УДК 004.94

ИНТЕРВАЛЬНАЯ АРИФМЕТИКА: ЭМПИРИКО-СТАТИСТИЧЕСКИЙ ПОДХОД К ОЦЕНКЕ РЕЗУЛЬТАТОВ ДЕЙСТВИЙ[1]

Петрушин В.Н., Никульчев Е.В.

Всероссийская государственная налоговая академия
Министерства финансов Российской Федерации
e-mail: PetrushinVN@mail.ru, nikulchev@vgna.ru

Аннотация

В докладе приведена методика поиска минимально возможного интервала оценки значений случайной величины с заданной надежностью. Рассмотрены вопросы построения гистограммы на основании выборки. Проведена интервальная оценка числовых характеристик без гипотезы о виде распределения. Приведены результаты интервальной арифметики с применением выборочных распределений.

Приведенные примеры оценок величин, их средних, результатов арифметических действий свидетельствуют о том, что эмпирический подход позволяет делать более точные вероятностные интервальные оценки для любого наблюдаемого распределения без его аналитической аппроксимации и связанных с ней погрешностями. Удалось совместить независимость классического интервального анализа от вида распределения случайной величины внутри сегмента значения и возможность выбора наиболее вероятных значений без оценки границ генеральной совокупности.

Введение. В процессе измерений исследователи крайне редко имеют с точными значениями. В силу погрешности приборов, невозможности количественно определить воздействие всех факторов на изучаемое явление или процесс наблюдаемая величина является случайной, и, как правило, ограниченной на каком-либо сегменте или интервале. Использование сегмента предполагает знание точных границ измеряемой величины, интервал этого не требует, но, к сожалению, искусственно расширяет количественные показатели включением невозможных значений.

Если ставится условие получить решение какой-либо расчетной задачи на базе экспериментальных данных с вероятностью $P = 1$ целесообразно применять интервальный анализ [1, 2, 3], достоинством которого является полное безразличие к виду функции распределения случайной величины. Но у этого вида оценок два существенных недостатка: при применении интервальной арифметики и алгебры происходит значительное расширение интервалов, и такие расчеты не пригодны для конструкторов и технологов; вторым существенным недостатком является отсутствие знания точных границ вариации случайной величины, что приводит, как указывалось выше, к появлению в косвенных измерениях конечного для дискретных и бесконечного для непрерывных вычисляемых величин числа невозможных решений.

В теории эксперимента отказываются от достоверности расчетов с вероятностью $P = 1$, исследователь самостоятельно определяет необходимую надежность результата. Такой подход требует знания функции плотности вероятности случайной величины или ее закона распределения. Чаще всего, традиционно, предполагается нормальное распределение и распределения, получаемые на базе нормального при нелинейных преобразованиях [4, 5]. При учете ограниченности диапазона возможных значений,

[1] Работа выполнена при финансовой поддержке РФФИ (грант № 11-07-00772-а).

используются другие распределения, например, бета-распределение [6, 7]. Интервальная оценка рассчитывается по принятому распределению с выбранной вероятностью. Традиционно для интервального анализа и теории эксперимента центром оценки выбирается медиана, которая в случае симметричных относительно нее по вероятности распределений совпадает с математическим ожиданием [8]. То есть из бесконечного для непрерывных и конечного для дискретных множество значений случайной величины выбирается далеко не оптимальное решение для ассиметричных относительно медианы распределения. Достаточно часто совершается вероятностная ошибка: надежностью результата провозглашается выбранная вероятность попадания в интервал, но при этом забывают, что теоретическое распределения принято тоже с какой-то вероятностью, поэтому надежность

$$\gamma = P(A) \cdot P_A(B),$$

где $P(A)$ — вероятность оправдания гипотезы о виде распределения; $P_A(B)$ — вероятность попадания в интервал наблюдаемой случайной величины при условии выполнении гипотезы. Гипотеза о виде распределения связана с расчетами параметров распределений по выборочным данным методом моментов или методом наибольшего правдоподобия, и несет за собой все погрешности этих расчетов.

Еще сложнее обстоит дело с оценками динамических процессов, нельзя гарантированно утверждать, что серия измерений проводится мгновенно, т. е. в одном пространственно-временном сечении. Динамика процесса влечет за собой динамику числовых характеристик случайной величины и динамику ее распределения. В силу приведенных причин у исследователя весьма не простой выбор между интервальным анализом и вероятностными методами расчетов. Авторы поставили себе цель совместить преимущества каждого из подходов при обработке экспериментальных данных.

1. Поиск минимально возможного интервала оценки значений случайной величины с заданной надежностью.

Предположим, что имеется ограниченная непрерывная случайная величина X с плотностью распределения вероятностей $f(x)$ или дискретная величина с законом распределения $P\ (X)$, которая аппроксимируется функцией плотности вероятностей $f(x)$. Задача состоит в поиске наиболее вероятных значений случайной величины с выбранной надежностью результата γ.

Рассмотрим сначала унимодальный случай, пусть все возможные значения $x \in [a,b]$. Найдем наиболее узкий интервал решений $(x_1, x_2) \subset [a,b]$, соответствующий выбранной надежности γ. То есть решаем задачу поиска условного экстремума функции $y = x_2 - x_1$ при условии $\int_{x_1}^{x_2} f(x)dx = \gamma$. Воспользуемся методом множителей Лагранжа, исследуем на минимум функцию $z = x_2 - x_1 + \lambda \left(\int_{x_1}^{x_2} f(x)dx - \gamma \right)$. Частные производные функции $z(x_1, x_2, \lambda)$:

$$\frac{\partial z}{\partial x_1} = -1 - \lambda f(x_1),$$

$$\frac{\partial z}{\partial x_2} = 1 + \lambda f(x_2),$$

$$\frac{\partial z}{\partial \lambda} = \int_{x_1}^{x_2} f(x)dx - \gamma.$$

Составим систему уравнений, используя необходимые условия экстремума функции нескольких переменных

$$\begin{cases} -1-\lambda f(x_1)=0, \\ 1+\lambda f(x_2)=0, \\ \int\limits_{x_1}^{x_2} f(x)dx-\gamma=0. \end{cases}$$

Решая систему уравнений, получаем

$$f(x_1)=-\frac{1}{\lambda}, \qquad f(x_2)=-\frac{1}{\lambda},$$

т. е., $f(x_1)=f(x_2)$, т. к. $f(x)>0$, то $\lambda<0$.

Отметим, что наиболее вероятные интервалы значений случайной величины соответствуют большим значениям *f*(*x*), следовательно, интервал (x_1,x_2) включает в себя моду, причем x_1 находится слева от нее, а x_2 — справа. Заметим, что в рассматриваемом случае производная $f'(x)$ положительна в точке x_1, и отрицательна в точке x_2. Исследуем знак квадратичной формы, для чего найдем вторые производные функции $z(x_1,x_2,\lambda)$:

$$\frac{\partial^2 z}{\partial x_1^2}=-\lambda f'(x_1), \qquad \frac{\partial^2 z}{\partial x_1 \partial x_2}=0, \qquad \frac{\partial^2 z}{\partial x_1 \partial \lambda}=-f(x_1),$$

$$\frac{\partial^2 z}{\partial x_2 \partial x_1}=0, \qquad \frac{\partial^2 z}{\partial x_2^2}=\lambda f'(x_2), \qquad \frac{\partial^2 z}{\partial x_2 \partial \lambda}=f(x_2),$$

$$\frac{\partial^2 z}{\partial \lambda \partial x_1}=-f(x_1), \qquad \frac{\partial^2 z}{\partial \lambda \partial x_2}=f(x_2)\,\frac{\partial^2 z}{\partial \lambda^2}=0.$$

Составим определитель из производных второго порядка:

$$\Delta=\begin{vmatrix} -\lambda f'(x_1), & 0, & -f(x_1), \\ 0, & \lambda f'(x_2), & f(x_2), \\ -f(x_1), & f(x_2) & 0. \end{vmatrix}=\lambda\left(f'(x_1)f^2(x_2)-f'(x_2)f^2(x_1)\right).$$

Так как $f'(x_1)>0,\ f'(x_2)<0$, то $\Delta<0$. Выполнено условия минимума. Для окончательного решения необходимо найти x_1,x_2 с помощью системы уравнений:

$$\begin{cases} f(x_1)=f(x_2), \\ \int\limits_{x_1}^{x_2} f(x)dx=\gamma. \end{cases}$$

Обращаем внимание на то, что в интервале (x_1,x_2) находится наиболее вероятное решение. Традиционно ищут аппроксимирующую функцию *f*(*x*) и далее исследуют аналитический или численный поиск решений. Как уже отмечалось, аппроксимация сопряжена со снижением надежности результата, либо при ее сохранении — увеличении длины интервала (x_1,x_2). Можно этого избежать, анализируя репрезентативную выборочную совокупность. Если надежность выбираемого интервала γ, то необходимо удалить из ранжированного вариационного ряда наименее вероятные по показателям частоты *k* вариант в начале и конце ряда. Значение *k* определяется из соотношения

$$1-\gamma=\frac{k}{n},$$

$$k=n(1-\gamma),$$

где *n* — объем выборки; $k\in\mathbf{N}$.

При проведении алгебраических и арифметических действий возникают новые выборки, и, соответственно, новые распределения. Поэтому, правила классического интервального анализа не обязаны выполняться при вероятностях результат ниже единицы.

Рассмотрим U-образное распределение. Отметим, что для него $f'(x_1) < 0,\ f'(x_2) > 0$, решая аналогичную приведенной выше, задачу поиска условного экстремума, находим $\max z = x_2 - x_1$, соответствующий условию

$$\int_{x_1}^{x_2} f(x)dx = 1-\gamma.$$

Тогда объединение полусегментов $[a, x_1) \bigcup (x_2, b]$ будет иметь наименьшую длину. Значения x_1, x_2 находим с помощью системы уравнений:

$$\begin{cases} f(x_1) = f(x_2), \\ \int_a^{x_1} f(x)dx + \int_{x_2}^{b} f(x)dx = \gamma. \end{cases}$$

Суммарная длина двух интервалов будет наименьшей и содержит наиболее вероятные значения *x*. Применение изложенного выше подхода состоит в удалении наименее вероятных *k* вариант вариационного ряда.

В случае, когда плотность распределения представляет собой монотонную функцию, то, очевидно, интервал отсчитывается от границы, соответствующей наибольшему значению этой функции.

Если плотность распределения являются многомодальной, то из области возможных значений случайной величины необходимо численными методами выбрать наиболее вероятные интервалы значения *X* и объединить их. На границе этих интервалов внутри [*a*, *b*], будут выполняться условия $f(x_1) = f(x_2) = ... = f(x_n)$. Сумма вероятностей попадания наблюдаемой величины в полученное объединение интервалов равняется надежности γ.

2. Построение гистограммы на основании выборки. На предлагаемый способ получения вероятностных интервальных решений существенное влияние может оказать способ группировки наблюдаемых значений. Число возможных полусегментов *k* при построении гистограммы может варьироваться в пределах, определяемых неравенствами [6]:

$$\begin{cases} 1+\log_2 n < k < \dfrac{n}{1+\left(t(\gamma; n_j)\right)^2}, \\ n_j > 1+\left(t(\gamma; n_j)\right)^2, \end{cases}$$

где *n* — объем выборки, γ — надежность оценки средней групповой в каждом полусегменте; n_j — число значений случайной величины в *j*-м полусегменте; $t(\gamma; n_j)$ — значение критерия Стьюдента.

Приведем пример: если $n = 100$, $\gamma = 0{,}95$, то $k \in (7{,}10)$, т. е. гистограмма строится по 8–9 полусегментам. Предложный способ получения вероятностных интервальных решений не зависит от выбора аппроксимирующей функции и избавлен от численных ошибок приближенной аппроксимации, принимая эмпирику без всяких выравниваний и сглаживания.

3. Интервальная оценка числовых характеристик без гипотезы о виде распределения.

В случае вероятностной интервальной оценки математического ожидания ее тоже можно сделать на основе экспериментальных данных, не прибегая при этом к традиционно применяемому распределению Стьюдента, влекущему за собой значительные погрешности [5]. Точной оценкой математического ожидания является δ-

фунция Дирака. При любых конечных объемам выборки приходим к интервальной оценке, т. к. точечная оценка имеет бесконечно малую вероятность. Длина интервала зависит от выбранной надежности результата и объема результатов наблюдений. При наличии репрезентативной выборки можно избежать аппроксимации эмпирической функции распределения вероятностей и связанных вместе с этим действием погрешности.

В процессе усреднения наблюдаемых значений происходит сглаживание распределения с ростом объемов усредняемых значений. Вероятности «хвостов» при этом достаточно быстро уменьшаются, стремясь в пределе к нулю. Процесс усреднения аналогичен повторной выборке.

Пусть имеется некоторая выборочная совокупность объема n, представим ее в виде таблицы:

$$\begin{matrix} x_i & x_1 & x_2 & \dots & x_k \\ n_i & n_1 & n_2 & \dots & n_k \end{matrix}$$

Среднее значение $\bar{X} = \sum_{i=1}^{k} x_i n_i \Big/ \sum_{i=1}^{k} n_i$, где $\sum_{i=1}^{k} n_i = n$.

Наша задача состоит в поиске интервала, дающего оценку средней с надежности γ. Пусть $\bar{X}_m$ — результат усреднения по m вариантам ранжированного вариационного ряда, причем $m < n$; $\bar{X}_r$ — наблюдаемая средняя по r измерениям, $r < n$. Каждая из этих величин имеет наибольшее и наименьшее значение для любой конкретной выборки. Для данной задачи важны $\min \bar{X}_m = \sum_{i=1}^{m} x_i n_i \Big/ \sum_{i=1}^{m} n_i$ и $\max \bar{X}_r = \sum_{i=r}^{k} x_i n_i \Big/ \sum_{i=r}^{k} n_i$, несложно доказать, что $\min \bar{X}_m < \bar{X} < \max \bar{X}_r$. Найдем вероятность этого неравенства, используя эмпирическое распределение:

$$P\left(\min \bar{X}_m < \bar{X} < \max \bar{X}_r\right) = 1 - P\left(\bar{X} \le \min \bar{X}_m\right) - P\left(\bar{X} \ge \max \bar{X}_r\right) = 1 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i\right)^n - \left(\frac{1}{n}\sum_{i=r}^{k} n_i\right)^n.$$

Вероятность этого неравенства равна надежности оценки γ, получаем уравнение:

$$1 - \frac{1}{n^n}\left(\sum_{i=1}^{m} n_i\right)^n - \frac{1}{n^n}\left(\sum_{i=r}^{k} n_i\right)^n = \gamma,$$

оно имеет не единственное решение, выбираем то, которое дает максимальный размах интервальной оценки. Поиск решения можно осуществить с помощью переборного алгоритма. Выбор в пользу наращивания m или r осуществляется по наименьшему показателю суммы частот. Следовательно, при $m = r = n/2$, получаем эмпирическое $\gamma = 1 - 2^{-(n-1)}$. Полученное значение γ даже при небольших выборках очень велико, и будет давать достаточно широкие диапазоны оценок, но достоинство их состоит в их высокой надежности. Таким образом, и для вероятностной оценки математического ожидания можно обойтись без аппроксимации и связанных с этим ошибок. Эмпирический подход позволяет оценить вероятность попадания случайной величины в интересующий исследователя интервал или интервалы. Это касается и числовых характеристик распределения. Интересна достаточная оценка выборочной дисперсии $D_{\text{в}}$:

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Рассмотрим два значения случайной величины x_i и x_j. Дисперсия этой пары

$$D_{ij} = \frac{1}{2}\left(\left(x_i - \frac{x_i + x_j}{2}\right)^2 + \left(x_j - \frac{x_i + x_j}{2}\right)^2\right) = \frac{\left(x_i - x_j\right)^2}{4},$$

$\max D_{ij}$ возникает в случае $|x_i - x_j| = R$, где R — размах выборки. Таким образом, $\max D_\text{В}$ соответствует маловероятному варианту максимума каждой пары слагаемых, т. е. половина вариационного ряда — левая граница выборочной совокупности, а вторая половина — правая граница. Отсюда

$$\max D_\text{В} = \frac{R^2}{4}.$$

Вероятность такого значения дисперсии вычислим, исходя из повторности испытаний:

$$P\left(D_\text{В} = \frac{R^2}{4}\right) = \frac{n!}{\left((n/2)!\right)^2}\left(\frac{n_1 n_k}{n}\right)^{n/2},$$

объем выборки n — должен быть числом четным, что в любом эксперименте не является сложной задачей для исследователя. Таким образом, дисперсии

$$0 \le D \le \frac{R^2}{4}.$$

Вероятность равенства дисперсии нулю

$$P(D=0) = \left(\frac{n_1}{n}\right)^n + \left(\frac{n_2}{n}\right)^n + \ldots + \left(\frac{n_k}{n}\right)^n = \frac{(n_1)^n + (n_2)^n + \ldots + (n_k)^n}{n^n}.$$

Дисперсия средней $D(\bar{x})$ будет подчиняться неравенству

$$0 \le D(\bar{x}) \le \frac{R^2}{4n}.$$

Эмпирический подход даёт возможность оценить объём повторной выборки, необходимый для фиксированной интервальной оценки среднего значения случайной величины. Пусть длина интервала оценки Δ, вероятность попадания в интервал γ, объём повторной выборки l. Воспользуемся вероятностным уравнением, приведённым выше.

$$1 - \frac{1}{n^l}\left(\sum_{i=1}^{m} n_i\right)^l - \frac{1}{n^l}\left(\sum_{i=r}^{k} n_i\right)^l = \gamma,$$

где $\sum_{i=1}^{m} n_i$ — сумма частот для всех $x < \bar{x} - \Delta/2$, $\sum_{i=r}^{k} n_i$ — сумма частот для всех $x > \bar{x} + \Delta/2$.

Учитывая симметризацию плотности распределения $\bar{x}$ и равенство значений $f(\bar{x} - \Delta/2)$ и $f(\bar{x} + \Delta/2)$, получаем $\sum_{i=1}^{m} n_i = \sum_{i=r}^{k} n_i$, тогда

$$1 - 2\left(\frac{1}{n}\sum_{i=1}^{m} n_i\right)^l = \gamma.$$

Логарифмируя, находим l

$$l = \frac{\ln\left(\frac{1-\gamma}{2}\right)}{\ln\left(\frac{1}{n}\sum_{i=1}^{m} n_i\right)}.$$

Предложенный подход пригоден для любого наблюдаемого распределения при поиске интервальных решений с любой надежностью.

4. Интервальная арифметика с применением выборочных распределений. Проиллюстрируем полученные результаты на примере двух выборок, приведенных в таблицах 1, 2.

Таблица 1.

Таблица 2.

x_i	n_i
2	1
3	5
4	10
5	8
6	6
7	4
8	2
9	1
10	1

y_j	k_j
3	1
4	3
5	10
6	8
7	7
8	5
9	4
10	3
11	2
12	1

Критерий Пирсона не отвергает гипотезу о нормальном распределении обеих выборочных совокупностей:

$$\chi^2_{\text{набл.}x} = 0,15; \qquad \chi^2_{\text{кр.}x}(0,95;38) = 1,635,$$

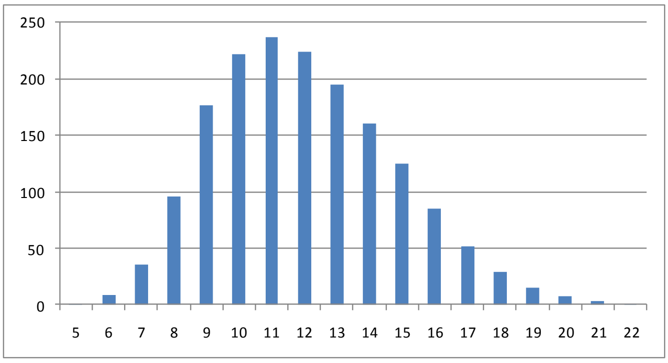
$$\chi^2_{\text{набл.}y} = 0,137; \qquad \chi^2_{\text{кр.}y}(0,95;44) = 2,167.$$

С вероятностью единица $x \in [2,10]$, $y \in [3,12]$, но, исходя из нормально распределения вероятность попадания в первый интервал 0,7234; во второй — 0,689. Убедительное доказательство того, что неограниченные по значению случайной величины распределения в данном случае не пригодны. Если взять какое-либо распределение ограниченной случайно величины, то необходимо расширять интервалы до нулевых значений плотности вероятностей, таким образом, наблюдаемые величины попадут в указанные интервалы с вероятностью меньше единицы. Средние значения $\bar{x} = 5,158$, $\bar{y} = 6,864$, с вероятностью единица они принадлежат своим интервалам. С вероятностью 0,99999 исходя из гипотезы нормального распределения средние принадлежат интервалам $\bar{x} \in (4,018; 6,306)$, $\bar{y} \in (5,638; 7,090)$. На основании эмпирической оценки $\bar{x} \in (5,079; 6,526)$ с вероятностью 0,999999999993; $\bar{y} \in (5,136; 8,588)$ с вероятностью 0,9999999999999. Если поставить задачу оценки интервалов исходя из нормального распределения с эмпирическими вероятностями, то они перекроют весь диапазон значений случайных величин. Следовательно, экспериментатору целесообразнее доверять эмпирической оценке. В случае не столь высоких требований по вероятности, например $\gamma = 0,99$, оценка по нормальному распределению дает интервалы $(4,446; 5,870)$, $(5,816; 7,912)$, оценка по эмпирическому распределению, соответственно, с той же надежностью (4,667;5,515) и (6,101;7,576). Исходя из нормального распределения, средние значения наблюдаемых выборочных совокупностей при выбранной вероятности 0,99 не различаются, тогда как эмпирический подход говорит об обратном.

Полученные эмпирические интервалы значений $\bar{x}$ и $\bar{y}$ убедительно свидетельствуют в пользу эмпирики.

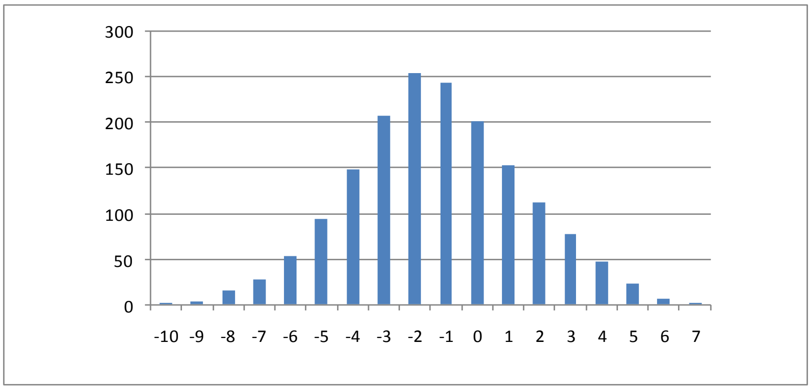
Займемся арифметикой: $x + y \in [5,22]$, $x - y \in [-10,7]$, $x \cdot y \in [6,120]$, $y/x \in [0.3, 6]$, согласно аксиоматике классической интервальной арифметики [1].

Приведем в таблицах эмпирические распределения суммы, разности, произведения и частного:



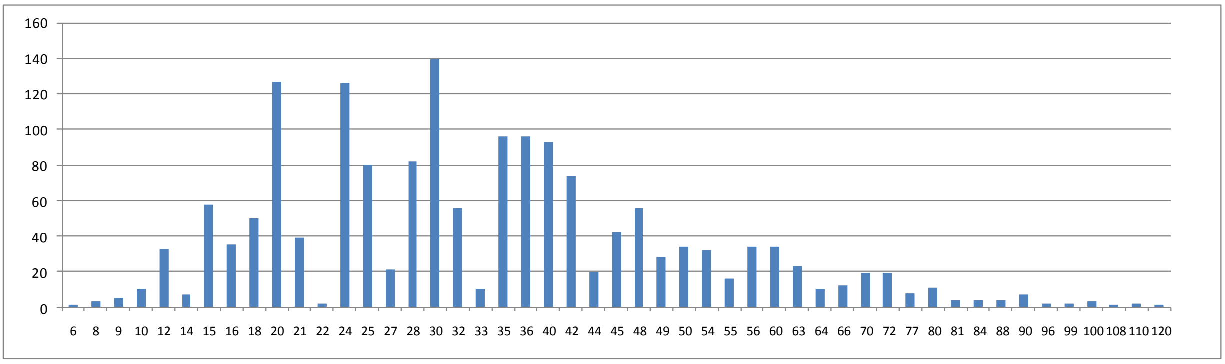
Сумма

значение	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
частота	1	8	35	96	177	222	237	224	195	160	125	85	52	29	15	7	3	1



Разность

значение	–10	–9	–8	–7	–6	–5	–4	–3	–2	–1	0	1	2	3	4	5	6	7
частота	1	4	15	28	53	94	149	208	255	244	201	153	112	77	47	23	7	1

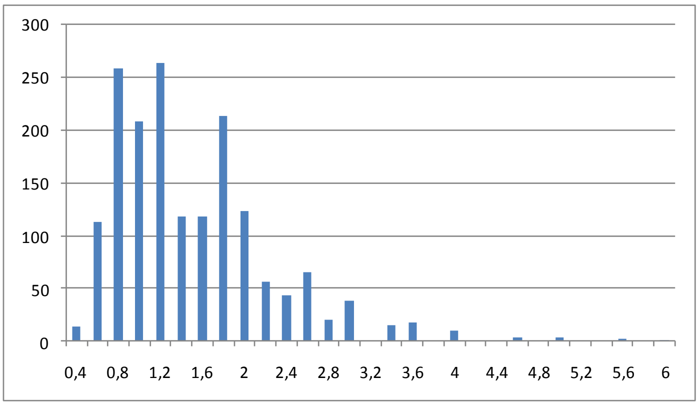


Произведение

значение	6	8	9	10	12	14	15	16	18	20	21	22	24	25	27	28	30	32
частота	1	3	5	10	33	7	58	35	50	127	39	2	126	80	21	82	140	56

значение	33	35	36	40	42	44	45	48	49	50	54	55	56	60	63	64	66	70
частота	10	96	96	93	74	20	42	56	28	34	32	16	34	34	23	10	12	19

значение	72	77	80	81	84	88	90	96	99	100	108	110	120
частота	19	8	11	4	4	4	7	2	2	3	1	2	1



Частное

значение	0,3	0,333	0,375	0,4	0,429	0,444	0,5	0,556	0,571	0,6	0,625	0,667
частота	1	1	2	3	4	3	22	10	12	16	20	26

значение	0,7	0,714	0,75	0,778	0,8	0,833	0,857	0,875	0,889	0,9	1	1,1
частота	7	40	26	7	29	60	32	55	5	4	208	2

значение	1,111	1,125	1,143	1,167	1,2	1,222	1,25	1,286	1,333	1,375	1,4	1,429
частота	3	8	20	42	65	2	106	16	46	4	56	12

значение	1,5	1,571	1,6	1,667	1,714	1,75	1,8	1,833	2	2,2	2,25	2,333
частота	107	8	40	68	4	70	32	12	123	16	40	35

значение	2,4	2,5	2,667	2,75	3	3,333	3,5	3,667	4	4,5	5	5,5	6
частота	8	40	25	20	38	15	7	10	10	4	3	2	1

Проверка критерием Пирсона показала, что гипотеза о нормальном распределении не отвергается для всех результатов арифметических операций. Отметим, что произведение и частное двух унимодальных случайных величин не является унимодальными.

В предположении нормального распределения с надежностью $\gamma = 0,99$ сумма $x + y \in (2.9421; 21.0999)$, полученный интервал выход левой границей за возможные значения суммы. Эмпирическая оценка с той же надежностью предлагает более узкий сегмент — [7; 19]. Еще более существенна эта разница при вероятностной оценки средних значений: в предположении нормального распределения среднего значения суммы с вероятностью 0,99 — $\overline{x + y} \in (11,8; 12,244)$, эмпирическая оценка этой средней с той же вероятностью более чем в двадцать раз уже — (11,983; 12,04).

В тех же предположениях и с той же надежностью $\gamma = 0,99$ разность $x - y \in (-10.372;\ 7.7843)$, либо эмпирически $x - y \in [-8;\ 5]$, $\overline{x - y}$, соответственно, (–1,516; –1,072) и (–1,317; –1,271). Вероятностные интервалы и для разности убедительно свидетельствуют в пользу эмпирического подхода.

Произведение $xy \in (-19.7674; 90.5714)$ в предположении нормального распределения с надежностью $\gamma = 0,99$, но интервал дает даже отрицательные результаты, которых, судя по выборке, не должно быть. Отсутствие унимодальности распределения произведения на основе эмпирического подхода приводит к оценке возможных его значений с выбранной надежностью к объединению сегментов $[9; 21] \cup [24; 90]$. Среднее значение произведения с той же вероятностью на основе гипотезы о нормальном распределении оценивается интервалом $\overline{xy} \in (34,053;\ 36,751)$, эмпирически — (35,179; 35,484). Эмпирический подход в случае произведения гораздо более оправдан, он выбирает наиболее надежное из возможных значений. Можно, конечно, представить распределение произведения как вероятностную смесь [6], но и это добавит погрешностей оценок по сравнению с эмпирикой.

Частное y/x на основе гипотезы о нормальном распределении должно с вероятностью $\gamma = 0,99$ находиться в интервале $y/x \in (-3,268; 5,3120)$. То есть опять возникает возможность отрицательных значений, которых нет. Отсутствие унимодальности распределения частного приводит в случае эмпирического подхода к весьма интересному решению — $y/x \in (0,42; 0,43) \bigcup [0,5;1] \bigcup [1,125,12] \bigcup [1,25; 4,5]$. Вероятности попадания частного в каждый промежуток разные, а сумма этих вероятностей равна 0,99. Интервальная оценка среднего значения частного с надежностью $\gamma = 0,99$ на основе гипотезу о нормальном распределении — (1,439; 1,557), эмпирический подход дает интервал — (1,487; 1,502).

Выводы. Приведенные примеры оценок величин, их средних, результатов арифметических действий свидетельствуют о том, что эмпирический подход позволяет делать более точные вероятностные интервальные оценки для любого наблюдаемого распределения без его аналитической аппроксимации и связанных с ней погрешностями. Удалось совместить независимость классического интервального анализа от вида распределения случайной величины внутри сегмента значения и возможность выбора наиболее вероятных значений без оценки границ генеральной совокупности.

Литература

1. Шокин Ю.И. Интервальный анализ / Ю.И. Шокин**.** – Новосибирск: Сибирское отделение изд-ва Наука, 1981.
2. Шарый С.П. Конечномерный интервальный анализ / С.П. Шарый [Электронный документ], режим доступа: http://www-sbras.nsc.ru/interval/Library/InteBooks/SharyBook.pdf.
3. Alefeld G. Interval analysis: theory and applications / G. Alefeld, G. Mayer // Journal of Computational Applied Mathematics. – 2000. – Vol. 121. – P. 421–464.
4. Смирнов Н. В. Курс теории вероятностей и математической статистики для технических приложений / Н. В. Смирнов, И. В. Дунин-Барковский— 3-е, стереотипное.— М.: Физматлит, 1969.— 512 с.
5. Краснов М. Л. Вся высшая математика. Т. 5. Теория вероятностей и математическая статистика / М. Л. Краснов, А. И. Киселев, Г. И. Макаренко, Е. В. Шикин, В. И. Заляпин, С. К. Соболев.— М.: Эдиториал УРСС, 2001. - 296 с.
6. Петрушин В.Н. Информационная чувствительность компьютерных алгоритмов / В.Н. Петрушин, М.В. Ульянов. — М.: Физматлит, 2010. — 224 с.
7. Гапонова Е.А. Модифицированный метод внутреннего оценивания множества решений интервальных систем линейных алгебраических уравнений / Е.А. Гапонова // Автореферат дис. .. к.т.н.: 05.13.17.— М.: МГУП, 2008.
8. Гмурман В.Е. Теория вероятностей и математическая статистика: Учеб. пособие для вузов / В. Е. Гмурман. — 9-е изд., стер. — М.: Высшая школа, 2003. — 479 с.