ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

2012 Управление, вычислительная техника и информатика № 4(21)

УДК.519.24

**Б.С. Добронец, О.А. Попова**

# ЧИСЛЕННЫЙ ВЕРОЯТНОСТНЫЙ АНАЛИЗ ДЛЯ ИССЛЕДОВАНИЯ СИСТЕМ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ[1]

Рассмотрено использование численного вероятностного анализа для исследования систем в условиях различного типа неопределенностей, включая эпистемистическую.

**Ключевые слова:** *численный вероятностный анализ, эпистемистическая неопределенность, вероятности второго порядка, гистограммы второго порядка.*

Изучение любых сложных систем всегда представляет собой трудную задачу, которая приобретает наивысший уровень сложности, если рассматриваемая система находится в условиях неопределенности и риска. Специфика сложности исследования таких систем обуславливается рядом факторов, которые с некоторой степенью общности можно отнести к четырем основным группам. Первая группа характеризуется внутренней сложностью системы как таковой. Вторая – внешней сложностью явлений и процессов, влияющих на систему и взаимодействующих с ней. Третья – недостаточностью информации и знаний о предыстории процесса функционирования системы. Четвертая группа факторов связана с разрабатываемыми и применяемыми технологиями анализа систем, которые в настоящее время отличаются чрезвычайной сложностью, расширением круга решаемых задач и диапазоном эксплуатационного использования. Обеспечение требуемой надежности и сложность исследования таких систем требует привлечения большого объема материальных, финансовых, интеллектуальных, временных и других ресурсов. При этом практика показывает, что привлекаемые ресурсы и вложения в их исследования не всегда пропорциональны требуемому уровню надежности и качеству функционирования систем. Поэтому актуальной является задача: там, где это возможно, снизить уровень сложности применяемой технологии исследования системы, не изменив, а по возможности улучшив ее показатели. Для снижения сложности, привносимой в исследования рассмотренными группами факторов, авторами разрабатывается технология изучения сложных систем на основе численного вероятностного анализа (ЧВА) [1,2], который представляет собой эффективный инструмент в условиях неопределенности и риска.

Предметом ЧВА является решение различных задач со стохастическими неопределенностями в данных с использованием численных операций над плотностями вероятностей случайных величин и функций со случайными аргументами. Для этого предлагается разнообразный инструментарий, включающий такие понятия, как гистограммная арифметика [1], вероятностные, естественные и гистограммные расширения [2], гистограммы второго порядка [3].

ЧВА представляет собой непараметрический подход и может успешно применяться для вероятностного описания систем в рамках визуально-интерактивного моделирования, повышая тем самым качество исследования систем.

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

На тестовых примерах и ряде практических задач доказаны преимущества данного подхода перед методом Монте-Карло [2].

## 1. Типы неопределенности

Для исследования систем в условиях неопределенности используются различные методы, в том числе подходы, основанные на описании законов распределения случайных величин [3]. В том случае, когда класс, к которому принадлежит искомое распределение (с точностью до численных значений конечного числа параметров), неизвестен, для восстановления распределения используются непараметрические методы [3].

Практика решения подобных задач показывает, что выбор соответствующего метода вероятностного описания системы должен обязательно учитывать характер и тип неопределенности, которые, в первую очередь, обуславливаются двумя основными факторами, а именно, изменчивостью физических явлений и процессов и частичным неведением о них [4]. Важно заметить, что при решении практических задач необходимо в понятии неопределенности (commom uncertainty), выделять две важные составляющие, а именно, элиторную (aleatory) и эпистемистическую (epistemic) неопределенности [5].

При этом первая составляющая, или элиторная неопределённость, в первую очередь определяется изменчивостью явлений и процессов, является неотъемлемым атрибутом случайных событий и характеризуется присущей поведению систем случайностью, которая может быть представлена частотными функциями распределения. Альтернативная терминология для элиторной неопределенности включает такие понятия, как изменчивость, стохастическая неопределенность, «объективная» неопределенность или неопределенность типа А. Теория вероятностей предназначена для моделирования, оценки и оперирования именно с элитерными неопределённостями, когда имеется возможность оперировать с повторными выборками. В свою очередь, вторая составляющая, или эпистемическая неопределённость, обусловлена недостатком знаний о системе и характеризуется неопределенностью самих вероятностных оценок. Эпистемическая неопределённость прямо связана с объёмом и достоверностью информации, на основании которой получаются эти оценки, характеризуется состоянием отсутствия знаний об изучаемом объекте, о предыстории рассматриваемого процесса, невозможностью использовать повторные выборки. Альтернативная терминология для эпистемистической неопределенности включает такие понятия, как «субъективная» неопределенность, неопределенность знаний.

## 2. Построение функции плотности в условиях элиторной неопределенности

Для исследования систем в условиях элиторной неопределенности используются различные методы, которые опираются на теорию вероятностей, в том числе подходы, основанные на описании законов распределения случайных величин. В этом случае важно отметить, что вероятностные оценки законов распределения носят детерминированный характер. В том случае, когда класс, к которому принадлежит искомое распределение (с точностью до численных значений конечного числа параметров) неизвестен, для восстановления распределения используются непараметрические методы [6]. Рассмотрим применение ЧВА в условиях элиторной неопределенности. Для этого рассмотрим гистограммный подход.

Пусть система описывается функциональной зависимостью $y = f(x)$, где $x$ – вектор $m$ входных неопределенных величин, а $y$ – выходное значение. Рассмотрим задачу построения функции плотности вероятности $y$ в условиях, когда известны повторные выборки для вектора $x$. В рамках ЧВА в случае независимых переменных $x_i$ для решения этой задачи необходимо знать их функции плотности вероятности [2].

Исторически одним из первых методов решения данной задачи был гистограммный подход [6], в основе которого лежало понятие гистограммы, где гистограмма $P$ – кусочно-постоянная функция, определенная сеткой $\{z_i,\ i = 0,1,\ldots,n\}$, на отрезке $[z_{i-1}, z_i]$, принимающая постоянное значение $p_i$.

Рассмотрим построение гистограммы $P$ для случайной величины $X$. Пусть в случае элиторной неопределенности для случайной величины $X$ известна повторная выборка $(X_1, X_2,\ldots,X_N)$. Обозначим через $n_j$ количество членов $X_i$ повторной выборки, попавших в интервал $[z_{i-1}, z_i]$, тогда

$$p_j = \frac{n_j}{N(z_j - z_{j-1})}.$$

Применим гистограммный подход, когда имеются две зависимые случайные величины $X$, $Y$ и рассмотрим построение оценки их совместной функции плотности вероятности. Предположим, что для этих целей нам известна повторная выборка $(X_1,Y_1), (X_2,Y_2),\ldots,(X_N,Y_N)$.

Оценим их совместную плотность вероятности кусочно-постоянной функцией $P_{XY}$, которая принимает на каждом прямоугольнике$[v_{i-1}, v_i]\times[w_{j-1},w_j]$ постоянное значение $p_{ij}$:

$$p_{ij} = \frac{n_{ij}}{N(v_i - v_{i-1})(w_j - w_{j-1})},$$

где $n_{ij}$ – количество членов выборки, попавших в прямоугольник $[v_{i-1}, v_i]\times$ $\times\ [w_{j-1},w_j]$.

Рассмотрим случай, когда $x_1, x_2,\ldots,x_m$ – зависимые переменные. Пусть известна повторная выборка $X_1 = (x_1, x_2,\ldots,x_m)_1$, $X_2 = (x_1,x_2,\ldots,x_m)_2,\ldots, X_N = (x_1,x_2,\ldots,x_m)_N$.

Построим гистограммную оценку $Py$ функции плотности вероятности для случайной величины $y$. Пусть гистограмма $Py$ определяется сеткой $\{z_i,\ i = 0,1,\ldots n\}$. Тогда на отрезке $[z_{i-1}, z_i]$ гистограмма принимает значение $P_j$

$$p_j = \frac{n_j}{N(z_j - z_{j-1})},$$

где $n_j$ – количество $y_i = f(X_i)$, попавших в в интервал $[z_{i-1}, z_i]$.

В рамках ЧВА имеется возможность выполнения численных операций над плотностями случайных величин и их функций. Для этого предлагается использовать гистограммную арифметику. В работах [1,2] рассмотрены вопросы, связанные с определением численных операций над случайными величинами, приведены тестовые примеры и решены ряд практических задач, например задача оценки инвестиционного проекта.

**Пример 1.** Рассмотрим задачу принятия решения об инвестировании проекта выпуска лекарственного препарата [1]. Для данного проекта $NPV$ вычислялось по следующей формуле:

$$NPV(r) = 0{,}8181818 \cdot 0{,}68 z_1 s_1 \sum_{i=1}^{3} \frac{c_i x_i}{(1+r)^i} - 3400000,$$

где $c_i$ – цена в $i$-й год, $x_i$ – продажи в $i$-й год, $s_1$ – себестоимость, $z_1$ – издержки, $r$ – ставка дисконтирования. Функции плотности вероятности для переменных $c_i$, $x_i$, $s_1$, $z_1$ были представлены гистограммами с $n = 50$. Сравнение вычисления $NPV$ с методом Монте-Карло показало, что при числе экспериментов $N = 1\,000\,000$ р езультаты совпадает с гистограммным расчетом с точностью до трех-четырех знаков после запятой. Численные эксперименты показали, 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: $IRR = r$, при котором $NPV(r) = 0$. Для вычисления $IRR$ необходимо решать нелинейные уравнения. В случае численного вероятностного анализа вычисление гистограммы корня нелинейного уравнения сводилось к вычислению соответствующих интегралов от гистограммных расширений [1].

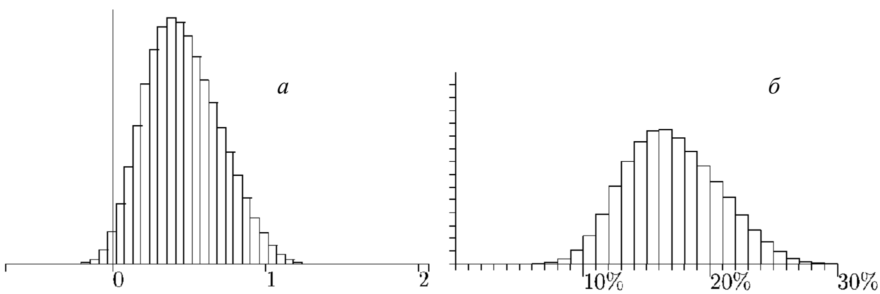


Рис. 1. Гистограммы $NPV$ и $IRR$

На рис. 1 приведены гистограммы $NPV$ (млн руб.) – *а* и $IRR$ (%) – *б*. 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) позволяет лицу, принимающему решение, увидеть возможные варианты негативных исходов реализации проекта, по сравнению со стандартным анализом, который дает только положительный ответ.

## 3. Анализ подходов и методов в условиях эпистемистической неопределенности

Для вероятностного описания систем в условиях элиторной неопределенности априори предполагается, что вероятностные характеристики имеют детерминированный характер. Если система находится в условиях эпистемистической неопределенности, то для ее анализа важен факт неопределенности вероятностных оценок. Анализ публикаций по данной тематике позволил выделить шесть основных направлений в изучении эпистемистической неопределенности. Дадим краткую характеристику каждого из подходов.

Первый подход отражает субъективную природу эпистемистической неопределенности, когда необходимая информация, снижающая уровень неопределенности и необходимая для получения вероятностных оценок, зависит от субъекта. В этом случае используются экспертные оценки. Вместо задания значения вероятности осуществления события эксперт задает некоторое множество таких значений и приписывает каждому значению вероятность его истинности. Чтобы получить точечную оценку вероятности события, рассчитывается математическое ожидание полученного распределения.

Подход второй – интервальный, связан с возможностью для построения вероятностных оценок использовать информацию о граничных (интервальных) значениях оцениваемых характеристик. Эти методы достаточно сложны и достаточно широко представлены в рамках интервального анализа.

Подход третий – анализ вероятностной чувствительности – имеет исключительно прикладной характер.

В конкретной предметной задаче производится субъективное оценивание вероятностей релевантных событий. На основе полученной информации решается задача и получаются требуемые конечные результаты.

После этого исходные значения вероятностей варьируются в заданных пределах и оценивается влияние этого варьирования на конечные результаты. Анализ вероятностной чувствительности нашел широкое распространение, особенно в задачах принятия решений.

Подход четвертый – теория Демпстера – Шефера (Dempster – Shafer Theory of Evidence) [5]. Эта теория является расширением теории вероятностей на случай, когда пространство случайных событий состоит не из синглетонов, а может включать в себя некоторые подмножества событий. Теория Демпстера – Шефера позволяет справиться с такими задачами, в которых теория вероятностей принципиально неприменима.

Подход пятый представляет собой метод «распространения» вероятностей и связан с понятием сети уверенностей. Он был предложен Ю. Пирлом и представляет эффективное средство моделирования исходных неопределенных ситуаций во многих приложениях искусственного интеллекта и теории принятия решений.

Шестой подход опирается на понятие «вероятность второго порядка» и известен как second-order probability. Данный подход представляет собой метод, позволяющий строить вероятностные оценки в случае эпистемистической неопределенности [7]. Концепция вероятностей второго порядка была изложена в 1996 году в работах Mosleh A. и V. M. Bier. Анализ публикаций показал, что несмотря на то, что данное направление достаточно активно развивается за рубежом, понятие «вероятность второго порядка» еще находится в стадии определения. Ниже приводится пример задачи, который иллюстрирует неопределенность вероятностных оценок или, следуя зарубежной терминологии, second-order probability.

**Пример 2.** Пусть случайная величина $S_t$ имеет треугольное распределение $P_t$ на отрезке [0,1], высота $h = 2$ и вершина в некоторой точке, $(t,2)$, $t$ – случайная величина с треугольным распределением на отрезке [0,25, 0,75] с вершиной (0,5, 4) (рис. 2, *а*).

## 4. Гистограммный подход в условиях эпистемистической неопределенности. Гистограммы второго порядка

Рассмотрим возможности применения ЧВА в случае эпистемической неопределённости, когда вероятностные оценки носят неопределенный характер. Для решения этой задачи можно использовать интервальные гистограммы и гистограммы второго порядка. В тех случаях, когда нет возможности получить точную функцию распределения случайной величины задают оценки плотности распределения сверху и снизу. Такие оценки удобно аппроксимировать *интервальными гистограммами* [3]. Гистограмму будем называть интервальной, если значения гистограммы принимают интервальные значения.

В случае эпистемической неопределённости наряду с интервальными гистограммами авторами предлагается использовать *гистограммы второго порядка*, т.е. такие гистограммы, каждый столбец которых – гистограмма [3]. Определим гистограмму второго порядка (ГВП) как кусочно-гистограммную функцию. ГВП так же, как и обычная гистограмма, определяется сеткой $\{z_i,\ i = 0,1,\dots,n\}$ и набором гистограмм $\{P_i\ ,\ i = 1,2,\dots,n\}$. На каждом отрезке $[z_{i-1}, z_i]$ ГВП принимает постоянное гистограммное значение $P_i$.

Рассмотрим построение ГВП в условиях эпистемистической неопределенности. Пусть мы имеем ряд гистограмм $\{Y_i\ ,\ i = 1,2,\dots,N\}$. Каждой $Y_i$ поставим в соответствие вероятность $p_i$: $\sum p_i. = 1$. Для простоты будем считать, что все гистограммы $Y_i$ заданы на сетке $\{z_i,\ i = 0,1,\dots n\}$ и на отрезке $[z_{k-1}, z_k]$ $Y_i$ принимает значение $Y_{ik}$. Таким образом, на каждом отрезке $[z_{k-1}, z_k]$ имеем случайную величину $Y_k$, принимающую значения $Y_{ik}$. с вероятностью $p_i$. Используя эти значения, мы можем на каждом отрезке $[z_{k-1}, z_k]$ восстановить гистограмму $Pz_k$.

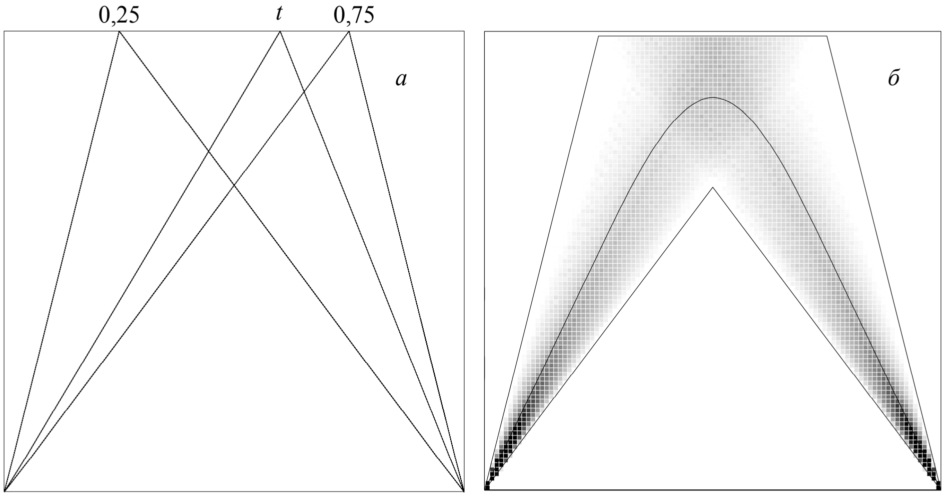


Рис. 2. Построение гистограммы второго порядка

Для примера 2 на рис. 2, *б* приведена гистограмма второго порядка, где оттенками серого показано распределение вероятностей. Интервальное распределение (максимальное и минимальное $P_t$ для всех $t$) изображено граничными линиями. Внутренняя линия определяет *«эффективную» плотность вероятности* гистограммы второго порядка – математическое ожидание плотностей вероятности $P_t$ в точке $x$.

Для осуществления численных операций над неопределенными переменными, заданными своими фунциями плотности в виде гистограмм второго порядка, в условиях эпистемистической неопределенности определим арифметику для ГВП.

Пусть $X, Y$ – ГВП, определяются сетками $\{v_i,\ i = 0,1,\ldots,n\}$, $\{w_i,\ i = 0,1,\ldots,n\}$ и наборами гистограмм $\{Px_i\}$, $\{Py_i\}$. Пусть $Z = X*Y$, где $* \in \{+,-,\cdot,/,\uparrow\}$. Построим Z как ГВП. Зададим сетку $\{z_i,\ i = 0,1,\ldots,n\}$, тогда гистограмма $Pz_i$ на отрезке $[z_{i-1}, z_i]$, следуя работе [1], определяется по формуле

$$Pz_i = \int_{\Omega_i} X(\xi)Y(\eta)d\xi d\eta,$$

где $\Omega_i = \{(\xi,\eta) \,|\, z_i \le \xi * \eta \le z_{i+1}\}$. Заметим, что на каждом прямоугольнике $[v_{i-1}, v_i]\times[w_{j-1},w_j]$ функция $X(\xi)Y(\eta)$ есть постоянная гистограмма $Px_i \cdot Py_i$. Интеграл от гистограммы по некоторой области есть значение гистограммы, умноженное на площадь области.

Проиллюстрируем, как работает гистограммная арифметика в случае сложения двух ГВП.

**Пример 2**. Пусть необходимо сложить две гистограммы второго порядка $X$ и $Y$. Гистограммы $X$ и $Y$ порождены равномерными случайными величинами, заданными соответственно на отрезках $[0, t_1]$ и $[t_2, 2]$, где $t_1$ – равномерная случайная величина, заданная на отрезке [1,2], $t_2$ – равномерная случайная величина, заданная на отрезке [0,1]. Результат сложения двух гистограмм представлен в виде гистограммы второго порядка $Z$, изображенной на рис. 3. Носителем $Z$ является отрезок [0,4], высота 1, значения плотности вероятности представлены оттенками серого.

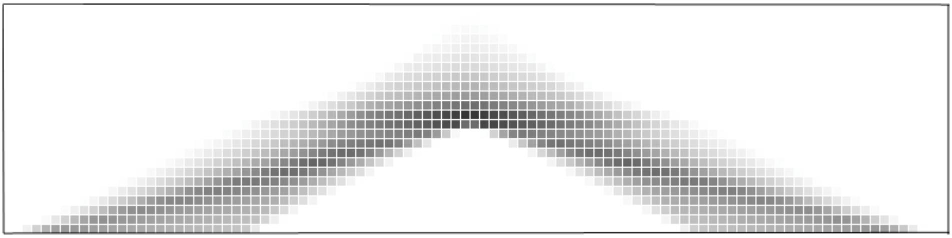

Рис. 3. Сумма двух гистограмм второго порядка $Z = X+Y$

## Заключение

В работе рассмотрено использование численного вероятностного анализа для исследования систем в условиях как элиторной, так и эпистемистической неопределенности. В первом случае функции плотности вероятности случайных переменных представляются в виде гистограмм, во втором – используются вероятности второго порядка. Для учета эпистемистической неопределенности предложено использовать гистограммы второго порядка. Численные примеры показали эффективность использования гистограммной арифметики для исследования систем.

ЛИТЕРАТУРА


1. *Добронец Б.С.*, *Попова О.А.* Численные операции над случайными величинами и их приложения // Журн. СФУ. Сер. Матем. и физ. 2011. Т. 4. № 2. С. 229–239.
2. *Добронец Б.С.*, *Попова О.А.* Численный вероятностный анализ и вероятностные расширения // Тр. XIV Международной ЭМ конференции по эвентологической математике и смежным вопросам. Красноярск, 2011. С. 67–69.
3. *Добронец Б.С.*, *Попова О.А.* Гистограммные временные ряды // Тр. X Международной конференции ФАМЭТ-2011. Красноярск: КГТЭН, СФУ, 2011. С.130–133.

4. *Воробьев О.Ю.* Современные теории неопределенности: эвентологический взгляд // Тр. VIII Международной конференции ФАМ. Красноярск: СФУ, 2009. С. 83–92.
5. *Swiler L.P.*, *Giunta A. A.* Aleatory and epistemic uncertainty quantification for engineering applications // Sandia Technical Report, SAND2007–2670.
6. *Крянев А.В.*, *Лукин Г.В.* Математические методы обработки неопределенных данных. М.: ФИЗМАТЛИТ, 2006. 216 с.
7. *Skyrms B.* Higher Order Degrees of Belief // Prospects for Pragmatism: Essays in Memory of F.P. Ramsey / D.H. Mellor, ed. Cambridge; New York: Cambridge University Press, 1980. P. 109–137.

*Добронец Борис Станиславович*
*Попова Ольга Аркадьевна*
Сибирский федеральный университет
E-mail: BDobronets@yandex.ru; olgaarc@yandex.ru
Поступила в редакцию 2 июля 2012 г.

*Dobronets Boris S.*, *Popova Olga A.* (Siberian Federal University. Krasnoyarsk). **Numerical probabilistic analysis for the study of systems with uncertainty.**

Keywords: numerical probabilistic analysis, epistemic uncertainty, second order probability, second order histogram.

In the paper consider the use of a numerical probabilistic analysis for the study of systems under aleatory and epistemic uncertainty. In the first case, the probability density functions of random variables are presented as histograms, the second uses of the second order probability. To account for uncertainty epistemic proposed to use the second order histogram. Numerical examples have shown the effectiveness of numerical probabilistic analysis for study systems.