УДК 519.254

ОБ ОДНОМ ПАРАДОКСЕ ЗАКОНА БОЛЬШИХ ЧИСЕЛ ДЛЯ МАКСИМАЛЬНЫХ СЕРИЙ В ПОСЛЕДОВАТЕЛЬНОЙ ВЫБОРКЕ

© 2009 А.Н. Плотников

Самарский государственный аэрокосмический университет

Поступила в редакцию 13.01.2009

Дополнены и уточнены ранее опубликованные результаты исследования закономерностей формирования серий в последовательной выборке. Получены предельные формы законов распределения длины максимальных серий, образуемых положением относительно медианы и отношением порядка между соседними индивидуальными значениями. Установлено, что предельная форма распределения длины максимальной серии отношений порядка ведет себя парадоксальным образом, циклически эволюционируя от вырожденного распределения с изолированной модой до бинарного с модой, симметрично расщепленной на два подряд стоящих значения.

Ключевые слова: последовательная выборка, длина максимальной серии, закон больших чисел, структура серий, критерии случайности.

1. Серии в последовательной выборке, представляющие собой группы подряд стоящих индивидуальных значений, расположенных по одну сторону от медианы (знаковые серии) или образующих монотонную последовательность (трендовые серии), как было показано в [1, 2], имеют отчетливую спектральную структуру с устойчивой воспроизводимостью. В [1] был получен точный закон распределения длины максимальной серии обоих типов. В [2] установлен закон распределения числа серий фиксированной длины и получены его нормальные (для коротких серий $l \le 7$) и Пуассоновские (для $l > 7$) асимптотики при большой длине последовательности. Последние из указанных результатов легко позволяют установить предельную форму закона распределения максимальной длины серий. При этом, как оказывается, предельная форма распределения максимальной трендовой серии обладает довольно неожиданным свойством, которое без преувеличения можно назвать "парадоксом вырождения". Этот парадокс заключается в том, что при неограниченном возрастании длины последовательности n длина максимальной трендовой серии перестает быть случайной, т.е вырождается в фиксированное значение: $P\{L_{n_1} = l_1(n_1)\} \approx 1$.

Причем, с дальнейшим возрастанием n эта изолированная мода расщепляется на две, и вторая мода $l_2 = l_1 + 1$, постепенно возрастая, сравнивается с первой:

$$P\{L_{n_2} = l_1(n_2)\} \approx P\{L_{n_2} = l_1(n_2) + 1\} \approx \frac{1}{2}.$$

Плотников Андрей Николаевич, кандидат технических наук, доцент кафедры производства летательных аппаратов и управления качеством в машиностроении. E-mail: metatlas@yandex.ru

Затем вторая мода поглощает первую (распределение опять вырождается), и это процесс закономерно циклически повторяется. При умеренно больших $n\left(\sim 10^3\right)$ этот эффект проявляет себя, как показано в [1], в виде апериодических колебаний дисперсии и моментов более высокого порядка.

Серии первого типа (знаковые) ведут себя не столь парадоксально. Их предельная форма рассеяния подобно выборочному размаху [3] является стационарной и смещается вправо по оси n со скоростью $\log_2 n$.

2. Асимптотику распределения длины максимальной серии L_n найдем, используя полученные в [2] Пуассоновские оценки для числа нерекуррентных серий фиксированной длины $\tilde{R}_l(n)$. Для знаковой серии длины l, расположенной выше медианы в последовательности длиной n эта оценка, согласно [2], составляет:

$$P\left\{\tilde{R}_l^+(n) = k\right\} = \exp\left(-k\frac{n}{2^{l+2}}\right).$$

Используя двойственность величин $\tilde{R}_l^+(n)$ и L_n^+, ряд распределения последней можно записать в виде:

$$p_l^+ = P\left\{L_n^+ = l\right\} = = P\left\{\tilde{R}_l^+(n) > 0\right\} P\left\{\tilde{R}_m^+(n) = 0, \forall m > l\right\}. \quad (1)$$

Подставив в (1) Пуассоновские вероятности, получим:

$$p_l^+ = \left[1 - \exp\left(-\frac{n}{2^{l+2}}\right)\right] \exp\left(-\sum_{m=l+1}^{\infty} \frac{n}{2^{m+2}}\right) = = \left[1 - \exp\left(-\frac{n}{2^{l+2}}\right)\right] \exp\left(-\frac{n}{2^{l+2}}\right) = = \exp\left(-\frac{n}{2^{l+2}}\right) - \exp\left(-\frac{n}{2^{l+1}}\right),\ 1 \le l \le n. \quad (2)$$

Асимптотический ряд (2) очевидно сходится к 1 как $\sum_{l=1}^{n} p_l^+ = \exp(-\frac{n}{2^{n+2}}) - \exp(-\frac{n}{4})$.

Для величины $L_n = \max\{L_n^+, L_n^-\}$, собственно и являющейся максимальной длиной знаковой серии, определённой согласно [1], ряд распределения найдём по формуле максимума двух независимых величин с одинаковым законом распределения:

$$p_l = 2p_l^+ \sum_{k<l} p_k^+ + {p_l^+}^2 . \qquad (3)$$

Подставляя в (3) члены ряда (2) и пренебрегая слагаемыми $\sim e^{-an}$, получим

$$p_l \approx \exp(-\frac{n}{2^{l+1}}) - \exp(-\frac{n}{2^l}) . \qquad (4)$$

Для рядов (2), (4) очевидно тождество $p_{l+1} = p_l^+$ или, что фактически то же самое, $p_l^+(n) = p_l(2n)$.

Таким образом, ряд (4) сдвинут на 1 вправо по оси l относительно ряда (2).

Переходя в (4) к величине

$$t = l - \log_2 n \qquad (5)$$

получаем независящую от n предельную форму (4):

$$f(t) = \exp(-\frac{1}{2^{t+1}}) - \exp(-\frac{1}{2^t}), -\infty < t < \infty , \qquad (6)$$

вид которой с последующим квантованием показан на рис.1. Числовые характеристики (6) равны:

$$\mu = \int_{-\infty}^{\infty} tf(t)dt = \frac{-1}{\ln^2(2)} \int_0^{\infty} \left(e^{-\frac{1}{2}x} - e^{-x} \right) \ln(x) \frac{dx}{x} = \\ = \frac{C_1}{\ln(2)} - \frac{1}{2} \approx 0.3327 , \qquad (7)$$

$$\sigma^2 = \int_{-\infty}^{\infty} t^2 f(t)dt - \mu^2 + \frac{1}{12} = \\ = \frac{C_2}{\ln^2(2)} - \frac{C_1}{\ln(2)} + \frac{1}{3} - \mu^2 + \frac{1}{12} \approx 3.5904 , \qquad (8)$$

где C_k, $k = 1,2$ - Эйлеровы интегралы [3, 4]: $C_k = (-1)^k \int_0^{\infty} e^{-x} \ln^k(x)dx$; $\frac{1}{12}$ в (8) – поправка на группировку (6). Таким образом, среднее значение длины максимальной знаковой серии составит

$$\mu_n \approx \frac{1}{3} + \log_2 n , \qquad (9)$$

дисперсия определяется согласно (8), а предельная форма рассеяния имеет вид (6).

3. По аналогичной схеме, используя Пуассоновские оценки для трендовых серий [2]:

$$P\{\tilde{R}_l^+(n) = k\} = \exp(-k(n-1)\frac{l^2 + l - 1}{(l+2)!}),$$

получаем:

$$P\{L_n^+ = l\} = p_l^+ = [1 - \exp(-(n-1)\frac{l^2 + l - 1}{(l+2)!})] \times \\ \times \exp[-(n-1) \sum_{m=l+1}^{\infty} \frac{m^2 + m - 1}{(m+2)!}] = \\ = \exp(-(n-1)\frac{l+1}{(l+2)!}) - \exp(-(n-1)\frac{l}{(l+1)!}), \qquad (10)$$

$2 \le l \le n.$

Сумма ряда (10) составит:

$$\sum_{l=2}^{n} p_l^+ = \exp(-\frac{n^2 - 1}{(n+2)!}) - \exp(-\frac{n-1}{3}) . \qquad (11)$$

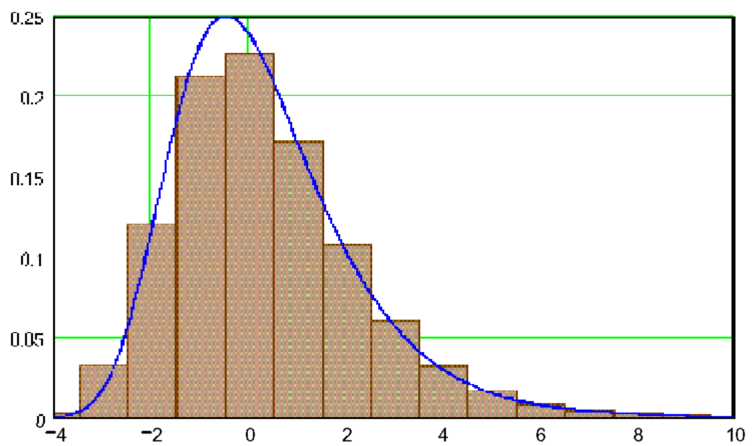

Рис. 1. Предельное распределение длины максимальной “знаковой” серии и его сглаживающая кривая

По аналогии с (3), (4), для $L_n = \max\{L_n^+, L_n^-\}$ получаем

$$P\{L_n = l\} = p_l = \exp(-2(n-1)\frac{l+1}{(l+2)!}) - \\ -\exp(-2(n-1)\frac{l}{(l+1)!}),\ 2 \le l \le n. \quad (12)$$

В дальнейшем, поскольку речь идет о больших n, различием между n и $n-1$ в (10)-(12) пренебрежем и положим $l > 10,\ 2n = \alpha l!$, где α – константа ~1.

Для соседних членов ряда (12) с номерами $l-1$ и l будет иметь:

$$p_{l-1} = \exp(-\alpha\frac{l}{l+1}) - \exp(-\alpha(l-1)) \approx e^{-\alpha}, \\ p_l = \exp(-\frac{1}{l+2}) - \exp(-\alpha\frac{l}{l+1}) \approx 1 - e^{-\alpha}. \quad (13)$$

Приравняв $p_{l-1} = p_l = \frac{1}{2}$, получим

$$\alpha = \ln 2. \quad (14)$$

Таким образом, последовательность

$$n_2(l) = \left[\frac{l!\cdot \ln 2}{2}\right] \quad (15)$$

соответствует наличию симметрично расщепленной двойной изолированной моды:

$$P\{L_{n_2} = l-1\} \approx P\{L_{n_2} = l\} \approx \frac{1}{2}$$

(остальные члены ряда (12) в сумме составляют ничтожно малую вероятность, и распределение становится "почти вырожденным").

Последовательность значений $n_1(l)$, соответствующих полному вырождению (изолированной моде) будем искать в виде:

$$2n_1(l) = l!\beta(l), \quad (16)$$

где $\beta(l)$ определим из условия $p_{l-1} \to \max$. Для члена ряда (12) с номером $l-1$ будем иметь:

$$p_{l-1} = \exp(-\beta\frac{l}{l+1}) - \exp(-\beta(l-1)) \approx e^{-\beta} - e^{-l\beta}. \quad (17)$$

Подстановкой $e^{-\beta} = u$ приходим к функции вида

$$u - u^l, \quad (18)$$

максимум которой достигается в точке

$$u = \left(\frac{1}{l}\right)^{\frac{1}{l-1}}, \quad (19)$$

откуда, после обратной подстановки, находим

$$\beta(l) \approx \frac{\ln l}{l-1} \approx \frac{\ln l}{l}. \quad (20)$$

Таким образом, значения

$$n_1(l) = [\frac{(l-1)!\ln l}{2}] \quad (21)$$

соответствуют фазе существования изолированной моды (полного вырождения закона распределения): $P\{L_{n_1} = l-1\} \approx 1$.

Область значений $n_1(l) < n < n_2(l)$ соответствует периоду "зарождения" и роста второй моды $\{L_n = l\}$. Затем, на интервале $n_2(l) < n < n_1(l+1)$ новая мода "поглощает" старую, и процесс циклически повторяется. Эво-

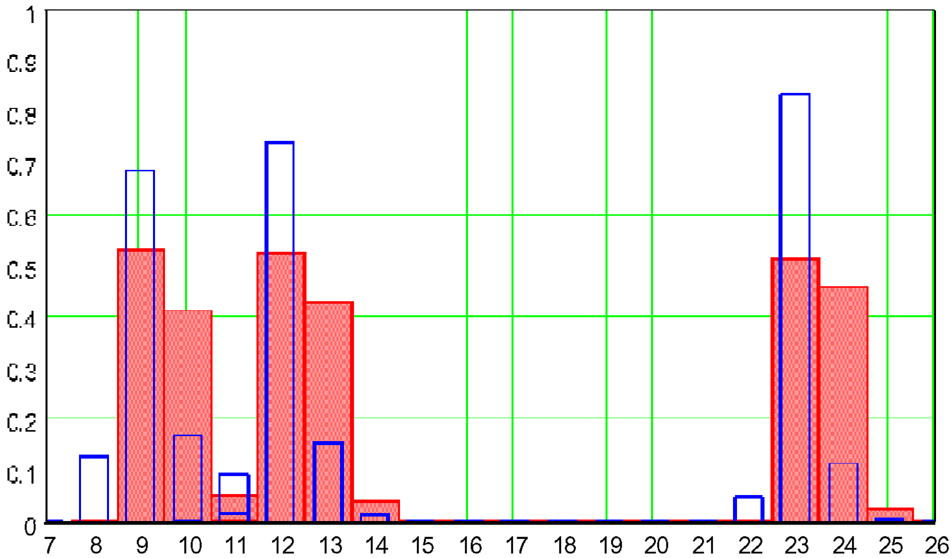

Рис. 2. Эволюция распределения длины максимальной трендовой серии. Фазы изолированной и симметрично расщепленной моды при n=4.178×10^5; 1.258×10^6; 6.143×10^8; 2.158×10^9; 4.108×10^{22}; 2.150×10^{23} (слева направо)

люция ряда (12) в зависимости от n показана на рис. 2. При этом фазы полного вырождения рядов (10), (12) совпадают между собой, и (3) преобразуется к виду $p_l = p_l^{+2} = 1$.

В фазе расщепления моды (12) соотношение (3) примет вид:

$$\begin{cases} p_{l-1} = p_{l-1}^{+\;2} \\ p_l = 2p_l^+ p_{l-1}^+ + p_l^{+2} \end{cases} . \qquad (22)$$

При этом в точке симметрии $p_{l-1}(n_2) = p_l(n_2) = \frac{1}{2}$ мода L_n^+ расщепляется в пропорции $p_{l-1}^+ = \frac{1}{\sqrt{2}}$, $p_l^+ = 1 - \frac{1}{\sqrt{2}}$. С возрастанием длины последовательности до $n_2^+ = 2n_2$ мода L_n^+ становится симметричной, а значения L_n перенормируются как $p_{l-1} = \frac{1}{4}$, $p_l = \frac{3}{4}$.

Таким образом, длина максимальной трендовой серии в большой последовательной выборке становится величиной, закономерно возрастающей с увеличением объема выборки (неслучайной). Тем самым, она не только является надежным критерием случайности выборки, но и, как представляется, сможет послужить ключом к объяснению некоторых, еще не до конца понятых закономерностей в реальных стохастических процессах, в частности, в процессах диффузии и им подобных.

СПИСОК ЛИТЕРАТУРЫ

1. *Плотников А.Н.* Закон распределения длины максимальной серии и его статистические приложения // Известия Самарского научного центра РАН. 2006. Т. 8. №4. С. 1047-1056.
2. *Плотников А.Н.* Об инвариантах структуры серий и критериях случайности последовательной выборки // Известия Самарского научного центра РАН. 2006. Т.8. №4. С.1142-1147.
3. *Крамер Г.* Математические методы статистики. М.: Мир, 1976.
4. *Корн Г., Корн Т.* Справочник по математике.М.: Наука, 1984.

ON ONE PARADOX OF LAW OF LARGE NUMBERS FOR MAXIMUM SERIES IN THE SERIES CIRCUIT OF CARPENTERS

Samara State Aerospace University

The previously published results of investigating laws governing the formation of series in the series circuit are augmented and refined. The limited forms of the laws of distribution of the length of the maximum series, formed by position relative to median and order relation between the adjacent individual values, are obtained. It is established that the limited form of the distribution of the length of a maximum series of the order relations behaves paradoxically, cyclically evolving from the singular distribution with the isolated mode to the binary with the mode, symmetrically split to two contract of those standing of value.

The keywords: series circuit, the length of a maximum series, law of large numbers, the structure of series, the criterion of chance.

Andrey Plotnikov, Candidate of Technics, Associate Professor at the Manufacturing of Flight Vehicles and Management of Quality in Machine Building Department.
E mail: metatlas@yandex.ru.