РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

АВТОМЕТРИЯ

№ 5 1995

УДК 519.24

Б. В. Бахарев

(Пущино Московской обл.)

ПРОВЕРКА СЛУЧАЙНОСТИ ПО ДЛИНЕ СЕРИИ В ПОСЛЕДОВАТЕЛЬНОСТИ, АППРОКСИМИРУЕМОЙ ИСПЫТАНИЯМИ БЕРНУЛЛИ

Применена теория серий комбинаторного анализа для вычисления доверительных уровней оценки неслучайной компоненты в последовательности значений корреляционной гистограммы межимпульсных интервалов всех порядков, строящейся по импульсным последовательностям нейронов мозга животных [2]. В случае независимых импульсных потоков каждое значение этой гистограммы аппроксимируется одним и тем же распределением Пуассона [3]. Принимая за «успех» случайный выход за доверительный интервал значения коррелограммы, находим вероятность появления хотя бы одной серии, состоящей не менее чем из r «успехов» в n независимых испытаниях Бернулли. Приводится точная формула, рекуррентное соотношение и численные оценки, связанные с разложением на простейшие дроби и в степенной ряд, а также процедура вычисления вероятности «успеха» p_r методом последовательных приближений при заданном уровне значимости.

Введение. Одно из главных применений теории серий — проверка случайности в последовательности данных одного наблюдения [1. С. 62]. Обычно о случайности судят по количеству серий. Подсчет серий может обнаружить перемешивание, большее или меньшее случайного, что позволяет говорить об определенных закономерностях: избыток серий указывает на сильное перемешивание, недостаток — на сильную группировку. При этом не надо знать вероятность появления элемента того или другого типа, важно, что все размещения равновероятны.

У нас же возникла задача проверки случайности экстремальных отклонений от среднего уровня в последовательности значений корреляционной гистограммы межимпульсных интервалов всех порядков, строящейся по импульсным последовательностям нейронов мозга животных [2]. В случае независимых импульсных потоков каждое значение гистограммы аппроксимируется распределением Пуассона [3] с одним и тем же средним значением. Определим доверительные уровни оценки неслучайной компоненты следующим образом. Будем считать «успехом» случайный выход за доверительный интервал значения коррелограммы, и пусть вероятность этого события будет p. При размытости пиков гистограммы часто бывает, что сильное отклонение имеют несколько соседних точек. Пусть вероятность того, что никакие r точек подряд из n значений гистограммы не выйдут за доверительный интервал, равна 0,95, т. е. будем считать вероятность случайного выброса, по крайней мере, в r соседних точках равной 0,05 (5 %-ный уровень значимости, принятый в биологии). Вычисляя таким образом p как функцию от r и n, построим доверительный интервал, используя нормализующее преобразование [4] и нормальную функцию распределения. Сравним последовательно данные гистограммы с доверительными уровнями для $r = 1, 2, 3, \ldots$. Если значения гистограммы выше верхнего доверительного уровня или ниже нижнего, можно говорить с 5 %-ным уровнем значимости о возбуждении или торможении и оценить величину сдвига корреляционной зависимости.

Таким образом, задача свелась к нахождению вероятности появления хотя бы одной серии, состоящей не менее чем из r «успехов» в n независимых испытаниях, с вероятностью «успеха» при каждом испытании p. Подобная задача может возникнуть, например, при передаче n сообщений по каналу связи, когда каждое сообщение с вероятностью p независимо от других искажается. Малая вероятность того, что будет искажено r и более сообщений подряд, позволит предположить, что сообщение все-таки будет расшифровано. Другой пример — выход из строя не менее r элементов подряд, например, осветительных или обогревательных приборов, что может вызвать технологические нарушения. Иногда бывает полезно оценить вероятность длинной серии. Малая вероятность такого события позволяет подозревать неслучайность данного наблюдения.

1. **Постановка задачи.** Случайная величина $X \in \mathbf{N}_n$ равна максимальной длине серии «успехов» в n испытаниях Бернулли. Вероятность «успеха» в одном испытании равна p, а вероятность «неудачи» $q = 1 - p$. Найти точные и приближенные значения вероятности $P_r(n)\{X \geq r\}$ и $P_r(n)\{X = r\}$.

2. **Точное решение.** Очевидно: $P_r(n)\{X \geq r\} = 1 - P_r(n)\{X < r\}$. Пусть k — число «успехов», k_i — число серий «успехов» длиной $i < r$ и

$$\sum_{i=1}^{r-1} k_i = l, \qquad \sum_{i=1}^{r-1} ik_i = k.$$

Задача сводится к размещению l серий «успехов» среди $n - k$ «неудач» так, чтобы между сериями была хотя бы одна «неудача», т. е. серия «успехов» может быть либо первой, либо последней, либо в промежутках между «неудачами». Число способов выбора l мест из $n - k + 1 \geq l$ есть $\binom{n-k+1}{l}$ [1. С. 54], а внутри каждого варианта возможно A расположений различными способами [1. С. 57]:

$$A = \frac{l!}{k_1!\,k_2!\,\dots\,k_{r-1}!}.$$

Таким образом, вероятность $P_r(n)\{X < r\}$ равна

$$\sum_{k=a}^{[(r-1)(n+1)/r]} \sum_{k_{r-1}=b}^{[k/(r-1)]} \dots \sum_{k_i=0}^{c_i} \dots \sum_{k_2=c}^{c_2} A\binom{n-k+1}{l} p^k q^{n-k}, \tag{1}$$

где

$$k_1 = k - \sum_{i=2}^{r-1} ik_i, \qquad a = b = 0, \qquad c = \max\left(0, 2k - n - 1 - \sum_{i=3}^{r-1}(i-1)k_i\right),$$

$$l = k - \sum_{i=2}^{r-1}(i-1)k_i, \qquad c_i = \left[\left(k - \sum_{j=i+1}^{r-1} jk_j\right)/i\right], \quad i = r-2, r-3, \dots, 2,$$

$[x]$ — наибольшее целое число, не превосходящее x. Суммирование проводится при $c \leq c_2$.

Если положить $a = r - 1$ и $b = 1$, то выражение (1) равно вероятности $P_{r-1}(n)\{X = r - 1\}$.

3. **Численные оценки.** 3.1. *Рекуррентное соотношение.* Пусть r — фиксированное положительное число, а событие ε означает появление хотя бы одной серии, по крайней мере, из r «успехов» в последовательности испытаний Бернулли. В зависимости от того, когда наступает первая «неудача» или не наступает вовсе в первых r испытаниях, имеем $r + 1$ взаимоисключающихся возможностей. Вероятность того, что первые $i - 1$ испытаний закончатся «успехами» и первая «неудача» произойдет в i-м испытании ($i \leq r$), а остальные

$n - i$ испытания приведут к появлению события ε, равна $p^{i-1}qP_r(n-i)$. Таким образом, вероятность $P_r(n)\{X \geq r\}$ удовлетворяет следующему рекуррентному соотношению:

$$P_r(n) = qP_r(n-1) + pqP_r(n-2) + \ldots +$$
$$+ p^{r-1}qP_r(n-r) + p^r, \quad n \geq r, \tag{2}$$

с начальными условиями

$$P_r(0) = P_r(1) = \ldots = P_r(r-1) = 0. \tag{3}$$

3.2. *Производящая функция.* Умножим левую и правую части (2) на s^n и просуммируем по $n = r, r+1, r+2, \ldots$. Обозначим через $R_r(s)$ производящую функцию распределения вероятностей:

$$R_r(s) = \sum_{n=0}^{\infty} P_r(n)s^n, \quad |s| < 1.$$

Тогда

$$R_r(s) = \frac{(ps)^r}{(1-s)\left[1 - qs\sum_{i=0}^{r-1}(ps)^i\right]} = \frac{U(s)}{V(s)}. \tag{4}$$

Для получения приближенной формулы используем разложение на простые дроби [1. С. 289]. Знаменатель $V(s)$ в выражении (4) имеет два положительных корня: $s_1 = 1$ и $s_2 = x_r$, причем $x_r > 1$ и по абсолютной величине меньше любого другого корня [1. С. 339]. Поэтому

$$P_r(n) \sim -\frac{U(s_1)}{V'(s_1)s_1^{n+1}} - \frac{U(s_2)}{V'(s_2)s_2^{n+1}} = 1 - \frac{1 - px_r}{q[1 - r(x_r - 1)]}\,\frac{1}{x_r^{n+1}},$$

где $\sim$ означает, что отношение правой и левой частей стремится к единице при $n \to \infty$. Так, при $r = 1$ будет $x_1 = 1/q$ и $P_1(n) = 1 - q^n$ равна вероятности, по крайней мере, одного «успеха» в n испытаниях. Точное значение $x_2 = [\sqrt{(1+3p)/q} - 1]/(2p)$. Корень x_r можно вычислить методом последовательных приближений, полагая $x_{r,0} = 1$ и $x_{r,\nu+1} = 1 + qp^r x_{r,\nu}^{r+1}$. Последовательность сходится к x_r монотонно слева. Приближенно x_r вычисляется по формуле

$$x_r = 1 + qp^r + (r+1)q^2p^{2r} + (r+1)(3r+2)q^3p^{3r}/2 + \ldots .$$

Погрешность, возникшая из-за отбрасывания $r - 1$ корней, отличных от x_r, по абсолютной величине меньше чем $2(r-1)p^{n+2}/[rq(1+p)]$ [1. С. 340].

Заметим, что

$$P_r(n)\{X = r\} = P_r(n)\{X \geq r\} - P_{r+1}(n)\{X \geq r+1\}.$$

3.3. *Разностное уравнение.* С помощью оператора смещения $\mathbf{E}$ соотношение (2) запишется в виде линейного обыкновенного неоднородного разностного уравнения порядка r с постоянными коэффициентами

$$\left(\mathbf{E}^r - q\sum_{i=1}^{r} p^{i-1}\mathbf{E}^{r-i}\right)P_r(n) = p^r, \tag{5}$$

с начальными условиями (3).

3.4. **Теорема 1.** Решение уравнения (5) имеет следующее разложение в степенной ряд относительно p:

$$P_r(n) = (n - r + 1)p^r - (n - r)p^{r+1} - \frac{(n - 2r + 1)(n - 2r)}{2} p^{2r} + (n - 2r)^2 p^{2r+1} -$$

$$- \frac{(n - 2r - 1)(n - 2r)}{2} p^{2r+2} + \frac{(n - 3r + 1)(n - 3r)(n - 3r - 1)}{6} p^{3r} - \dots, \quad r > 2,$$

$$P_2(n) = (n - 1)p^2 - (n - 2)p^3 - \frac{(n - 3)(n - 4)}{2} p^4 + (n - 4)^2 p^5 +$$

$$+ \frac{(n - 5)(n^2 - 16n + 54)}{6} p^6 - \frac{(n - 6)^2(n - 7)}{2} p^7 + \dots,$$

$$P_1(n) = 1 - q^n = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j+1} p^{j+1},$$

причем все коэффициенты либо положительные, либо нуль.

Разложение полезно при $(n - r)qp^r \ll 1$, так как дает представление о порядке величины вероятности $P_r(n)\{X \geq r\}$.

4. **Доказательство теоремы 1.**

Представление 1. Решение уравнения (5) представимо в виде степенного ряда относительно p:

$$P_r(n) = \sum_{j=0}^{n-r} \alpha_{j,r}(n - r)p^{r+j}. \tag{6}$$

Доказательство. Минимальная степень p равна r, так как предполагается, что последовательность испытаний Бернулли содержит не менее r «успехов». При $n = r$ будет только первый член, причем $\alpha_{0,r}(0) = 1$. Максимальная степень p определяется длиной последовательности испытаний n. Поэтому при $j > n - r$ все коэффициенты $\alpha_j(n - r) = 0$.

Лемма 1.

$$\nabla \alpha_{0,r}(n) = 1, \qquad \alpha_{0,r}(n) = n + 1. \tag{7}$$

Доказательство. Подставим (6) в (5) и приравняем к единице коэффициент при p^r в левой части уравнения (5). Получим разностное уравнение 1-го порядка

$$\alpha_{0,r}(n) - \alpha_{0,r}(n - 1) = 1$$

с начальным значением $\alpha_{0,r}(0) = 1$, решение которого (7) [5. С. 585].

Лемма 2. Для $j > 0$ начальные значения коэффициентов в выражении (6) равны:

$$\alpha_{j,r}(j) = -\sum_{i=\max(0,\, j-r)}^{j-1} \alpha_{i,r}(i). \tag{8}$$

Доказательство. Подставим (6) в (5) и приравняем к нулю коэффициенты при p^{r+j}, и с учетом того, что $\alpha_{j-i}(j - i - 1) = 0$, получим (8).

Следствие из леммы 2.

$$\alpha_{j,r}(j) = \begin{cases} 1, & \text{если остаток от деления } j \text{ на } r + 1 \text{ равен } 0, \\ -1, & \text{если остаток от деления } j \text{ на } r + 1 \text{ равен } 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Лемма 3. При $r > 1$

$$\nabla \alpha_{1,r}(n) = -1, \qquad \alpha_{1,r}(n) = -n. \tag{9}$$

Доказательство. Если $r > j > 0$, то в результате подстановки (6) в (5) получим разностное уравнение:

$$\sum_{i=0}^{j} \nabla \alpha_{i,r}(n - j + i) = 0. \tag{10}$$

Для $j = 1$ с учетом (7) и начального значения $\alpha_{1,r}(1) = -1$ получим (9).

Лемма 4. При $1 < j < r$

$$\nabla \alpha_{j,r}(n) = 0, \qquad \alpha_{j,r}(n) = 0. \tag{11}$$

Доказательство. $\alpha_{j,r}(j) = 0$, так как при $j < r$ остаток от деления j на $r + 1$ равен $j > 1$. Подставляя (7) и (9) в (10), получим (11).

Лемма 5. Если $2 \le r \le j$, коэффициенты $\alpha_{j,r}(n)$ удовлетворяют следующему разностному уравнению 1-го порядка:

$$\nabla \alpha_{j,r}(n) = -\sum_{i=j-r+1}^{j-1} \nabla \alpha_{i,r}(n - j + i) - \alpha_{j-r,r}(n - r). \tag{12}$$

Доказательство. В результате подстановки (6) в (5) и с учетом того, что $j \ge r$, получим разностное уравнение

$$\sum_{i=j-r+1}^{j} \nabla \alpha_{i,r}(n - j + i) + \alpha_{j-r,r}(n - r) = 0,$$

из которого следует (12). Это уравнение позволяет вычислять старшие коэффициенты через младшие.

Лемма 6.

$$\nabla \alpha_{r,r}(n) = -(n - r), \qquad \alpha_{r,r}(n) = -\frac{(n - r + 1)(n - r)}{2}, \qquad r = 2, 3, \dots . \tag{13}$$

Доказательство. Для $j = r$ уравнение (12) с учетом (11) перепишется в виде

$$\nabla \alpha_{r,r}(n) = -\nabla \alpha_{1,r}(n - r + 1) - \alpha_{0,r}(n - r).$$

Подставляя (7) и (9) и учитывая, что $\alpha_{r,r}(r) = 0$, получим (13).

Лемма 7.

$$\nabla \alpha_{r+1,r}(n) = 2(n - r) - 1, \qquad \alpha_{r+1,r}(n) = (n - r)^2, \qquad r = 2, 3, \dots . \tag{14}$$

Доказательство. Для $j = r + 1$ уравнение (12) с учетом (11) будет

$$\nabla \alpha_{r+1,r}(n) = -\nabla \alpha_{r,r}(n - 1) - \alpha_{1,r}(n - r).$$

Подставляя (9) и (13) и учитывая, что $\alpha_{r+1,r}(r + 1) = 1$, получим (14).

Лемма 8.

$$\nabla \alpha_{r+2,r}(n) = -(n - r) + 1, \qquad \alpha_{r+2,r}(n) = -\frac{(n - r - 1)(n - r)}{2}, \qquad r = 3, 4, \dots,$$

$$\nabla \alpha_{4,2}(n) = -\frac{n^2 - 11n + 26}{2}, \qquad \alpha_{4,2}(n) = -\frac{(n - 3)(n^2 - 12n + 26)}{6}. \tag{15}$$

Д о к а з а т е л ь с т в о. Если $r > 2$, то для $j = r + 2$ уравнение (12) с учетом (11) запишется так:

$$\nabla \alpha_{r+2,r}(n) = -\nabla \alpha_{r+1,r}(n-1) - \nabla \alpha_{r,r}(n-2) - \alpha_{2,r}(n-r).$$

При $r = 2$ уравнение (12) будет

$$\nabla \alpha_{4,2}(n) = -\nabla \alpha_{3,2}(n-1) - \alpha_{2,2}(n-2).$$

Подставляя (11), (13), (14) и учитывая, что $\alpha_{r+2,r}(r+2) = -1$, получим (15).

Лемма 9.

$$\nabla \alpha_{r+3,r}(n) = 0, \qquad \alpha_{r+3,r}(n) = 0, \quad r = 4, 5, \ldots,$$

$$\nabla \alpha_{6,3}(n) = \frac{(n-6)(n-7)}{2}, \qquad \alpha_{6,3}(n) = \frac{(n-5)(n-6)(n-7)}{6}, \tag{16}$$

$$\nabla \alpha_{5,2}(n) = -\frac{(n-5)(3n-14)}{2}, \qquad \alpha_{5,2}(n) = -\frac{(n-4)^2(n-5)}{2}.$$

Д о к а з а т е л ь с т в о. Перепишем уравнение (12) для $j = r + 3$

$$\nabla \alpha_{r+3,r}(n) = -\sum_{i=4}^{r+2} \nabla \alpha_{i,r}(n-r-3+i) - \alpha_{3,r}(n-r)$$

и рассмотрим три случая: $r > 3$, $r = 3$ и $r = 2$. Воспользуемся результатами (11), (13)—(15) и с учетом того, что $\alpha_{r+3,r}(r+3) = 0$, получим (16).

Лемма 10. При $r + 3 \le j < 2r$, $r = 4, 5, \ldots,$

$$\nabla \alpha_{j,r}(n) = 0, \qquad \alpha_{j,r}(n) = 0. \tag{17}$$

Д о к а з а т е л ь с т в о. Для $j = r + m$ уравнение (12) будет

$$\nabla \alpha_{r+m,r}(n) = -\sum_{i=m+1}^{r+m-1} \nabla \alpha_{i,r}(n-r-m+i) - \alpha_{m,r}(n-r), \quad m = 3, 4, \ldots, r-1.$$

Поскольку $r + 2 \le r + m - 1$ и $r \ge m + 1$, сумма содержит выражение

$$\sum_{i=r}^{r+2} \nabla \alpha_{i,r}(n-r-m+i),$$

которое обращается в нуль после подстановки в него (13)—(15). Учитывая (11) и то, что $\alpha_{r+m,r}(r+m) = 0$, получим (17).

Аналогично рассуждая, можно вычислить сколь угодно коэффициентов в разложении (6). Мы же ограничимся еще одним, чтобы по нему дать возможность судить о точности вычисления $P_r(n)$.

Лемма 11. При $r > 2$

$$\nabla \alpha_{2r,r}(n) = \frac{(n-2r)(n-2r-1)}{2}, \qquad \alpha_{2r,r}(n) = \frac{(n-2r+1)(n-2r)(n-2r-1)}{6}.$$

Д о к а з а т е л ь с т в о. Для $j = 2r$ уравнение (12) с учетом (17) будет

$$\nabla \alpha_{2r,r}(n) = -\nabla \alpha_{r+2,r}(n-r+2) - \nabla \alpha_{r+1,r}(n-r+1) - \alpha_{r,r}(n-r).$$

Подставляя (13)—(15) и учитывая, что $\alpha_{2r,r}(2r) = 0$, получим искомые выражения.

5. Численная оценка р. В заключение приведем процедуру вычисления вероятности «успеха» p_r методом последовательных приближений, когда вероятность $P_r(n)\{X \geq r\} = F_{r,n}(p_r) = \alpha$.

Если $r = 1$, то $p_1 = 1 - \exp((\ln(1 - \alpha))/n)$. При $r > 1$ в качестве начального приближения можно взять $p_{r,0} = \exp((\ln\alpha/(n - r))/r)$ и вычислить $p_{r,\nu+1} = p_{r,\nu} - f(p_{r,\nu})$, где

$$f(p) = (F_{r,n}(p) - \alpha)/F'_{r,n}(p) =$$

$$= \frac{1 - px_r - (1 - \alpha)q[1 + r(1 - x_r)]x_r^{n+1}}{(1 - x_r)/q + x'\{(rq - p)/[1 + r(1 - x_r)] - (1 - px_r)(n + 1)/x_r\}},$$

$$x' = \frac{(x_r - 1)x_r(rq - p)}{qp[1 + r(1 - x_r)]}.$$

СПИСОК ЛИТЕРАТУРЫ

1. **Феллер В.** Введение в теорию вероятностей и ее приложения. М.: Мир, 1984. Т. 1.
2. **Бахарев Б. В.** Оценка значимости экстремумов корреляционных функций импульсных последовательностей // Журн. высш. нервн. деятельности им. И. П. Павлова. 1991. **41**, вып. 3.
3. **Brillinger D. R.** Estimation of the second-order intensities of a bivariate stationary point process // J. R. Statist. Soc. 1976. **38**, N 1. P. 60.
4. **Бахарев Б. В., Ковалев А. Э.** Оптимизация нормализующего преобразования пуассоновского процесса для оценки доверительного интервала неслучайного отклонения // Математические заметки. 1992. **51**, вып. 2.
5. **Корн Г., Корн Т.** Справочник по математике. М.: Наука, 1968.

Поступила в редакцию 17 мая 1995 г.