
**ИССЛЕДОВАНИЯ, МОДЕЛИ, МЕТОДЫ
И МЕТОДИКИ ИЗМЕРЕНИЙ**

УДК 519.24 + 519.234.7

© А. Л. Буляница

УСЛОВИЯ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ МЕТОДА ПРОСТОГО ИНТЕРВАЛЬНОГО ОЦЕНИВАНИЯ

Метод простого интервального оценивания предполагает применение экстремальных порядковых статистик для оценивания математического ожидания случайной величины. Ранее было сформулировано, что необходимым условием большей эффективности этой оценки по сравнению с оценкой выборочного среднего является ограниченность помехи. В работе показано, что данное условие не является достаточным и должно быть дополнено требованием невырожденности помехи на границе диапазона.

ВВЕДЕНИЕ

Метод ПИО — простого интервального оценивания [1] — получил практическое развитие в области анализа больших массивов физико-химических данных [2]. Главным образом, его использование ориентировано на решение задач дискриминации или/и классификации.

Одним из его этапов является интервальное оценивание математического ожидания ограниченной промежутком $[-\beta, \beta]$ случайной величины с плотностью распределения вероятностей $f(x)$ как $[\max(x_i - \beta); \min(x_i + \beta)]$. Т. е. оценка использует экстремальные порядковые статистики. Необходимым условием ее эффективности является, по мнению авторов работ [2, 3], ограниченность погрешностей. В противном случае, как известно, при неограниченных погрешностях существуют наиболее эффективные оценки метода максимального правдоподобия, отличные от ПИО-оценок. Так, для случая нормальной (гауссовой) погрешности наиболее эффективна оценка выборочного среднего, для случая симметричной экспоненциальной (лапласовой) погрешности наибольшей эффективностью обладает медианная порядковая статистика.

Полагая ограниченность не только необходимым, но и достаточным условием эффективности ПИО-оценки, в работе [2] делается вывод: ширина доверительного интервала δ для математического ожидания на основе оценок выборочного среднего и медианы — $\delta = O(1/\sqrt{n})$, в то время как для ПИО-оценки (экстремальной порядковой статистики) — $\delta = O(1/n)$, и это гарантирует большую эффективность ПИО-оценок для объемов выборки, начиная с некоторого достаточно большого n .

В работе более развернуто обосновывается ранее сформулированная автором идея [4], что тре-

бование ограниченности помехи не является достаточным, и на помеху должно быть наложено еще одно нетривиальное ограничение.

АНАЛИЗ ВЕЛИЧИНЫ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ПИО-ОЦЕНКИ (ЭКСТРЕМАЛЬНОЙ ПОРЯДКОВОЙ СТАТИСТИКИ)

Без значимой потери общности полагаем ограниченное распределение симметричным с диапазоном $[-1, 1]$ и зададим доверительную вероятность $P = 1 - 2Q$.

В соответствии с известными формулами (бета-распределениями), плотность распределения вероятностей произвольной k -й порядковой статистики из выборки n измерений будет $f_n^k(x) = n C_{n-1}^{k-1} F^{k-1}(x) (1 - F(x))^{n-k} f(x)$. Соответственно для максимальной (n -й) статистики плотность распределения вероятностей будет $f_n^n(x) = n F^{n-1}(x) f(x)$, следовательно, функция распределения есть $F_n^n(x) = F^n(x)$, где $F(x)$ — функция распределения исходной величины (элемента выборки).

Минимальная граница доверительного интервала для n -й статистики определена условием: при $x = 1 - \delta/2$ функция распределения есть Q , т. е. $F_n^n(1 - \delta/2) = Q$. Аналогично, как следствие симметрии, максимальная граница нижней (первой) порядковой статистики определяется условием $F_n^1(-1 + \delta/2) = Q$.

Тем самым для смещенных на ± 1 порядковых статистик доверительный интервал, соответствующий вероятности $1 - 2Q$, является симметричным относительно нуля интервалом $[-\delta/2; \delta/2]$. Тогда граница доверительного интервала ПИО-оценки (n -порядковой статистики) определена ус-

ловием $F(1 - \delta/2) = Q^{1/n}$. Оценим δ при $n \rightarrow \infty$, представив функцию распределения $F(x)$ многочленом Тейлора 2-го порядка с $x_0 = 1$. Левая часть равенства примет вид: $F(1 - \delta/2) = 1 - f(1) \cdot \delta/2 + F''(1) \cdot \delta^2/8 + o(\delta^2)$. Т. к. при $n \rightarrow \infty$ $1 - Q^{1/n} \cong \ln(1/Q) \cdot 1/n = O(1/n)$ в соответствии с "замечательным" пределом, то $\delta = O(1/n)$ гарантируется только при $f(1) > 0$. В противном случае $\delta = O(1/\sqrt{n})$, и ПИО-оценка может иметь меньшую эффективность по сравнению с альтернативными оценками медианы и выборочного среднего.

Заметим, что в случае $f(1) = 0$ обязано выполняться $f'(1) = F''(1) < 0$, т. к. для точек внутри интервала $(-1; 1)$ плотность распределения вероятностей положительна. Тогда всегда возможно вычислить оценку доверительного интервала на основе параболического приближения.

Подтверждающим примером является распределение Симпсона: $f(x) = 1 - |x|$ при $|x| \leq 1$, т. е. $f(1) = 0$. Данная случайная величина обладает математическим ожиданием ноль (в силу симметричности и ограниченности) и дисперсией $1/6$. Очевидно, что дисперсия выборочного среднего будет в n раз меньше. Полагая квантиль распределения Стюдента $t_{Q,n}$, получим ширину доверительного интервала оценки выборочного среднего $\delta = 2t_{Q,n}\sqrt{1/6n}$, а ПИО-оценки — $\delta = \sqrt{-8\ln(Q)/n}$, т. е. возможно большую. Аналогичным свойством будут обладать косинусное распределение с параметром 1 и ряд других, вырожденных на границе диапазона.

Тем самым для гарантированного достижения большей эффективности ПИО-оценки, начиная с какого-либо, возможно, достаточно большого номера n , требуется удовлетворение еще одного нетривиального условия — невырожденность случайной величины на границе диапазона.

Следует заметить, что заведомо вырожденной на границе диапазона будет, например, разностная помеха при ограниченности исходной помехи. Как известно, соответствующая плотность распределения вероятностей разностной помехи определится как интегральная свертка двух одинаковых сдвинутых друг относительно друга плотностей распределения вероятностей $f(x)$. Аналитически с учетом ограниченности исходной помехи диапазоном $[-1; 1]$ ее плотность распределения вероятностей $p(x)$ определится как $p(x) =$

$$= \int_{x-1}^1 f(t)f(t-x)dt \text{ при } x \geq 0.$$
 Нетрудно показать, что при $x = 2$ пределы интегрирования совпадут, и, по определению, подобный интеграл есть ноль.

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

За основу возьмем стандартный датчик равномерно распределенной случайной величины $R[0; 1]$, реализуемый функцией RANDOM на языке C++. Этот датчик имеет свои недостатки, особенно в случае генерации коротких серий равномерных случайных величин. Однако его качество достаточно для дальнейших оценок, не требующих особо высокой точности.

На основе генерированной равномерной случайной величины ξ по известному методу формируется случайная величина, удовлетворяющая распределению Симпсона, по алгоритму:

$$s = \sqrt{2\xi} - 1; \quad 0 \leq \xi < 1/2$$

и

$$s = 1 - \sqrt{2(1 - \xi)}; \quad 1/2 \leq \xi \leq 1,$$

а также удовлетворяющая косинусному распределению с параметром 1, имеющая плотность распределения вероятностей $f(x) = \frac{\pi}{4} \cos(\pi x/2)$, $|x| \leq 1$ и функцию распределения $F(x) = (1 + \sin(\pi x/2))/2$, $|x| \leq 1$, формируемая по алгоритму $c = 2\arcsin(2\xi - 1)/\pi$.

Для расчетов доверительного интервала ПИО-оценки требуется знание $F''(1)$: при распределении Симпсона $F''(1) = -1$, при косинусном распределении $F''(1) = -\pi^2/8$. Далее использован написанный выше многочлен Тейлора 2-го порядка.

Для расчета доверительного интервала оценки выборочного среднего требуется знание коэффициента Стюдента $t = 0.676$ (более 120 степеней свободы), что практически совпадает с соответствующим квантилем нормального распределения $t = 0.675$ для $Q = 25\%$, а также дисперсию исходной случайной величины: для распределения Симпсона $1/6$, для косинусного распределения $1 - 8/\pi^2$.

Результаты имитационного моделирования представлены ниже. В соответствии с двумя законами распределения сгенерированы 400 серий по 400 измерений. В рамках каждой серии оценивается выборочное среднее значение, а также первый и последний члены вариационного ряда (экстремальные статистики). На основании измерений по всем сериям для каждой из указанных величин вычислен 50 %-й доверительный интервал ($Q = 25\%$) ПИО-оценки и оценки выборочного среднего.

В табл. 1 сведены результаты моделирования (названы "Модель") и теоретические расчеты ("Теория"). Данные таблицы отражают тенденцию: эффективность ПИО-оценки оказалась примерно в 5–6 раз хуже по сравнению с оценкой

Табл. 1. Доверительные интервалы оценок выборочного среднего и ПИО-оценок

Оценка		Закон	
		Симпсона	Косинусный
ПИО-оценка	Модель	0.1674	0.1507
	Теория	0.1665	0.1499
Оценка выборочного среднего	Модель	0.02591	0.02712
	Теория	0.02755	0.02938

Табл. 2. Отношение доверительных интервалов оценок ПИО и выборочного среднего для симпсоновской помехи

Q	$P, \%$	k
0.25	50	6.042
0.20	60	5.219
0.15	70	4.606
0.10	80	4.100
0.05	90	3.645
0.03	94	3.449
0.02	96	3.336
0.01	98	3.196
0.005	99	3.097
0.001	99.8	2.946

выборочного среднего. Однако выбранная доверительная вероятность не слишком велика. В табл. 2 для различных доверительных вероятностей P (величин Q) приведены расчетные значения относительной эффективности ПИО-оценки $k = \delta_{\text{ПИО}} / \delta_{\text{ср}}$ (отношение доверительных интервалов ПИО-оценки и оценки выборочного среднего). Данные относятся к распределению Симпсона при $k = \sqrt{\frac{8 \ln(1/Q)}{n}} / \left(2t \sqrt{\frac{1}{6n}} \right)$, или $k = \sqrt{\frac{12 \ln(1/Q)}{t^2}}$, однако для другого рассматриваемого случая — косинусного распределения — ситуация количественно схожая.

При расчете использованы данные о квантилях нормального распределения [5]. Наблюдается тенденция уменьшения выигрыша по эффективности оценки выборочного среднего по сравнению с ПИО-оценкой по мере роста доверительной веро-

ятности P . Далее определим асимптотический предел этого отношения по мере стремления $Q \rightarrow 0$ (или $P \rightarrow 100 \%$). Для решения указанной задачи достаточно рассмотреть асимптотическое отношение $\ln(1/Q)/t^2$. Взяв за основу аппроксимацию интеграла вероятностей, приведенную в работе [6], для больших t (малых Q), получим приближенную оценку в виде

$$Q = \frac{\exp(-t^2/2)}{t\sqrt{2\pi}} (1 + O(t^{-2})).$$

Переход к натуральному логарифму в правой и левой частях ведет к приближенному равенству

$$\ln(Q) \cong -t^2/2 - \ln(\sqrt{2\pi}t), \text{ или}$$

$$\frac{\ln(1/Q)}{t^2} \cong 1/2 + \frac{\ln(\sqrt{2\pi}t)}{t^2}.$$

Исследование предела при $t \rightarrow \infty$ (что соответствует $Q \rightarrow 0$) после использования правила Лопиталя дает оценку искомого отношения как $1/2$. Тем самым асимптотически величина k стремится к $\sqrt{6}$.

Аналогичными рассуждениями получим, что асимптотическая эффективность ПИО-оценки для косинусного распределения будет соответствовать

$$k = \sqrt{\frac{8}{\pi^2 - 8}} \cong 2.069.$$

Следовательно, в рассмотренных выше случаях ПИО-оценка по меньшей мере вдвое менее эффективна по сравнению с оценкой выборочного среднего.

ЗАКЛЮЧЕНИЕ

Сопоставление приведенных в таблицах данных позволяет сделать однозначный вывод о значимо большей эффективности оценки выборочного среднего по сравнению с ПИО-оценкой в указанных случаях ограниченных помех, вырожденных на границе диапазона — треугольной помехи Симпсона и помехи, удовлетворяющей косинусному распределению с параметром 1.

Таким образом, для гарантированного обеспечения большей эффективности ПИО-оценки по сравнению с оценкой выборочного среднего принцип ограниченности помехи должен быть дополнен условием ее невырожденности на границе диапазона.

СПИСОК ЛИТЕРАТУРЫ

1. Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке на-

блюдений // Сиб. мат. журн. 1962. Т. 3, № 5. С. 701–709.

2. Померанцев А.Л., Родионова О.Е. Построение многомерной градуировки методом ПИО // Журн. аналит. химии. 2006. Т. 61, № 10. С. 1032–1047.

3. Rajko R. Treatment of Model Error in Calibration by Robust and Fuzzy Procedures // Anal. Letters. 1994. V. 27. P. 215–228.

4. Буляница А.Л. Условия эффективности метода простого интервального оценивания // Доклады 8-го Всеросс. симп. по прикладной и промышленной математике, 07–10 окт. 2007 г., Адлер (в печати). Опубл. (www.tvp.ru/conferen/conferfr.ru).

5. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.

6. Анго А. Математика для электро- и радиоинженеров. М.: Наука, 1964. С. 360–362.

*Институт аналитического приборостроения РАН,
Санкт-Петербург*

Материал поступил в редакцию 28.03.2008.

EFFICIENCY USING CONDITIONS FOR SINGLE INTERVAL ESTIMATION METHOD

A. L. Bulyanitsa

Institute for Analytical Instrumentation RAS, Saint-Petersburg

Single interval estimation' method allow to use extremal range statistics for mathematical expectation' calculation of random value. Early was determine an infinite this value as necessary condition of more efficiency this estimation comparison with average-sampling estimation. In this work demonstrate, that this condition is not sufficient and it must be supple the requirement of undegeneration at the border range.