УДК 621.317.08

ОПРЕДЕЛЕНИЕ ЭФФЕКТИВНЫХ ОЦЕНОК ЦЕНТРА РАСПРЕДЕЛЕНИЯ ПРИ СТАТИСТИЧЕСКОЙ ОБРАБОТКЕ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

ЗАХАРОВ И.П., ШТЕФАН Н.В.

При статистической обработке результатов наблюдений очень важно для оценивания центра распределения использовать наиболее эффективную оценку, поскольку это позволяет более точно оценить другие параметры распределения и добиться той же точности вычислений при меньшем числе наблюдений. Предлагается метод определения эффективной оценки центра распределения в виде взвешенной суммы среднего арифметического, медианы и центра размаха.

При статистической обработке экспериментальных данных наиболее часто используются следующие оценки центра распределения $\hat{M}_x$:

– среднее арифметическое результатов наблюдений

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \qquad (1)$$

где n – число наблюдений; x_i – результат отдельного наблюдения;

– оценка медианы

$$Me = \begin{cases} x_{\frac{n+1}{2}}, & n - \text{нечетное}, \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & n - \text{четное}; \end{cases} \qquad (2)$$

– центр размаха

$$x_{cp} = \frac{x_{min} + x_{max}}{2} \,, \qquad (3)$$

здесь x_{min} – минимальный член выборки; x_{max} – максимальный член выборки.

Под наиболее эффективной понимается оценка, обладающая наименьшей дисперсией. Эффективность - одно из основных требований к оценке, поскольку определение всех последующих оценок базируется на оценке центра распределения и погрешность в его определении может повлечь за собой неправильную оценку среднего квадратичного отклонения (СКО), границ доверительного интервала, вида распределения и т.п. При этом использование более эффективных оценок позволяет оценить параметр по меньшему числу наблюдения при той же точности, что особенно актуально в настоящее время для построения процессорных средств измерений.

Наиболее часто за оценку центра распределения принимается среднее арифметическое результатов наблюдений. Однако [1-3] такая оценка наиболее эффективна только для нормального распределения. Она менее эффективна, чем центр размаха для ограниченных распределений (кроме треугольного), и менее эффективна, чем медиана для распределения Лапласа, т.е. эффективность оценок зависит от вида распределения. В литературе приводится методика, согласно которой эффективная оценка центра распределения выбирается в зависимости от эксцесса распределения, оценка которого вычисляется по формуле

$$\hat{E} = \frac{\sum_{i=1}^{n}\left(x_i - \hat{M}_x\right)^4}{n\hat{\sigma}_x^4} - 3 \,, \qquad (4)$$

где $\hat{\sigma}_x$ – оценка среднего квадратического отклонения результатов наблюдения.

Согласно этой методике для распределений со значениями эксцесса меньше –0,5 наиболее эффективной оценкой является центр размаха, со значениями эксцесса от –0,5 до 1 – среднее арифметическое результатов наблюдений, со значением эксцесса, большим 1 – медиана. Указанные выше границы выбраны приблизительно, поэтому они нуждаются в дополнительном исследовании.

В связи с этим возникает задача отыскания такой $\hat{M}_x$, которая была бы эффективной для любого вида распределения и изменяла свой вид в зависимости от формы распределения.

Нами предлагается для оценивания центра распределения оценка в виде взвешенной суммы всех трех оценок:

$$x_{вз} = k_1\overline{x} + k_2 x_{cp} + k_3 Me \,, \qquad (5)$$

где k_1, k_2 и k_3 – коэффициенты влияния, значения которых зависят от формы закона распределения. При этом для того, чтобы оценка (5) была несмещенной, необходимо выполнять условия $k_1 + k_2 + k_3 = 1$.

Эффективность оценки (4) была определена численным методом с помощью моделирования на ЭВМ. В его основу положен метод математического моделирования (метод Монте-Карло) [4], суть которого заключается в следующем:

1. Получаем генеральную совокупность случайных чисел объемом N, подчиняющуюся заданному распределению.

2. Из этой генеральной совокупности формируем выборки малого объема n. Для каждого распределения на ЭВМ формируем не менее 1000 выборок с объемом 3,5,7,...,21 наблюдений.

3. Для каждой малой выборки объемом n вычисляем оцениваемый параметр, в результате чего получаем выборку из значений этого параметра

объемом $m = \frac{N}{n}$ для заданного числа наблюдений n и заданного распределения.

4. По этой выборке рассчитываем оценки точечных характеристик – математического ожидания и дисперсии, т.е. задаваясь различными сочетаниями коэффициентов, находим значения оценки (5) и по ним вычисляем оценку их дисперсии. Определяем минимальное значение дисперсии оценки (5) и фиксируем соответствующее ему сочетание коэффициентов.

Для исследования были взяты распределения различных классов [1] с разными значениями эксцесса: арксинусоидальное (эксцесс -1,5), равномерное (эксцесс -1,2), трапецеидальные с отношением оснований 0,75 (эксцесс -1,15), 0,5(эксцесс -0,98), 0,25 (эксцесс -0,73),треугольное распределение (эксцесс -0,6), распределения типа шапо, полученные как композиция равномерного и нормального распределений (эксцессы -0,4 и -0,2), равномерного распределения и распределения Лапласа (эксцессы 0,4; 0,87; 1,32; 1,99;2,44), нормальное распределение (эксцесс 0), распределение Лапласа (эксцесс 3).

Для моделирования равномерного и нормального распределений был использован встроенный генератор случайных чисел Microsoft Mathcard 7.0. Треугольное, арксинусоидальное распределения и распределение Лапласа получены из равномерного распределения с границами (0;1) методом обратных функций. Трапецеидальные распределения получены как композиция двух равномерных распределений с различными границами. Распределения типа шапо получены как композиция равномерного и нормального распределений (получены распределения с отрицательными эксцессами) или равномерного и распределения Лапласа (получены распределения с положительными эксцессами).

Результаты вычислений приведены в табл. 1.

Зависимости коэффициентов от эксцесса приведены на рисунке.

Как видно из рисунка, на оценку центра распределения оказывают влияние не более двух оценок: для распределений с отрицательными эксцессами сказывается центр размаха и среднее арифметическое, а для распределений с положительным эксцессом – среднее арифметическое и медиана.

Таблица 1

E	k_1	k_2	k_3
-1,5	0	1	0
-1,2	0	1	0
-1,15	0	1	0
-0,99	0,2	0,8	0
-0,73	0,45	0,55	0
-0,6	0,6	0,4	0
-0,4	0,8	0,2	0
-0,2	1	0	0
0	1	0	0
0,39	1	0	0
0,87	1	0	0
1,32	0,8	0	0,2
1,99	0,45	0	0,55
2,44	0,25	0	0,75
3	0	0	1

Зависимости коэффициентов влияния от эксцесса могут быть аппроксимированы выражениями, приведенными в табл. 2 с погрешностью, не превышающей 2%.

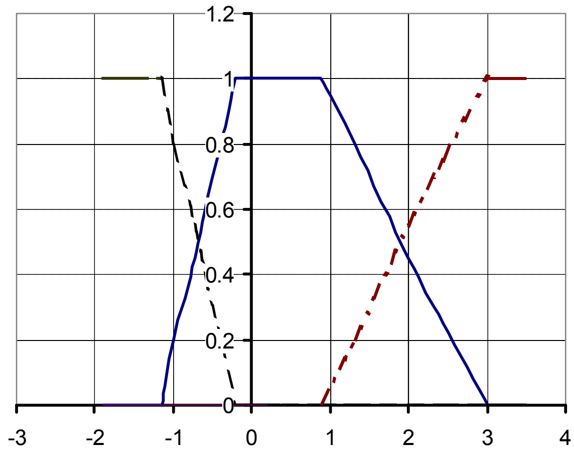

Таблица 2

Область изменения эксцесса	k_1	k_2	k_3
$E \le -1{,}15$	0	1	0
$-1{,}15 < E \le -0{,}2$	$1{,}05E + 1{,}22$	$-0{,}05E - 0{,}22$	0
$-0{,}2 < E \le 0{,}9$	1	0	0
$0{,}9 < E < 3$	$-0{,}47E + 1{,}41$		$0{,}47E - 0{,}41$
$E \ge 3$	0	0	1

Установлено, что для распределений, имеющих совершенно различную форму, но близкие значения эксцесса (например, для равномерного распределения с эксцессом –1,2 и распределения типа двойной шлем с эксцессом –1,18; трапецеидального распределения с эксцессом -0,73 и распределения типа русский шлем с эксцессом -0,75), дисперсия оценки (5) достигала минимума при одинаковых сочетаниях коэффициентов. Это говорит о том, что есть возможность определять значения коэффициентов по характеристике формы распределения - эксцессу и предложить следующую методику оценки центра распределения:

1. Определяется оценка эксцесса распределения.

2. По рисунку или с помощью табл. 2 рассчитываются значения коэффициентов.

3. По формуле (5) определяется эффективная оценка центра распределения.

Сравним эффективность предложенной оценки с остальными для распределений с различными эксцессами. Для этого определим величину, называемую сравнительной эффективностью, представляющую собой отношение дисперсий предложенной средневзвешенной оценки к дисперсиям других оценок. Результаты этих вычислений приведены в табл. 3.

Как видно из табл. 3, оценка центра в виде взвешенной суммы более эффективна, чем центр размаха для распределений с эксцессом, меньшим -0,5, и более эффективна, чем медиана для распределений с эксцессом, большим 1.

Таблица 3

E	$\frac{D[x_{вз}]}{D[\bar{x}]}$	$\frac{D[x_{вз}]}{D[x_{ср}]}$	$\frac{D[x_{вз}]}{D[Me]}$
-1,5	0,02	1	0,005
-1,2	0,2	1	0,08
-1,15	0,4	1	0,15
-0,99	0,63	0,85	0,25
-0,73	0,81	0,75	0,4
-0,6	0,86	0,67	0,54
-0,4	0,97	0,5	0,55
-0,2	1	0,4	0,6
0	1	0,31	0,64
0,39	1	0,25	0,65
0,87	1	0,2	0,67
1,32	0,96	0,18	0,8
1,99	0,9	0,12	0,85
2,44	0,8	0,1	0,93
3	0,67	0,07	1

Таким образом, предложенную средневзвешенную оценку можно использовать в качестве эффективной оценки центра распределения.

Литература: 1. *Новицкий П. В., Зограф И. А.* Оценка погрешностей результатов измерений. Л.:Энергоатомиздат. Ленингр. отд-ние, 1991. 304 с. 2.*Крамер Г.* Математические методы статистики. М.: Мир, 1976. 648 с. 3. *Закс Л.* Статистическое оценивание. М.: Статистика, 1976. 598 с. 3. *Дунин-Барковский И.В., Смирнов Н.В.* Теория вероятностей и математическая статистика в технике (общая часть). М.: Государственное издательство технико-теоретической литературы, 1955. 557 с. 4. *Ермаков С. М., Михайлов Г. А.* Курс статистического моделирования. М.: Наука, 1976. 472с.

Поступила в редколлегию 16.03.2002

Рецензент: д-р физ.-мат. наук, проф. Яковлев С.В.

Захаров Игорь Петрович, канд. техн. наук, доцент кафедры МИТ ХНУРЭ. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 40-93-31, 27-60-17.

Штефан Наталья Владимировна, младший научный сотрудник кафедры МИТ ХНУРЭ. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 40-93-31, 27-60-17.