УДК 519.711.3

А. Ф. Турбін – доктор фізико-математичних наук, професор кафедри вищої математики Національного педагогічного університету імені М. П. Драгоманова;
О. Г. Ханін – кандидат фізико-математичних наук, доцент кафедри математичного аналізу Волинського національного університету імені Лесі Українки

Середньоквадратичне відхилення від еталону моделі випадкової величини, побудованої за допомогою S-статистики

Роботу виконано на кафедрі математичного аналізу ВНУ ім. Лесі Українки

Статтю присвячено одному з аспектів моделювання складних систем та явищ – моделюванню випадкових впливів за певною кількістю попередніх спостережень їх фактичних значень. У статті отримано середньо-квадратичну оцінку якості моделі, побудованої за допомогою S-статистики.

Ключові слова: інтерполяційний кубічний сплайн, S-статистика, середньоквадратичне відхилення, моделювання, рівномірно розподілена випадкова величина, функція та щільність розподілу генеральної сукупності.

Турбін А. Ф., Ханин А. Г. Среднеквадратическое отклонение от эталона модели случайной величины, построенной с помощью S-статистики. Статья посвящена одному из аспектов моделирования сложных систем и явлений – моделированию случайных воздействий по определенному количеству предварительных наблюдений их фактических значений. В статье получена среднеквадратическая оценка качества модели, построенной с помощью S-статистики.

Ключевые слова: интерполяционный кубический сплайн, S-статистика, среднеквадратическое отклонение, моделирование, равномерно распределенная случайная величина, функция и плотность распределения генеральной совокупности.

Turbin A. F., Khanin O. G. Mean-Square Deviation From Etalon of the Random Variable Model, Constructed by Using of S-Statistics. The article is devoted to one aspect of complex systems and phenomena modeling – modeling of random effects based on a certain number of previous observations of their actual values. The mean-square quality estimation of the model, based on S-statistics, was obtained in the paper.

Key words: cubic spline interpolation, S-statistics, mean-squre deviation, simulation, uniformly distributed random variable, general function and density of distribution .

Постановка наукової проблеми та її значення. При побудові та подальшому дослідженні комп’ютерних моделей складних фізичних явищ, економічних або технічних систем, які об’єктивно підпорядковані випадковим впливам, ймовірнісні характеристики яких апріорі невідомі, необхідно вміти здійснювати моделювання цих впливів. Зазвичай варто адекватно моделювати випадкову величину з невідомим розподілом по певній кількості спостережень над її значеннями. Один з алгоритмів такого моделювання запропоновано в [3]. Однак для його практичного використання під час моделювання складних систем та явищ надзвичайно важливо розуміти, наскільки змодельовані випадкові величини відрізняються від еталонної випадкової величини, значення якої спостерігали до початку процесу моделювання, якою мірою отримані випадкові величини можна вважати підпорядкованими тому самому розподілу, якому підпорядкована еталонна величина, тобто яким чином вони належать тій самій генеральній сукупності.

Аналіз останніх досліджень. Початок активного застосування методів сплайн-функцій до задач статистичного оцінювання відноситься до 70−80-х років минулого сторіччя [6, 57−62], [7, 351−365]. Асимптотичним властивостям непараметричних сплайн-оцінок (S-статистик) генеральної щільності розподілу, а також їх застосуванню до моделювання випадкових величин присвячено роботи А. Ф. Турбіна та О. Г. Ханіна.

Мета наукової розвідки − оцінка величини середньоквадратичного відхилення розподілу змодельованих за допомогою S-статистик випадкових величин від невідомого теоретичного розподілу, якому була підпорядкована спостережена (еталонна) випадкова величини.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Нехай є вибірка $X_1, \ldots, X_n$ з генеральної сукупності з невідомою функцією розподілу F(x), яка строго зростає на відрізку [0,1]. Сутність алгоритму моделювання випадкових величин, які підпорядковані розподілу F(x), за допомогою S-статистик полягає у такому [3].

Нехай $X_{(1)}, \ldots, X_{(n)}$ – побудований за вибіркою варіаційний ряд. Відрізок [0,1] розбивається рівномірною сіткою Δ_n: $0 = y_0 < y_1 <\ldots< y_n = 1$, де $y_{i+1}-y_i = 1/n$. По цій сітці та крайових умовах типу $s_n'(0) = na_n$, $s_n'(1) = nb_n$, де $a_n \to 0$, $b_n \to 0$ при $n \to \infty$ будується S-статистика [2] − інтерполяційний кубічний сплайн $s_n(x)$ дефекту 1 на відрізку [0, 1], який задовільняє умови інтерполяції $s_n(y_i) = X_{(i)}$, i = 0, 1, …, n (якщо за означенням $X_{(0)} = 0$). В якості моделі спостереженої випадкової величини з розподілом F(x) розглядається випадкова величина $s_n(\xi)$, де ξ рівномірно розподілена на [0, 1] випадкова величина.

Ми отримали оцінку швидкості збіжності у середньоквадратичному послідовності $\{s_n(\xi),\ n \geq 1\}$ до випадкової величини $F^{(-1)}(\xi)$, де функція $F^{(-1)}(x)$ обернена до теоретичної функції розподілу F(x). Зауважимо, що випадкова величина $F^{(-1)}(\xi)$ має своєю функцією розподілу F(x).

Теорема. Нехай виконуються три умови:

1. F(x) строго зростає на [0, 1].
2. F(x) належить класу функцій $W_p^q[0,1]$, $1 \leq p \leq 2$, q=1,…,4.
3. Теоретична щільність розподілу f(x) відділена на [0, 1] від нуля, тобто $\exists\, m > 0\ \forall\, x \in [0, 1]$: $f(x) \geq m$.

Тоді існує така константа D, що

$$M(s_n(\xi) - F^{(-1)}(\xi))^2 \leq D/n,$$

де ξ – рівномірно розподілена на [0, 1] випадкова величина.

Таким чином, при виконанні достатньо широких умов 1−3 середньоквадратична швидкість збіжності змодельованої випадкової величини до випадкової величини, значення якої реально спостерігали до початку процесу моделювання, обернено пропорційна кількості цих спостережень (об’єму вибірки).

Зауважимо, що значення константи D залежить тільки від параметрів p, q і теоретичної функції розподілу F(x). Якщо вимагається, щоб функція F(x) належала простору $W_{p,M}^q[0,1]$, що складається з функцій з класу $W_p^q[0,1]$, q-а похідна яких обмежена за абсолютною величиною константою M, то в як константу D візьмемо значення M.

У подальшому будемо використовувати також той факт, що функція $F^{(-1)}(x)$ є $W_p^q[0,1]$ з тими самими параметрами p і q, що і функція F(x).

Доведення.

$$M(s_n(\xi) - F^{(-1)}(\xi))^2 = \int_0^1 ((s_n(x) - F^{(-1)}(x))^2 dx.$$

Позначимо через $\|g\|_p$ норму функції g(x) у просторі $L_p[0, 1]$. Тоді

$$\int_0^1 ((s_n(x) - F^{(-1)}(x))^2 dx = (\|s_n(x) - F^{(-1)}(x)\|_2)^2 \leq (\|s_n(x) - s(x)\|_2)^2 + (\|s(x) - F^{(-1)}(x)\|_2)^2,$$

де s(x) – інтерполяційний кубічний сплайн дефекту 1 на [0,1], який інтерполює у вузлах сітки Δ_n значення функції $F^{(-1)}(x)$ і задовольняє крайові умови вигляду:

$$s'(0) = dF^{(-1)}(x)/dx|_{x=0},\ s'(1) = dF^{(-1)}(x)/dx|_{x=1}.$$

За виконання умов теореми є така оцінка [5]:

$$(\|s(x) - F^{(-1)}(x)\|_2)^2 \leq C(p)(\|d^q F^{(-1)}(x)/dx^q\|_p)^2 n^{-2q-1+2p}, \tag{1}$$

де C(p) – константа, яка залежить лише від значення p.

Оцінимо тепер величину $M(\|s_n(x) - s(x)\|_2)^2$. Скористаємося представленням інтерполяційного кубічного сплайну, отриманому в [4]. При $x \in [y_i, y_{i+1})$, i = 0, ..., n−1:

$$s_n(x) - s(x) = (X_{(i)} - F_i^{(-1)})h_{n,1}(x) + (X_{(i+1)} - F_{i+1}^{(-1)})\, h_{n,2}(x) + \sum_{k=1}^{n} l_k\left(\frac{[nx]}{x}\right)(X_{(k)} - X_{(k-1)} - F_k^{(-1)} + F_{k-1}^{(-1)})\, h_{n,3}(x) +$$

$$+\sum_{k=1}^{n} l_k\left(\frac{[nx]+1}{x}\right)(X(k) - X(k-1) - Fk(-1) + Fk-1(-1))\, hn,4(x), \tag{2}$$

де **[a]** – ціла частина числа **a, {a}** – дробна частина числа **a,** і за означенням

$F_i^{(-1)}=F^{(-1)}(y_i)= F^{(-1)}(i/n)$, $h_{n,1}(x)=2\{nx\}^3+1$, $h_{n,2}(x) = 3\{nx\}^2-2\{nx\}^3$, $h_{n,3}(x)= \{nx\}^3-2\{nx\}^2+\{nx\}$,

$h_{n,4}(x) =\{nx\}^3 - \{nx\}^2$.

Зі співвідношення (2) випливає, що при $x\epsilon[y_i, y_{i+1})$, $i=0,1,\ldots, n-1$:

$$(s_n(x) - s(x))^2 \le 6((X_{(i)} - F_i^{(-1)})^2 h^2_{n,1}(x) + (X_{(i+1)} - F_{i+1}^{(-1)})^2 h^2_{n,2}(x) + (\sum_{k=1}^{n} l_k(\frac{[nx]}{x})(X_{(k-1)} - F_{k-1}^{(-1)}))^2 h^2_{n,3}(x) +$$

$$+(\sum_{k=1}^{n} l_k(\frac{[nx]}{x})(X_{(k)} - F_k^{(-1)}))^2 h^2_{n,3}(x) + (\sum_{k=1}^{n} l_k(\frac{[nx]+1}{x})(X_{(k-1)} - F_{k-1}^{(-1)}))^2 h^2_{n,4}(x) +$$

$$+ (\sum_{k=1}^{n} l_k(\frac{[nx]+1}{x})(X_{(k)} - F_k^{(-1)}))^2 h^2_{n,4}(x)). \tag{3}$$

Розглянемо перший доданок суми (3):

$$(X_{(i)} - F_i^{(-1)})^2=(F^{(-1)}(F(X_{(i)})) - F_i^{(-1)})^2=(\frac{dF(x)^{(-1)}}{dx}|_{x=\Theta_i} \cdot (F(X_{(i)}) - \frac{i}{n}))^2=(F(X_{(i)}) - \frac{i}{n}))^2/f^2(F^{(-1)}(\Theta_i)) \le$$

$$\le \frac{1}{m_2} \cdot (F(X_{(i)}) - \frac{i}{n})^2, \text{ де } \Theta_i=\Theta(X_{(i)}) \in [0, 1]. \tag{4}$$

Згідно з [1], випадкова величина $F(X_{(i)})$, $i = 1, \ldots, n$, має бета-розподіл *Be(i, n − i + 1)*. Тому при $n \ge 5$, $i = 1, \ldots, n$.

$$M(F(X_{(i)}) - \frac{i}{n})^2 = MF^2(X_{(i)}) - \frac{2i}{n} MF(X_{(i)}) + \frac{i^2}{n^2} = \frac{i}{n+1} \cdot \frac{i+1}{n+2} - 2\frac{i}{n} \cdot \frac{i}{n+1} + \frac{I^2}{n^2} \le \frac{n^2}{4(n-1)(n+1)(n+2)} \le \frac{1}{4n}. \tag{5}$$

Зрозуміло, що при $i = 0$ $M(F(X_{(i)}) - \frac{i}{n})^2 = 0$. Зауважимо, що

$$M(\sum_{k=1}^{n} l_k(\frac{[nx]}{x})(X_{(k)} - F_k^{(-1)}))^2 = \sum_{k=1}^{n} l_k^2(\frac{[nx]}{x})M(X_{(k)} - F_k^{(-1)})^2 +$$

$$+ \sum_{u,v=1}^{n} l_u(\frac{[nx]}{x}) \cdot l_v(\frac{[nx]}{x})M((X_{(u)} - F_u^{(-1)})(X_{(v)} - F_v^{(-1)})).$$

За нерівністю Коші-Шварца

$$M((X_{(u)} - F_u^{(-1)})(X_{(v)} - F_v^{(-1)})) \le (M(X_{(u)} - F_u^{(-1)})^2)^{\frac{1}{2}}(M(X_{(v)} - F_v^{(-1)})^2)^{\frac{1}{2}}.$$

Крім того, в [4] показано, що $\forall\ n \in N$, $x \in [0, 1]$:

$$\sum_{k=1}^{n} |l_k(\frac{[nx]}{x})| \le C.$$

Тому, згідно (4) та (5)

$$M\left(\sum_{k=1}^{n} l_k(\frac{[nx]}{x})(X_{(k)}-F_k^{(-1)})\right)^2 \le \frac{C^2}{2nm^2}.$$

Нехай за означенням $K_i = \int_0^1 h_{n,i}^2(x)dx$. Тоді, з урахуванням (3) маємо таку оцінку:

$$M(\|s_n(x) - s(x)\|_2)^2 \le \frac{3}{nm^2}(1/2 + C^2)(K_1 + K_2).$$

Із врахуванням (1) отримаємо, що при виконанні умов теореми

$$M(s_n(\xi) - F^{(-1)}(\xi))^2 \le C(p)\left|\left|\frac{d^qF^{(-1)}(x)}{dx^q}\right|\right|_p^2 n^{-2q-1+\frac{2}{p}} + \frac{3}{m^2}(1/2+C^2)(K_1 + K_2)n^{-1},$$

звідки і випливає твердження теореми.

Список використаної літератури

1. Дейвид Г. Порядковые статистики / Г. Дейвид. – М. : Наука, 1979. – 240 с.
2. Турбин А. Ф. Асимптотические свойства S-статистик / А. Ф. Турбин, А. Г. Ханин // Докл. АН УССР. Сер. А. – 1987. – № 11. – С. 14–17.
3. Турбин А. Ф. Моделирование непрерывных случайных величин с помощью интерполяционных и сглаживающих сплайнов / А. Ф. Турбин, А. Г. Ханин // Асимптотические методы в вероятностных задачах. – Киев : Ин-т математики АН УССР, 1988. – С. 91–95.
4. Ханин А. Г. Сплайн-оценка интеграла от квадрата генеральной плотности и ее асимптотические свойства / А. Г. Ханин. – Киев : Ин-т математики АН УССР, 1989. – 51 с.
5. Schumaker L. Spline functions: basic theory / L. Schumaker. – New-York : Wiley, 1981. – 553 p.
6. Smith P. L. Splines As a Useful and Convenient Statistical Tool // The American Statistician. – 1979. – Vol. 33. – No. 2. – P. 57–62.
7. Wegman E. J. Splines in Statistics / E. J. Wegman, J.W. Wright // Journal of the American Statistical Association. – 1983. – Vol. 78. – No. 382. – P. 351–365.

Статтю подано до редколегії
10.10.2011 р.