ЧИСЛЕННЫЕ ОПЕРАЦИИ НАД ПЛОТНОСТЯМИ ВЕРОЯТНОСТЕЙ СЛУЧАЙНЫХ ВЕЛИЧИН*

Б.С. Добронец
Сибирский Федеранльный Университет, Красноярск
e-mail: `BDobronets@sfu-kras.ru`

О.А. Попова
Сибирская-автомобильно-дорожная академия,
e-mail: `olgaarc@yandex.ru`

В работе приведено описание и свойства основных операций над плотностями вероятности случайных величин. Представлены алгоритмы и примеры решений систем линейных алгебраических и нелинейных уравнений, примеры использования развиваемых методов в практике принятия решений и исследования операций. Показано, что данный подход позволяет существенно повысить качество принимаемых решений и сократить объем вычислений.

Введение

При решении большого числа практических задач возникает необходимость нахождения не только самих решений, но и различных оценок, таких как оценки погрешности, границы множеств решений и т. п [1].

Важное направление — вероятностное представление входных данных, например, в виде гистограммных чисел и разработка численных операций над ними [2, 3]. Одна из первых работ в этой области [4].

Наличие информации о плотности вероятности случайных величин приводит к возможности при расчетах учитывать и получать результаты в виде случайных величин с построенной плотностью вероятности.

Сравнение подходов интервальной математики и численных операций над плотностями вероятности случайных величин, показывает что практически значимые плотности вероятности занимают лишь небольшую часть полученных интервалов. Поэтому, в таких случаях, границы решений мало информативны. Часто при работе со случайными величинами ограничиваются вычислением лишь небольшого числа характеристик: математическое ожидание и дисперсия.

Численные операции над плотностями вероятности случайных величин позволяют существенно поднять точность расчетов при сравнительно небольшом объеме вычислений.

*Работа выполнена при поддержке гранта ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы, ГК № 02.740.11.0621.

1. Численные операции

Обозначим, через $\boldsymbol{R}$ — множество $\{\boldsymbol{x}\}$ случайных величин, заданных своими плотностями вероятности p_x, соответственно $\boldsymbol{R}^n$ — пространство случайных векторов.

Наряду с общими представлениями случайных величин своими плотностями в виде непрерывных функций, будем рассматривать случайные величины плотность распределения которых представляет гистограмму. Гистограмма P — кусочно-постоянная функция определяется сеткой $\{x_i | i = 0, ..., n\}$, на отрезке $[x_{i-1}, x_i]$ гистограмма принимает постоянное значение p_i.

Рассмотрим вопрос построения гистограммы P по некоторой p_x. Тогда значение p_i на отрезке $[x_{i-1}, x_i]$ определиться как среднее

$$p_i = \int_{x_{i-1}}^{x_i} p_x(\xi) d\xi / (x_i - x_{i-1}).$$

Имеется непрерывная случайная величина $\boldsymbol{x}$ с плотностью распределения p_x. Другая случайная величина $\boldsymbol{y}$ связана с нею функциональной зависимостью:

$$\boldsymbol{y} = f(\boldsymbol{x}).$$

Плотность распределения величины $\boldsymbol{y}$ в случае монотонной функции f согласно [5] определяется следующим образом:

$$p_y(y) = f(f^{-1}(y)) |(f^{-1})'(y)|,$$

где f^{-1} — функция обратная к f. Однако, пользоваться таким представлением в ряде случаев довольно затруднительно.

Рассмотрим построение гистограммы p_y по известной гистограмме p_x. Пусть p_x задается сеткой $\{x_i | i = 0, ..., n\}$. Вычислим сетку для гистограммы p_y $\{y_i | i = 0, ..., n\}$ как $y_i = f(x_i)$. Тогда, вероятность попадания случайной величины $\boldsymbol{x}$ в отрезок $[x_{i-1}, x_i]$ равна вероятности попадания случайной величины $\boldsymbol{y}$ в отрезок $[y_{i-1}, y_i]$. Следовательно, гистограмма p_y будет определяться сеткой $\{y_i | i = 0, ..., n\}$ и соответствующими значениями $p_{yi} = p_{xi}$.

Рассмотрим общий случай, когда f не является монотонной. Пусть $\{y_i | i = 0, ..., n\}$ — сетка для гистограммы p_y. Фиксируем отрезок $[y_{i-1}, y_i]$, найдем p_{yi}. Пусть $[a_j, b_j]$, $j = 1, ..., m$ такие, что $[y_{i-1}, y_i] = f([a_j, b_j])$. Тогда вероятность p_{yi} попадания случайной величины $\boldsymbol{y}$ в отрезок $[y_{i-1}, y_i]$ определяется, как

$$p_{yi} = \sum_{j=1}^{m} \int_{a_j}^{b_j} p_x(\xi) d\xi.$$

Рассмотрим задачу определения закона распределения функции нескольких случайных аргументов.

Приведем метод решения этой задачи для случая функции двух аргументов. Пусть имеется система двух непрерывных случайных величин $(\boldsymbol{x}, \boldsymbol{y})$ с плотностью распределения $p(x, y)$. Случайная величина $\boldsymbol{z}$ связана с $\boldsymbol{x}$ и $\boldsymbol{y}$ функциональной зависимостью:

$$\boldsymbol{z} = f(\boldsymbol{x}, \boldsymbol{y}).$$

Тогда закон распределения P_z величины $\boldsymbol{z}$ [5]:

$$P_z(r) = P(f(\boldsymbol{x}, \boldsymbol{y}) < r).$$

Пусть, как и выше, гистограмма p_z определяется сеткой $\{z_i | i = 0, ..., n\}$. Определим область $\Omega_i = \{(x, y) | z_i < f(x, y) < z_{i+1}\}$. Тогда p_{zi} имеет вид

$$p_{zi} = \int\int_{\Omega_i} p(x, y) dx dy / (z_{i+1} - z_i).$$

Пусть имеется система двух непрерывных случайных величин $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ с плотностью распределения $p(x_1, x_2)$. Известны аналитические формулы для определения плотности вероятности результатов арифметических действий над случайными величинами [6]. Однако эти формулы не всегда удобны для численных расчетов.

Основные принципы разработки гистограммных операций продемонстрируем на примере операции сложения. Пусть $\boldsymbol{z} = \boldsymbol{x}_1 + \boldsymbol{x}_2$, и носители $\boldsymbol{x}_1$ — $[a_1, a_2]$, $\boldsymbol{x}_2$ — $[b_1, b_2]$, $p(x_1, x_2)$ — плотность распределения вероятностей случайного вектора $(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Заметим, что прямоугольник $[a_1, a_2] \times [b_1, b_2]$ — носитель плотности распределения вероятностей $p(x_1, x_2)$ и плотность вероятности $\boldsymbol{z}$ отлична от нуля на интервале $[a_1 + b_1, a_2 + b_2]$. Обозначим $z_i,\ i = 0, 1, \dots, n$ — точки деления этого интервала на n отрезков. Тогда вероятность попадания величины $\boldsymbol{z}$ в интервал $[z_i, z_{i+1}]$ определяется по формуле

$$P(z_i < \boldsymbol{z} < z_{i+1}) = (\int\int_{\Omega_i} p(x_1, x_2) dx_1 dx_2,$$

где $\Omega_i = \{(x_1, x_2) | z_i \leq x_1 + x_2 \leq z_{i+1}\}$ [4]. И окончательно p_{zi} имеет вид

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

Рассмотренный выше подход обобщается на случай большего числа переменных. Пусть требуется найти гистограмму p_z суммы

$$\boldsymbol{z} = a_1 \boldsymbol{x}_1 + a_2 \boldsymbol{x}_2 + ... + a_n \boldsymbol{x}_n$$

и пусть $p(x_1, x_2, ..., x_n)$ — плотность распределения вероятностей случайного вектора $(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n)$. Тогда вероятность попадания $\boldsymbol{z}$ в интервал (z_i, z_{i+1}) соответственно равна [6]

$$P(z_i < \boldsymbol{z} < z_{i+1}) = \int ... \int_{\Omega_i} p(x_1, x_2, ..., x_n) dx_1 dx_2 .... dx_n,$$

где $\Omega_i = \{(x_1, x_2, ..., x_n) | z_i < a_1 x_1 + a_2 x_2 + ... + a_n x_n < z_{i+1}\}$, p_{zi} имеет вид

$$p_{zi} = \int ... \int_{\Omega_i} p(x_1, x_2, ..., x_n) dx_1 dx_2 .... dx_n / (z_{i+1} - z_i).$$

Рассмотрим, в качестве примера, операцию $\max(\boldsymbol{x}, \boldsymbol{y})$. Вероятность того, что $\max(\boldsymbol{x}, \boldsymbol{y}) < z$ — определяется через функцию распределения F

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

Используя функцию распределения F, можно построить гистограмму для $\max(\boldsymbol{x}, \boldsymbol{y})$ на сетке $\{z_i\}$. Тогда $p_i = F(z_{i+1}) - F(z_i)$.

2. Решение уравнений

Рассмотрим системы линейных или нелинейных уравнений

$$f_i(\boldsymbol{x}, \boldsymbol{k}) = 0, i = 1...n,$$

где $\boldsymbol{x} \in \boldsymbol{R}^n$ — случайный вектор решения, $\boldsymbol{k} \in \boldsymbol{R}^m$ — случайный вектор коэффициентов, $p_k(\xi_1 \xi_2 ... \xi_m)$ — функция плотности вероятности вектора коэффициентов и K — носитель p_k. Множество решений X имеет вид

$$X = \{x | f(x, k) = 0, k \in K\}.$$

Каждому $x \in X$ можно сопоставить подмножество коэффициентов $K_x \subset K$

$$K_x = \{k | f(x, k) = 0.\}$$

Пусть необходимо найти вероятность $P(X_0)$ попадания решения x в некоторое подмножество $X_0 \subset X$. Сопоставим X_0 множество коэффициентов $K_0 = \{K_x | x \in X_0\}$

Тогда вероятность

$$P(X_0) = \int_{K_0} p_k(\xi_1, \xi_2, ..., \xi_m) d\xi_1 d\xi_2 ... d\xi_m.$$

Нахождение подмножеств $K_x \subset K$ в общем случае не простая задача. Но для ряда частных случаев вполне реализуемая.

Рассмотрим решение систем линейных алгебраических уравнений

$$Ax = b.$$

Для простоты изложения предположим, что случайный вектор $\boldsymbol{b}$ состоит из независимых компонент $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, каждая из которых, случайная величина равномерно распределенная на отрезке $[0, 1]$. Тогда носитель плотности вероятности вектора $\boldsymbol{b}$ — квадрат $[0, 1]^2$. Пусть матрица матрица A имеет вид

$$A = \begin{pmatrix} a_{11} & -1 \\ -1 & 2 \end{pmatrix},$$

$\boldsymbol{a}_{11}$ — случайная дискретная величина равномерно распределенная на отрезке $[2, 4]$.

В левой части рис. 1 приведено кусочно-постоянное с шагом 0.1 приближение совместной плотности вероятности вектора $\boldsymbol{x}$. Сплошной линией приведена граница множества решений исходной системы.

Рассмотрим пример системы нелинейных уравнений

$$\boldsymbol{a}x^2 + \boldsymbol{b}y^2 - 4 = 0,$$

$$xy - \boldsymbol{c} = 0,$$

где $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ — равномерные случайные величины, плотности вероятности которых имеют носители $[1, 1.1]$, $[2, 2.1]$, $[0.505, 0.51]$. В правой части рис. 1 приведено кусочно-постоянное приближение совместной плотности вероятности вектора $(\boldsymbol{x}, \boldsymbol{y})$ решения системы нелинейных уравнений.

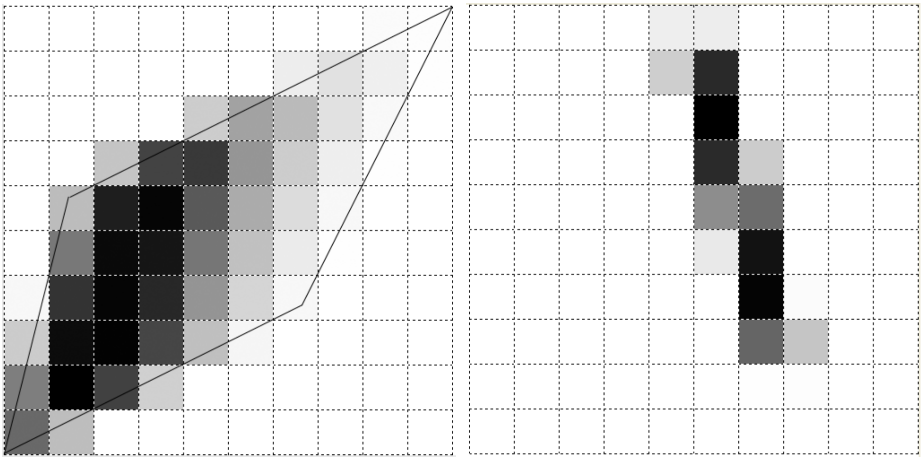

Рис. 1. Приближение плотности вероятности векторов решений

Рассмотрим задачу нахождения корня одномерного уравнения $f(x,k) = 0$. Предположим, что корень локализован на отрезке $[a,b]$. Заметим, что x будет представлять собой случайную величину и необходимо найти плотность распределения $\boldsymbol{x}$ этой случайной величины. Далее, мы будем предполагать, что можно с достаточной точностью для любого $z \in [a,b]$ вычислять плотность распределения ϕ_z случайной величины $f(z,\boldsymbol{k})$.

Тогда, $P(z)$ есть вероятность, что корень лежит левее (правее) точки z:

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

3. Прогноз временных рядов

Пусть имеется ряд наблюдений случайной величины $x(t_0), x(t_1), \ldots, x(t_n)$, которые произведены в моменты времени t_i

$$x(t_i) = f(t_i) + \eta(t_i), i = 0,1,2,...,n,$$

где $f(t)$ — некоторая функция, η — случайная величина с математическим ожиданием равным нулю. Предполагаем, что плотности вероятности $P_\eta(t)$ случайной величины η может зависеть от времени.

Для оценки f и $P_\eta(t)$ будем использовать гистограммный подход. Для этих целей начиная с момента времени t_l строим специальным способом гистограммы G_i используя значения с x_{i-l} по x_{i+l}. По полученным гистограммам вычисляем математическое ожидание $m_i = M[G_i]$.

Далее по полученным величинам m_i, строится прогноз для $f(t)$ и по гистограммам G_i приближение для $P_\eta(t)$.

Таким образом, данный подход позволяет получить оценку не только $f(t)$, но и гистограммную оценку плотности $P_\eta(t)$ случайной величины η.

4. Приложения

В качестве примеров использования численных операций над гистограммными переменными были рассмотрены задачи принятия решения об инвестировании проекта выпуска лекарственного препарата [7]. На основе вероятностных данных построены гистограммы NPV и IRR для проекта. Из анализа гистограмм NPV и IRR видно, что вероятны как крайне негативные исходы, так и возможна значительная прибыль в сравнении со стандартным анализом. Приведенный пример показывает, что применение гистограммной арифметики в рамках технологии визуально-интерактивного моделирования (ВИМ) [8] позволяет ЛПР увидеть возможные варианты негативных исходов реализации проекта, по сравнению со стандартным анализом, который дает только положительный ответ.

В заключение отметим, что проведенные авторами теоретические и практические исследования позволяют сделать два основных вывода:

• гистограммная арифметика может рассматриваться, как численный метод вероятностного анализа, позволяющий работать с неопределенными данными в рамках различных практических приложений [9, 10];

• гистограммная арифметика может использоваться, как инструмент ВИМ – технологии, что значительно повышает качество анализа возможных вариантов решений и дает в руки ЛПР удобное средство для принятия решений.

Список литературы

[1] Добронец Б.С. Интервальная математика. Красноярск: Изд-во КГУ, 2004. 216 с.

[2] Berleant D. Automatically verified reasoning with both intervals and probability density functions // Interval Computations. 1993 No. 2, P. 48–70

[3] Li W., Hym J. Computer arithmetic for probability distribution variables // Reliability Engineering and System Safety. 2004. N 85. P. 191–209.

[4] Герасимов В.А., Добронец Б.С., Шустров М.Ю. Численные операции гистограммной арифметики и их применения // АиТ. 1991, №2. С. 83–88.

[5] Вентцель Е.С. Теория вероятностей. М.: Наука, 1969. 576 с.

[6] Гнеденко Б. В. Курс теории вероятностей. М.: Наука, 1988. 448 с.

[7] Добронец Б.С., Попова О.А. Применение гистограммной математики в задачах принятия экономических решении // Труды IX международной ФАМЭТ 2010 конференции. — Красноярск: КГТЭН, СФУ, 2010. С.127-130

[8] Добронец Б.С., Попова О.А. Гистограммая арифметика для визуально-интерактивного моделирования в задачах принятия экономических решений // Актуальные проблемы анализа ипостроения информационных систем и процессов: сб. статей Международной научно-технической конференции. Таганрог: Издательство Технологичсекого института ЮФУ, 2010. С. 53–57.

[9] Добронец Б.С., Попова О.А. Применение гистограммной математики в экономических задачах исследования операций. — VI Московская международная конференция по исследованию операций (ORM2010): Москва, 19–23 октября 2010 г.: Труды / Отв. ред. П.С. Краснощеков, А.А. Васин. — М.: МАКС Пресс, 2010. С. 90–92.

[10] Добронец Б.С., Попова О.А. Гистограмный подход к согласованию интересов в условиях неопределенности и задачах принятия решений // Труды XIV международной ЭМ'2010 конференции.— Красноярск: КГТЭН, СФУ, 2010. С.84–88.