

ЭЛЕМЕНТЫ ЧИСЛЕННОГО ВЕРОЯТНОСТНОГО АНАЛИЗА

Определяется понятие численного вероятностного анализа. Рассматриваются его элементы с точки зрения теории и возможности их применения на практике. Изучается понятие вероятностного расширения. Описываются подходы к вычислению функций от случайных аргументов и операций над ними. Приводятся примеры использования элементов анализа в практике решения экономических задач. Показывается, что в ряде случаев данный подход можно рассматривать как реальную альтернативу методу Монте-Карло, которая позволяет существенно повысить точность вычислений и сократить их объем.

Ключевые слова: численный вероятностный анализ, вероятностное расширение, численные операции над случайными величинами, гистограммная арифметика, функции случайных аргументов, стохастические линейные и нелинейные уравнения.

В последнее время возрос интерес и увеличилась исследовательская активность в области теории и практики анализа данных больших объемов. Такие объемы информации, с одной стороны, позволяют получить более точное описание объекта исследования, а с другой – превращают поиск решений в сложную задачу, требующую применения современных математических методов обработки и анализа информации, в том числе разработки численных методов, позволяющих определять функции от случайных аргументов и осуществлять операции над ними.

В настоящее время в практике принятия управленческих решений, например при оценке инвестиционных проектов в условиях высокой неопределенности и риска с успехом, используют метод Монте-Карло [1; 2]. Однако при всех положительных качествах этот метод обладает рядом недостатков, самыми существенными из которых являются низкая скорость сходимости и высокие вычислительные затраты, что вносит дополнительные сложности в работу с данными большого объема.

В рамках обозначенного направления анализа данных рассмотрим понятие численного вероятностного анализа как раздела вычислительной математики, предметом которого является решение задач со стохастическими неопределенностями в данных с использованием численных операций над плотностями вероятностей случайных величин и их функций. Одним из основных элементов численного вероятностного анализа является гистограммная арифметика, применение которой позволяет снизить уровень неопределенности в данных и получить дополнительную информацию о распределении случайных величин [3]. Идея гистограммного подхода изложена в работах [4; 5] и заключается в следующем: наряду с общим представлением плотностей вероятностей случайных величин в виде непрерывных функций можно рассматривать случайные величины, плотность распределения которых представляет гистограмму. Например, для одномерной случайной величины гистограмма P – кусочно-постоянная функция, которая определяет-

ся сеткой $\{x_i | i = 0, \dots, n\}$ и на каждом отрезке $[x_i, x_{i+1}]$ принимает постоянное значение P_i , $h = \max_{i=0}^{n-1} \{x_{i+1} - x_i\}$.

Важным понятием численного вероятностного анализа является понятие вероятностного расширения. Определим это понятие в рамках гистограммного подхода. Для этого рассмотрим задачу определения закона распределения функции нескольких случайных аргументов.

Пусть имеется система непрерывных случайных величин $(x_1, \dots, x_n)$ с плотностью распределения $p(x_1, \dots, x_n)$. Случайная величина z связана с системой $(x_1, \dots, x_n)$ функциональной зависимостью

$$z = f(x_1, \dots, x_n).$$

Тогда плотность вероятности случайной величины z будем называть *вероятностным расширением функции*.

На основе понятия вероятностного расширения определим гистограммное вероятностное расширение. Пусть гистограмма F определяется сеткой $\{z_i | i = 0, \dots, n\}$. Зададим область $\Omega_i = \{(x_1, \dots, x_n) | z_i < f(x_1, \dots, x_n) < z_{i+1}\}$. Тогда значение гистограммы F_i на отрезке $[z_i, z_{i+1}]$ имеет вид [3]:

$$F_i = \int_{\Omega_i} p(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n / (z_{i+1} - z_i). \quad (1)$$

Гистограмму F , построенную по (1), будем называть *вероятностным гистограммным расширением*.

Далее построим гистограммные вероятностные расширения для арифметических операций над случайными величинами. Пусть P – гистограмма плотности вероятности $z = x * y$, где $*$ $\in \{+, -, \cdot, /, \uparrow\}$. Тогда на интервале $[z_i, z_{i+1}]$

$$P_i = \int_{\Omega_i} p(x_1, x_2) dx_1 dx_2 / (z_{i+1} - z_i),$$

где $\Omega_i = \{(x_1, x_2) | z_i \leq x_1 * x_2 \leq z_{i+1}\}$ [4].

Операция $\max(x, y)$ определяется через функцию распределения F :

$$F(z) = \int_{-\infty}^z P_x(\xi) d\xi \int_{-\infty}^z P_y(\xi) d\xi.$$

Пусть $f(x_1, \dots, x_n)$ – рациональная функция. Тогда для вычисления гистограммы F заменим арифметические операции гистограммными, а переменные $x_1, x_2, \dots, x_n$ – их гистограммными значениями. Полученную гистограмму F будем называть *естественным гистограммным расширением*.

Теорема 1. Пусть $f(x_1, \dots, x_n)$ – рациональная функция, каждая переменная которой встречается только один раз и $x_1, \dots, x_n$ – независимые случайные величины. Тогда естественное гистограммное расширение аппроксимирует вероятностное расширение с точностью $O(h^a)$.

Доказательство проведем по индукции. Для $n = 2$ утверждение справедливо [4]. Пусть это утверждение справедливо и для $n = k$ и гистограмма F_k аппроксимирует плотность вероятности функции $f(x_1, \dots, x_k)$ с некоторой точностью $O(h^a)$. Покажем, что это справедливо и для $n = k + 1$. Действительно, $F_{k+1} = F_k * x_{k+1}$, но

$$\begin{aligned} F_{k+1} - f(x_1, \dots, x_k, x_{k+1}) &= \\ &= (F_k - f(x_1, \dots, x_k)) * x_{k+1} \leq Ch^a * \text{supp} \{x_{k+1}\}. \end{aligned}$$

Теорема доказана.

Рассмотрим пример, который иллюстрирует теорему 1. Пусть функция представлена в виде

$$f(x, y) = xy + x + y + 1 = (x + 1)(y + 1).$$

Заметим, что только второе представление функции в виде произведения двух сомножителей подпадает под условие теоремы 1 и, следовательно, естественное гистограммное расширение будет аппроксимировать вероятностное с точностью $O(h^a)$.

Теорема 2. Пусть для функции $f(x_1, \dots, x_n)$ возможна замена переменных, такая что $f(z_1, \dots, z_k)$ – рациональная функция от переменных $z_1, \dots, z_k$, удовлетворяющая условиям теоремы 1, и z_i – функции от множества переменных $x_i, i \in \text{Ind}_i$, причем множества Ind_i попарно не пересекаются. Пусть для каждой функции z_i можно построить вероятностное расширение. Тогда естественное расширение $f(z_1, \dots, z_k)$ будет аппроксимировать вероятностное расширение $f(x_1, \dots, x_n)$.

Пример 1. Пусть $f(x_1, x_2) = (-x_1^2 + x_1) \sin(x_2)$. Тогда $z_1 = (-x_1^2 + x_1)$ и $z_2 = \sin(x_2)$. Заметим, что можно построить вероятностные расширения функций z_1, z_2 , где $f = z_1 z_2$ – рациональная функция, подпадающая под условия теоремы 1. Следовательно, естественное расширение этой функции будет аппроксимировать вероятностное расширение $f(x_1, x_2)$.

Рассмотрим случай, когда для $f(x_1, \dots, x_n)$ необходимо найти вероятностное расширение f , но не удается построить замену переменных согласно теореме 2. Пусть для определенности только x_1 встречается несколько раз. Заметим, что если подставить вместо случайной величины x_1 детерминированную t , то для функции $f(t, x_2, \dots, x_n)$ можно построить естественное вероятностное расширение. Пусть t – дискретная случайная величина, аппроксимирующая x_1 следующим образом: t принимает значения t_i с вероятностью P_i – и пусть для каждой $f(t_i, x_2, \dots, x_n)$ можно построить естественное вероятностное расширение φ_i . Тогда вероятностное расширение f функции $f(x_1, \dots, x_n)$ можно аппроксимировать плотностью вероятности φ следующим образом:

$$\varphi(\xi) = \sum_{i=1}^n P_i \varphi_i(\xi).$$

Пример 2. Пусть $f(x, y) = x^2 y + x$, где x, y – равномерные случайные величины, заданные на $[0, 1]$. Заменим x дискретной случайной величиной t , $\{t_i | t_i = (i - 0.5) / n, i = 1, 2, \dots, n\}$, $P_i = 1/n$. Далее вычислим естественные вероятностные расширения φ_i (табл. 1). Сравнение φ и f , вероятностного расширения $f(x, y)$ показывает хорошее приближение: φ аппроксимирует f с порядком $\alpha = 1,4998$.

Таблица 1

Погрешность аппроксимации вероятностных расширений

n	$\ f - \varphi\ _2$
10	$1,2887825282E - 03$
20	$4,5592973952E - 04$
40	$1,6120775967E - 04$
80	$5,6996092139E - 05$
160	$2,0151185588E - 05$

Важно отметить, что число арифметических операций для вычисления $x * y$ имеет порядок $O(n^2)$, где n – размерность сетки гистограмм. Гистограмма как кусочно-постоянная функция приближает плотность вероятности с точностью $O(1/n)$. Однако уже средние точки гистограмм приближают плотность вероятности с точностью $O(1/n^2)$. Как показано в [4], нижняя оценка скорости сходимости гистограммных расширений к вероятностным имеет значение $\alpha = 1$ (реально это значение несколько выше). Численные же эксперименты подтвердили, что в условиях теоремы 1 при правильном согласовании сеток гистограмм и использовании более тонких способов интегрирования гистограммные расширения имеют скорость сходимости $\alpha \approx 2$.

Сравним полученные результаты с методом Монте-Карло. Известно, что метод Монте-Карло имеет сходимость $O(1/\sqrt{N})$, где N – число повторов [1]. Гистограммные же расширения имеют скорость сходимости $O(1/n^a)$.

Пусть необходимо достигнуть точности ε . Число операций метода Монте-Карло при этом составит $O(\varepsilon^{-2})$, число операций с гистограммной арифметикой – $O(\varepsilon^{-2/\alpha})$, следовательно гистограммная арифметика при $\alpha > 1$ эффективнее метода Монте-Карло примерно в $O(\varepsilon^{-2(1-1/\alpha)})$ раз.

Проиллюстрируем эффективность гистограммной арифметики на численном примере.

Пример 3. Пусть требуется найти сумму трех равномерно распределенных на $[0, 1]$ случайных величин. Вычисления при $n = 30$ показывают, что ошибка при использовании численных операций гистограммной арифметики в норме l_2 составляет $1,1680E - 03$; ошибка метода Монте-Карло при числе бросаний $N = 10^5$ равна $5,4601E - 03$, а при числе бросаний $N = 10^6$ – $2,0284E - 03$, что сравнимо с погрешностью гистограммной арифметики. При этом число арифметических операций гистограммной арифметики определяется величиной порядка 30^2 или 10^3 , а метода Монте-Карло – величиной порядка 10^6 . Таким образом, в данном примере гистограммная арифметика эффективнее метода Монте-Карло примерно в тысячу раз.

Для решения ряда практических задач необходимо решать системы линейных и нелинейных уравнений. Методы их решений рассмотрены в работе [3].

В случае нахождения корня одномерного уравнения $f(x, k) = 0$, где k – случайный вектор параметров, предположим, что корень локализован на отрезке $[a, b]$, φ_z – вероятностное расширение $f(z, k)$, $z \in [a, b]$. Тогда $P(z)$ есть вероятность, что корень лежит левее (правее) точки z :

$$P(z) = \int_{-\infty}^0 \varphi_z(\xi) d\xi. \quad (2)$$

Применение численного вероятностного анализа рассмотрим на примере.

Пример 4. Пусть завод производит некоторую продукцию и осуществляет ее реализацию. По наблюдениям известна гистограмма Y_i покупок i -го товара, например, за неделю. Это означает, что известна плотность вероятности продажи товара p_i , $Y_i \in [y_i, \bar{y}_i]$.

Пусть доход D_i линейно зависит от продажи товара $D_i = \alpha Y_i$. Будем считать, что производство i -го товара не зависит от производства других товаров. Тогда общий доход будет просто суммой всех доходов по продажам всех товаров. Поэтому мы

можем рассмотреть производство только одного вида продукции и для простоты индекс опустим.

Применим гипотезу: завод несет убытки, если товар произведен в объеме R , но продано $Y < R$. Величина убытков U линейно зависит от разницы $\beta(Y - R) \leq 0$.

Пусть завод произвел товар объемом R . Оценим возможный доход:

$$D = \alpha \int_y^R p(y) y dy,$$

и возможный убыток

$$U = \beta \int_y^R p(y) (y - R) dy.$$

Заметим, что в этой постановке нам нужно максимизировать сумму $D + U \rightarrow \max$:

$$\begin{aligned} f(R) = (D + U)(R) &= \alpha \int_y^R p(y) y dy + \beta \int_y^R p(y) (y - R) dy = \\ &= (\alpha + \beta) \int_y^R p(y) y dy - \beta R \int_y^R p(y) dy \rightarrow \max. \end{aligned}$$

Приведем графики зависимости «производство–прибыль» при различных значениях параметров α и β (рис. 1). В расчетах величина $\alpha = 1$, кривые 1–5 соответствуют значениям $\beta = 0,3$ (кривая 1), $\beta = 0,4$ (кривая 2), $\beta = 0,5$ (кривая 3), $\beta = 0,7$ (кривая 4), $\beta = 1$ (кривая 5).

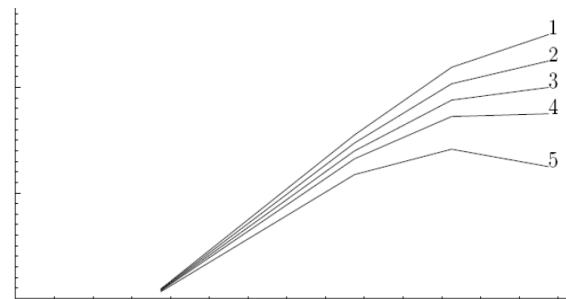


Рис. 1. Графики зависимости «производство–прибыль»

Таким образом, при небольших значениях β увеличение производства покрывает возможные убытки перепроизводства, но при сравнительно больших значениях β величину производства R следует ограничивать исходя из максимального значения $D + U$.

В качестве примера использования численных операций над гистограммными переменными рассмотрим задачу принятия решения об инвестировании проекта выпуска лекарственного препарата [2].

Пример 5. Компания рассматривает вопрос о приобретении для последующего производства патента нового лекарственного препарата. Стоимость патента составляет 3,4 млн долл. Решение принимается на основе анализа дисконтированных денежных потоков

по значениям рассчитанных показателей NPV и IRR. Горизонт расчетов – три года. Согласно прогнозам, компания в первый, второй и третий годы реализации проекта продаст соответственно 802 тыс., 967 тыс. и 1 132 тыс. упаковок лекарства по цене 6, 6,05 и 6,10 долл. за упаковку. Ставка налога на прибыль равна 32 %, ставка дисконтирования – 10 %, себестоимость – 55 %, операционные издержки – 15 % от цены препарата.

По результатам расчетов IRR проекта составляет 15 %, а NPV – 344,8 тыс. долл. Стандартная финансовая модель приведена в табл. 2.

В данном случае мы имеем дело с высоким уровнем рыночной неопределенности, поэтому стандартная финансовая модель не дает достаточных оснований для принятия решения. Для одновременного учета неопределенности в цене, продажах, себестоимости и издержках применяется численный вероятностный анализ. Основные параметры финансовой модели: цена и объем продаж – моделируются как случайные переменные, имеющие вероятностное распределение. Численный вероятностный анализ позволит понять, какие факторы повлияют на финансовые результаты проекта в наибольшей степени.

Для моделирования цены продажи (отдельно за первый, второй и третий год реализации проекта) используется треугольное распределение.

Данное распределение имеет три параметра: минимальное значение, максимальное значение и наиболее вероятное значение. Цена продажи в первый год имеет минимальное значение 5,90 долл., максимальное значение – 6,10 долл., наиболее вероятное значение – 6,00 долл. Аналогично цена продажи во второй год имеет треугольное распределение с параметрами 5,95, 6,05, 6,15 долл. Цена продажи на третий год имеет треугольное распределение с параметрами 6,00, 6,10, 6,20 долл.

Объем продаж моделируется как случайная переменная с нормальным распределением. При этом в первый год нормальное распределение имеет среднее значение (математическое ожидание) 802 тыс. долл. и стандартное отклонение 25 тыс. долл., во второй год – 967 тыс. долл. и 30 тыс. долл. и в третий год – 1 132 тыс. долл. и 25 тыс. долл. соответственно. Предполагается, что себестоимость (процент от продаж)

имеет треугольное распределение с минимальным значением 50 %, максимальным значением 65 % и наиболее вероятным значением 55 %.

Следует отметить, что в данном случае треугольное распределение имеет не симметричную форму, а немного скошено вправо. Поэтому велика вероятность того, что себестоимость будет завышена, а не занижена по сравнению с наиболее вероятным значением. Операционные издержки (процент от продаж) моделируются с нормальным распределением с ожиданием 15 % и стандартным отклонением 2 %.

Для данного проекта NPV вычислялся по формуле

$$NPV(r) = 0,8181818 \cdot 0,68z_1s_1 \times$$

$$\times \sum_{i=1}^3 \frac{c_i x_i}{(1+r)^i} - 3\,400\,000,$$

где c_i – цена в i -й год; x_i – продажи в i -й год; s_1 – себестоимость; z_1 – издержки. Заметим, что в этой формуле каждая случайная величина встречается только один раз.

Расчеты велись при размерности гистограмм $n = 50$. Сравнение вычисления NPV методом Монте-Карло показало, что при числе экспериментов $N = 1\,000\,000$ результаты совпадают с гистограммным расчетом с точностью до трех-четырех знаков после запятой. Численные эксперименты продемонстрировали, что при этом гистограммная арифметика более чем в триста раз быстрее.

Внутренняя норма доходности IRR определяет максимально приемлемую ставку дисконта, при которой можно инвестировать средства без каких-либо потерь для собственника: $IRR = r$, и при которой $NPV(r) = 0$.

Для вычисления IRR необходимо решать нелинейные уравнения. В случае гистограммного анализа использовалась формула (2). Таким образом, вычисление гистограммы корня нелинейного уравнения свелось к вычислению соответствующих интегралов от гистограммных расширений.

Анализ гистограмм NPV и IRR (рис. 2) показывает, что вероятны как крайне негативные исходы, так и значительная прибыль по сравнению со стандартным анализом.

Таблица 2

Стандартная финансовая модель

Показатель	Год 0	Год 1	Год 2	Год 3
Цена упаковки, долл.	–	6,00	6,05	6,10
Количество проданных штук	–	802 000	967 000	1 132 000
Выручка, долл.	–	4 812 000	5 850 350	6 905 200
Себестоимость, долл.	–	2 646 600	3 217 693	3 797 860
Валовая прибыль, долл.	–	2 165 400	2 632 658	3 107 340
Операционные издержки, долл.	–	324 810	394 899	466 101
Чистый доход до налогов, долл.	–	840 590	2 237 759	2 641 239
Налоги, долл.	–	588 989	716 083	845 196
Стартовые инвестиции, долл.	3 400 000	–	–	–
Чистый доход, долл.	3 400 000	251 601	1 521 676	1 796 043

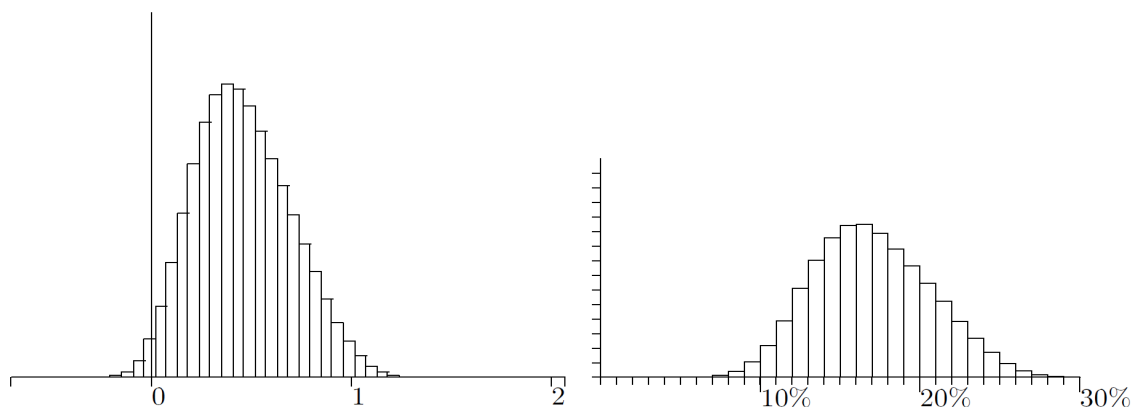


Рис. 2. Гистограммы NPV и IRR

На основании последнего примера можно сделать вывод о том, что применение гистограммной арифметики в рамках технологии визуально-интерактивного моделирования (ВИМ) [6] позволяет лицу, принимающему решение, увидеть возможные варианты негативных исходов реализации проекта в отличие от стандартного анализа, который дает только положительный ответ.

Таким образом, проведенные авторами теоретические и практические исследования приводят к двум основным выводам:

- гистограммная арифметика может рассматриваться как элемент численного вероятностного анализа, который позволяет работать с неопределенными данными в рамках различных практических приложений;
- гистограммная арифметика может использоваться как инструмент технологии визуально-интерактивного моделирования, что значительно повышает качество анализа возможных вариантов решений и дает удобное средство для их принятия.

Библиографические ссылки

1. Соболев И. М. Численные методы Монте-Карло. М. : Наука, 1973.
2. Лукашов А. В. Метод Монте-Карло для финансовых аналитиков: краткий путеводитель // Упр. корпоратив. финансами. 2007. № 1. С. 22–39.
3. Добронев Б. С., Попова О. А. Численные операции над случайными величинами и их приложения // J. of Siberian Federal Univ. Mathematics & Physics. 2011. Т. 4, № 2. С. 229–239.
4. Герасимов В. А., Добронев Б. С., Шустров М. Ю. Численные операции гистограммной арифметики и их применения // Автоматика и телемеханика. 1991. № 2. С. 83–88.
5. Добронев Б. С. Интервальная математика / Краснояр. гос. ун-т. Красноярск, 2004.
6. Добронев Б. С., Попова О. А. Применение гистограммной математики в экономических задачах исследования операций // Тр. VI Моск. междунар. конф. по исслед. операций / отв. ред. П. С. Краснощевков, А. А. Васин. М. : МАКС-Пресс, 2010. С. 90–92.

B. S. Dobronets, O. A. Popova

ELEMENTS OF A NUMERICAL PROBABILISTIC ANALYSIS

In this paper the authors define the notion of a numerical probabilistic analysis. Its elements are considered in terms of theories and their applicability in practice. The authors explore the notion of probabilistic extensions, present overview of the approaches to the calculation of functions of random arguments and operations on them, and provide with examples of elements of the analysis in the practice of solving economic problems. It is shown that this approach can be considered in some cases, as a real alternative to the Monte Carlo method that can significantly improve the calculation accuracy and to reduce their volume.

Keywords: numerical probabilistic analysis, probabilistic extension, numerical operations on random variables, histogram arithmetic, functions of random arguments, stochastic linear and nonlinear equations.

© Добронев Б. С., Попова О. А., 2012