОС  $G_N^{(p)}$  для положительных значений  $p$  является монотонной функцией, т. е.

$$G_N^{(t)} < G_N^{(s)} \quad \text{для } t < s \text{ и } (y_j \neq 0). \quad (16)$$

Если набор  $(y_j)$  является положительным, то выражение (12) можно легко обобщить на любое реальное значение  $p$ . Это обобщение позволяет расширить класс моментов и рассмотреть наряду с целочисленными моментами произвольные дробные моменты индекса  $p \in (0, \infty)$ . Если исходная выборка содержит также отрицательные значения, то в этом случае можно расщепить исходную последовательность на две положительные части в соответствии с определением

$$y_j^{(p)} = \frac{1}{2} (y_j + |y_j|), \quad y_j^{(n)} = -\frac{1}{2} (y_j - |y_j|). \quad (17)$$

Если есть необходимость, то нули, содержащиеся в выражениях (17), могут быть удалены. До этого мы рассмотрели случай, когда  $k$  добавляемых точек были расположены справа от исходной выборки. По аналогии можно рассмотреть ситуацию, когда предсказание статистической устойчивости вызывается точками, расположенными в «прошлом», т. е. слева от исходной выборки. Но анализ уравнения (3) показывает, что можно провести автоматическую редукцию исходной выборки к набору  $k$  статистически устойчивых точек, расположенных *внутри* исходной выборки. Если хотим сохранить систему (3) вновь в качестве инвариантного условия сохранения значений  $k$  целочисленных моментов, то условие нахождения этих точек примет вид

$$\frac{Y_1^p + Y_2^p + \dots + Y_k^p}{k} = \Delta_N^{(p)}, \quad p = 1, 2, \dots, k. \quad (18)$$

Система уравнений (18) формально тождественна системе (3), но с одним *принципиальным* отличием. Она сводит исходную выборку  $(y_1, y_2, \dots, y_N)$  к новому набору статистически устойчивых точек, расположенных *внутри* исходной выборки. В частности, для  $k = 1$  редуцированная точка совпадает со средним арифметическим (6); для  $k = 2$  редуцированные точки – с относительной дисперсией, отсчитанной от среднего значения (см. выражение (7)). Поэтому для  $k = 1, 2$  этот подход подтверждает «интуитивное» сведение исходной выборки к двум статистически устойчивым величинам, выраженным в терминах первого и второго моментов. Для  $k = 3, 4$  уравнения (10) и (11) восстанавливают другие статистически устойчивые точки, которые используют естественное определение меры асимметрии и эксцесса [18].

Для  $k > 4$  можно продолжить поиск возможных статистических различий между двумя сравниваемыми выборками, если набор первых четырёх величин, выраженный в терминах статистически устойчивых точек, оказывается недостаточным для этой цели.

На практике трудно ожидать полного совпадения набора из  $k$  точек, получаемых из эквивалентного условия

$$G_N^{(p)} \equiv [\Delta_N^{(p)}]^{1/p} = [\Delta_k^{(p)}]^{1/p} \equiv G_k^{(p)}, \quad (19)$$

особенно в случаях, когда набор  $k$  достаточно велик ( $k > 100$ ). В этом случае необходимо разработать специальные методы для сравнения двух ФОС.

## 2. Приближенные выражения для ФОС. Дробные и комплексные моменты

Для приближенного решения уравнения (19), когда значения  $k > 100$ , и для более полного сравнения двух различных ФОС  $G_{N_1}^{(p)}$  и  $G_{N_2}^{(p)}$  (когда  $N_{1,2}$  отличны друг от друга) полезно найти аналитическое выражение (подгоночную функцию), которая может подогнать заданную ФОС с приемлемой точностью. Для того чтобы найти искомую аналитическую функцию, необходимо рассмотреть следующее выражение

$$\begin{aligned} F_N^{(p)} &= \frac{1}{N} \sum_{j=1}^N \left( \frac{y_j}{y_{\max}} \right)^p \equiv \frac{1}{N} \sum_{j=1}^N (\theta_j)^p = \\ &= \frac{1}{N} \sum_{j=1}^N \exp[p \ln(\theta_j)] \cong A_0 + \sum_{n=1}^s A_n \exp(-\lambda_n p). \end{aligned} \quad (20)$$

Параметр  $\theta_j = (y_j/y_{\max})$ , выражающий отношение амплитуды  $y_j$  к максимальному значению, всегда меньше единицы. Поэтому выражение (20) может быть записано в виде конечной комбинации экспоненциальных функций. Коэффициент  $A_0$  отражает вклад величин  $y_j$ , которые близки к максимальному значению  $(y_j/y_{\max}) \leq 1$ ; другие коэффициенты  $A_n$  соответствуют вкладу совокупности случайных амплитуд  $y_j$  с отношением  $(y_j/y_{\max}) \ll 1$ . Основываясь на выражении (20), легко записать приближенное выражение для ФОС, определённое формулой (12)

$$G_N^{(p)} = g \left[ 1 + \sum_{n=1}^s a_n \exp(-\lambda_n p) \right]^{1/p} \equiv y_{\max} \cdot (y_s(p))^{1/p}, \quad (21)$$

где  $a_n = A_n/A_0$ . Формула (21) позволяет «прочитать» произвольную случайную последовательность в терминах редуцированных *количественных* параметров



$(a_n, \lambda_n)$ , включая также площадь под кривой (ППК)  $G_N^{(p)}$  и величину  $y_{\max}$ . Для вычисления заданных подгоночных параметров  $(a_n, \lambda_n)$  ( $n = 1, 2, \dots, s$ ) можно использовать метод собственных координат (СК), который позволяет представить функцию (20) в виде совокупности прямых линий и найти значения подгоночных параметров, соответствующих *глобальному* минимуму. Не вдаваясь в детали этого эффективного метода, достаточно полно представленного в работах [23–25], можно записать только основное линейное соотношение (ОЛС) для функции  $y_s(p)$ , содержащей три экспоненциальные функции ( $s = 3$ ). Многочисленные модельные расчёты показывают, что при практическом применении выражения (21) для достижения приемлемой точности число различных экспонент не должно превышать трёх или четырёх. К примеру, ОЛС для трёх-экспоненциальной функции ( $s = 3$ ) может быть записано в виде

$$Y(p) = \sum_{q=1}^6 C_q X_q(p). \quad (22)$$

Здесь

$$\begin{aligned} Y(p) &= F_N^{(p)} - \langle \dots \rangle, \\ X_1(p) &= \int_{p_0=1}^p F_N^{(u)} du - \langle \dots \rangle, \quad C_1 = \lambda_1 + \lambda_2 + \lambda_3, \\ X_2(p) &= \int_{p_0=1}^p (p-u) F_N^{(u)} du - \langle \dots \rangle, \quad C_2 = -(\lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3), \\ X_3(p) &= \frac{1}{2} \int_{p_0=1}^p (p-u)^2 F_N^{(u)} du - \langle \dots \rangle, \quad C_3 = \lambda_1 \lambda_2 \lambda_3. \end{aligned} \quad (23)$$

$$\begin{aligned} X_4(p) &= p^3 - \langle \dots \rangle, \\ X_5(p) &= p^2 - \langle \dots \rangle, \\ X_6(p) &= p - \langle \dots \rangle. \end{aligned} \quad (24)$$

Величины  $C_4, C_5, C_6$  содержат неизвестные значения производных в начальной точке  $y_3^{(r)}(p_0)$  ( $r = 3, 2, 1$ ) и несущественны для расчета искомых корней  $\{\lambda_r\}$  ( $r = 3, 2, 1$ ) методом наименьших квадратов (МНК). Символ  $\langle \dots \rangle$  в (23)–(24) определяет соответствующее арифметическое среднее, которое вычитается из соответствующей функции  $Y(p), X_q(p)$  для того, чтобы обеспечить основное требование  $\langle \varepsilon \rangle = 0$  метода МНК [26]. Неизвестные константы  $A_n$  ( $n = 0, 1, \dots, s$ ) также находятся методом МНК из соотношения (20). Здесь уместно подчеркнуть следующие важные достоинства ФОС, которые могут найти широкие применения при анализе случайных последовательностей различной природы:

а) Приближенное аналитическое выражение (21) обеспечивает «универсальную» количественную редукцию произвольной случайной последовательности к набору параметров  $(A_n, \lambda_n)$ , включая также ППК и величину  $y_{\max}$ .

б) Вычисленные подгоночные параметры позволяют разделить амплитуды случайной последовательности  $y_j$  на некие оптимальные статистические группы (кластеры)  $n$  с параметрами  $(A_n, \lambda_n)$ , что соответствует редуцированному (сокращённому) описанию рассматриваемой случайной последовательности.

в) В случаях, когда объём выборки достаточно велик ( $N > 100$ ), такое редуцированное представление становится достаточно информативным по отношению

к некоторому внешнему фактору по сравнению с численной оценкой  $k$  корней системы уравнений (18). Для этого случая ( $N > 100$ ) целесообразно подогнать и сравнить исходную функцию  $G_N^{(p)}$  с  $G_k^{(p)}$ , входящую в выражение (19) отдельно, а затем сравнить их близость в терминах подгоночных параметров ( $A_n, \lambda_n$ ). Сравнение высших моментов, формирующих две случайные и различные выборки ( $k \neq N$ ), является более точным. При  $k = 1, 2$  получаем общепринятое сокращенное описание, выраженное в терминах среднего арифметического и стандартного отклонения. Но такое «общепринятое» описание, выраженное величинами только целых моментов  $\Delta_1, \Delta_2$ , может оказаться *неудовлетворительным* для статистического разделения случайных последовательностей в большинстве случаев.

г) Численные расчеты показывают, что представление  $G_k^{(p)}$  по отношению к  $G_N^{(p)}$  является весьма информативным при сравнении случайных последовательностей. Если две случайные последовательности статистически близки друг другу (в смысле совпадения совокупности моментов), то зависимость  $G_k^{(p)}$  ( $G_N^{(p)}$ ) в пространстве параметра  $p$  должна представляться прямой линией с наклоном, равным единице, и отсечкой, близкой к нулю. Возможные отклонения от этого простого критерия дают возможность детектировать нечто «странное» (присутствие сигнала или изменения прежнего статистического поведения) в сравниваемых статистических последовательностях. Примеры применения такого удобного представления рассмотрены в разд. 6. Этот простой анализ помогает в детектировании самоподобных (фрактальных) компонент, если они присутствуют в анализируемой случайной последовательности.

Допустим, что существует отрезок некоторой последовательности, амплитуды которого по отношению к сравниваемой последовательности приближенно удовлетворяют следующему равенству

$$\frac{y_j}{u_m} \cong \lambda, \quad (\lambda \neq 1), \quad j, m = l, l+1, \dots, M. \quad (25)$$

Для такого отрезка легко показать, что две ФОС становятся пропорциональными друг другу, т. е.

$$G(y)_{M-l+1}^{(p)} = \lambda G(u)_{M-l+1}^{(p)}. \quad (26)$$

Условие (26) расширяет понятие корреляции, введенное Пирсоном [26], и помогает обнаружить отрезок случайной последовательности, обладающий самоподобным (фрактальным) поведением. Интересно отметить, что самоподобное поведение идентифицируется для многих случайных последовательностей, если они должным образом упорядочиваются и представляются в виде ранжированного графика [15]. Смотри ниже выражение (34), которое может быть использовано в качестве «универсальной» подгоночной функции для распознавания фрактального поведения, скрытого в случайных последовательностях, имеющих различную статистическую природу.

Уравнение (26) позволяет определить статическую близость двух сравниваемых последовательностей. Если две последовательности имеют соответственно  $N_{1,2}$  случайных точек и в пространстве моментов удовлетворяют условию ( $\lambda \cong 1$  и  $b$  – две произвольные константы)

$$G_{N_1}^{(p)} = \lambda G_{N_2}^{(p)} + b, \quad (27)$$

то такие две последовательности могут быть определены как статистически близкие между собой. Когда величина  $\lambda$  принимает произвольное значение и отсечка  $b \cong 0$ , то такие два отрезка могут быть определены как самоподобные (коррелированные) друг другу. Если величины  $\lambda$  и  $b$  начинают зависеть от параметра  $p$ , то

такие две последовательности могут быть определены как статистически различные друг другу.

д) «Универсальное» описание последовательностей с помощью ФОС можно обобщить не только на действительные, но и на комплексные значения величины  $p$ . Такое двумерное обобщение может быть полезным для выявления более тонких различий, присутствующих в одномерных последовательностях. Когда параметр  $p$  выражается комплексным числом  $p = u + iv$ , то ФОС (12) преобразуется в поверхность вида

$$\begin{aligned} G_N^{(p)} &= \operatorname{Re} G_N^{(p)} + i \operatorname{Im} G_N^{(p)} = \left[ \frac{1}{N} \sum_{j=1}^N y_j^{u+iv} \right]^{\frac{1}{u+iv}} = \\ &= [M_c(u, v) + i M_s(u, v)]^{\frac{1}{u+iv}} = \exp \left[ \frac{1}{u+iv} \ln \left( \frac{1}{N} \sum_{j=1}^N \exp [(u+iv) \ln y_j] \right) \right]. \end{aligned} \quad (28)$$

После некоторых алгебраических преобразований действительная и мнимая части ФОС могут быть представлены следующими соотношениями

$$M_c(u, v) = \frac{1}{N} \sum_{j=1}^N y_j^u \cos(v \ln(y_j)), \quad (29)$$

$$M_s(u, v) = \frac{1}{N} \sum_{j=1}^N y_j^u \sin(v \ln(y_j));$$

$$\operatorname{Re} G_N^{(p)}(u, v) = \exp [F(u, v)] \cos [\Phi(u, v)],$$

$$\operatorname{Im} G_N^{(p)}(u, v) = \exp [F(u, v)] \sin [\Phi(u, v)];$$

$$F(u, v) = \frac{u \cdot \ln \left[ \sqrt{G_1^2(u, v) + G_2^2(u, v)} \right]}{u^2 + v^2} + \frac{v \cdot \varphi(u, v)}{u^2 + v^2}, \quad (30)$$

$$\Phi(u, v) = -\frac{v \cdot \ln \left[ \sqrt{G_1^2(u, v) + G_2^2(u, v)} \right]}{u^2 + v^2} + \frac{u \cdot \varphi(u, v)}{u^2 + v^2};$$

$$G_1(u, v) = \ln \left[ \frac{1}{N} \sum_{j=1}^N (y_j)^u \cos(v \cdot \ln(y_j)) \right] = \ln (M_c(u, v)),$$

$$G_2(u, v) = \ln \left[ \frac{1}{N} \sum_{j=1}^N (y_j)^u \sin(v \cdot \ln(y_j)) \right] = \ln (M_s(u, v)), \quad (31)$$

$$\varphi(u, v) = \operatorname{arctg} \left[ \frac{G_2(u, v)}{G_1(u, v)} \right].$$

Последняя группа выражений требует специального исследования, чтобы выявить такие ситуации, когда применение комплексных моментов является оправданным и информативным.

е) Определение (12) допускает дальнейшее обобщение для двумерных последовательностей:

$$G(p, q) = \left[ \frac{1}{N_1 N_2} \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} (Y_{j_1, j_2})^{p+q} \right]^{\frac{1}{p+q}}. \quad (32)$$

Эта поверхность может быть проанализирована аналогично выражению (12), если заменить одномерный параметр  $p$  на  $p+q$ . Для многомерных последовательностей, имеющих  $m$  различных компонент, по аналогии с выражением (32) можно записать следующее выражение

$$G\left(\sum_{i=1}^m p_i\right) = \left[ \frac{1}{N_1 N_2 \cdots N_m} \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_m} (Y_{j_1, j_2, \dots, j_m})^{\sum_{i=1}^m p_i} \right]^{\left(\sum_{i=1}^m p_i\right)^{-1}}. \quad (33)$$

В работе [15] была идентифицирована некая «универсальная» функция, которая описывает распределение случайных точек в широком классе случайных последовательностей. Это распределение соответствует огибающей упорядоченных амплитуд (определяемых как ранжированный график), получаемой для детрендрованных (т. е. не имеющих выраженного тренда) случайных последовательностей и имеющих относительно большую выборку случайных точек ( $N \geq 1000$ ). Были проанализированы множества реальных последовательностей (медицина, метеорология, экономика и пр.) и было показано, что огибающая упорядоченных амплитуд (ранжированная последовательность) описывается функцией

$$y(t) = A_1 t^{\nu_1} + A_2 t^{\nu_2}. \quad (34)$$

Вычисленные подгоночные параметры этой функции ( $A_1(f)$ ,  $A_2(f)$ ,  $\nu_1(f)$ ,  $\nu_2(f)$ ) по отношению к некоему внешнему фактору ( $f$ ) могут быть проанализированы в терминах целых моментов (уравнение 5)) или с помощью ФОС (12). СДМ, основанная на использовании высших (дробных) моментов, может быть использована для построения *калибровочных кривых*, которые применяются для детектирования *значимого* подгоночного параметра, характеризующего произвольную случайную последовательность (содержащую возможный тренд) по отношению к контролируемому *внешнему* фактору (концентрация добавки, величина внешнего поля, температура, давление, рН-фактор и т. д.). Здесь уместно вновь подчеркнуть, что СДМ не использует *никаких* модельных представлений, основанных на традиционных представлениях гауссовской статистики и её модификациях. Необходимо отметить также, что возможные приложения СДМ основаны на соотношении между идентифицированным числом добавленных точек ( $k$ ) и исходным объёмом выборки ( $N$ ). Если число различимых точек ограничено ( $k/N \ll 1$ ), то количественная схема, основанная на вычислении ограниченного числа целых моментов в соответствии с формулами (6)–(11), является вполне приемлемой. Если же необходимо сравнить большие выборки ( $k/N \leq 1$ ,  $N \gg 1$ ), когда значительный набор начальных моментов близок друг к другу, то подход, основанный на приближенном вычислении ФОС (21) с последующим вычислением подгоночных параметров с помощью метода СК, является наиболее предпочтительным. Иногда для построения искомой калибровочной кривой достаточно использовать соотношение (27), а затем взять ППК в качестве количественного параметра, вычисленного для относительной разности  $(G_{N_2}^{(p)} - G_{N_1}^{(p)})/G_{N_1}^{(p)}$ . Примеры такого упрощенного подхода при детектировании внешнего фактора рассмотрены ниже в разд. 5.

### 3. Связь дробных моментов с выражением для неэкстенсивной энтропии

Вернемся вновь к рассмотрению отрезка случайной последовательности как некоего набора  $j = 1, 2, \dots, N$  точек, локализованных в полосе  $y_{\min} \leq \{y_j\} \leq y_{\max}$ . Для этих  $N$  случайно расположенных точек абсолютное значение величины момента порядка  $p$  определяется более общим выражением

$$\Delta_w^{(p)} = \sum_{j=1}^N w_j y_j^p, \quad (35)$$

где значения  $w_j$  – дискретный набор *априорных* вероятностей, удовлетворяющих стандартному условию нормировки

$$\sum_{j=1}^N w_j = 1. \quad (36)$$

ФОС, определённая ранее выражением (12), принимает следующий вид

$$G_w^{(p)} = \left[ \sum_{j=1}^N w_j y_j^p \right]^{1/p} \equiv \left( \Delta_w^{(p)} \right)^{1/p}. \quad (37)$$

Последнее выражение позволяет получить выражение для обобщенной энтропии, предложенной Цаллисом [27–30]

$$S_q = -k_B \frac{1 - \sum_{j=1}^m w_j^q}{1 - q} \equiv -k_B \frac{1 - F_q}{1 - q}. \quad (38)$$

Здесь  $w_j$  определяет некий набор вероятностей, характеризующих часть отрезка случайного ряда, параметр  $q$  определяет величину неэкстенсивного параметра,  $k_B$  – постоянная Больцмана.

Для того чтобы получить изящное выражение для неэкстенсивной энтропии, которое могло быть применимо для описания случайной последовательности *произвольной* природы, необходимо определить более точно выражение для набора вероятностей  $w_j$ , связанных с неким отрезком случайной последовательности. Любая случайная выборка может иметь положительные или отрицательные значения случайных амплитуд, сосредоточенных в полосе  $y_{\min} \leq \{y_j\} \leq y_{\max}$ . Вначале определяем две случайные, но *положительные* последовательности, рассчитанные относительно модуля  $|y_j|$

$$y_j^+ = \frac{1}{2} (y_j + |y_j|), \quad y_j^- = -\frac{1}{2} (y_j - |y_j|). \quad (39)$$

Допустим, что числа точек для этих случайных последовательностей определяются, соответственно, величинами  $N_{\pm}$ , подчиненными условию:  $N_+ + N_- = N$ . Для этих  $N$  случайно локализованных точек вероятность появления случайной амплитуды  $y_j$  может быть естественно определено выражениями

$$w_j^+ = \frac{y_j^{(+)}}{S^{(+)}} , \quad w_j^- = \frac{y_j^{(-)}}{S^{(-)}}, \quad (40)$$

где суммы

$$S^{(\pm)} = \sum_{j=1}^{N_{\pm}} y_j^{(\pm)}$$

выбраны таким образом, чтобы обеспечить нормировку каждой вероятности из (40) на единицу.

Снова вернемся к выражению (38). С учетом априорных вероятностей, определенных выражениями (37), функция  $F_q$  может быть переписана в виде

$$F_q^{(s)} = \sum_{j=1}^m (w_j^{(s)})^q = \sum_{j=1}^m \left( \frac{y_j^{(s)}}{S^{(s)}} \right)^q. \quad (41)$$

Здесь новый индекс  $s$  определяет, соответственно, положительную и отрицательную ( $s = \pm$ ) ветви случайной последовательности. Сравнивая сумму в последнем выражении с определением высших моментов (35) и вероятностями из определения (40), можно установить искомое соотношение между фактором неэкстенсивности  $q$  и величиной дробного момента порядка  $q - 1$ :

$$\begin{aligned} F_q^{(s)} &= \frac{1}{S_{\pm}^{q-1}} \sum_{j=1}^{N_{\pm}} \left( \frac{y_j^{(s)}}{S_{\pm}} \right) (y_j^{(s)})^{q-1} = \\ &= \frac{1}{(S^{(s)})^{q-1}} \sum_{j=1}^{N_{\pm}} w_j^{(s)} (y_j^{(s)})^{q-1} = \frac{1}{(S^{(s)})^{q-1}} \Delta_{w^{(s)}}^{(q-1)} \equiv \Delta_{w^{(s)}}^{(q-1)}. \end{aligned} \quad (42)$$

Последнее выражение справа справедливо для нормированных случайных последовательностей  $y_j^{(s)} \implies y_j^{(s)}/S^{(s)}$ . Используя похожие математические выкладки, приведшие к выражению (20), можно получить следующее выражение

$$\begin{aligned} F_q^{(s)} &= \sum_{j=1}^m (w_j^{(s)})^q = \sum_{j=1}^m w_j \exp \left[ - (q-1) \ln \left( \frac{1}{w_j^{(s)}} \right) \right] = \\ &= \sum_{j=1}^m w_j \exp \left[ - (q-1) \ln \left( \frac{S_{\pm}}{y_j^{(s)}} \right) \right] \cong A_0^{(s)} + \sum_{n=1}^k A_n^{(s)} \exp \left[ - (q-1) \lambda_n^{(s)} \right]. \end{aligned} \quad (43)$$

Можно дать следующую интерпретацию выражению (43). Обычно для характеристики некоторой случайной последовательности используются арифметическое среднее и величина стандартного отклонения. Но как было показано выше, эти величины являются приближенными и справедливы только для первых *двух* моментов (7) с двумя добавленными точками. Но, как показывает новая статистика, для характеристики случайной последовательности можно использовать *кластеризационную* интерпретацию. В соответствии с этой трактовкой случайные амплитуды, имеющие примерно одинаковые величины априорных вероятностей, могут быть сгруппированы в кластеры. Если вероятность некоторой группы кластеров близка к единице  $w_j^{(s)} \cong 1$ , то их вклад характеризуется амплитудой  $A_0^{(s)}$ . В другом предельном случае, когда вероятности некоторых амплитуд близки к нулю ( $w_j^{(s)} \cong 0$ ), то вклад этой  $k$ -й последней кластерной группы в выражении (43) становится пренебрежимо малым, т. е.  $A_k^{(s)} \cong 0$ . Другие промежуточные кластерные группы ( $n = 1, 2, \dots, k-1$ ) с вероятностью  $\lambda_n^{(s)} = \ln(1/w_n^{(s)})$  будут ранжированы между этими двумя предельными кластерами с амплитудами  $A_n^{(s)}$ , определяемыми

методом СК. Другими словами, *произвольная* случайная последовательность имеет свою неэкстенсивную энтропию, характеризуемую величиной  $q$ , *совпадающей с величиной*  $(q - 1)$  момента, и сгруппированную в соответствии с кластерами, каждый из которых характеризуется величиной  $\lambda_n^{(s)} = \ln(1/w_n^{(s)})$  и  $A_n^{(s)}$ . Если все значения  $w_j^{(s)}$  в последнем выражении строго положительны, то значения  $q$  определены в интервале  $(-\infty < q - 1 < \infty)$ . Если для некоторых кластеров  $w_j^{(s)} = 0$  (бинарные последовательности, к примеру), то значения величины  $q$  должны быть локализованы в интервале  $(1 < q < \infty)$ . Принимая во внимание выражение (43), формула (38) для обобщенной неэкстенсивной энтропии принимает следующий вид

$$S_q^{(\pm)} = -k_B \frac{1 - \sum_{j=1}^m w_j^q}{1 - q} = k_B \frac{1 - \Delta_{w_{\pm}}^{(q-1)}}{q - 1} \cong \frac{1 - \left[ A_0^{(\pm)} + \sum_{n=1}^k A_n^{(\pm)} \exp \left[ - (q - 1) \lambda_n^{(\pm)} \right] \right]}{q - 1}. \quad (44)$$

Случайная последовательность может быть охарактеризована определенным набором параметров  $A_0^{(\pm)}$ ,  $A_n^{(\pm)}$ ,  $\lambda_n^{(\pm)}$ . Эти параметры, как следует из (44), имеют те же самые значения дробного момента, что и для энтропии Цаллиса. Число кластеров  $n$  может быть идентифицировано методом СК. Поэтому энтропия Цаллиса получает дополнительное истолкование и может быть использована для описания *произвольной* случайной последовательности. Предполагается, что параметр  $q$  локализован в интервале  $1 < q < \infty$ . Предельное значение энтропии (44) при  $q = 1$  определяется стандартным выражением

$$S_1^{(\pm)} = -k_B \sum_{j=1}^N w_j \ln(w_j). \quad (45)$$

#### 4. Статистическая защита ценных документов

Как хорошо известно, надежная защита денежных банкнот, пластиковых карточек, картин, старинных манускриптов и прочих ценных документов от их подделок мошенниками представляет собой серьёзную проблему в современном мире. Не вдаваясь в её детальный анализ, в этом разделе рассмотрим схематически возможное приложение метода СДМ для *статистической* защиты некоторой поверхности, которая должна быть надежно защищена от подделки (возможного копирования).

Предположим, что поверхность на рис. 1. представляет собой поверхность пластиковой карточки и случайные точки (отмеченные крестиками), нанесенные с помощью специальной маски, представляют собой некую статистическую защиту этой поверхности. Технология нанесения этих точек здесь не обсуждается, так как представляет собой отдельную и нетривиальную задачу. Число маркированных точек, равное 40, и место их локализации на заданной поверхности составляют некую конфиденциальную информацию. Существует некий потенциальный «мошенник», который хочет создать фальшивую карточку и тем самым заменить ею настоящую. Допустим, что он информирован о месте нанесения защитных точек и даже знает точное их число. В этом случае специальное устройство, «настроенное» на чтение истинных точек, легко распознаёт фальшивую карточку, ориентируясь на

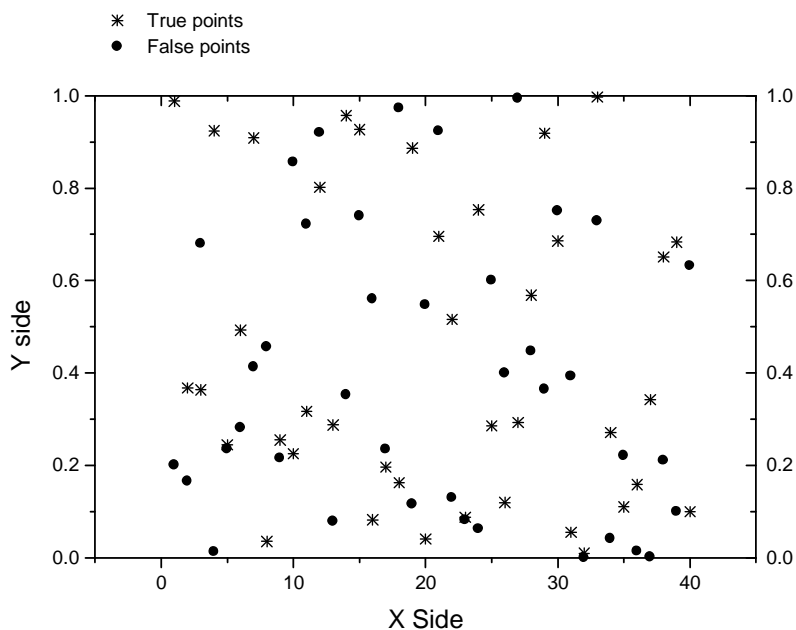


Рис. 1. Воображаемая поверхность пластиковой карточки, содержащая 40 истинных точек, отмеченных звездочками, и 40 имитированных (фальшивых) точек, отмеченных черными кружочками

тот факт, что места нанесения истинных и фальшивых точек различаются между собой. Эти отличия легко детектируются в пространстве коэффициентов характеристического полинома и по числу отличающихся между собой корней. Эти различия показаны, соответственно, на рис. 2 и 3.

Допустим теперь более серьезную ситуацию, когда потенциальный мошенник *знает* локализацию на поверхности всех истинных точек, за исключением *одной*. Эта возможная ситуация представлена на рис. 4. Возможно ли отличить статистическое различие всего в одну точку в данной ситуации? Ответ на этот вопрос является положительным и представлен на рис. 5 и 6. Вновь, как и в первом случае, можно эмпирически определить порядок характеристического полинома, который выявляет это различие. В рассматриваемом случае он равен 20. Появление комплексно-сопряженных корней делает эти различия особенно яркими. Если эта случайная последовательность отбирается как выборка из более расширенной совокупности и детектор, настроенный на «чтение» истинной карточки, «знает» объём выборки, отобранный для распознавания, зависящий от других случайных, но *управляемых* факторов, то подделка такой карточки становится попросту невозможной. Эта простая идея, лежащая в основе статистической защиты, для осуществления её в реальности естественно нуждается в определенной модификации и адаптации и может быть использована в совокупности с другими мерами защиты, используемыми для защиты ценных документов.

Интересно сравнить первый подход, основанный на вычислении искомых корней характеристического полинома (5), с вычислением ФОС, связанной с расчетом подгоночных параметров, фигурирующих в выражении (30). Результаты расчета подгоночной функции для второго случая, когда случайные последовательности различаются всего на одну точку, представлены на рис. 7.



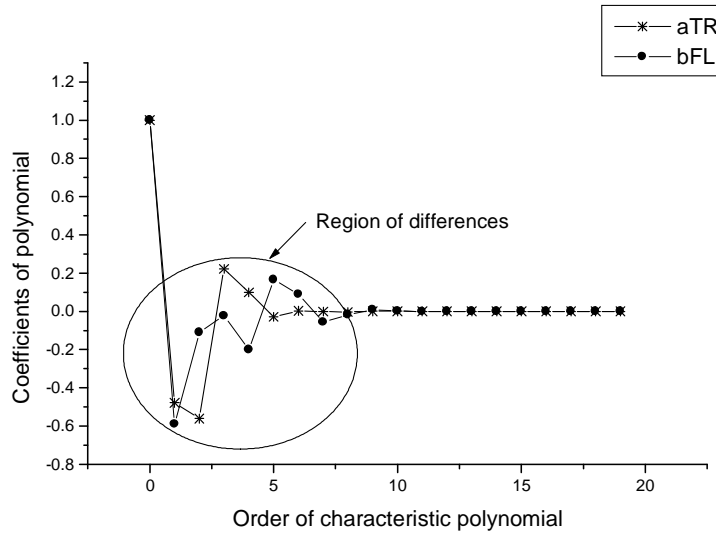


Рис. 2. Показано качественное различие между двумя характеристическими полиномами, построенными при разных коэффициентах, вычисленных по формуле (4). Порядок характеристического полинома в этом случае равен 19

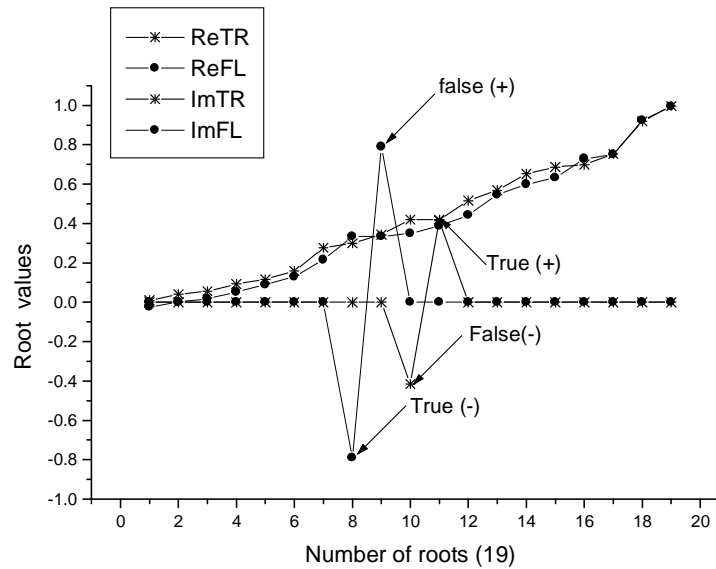


Рис. 3. Различие между коэффициентами двух полиномов (рис. 2) становится более заметным, если вычислить 19 корней этих полиномов. Эти различия показаны стрелками. В результате расчетов могут появиться «нефизические» комплексные корни. В этом случае они могут быть использованы для сравнения, так как численная программа, применяемая для их расчета, одна и та же самая для обоих случаев



Рис. 4. Попытка потенциального мошенника имитировать все защитные (истинные) точки, за исключением одной, показанной стрелкой. Возможно ли различить две случайные последовательности, отличающиеся всего лишь одной точкой?

Этот важный результат показывает, что случайные последовательности (отличающиеся всего на одну точку!) могут быть (если в этом есть необходимость) отделены друг от друга и, следовательно, ФОС может быть использована в качестве весьма чувствительного детектора для количественного «прочтения» и последующего разделения случайных последовательностей самой различной природы в терминах подгоночных параметров  $A_n$  и  $\lambda_n$ .

## 5. Модельные эксперименты

**5.1. Детектирование преобладающего фактора (сверхслабого сигнала).** Проблема детектирования сверхслабых сигналов, когда амплитуда возможного сигнала ( $S$ ) меньше, чем средняя амплитуда шума ( $N$ ) ( $S/N < 1$ ), является важной при анализе случайных последовательностей. Не вдаваясь в детальное обсуждение этой проблемы, покажем, как использовать ФОС для детектирования искажений, вызванных влиянием некоторого малого преобладающего фактора. Рассмотрим некий случайный модельный шум, имеющий равномерное распределение, с амплитудами, локализованными в интервале  $nin_j \in [0, 1]$ . Число точек равно  $j = 1, 2, \dots, N$  ( $N = 400$ ), переменная  $x_j$  локализована в интервале  $[0, 10.0]$ . К этому начальному шуму добавляем сигнал с малой амплитудой  $a \in [0.01, -0.1]$ , выраженной в виде функции  $\sin^2(\Omega \cdot x_j)$  ( $\Omega = 1$ ). Математическая схема этого численного эксперимента может быть выражена в виде

$$ns_j = nin_j(1 + a \sin^2(\Omega x_j)). \quad (46)$$

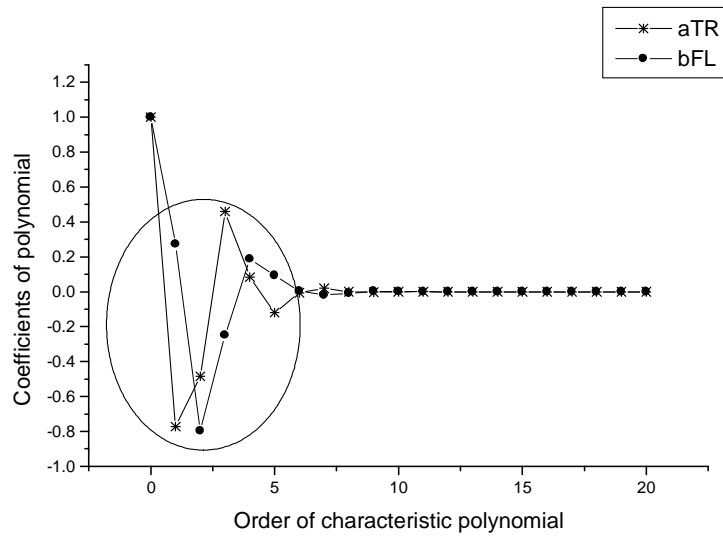


Рис. 5. Этот рисунок дает утвердительный ответ и показывает, что искомое различие может быть выявлено на коэффициентах сравниваемых полиномов (в этом случае порядок полинома равен двадцати). Как видно из этого рисунка, исходное различие в одну точку порождает в этом представлении 7 различных коэффициентов! Порядок полинома находится эмпирически

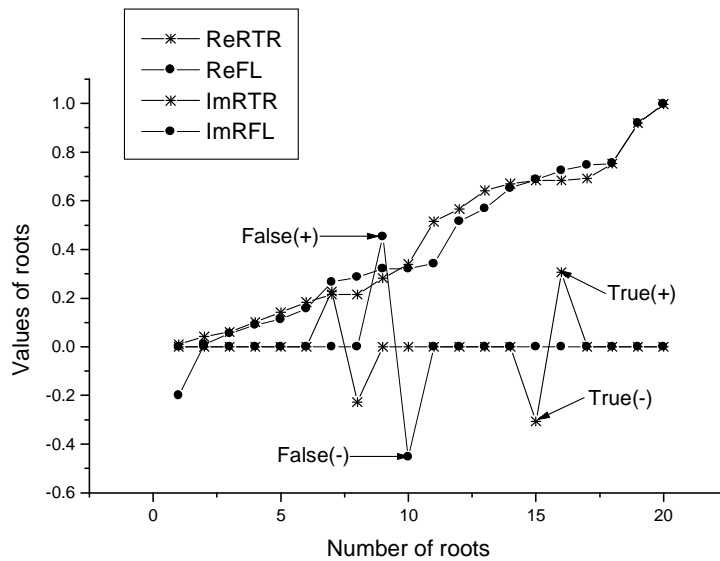


Рис. 6. Различия, представленные на предыдущем рисунке, становятся более явственными, если вычислить корни этих полиномов. Как и ранее, комплексные корни выявляют наибольшие различия, которые порождаются всего лишь *одной точкой* (!), показанной на рис. 4

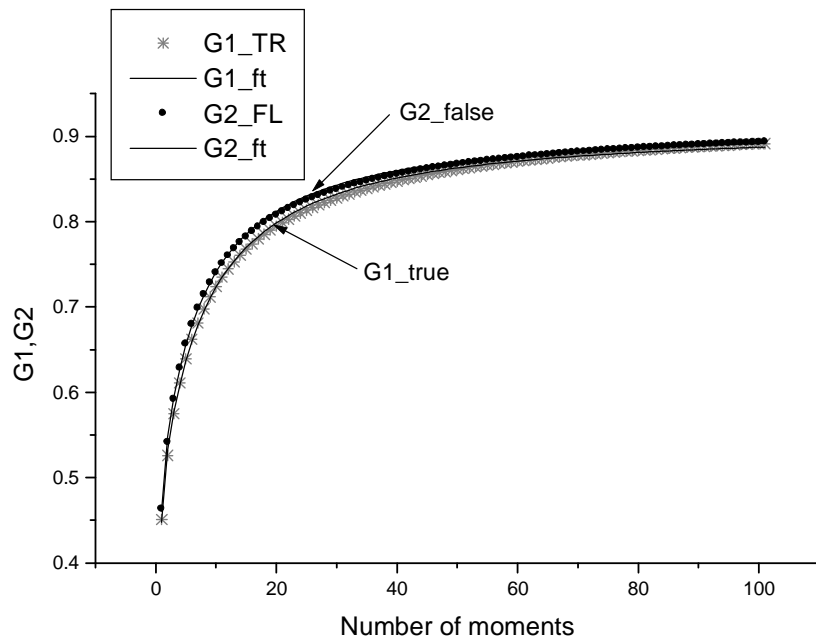


Рис. 7. Для случайных последовательностей, представленных на рис. 4, можно рассчитать ФОС и подогнать её по формулам (21)–(23). Подгоночная функция содержит три экспоненты и соответствующие параметры равны (в скобках даны величины, соответствующие «фальшивой» функции):  $A_1 = 0.5127$  (0.4886),  $A_2 = 0.1740$  (0.1765),  $A_3 = 0.0631$  (0.0876),  $\lambda_1 = -0.6983$  (-0.6959),  $\lambda_2 = -0.2546$  (-0.2548),  $\lambda_3 = -0.0921$  (-0.0920). Эти величины могут быть использованы в качестве *универсальной* количественной меры для дифференциации двух близких последовательностей

Так как амплитуда добавленного сигнала очень мала, то сигнал визуально не может быть обнаружен, как это видно из рис. 8. Проблема может быть сформулирована следующим образом: *возможно ли обнаружить влияние малого преобладающего фактора, который изначально выражен в виде функции  $\sim \sin^2(\Omega x_j)$ , а затем выразить в форме возрастающей функции  $f(a)$  по отношению к монотонному изменению амплитуды  $a$ ?*

Для того чтобы найти искомую зависимость (которая может быть определена как *калибровочная кривая*), мы выполнили следующую процедуру:

1. В соответствии с определением (12) были вычислены ФОС, соответствующие исходному шуму и шуму, модифицированному малым сигналом. Наиболее удобным представлением для выявления влияния малого фактора является зависимость  $G_{s,N}^{(p)}$  (содержащий возможный сигнал) против  $G_{0,N}^{(p)}$  (ФОС для сегмента без сигнала), вычисленного в пространстве моментов. Рис. 9 показывает монотонное изменение шума (в зависимости от величины внешней амплитуды  $a$ ) в присутствии сигнала  $G_{s,N}^{(p)}$  от функции  $G_{0,N}^{(p)}$ , соответствующей «чистому» шуму.

2. Для того чтобы выразить эти монотонные изменения *количественно* в зависимости от амплитуды  $a$ , нужно вычислить площадь под кривой (ППК), образо-

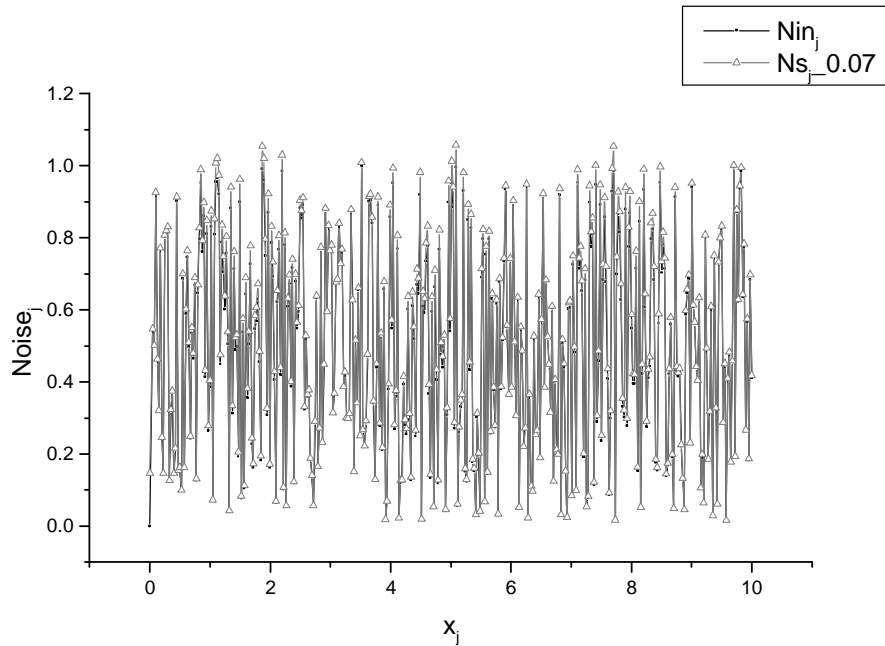


Рис. 8. Изображены две случайные последовательности. Одна (отмеченная серым цветом) содержит малое возмущение (сигнал) с амплитудой  $a = 0.07$ . Визуально присутствие сигнала незаметно. Необходимы специальные методы для выявления таких малых возмущений. Присутствие сигнала может быть замечено с помощью ФОС, вычисленной и представленной особым образом в пространстве дробных моментов

ванной относительной разностью

$$D(p) = \left( G_{s,N}^{(p)} - G_{0,N}^{(p)} \right) / G_{0,N}^{(p)}. \quad (47)$$

Величина ППК ( $S$ ), вычисленная для функции  $D(p)$ , показана на рис. 8. Эта кривая, отражающая монотонное поведение фактора  $S(a)$ , может быть использована в качестве калибровочной кривой.

На практике трудно ожидать, что начальный «шум» и шум, искаженный присутствием малого сигнала, совпадают друг с другом. Поэтому с самого начала необходимо сравнить две случайные последовательности, а лишь после того «подозрительную» последовательность, искаженную присутствием сигнала (внешнего фактора). В качестве начального шума мы выбрали шум, подчиняющийся равномерному распределению. Калибровочная кривая, вычисленная для этого случая, показана на рис. 10. Она сдвинута по сравнению с идеальным случаем, когда исходный шум двух случайных последовательностей идентичен друг другу. Аналогичным образом мы рассмотрели случаи, когда амплитуда сигнала ( $a$ ) в десять раз меньше и сосредоточена в интервале  $(0.001, 0.002, \dots, 0.01)$ . Все возможные ситуации математически могут быть выражены в виде

$$ns(x_j) = nin_1(x_j) + a \sin^2(\Omega x_j + \Phi(x_j)) \exp \left[ -\frac{(x_j - \langle x \rangle)^2}{\sigma} \right] nin_2(x_j). \quad (48)$$

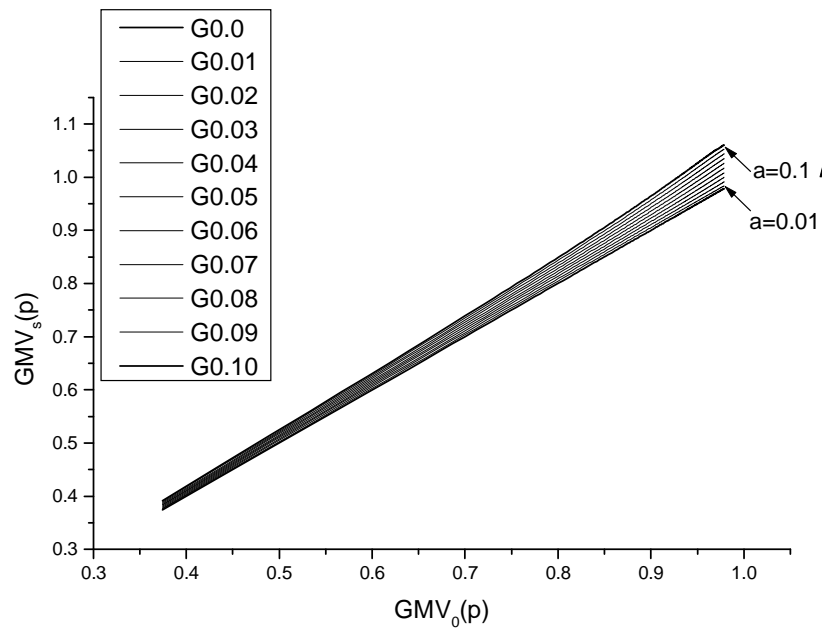


Рис. 9. Показана начальная стадия детектирования слабого возмущения. Возможные искажения, вызванные присутствием малого сигнала, могут быть обнаружены на графике  $G_{Ns}^{(p)}$  против  $G_{0N}^{(p)}$ . Основной график  $G_{0N}^{(p)}/G_{0N}^{(p)}$  образует прямую линию с единичным наклоном и отсечкой, равной нулю. Малые возмущения вызывают вариацию этих параметров. Если внешний фактор (в этом конкретном случае – величина амплитуды малого сигнала) вызывает монотонные искажения, то искомая калибровочная кривая может быть вычислена

Здесь  $nin_{1,2}(x_j)$  и  $\Phi(x_j)$  – случайные различные функции, экспоненциальная функция указывает на степень локализации малого сигнала. Вычисления функции (48) показывают, что, несмотря на малую величину амплитуды  $a \in [10^{-3}, 10^{-2}]$ , возможно вычислить калибровочную кривую при различном влиянии таких факторов, как  $\Omega$ ,  $\sigma$  и произвольном выборе функций  $nin_{1,2}(x_j)$  и  $\Phi(x_j)$ . Соответствующие калибровочные кривые показаны на рис. 11 с необходимыми пояснениями, приведенными в подписи к этому рисунку. Проведенные расчеты показывают, что локализация искомого сигнала (зависящая от величины  $\sigma$ ) и монотонное поведение амплитуды сигнала (выраженной величиной  $a$ ) являются важными факторами при построении калибровочной кривой. Эти наблюдения, полученные на модельных данных, должны быть приняты во внимание при обработке реальных данных, когда необходимо «поймать» влияние малого внешнего фактора, полностью скрытого внутри анализируемой случайной последовательности.

**5.2. Детектирование статистической близости.** Статистика дробных моментов, выраженная в терминах ФОС, предоставляет уникальную возможность сравнить различные распределения и оценить их статистическую близость друг другу. Если для двух случайных последовательностей часть целых моментов ( $p = 0, 1, \dots, k$ ) близка друг другу (смотри определение (27)), то такие последовательности, представленные зависимостью  $G_{N_1}^{(p)} [G_{N_2}^{(p)}]$ , представляют собой прямую

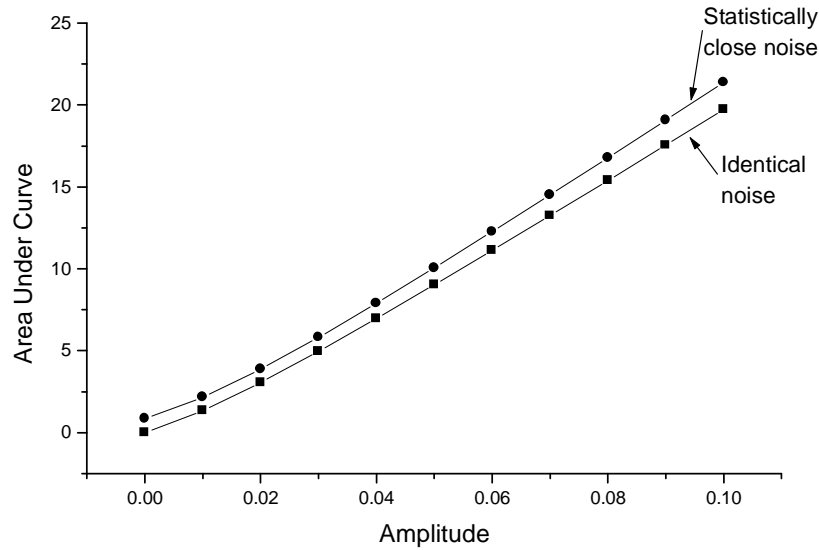


Рис. 10. Показана заключительная стадия детектирования малого монотонного возмущения. Чувствительным фактором может служить величина ППК, вычисленная для относительной разности  $(G_{sN}^{(p)} - G_{0N}^{(p)}) / G_{0N}^{(p)}$ . Именно вариации этой площади помогают рассчитать искомую калибровочную кривую. Нижняя кривая соответствует идеальной ситуации: исходный шум в обоих случаях один и тот же. Верхняя кривая соответствует более реалистическому случаю: исходные шумы различны, но подчиняются одному и тому же (равномерному) распределению

линию. К примеру, рис. 11 демонстрирует статистически близкие последовательности. Если распределения случайных точек различны, то можно ожидать, что их «статистические рисунки», выраженные в форме ФОС, также должны быть различными. Рис. 12 иллюстрирует эту идею, выраженную в виде функции  $G_{ND}^{(p)}$  (зависимость построена для нормального распределения), против ФОС, рассчитанных для других распределений. Как можно видеть из этого рисунка, поведения совокупности моментов для различных распределений разительно отличаются друг от друга. Результаты этого предварительного анализа могут быть использованы для обнаружения малого преобладающего фактора в тех случаях, когда его влияние изменяет исходное распределение.

## 6. Обработка реальных данных

Изначально на основе сыворотки новорожденного теленка (фирма «Сигма Олдрич») были приготовлены образцы с различной концентрацией глюкозы: 100 mM, 10 mM, 1 mM, 0.1 mM, 0.01 mM, 0.001 mM и 0.0001 mM (концентрация даётся в миллимолях). Диэлектрическая проницаемость образцов сыворотки была измерена при температуре 25°C в диапазоне частот  $2.10^8 \div 2.10^{10}$  Hz с использованием НР диэлектрического спектрометра. Для каждой пробы с заданной концентрацией глюкозы было проведено 20 идентичных измерений. Типичные зависимости комплексной проницаемости  $\text{Re}(\varepsilon)$  и  $\text{Im}(\varepsilon)$  от логарифма частоты для чистой сыворотки без примеси глюкозы показаны на рис. 13. Проблема может быть сфор-

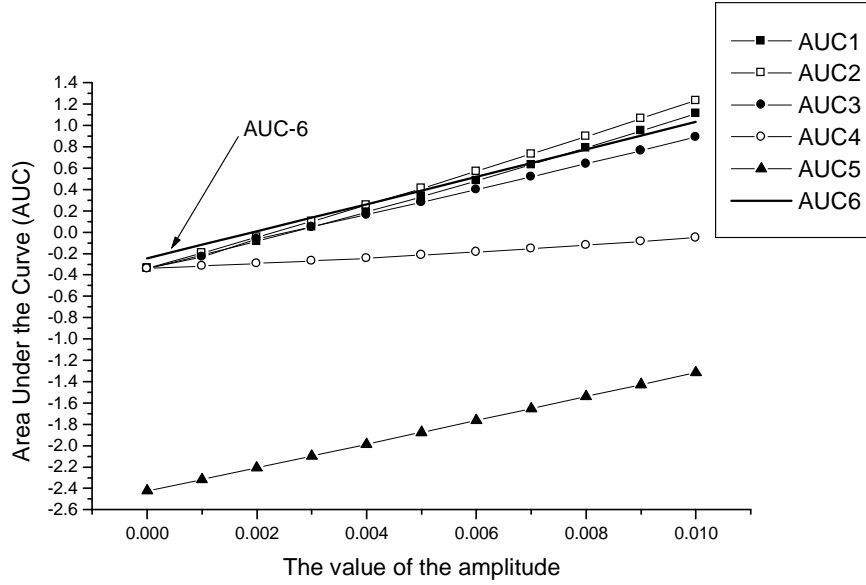


Рис. 11. Здесь показаны различные калибровочные кривые, полученные из выражения (48). Кривая **AUC-1** соответствует случаю:  $(nin_2(x_j) \equiv \exp[(x_j - \langle x \rangle)^2 / \sigma] \equiv \Phi(x_j) \equiv 1)$ ; кривая **AUC-2** ( $nin_2(x_j) \equiv \exp[(x_j - \langle x \rangle)^2 / \sigma] \equiv 1$ ,  $\Phi(x_j)$  совпадает с равномерным распределением); кривая **AUC-3** ( $nin_2(x_j) \equiv \Phi(x_j) \equiv 1$ ,  $\sigma = \langle x \rangle^2$ ); кривая **AUC-4** ( $nin_2(x_j) \equiv \Phi(x_j) \equiv 1$ ,  $\sigma = 1$ ); кривая **AUC-5** ( $\exp[(x_j - \langle x \rangle)^2 / \sigma] \equiv \Phi(x_j) \equiv 1$ ,  $nin_2(x_j)$  совпадает с нормальным распределением); последняя кривая **AUC-6** соответствует тому же набору параметров, что и кривая **AUC-5**, но число случайных точек в десять раз больше,  $N = 5000$ . Для всех остальных кривых, за исключением **AUC-6**,  $\Omega = 1$ ,  $N = 500$

мулирована следующим образом: возможно ли вычислить калибровочную кривую без привлечения модельных представлений (при отсутствии информации о подгочной функции) к такой сложной системе, как смесь сыворотки с глюкозой?

Эта проблема может быть решена в рамках метода СДМ при условии, что добавляемая жидкость (глюкоза в нашем случае) просто монотонно сдвигает изначальную кривую и не взаимодействует химически или физически с исходным базовым компонентом (сывороткой). Чтобы увидеть визуально возможные искажения в спектре, обусловленные добавлением глюкозы, удобно вычислить вторые базовые *относительные* моменты для величин  $\text{Re}(\varepsilon)$  и  $\text{Im}(\varepsilon)$  по формулам

$$\Phi(\text{Re}(\varepsilon)) = \left[ \frac{1}{M} \sum_{m=1}^M (\text{Re}(\varepsilon)_m - \langle \text{Re}(\varepsilon) \rangle)^2 \right]^{\frac{1}{2}},$$

$$\Phi(\text{Im}(\varepsilon)) = \left[ \frac{1}{M} \sum_{m=1}^M (\text{Im}(\varepsilon)_m - \langle \text{Im}(\varepsilon) \rangle)^2 \right]^{\frac{1}{2}}, \quad (49)$$

$$\langle f \rangle = \frac{1}{M} \sum_{m=1}^M f_m.$$



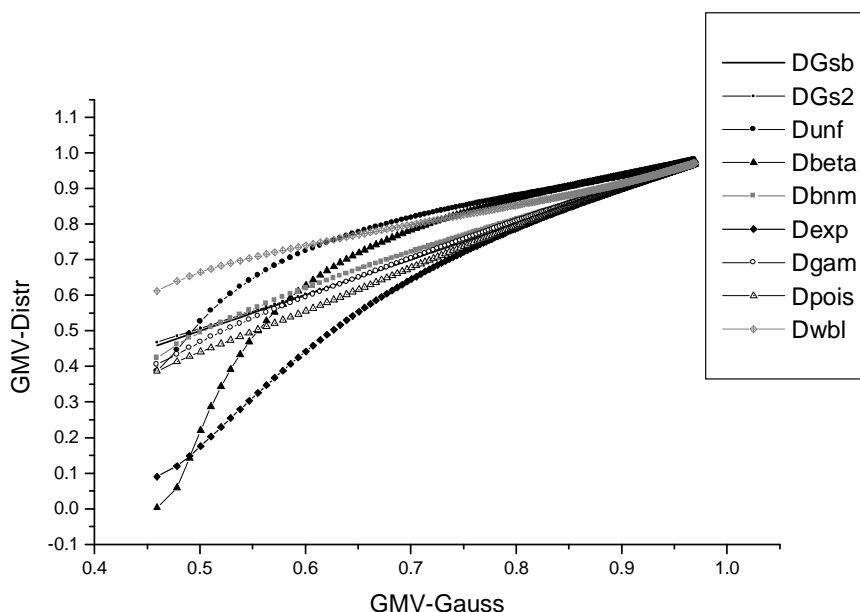


Рис. 12. Этот рисунок иллюстрирует тот факт, что случайные последовательности, принадлежащие различным распределениям, могут быть визуально отделены друг от друга, если представить их в пространстве дробных моментов. Здесь все случайные последовательности, принадлежащие различным распределениям, нормированы в интервале  $[0, 1]$ . Все дробные моменты вычислены в интервале  $[0.1, 200]$  с шагом  $h_p = 0.4$ . Обозначения соответствуют следующим распределениям: «DGsb», «DGs2» – нормальные распределения со средним значением, равным 0.5, и единичной дисперсией; «Dunf» соответствует равномерному распределению в интервале  $[0.01, 1]$ ; «Dbeta» соответствует бета-распределению с параметрами  $a = 0.1$  и  $b = 0.5$ ; «Dbnm» – это биномиальное распределение с  $m = 50$  и  $p = 0.5$ . Можно заметить, что это распределение близко к нормальному. «Dexp» – экспоненциальное распределение с параметром  $\lambda = 50$ ; «Dgam» описывает гамма-распределение с параметром  $s = 50$ ; «Dpois» – распределение Пуассона с  $s = 25$ ; «Dwbl» – распределение Вэйбулла с параметром  $w = 15$ . Анализируя этот рисунок, можно сказать, что каждое распределение имеет свой неповторимый статистический рисунок

Базовые относительные моменты определены в работах [15, 16]. Здесь индекс  $m$  означает номер текущего измерения, индекс  $M$  соответствует полному числу измерений (в нашем случае  $M = 20$ ).

Рис. 14 и 15 показывают частотное поведение вычисленных значений относительных флуктуаций. Их визуальный анализ показывает, что эти функции располагаются *монотонно* по отношению к различным концентрациям глюкозы. Этот факт позволяет разделить эти функции в пространстве моментов. Рассматривая каждую функцию как случайную последовательность и осуществляя ту же численную процедуру, описанную в разд. 6, можно вычислить искомую калибровочную кривую. Для этого случая достаточно рассчитать ППК для относительной разности  $D(p)$ , определенную выражением (47). Соответствующие калибровочные кривые, вычисленные для действительной и мнимой частей относительных моментов второго порядка, показаны на рис. 16. Можно заметить, что эти кривые являются монотонными кривыми, но они не могут быть подогнаны под прямую линию,

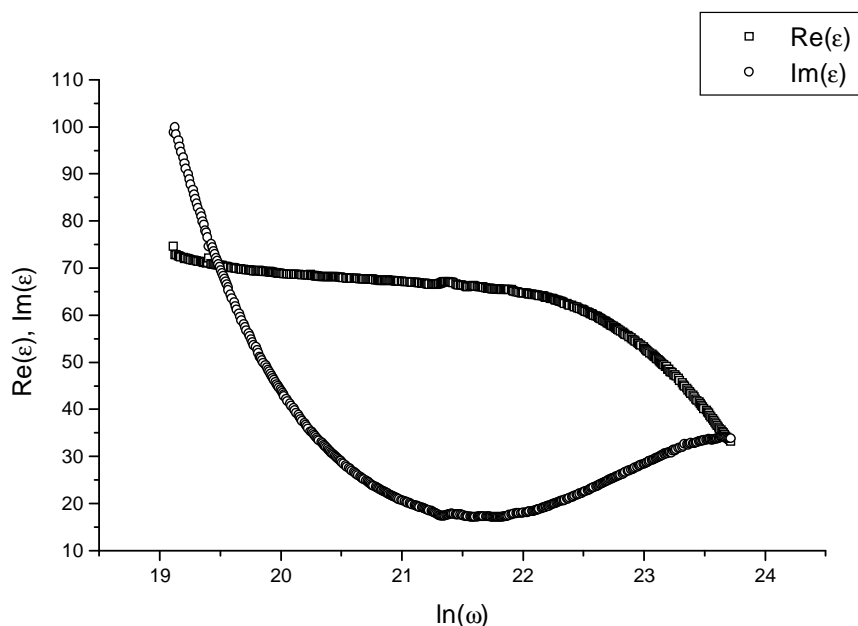


Рис. 13. Типичная кривая, демонстрирующая поведение комплексной восприимчивости измеренная для такой сложной жидкости, как сыворотка новорожденного теленка при ( $T = 25^{\circ}\text{C}$ ). Эти кривые (если их подгоночная функция неизвестна) могут быть рассмотрены как случайные кривые. Другие кривые, измеренные для различных концентраций глюкозы, не показаны, так как они весьма близки друг другу

потому что эффект разделения, представленный на рис. 14 и 15, не является *равномерным*.

СДМ предоставляет новую возможность для проверки устойчивости и статистической близости различных измерений, полученных с помощью разных экспериментальных приборов. Кроме того, метод СДМ позволяет рассмотреть каждое измерение как некое уникальное «действие», а затем выразить результат каждого измерения *количественно* в терминах ППК с использованием выражения (47). Такого рода зависимости могут дать дополнительную информацию при анализе различных измерений. В частности, можно оценить более точно в пространстве моментов области неопределенности и рассмотреть распределение отклонений, полученных в течение всего эксперимента. Соответствующие зависимости ППК от номера измерения ( $m = 1, 2, \dots, 20$ ), полученные для чистой сыворотки и раствора с концентрацией глюкозы ( $c = 10^2$ ), представлены на рис. 17.

## 7. Основные результаты

В этом разделе можно вкратце сформулировать основы статистики дробных моментов (СДМ), которая, несомненно, найдет широкое применение при количественном сравнении случайных последовательностей различной природы.

1. Набор величин высших моментов, которые оставляют неизменными величину начального отрезка случайной последовательности (условие (2)), позволяет по-новому оценить и сравнить стабильность и близость случайных рядов и вы-

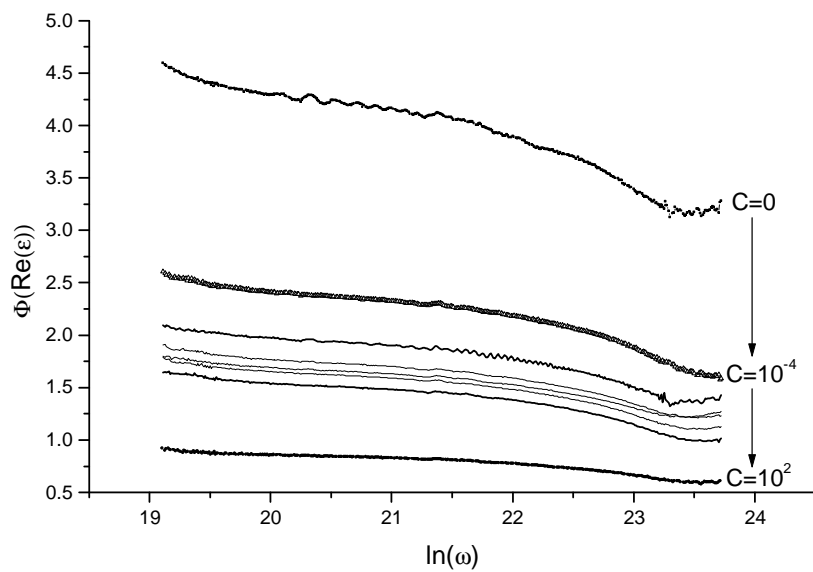


Рис. 14. Набор кривых, соответствующих второму относительному моменту, вычисленных для  $\text{Re}(\epsilon)$ . Концентрация глюкозы монотонно увеличивается от верхней кривой к нижней. Правда, следует отметить, что это увеличение неравномерное

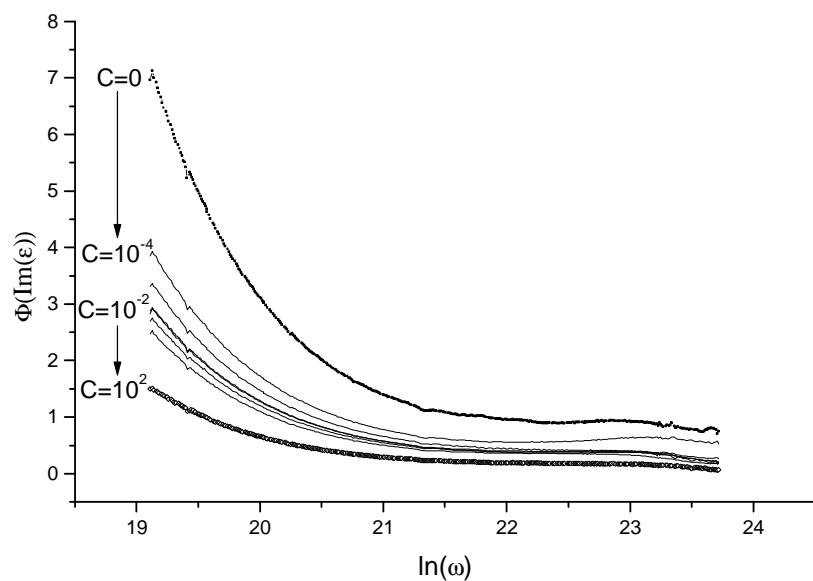


Рис. 15. Набор кривых, соответствующих второму относительному моменту, вычисленных для  $\text{Im}(\epsilon)$ . Как и на прежнем рисунке, концентрация глюкозы монотонно и неравномерно увеличивается от верхней кривой к нижней

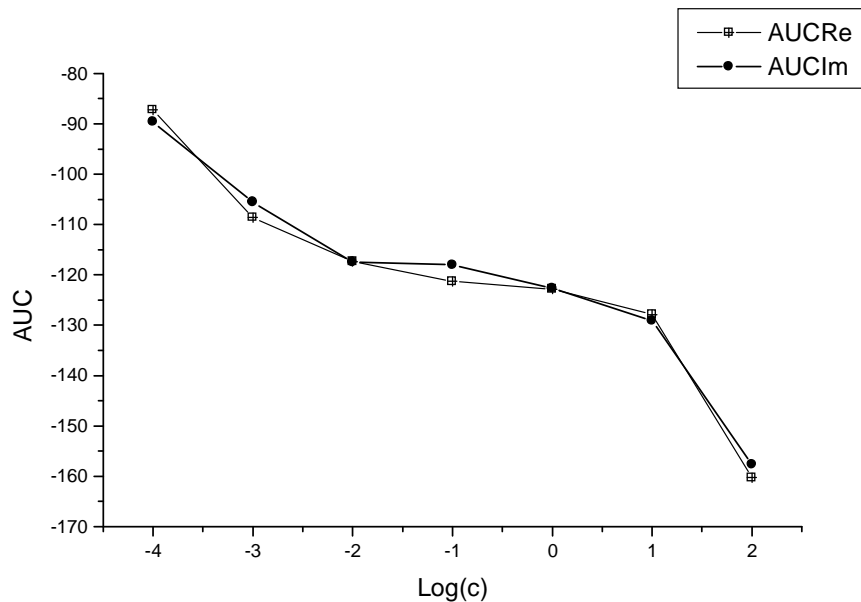


Рис. 16. Две калибровочные кривые, рассчитанные для ППК, изображенных на рис. 14 и 15. Эти кривые в силу неравномерности их возрастания не могут быть подогнаны под прямую линию

разить эти характеристики в терминах целых ( $k \ll N$ ) или дробных моментов ( $k \cong N \gg 1$ ). Это условие в общем виде выражается соотношением (27).

2. Решения уравнения (3) для относительно малых значений  $k$  обладают уникальной чувствительностью и помогают детектировать одну (!) «чужую» точку, которая различает две сравниваемые последовательности. Примеры по использованию такой чувствительности рассмотрены в разд. 4.

3. ФОС, определенная выражением (12), помогает преобразовать случайную последовательность в детерминированную функцию (21). Подгоночные параметры этой функции помогают сравнить *количественно* две случайные последовательности друг с другом. Это важное достоинство метода СДМ, несомненно, найдет широкое применение при расчете различных калибровочных кривых и в анализе случайных последовательностей самой различной природы.

4. Целочисленные моменты могут быть легко обобщены на дробные и даже комплексные моменты, что позволяет расширить возможности новой СДМ в целом.

5. СДМ тесно связана с выражением для неэкстенсивной энтропии (44), что, несомненно, расширяет границы этого определения и помогает оценить количественно стохастические (энтропийные) свойства произвольной случайной последовательности с помощью набора дробных моментов.

6. Поучительные примеры, рассмотренные в этой статье, должны убедить потенциального читателя в преимуществах метода СДМ, которая позволяет решить некоторые новые проблемы: 1) дифференцирование статистической устойчивости и её возможное предсказание; 2) количественное сравнение различных случайностей между собой; 3) установление связи между ФОС и различными распределениями; 4) количественное детектирование влияния некоторого внешнего доминантного фактора (малого сигнала).

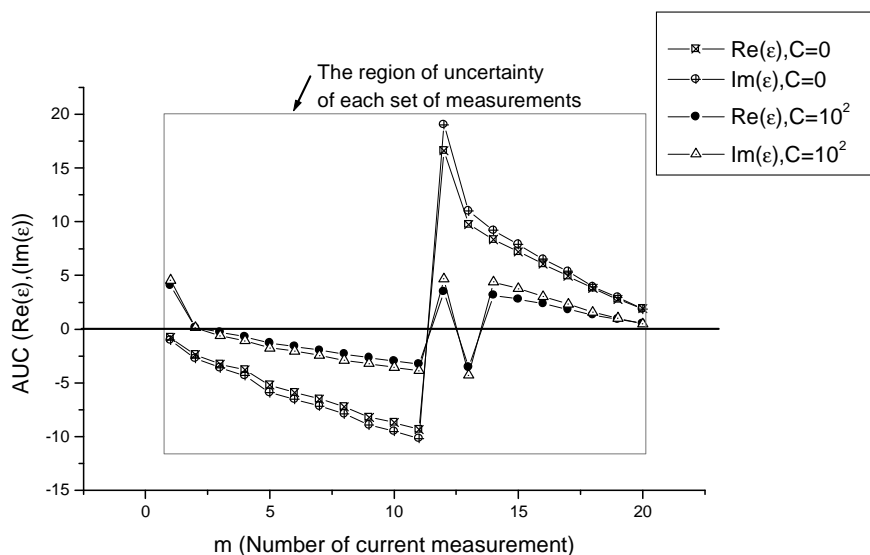


Рис. 17. Анализ каждого набора повторяющихся измерений, осуществляемого с помощью ФОС и последующего расчета ППК, позволяет найти область неопределенности (на рисунке она заключена в рамку) заданного числа измерений и получить, таким образом, дополнительную и уникальную информацию об особенностях измерительного прибора, участвующего в процессе измерения. Эти кривые показаны для двух случаев: чистая сы-воротка (величина ППК отмечена центрированным прямоугольником и ромбом) и для концентрации  $C = 10^2$  (величины ППК отмечены черным ромбом и треугольником)

Мы перечислили задачи, которые лежат на поверхности, нет сомнения в том, что дальнейшие исследования в этом направлении позволят существенно расширить круг проблем, которые могут быть решены в рамках новой статистики дробных моментов.

**Приложение: вывод системы уравнений (3), (15) и их возможные обобщения**

Вначале имеет смысл рассмотреть случай  $k = 1$ . Для этого случая имеем

$$\begin{aligned} \Delta_N^{(1)} &= \frac{1}{N} (y_1 + y_2 + \dots + y_N) = \frac{N-1}{N} \Delta_{N-1}^{(1)} + \frac{1}{N} y_N = \\ &= \Delta_{N-1}^{(1)} + \frac{1}{N} (y_N - \Delta_{N-1}^{(1)}) \end{aligned}$$

Если потребовать совпадения

$$\Delta_N^{(1)} = \Delta_{N-1}^{(1)},$$

то получается

$$y_N = \Delta_{N-1}^{(1)}.$$

Для  $k = 2$  имеем

$$\begin{aligned}\Delta_N^{(1)} &= \frac{1}{N} (y_1 + y_2 + \dots + y_N) = \frac{N-2}{N} \Delta_{N-2}^{(1)} + \frac{1}{N} (y_{N-1} + y_N) = \\ &= \Delta_{N-2}^{(1)} + \frac{1}{N} (y_{N-1} + y_N - 2\Delta_{N-2}^{(1)}). \quad (\text{A1})\end{aligned}$$

Заменяя  $y_j$  на  $y_j^2$ , получим по аналогии (A1)

$$\begin{aligned}\Delta_N^{(2)} &= \frac{1}{N} (y_1^2 + y_2^2 + \dots + y_N^2) = \frac{N-2}{N} \Delta_{N-2}^{(1)} + \frac{1}{N} (y_{N-1}^2 + y_N^2) = \\ &= \Delta_{N-2}^{(2)} + \frac{1}{N} (y_{N-1}^2 + y_N^2 - 2\Delta_{N-2}^{(2)})\end{aligned}$$

Из условия  $\Delta_N^{(p)} = \Delta_{N-2}^{(p)}$ , ( $p = 1, 2$ ) следует

$$\frac{y_{N-1} + y_N}{2} = \Delta_{N-2}^{(1)}, \quad \frac{y_{N-1}^2 + y_N^2}{2} = \Delta_{N-2}^{(2)}.$$

Для произвольного  $k$  аналогичные вычисления приводят к следующему соотношению

$$\begin{aligned}\Delta_N^{(p)} &= \frac{1}{N} (y_1^p + y_2^p + \dots + y_N^p) = \frac{N-k}{N} \Delta_{N-k}^{(p)} + \frac{1}{N} (y_{N-k+1}^p + \dots + y_N^p) = \\ &= \Delta_{N-k}^{(p)} + \frac{1}{N} (y_{N-k+1}^p + \dots + y_N^p - k\Delta_{N-k}^{(p)}), \quad p = 1, 2, \dots, k.\end{aligned}$$

Из условия  $\Delta_N^{(p)} = \Delta_{N-k}^{(p)}$ , ( $p = 1, 2, \dots, k$ ) получаем

$$\frac{y_{N-k+1}^{(p)} + \dots + y_N^{(p)}}{k} = \Delta_{N-k}^{(p)}, \quad p = 1, 2, \dots, k.$$

Это соотношение совпадает с системой уравнений (15) для набора из  $k$  точек (амплитуд,) локализованных *внутри*  $N$  точек. Для набора  $k$  точек, расположенных *вне* заданной случайной последовательности, необходимо заменить  $N-k \rightarrow N$ . Эти простые математические выкладки допускают следующее обобщение. Возможно произвести замену абсолютной величины случайной переменной  $y_j^p \rightarrow (y_j - a)^p$ , где  $a$  – произвольная константа. Когда  $a = \mu \equiv \Delta_N^{(1)}$ , то абсолютное значение момента совпадает с величиной относительного момента, определенного соотношением (8).

Автор хочет выразить свою глубокую благодарность доктору Джеффу Смитту (Университет де Монфорга, г. Лэстер, Великобритания) за предоставленную возможность использовать экспериментальные данные по комплексной восприимчивости сывортки для их последующего анализа методом СДМ.

Частично эта работа выполнена в рамках проекта РТ (№ 05-5.2-334/2005).

### Summary

*R.R. Nigmatullin.* The statistics of higher (fractional) moments: new method of quantitative “reading” of any arbitrary random sequence.

The statistical meaning of higher  $\Delta_N^{(p)}$  ( $p = 1, 2, \dots$ ) and fractional  $\Delta_N^{(p)}$  ( $0 < p < \infty$ ) moments for an arbitrary random sequence of the length  $N$  has been found. The higher moments help to reduce the sequence analyzed to a finite set of  $k$  *statistically stable* parameters, keeping invariant the values of the first  $k$ th moments  $\Delta_k^{(p)}$  ( $p = 1, 2, \dots, k$ ). The conditions of statistical stability and proximity expressed in terms of higher moments  $\Delta_N^{(p)} = \Delta_{N+k}^{(p)}$  ( $p = 1, 2, \dots, k$ ) help to find  $k$  unknown stable points and predict possible future behavior of the random sequence analyzed. The generalized mean value (GMV)-function defined as  $G_N^{(p)} = \left(\Delta_N^{(p)}\right)^{1/p}$  is turned to be very effective in analysis of statistically close random sequences or containing large numbers of measured points ( $N \gg 1$ ). The approximate analytical expression for an arbitrary  $p$  value from the range ( $-\infty < p < \infty$ ) entering into  $G_N^{(p)}$  has been found. It gives a possibility to transform *any* random sequence to the determined GMV curve and express *quantitatively* the reduced characteristics of any random sequence in terms of a “universal” set of the fitting parameters defined by the determined GMV-function. Statistical proximity factor can be used for construction of *calibration curves*, when it is necessary to compare one random sequence with another one to respect of variations of some given external factor (small signal). The higher moments are easily generalized for the fractional and even complex moments. In turn, the GMV-function can be also generalized and then calculated for 2D and 3D random sequences. The approach developed in this paper is free from *any* model assumption and can be extremely helpful in comparison of different random sequences using for these purposes the “unified” quantitative language based on introduction of the given set of fractional moments. The relationship between the value of the fractional moment and non-extensive parameter  $q$  entering into the definition of the non-extensive Tsallis entropy has been found. A possible model of statistical detection of plastic cards and other valuable documents demonstrating the effectiveness of the statistics of fractional moments has been considered. Some instructive examples in detection of superweak signals embedded into the basic random sequence ( $S/N = 10^{-2}, 10^{-3}$ ) based on model and real data confirm the effectiveness of new approach and can serve a new basis for numerous practical applications. Analysis of dielectric spectroscopy data by means of fractional moments gives unique possibility to compare *quantitatively* each measurement with each other and express the influence of a neutral additive in terms of calibration curve *without* concrete knowledge of the corresponding fitting function.

### Литература

1. *Feder E.* Fractals. – N. Y., London: Plenum Press, 1988.
2. *Daubechies I.* Wavelets and their Applications // Comm. Pure Appl. Math. – 1988. – V. 41. – P. 909.
3. *Daubechies I.* The basic aspects of wavelet theory // IEEE Trans. Inform Theory. – 1990. – V. 36. – P. 961.
4. *Daubechies I.* Ten Lectures on wavelets. – Philadelphia: CBMS Lecture Notes Series, 1991.
5. *Caufman R.* Wavelets and their applications. – Boston: John and Barlett Publishing, 1992.
6. *Yulmetyev R, Hanggi P, Gafarov F.* Stochastic dynamics of time correlation in complex systems with discrete time // Phys. Rev. E. – 2000. – V. 62. – P. 6178.
7. *Yulmetyev R, Hanggi P, Gafarov F.* Quantification of heart rate variability by discrete non-stationary non-Markov stochastic processes // Phys. Rev. E. – 2002. – V. 65, No 4, Art. 046107.

8. *Yulmetyev R.M., Gafarov F.M., Yulmetyeva D.G., Emelyanova N.A.* Intensity approximation of random fluctuation in complex systems // *Physica A.* – 2002. – V. 303. – P. 425.
9. *Timashev S.F.* A new dialogue with nature. Stochastic and chaotic dynamics in the lakes // *STOCHAOS* / Eds. D.S. Broomhead, E.A. Luchinskaya, P.V.E. McClintock, T. Mulin. – Melville, N. Y.: AIP Conference Proceedings, 2000. – P. 238.
10. *Timashev S.F.* Self-Similarity in Nature. – Ibid. – P. 562.
11. *Timashev S.F.* Science of complexity: Phenomenological basis and possibility of application to problems of chemical engineering // *Theoretical Foundation of Chem. Engineering.* 2000. – V. 34. – P. 301–312.
12. *Nigmatullin R.R.* Recognition of nonextensive statistic distribution by the eigen-coordinates method // *Physica A.* – 2000. – V. 285. – P. 547.
13. *Nigmatullin R.R.* Detection of weak signals based on a new class of transformations of random series // *Physica A.* – 2001. – V. 289. – P. 18.
14. *Nigmatullin R.R., Toboev V.A., Smith G., Butler P.* Statistical detection of the hidden distortions in diffusive spectra // *J. Phys. D: Appl. Physics.* – 2003. – V. 36. – P. 1044.
15. *Nigmatullin R.R., Smith G.* Fluctuation-noise spectroscopy and a “universal” fitting functions of amplitudes of random sequences // *Physica A.* – 2003. – V. 320. – P. 291.
16. *Nigmatullin R.R., Smith G.* The generalized mean value function approach: new statistical tool for the detection of weak signals in spectroscopy // *J. Phys. D: Appl. Physics.* – 2005. – V. 38. – P. 328.
17. *Mukundan R., Ramakrishnan K.R.* Moment functions in image analysis. Theory and applications. – Singapore: World Scientific, 1998.
18. *Giuliani A., Colafranceschi M., Webber Ch. (Jr.), Zbilut G.A.* Complexity score derived from the principal component analysis of nonlinear order measures // *Physica A.* – 2001. – V. 301. – P. 567.
19. *Abramovitz M., Stegan A.* Handbook of Mathematical Functions. – N. Y.: Dover, 1972.
20. *Mishina A.P., Proskuryakov I.V.* Advanced Algebra. – M.: Fizmatgiz, 1962 (in Russian).
21. *Belanov A.A.* Solution of algebraic equations by Lobachevsky method. – M.: Nauka, 1989 (in Russian).
22. *Nigmatullin R.R.* Eigen-coordinates: new method of identification of analytical functions in experimental measurements // *J. of Appl. Magn. Resonance.* – 1998. – V. 14. – P. 601.
23. *Abdul-Gader Jafar M.M., Nigmatullin R.R.* Identification of a new function model for the AC-impedance of thermally evaporated (undoped) selenium films using the Eigen-coordinates method // *Thin Solid Films.* – 2001. – V. 396. – P. 280.
24. *Nigmatullin R.R., Abdul-Gader Jafar M.M., Shinyashiki N., Sudo S., Yagihara S.* Recognition of a new universal permittivity for glycerol by the use of the Eigen-coordinates method // *J. of Non-Crystalline Solids.* – 2002. – V. 305. – P. 96.
25. *Al-Hasan M., Nigmatullin R.R.* Identification of the generalized weibull distribution in wind speed data by the Eigen-coordinates method // *Renewable Energy.* – 2003. – V. 28. – P. 93.
26. *Kendall M.G., Stuart A.* The advanced theory of statistics. V. 1. – N. Y., London, Sydney, Toronto: Ch. Griffin & Co. LTD, 1962.
27. *Tsallis C.* Possible generalization of Boltzmann–Gibbs statistics // *J. Stat. Phys.* – 1988. – V. 52. – P. 479.



- 
28. *Tsallis C.* Nonextensive thermostatics and fractal // *Fractals*. – 1995. – No 3. – P. 541.
  29. *Tsallis C.* Classical and quantum complexity and nonextensive thermodynamics // *Chaos, Solitons and Fractals*. – 2002. – V. 13. – P. 371.
  30. *Tsallis C.* Nonextensive statistic: theoretical, experimental and computational evidences and connections // *Braz. J. Phys.* – 1999. – V. 29. – P. 1.

Поступила в редакцию  
25.08.05

---

**Нигматуллин Равиль Рашидович** – доктор физико-математических наук, профессор кафедры теоретической физики Казанского государственного университета.

E-mail: *nigmat@knet.ru*