EESTI NSV TEADUSTE AKADEEMIA TOIMETISED.
FÜÜSIKA * MATEMAATIKA
ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ФИЗИКА * МАТЕМАТИКА
PROCEEDINGS OF THE ACADEMY OF SCIENCES OF THE ESTONIAN SSR.
PHYSICS * MATHEMATICS

1987, 36, 4

УДК 519.24

В. ЙОАЛА, В. ОЛЬМАН

МЕТОД ОЦЕНИВАНИЯ КОЭФФИЦИЕНТОВ ЛИНЕЙНОЙ ОДНОМЕРНОЙ РЕГРЕССИИ

V. JOALA, V. OLMAN. LINEAARSE ÜHEMÕÕTMELISE REGRESSIOONI KOEFITSIENTIDE HINDAMISE MEETOD

V. JOALA, V. OLMAN. AN ESTIMATION METHOD OF COEFFICIENTS OF ONE-DIMENSIONAL LINEAR REGRESSION

(Представил Н. Алумяэ)

Пусть наблюдается процесс

$$x(t)=a_0+a_1t+u_t, \quad t=1, \ldots, M,$$

где случайные величины u_t некоррелированы и имеют средние значения $Eu_t=0$ и дисперсии $Eu_t^2=\sigma^2>0$.

Для оценивания коэффициентов a_0 и a_1 обычно применяют метод наименьших квадратов [1]. С этой целью минимизируется по a_0 и a_1 следующая функция

$$\sum_{t=1}^{M} u_t^2=\sum_{t=1}^{M} [x(t)-a_0-a_1t]^2.$$

Оценки $\bar{a}_0$ и $\bar{a}_1$ коэффициентов a_0 и a_1 определяются следующим образом:

$$\bar{a}_1=\frac{12\sum\limits_{t=1}^{M} tx(t)-6(M+1)\sum\limits_{t=1}^{M} x(t)}{M(M^2-1)}, \tag{1}$$

$$\bar{a}_0=\frac{1}{M}\sum_{t=1}^{M} x(t)-\bar{a}_1\frac{M+1}{2}. \tag{2}$$

В литературе [2, 3] приведены формулы для определения дисперсий оценок $\bar{a}_0$ и $\bar{a}_1$.

$$D(\bar{a}_1)=\frac{\sigma^2}{\sum\limits_{t=1}^{M}(t-\bar{t})^2}=\frac{12}{M^3-M}\sigma^2,$$

$$D(\bar{a}_0)=\frac{\sigma^2}{M}+(\bar{t})^2D(\bar{a}_1)=2\frac{2M+1}{M(M-1)}\sigma^2,$$

где $\bar{t}=\frac{1}{M}\sum\limits_{t=1}^{M} t=\frac{M+1}{2}$, Как видно из формул (1) и (2), для вычисления оценок $\hat{a}_0$ и $\hat{a}_1$ необходимо вычислить суммы

$$\sum_{t=1}^{M} x(t) \quad \text{и} \quad \sum_{t=1}^{M} t x(t),$$

которые требуют выполнения M умножений и $2M$ сложений.

Оценка метода наименьших квадратов является наилучшей среди нелинейных несмещенных оценок. В настоящей работе приведен линейный метод оценивания, который при незначительной потере в точности позволяет существенно сократить время вычислений оценок.

Метод основан на вычислении последовательных разностей элементов исходного ряда с шагом N. Для оценивания коэффициента a_1 усредняется k разностей ($k>0$)

$$x(t+N)-x(t), \quad t=1, \ldots, k.$$

I. Оценка

$$\hat{a}_1=\frac{1}{Nk}\sum_{t=1}^{k}[x(t+N)-x(t)]=\frac{1}{Nk}\left[\sum_{t=N+1}^{N+k} x(t)-\sum_{t=1}^{k} x(t)\right]$$

является несмещенной и состоятельной оценкой коэффициента a_1 для любых N и k ($N+k\leqslant M$).

Доказательство. Представим оценку $\hat{a}_1$ в виде

$$\hat{a}_1=a_1+\frac{1}{Nk}\left[\sum_{t=1}^{k} u_{t+N}-\sum_{t=1}^{k} u_t\right],$$

откуда легко следует, что

$$E(\hat{a}_1)=a_1.$$

Дисперсию оценки $\hat{a}_1$ выведем в двух случаях.

1) При $N+1>k$ получаем

$$E(\hat{a}_1-a_1)^2=\frac{1}{N^2k^2}E\left\{\sum_{t=1}^{k}(u_{t+N}-u_t)\right\}^2=\frac{2}{N^2k}\sigma^2.$$

Можно показать, что минимальное значение полученной дисперсии при ограничении $N+k\leqslant M$ достигается при $k=M/3$.

2) В результате несложных преобразований получаем, что при $N+1<k$

$$E(\hat{a}_1-a_1)^2=\frac{2}{Nk^2}\sigma^2.$$

Минимум полученной дисперсии достигается здесь при $k=\frac{2M}{3}$.

При $N+k=M$ ввиду минимизации вычислений выбираем первый вариант ($k=M/3$ и $N=2M/3$) и отсюда получаем

$$\min_{N,k} D(\hat{a}_1)=\frac{13{,}5}{M^3}\sigma^2.$$

II. Оценка

$$\hat{a}_0=\frac{1}{M}\left[\sum_{t=1}^{M} x(t)-\hat{a}_1\sum_{t=1}^{M} t\right]$$

является несмещенной и состоятельной оценкой коэффициента a_0.

Доказательство.

$$\hat{a}_0 = a_0 + \frac{1}{M}\left\{\sum_{t=1}^{M}\left[u_t - \frac{t}{Nk}\left(\sum_{t=1}^{k} u_{t+N} - \sum_{t=1}^{k} u_t\right)\right]\right\},$$

$$E(\hat{a}_0) = a_0,$$

$$E(\hat{a}_0 - a_0)^2 = \frac{35M^2 + 54M + 27}{8M^3}\sigma^2.$$

Окончательные формулы вычисления оценок $\hat{a}_0$ и $\hat{a}_1$ следующие:

$$\hat{a}_1 = \frac{4{,}5}{M^2}(S_3 - S_1), \tag{3}$$

$$\hat{a}_0 = \frac{1}{M}\left(S_1 + S_2 + S_3 - \hat{a}_1 \frac{M(M+1)}{2}\right), \tag{4}$$

где $S_1 = \sum_{t=1}^{k} x(t)$, $S_2 = \sum_{t=k+1}^{M-k} x(t)$ и $S_3 = \sum_{t=M-k+1}^{M} x(t)$,

а k ближайшее целое число от $M/3$.

По приведенным в [4] данным, вычисление оценок $\hat{a}_0$ и $\hat{a}_1$ по формулам (3) и (4) при использовании арифметики с плавающей запятой на вычислительных машинах СМ-3 и СМ-4 занимает примерно в 3,2 раза меньше времени, чем вычисление этих оценок по формулам (1) и (2).

Таким образом, показано, что дисперсии оценок предлагаемого метода имеют такой же порядок по M, как и дисперсии оценок метода наименьших квадратов, и что при существенном сокращении времени вычислений, точность уменьшается незначительно, т. е. оценки (3) и (4) имеют при $M \to \infty$ соответственно на 12,5 и 9,4% большую дисперсию, чем оценки метода наименьших квадратов (1) и (2).

ЛИТЕРАТУРА

1. *Андерсон Т.* Статистический анализ временных рядов. М., «Мир», 1976.
2. *Бендат Дж., Пирсол А.* Измерение и анализ случайных процессов. М., «Мир», 1971.
3. *Джонсон Н., Лион Ф.* Статистика и планирование эксперимента в технике и науке. Методы обработки данных. М., «Мир», 1980.
4. Малые ЭВМ и их применение (ред. Б. Н. Наумов). М., «Статистика», 1980.

Институт кибернетики
Академии наук Эстонской ССР

Поступила в редакцию
2/III 1987