УДК 519.217

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ СЕРИЙ В СТАЦИОНАРНОЙ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Кирьянова В. С., Плотников А. Н.

Самарский государственный аэрокосмический университет имени академика С. П. Королёва (национальный исследовательский университет), г. Самара

Положения индивидуальных значений временного ряда относительно медианы образуют двоичную последовательность ($\mathbf{0}$ - ниже, $\mathbf{1}$ - выше медианы). В более общей постановке возникает последовательность испытаний по схеме Бернулли. Дальнейшее обобщение приводит к двоичной цепи Маркова. Характерными структурами такой последовательности являются серии – отрезки вида $\mathbf{0}\underbrace{\mathbf{1}\dots\mathbf{1}}_{l}\mathbf{0}$ и их зеркальные копии. Вывод закона распределения числа серий проще всего осуществить с помощью индикаторных переменных.

Введём в рассмотрение индикаторные переменные $U_n^{(l)}$, принимающие значения $\mathbf{1}$, если очередная простая серия единиц длиной l завершилась на n-ом шаге и $\mathbf{0}$ – в противном случае, как при анализе биномиального закона. Соответствующие вероятности обозначим $u_n^{(l)}$ так, что $\mu_U = \mu_{U^2} = u$. Следуя Феллеру, воспользуемся рекуррентно-рекурсивной схемой. Последовательность испытаний будем полагать неограниченно продолжающейся, а за начало отсчёта $n = 0$ примем момент образования очередной серии, т.е. доопределим $u_0^{(l)} = 1$. С учётом последнего уточнения ряд $u_n^{(l)}$ получим, перемножая вероятности переходов в цепочке-серии

$$u_n^{(l)} = \begin{cases} 1, & n = 0 \\ 0, & 1 \le n \le l \\ \gamma\beta^{\,l-1}, & n = l+1 \\ \gamma\beta^{\,l-1}(1-\beta), & n \ge l+2 \end{cases},$$

где $\gamma = p_0(1-\alpha) = \dfrac{(1-\alpha)(1-\beta)}{2-\alpha-\beta}$ – предельная вероятность перехода $*0 \to *01$. Число серий и индикаторные переменные связаны между собой соотношением $R_l^{(n)} = \sum\limits_{i=0}^{n} U_i^{(l)}$.

В силу центральной предельной теоремы последняя сумма сходится по распределению к нормальному закону. Причём её среднее и дисперсия пропорциональны n с точностью до аддитивной константы. Поэтому при вычислении параметров нормальной асимптотики конечной целью являются пределы $\mu_l = \lim\limits_{n\to\infty} \dfrac{1}{n} M\left[\sum\limits_{i=0}^{n} U_i^{(l)}\right]$, $\sigma_l^2 = \lim\limits_{n\to\infty} \dfrac{1}{n} D\left[\sum\limits_{i=0}^{n} U_i^{(l)}\right]$. Непосредственным подсчётом получаем

$$\mu_l = \gamma\beta^{\,l-1}(1-\beta),\ \sigma_l^2 = \gamma\beta^{\,l-1}(1-\beta)\{1-\beta^{\,l-1}(1-\beta)[(2l+5)\gamma - 2(1-\alpha^2)]\}.$$

Таким образом, параметры нормальной асимптотики $R_l^{(n)}$ имеют вид $R_l^{(n)} \sim N(n\mu_l, \sigma_l\sqrt{n})$.

Устремив $l \to \infty$, замечаем, что $\sigma_l^2 \to \mu_l$ ($\frac{\mu_l - \sigma_l^2}{\mu_l} \sim O\left(l\beta^l\right)$). Следовательно, для длинных серий наряду с нормальной справедлива и пуассоновская асимптотика.

При исследовании зависимости между величинами $R_l^{(n)}, l = 1,2,...$ воспользуемся нормальной асимптотикой. Взаимосвязь в системе нормальных величин исчерпывается линейной, которую удобнее всего представлять в виде ковариационной, либо корреляционной матрицы. По аналогии с дисперсией введём в рассмотрение дисперсию суммы $\sigma_{k,l}^2 = \lim\limits_{n\to\infty} \frac{1}{n} D\left[R_k^{(n)} + R_l^{(n)}\right], l > k$. Ряд u_n для составного события в силу несовместности комбинируемых событий будет иметь вид

$$u_n \sim \underset{0}{1}, \underset{1}{0}, ..., \underset{k}{0}, \underset{k+1}{\mu_k^0}, \underset{k+2}{\mu_k^1}, ..., \underset{l+1}{\mu_k^1 + \mu_l^0}, \underset{l+2}{\mu_k^1 + \mu_l^1}, ...,$$

где обозначено $\mu_k^0 = \gamma\beta^{k-1}$, $\mu_k^1 = \gamma\beta^{k-1}(1-\beta)$.

Произведя последовательность вычислений, аналогичную предыдущей, находим $\lim\limits_{n\to\infty} \frac{1}{n} COV\left[R_k^{(n)}, R_l^{(n)}\right] = \frac{\sigma_{k,l}^2 - \sigma_k^2 - \sigma_l^2}{2} = \gamma\beta^{k+l-2}(1-\beta)^2\left[2\left(1-\alpha^2\right) - \gamma(k+l+5)\right]$. Поделив на $\sigma_k\sigma_l$, получаем предельный коэффициент корреляции

$$\rho_{k,l} = \frac{(1-\beta)\beta^{\frac{k+l}{2}-1}}{\sqrt{1-\beta^{k-1}(1-\beta)\left((2k+5)\gamma - 2\left(1-\alpha^2\right)\right)}} \times \frac{\left[2\left(1-\alpha^2\right) - \gamma(k+l+5)\right]}{\sqrt{1-\beta^{l-1}(1-\beta)\left((2l+5)\gamma - 2\left(1-\alpha^2\right)\right)}}. \qquad (1)$$

Как видно из (1), при $k, l \to \infty$ $\rho_{k,l} < 0$, $\left|\rho_{k,l}\right| = O\left((k+l)\beta^{\frac{k+l}{2}}\right)$.

Таким образом, числа длинных серий успехов в двоичной марковской последовательности имеет нормально-пуассоновскую асимптотику и образует взаимно независимую систему величин.

Следует отметить интересный факт. Знак корреляции коротких серий определяется, согласно (1), величиной $2\left(1-\alpha^2\right) - \gamma(k+l+5)$. Стало быть при $k + l < \frac{2\left(1-\alpha^2\right)}{\gamma} - 5$ корреляция положительна, при смене знака неравенства – отрицательна, а в случае равенства $R_k^{(n)}$ и $R_l^{(n)}$ независимы. В качестве иллюстрации ниже приведён начальный фрагмент матрицы $\rho_{k,l}, k, l \in \{1,2,...,7\}$ для симметричной цепи Бернулли (орлянки):

$$\rho = \begin{pmatrix} 1 & -0{,}21 & -0{,}218 & -0{,}2 & -0{,}173 & -0{,}145 & -0{,}118 \\ -0{,}21 & 1 & -0{,}214 & -0{,}184 & -0{,}153 & -0{,}124 & -0{,}099 \\ -0{,}218 & -0{,}214 & 1 & -0{,}153 & -0{,}123 & -0{,}098 & -0{,}078 \\ -0{,}2 & -0{,}184 & -0{,}153 & 1 & -0{,}097 & -0{,}076 & -0{,}059 \\ -0{,}173 & -0{,}153 & -0{,}123 & -0{,}097 & 1 & -0{,}059 & -0{,}045 \\ -0{,}145 & -0{,}124 & -0{,}098 & -0{,}076 & -0{,}059 & 1 & -0{,}034 \\ -0{,}118 & -0{,}099 & -0{,}078 & -0{,}059 & -0{,}045 & -0{,}034 & 1 \end{pmatrix}.$$