

## ЗАКОН РАСПРЕДЕЛЕНИЯ ДЛИНЫ МАКСИМАЛЬНОЙ СЕРИИ И ЕГО СТАТИСТИЧЕСКИЕ ПРИЛОЖЕНИЯ

© 2006 А.Н. Плотников

Самарский государственный аэрокосмический университет

В статье обсуждается целесообразность использования длины максимальной серии в качестве статистики критерия случайности. Установлены и экспериментально подтверждены законы распределения указанной статистики для двух основных типов серий в случайной последовательности количественных данных. Получены таблицы доверительных интервалов и методики интервальной оценки длины максимальной серии, предназначенные для статистического анализа в прикладных исследованиях.

При статистическом анализе массовых случайных явлений самой различной природы одной из актуальных и первоочередных проблем является оценка того, на сколько обоснованно может считаться случайной последовательность результатов наблюдений (либо отклонений от априорной функциональной зависимости). Под случайностью чаще всего понимается стационарность “в узком смысле” исследуемого процесса, то есть что полученная последовательность представляет собой независимые в совокупности реализации случайной величины (СВ) с неизменным законом распределения. На практике, опять таки чаще всего, исходят из того, что условия случайности соответствуют действительности. Вопросу оценки правдоподобия этого исходного положения, то есть гипотезы случайности посвящен достаточно основательно разработанный раздел математической статистики, объединяемый понятием критерии случайности [1,2]. В качестве статистики критериев случайности используются целочисленные СВ: длина серий, число серий, число инверсий и число циклов случайной последовательности. Наиболее полно исследованы законы распределения двух последних из перечисленных статистик. В частности, установлена их достаточно быстрая нормализация (при длине последовательности  $n > 20$ ). Несомненным достоинством этих статистик является их универсальность по отношению к закону распре-

деления исходной непрерывной СВ, то есть рассматриваются не сами выборочные значения, а их порядковые номера в вариационном ряду. Вместе с тем, как и все статистики кумулятивно-суммирующего типа, они позволяют лишь усреднено оценить характер всей последовательности в целом и могут быть недостаточно чувствительны к ее локальному упорядочению. Для выявления локальных отклонений, как известно, более пригодны критерии, построенные на статистиках селективно-порядкового типа.

Длина максимальной серии, хотя и не в полной мере, все же позволяет отследить оба указанных типа отклонения от случайности. А именно, наличие одной или более аномально длинных серий может свидетельствовать об искусственной группировке по месту их расположения в последовательности. Слишком короткие серии, в свою очередь, указывают на искусственное перемешивание всей последовательности в целом.

Относительно закономерностей формирования серий можно сказать следующее. В [3] было показано, что в последовательной выборке непрерывной СВ существуют два принципиально различающихся типа серий. По предложенной там же классификации “знаковые” – группа последовательных точек, расположенных по одну сторону от центральной линии (медианы) и “трендовые” – группа точек, образующих монотонную последовательность.

Серии первого типа сводятся к последовательным испытаниям по схеме Бернулли и подробно изучены в теории рекуррентных случайных событий [2]. Однако основное внимание в исследованиях было уделено законам распределения числа серий и времени возвращения серии фиксированной длины. Для вероятностей длинных серий в литературе можно найти лишь асимптотические Пуассоновские оценки [1, 2].

Задача же об установлении закона распределения длины максимальной серии, несмотря на свою очевидную “хрестоматийность”, осталась несколько в стороне от “магистрального” направления развития теории. Более того, информация о сериях второго типа по существу ограничена лишь указанием о неприменимости в данном случае схемы Бернулли, хотя такие серии представляют ни чуть не меньший интерес. Такое положение выглядит довольно странным, тем более, что задача допускает точное решение вполне элементарными методами.

Рассмотрим закономерности формирования знаковых серий. Как известно, последовательность  $n$  испытаний по схеме Бернулли порождает пространство элементарных событий, содержащее  $2^n$  различных равновероятных комбинаций.

Введем в рассмотрение целочисленные С. В.:  $L_n$  – длину максимальной серии успехов (либо неудач) в последовательности испытаний длиной  $n$  и  $T_l$  – длину последовательности, оканчивающейся первой серией длиной  $l$ . Следуя [2], рассмотрим следующие последовательности вероятностей, связанных с  $L_n$  и  $T_l$ :

вероятность “первого вхождения” серии длиной  $l$  на  $n$ -ом шаге  $f_n^{(l)} = P\{L_n = l; L_{n-1} = l-1\} = P\{T_l = n\}, l \geq 2, n \geq l$

вероятность отсутствия серии длиной  $l$  в последовательности длиной  $n$

$$q_n^{(l)} = P\{L_n < l\} = P\{T_l > n\} = \sum_{k=n+1}^{\infty} f_k^{(l)}$$

В терминах теории рекуррентных собы-

тий  $f_n^{(l)}$  является рядом распределения времени возвращения  $T_l$ , а вероятности  $q_n^{(l)}$  принято называть хвостами распределения времени возвращения.

Ряд распределения  $L_n$  связан с хвостами распределения  $T_l$  очевидным соотношением:

$$P\{L_n = l\} = v_n^{(l)} = q_n^{(l+1)} - q_n^{(l)}. \quad (1)$$

Как показано в теории рекуррентных событий [2], для вероятностей  $q_n^{(l)}$  серии успехов длиной  $l$  справедливо следующее рекуррентное соотношение:

$$\begin{cases} q_n^{(l)} = (1-p) \sum_{k=0}^{l-1} p^k q_{n-k-1}^{(l)}, \\ q_n^{(l)} = 1, n = 0, l-1, \end{cases}, \quad (2)$$

где  $p$  – вероятность успеха в одном испытании.

Соотношение (2) получается путем разложения события, соответствующего вероятности  $q_n^{(l)}$  на полную группу несовместных событий по допустимым значениям длины “финишной” серии успехов:  $L_n^* = \{0, 1, \dots, l-1\}$ .

В качестве иллюстрации рассмотрим вероятности появления трех шестерок подряд при бросании симметричной игральной кости (табл. 1).

$$p = \frac{1}{6}, l = 3; q_0 = q_1 = q_2 = 1;$$

$$q_n = \frac{5}{6}(q_{n-1} + \frac{1}{6}q_{n-1} + \frac{1}{36}q_{n-2})$$

Там же в [2], в разделе “более общие рекуррентные события”, рассматриваются события заключающиеся в том, что последовательность содержит либо серию успехов длиной  $l_1$ , либо серию неудач длиной  $l_0$ , либо обе серии одновременно. Элементарный способ непосредственного вычисления вероятностей  $q_n$  таких событий заключается в следующем. Разобьем полную группу событий, соответствующих  $q_n^{(l)}$ , на две подгруппы: {... последовательность оканчивается неудачей}, {... последовательность оканчивается успехом}.

Таблица 1. Ряд распределения  $f_n$  и хвосты  $q_n$  времени возвращения серии успехов  $T_3$

$n$	$q_n$	$f_n$
1	1	0
2	1	0
3	0.995	0.005
4	0.992	0.003
5	0.988	0.004
6	0.984	0.004
7	0.980	0.004
8	0.976	0.004
9	0.972	0.004
10	0.968	0.004
...	...	...

Тогда вероятность  $q_n$  можно представить в виде суммы:  $q_n = q_{0,n}^{(l_0)} + q_{1,n}^{(l_1)}$ , а для слагаемых будет справедлива система двух рекуррентных соотношений:

$$\begin{cases} q_{1,n}^{(l_1)} = \sum_{k=1}^{l_1-1} p^k q_{0,n-k}^{(l_0)}, n \geq 1 \\ q_{0,n}^{(l_0)} = \sum_{k=1}^{l_0-1} (1-p)^k q_{1,n-k}^{(l_1)}, n \geq 1, \end{cases} \quad (3)$$

$$q_{1,n}^{(l_1)} = q_{0,n}^{(l_0)} = 0, n < 0; \quad q_{1,0}^{(l_1)} = q_{0,0}^{(l_0)} = 1.$$

При условиях предыдущего примера и длине серии неудач  $l_0 = 5$  получим результат, представленный в табл. 2.

$$p = \frac{1}{6}, l_1 = 3, l_0 = 5$$

$$q_{1,0} = q_{0,0} = 1, q_{1,n} = q_{0,n} = 0, n < 0$$

$$\begin{cases} q_{1,n} = \frac{1}{6}(q_{0,n-1} + \frac{1}{6}q_{0,n-2}) \\ q_{0,n} = \frac{5}{6}(q_{1,n-1} + \frac{5}{6}(q_{1,n-2} + \frac{5}{6}(q_{1,n-3} + \frac{5}{6}q_{1,n-4}))) \end{cases}$$

Аналогичным образом можно получить рекуррентные формулы для вероятностей  $q_n$  любой конечной комбинации успехов и неудач. В частности, например такой, как “два из трех последовательных значений выборки из нормальной совокупности лежат вне центральной  $4\sigma$ -ой полосы” и т.д.

В случае, наиболее интересном с практической точки зрения, при  $p = \frac{1}{2}, l_1 = l_0 = l$

Таблица 2. Ряд распределения  $f_n$  и хвосты  $q_n$  времени возвращения, комбинированной серии успехов/неудач  $T_{3,5}$

$n$	$q_{1,n}$	$q_{0,n}$	$q_n$	$f_n$
1	0.167	0.833	1	0
2	0.167	0.833	1	0
3	0.162	0.833	0.995	0.005
4	0.162	0.830	0.992	0.003
5	0.161	0.425	0.586	0.406
6	0.094	0.421	0.515	0.071
7	0.082	0.362	0.444	0.071
8	0.072	0.305	0.377	0.067
9	0.061	0.249	0.310	0.067
10	0.050	0.194	0.244	0.066
...	...	...	...	...

очевидно, что  $q_{1,n}^{(l)} = q_{0,n}^{(l)} = \frac{1}{2} q_n^{(l)}$ , и для  $q_n^{(l)}$  получается соотношение приведенное в [3]:

$$q_n^{(l)} = \sum_{k=1}^{l-1} \left(\frac{1}{2}\right)^k q_{n-k}^{(l)}. \quad (4)$$

Такое свойство хвостов распределения времени возвращения  $q_n^{(l)}$  обусловлено тем, что события, удовлетворяющие условию  $\{L_n^* < l\}$  образуют эргодическое множество состояний. Событие  $\{L_n^* = l\}$ , соответственно, является поглощающим состоянием, и процесс формирования последовательности можно представить в виде ориентированного графа. Например, для комбинированной серии длиной  $l=4$  граф имеет вид, показанный на рис. 1. Вероятности перехода, изображенные на графе стрелками, стационарны и равны  $p$ , если последовательность оканчивается “1” (успех), либо  $1-p$ , если последовательность оканчивается “0” (неудача).

Соотношение (3) с учетом (1) позволяет получить исчерпывающую информацию как о  $T_l$ , и о  $L_n$ . Как показывают исследования с помощью производящих функций, числовые характеристики времени возвращения вычисляются наиболее просто. В частности, среднее время возвращения серии успехов

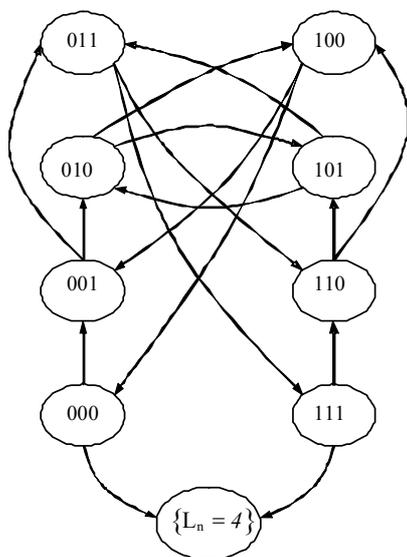


Рис. 1. Граф вероятностной цепи формирования знаковой серии  $l=4$  (трендовой серии  $l=5$ )

длиной  $l$  при  $p = \frac{1}{2}$  составляет

$$M[T_l] = \mu_l = 2 \cdot (2^l - 1).$$

Среднее время возвращения комбинированной серии при тех же условиях составляет  $\mu_l = 2^l - 1$ .

В рассматриваемой постановке задачи искомыми являются средние значения СВ  $L_n$ , связанные с вероятностями  $q_n^{(l)}$  соотношением:

$$M[L_n] = n - \sum_{l=2}^n q_n^{(l)}. \quad (5)$$

Вид точной зависимости  $M[L_n]$  для частного случая (4) в сравнении с результатами имитационного статистического эксперимента представлен на рис. 2. Экспериментальные точки на графиках рис. 2, 3 представляют собой средние по 200 реализациям значения  $L_n$  в последовательной выборке  $X_i, i = 1, n$  из совокупности  $N(0;1)$ , а также в последовательностях  $Y_i = X_i - \frac{1}{n} \sum_{i=1}^n X_i, i = 1, n$

и  $Z_i = X_{i+1} - X_i, i = 1, n-1$ . Дисперсия при  $n \approx 50$  практически стабилизируется на постоянном уровне (рис. 4). Общий вывод заключается в том, что с ростом  $n$  закон распределения  $L_n$  быстро выходит на установившийся режим, то есть среднее значение смещается вправо с логарифмической скоростью, а форма рассеяния и, соответственно, ширина интервала рассеивания сохраняются практически неизменными. Границы 90% доверительных интервалов  $L_n$  для  $n = 2 \div 1000$  приведены в таблице 3 (в скобках даны границы 95% доверительных интервалов). Фактически, поскольку СВ  $L_n$  целочисленная, речь идет не об интервале в привычном смысле, а о конечном множестве “практически реализуемых” значений. Мощность этого множества практически не зависит от  $n$  и составляет 7ч8 и 3ч4 для серий первого и второго типа соответственно. На-

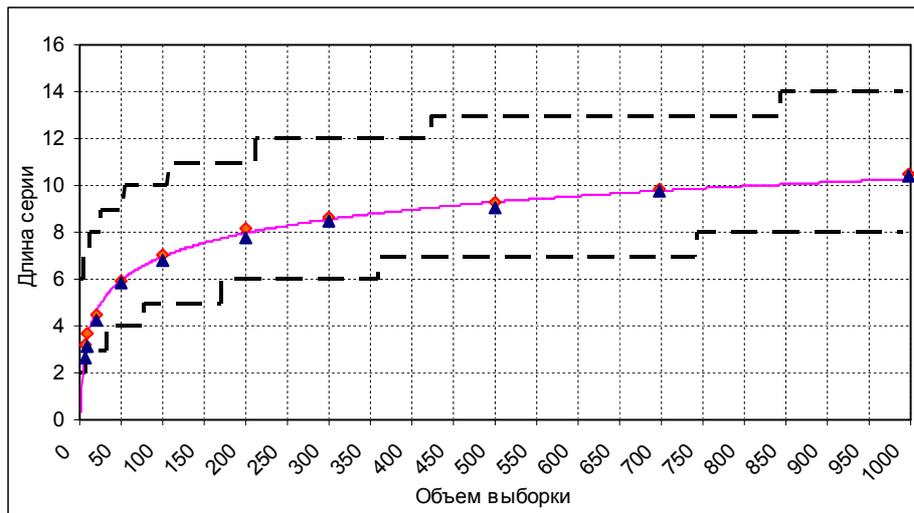


Рис. 2. Среднее значение и границы доверительного ( $\gamma \geq 0,9$ ) интервала длины максимальной “знаковой” серии:

- экспериментальные значения в исходной выборке;
- ▲ - экспериментальные значения в центрированной выборке

пример, требуется найти 90% доверительный интервал для  $L_n$  при  $n = 40$ . По таблице находим  $4 \leq L_n \leq 9$ . Для  $\gamma = 95\%$  при тех же значениях  $n$  интервал составит уже  $3 \leq L_n \leq 10$ . При значениях  $n > 1000$  для определения границ доверительных интервалов можно пользоваться простым соотношением

$$\log_2 \frac{5n}{4} - 2,5 \leq l \leq \log_2 \frac{5n}{4} + 4,5.$$

Таким образом для серий в случайной последовательности, к которой применима схема Бернулли, можно считать полностью установленными законы распределения как числа серий [3], так и длины максимальной серии.

Однако существует необходимость теоретического уточнения, обусловленная тем, что рассматриваемая чаще всего на практике

последовательность  $Y_i = X_i - \frac{1}{n} \sum_{i=1}^n X_i$  не является независимой [3]. Значения  $Y_i$  связа-

ны очевидным соотношением  $\sum_{i=1}^n Y_i = 0$ . Поэтому число степеней свободы в ней составляет  $n - 1$ . Такое же значение имеет ранг ковариационной матрицы  $\bar{Y}$ . При этом лю-

бой из миноров порядка  $n - 1$  представляет собой блочную матрицу:

$$K_n = \sigma^2 \left\{ \begin{array}{c} \left. \begin{array}{ccc} \frac{n-1}{n} & & \\ & \ddots & -\frac{1}{n} \\ & & \frac{n-1}{n} \end{array} \right\} \right\}^{n-1}, \\ \left. \begin{array}{ccc} & & \\ & -\frac{1}{n} & \\ & & \ddots \\ & & & \frac{n-1}{n} \end{array} \right\}^{n-1} \end{array} \right\} n > 2.$$

В дальнейшем, без ограничения общности, дисперсию можно положить равной 1.

Определитель такой матрицы равен

$|K_n| = \frac{1}{n}$ . Обратная матрица имеет аналогичный вид. На главной диагонали стоят “2”, на остальных местах – “1”.

Например, для  $n = 5$  матрицы выглядят следующим образом:

$$K_5 = \begin{pmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix},$$

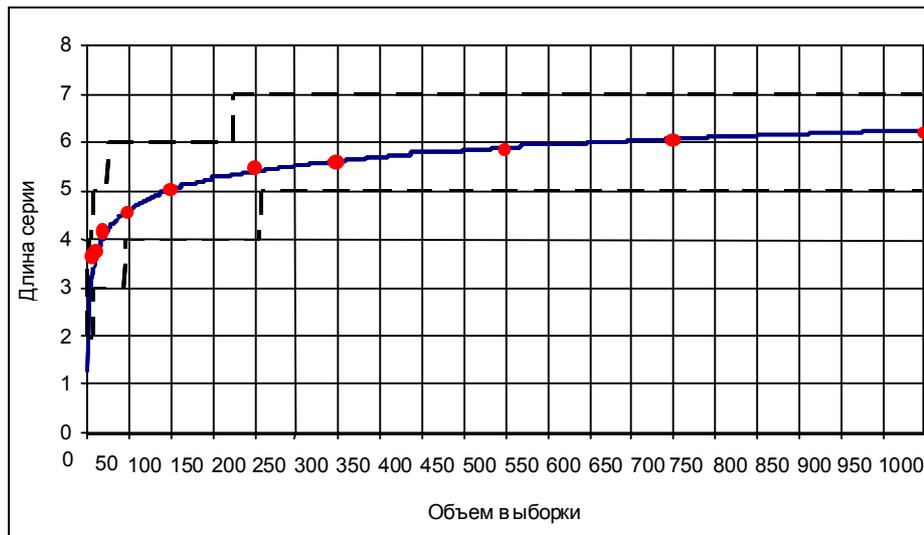


Рис. 3. Среднее значение и границы доверительного интервала ( $\gamma \geq 0,9$ ) длины максимальной “трендовой” серии:  
● - результаты эксперимента

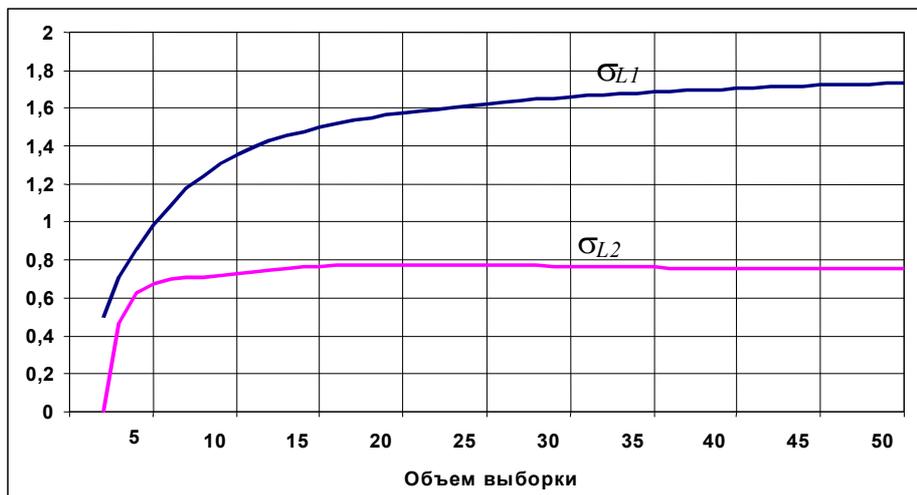


Рис. 4. Стандартное отклонение длины максимальной серии в зависимости от объема выборки.

$$K_5^{-1} = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}.$$

Таким образом, плотность совместного распределения подгруппы любых  $n - 1$  из  $n$  величин  $Y_i$  при нормальном исходном распределении составит:

$$f(y_1, y_2, \dots, y_{n-1}) = \frac{\sqrt{n}}{(2\pi)^{\frac{n-1}{2}}} \exp \left[ - \left[ \sum_{i=1}^{n-1} y_i^2 + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} y_i y_j \right] \right],$$

$n \geq 3.$  (6)

Для сравнения, ковариационная матрица последовательности  $X$  единичная  $n \times n$ , а плотность совместного распределения равна:

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp \left[ - \frac{1}{2} \sum_{i=1}^n x_i^2 \right],$$

$n \geq 1.$

Принципиальное отличие последовательности  $Y$  заключается в том, что размерность пространства элементарных событий составляет не  $2^n$ , а  $2^n - 2$ , и имеет место отрицательная корреляция [3]. При этом среднее и дисперсия статистики  $L_n$  в последовательности  $Y$  имеют несколько меньшие зна-

чения. Однако доверительные интервалы  $L_n$  уже не будут сильно отличаться от аналогичных интервалов для последовательности  $X$ , поскольку количественное различие между законами распределения быстро сходит на нет с ростом  $n$ . Например, для любых двух значений  $Y$  при объеме выборки  $n$  справедливы следующие соотношения:

$$P\{Y_i > 0, Y_j > 0\} = \frac{1}{4} - \frac{1}{2\pi} \operatorname{arctg} \frac{1}{\sqrt{n(n-2)}}.$$

$$P\{Y_i > 0, Y_j < 0\} = \frac{1}{4} + \frac{1}{2\pi} \operatorname{arctg} \frac{1}{\sqrt{n(n-2)}}.$$

Следовательно, при  $n \gg l$  вероятности в последовательности  $Y$  отличаются от аналогичных в последовательности  $X$  на величину порядка  $\sim \frac{1}{n}$ .

Средние по 200 реализациям значения  $\overline{L}_n$  в последовательности  $Y$  показаны маркерами  $\blacktriangle$  на рис. 2.

При исследовании закономерности образования “трендовых” серий используем тот факт, что соотношение порядка между независимыми выборочными значениями непрерывной С.В. инвариантно к закону ее распределения.

Как известно, в общем случае закон распределения непрерывной СВ  $F_X(x) = P\{X < x\}$  представляет собой монотонно возрастающее (монотонно не убывающее) отображение некоторого подмножества числового интервала  $(-?; +?)$  на отрезок  $[0; 1]$ . При этом очевидно, что из всего многообразия законов распределения существует единственный –  $R[0; 1]$ , который отображает выше указанный отрезок сам на себя. Именно на этом факте построены критерии согласия А.Н. Колмогорова и критерий принадлежности двух выборок одной генеральной совокупности Н.В. Смирнова.

В силу своей универсальности (преобразование  $Y_i = X_i - \frac{1}{n} \sum_{i=1}^n X_i$  так же никак не влия-

ет на соотношение порядка), закономерность образования трендовых серий является, в некотором смысле, одним из фундаментальных количественных соотношений в материальном мире и, поэтому актуальна для всех без исключения массовых случайных явлений.

Из этих рассуждений общего характера, тем не менее, можно вполне обоснованно сделать конструктивное и весьма существенное заключение, что не ограничивая общность результатов и выводов, достаточно рассмотреть закономерность “трендовых” серий в случайной выборке из совокупности  $R[0; 1]$ , сведя тем самым к возможному минимуму объем вычислений (при любом объеме выборки плотность совместного распределения  $\equiv 1$ ).

Используя принятую в [3] мнемонику, а именно, обозначая “0” отрицательные значения последовательных разностей, “1” – положительные, вероятность любого отрезка последовательности теперь однозначно определится следующим образом:

$$P\{\underbrace{001\dots 01}_{n-1}\} = \int_0^1 dx_n \int_0^{x_n} dx_{n-1} \int_{x_{n-1}}^1 dx_{n-2} \dots \int_0^{x_4} dx_3 \int_{x_3}^1 dx_2 \int_{x_2}^1 dx_1.$$

Например, сравнивая с вычисленными в [3]  $P\{00\} = \frac{1}{6}$ ,  $P\{01\} = \frac{1}{3}$ , для выборки из нормальной совокупности убеждаемся в их совпадении с выборкой из  $R[0; 1]$ :

$$P\{00\} = \int_0^1 dx_3 \int_{x_3}^1 dx_2 \int_{x_2}^1 dx_1 = \frac{1}{6},$$

$$P\{01\} = \int_0^1 dx_3 \int_0^{x_3} dx_2 \int_{x_2}^1 dx_1 = \frac{1}{3}.$$

Рассмотрим более детально структуру этих вероятностей. Каждая из них может быть представлена в виде:

$$P = \int_0^1 \varphi(x) dx, \quad (7)$$

где  $\varphi(x)$  – степенной многочлен, порядок которого на 1 меньше длины последовательности. В свою очередь, каждый такой много-

член однозначно определяется набором (вектором) своих коэффициентов.

Используя уже рассмотренную схему рассуждений, разложим вероятности  $q_n^{(l)}$  трендовых серий и, следовательно, соответствующие многочлены, на сумму двух слагаемых. Для наиболее интересного симметричного случая  $l_0 = l_1$  система рекуррентных уравнений будет иметь следующий вид:

$$\begin{cases} \varphi_{1,n}^{(l)}(x) = \int_0^x \varphi_{0,n-1}^{(l)}(x)dx + \int_0^x \int_0^x \varphi_{0,n-2}^{(l)}(x)dx + \dots + \underbrace{\int_0^x \dots \int_0^x \varphi_{0,n-l+2}^{(l)}(x)dx}_{l-2} \\ \varphi_{0,n}^{(l)}(x) = \int_x^1 \varphi_{1,n-1}^{(l)}(x)dx + \int_x^1 \int_x^1 \varphi_{1,n-2}^{(l)}(x)dx + \dots + \underbrace{\int_x^1 \dots \int_x^1 \varphi_{1,n-l+2}^{(l)}(x)dx}_{l-2} \end{cases} \quad (8)$$

$$\varphi_{0,1}^{(l)}(x) = \varphi_{1,1}^{(l)}(x) \equiv 1.$$

$$\varphi_n^{(l)}(x) = \int_0^x \varphi_{n-1}^{(l)}(x)dx + \dots + \underbrace{\int_0^x \dots \int_0^x \varphi_{n-l+2}^{(l)}(x)dx}_{l-2},$$

Из условия симметрии с очевидностью следует, что  $\varphi_{0,n}^{(l)}(x) = \varphi_{1,n}^{(l)}(1-x)$ . Таким образом, система (8) фактически сводится к одному уравнению:

$$\varphi_2^{(l)}(x) = \dots \varphi_{l-1}^{(l)}(x) = x, \quad (9)$$

где обозначено  $\varphi_n^{(l)}(x) = \varphi_{1,n}^{(l)}(x)$ ,  $\varphi_n^{(l)}(x) = \varphi_n^{(l)}(1-x)$ .

Порядок рекуррентного многочлена  $\varphi_n^{(l)}(x)$  равен  $n-1$ , а его коэффициенты имеют вид правильных простых дробей.

Например, при  $l=3$  рекуррентное уравнение имеет первый порядок:

$$\begin{cases} \varphi_n(x) = \int_0^x \varphi_{n-1}(x)dx, n \geq 3 \\ \varphi_2(x) = x. \end{cases} \quad (10)$$

Откуда получаем:

$$\varphi_3(x) = x - \frac{1}{2}x^2; \quad \varphi_4(x) = \frac{1}{2}x - \frac{1}{6}x^3;$$

$$\varphi_5(x) = \frac{1}{3}x - \frac{1}{6}x^3 + \frac{1}{24}x^4;$$

$$\varphi_6(x) = \frac{5}{24}x - \frac{1}{12}x^3 + \frac{1}{120}x^5;$$

$$\varphi_7(x) = \frac{16}{120}x - \frac{1}{18}x^3 + \frac{1}{120}x^5 - \frac{1}{720}x^6; \dots$$

Соответствующий ряд вероятностей, определяемых как  $q_n^{(3)} = 2 \int_0^1 \varphi_n^{(3)}(x)dx$ , составит:

$$\frac{2}{3}; \frac{5}{12}; \frac{16}{60}; \frac{61}{360}; \frac{272}{2520}; \dots$$

При  $l=5$  порядок рекуррентного уравнения составляет  $l-2=3$ :

$$\begin{cases} \varphi_n(x) = \int_0^x \varphi_{n-1}(x)dx + \int_0^x \int_0^x \varphi_{n-2}(x)dx + \int_0^x \int_0^x \int_0^x \varphi_{n-3}(x)dx, n \geq 5 \\ \varphi_2(x) = \varphi_3(x) = \varphi_4(x) = x \end{cases} \quad (11)$$

Последовательность многочленов и вероятностей  $q_n^{(5)}$  составляет:

$$\varphi_5 = x - \frac{1}{24}x^4;$$

$$\varphi_6 = \frac{23}{24}x + \frac{1}{12}x^2 - \frac{1}{12}x^3 - \frac{1}{120}x^5; \dots$$

$$\varphi_7 = \frac{11}{12}x + \frac{1}{16}x^2 - \frac{1}{18}x^3 - \frac{1}{48}x^4$$

$$q_n^{(5)} = \frac{59}{60}; \frac{349}{360}; \frac{2408}{2520}; \dots$$

Общий вид формул для рекуррентного многочлена, рекуррентных соотношений для его коэффициентов и вероятностей  $q_n^{(l)}$  выглядит следующим образом:

$$\begin{cases} \varphi_n^{(l)}(x) = \sum_{k=1}^{n-1} a_k^{(l)}(n)x^k, n \geq l \\ a_k^{(l)}(n) = \sum_{i=1}^{\min(k,l-2)} \frac{(-1)^{k-i}(k-i)!}{k!} \sum_{m=k-i}^{n-i-1} C_m^{k-i} a_m^{(l)}(n-i), n \geq l, k \leq n-1 \\ q_n^{(l)} = 2 \sum_{k=1}^{n-1} \frac{a_k^{(l)}(n)}{k+1} \end{cases} \quad (12)$$

Точные значения  $M[L_n]$  и  $\sigma_{L_n}$  трендовых серий в сравнении с результатами статистического эксперимента приведены на рисунках 3, 4. Границы доверительных интервалов для  $L_n$  трендовых серий указаны в табл. 4. Устройство и порядок работы с ней

Таблица 3

Объем выборки, $n$	$M[L_2n]$	$\sigma_{L_2}$	НГ	ВГ
2	2,00	0,00	2	2
3	2,33	0,47	2	3
4	2,67	0,62	2	4
5	2,90	0,68	2	4
6	3,08	0,70	2	4(5)
7	3,22	0,71	2	4(5)
8	3,33	0,71	2	5
9	3,42	0,72	3(2)	5
10	3,50	0,73	3(2)	5
12	3,63	0,75	3	5
14	3,74	0,76	3	5
16	3,83	0,77	3	5
18	3,92	0,78	3	5(6)
20	3,99	0,78	3	5(6)
25	4,14	0,77	3	5(6)
30	4,27	0,77	3	6
40	4,45	0,76	3	6
50	4,58	0,76	4	6
100	4,99	0,76	4	6(7)
200	5,39	0,73	4	7
300	5,61	0,72	5	7
500	5,88	0,73	5	7
700	6,06	0,71	5	7(8)
1000	6,25	0,69	5	7(8)

Таблица 4

Объем выборки, $n$	$M[L_1n]$	$\sigma_{L_1}$	НГ	ВГ
2	1,50	0,50	1	2
3	2,00	0,71	1	3
4	2,38	0,86	1	4
5	2,69	0,98	1	5
6	2,94	1,09	2(1)	5(6)
7	3,16	1,18	2	5(6)
8	3,34	1,25	2	6
9	3,51	1,30	2	6(7)
10	3,66	1,35	2	6(7)
12	3,92	1,43	2	7
14	4,15	1,48	2	7(8)
16	4,34	1,52	3(2)	7(8)
18	4,51	1,55	3(2)	7(8)
20	4,66	1,58	3	8
25	4,98	1,63	3	8(9)
30	5,24	1,66	3	8(9)
40	5,66	1,70	4(3)	9(10)
50	5,98	1,73	4	9(10)
100	6,98	1,79	5(4)	10(11)
200	7,98	1,83	6(5)	11(12)
300	8,56	1,84	6	12(13)
500	9,30	1,85	7	13(14)
700	9,78	1,86	7	13(14)
1000	10,30	1,86	8	14(15)

полностью аналогичны табл. 3. Например, при объеме последовательной выборки количественных данных (результатов измерений)  $n = 100$  максимальная длина монотонно убывающей или возрастающей последовательности (трендовой серии), согласующаяся с гипотезой об отсутствии закономерно обусловленной изменчивости, должна лежать в границах  $4 \leq L_{100} \leq 6$  при уровне значимости  $1 - \gamma = 10\%$ . При уровне значимости  $1 - \gamma = 5\%$  интервал составляет  $4 \leq L_{100} \leq 7$ .

Таким образом, закономерность формирования трендовых серий имеет качественно сходный характер со знаковыми, а именно, среднее значение длины максимальной серии с увеличением  $n$  неограниченно возрастает. Аналогичным образом ведут себя и границы интервала рассеивания, причем, ширина интервала остается практически постоянной. Различия заключаются в том, что среднее значение с увеличением  $n$  растет не как двоичный логарифм, а существенно мед-

леннее – как обратная гамма-функция:  $n \approx \Gamma(M[L_n] + 1)$ . При этом интервал ее рассеивания примерно вдвое уже, чем знаковой:  $\Gamma^{-1}(n) - 2 \leq l \leq \Gamma^{-1}(n) + 1$ .

Полученные асимптотические оценки средних значений длины максимальной серии являются эмпирическими. Однако они дают поразительно высокую точность, по крайней мере, в исследованном диапазоне  $n \leq 10^5$ . Последнее обстоятельство, конечно же, не лишает актуальности задачу их строгого аналитического исследования, также как и числовых характеристик более высокого порядка. Тем не менее, как представляется, есть основания полагать, что установленные закономерности, главным образом механизм формирования последовательности в немарковской цепи (с зависимыми вероятностями перехода), являются нетривиальным теоретико-вероятностным результатом и могут найти интересные приложения в смежных с теорией вероятностей областях.

Кроме того, учитывая разработку критерия длины максимальной серии до простой в обращении типовой методики, представляется возможным и целесообразным, чтобы этот критерий был включен в инструментарий специалистов, занимающихся статистическим анализом в прикладных исследованиях.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Дунин-Барковский И.В., Смирнов Н.В.* Те-

ория вероятностей и математическая статистика в технике (общая часть). М.: ГИТТЛ, 1955.

2. *Феллер В.* Введение в теорию вероятностей и ее приложения т.1 (Дискретные распределения). М.: Мир, 1984.

3. *Юнак Г.Л., Годлевский В.Е., Плотников А.Н.* Об интерпретации серий на контрольных картах // *Методы менеджмента качества.* 2005. №4.

### **DISTRIBUTION LAW OF THE SIZE OF MAXIMAL SERIES AND ITS STATISTICAL APPLICATION**

© 2006 A.N. Plotnikov

Samara State Aerospace University

The article deals with the problem of maximal series size as the criterion of its randomness use appropriateness. Distribution laws of mentioned statistic for two main series types in random number sequence have been obtained and experimentally justified. Also were received tables of confidence intervals and the technique of interval estimate of maximal series size for statistical analysis in applied research.