КРАТКОЕ ОПИСАНИЕ КОМПЬЮТЕРНОЙ СИСТЕМЫ EDSRM (Economical Double-Sided Rejection Method) – см. http://nikita-e.ru:5000/ или http://nikita-e.ru:5000/index.html – И ОСНОВНЫЕ АЛГОРИТМЫ СИСТЕМЫ

1. Предназначение и применения системы EDSRM

Система **EDSRM** предназначена для демонстрации эффективности и исследования границ применимости экономичного двустороннего метода исключения с кусочно-постоянными мажорантой и минорантой в применении к моделированию (реализации выборочных значений на компьютере) одномерных случайных величин, распределенных на конечных интервалах согласно монотонным плотностям распределения.

Особо отметим, что система **EDSRM** является определенным дополнением к разработанной ранее системе **NMPUD** (Numerical Modelling of Probabilistic Univariate Distributions; ссылка https://nmpud.netlify.app**;** см., например, [1]) по изучению адекватности и экономичности формул метода обратной функции распределения (см., например, разделы 2.5 и 2.6 учебника [2]) и простейших вариантов метода дискретной суперпозиции (см., например, раздел 11.1 учебника [2]). Конкретнее, в системе **EDSRM** происходит, в том числе, сравнение времен реализации формул метода обратной функции распределения и соответствующих реализаций универсального экономичного алгоритма двустороннего метода исключения для одномерных случайных величин с монотонными плотностями распределения.

Системы **EDSRM** и **NMPUD** используются в первую очередь как полезные (и даже необходимые) инструменты по выбору нужных экономично моделируемых вероятностных распределений для исследователей, занимающихся созданием и (или) использованием численных стохастических моделей для решения актуальных прикладных задач.

Кроме того, системы **EDSRM** и **NMPUD** активно применяются в учебном процессе, причем сразу в нескольких направлениях.

Во-первых, системы **EDSRM** и **NMPUD** используются для ознакомления школьников старших классов с основами численного статистического моделирования (в рамках функционирования лаборатории математического моделирования лицея № 130 города Новосибирска) – на уровне моделирования случайных величин, при этом происходит закрепление школьного (и весьма полезного для сдачи профильного Единого государственного экзамена по математике) материала по таким темам как «Свойства элементарных функций», «Начала математического анализа (дифференцирование и интегрирование функций)», «Начала теории вероятностей и математической статистики (понятие плотности распределения, закон больших чисел и его применение в статистике и др.)».

Во-вторых, системы **EDSRM** и **NMPUD** используются для иллюстрирования соответствующих разделов университетских курсов: «Методы Монте-Карло» (4 курс бакалавриата направления «Прикладная математика и информатика» механико-математического факультета Новосибирского университета – НГУ) [2], части второй «Основы метода Монте-Карло» дисциплины «Символьные и численные расчеты в физических приложениях» (первый год обучения, магистратура физического факультета НГУ), междисциплинарного курса «Теоретические основы и прикладные аспекты стохастического моделирования» (НГУ) и др.

В-третьих, системы **EDSRM** и **NMPUD** применяется для выполнения и проверки студентами примеров из творческого семестрового домашнего задания по конструированию моделируемых плотностей (в курсах «Методы Монте-Карло» и «Символьные и численные расчеты в физических приложениях» в НГУ).

В-четвертых, системы **EDSRM** и **NMPUD** рассматриваются их авторами как важнейший раздел разрабатываемого электронного учебника по методам Монте-Карло.

Разработка систем **NMPUD** и **EDSRM** поддержана Благотворительным фондом В. Потанина (в рамках Грантовых конкурсов 2019 и 2023 года, соответственно).

2. Основное наполнение системы EDSRM

Как отмечено выше, описываемая система **EDSRM** существенно обогащает возможности, реализованные в системе **NMPUD**, по компьютерному моделированию одномерных случайных величин $\xi \in (a, b)$, имеющих монотонные плотности распределения

$$f_\xi(u), \;\; u \in (a, b) \tag{1}$$

на конечном интервале (т. е. здесь $-\infty < a < b < +\infty$). В первую очередь, нас будут интересовать т. н. элементарные плотности (см. раздел 2.6 учебника [2]), для которых выводятся аналитические (программируемые, представляющие собой композиции элементарных функций) формулы метода обратной функции распределения для численного моделирования выборочных значений ξ_0 случайных величин ξ вида

$$\xi_0 = \psi_\xi(\alpha_0) = F_\xi^{-1}(\alpha_0); \tag{2}$$

здесь $F_\xi(x) = \mathbf{P}\{\xi < x\}$ – функция распределения случайной величины ξ, а $\alpha_0 \in U(0,1)$ – стандартное случайное число (см. раздел 2.4 учебника [2]), т. е. выборочное значение случайной величины $\alpha \in U(0,1)$, равномерно распределенной в интервале $(0,1)$; это выборочное значение реализуется на компьютере с помощью специальной подпрограммы (именуемой в языках программирования как *RAND* или *RANDOM*).

В системе **EDSRM** изучаются случаи, когда формулы (2) являются трудоемкими. В этих случаях наглядно показывается целесообразность использования экономичного (с уравниванием вероятностей) двустороннего метода исключения с кусочно-постоянными мажорантой и минорантой (см. раздел 11.6 учебника [2] и раздел 3 данной инструкции). Сравнение затрат на использование формул (2) метода обратной функции распределения и приведенного ниже алгоритма 4 (и его модификаций для убывающих и кусочно-монотонных плотностей) показывается соответствующими диаграммами в разделе **«Трудоемкость»** системы **EDSRM**.

3. Описание алгоритма двустороннего метода исключения с кусочно-постоянными мажорантой и минорантой для случайной величины с монотонной плотностью распределения

3.1. Утверждения, обосновывающие мажорантный метод исключения

В компьютерной системе **EDSRM** рассматривается и оптимизируется (по времени получения одного выборочного значения ξ_0) алгоритм для моделирования одномерной случайной величины $\xi \in (a, b); -\infty < a < b < +\infty$ с монотонной плотностью распределения такой, что

$$f_\xi(u) = Hf(u); \;\;\; u \in (a, b); \;\; H = \frac{1}{\int_a^b f(u)\, du}; \tag{3}$$

здесь H – соответствующая нормирующая константа.

Идея алгоритма основана на следующих хорошо известных фактах.

УТВЕРЖДЕНИЕ 1 (см., например, раздел 11.5 книги [1]). *Если случайная точка* (ξ, η) *равномерно распределена в области*

$$G = \{(u, v): u \in (a, b); \;\; 0 < v < f(u)\}, \tag{4}$$

т. е. в «подграфике» функции $f(u)$ из соотношения (3) *(обозначение $(\xi,\eta) \in U(G)$); при этом $\xi \in (a,b)$ и $\eta \in (0, f(\xi))$.*

Тогда случайная величина ξ распределена согласно плотности (3).

УТВЕРЖДЕНИЕ 2 (см., например, раздел 2.4 книги [1]). *Если d-мерная случайная точка $\boldsymbol{\alpha} \in U\left(G^{(1)}\right)$ равномерно распределена в области $G^{(1)} \subset \mathbb{R}^d$ конечного объема* (площади) $\bar{G}^{(1)} = \int_{G^{(1)}} d\boldsymbol{u}$, *то она также равномерно распределена в любой подобласти $G \subseteq G^{(1)}$ объема $\bar{G}$ при условии попадания в эту подобласть; при этом $\mathbf{P}\{\boldsymbol{\alpha} \in G\} = \bar{G}/\bar{G}^{(1)}$.*

УТВЕРЖДЕНИЕ 3 (см., например, раздел 11.5 книги [1]). *Пусть случайная величина $\xi^{(1)} \in (a,b)$ распределена согласно плотности*

$$f_{\xi^{(1)}}(u) = H^{(1)} f^{(1)}(u), \;\; H^{(1)} = \frac{1}{\int_a^b f^{(1)}(u)\,du}, \tag{5}$$

а условное распределение при фиксированном $\xi^{(1)} = \xi_0^{(1)}$ случайной величины η является равномерным на интервале $\left(0, f^{(1)}\left(\xi_0^{(1)}\right)\right)$. Тогда случайная точка $\left(\xi^{(1)},\eta\right)$ равномерно распределена в «подграфике»

$$G^{(1)} = \left\{(u,v): u \in (a,b);\;\; 0 < v < f^{(1)}(u;\boldsymbol{\lambda})\right\} \tag{6}$$

функции $f^{(1)}(u)$, т. е. $\left(\xi^{(1)},\eta\right) \in U\left(G^{(1)}\right)$.

3.2. Двусторонний метод исключения

Применение компьютерной системы **EDSRM** оказывается особо эффективным для весьма распространенных случаев моделирования одномерных случайных величин $\xi \in (a,b)$, у которых функции $f(u)$ из формулы плотности (3) и (или) моделирующие формулы вида $\xi_0 = \psi_\xi(\bar{\alpha}_0)$ (здесь $\bar{\alpha}_0 \in U(0,1)$ – используемые при моделировании наборы стандартных случайных чисел) содержат трудоемкие (для вычисления на компьютере) операции.

Рассмотрим функцию (мажоранту) $f^{(1)}(u) \geq f(u), u \in (a,b)$ такую, что:

– вычисление значений функции $f^{(1)}(u)$ на компьютере является экономичным,

– существует и является экономичным моделирующий алгоритм (формула) $\xi_0^{(1)} = \psi_{\xi^{(1)}}(\bar{\alpha}_0), \bar{\alpha}_0 \in U(0,1)$ получения выборочного значения $\xi_0^{(1)}$ случайной величины $\xi^{(1)}$, имеющей плотность распределения (5).

Рассмотрим также миноранту $f^{(2)}(u) \leq f(u), u \in (a,b)$, вычисление значений которой является экономичным.

АЛГОРИТМ 1 (двусторонний метод исключения; см. раздел 11.6 книги [1]). *1. Моделируем выборочное значение $\xi_0^{(1)}$ согласно алгоритму (формуле) $\xi_0^{(1)} = \psi_{\xi^{(1)}}(\bar{\alpha}_1)$, а также значение $\eta_0 = \alpha_2 f^{(1)}\left(\xi_0^{(1)}\right) \in U\left(0, f^{(1)}\left(\xi_0^{(1)}\right)\right)$; здесь $\bar{\alpha}_1, \alpha_2 \in U(0,1)$. Согласно утверждению* 3, *точка $\left(\xi_0^{(1)},\eta_0\right)$ равномерно распределена в подграфике* (6) *мажоранты $f^{(1)}(u)$.*

2. Сравниваем сначала значение η_0 со значением миноранты $f^{(2)}(u)$ в точке $\xi_0^{(1)}$, т. е. проверяем неравенство $\eta_0 < f^{(2)}\left(\xi_0^{(1)}\right)$. Если оно выполнено, то пара $\left(\xi_0^{(1)},\eta_0\right)$ принадлежит «подграфику» функции $f^{(2)}(u)$, а значит, и области (4). *Тогда, согласно утверждениям* 1 *и* 2, *можно положить $\xi_0 = \xi_0^{(1)}$.*

Если же $\eta_0 \geq f^{(2)}\left(\xi_0^{(1)}\right)$, то проверяем неравенство

$$\eta_0 < f\left(\xi_0^{(1)}\right). \qquad (7)$$

Если неравенство (7) *выполнено, то* $\left(\xi_0^{(1)}, \eta_0\right) \in G$*, и тогда, согласно утверждениям* 1 *и* 2*, можно положить* $\xi_0 = \xi_0^{(1)}$*. В противном случае (т. е. при* $\eta_0 \geq f\left(\xi_0^{(1)}\right)$*) повторяем пункт* 1 *и т. д.*

В случае, когда все три функции $f(u), f^{(1)}(u), f^{(2)}(u)$ близки, проверка неравенства (7), связанная с трудоемким вычислением значения $f\left(\xi_0^{(1)}\right)$, будет происходить относительно редко, и алгоритм 1 может оказаться намного эффективнее (экономичнее) других моделирующих методов (например, метода обратной функции распределения (2)).

3.3. Построение кусочно-постоянных мажорант и минорант

При выборе функций $f^{(1)}(u)$ и $f^{(2)}(u)$ из алгоритм 1 наиболее универсальным и целесообразным является использование кусочно-постоянных приближений функции $f(u)$ снизу и сверху:

$$f^{(1)}(u) \equiv A_i^{(+)}; f^{(2)}(u) \equiv A_i^{(-)}; u \in \Delta_i = (u_{i-1}; u_i];\ \ i = 1, \dots, M; \qquad (8)$$

здесь

$$a = u_0 < u_1 < u_2 < \dots < u_{M-1} < u_M = b \qquad (9)$$

и $0 \leq A_i^{(-)} \leq f(u) \leq A_i^{(+)}$ при $u \in \Delta_i$.

В свою очередь, выбор констант $A_i^{(-)}$ и $A_i^{(+)}$ существенно упрощен для случая, когда функция $f(u)$ является монотонной (кусочно-монотонной) и ограниченной на конечном интервале (a, b).

Рассмотрим, для определенности, возрастающую функцию $f(u) \uparrow$ (построение аналогов всех приведенных ниже вычислительных алгоритмов для случая монотонно убывающей функции $f(u) \downarrow$ не вызывает никаких затруднений). В этом случае имеем $A_i^{(-)} = f(u_{i-1})$ и $A_i^{(+)} = f(u_i)$.

В этом случае для мажоранты $f^{(1)}(u)$ вида (8) плотность (5) случайной величины $\xi^{(1)} \in (a, b)$ можно представить в виде составной плотности

$$f_{\xi^{(1)}}(u) = \sum_{i=1}^{M} p_i g_i(u) \chi^{(\Delta_i)}(u); \qquad (10)$$

$$p_i = \frac{f(u_i)(u_i - u_{i-1})}{\sum_{j=1}^{M} f(u_j)(u_j - u_{j-1})};\ g_i(u) = \left\{\frac{1}{u_i - u_{i-1}} \text{ при } u \in \Delta_i \text{ и } 0 \text{ иначе}\right\};$$

здесь $\chi^{(\Delta_i)}(u)$ – индикатор полуинтервала Δ_i, т. е. $\chi^{(\Delta_i)}(u) \equiv 1$ при $u \in \Delta_i$ и 0 иначе.

3.4. Уравнивание вероятностей

Для построения экономичной версии алгоритма 1 мы применили следующий прием из работы [3], позволяющий получить равные вероятности $\{p_i\}$ в соотношении (10).

Выберем сетку (9) таким образом, что площади $S_i = f(u_i)(u_i - u_{i-1})$ одинаковы для всех $i = 1, \dots, M$.

По аналогии с работой [3], отметим, что такая сетка может быть построена с помощью алгоритма «подбора», соответствующего следующей процедуре.

АЛГОРИТМ 2. *Определим процедуру-функцию от* $u_{M-1} = r$ *(наибольшего из левых концов полуинтервалов* $(u_{i-1}, u_i]$*) – обозначим ее* $z^{(rej)}(r)$ *– как следующую последовательность команд в коде* Maple*:*

$z^{(rej)}(r)$: $u_{M-1} = r$; $S = (b - r)f(b)$; **for** i **from** $M - 2$ **by** -1 **to** 1 **do** $u_i = u_{i+1} - \frac{S}{f(u_{i+1})}$ **od** ; **return** $(S - (u_1 - a)f(u_1))$

«Двигаем» (подбираем) параметр r пока с высокой точностью не будем иметь

$$z^{(rej)}(r) \approx 0. \tag{11}$$

Отметим, что в работе [3] не описаны точные процедуры получения параметра r, обеспечивающих выполнение соотношения (11) с указанной в [3] точностью $\varepsilon = 10^{-K}$ для $K = 16$. Предлагается следующая процедура приближения.

АЛГОРИТМ 3. *Выберем некоторое целое $r_0 \in (a, b)$ и вычислим $z^{(rej)}(r_0)$.*

Пусть, для определенности, $z^{(rej)}(r_0) > 0$, т. е. площадь S «излишне большая» и параметр $r = r_0$ нужно увеличить.

Возьмем $\varepsilon^{(1)} = 10^{-1}$ и начнем прибавлять эту величину к r_0 до тех пор пока первый раз не получим точку r_1 такую, что $z^{(rej)}(r_1) < 0$.

Теперь возьмем $\varepsilon^{(2)} = 10^{-2}$ и начнем вычитать эту величину из r_1 до тех пор пока первый раз не получим точку r_2 такую, что $z^{(rej)}(r_2) > 0$ и т. д.

Точка $r = r_{K+1}$ (например, для $K = 16$) обеспечит выполнение соотношения (11) *с точностью $\varepsilon = 10^{-K}$ (это можно проверить, сравнивая $z^{(rej)}(r_{K+1})$ с нулем).*

Результаты построения неравномерной сетки для уравнивания вероятностей с помощью алгоритмов 2 и 3 показаны на соответствующих графиках в разделе **«Сетка»** системы **EDSRM**.

3.5. Выбор параметров алгоритма 1 с мажорантой и минорантой (8), (9)

Особо отметим, что моделирование значения $\xi_0^{(1)}$ из первого пункта алгоритма 1 согласно плотности (10) с приравненными, согласно алгоритмам 2 и 3, вероятностями $\{p_i\}$ может быть реализовано по рекордно экономичным формулам

$$m = [M\alpha_1] + 1;\ \xi_0^{(1)} = u_{m-1} + (u_m - u_{m-1})(M\alpha_1 - m + 1); \alpha_1 \in U(0,1), \tag{12}$$

что и обеспечивает эффективность следующей версии алгоритма 1.

АЛГОРИТМ 4. *Предварительно формируем массив $\boldsymbol{U} = \{a = u_0, u_1, u_2, \dots, u_{M-1}, u_M = b\}$ узлов сетки* (9) *согласно алгоритмам 2 и 3, а также массивы минимумов и максимумов функции $f(u)$ из формулы* (3) *на элементах разбиения Δ_i* из формулы (8)*:*

$$\boldsymbol{MIN} = \left\{A_1^{(-)} = f(u_0), A_2^{(-)} = f(u_1), \dots, A_M^{(-)} = f(u_{M-1})\right\} \text{ и}$$
$$\boldsymbol{MAX} = \left\{A_1^{(+)} = f(u_1), A_2^{(+)} = f(u_2), \dots, A_M^{(+)} = f(u_M)\right\}.$$

В свою очередь, для реализации каждого выборочного значения ξ_0 случайной величины ξ, *распределенной согласно монотонно возрастающей плотности* (3)*, реализуем на компьютере следующие действия.*

1. Моделируем выборочное значение $\xi_0^{(1)}$ согласно формулам (12)*, а также значение $\eta_0 = \alpha_2 A_m^{(+)} \in U\left(0, A_m^{(+)}\right)$; здесь $\alpha_2 \in U(0,1)$. Согласно утверждению* 3*, точка $\left(\xi_0^{(1)}, \eta_0\right)$ равномерно распределена в подграфике* (6) *кусочно-постоянной функции $f^{(1)}(u)$.*

2. Сравниваем сначала значение η_0 с соответствующим значением кусочно-постоянной миноранты $f^{(2)}(u)$, т. е. проверяем неравенство $\eta_0 < A_m^{(-)}$. Если оно выполнено, то пара $\left(\xi_0^{(1)}, \eta_0\right)$ принадлежит подграфику функции $f^{(2)}(u)$, а значит, и области (4)*. Тогда, согласно утверждениям* 1 *и* 2*, можно положить $\xi_0 = \xi_0^{(1)}$.*

3. Если же $\eta_0 \geq A_m^{(-)}$*, то проверяем неравенство* $\eta_0 < f\left(\xi_0^{(1)}\right)$*. Если оно выполнено, то* $\left(\xi_0^{(1)}, \eta_0\right) \in G$*, и тогда, согласно утверждениям* 1 *и* 2*, можно положить* $\xi_0 = \xi_0^{(1)}$.

В противном случае (т. е. при $\eta_0 \geq f\left(\xi_0^{(1)}\right)$*) повторяем пункт* 1 *и т. д. – до получения значения* ξ_0.

Для широкого класса распределений удалось найти оптимальное (обеспечивающее примерно одинаковое время работы алгоритма 1 с формулой (12) – для всех рассмотренных распределений) значение $M = 330$ (это количество узлов вводимой – для построения кусочно-постоянных мажоранты и миноранты (8), (9) – неравномерной сетки). Впрочем, в разделе **«Параметризация»** системы **EDSRM** предусмотрена возможность уточнить выбираемое значение параметра M для того или иного конкретного распределения с заданной монотонной плотностью.

Особо отметим, что пользователи системы могут воспользоваться кодами алгоритма 4 и его аналогов для убывающих и кусочно-монотонных плотностей (3), которые собраны в **библиотеке системы EDSRM** и снабжены соответствующими комментариями.

ЛИТЕРАТУРА

1. Cherkashin D. A., Voytishek A. V. Using the inverse distribution function method and the modified superposition method in the NMPUD computational system // Journal of Physics: Conference Series. – 2021. – Vol. 2099, No. 012071

2. Войтишек А. В. Лекции по численным методам Монте-Карло. Новосибирск: ИПЦ НГУ, 2018.

3. Marsaglia G., Tsang W. W. The ziggurat method for generating random variables // Journal of Statistical Software. – 2000. – Vol. 5. – Issue 8.

Автор и администратор системы EDSRM

(с ним следует консультироваться по вопросам использования системы)

Брызгалов Виктор Леонидович – стажер ЛММ; e-mail: finepowerc4acc@gmail.com

Научный руководитель разработки системы EDSRM

Войтишек Антон Вацлавович – заведующий лабораторией математического моделирования (ЛММ) лицея № 130; e-mail: vav@osmf.sscc.ru