АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ЧИТИНСКИЙ ИНСТИТУТ ПРИРОДНЫХ РЕСУРСОВ

Математические МОДЕЛИ рационального ПРИРОДО-ПОЛЬЗОВАНИЯ

СБОРНИК НАУЧНЫХ ТРУДОВ

Ответственные редакторы
доктор физико-математических наук
В. В. Пененко
доктор биологических наук
И. Б. Токин

СОДЕРЖАНИЕ

В. М. БУРЕ

КРИТЕРИЙ СЛУЧАЙНОСТИ, ОСНОВАННЫЙ НА ЗНАКАХ РАЗНОСТЕЙ ДЛЯ ВРЕМЕННЫ́Х РЯДОВ С КОНЕЧНЫМ МНОЖЕСТВОМ ДОПУСТИМЫХ ЗНАЧЕНИЙ ПЕРЕМЕННОЙ

В некоторых биологических приложениях временны́е ряды носят сугубо дискретный характер в том смысле, что переменная, меняющаяся во времени, может принимать лишь дискретные значения, причем часто совокупность этих значений конечна. Например, динамика урожайности сельскохозяйственной культуры или динамика численности популяции на заданной площади являются дискретными временны́ми рядами, при этом очевидно, что множество допустимых значений конечно.

Простейшей гипотезой, которую можно выдвинуть относительно колеблющегося ряда, является предположение о том, что колебания случайны [1]. Такая задача возникает, например, при исследовании остатков, полученных вычитанием из исходного ряда систематических элементов, когда требуется установить, не осталось ли в них какой-либо систематизации. В случайных рядах наблюдения независимы и могут следовать в любом порядке. Выбор критерия проверки гипотезы зависит от того, какие выдвигаются альтернативные гипотезы. Когда в качестве альтернативной гипотезы рассматривается гипотеза о наличии линейного тренда, обычно используют критерий, основанный на знаках разностей [1], который состоит в подсчете числа положительных разностей первого порядка в ряде, т. е. числа точек возрастания ряда. Этот критерий приведен в [1] в предположении, что переменная может принимать любые вещественные значения. Покажем, как изменится нормированная статистика критерия, если исследуемый временно́й ряд является дискретным, причем множество допустимых значений переменной конечно. После чего, используя центральную предельную теорему для m-зависимых случайных величин [2], докажем асимптотическую нормальность нормированной статистики.

Итак, имеем последовательность дискретных случайных величин, принимающих конечное множество значений: $\xi_1, \xi_2, \ldots, \xi_t, \ldots, \xi_n$; $\xi_t \in \{x_1, x_2, \ldots, x_N\}$, где $x_i < x_{i+1}$. Проверяемая гипотеза H_0 состоит в предположении о независимости и равномерной распределенности случайных величин ξ_t, т. е.

$$P\{\xi_t = x_i\} = 1/N,$$

$$P\{\xi_{i_1} = x_{i_1}, \xi_{i_2} = x_{i_2}, \ldots, \xi_{i_k} = x_{i_k}\} =$$

$$= P\{\xi_{i_1} = x_{i_1}\} \cdot P\{\xi_{i_2} = x_{i_2}\} \cdot \ldots \cdot P\{\xi_{i_k} = x_{i_k}\},$$

где k — произвольное целое число. Определим счетную переменную

В предположении истинности гипотезы H_0 вычислим вероятность $P\{I_i = 1\}$. Формируя выборку из двух элементов с возвращением, имеем N^2 исходов, а число пар (y_1, y_2) таких, что $y_1 < y_2$, есть число сочетаний $\binom{N}{2}$. Таким образом,

$$P\{I_i = 1\} = (N-1)/2N;$$

$$P\{I_i = 0\} = P\{\xi_i \geqslant \xi_{i+1}\} = 1 - \frac{N-1}{2N} = \frac{N+1}{2N}.$$

Следовательно,

$$E(I_i) = \frac{N-1}{2N}, \quad E\left(\sum_{i=1}^{n-1} I_i\right) = (n-1)\frac{N-1}{2N}.$$

Лемма. *Если гипотеза H_0 истинна, то случайные величины I_i и I_{i+1} являются зависимыми, в то время как I_i и I_{i+k} при $k>1$ — независимые случайные величины.*

Доказательство. Очевидно

$$P\{I_i = 1,\ I_{i+1} = 1\} = P\{\xi_i < \xi_{i+1} < \xi_{i+2}\} =$$

$$= \binom{N}{3}\Big/N^3 = \frac{(N-1)(N-2)}{6N^2} \neq \frac{(N-1)^2}{(2N)^2},$$

откуда следует зависимость случайных величин I_i и I_{i+1}. При $k>1$ в предположении истинности гипотезы H_0 имеем

$$P\{I_i = 1, I_{i+k} = 1\} = P\{\xi_i < \xi_{i+1},\ \xi_{i+k} < \xi_{i+k+1}\} =$$

$$= \binom{N}{2}\binom{N}{2}\Big/N^4 = P\{I_i = 1\}P\{I_{i+k} = 1\}.$$

Отсюда видно, что случайные величины I_i, I_{i+k} независимы, так как из независимости событий A и B следует независимость событий $\bar{A}$ и B, A и $\bar{B}$, $\bar{A}$ и $\bar{B}$. Лемма доказана.

Вычислим дисперсию суммы случайных величин I_i:

$$D\left(\sum_{i=1}^{n-1} I_i\right) = \sum_{i=1}^{n-1} D(I_i) + 2\sum_{i=1}^{n-2} E\{(I_i - EI_i)(I_{i+1} - EI_{i+1})\} =$$

$$= (n-1)D(I_1) + 2(n-2)\{E(I_1 \cdot I_2) - E(I_1)E(I_2)\} =$$

$$= (n-1)\left[\frac{N-1}{2N} - \frac{(N-1)^2}{4N^2}\right] + 2(n-2)\left[\frac{(N-1)(N-2)}{6N^2} - \frac{(N-1)^2}{4N^2}\right] =$$

$$= n\frac{N^2-1}{12N^2} + \frac{N^2-1}{12N^2}.$$

Из леммы следует, что последовательность случайных величин $I_1, I_2, \ldots, I_t, \ldots, I_n$ является последовательностью m-зависимых (m-связанных) случайных величин при $m=1$ в смысле определения, данного в [2]. Там же имеется теорема, в которой показано, что к стационарной последовательности m-связанных случайных величин применима центральная предельная теорема. Используя эту теорему, можно доказать следующее утверждение.

Теорема. *Пусть*

$$\zeta_n = \left[\sum_{i=1}^{n-1} I_i - E\left(\sum_{i=1}^{n-1} I_i\right)\right] \Big/ \sqrt{D\left(\sum_{i=1}^{n-1} I_i\right)}.$$

Если гипотеза H_0 истинна, то статистика ζ_n асимптотически нормальна с параметрами 0 и 1, т. е. для любого вещественного x имеем

$$P\{\zeta_n < x\} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-u^2/2}\, du.$$

С помощью теоремы можно построить асимптотическую критическую область для гипотезы H_0 с заданной величиной ошибки 1-го рода.

СПИСОК ЛИТЕРАТУРЫ

1. **Кендэл М.** Временные ряды.— М.: Финансы и статистика, 1981.
2. **Ибрагимов И. А., Линник Ю. В.** Независимые и стационарно связанные величины.— М.: Наука, 1965.