УДК 535.421+519.2

С. С. Мельник, О. В. Усатенко
Институт радиофизики и электроники им. А. Я. Усикова НАН Украины
12, ул. Ак. Проскуры, Харьков, 61085, Украина
E-mail: olegusatenko@mail.ru

СИМВОЛЬНЫЕ ЦЕПИ МАРКОВА С МУЛЬТИЛИНЕЙНОЙ ФУНКЦИЕЙ ПАМЯТИ

Задача конструирования различных радиотехнических устройств, таких как фильтры, линии задержки, случайные антенны с заданной диаграммой направленности, требует разработки методов генерации случайных последовательностей (значений параметров этих систем), обладающих заданными корреляционными свойствами, поскольку спектральные характеристики перечисленных и аналогичных им систем выражаются через фурье-компоненты корреляторов. Адекватным математическим аппаратом для решения таких задач являются цепи Маркова высших порядков. Статистические характеристики этих объектов полностью определяются их функцией условной вероятности, которая в общем случае может иметь весьма сложный вид. Целью данной работы является представление функции условной вероятности случайных символьных последовательностей с дальними корреляциями в виде, удобном для численной генерации последовательностей. Мы предполагаем, что пространство состояний системы является конечным абстрактным множеством. Производится разложение функции условной вероятности на независимые слагаемые, выраженные через так называемую матрицу-функцию памяти. Развитая теория открывает путь для построения более последовательного и тонкого подхода к описанию систем с дальними корреляциями. В предельном случае слабых (по значению, но не по расстоянию) корреляций функции памяти однозначно выражаются через корреляционные функции высших порядков, что позволяет генерировать случайные последовательности, обладающие сложными заданными дальними корреляциями. В качестве примера использования полученных аналитических результатов приводится численная реализация метода построения случайной последовательности с заданными конкурирующими матричными корреляторами второго и третьего порядков. Ил. 1. Библиогр.: 34 назв.

Ключевые слова: случайные последовательности, корреляционные функции, функция условной вероятности, цепи Маркова высших порядков.

Радиочастотные фильтры, линии задержки, фазированные антенные решетки (случайные антенны с заданной диаграммой направленности) [1] и дифракционные излучатели являются несколькими из многочисленных примеров систем, при конструировании которых могут быть использованы стохастические методы. Это возможно, когда в основе структуры этих устройств имеются случайные элементы, например, не эквидистантно расположенные излучающие элементы антенны. Такие элементы могут быть случайными как статически, так и динамически. При конструировании случайных систем с заданными, предписанными спектральными свойствами может использоваться метод подгонки, когда случайно выбирают один из элементов системы и, изменяя его характеристики, выбирают то его значение, которое наилучшим образом удовлетворяет критерию, по которому строится система. Такой метод является чрезвычайно трудоемким с точки зрения компьютерных вычислений, если речь идет о построении системы, число случайных элементов которой измеряется сотнями, и вовсе неприменимым, если число элементов существенно превышает несколько сотен. И в таком случае начинает хорошо «работать» статистический метод построения случайной последовательности параметров системы. Статистический подход дает инструмент для проектирования устройств и приложений со случайными компонентами в их структуре.

Статистические методы успешно развивались на протяжении последних десятилетий при изучении неупорядоченных систем, в теории андерсоновской локализации и слоистых систем [2], построении фильтров и линий задержек и многих других [3].

Исследования случайных систем в физических и инженерных науках можно разделить на две части. Первая занимается исследованием, анализом и предсказанием свойств таких систем, в то время как вторая, значительно меньшая, разрабатывает методы построения или генерации случайных процессов с заданными статистическими свойствами.

Стандартный метод описания статистических свойств заданной случайной последовательности требует оценки функции вероятности реализации L-слов (подпоследовательностей длиной L), или, что то же самое, функции условной вероятности появления определенного символа после L-слова. Надежные оценки для такой вероятности могут быть реализованы только при малых L, так как число m^L (m есть размер конечного алфавита) различных слов длиной L должно быть намного меньше, чем общее число $M–L$ слов во всей последовательности длиной M,

$$m^L << M - L \approx M. \qquad (1)$$

Как правило, корреляционная длина представляющих интерес естественных последовательностей имеет тот же порядок, что и длина самой последовательности. Неравенство (1) не может быть выполнено. Максимальная длина $L_{\max}$ слов, вероятности которых можно корректно оценить, равна 4–5 символов для реального естест-

 © ИРЭ НАН Украины, 2015

венного текста длиной 10^6 (записанного алфавитом, содержащим 27–30 букв и символов), или порядка 20 для огрубленного текста, представленного двоичной последовательностью. Следовательно, дальние корреляции, которые могут существовать в последовательностях, не могут быть учтены такого рода теориями.

Целью данной работы является представление функции условной вероятности случайных символьных последовательностей с дальними корреляциями в виде, удобном для реальной численной генерации.

В некотором смысле символьные последовательности являются более общими конструкциями, чем числовые. Действительно, предположим, что существует взаимно-однозначное соответствие $a_i \leftrightarrow \varepsilon_i$ между буквами символьной последовательности S и значениями числовой последовательности. Тогда обычная «числовая» корреляционная функция

$$C_\varepsilon(r) = \overline{(\varepsilon_i - \bar{\varepsilon})(\varepsilon_{i+r} - \bar{\varepsilon})} \tag{2}$$

последовательности ε_i может быть выражена через символьный коррелятор

$$C_\varepsilon(r) = \sum_{\alpha\beta\in A} \varepsilon^\alpha \varepsilon^\beta C_{\alpha\beta}(r). \tag{3}$$

Здесь $\sum\limits_{\alpha\in A}$ означает суммирование по всем возможным символам алфавита A, а определение символьного коррелятора приведено ниже в (15). Мы видим, что число параметров символьной корреляционной функции является большим, чем числовой, вследствие чего может более подробно описывать случайные последовательности.

На данный момент известно множество алгоритмов генерации последовательностей с дальними корреляциями: генерация Мандельброта дробного гауссового шума [4], процедура Восса последовательных случайных добавок [5], коррелированные блуждания Леви [6], метод Ли дефляции-модификации [7], метод свертки [8–20] (и его вариант – метод фильтрации Фурье [16]), метод цепей Маркова [20] и т. д. Отметим, что упомянутые методы хорошо зарекомендовали себя при генерации случайных последовательностей, когда значение каждого члена случайной последовательности (пространство состояний) принадлежит множеству вещественных чисел $R = (-\infty, \infty)$. В последнее десятилетие достигнут существенный прогресс в генерации бинарных (двухсимвольных) последовательностей. Методы генерации многосимвольных последовательностей, о которых идет речь в настоящей работе, практически не разработаны.

Структура работы следующая. В следующем разделе вводятся некоторые определения и обозначения. В разд. 2 мы формулируем проблему: приводим определение аддитивных цепей Маркова высших порядков, описываем процедуру декомпозиции функции условной вероятности (ФУВ), для билинейной ФУВ получаем уравнение, связывающее тройной коррелятор с соответствующими функциями памяти, и находим решение для функции памяти, выражая его через корреляционные функции второго и третьего порядков в предположении слабости корреляций. В разд. 3 рассматривается общая k-линейная форма функции условной вероятности и приводится решение для функции памяти k-го порядка, выраженной через соответствующий k-й коррелятор. В завершение, в разд. 4, приводится пример численной реализация метода построения случайной последовательности с заданными конкурирующими матричными корреляторами второго и третьего порядков.

1. Символьные марковские цепи. Рассмотрим полубесконечную случайную стационарную эргодическую последовательность S абстрактных символов (букв) a_i

$$S = a_0, a_1, a_2, \dots, \tag{4}$$

выбранных из конечного алфавита – множества, состоящего из m элементов:

$$A = \alpha^1, \alpha^2, \alpha^3, \dots, \alpha^m, a_i \in A. \tag{5}$$

Мы используем обозначение a_i, чтобы указать положение i-го символа a в цепи, и унифицированное обозначение α^k, чтобы подчеркнуть значение, которое принимает символ $a \in A$. Мы также используем персонифицированные обозначения для символов a алфавита, $A = \{\alpha, \beta, \dots, \omega\}$.

Последовательность S является N-шаговой цепью Маркова, если она обладает следующим свойством: вероятность того, что символ a_i, $i \geq N$, имеет определенное значение $\alpha^k \in A$ при условии, что все предыдущие символы заданы, зависит только от N предыдущих символов,

$$\begin{aligned} &P(a_i = \alpha^k \mid \dots, a_{i-2}, a_{i-1}) = \\ &= P(a_i = \alpha^k \mid a_{i-N} \dots, a_{i-2}, a_{i-1}). \end{aligned} \tag{6}$$

Рассматриваемые символьные последовательности называются также цепями Маркова высшего порядка [21–25], если $N \geq 2$, либо многошаговыми [20, 26, 27] цепями Маркова. Значение N называют порядком или длиной памяти марковской цепи, а символ $a_i = \alpha^k$ – генерируемым. Заметим, что определение (6) справедливо для $i \geq N$; для $i < N$ следует использовать хорошо известные условия совместности для функции условной вероятности более низкого порядка [28].

Заметим, что независимость функции $P(a_i = \alpha \mid a_{i-N}^{i-1})$ от индекса i обеспечивает однородность и стационарность рассматриваемой последовательности.

Мы предполагаем, что цепь является эргодической. В соответствии с теоремой Маркова (см., например, [28]), это свойство верно для всех однородных цепей Маркова, если строгие неравенства

$$0 < P(a_i = \alpha \mid a_{i-N}^{i-1}) < 1, \quad i \ge N,$$

налагаемые на функцию условной вероятности, выполняются для всех возможных значений аргументов функции (6). Здесь и далее мы используем краткое обозначение a_{i-N}^{i-1} для N-слова $a_{i-N} \dots, a_{i-2}, a_{i-1} = a_{i-N}^{i-1}$. Из эргодичности следует, что корреляции между любыми блоками символов в цепи стремятся к нулю, если расстояние между ними стремится к бесконечности. Другим следствием эргодичности является возможность использовать одну случайную последовательность как эквивалентное представление ансамбля и использовать усреднение по последовательности (7) вместо усреднения по ансамблю, т. е. среднее значение любой функции $f(a_{r_1}, a_{r_1+r_2}, \dots, a_{r_1+\dots+r_s})$, зависящей от s аргументов и определенной на подмножестве S, является статистическим (арифметическим, Чезаро) средним по цепи:

$$\overline{f(a_{r_1}, a_{r_1+r_2}, \dots, a_{r_1+\dots+r_s})} = \\ = \lim_{N\to\infty} \frac{1}{M} \sum_{i=0}^{M-1} f(a_{i+r_1}, a_{i+r_1+r_2}, \dots, a_{i+r_1+\dots+r_s}). \tag{7}$$

Если последовательность, статистические свойства которой мы хотим анализировать, задана, то ФУВ N-го порядка может быть найдена стандартным методом

$$P(a_i = \alpha \mid a_{i-N} \dots, a_{i-1}) = \frac{P(a_{i-N} \dots, a_{i-1}, \alpha)}{P(a_{i-N} \dots, a_{i-1})}, \tag{8}$$

где $P(a_{i-N} \dots, a_{i-1}, \alpha)$ и $P(a_{i-N} \dots, a_{i-1})$ – вероятности реализации $(N+1)$-подпоследовательности $a_{i-N} \dots, a_{i-1}, \alpha$ и N-подпоследовательности $a_{i-N} \dots, a_{i-1}$, соответственно. Здесь мы опустили и далее часто будем опускать индекс k в α^k для упрощения записи.

2. Формулировка задачи

2.1. *Оценка функции условной вероятности.* Функция условной вероятности полностью определяет все статистические свойства случайной цепи и метод ее построения. Из уравнения (8) видно, что ФУВ определена, если известны вероятности возникновения $(N+1)$-слов. Эти вероятности можно вычислить с помощью принципа максимального правдоподобия,

$$P(a_{i-N} \dots, a_{i-1}, a) = \frac{n(a_{i-N} \dots, a_{i-1}, a)}{M - L + 1}. \tag{9}$$

Здесь $n(a_{i-N} \dots, a_{i-1}, a)$ – количество реализаций слова $a_{i-N} \dots, a_{i-1}, a$ в последовательности длиной M. Таким образом, мы вычисляем вероятность $P(a_{i-N} \dots, a_{i-1})$ как частоту реализации слов и получаем инструмент построения случайной цепи.

Этот инструмент хорошо работает, если длина изучаемой случайной цепи M и $L_{\max}$ удовлетворяют неравенству

$$N < L_{\max} \approx \frac{\ln M}{\ln m}. \tag{10}$$

При $N > L_{\max}$ метод адекватно восстанавливает статистические свойства случайной цепи только на масштабах $L << L_{\max}$.

2.2. *Аддитивные цепи Маркова высших порядков.* Это неудовлетворительное заключение предыдущего подраздела побудило нас предложить метод [29, 30], основанный на аддитивных цепях Маркова высших порядков, где функция условной вероятности принимает вид

$$P(a_i = \alpha \mid a_{i-N} \dots, a_{i-2}, a_{i-1}) = \\ = p_\alpha + \sum_{r=1}^{N} \sum_{\beta\in A} F_{\alpha\beta}(r)[\delta(a_{i-r}, \beta) - p_\beta], \tag{11}$$

где p_α – относительное число символов α в цепи, или вероятность их возникновения,

$$p_\alpha = \overline{\delta(a_i, \alpha)}, \tag{12}$$

где $\delta(...)$ – дельта-символ Кронекера, который играет роль характеристической функции случайной переменной a_i и переводит символ в число. Аддитивность цепи означает, что предыдущие символы вносят независимый вклад в вероятность возникновения генерируемого символа $a_i = \alpha$. Первый член в правой части уравнения (11) отвечает за воспроизведение статистических свойств некоррелированной последовательности, второй – учитывает и корректно воспроизводит парные корреляции между символами случайной последовательности.

Было предложено два метода поиска функции $F_{\alpha\beta}(r)$, называемой функцией памяти, для последовательности с известной парной корреляционной функцией. Первый метод [31] основан на минимизации «расстояния» между функцией условной вероятности, содержащей функцию памяти, и данной последовательностью S символов с известной корреляционной функцией,

$$Dist = \overline{[\delta(a_i, \alpha) - P(a_i = \alpha \mid a_{i-N}^{i-1})]^2}. \tag{13}$$

Для любых значений $\alpha, \beta \in A$ и $r \ge 1$ минимизация относительно $F_{\alpha\beta}(r)$ дает зависимость между корреляционной функцией и функцией памяти:

$$C_{\alpha\beta}(r) = \sum_{r'=1}^{N} \sum_{\gamma\in A} C_{\alpha\gamma}(r - r') F_{\beta\gamma}(r'). \tag{14}$$

Второй метод получения (14) состоит в прямом вычислении вероятностей реализации определенной комбинации символов, задающих корреляционную функцию, и аналогичном представленному в [26]. Его модификация используется для получения уравнений (28) и (29).

Двухточечная символьная корреляционная функция (14) определяется следующим образом:

$$C_{\alpha\beta}(r)=\overline{[\delta(a_i,\alpha)-p_\alpha][\delta(a_{i+r},\beta)-p_\beta]}. \tag{15}$$

Эта функция обладает следующими свойствами:

$$C_{\alpha\beta}(r)=C_{\beta\alpha}(-r),\\ \sum_{\alpha\in A}C_{\alpha\beta}(r)=\sum_{\beta\in A}C_{\alpha\beta}(r)=0. \tag{16}$$

Следующий шаг, который мы делаем в настоящей работе, направленный на улучшение качества предсказания символов случайной цепи, а тем самым и качества генерации, основан на использовании k-линейной функции условной вероятности, $k\geq 2$.

2.3. *Мультилинейная функция условной вероятности.* Уравнение (11) можно рассматривать как приближенное модельное выражение, упрощающее общий вид функции условной вероятности. Вообще говоря, условная вероятность (8) символьной последовательности случайных переменных $a_i\in A$ может быть представлена точно в виде конечного полиномиального ряда, содержащего N дельта-символов Кронекера, определенным образом выделенного из ФУВ:

$$P(.|.)=P(a_i=\alpha\,|\,a_{i-N}\ldots,a_{i-2},a_{i-1})=p_\alpha+\sum_{\beta_1\ldots\beta_N\in A}\sum_{r_1\ldots r_N}F_{\alpha;\beta_1\ldots\beta_N}(r_1\ldots r_N)\prod_{s=1}^{N}\delta(a_{i-r_s},\beta_s). \tag{17}$$

Здесь аргументы $r_1\ldots r_N$ функции $F_{\alpha;\beta_1\ldots\beta_N}(r_1\ldots r_N)$, которые предполагаются расположенными в порядке возрастания, $r_1\leq r_2\leq\ldots\leq r_{N-1}\leq r_N$, обозначают расстояния между «генерируемым» символом $a_i=\alpha$ и символами $a_{i-1}=\beta_1,\ldots,a_{i-N}=\beta_N$. Ясно, что имеет место взаимно-однозначное соответствие между $P(a_i=\alpha\,|\,a_{i-N}^{i-1})$ и функцией $F_{\alpha;\beta_1\ldots\beta_N}(r_1\ldots r_N)$, которая называется обобщенной функцией памяти. Характеристические функции $\delta(a_{i-r_s},\beta_s)$ играют роль базиса, а обобщенные функции памяти – являются координатами ФУВ. Надеемся, что читатель обратил внимание на различие между обозначениями $\alpha^k, k\in(1,\ldots,m)$, и α_s, где индекс $s\in(1,\ldots,N)$ нумерует различные наборы символов.

Суммы в (17) содержат множество идентичных слагаемых с одинаковыми аргументами $r_i=r_{i+1}$, $i=1,\ldots,N-1$, смысл которых следует прояснить. Для удобства их рассмотрения полезно выделить их, собрав члены одинакового порядка по k в $Q^{(k)}(.|.)$:

$$P^{(k)}(.|.)=P^{(k-1)}(.|.)+Q^{(k)}(.|.),\ k=2,\ldots,N, \tag{18}$$

со следующими «начальными условиями», см. уравнение (11):

$$P^{(0)}=Q^{(0)}=p_\alpha,\\ Q^{(1)}=\sum_{r=1}^{N}\sum_{\beta\in A}F_{\alpha\beta}(r)[\delta(a_{i-r},\beta)-p_\beta]. \tag{19}$$

Вид члена $Q^{(1)}$ следует из (17), когда из N аргументов $r_1,\ldots,r_N$ $(N-1)$ совпадают. Главный член k-го порядка, называемый в мультилинейной алгебре k-линейной формой, равен

$$Q^{(k)}(.|.)=\sum_{\beta_1\ldots\beta_k\in A}\ \sum_{r_k=r_{k-1}+1}^{N}\ldots\sum_{r_1=1}^{N}F_{\alpha;\beta_1\ldots\beta_k}(r_1\ldots r_k)\{\prod_{s=1}^{k}[\delta(a_{i-r_s},\beta_s)-p_{\beta_s}]-C_{\beta_k\ldots\beta_1}(r_k-r_{k-1},\ldots,r_k-r_1)\}. \tag{20}$$

Символьная корреляционная функция k-го порядка равна

$$C_{\alpha_1,\ldots\alpha_k}(r_1,\ldots,r_{k-1})=\\ =\overline{[\delta(a_0,\alpha_1)-p_{\alpha_1}]\ldots[\delta(a_{r_1+\ldots+r_{k-1}},\alpha_k)-p_{\alpha_k}]}. \tag{21}$$

Функция $C_{\alpha_1\ldots\alpha_k}$ зависит, вообще говоря, от k аргументов, но рассматриваемая марковская цепь – однородна, поэтому, фактически, функция $C_{\alpha_1\ldots\alpha_k}$ зависит только от k–1 аргументов – расстояний $r_1,r_2\ldots,r_{k-1}$ между индексами соседних элементов. Тривиальное свойство функции $C_{\alpha_1,\ldots\alpha_k}(r_1,\ldots,r_{k-1})$, эквивалентное (16), имеет вид

$$\sum_{a_m}C_{\alpha_1,\ldots\alpha_k}(r_1,\ldots,r_{k-1})=0. \tag{22}$$

Это свойство отвечает за выполнение условия нормировки $\sum_{\alpha} P(a_i = \alpha \mid a_{i-N}^{i-1}) = 1$.

Определение (21) корректно для $r_i > 0,\ i = 1,\ldots,k-1$. Если некоторые из аргументов $r_1,\ldots,r_{k-1}$ корреляционной функции в (21) отрицательны, то их следует упорядочить [32] в соответствии с исходным определением (21). Например,

$$\begin{aligned} &C_{\alpha\beta\gamma}(3,-2) = \\ &= \overline{[\delta(a_0,\alpha)-p_\alpha][\delta(a_3,\beta)-p_\beta][\delta(a_1,\gamma)-p_\gamma]} = \\ &= \overline{[\delta(a_0,\alpha)-p_\alpha][\delta(a_1,\gamma)-p_\gamma][\delta(a_3,\beta)-p_\beta]} = \\ &= C_{\alpha\gamma\beta}(1,2). \end{aligned} \tag{23}$$

Мы использовали этот метод для представления корреляционной функции в выражении (20) в упорядоченном виде.

Последний член $Q^{(N)}$ в уравнении (18) содержит полностью различные аргументы $r_1,\ldots,r_k$. Для каждого фиксированного набора символов $\alpha;\beta_1,\ldots,\beta_N$ существует всего одна функция-константа $F_{\alpha;\beta_1\ldots\beta_N}(1\ldots N)$.

Мультилинейная функция условной вероятности $P^{(k)}(.|.)$ в форме (20) может корректно воспроизводить корреляции N-шаговой марковской цепи вплоть до корреляционных функций k-го порядка. При значении $k = N$ функция $P^{(N)}(.|.)$ представляет точно функцию $P(a_i = \alpha \mid a_{i-N}^{i-1})$. Заметим, что мы добавили в выражение (20) член $C_{\alpha_1,\ldots\alpha_k}$, чтобы обеспечить равенство $\overline{Q^{(k)}(.|.)} = 0$ для всех $k = 1, 2,\ldots, N$.

Таким образом, функция условной вероятности $P(a_i = \alpha \mid a_{i-N}^{i-1})$ представлена как разложение мультилинейной формы на k-линейные подпространства. Хосейни, Леб и Зейдека [33] строго доказали, что условная вероятность может быть записана как линейная комбинация мономиалов марковской цепи. Ранее эта идея была высказана в статье Бесага [34]. Мы изложили эту идею на физическом уровне строгости и терминологии.

Полезность описанной процедуры разложения можно объяснить следующим образом. Во-первых, величины $Q^{(k)}$ в ФУВ главным образом отвечают за воспроизводство корреляционных свойств k-го порядка, см. ниже (35). И, во-вторых, разложения (17) и (18) могут рассматриваться как асимптотические последовательные приближения для ФУВ.

2.4. *Билинейная ФУВ.* Модель аддитивных цепей Маркова высших порядков хорошо изучена [20]. В этом разделе исследуются N-шаговые марковские цепи (высших порядков) с билинейной функцией условной вероятности. Правая часть выражения (11) содержит два первых члена асимптотического приближения точного выражения (17). Следующий член $Q^{(2)}$ равен

$$\begin{aligned} &Q^{(2)}(.|.) = Q^{(2)}(a_i = \alpha \mid a_{i-N}\ldots,a_{i-2},a_{i-1}) = \\ &= \sum_{r_2=r_1+1}^{N}\sum_{r_1=1}^{N-1}\sum_{\beta,\gamma\in A} F_{\alpha;\beta\gamma}(r_1,r_2)\{[\delta(a_{i-r_1},\beta)-p_\beta][\delta(a_{i-r_2},\gamma)-p_\gamma]-C_{\gamma\beta}(r_2-r_1)\}. \end{aligned} \tag{24}$$

Функция условной вероятности, содержащая члены из (11) и (24), задает билинейную марковскую цепь, или, более точно, марковскую цепь с билинейной функцией памяти, см. определения (11), (19) и (24),

$$P(.|.) = Q^{(0)}(.|.) + Q^{(1)}(.|.) + Q^{(2)}(.|.). \tag{25}$$

Для корреляционных функций N-шаговой марковской цепи можно найти рекуррентные соотношения. Для этого, во-первых, надо вычислить среднее значение символа $a_{r_1+\ldots+r_{k-1}}$ в (21). Учитывая тождество $P(a=\alpha\mid .) + P(a\neq\alpha\mid .) = 1$, можно переписать (21) для произвольного значения $k \geq 2$ в виде

$$\begin{aligned} &\overline{[\delta(a_0,\alpha_1)-p_{\alpha_1}]\ldots[\delta(a_{r_1+\ldots+r_{k-1}},\alpha_k)-p_{\alpha_k}]} = \\ &= \overline{[\delta(a_0,\alpha_1)-p_{\alpha_1}]\ldots[P(a_{r_1+\ldots+r_{k-1}}=\alpha_k\mid .)-p_{\alpha_k}]}, \end{aligned} \tag{26}$$

где ФУВ $P(a_{r_1+\ldots+r_{k-1}}=\alpha_k\mid .)$ определена с помощью соотношения (17). Таким образом, мы можем получить фундаментальное рекуррентное соотношение, связывающее корреляционные функции различных порядков k. Здесь мы ограничимся представлением этих выражений для корреляционных функций $C_{\alpha\beta}(r)$ и $C_{\alpha\beta\gamma}(r_1,r_2)$, и билинейной функцией условной вероятности (25):

$$P(a_i=\alpha\,|\,.)=p_\alpha+\sum_{r_1=1}^{N}\sum_{\beta\in A}F_{\alpha\beta}(r_1)[\delta(a_{i-r_1},\beta)-p_\beta]+$$
$$+\sum_{r_1=1}^{N-1}\sum_{r_2=r_1+1}^{N}\sum_{\beta,\gamma\in A}F_{\alpha;\beta\gamma}(r_1,r_2)\{[\delta(a_{i-r_1},\beta)-p_\beta][\delta(a_{i-r_2},\gamma)-p_\gamma]-C_{\gamma\beta}(r_2-r_1)\}. \quad (27)$$

Для парной корреляционной функции $C_{\alpha\beta}(r)$ получаем

$$C_{\alpha\beta}(r)=\sum_{r_1=1}^{N}\sum_{\gamma\in A}C_{\alpha\gamma}(r-r_1)F_{\alpha\beta}(r_1)+\sum_{r_2=r_1+1}^{N}\sum_{r_1=1}^{N-1}\sum_{\gamma,\varepsilon\in A}F_{\beta:\gamma\varepsilon}(r_1,r_2)C_{\alpha\varepsilon\gamma}(r-r_2,r_2-r_1)\}. \quad (28)$$

Аналогично для трехточечной корреляционной функции $C_{\alpha\beta\gamma}(r_1,r_2)$ имеем соотношение

$$C_{\alpha\beta\gamma}(r_1,r_2)=\sum_{r_1'=1}^{N}\sum_{\eta\in A}C_{\alpha\beta\eta}(r_1,r_2-r_1')F_{\gamma\eta}(r_1')+$$
$$+\sum_{r_2'=r_1'+1}^{N}\sum_{r_1'=1}^{N-1}\sum_{\eta,\varepsilon\in A}F_{\gamma:\eta\varepsilon}(r_1',r_2')[C_{\alpha\beta\varepsilon\eta}(r_1,r_2-r_2',r_2'-r_1')-C_{\alpha\beta}(r_1)C_{\varepsilon\eta}(r_2'-r_1')]. \quad (29)$$

Можно показать, что соотношение (28) следует из (29). Для проверки этого утверждения следует в (29) два из трех символов положить равными, например, $\alpha=\beta,$ а координату $r_1=0.$ Чтобы показать эквивалентность (12) и (29), полагаем все символы в (29) равными друг другу, $\alpha=\beta=\gamma$ и $r_1=r_2=0.$ Другой способ получения (28) и (29) основан на минимизации «расстояния» (13).

Первый член правой части уравнения (11) отвечает за правильное восстановление статистических свойств некоррелированных последовательностей. Функция условной вероятности в виде (11) может воспроизводить корректно парные (двухточечные) корреляции в цепи. Выражения (28) и (29) свидетельствуют о том, что корреляции (и любые корреляционные свойства) высших порядков не являются независимыми. Мы не можем их контролировать и корректно воспроизводить с помощью функции памяти $F_{\alpha\beta}(r),$ так как она полностью определяется парной корреляционной функцией. Когда мы генерируем аддитивную марковскую цепь, функция памяти второго порядка $F_{\gamma;\eta\varepsilon}(r_1',r_2')$ равняется нулю. Корреляционная функция третьего порядка $C_{\alpha\beta\gamma}(r_1,r_2),$ заданная соотношением (29), зависит от функции памяти $F_{\beta\gamma}(r_1),$ которая, в свою очередь, есть функция парного коррелятора в соответствии с (28) (где также $F_{\gamma;\eta\varepsilon}(r_1',r_2')=0$).

Аналогичные заключения можно сделать и о функциях памяти k-го порядка. Функция условной вероятности, выраженная через функцию памяти k-го порядка, позволяет генерировать последовательности, в которых мы можем контролировать корреляционные функции k-го порядка. N-шаговая марковская цепь с N-линейной функцией памяти позволяет корректно восстанавливать свойства цепей вплоть до корреляционных функций N-го порядка.

2.5. *Решение уравнений* (28) *и* (29). Уравнения (28) и (29) могут быть решены аналитически только в некоторых частных случаях (одно- и двухшаговых цепей, марковских цепей со ступенчатой функцией памяти). Приведем их приближенные решения в предположении, что корреляции в последовательности не слишком сильные (по амплитуде, но не по расстоянию), и что алфавит A содержит много букв. Чтобы сформулировать эти условия, введем нормированную символьную корреляционную функцию, определенную как

$$K_{\alpha\beta}(r)=\frac{C_{\alpha\beta}(r)}{C_{\alpha\beta}(0)},$$
$$C_{\alpha\beta}(0)=p_\alpha\delta(\alpha,\beta)-p_\alpha p_\beta. \quad (30)$$

Если корреляции в цепи не являются сильными, естественно предположить, что все компоненты нормированной корреляционной функции при $r\neq 0$ малы по сравнению с $K_{\alpha\beta}(0)=1.$

Пренебрегая вторым членом в правой части (28) (корректность этой процедуры объясняется ниже, после (37)), получаем уравнение (14). Его решение может быть записано в виде

$$F_{\alpha\beta}(r)=\frac{1}{p_\alpha}C_{\alpha\beta}(r), \quad (31)$$

если в определении (30) величины $C_{\alpha\beta}(0)$ можно пренебречь членом $p_\alpha p_\beta$ по сравнению с p_α. Это возможно, если размер $|A|$ алфавита A удовлетворят условию

$$|A| = m >> 1, \tag{32}$$

т. е. все вероятности p_α малы по сравнению с единицей.

Анализ выражения (29) показывает, что в нем можно пренебречь первым членом в правой части уравнения по сравнению с членом $C_{\alpha\beta\gamma}(r_1, r_2)$. Это связано с тем, что функция $F_{\gamma\eta}(r_1')$ содержит только малую недиагональную компоненту коррелятора $C_{\eta\gamma}(r_1')$. По этой же причине член $C_{\alpha\beta}(r_1)C_{\varepsilon\eta}(r_2' - r_1')$ мал по сравнению с функцией $C_{\alpha\beta\varepsilon\eta}(r_1, r_2 - r_2', r_2' - r_1')$. Последнее утверждение следует из оценки коррелятора $C_{\alpha\beta\varepsilon\eta}(r_1, r_2 - r_2', r_2' - r_1')$. Его наибольшая компонента, удовлетворяющая условиям $r_2' \geq r_1' + 1$, $r_1' \geq 1$, равна $C_{\alpha\beta\alpha\beta}(r_1, -r_1, r_1)$ при $r_1' = r_2$, $r_2' = r_1 + r_2$. Таким образом, соотношение (29) сводится к следующему:

$$C_{\alpha\beta\gamma}(r_1, r_2) = F_{\gamma;\alpha\beta}(r_2, r_1 + r_2)\, C_{\alpha\beta\alpha\beta}(r_1, -r_1, r_1). \tag{33}$$

Учитывая (32) и пренебрегая корреляциями при вычислении $C_{\alpha\beta\alpha\beta}(r_1, -r_1, r_1)$, получаем

$$F_{\gamma;\alpha\beta}(r_2, r_1 + r_2) = \frac{1}{p_\alpha p_\beta} C_{\alpha\beta\gamma}(r_2 - r_1, r_1). \tag{34}$$

Выражение (17) для функции условной вероятности символьной марковской цепи высокого порядка с билинейной функцией памяти (в первом приближении по малым параметрам $|C_{\alpha\beta}(r)| << 1, r \neq 0$ и в случае многобуквенного алфавита, $p_\alpha << 1$)) принимает вид

$$\begin{gathered} P(a_i = \alpha \mid a_{i-N}^{i-1}) \approx p_\alpha + \sum_{r=1}^{N} \sum_{\beta \in A} \frac{1}{p_\beta} C_{\beta\alpha}(r_1)[\delta(a_{i-r_1}, \beta) - p_\beta] + \\ + \sum_{r_1=1}^{N-1} \sum_{r_2=r_1+1}^{N} \sum_{\beta,\gamma \in A} \frac{1}{p_\beta p_\gamma} C_{\beta\gamma\alpha}(r_2 - r_1, r_1)[\delta(a_{i-r_2}, \beta) - p_\beta][\delta(a_{i-r_1}, \gamma) - p_\gamma]. \end{gathered} \tag{35}$$

Это соотношение дает инструмент для построения случайных слабокоррелированных последовательностей с заданными корреляционными функциями второго и третьего порядков.

3.1. *k-линейная форма ФУВ.* Уравнения (31) и (34) показывают, что можно надеяться получить аналогичные выражения для обобщенных функций памяти *k*-го порядка, выраженных через корреляционные функции. Действительно, результат анализа и вычислений выглядит следующим образом:

$$\begin{gathered} F_{\alpha_{k+1};\alpha_k \ldots \alpha_1}(r_k, r_k + r_{k-1}, \ldots, r_k + \ldots + r_1) = \\ = \frac{1}{p_{\alpha_1} \cdots p_{\alpha_k}} C_{\alpha_1 \ldots \alpha_{k+1}}(r_1, \ldots, r_k). \end{gathered} \tag{36}$$

Или, после изменения «системы отсчета» функций памяти и корреляционных функций, получаем

$$\begin{gathered} F_{\alpha;\alpha_1 \ldots \alpha_k}(r_1, \ldots, r_k) = \\ = \frac{1}{p_{\alpha_1} \cdots p_{\alpha_k}} C_{\alpha_k \ldots \alpha_1 \alpha}(r_k - r_{k-1}, \ldots, r_2 - r_1, r_1). \end{gathered} \tag{37}$$

Для получения этого результата мы применяем тот же метод, что был использован для (31) и (34). Перечислим основные шаги этой процедуры:

1) вычисляем корреляционную функцию $(k+1)$ порядка $C_{\alpha_1 \ldots \alpha_{k+1}}(r_1, \ldots, r_k)$, $1 \leq k \leq N$, используя (26);

2) при вычислении используем $P(.|.) = P^{(N)}(.|.) = \sum_{k=0}^{N} Q^{(k)}(.|.)$;

3) в результате корреляционная функция представляется в виде суммы функции памяти разных порядков от 1 до N с коэффициентами $C_{\alpha'_1 \alpha'_2 \ldots}(r_1', r_2' \ldots)$;

4) в главном слагаемом суммы, содержащем

$$\sum_{r',\beta} F_{a_{k+1};\beta_1 \ldots \beta_k}(r_1', \ldots, r_k')\, C_{\alpha_1 \ldots \alpha_k \beta_k \ldots \beta_1}(r_1, \ldots, r_{k-1}, r_k - r_k', r_{k-1}' - r_{k-2}', \ldots, r_2' - r_1'),$$

находим максимальный член $C_{\alpha_1 \ldots \beta_1}(r_1, \ldots, r_2' - r_1')$. Его значение максимально, если в двух возрастающих последовательностях $0, r_1, \ldots, r_{k-1}$ и $r_k - r_k', r_{k-1}' - r_{k-2}', \ldots, r_2' - r_1'$ (переписанных в соответствии с процедурой упорядочения как $0, r_1, r_1 + r_2, \ldots, r_1 + r_2 + \ldots + r_{k-1}$

и

$r_1 + \ldots + r_k - r'_k, r_1 + \ldots + r_k - r'_{k-1}, \ldots, r_1 + \ldots + r_k - r'_1)$, имеет место взаимно-однозначное соответствие их элементов $r'_1 = r_k, r'_2 = r_k + r_{k-1}$ и $r'_k = r_k + \ldots + r_1$;

5) пренебрегая в вычислениях корреляциями, получаем

$C_{\alpha_1 \ldots \beta_1}(r_1, \ldots, r'_2 - r'_1) = \prod_{s=1}^{k} p_{\alpha_s} \delta(\alpha_s, \beta_s)$;

6) все остальные члены, содержащие $F_{a_{s+1}; \beta_1 \ldots \beta_s}(r'_1, \ldots, r'_s)$ с $s \neq k$, являются малыми относительно учтенных; они содержат дополнительные малые множители p^r, $r \geq 1$.

Таким образом, функция условной вероятности (17) для символьной цепи Маркова высокого порядка (в первом приближении по малым параметрам $|C_{\alpha\beta}(r)| << 1$, $r \neq 0$ и в случае многобуквенного алфавита $p_\alpha << 1$) выражается через «экспериментально» измеренные величины, т. е. корреляционные функции. Учитывая свойство (22), $\sum_{\alpha_m} C_{\alpha_1, \ldots \alpha_k}(r_1, \ldots, r_{k-1}) = 0$, конечный результат удобно представить в виде

$$P(a_i = \alpha \mid a_{i-N}^{i-1}) \approx p_\alpha + \sum_{r=1}^{N} \frac{1}{p_{\alpha_{i-r}}} C_{a_{i-r}\alpha}(r) + \sum_{r_1=1}^{N-1} \sum_{r_2=r_1+1}^{N} \frac{1}{p_{\alpha_{i-r_1}} p_{\alpha_{i-r_2}}} C_{a_{i-r_2} a_{i-r_1} \alpha}(r_2 - r_1, r_1) + \ldots$$
$$\ldots + \sum_{r_1=1}^{N-k} \ldots \sum_{r_k=r_{k-1}+1}^{N} C_{a_{i-r_k} \ldots a_{i-r_1} \alpha}(r_k - r_{k-1}, \ldots, r_2 - r_1, r_1) \prod_{s=1}^{k} p_{a_{i-r_s}}^{-1} + \ldots \quad (38)$$
$$\ldots + C_{a_{i-r_N} \ldots a_{i-r_1} \alpha}(r_N - r_{N-1}, \ldots, r_2 - r_1, r_1) \prod_{s=1}^{N} p_{a_{i-r_s}}^{-1}.$$

Формула (38) дает инструмент построения слабокоррелированной случайной последовательности символов с заданными корреляционными функциями $C_{\alpha_1, \ldots \alpha_k}(r_1, \ldots, r_{k-1}), k \leq N$.

Как видим, если корреляции слабые, все члены ФУВ независимы друг от друга. Таким образом, если мы строим, например, последовательность, используя члены нулевого порядка и билинейный, то остальные корреляторы равны нулю, за исключением коррелятора третьего порядка. В общем случае это не так. При генерации последовательности с произвольной функцией памяти возникают корреляции всех порядков.

3.2. *Случай, когда вероятности p_α не малы.* Если изучаемая последовательность не удовлетворяет условию $p_\alpha << 1$, как, например, в случае нуклеотидной последовательности ДНК, где все четыре вероятности *p* реализации различных нуклеотидов приблизительно равны 1/4, использовать выражение (38) для получения ФУВ по корреляционным функциям оказывается невозможно. Чтобы обойти эту трудность, следует уменьшить каким-либо способом вероятности *p*. Для этой цели можно использовать идеи, изложенные в [38, 39]. В нескольких работах Хименес-Монтанье, Эбелинг и другие предложили схемы кодировки случайной последовательности символов посредством рекурсивных парных подстановок, при которых наиболее распространенная комбинация пары символов во всех местах своего появления в последовательности замещается новым символом, при условии, что количество появлений комбинации в последовательности больше двух. Эта операция проводится до тех пор, пока не окажется, что в последовательности не находится ни одной пары символов, которая встречалась бы более двух раз. Каждая успешная подстановка сопровождается уменьшением вероятности $p_{\tilde{\alpha}}$, где $\tilde{\alpha}$ принадлежит новому расширенному алфавиту.

Стандартный путь нахождения функций памяти по корреляционным функциям, как было отмечено выше, состоит в численном решении уравнений (28) и (29) или аналогичных им уравнений, если речь идет о восстановлении корреляционных свойств последовательности с более высокой точностью, чем это позволяют сделать корреляторы третьего порядка.

4. Пример численной генерации случайной последовательности. В настоящем разделе мы приводим пример численного построения случайной символьной последовательности с пространством состояний $|A| = 4$.

Задача состоит в построении случайной последовательности с заданными статистическими свойствами, в данном случае с вероятностями реализации символов p_α и корреляционными функциями второго $C_{\alpha\beta}(r)$ и третьего порядков $C_{\alpha\beta\gamma}(r_1, r_2)$, $\alpha, \beta, \gamma = \{1, 2, 3, 4\}$.

Подставляя заданные величины в систему уравнений (29) и решая ее относительно

$F_{\alpha\beta}(r)$ и $F_{\alpha\beta\gamma}(r_1,r_2)$, находим значения функции памяти, которые, в свою очередь, задают ФУВ и метод построения цепи.

Если требуется построить случайную последовательность, обладающую теми же корреляционными свойствами, что и данная цепь, то есть решить обратную задачу, то следует произвести те же действия, предварительно рассчитав численно значения p_α, $C_{\alpha\beta}(r)$ и $C_{\alpha\beta\gamma}(r_1,r_2)$ по данной цепи в предположении, что она является репрезентативной.

В качестве примера решения прямой задачи рассмотрим генерацию случайной последовательности с билинейной ФУВ (35), где матрицы-корреляторы второго и третьего порядков заданы аналитически в виде следующих выражений:

$$\begin{aligned} &C_{\alpha\beta}(r)=C^{(2)}_{\alpha\beta}C^{(2)}(r),\\ &C_{\alpha\beta\gamma}(r_1,r_2)=C^{(3)}_{\alpha\beta\gamma}C^{(3)}(r_1,r_2). \end{aligned} \tag{39}$$

Здесь корреляторы разделены на независимые множители, задающие отдельно зависимость корреляций от расстояния и зависимость матричной части от значений символов.

Матричные множители для простоты выбираем симметричными по символам – они не будут зависеть от конкретных значений символов, а только от того, оказались ли какие-то из этих значений равны друг другу, или все различны. Иными словами, они имеют диагональный вид,

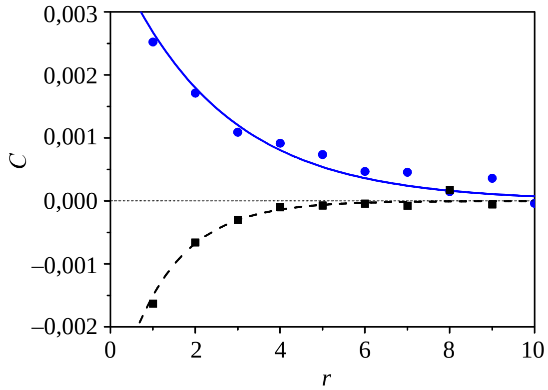
$$\begin{aligned} &C^{(2)}_{\alpha\beta}=\delta(\alpha,\beta)+A,\\ &C^{(3)}_{\alpha\beta\gamma}=\delta(\alpha,\beta,\gamma)+\\ &+B(\delta(\alpha,\beta)+\delta(\beta,\gamma)+\delta(\alpha,\gamma))+C, \end{aligned} \tag{40}$$

где константы A, B и C выбраны таким образом, чтобы выполнялась нормировка корреляторов (22).

Зависимость от расстояния выбираем для обоих корреляторов в виде экспоненциальных функций, причем коррелятор второго порядка выбираем положительным, а третьего – отрицательным,

$$\begin{aligned} &C^{(2)}(r)=c_2e^{-k_2r},\\ &C^{(3)}(r_1,r_2)=c_3e^{-k_3r_1}e^{-k_3(r_1+r_2)},\\ &c_2=0{,}004,\quad c_3=-0{,}005,\\ &k_2=k_3=0{,}4. \end{aligned} \tag{41}$$

На рисунке показан результат сравнения этих аналитически заданных корреляционных функций (сплошная и пунктирная линии) и рассчитанных численно (символы) по случайной последовательности длиной $M=10^6$, построенной с помощью билинейной функции условной вероятности (35) на алфавите размером $m=4$. На графике показаны диагональные значения корреляторов $C_{00}(r)$ и $C_{000}(r,1)$. График демонстрирует хорошее, в пределах погрешности вычислений, совпадение восстановленных корреляторов с заданными аналитическими выражениями.



Аналитические и численно рассчитанные значения диагональных корреляционных функций $C_{00}(r)$ (сплошная линия и символы) и $C_{000}(r,1)$ (пунктирная линия и символы), заданных выражениями (39)–(41)

Выводы. Предложен метод генерации эргодических стационарных символических случайных последовательностей с конечным абстрактным пространством состояний на основе многошаговых марковских цепей. Произведена декомпозиция функции условной вероятности на независимые слагаемые, выраженные через матричные функции памяти. Для последних получены цепочки уравнений, позволяющих выразить функции памяти через корреляционные функции.

Показано, что в предельном случае слабых корреляций функции памяти выражаются через корреляционные функции высших порядков с помощью простых соотношений, что позволяет генерировать случайные последовательности, обладающие сложными заданными дальними корреляциями.

Приводится пример численной реализации метода построения случайной последовательности с заданными конкурирующими матричными корреляторами второго и третьего порядков.

Таким образом, решена важная задача, возникающая при конструировании радиотехнических устройств с заданными спектральными свойствами.

Библиографический список

1. *Амитей Н.* Теория и анализ фазированных антенных решеток / Н. Амитей, В. Галиндо, Ч. Ву; пер. с англ. под ред. А. Ф. Чаплина. – М.: Мир, 1974. – 455 с.

2. *Izrailev F. M.* Anomalous localization in low-dimensional systems with correlated disorder / F. M. Izrailev, A. A. Krokhin, N. M. Makarov // Phys. Rep. – 2012. – 512, Iss. 3. – P. 125–254.
3. *Лукин К. А.* Получение изображений с помощью неподвижной антенной решетки, шумовых сигналов и метода синтезирования апертуры / К. А. Лукин, А. А. Могила, П. Л. Выплавин // Радиофизика и электрон.: сб. науч. тр. / Ин-т радиофизики и электрон. НАН Украины. – Х., 2007. – 12, № 3. – С. 526–531.
4. *Mandelbrot B. B.* A Fast Fractional Gaussian Noise Generator / B. B. Mandelbrot, J. R. Wallis // Water Resour. Res. – 1971. – 7, Iss. 3. – P. 543–553.
5. *Voss R. F.* Fundamental Algorithms in Computer Graphics / R. F. Voss – Berlin: Springer, 1985. –805 p.
6. *Shlesinger M. F.* Strange kinetics / M. F. Shlesinger, G. M. Zaslavsky, J. Klafter // Nature. – 1993. – 363, N 6424. – P. 31–37.
7. *Li W.* Spatial 1/f Spectra in Open Dynamical Systems / W. Li // Europhys. Lett. – 1989. – 10. – P. 395–400.
8. *Rice S. O.* Mathematical analysis of random noise / S. O. Rice // Bell Syst. Tech. J. – 1944. – 23. – P. 282–332.
9. *Wax N.* Selected Papers on Noise and Stochastic Processes / N. Wax – N. Y.: Dover, 1953. – 343 p.
10. *Saupe D.* The Science of Fractal Images / D. Saupe. – N. Y.: Springer, 1988. – 312 p.
11. *West C. S.* Observations of backscattering enhancement from polaritons on a rough metal surface / C. S. West, K. A. O'Donnell // J. Opt. Soc. Am. A. – 1995. – 12, N 2. – P. 390–397.
12. *Izrailev F. M.* Localization and the Mobility Edge in One-Dimensional Potentials with Correlated Disorder / F. M. Izrailev, A. A. Krokhin // Phys. Rev. Lett. – 1999. – 82, N 20. – P. 4062–4065.
13. *Izrailev F. M.* Anomalous transport in low-dimensional systems with correlated disorder / F. M. Izrailev, N. M. Makarov // J. Phys. A: Math. Gen. – 2005. – 38, N 49. – P. 10613–10637.
14. *Cakir R.* Dynamical origin of memory and renewal / R. Cakir, P. Grigolini, A. A. Krokhin // Phys. Rev. E. – 2006. – 74, Iss. 2. – P. 021108(6 p.).
15. *Romero A.* Generation of short and long range temporal correlated noises / A. Romero, J. Sancho // J. Computational Phys. – 1999. – 156, Iss. 1. – P. 1–11.
16. *Correlations* in binary sequences and a generalized Zipf analysis / A. Czirok, R. N. Mantegna, S. Havlin, H. E. Stanley // Phys. Rev. E. – 1995. – 52, Iss. 1. – P. 446–452.
17. *Method* for generating long range correlations for large systems / H. A. Makse, S. Havlin, M. Schwartz, H. E. Stanley // Phys. Rev. E. – 1996. – 53, Iss. 5. – P. 5445–5449.
18. *Izrailev F. M.* Generation of correlated binary sequences from white noise / F. M. Izrailev, A. A. Krokhin, N. M. Makarov, O. V. Usatenko // Phys. Rev. E. – 2007. – 76, Iss. 2. – P. 027701(4 p.).
19. *Izrailev F. M.* Anomalous localization in low-dimensional systems with correlated disorder / F. M. Izrailev, A. A. Krokhin, N. M. Makarov // Phys. Rep. – 2012. – 512, Iss. 3. – P. 125–254.
20. *Random* Finite-Valued Dynamical Systems: Additive Markov Chain Approach / O. V. Usatenko, S. S. Apostolov, Z. A. Mayzelis, S. S. Melnik. – Cambridge: Cambridge Scientific Publ., 2010. – 176 p.
21. *Raftery A.* A model for high-order Markov chains / A. Raftery // J. R. Stat. Soc. B. – 1985. – 47, N 3. – P. 528–539.
22. *Ching W. K.* Higher-order Markov chain models for categorical data sequences / W. K. Ching, E. S. Fung, M. K. Ng // Naval Res. Logist. – 2004. – 51, N 4. – P. 557–574.
23. *Li W. K.* Some results on high order Markov chain models / W. K. Li, M. C. O. Kwok // Comun. Stat. Simul. Comput. – 1990. – 19. – P. 363–380.
24. *Bacterial* genomes lacking long-range correlations may not be modeled by low-order Markov chains: the role of mixing statistics and frame shift of neighboring genes / J. Cocho, P. Miramontes, R. Mansilla, W. Li // Comput. Biol. Chem. – 2014. – 53, pt. A. – P. 15–25.
25. *Parsimonious* Higher-Order Hidden Markov Models for Improved Array-CGH Analysis with Applications to Arabidopsis thaliana / M. Seifert, A. Gohr, M. Strickert, I. Grosse // PLoS Computat. Biol. – 2012. – 8, Iss. 1. – P. e1002286.
26. *Competition* of Two Types of Correlations / S. S. Melnyk, O. V. Usatenko, V. A. Yampol'skii, V. A. Golick // Phys. Rev. E. – 2005. – 72, Iss. 2. – P. 026140(7 p.).
27. *Usatenko O. V.* Binary N-Step Markov Chains and Long-Range Correlated Systems / O. V. Usatenko, V. A. Yampol'skii // Phys. Rev. Lett. – 2003. – 90, Iss. 11. – P. 110601(4 p.).
28. *Shiryaev A. N.* Probability / A. N. Shiryaev. – N. Y.: Springer, 1996. – 624 p.
29. *Melnik S. S.* Entropy and long-range correlations in DNA sequences / S. S. Melnik, O. V. Usatenko // Computational Biology and Chemistry – 2014. – 53, pt. A. – P. 26–31.
30. *Melnik S. S.* Entropy of finite random binary sequences with weak long-range correlations / S. S. Melnik, O. V. Usatenko // Pys. Rev. E – 2014. – 90, Iss. 5. – P. 052106(8 p.).
31. *Melnyk S. S.* Memory functions of the additive Markov chains: applications to complex dynamic systems / S. S. Melnyk, O. V. Usatenko, V. A. Yampol'skii // Physica A. – 2006. – 361, Iss. 2. – P. 405–415.
32. *High* Order Correlation Functions of Binary Multi-Step Markov Chains / S. S. Apostolov, Z. A. Mayzelis, O. V. Usatenko, V. A. Yampol'skii // Int. J. Mod. Phys. B. – 2008. – 22, N 22. – P. 3841–3853.
33. *Hosseinia R.* A Characterization of Categorical Markov Chains / R. Hosseinia, N. Leb, J. Zideka // Journal of Statistical Theory and Practice – 2011. Vol. 5, I. 2, – P. 261–284.
34. *Besag J.* Spatial interactions and the statistical analysis of lattice systems / J. Besag // J. Royal Statistical Soc. B. – 1974. – 36, N 2. – P. 192–225.

Рукопись поступила 09.07.2015.

S. S. Melnyk, O. V. Usatenko

SYMBOLIC MARKOV CHAINS WITH MULTILINEAR MEMORY FUNCTION

The problem of designing various radio devices, such as filters, delay lines, random antennas with a given radiation pattern and so on, requires the development of methods for constructing random sequences of the parameters of these systems having specified correlation properties. An adequate mathematical approach for solving such problems is the Markov chains of higher orders. Statistical characteristics of these objects are completely determined by their conditional probability function that, in general, can be very complicated. The purpose of this paper is to present the decomposition procedure for the conditional probability function of random sequences with long-range correlationtions in a form convenient for their numerical generation. Here we restrict ourselves to the case of the state space of the system of such kind, when random values of its elements belong to the finite abstract set. The developed theory opens the way to build a more consistent and nuanced approach for the description of systems with long-range correlations. In the limiting case of weak (by value, but not the distance) correlations memory function is uniquely expressed in terms of higher-order correlation functions, allowing us to generate a random sequence with a given long-range correlations. As an example of the analytical results obtained, which can be used in practical applications, we present an example of the numerical realization of the method of construction of random sequence with specified correlators of the second and third orders.

Key words: random sequences, correlation functions, the function of conditional probability, the high order Markov chains.

С. С. Мельник, О. В. Усатенко

СИМВОЛЬНІ ЛАНЦЮГИ МАРКОВА З МУЛЬТИЛІНІЙНОЮ ФУНКЦІЄЮ ПАМ'ЯТІ

Задача конструювання різних радіотехнічних пристроїв, таких як фільтри, лінії затримки, випадкові антени із заданою діаграмою спрямованості, потребує розробки методів генерації випадкових послідовностей (значень параметрів систем), які мають задані кореляційні властивості. Адекватним математичним апаратом для розв'язання задач такого типу є ланцюги Маркова вищих порядків. Статистичні характеристики цих об'єктів повністю визначаються їх функцією умовної імовірності, яка у загальному випадку може мати достатньо складний вигляд. Метою роботи є подання функції умовної імовірності випадкових символьних послідовностей з далекими кореляціями у вигляді, зручному для чисельної генерації послідовностей. Ми припускаємо, що простір станів системи є кінцевою абстрактною множиною. Виконується розклад функції умовної імовірності на незалежні складові, подані через так звану матрицю-функцію пам'яті. Розвинена теорія відкриває можливості для побудови більш послідовного і тонкого підходу до опису систем з далекими кореляціями. У випадку слабких (за величиною, але не за відстанню) кореляцій функції пам'яті однозначно виражаються через кореляційні функції вищих порядків, що дозволяє генерувати випадкові послідовності, які мають задані складні далекі кореляції. В якості прикладу використання отриманих аналітичних результатів наводиться чисельна реалізація методу побудови випадкової послідовності із заданими конкуруючими матричними кореляторами другого і третього порядків.

Ключові слова: випадкові послідовності, кореляційні функції, функція умовної імовірності, ланцюги Маркова вищих порядків.