

ОБ ОПТИМАЛЬНОМ ОБЪЕМЕ ВЫБОРКИ ПРИ ОЦЕНКЕ ФРАКТАЛЬНОЙ РАЗМЕРНОСТИ РАНДОМИЗИРОВАННЫХ СИСТЕМ ИТЕРИРОВАННЫХ ФУНКЦИЙ

П. В. Москалев¹, А. Г. Буховец²

¹ Московский государственный технологический университет “СТАНКИН”,
Россия, Москва, 127055, Вадковский пер., 1

² Воронежский государственный аграрный университет имени Императора Петра I,
Россия, Воронеж, 394087, ул. Мичурина, 1
E-mail: moskaleff@mail.ru; abuhovets@mail.ru

Поиск общих подходов к описанию физических процессов и эффективных характеристик в таких гетерогенных системах как аэрозоли, пены, эмульсии, золи и пористые тела привели к выделению теории дисперсных систем в особый раздел, сначала в физической химии, а затем и в математическом моделировании. В качестве одного из эффективных средств моделирования пространственной структуры таких систем, начиная с последних десятилетий XX века могут рассматриваться рандомизированные системы итерированных функций (RIFS), описанные М. Барнсли в 1988 году [1].

Как показано в [2] для построения реализации RIFS применимы два алгоритма, которые условно можно обозначить как последовательный и параллельный. Последовательный алгоритм основан на формуле деления отрезка в заданном отношении и записывается в виде итерационной схемы вида:

$$x_{i+1} = \frac{x_i + \mu z_{ij}}{1 + \mu}, \quad \text{для } i = 1, 2, \dots, N, \quad (1)$$

где $x_i \in X \subset \mathbf{R}^n$ – простая бесповторная выборка N точек множества X , которое можно считать реализацией РСИФ; $z_{ij} \in Z \subset \mathbf{R}^n$ – выборка с повторениями N точек множества Z , соответствующая закону распределения $P = \{p_j | z_j\}$ при том, что $p_1 + p_2 + \dots + p_{km} = 1$; $\mu \in \mathbf{R}^+$ – положительный параметр RIFS, соответствующий коэффициенту разбиения итерационного отрезка.

На рис. 1 показаны примеры выборочных реализаций X_{km} , порождаемых равномерно распределенными ($p_1 = p_2 = \dots = p_k = 1/km$) квадратными множествами Z_{km} , с различным числом точек $km = 4, 8, 12$ при различных значениях параметра $\mu = 1, 2, 3$.

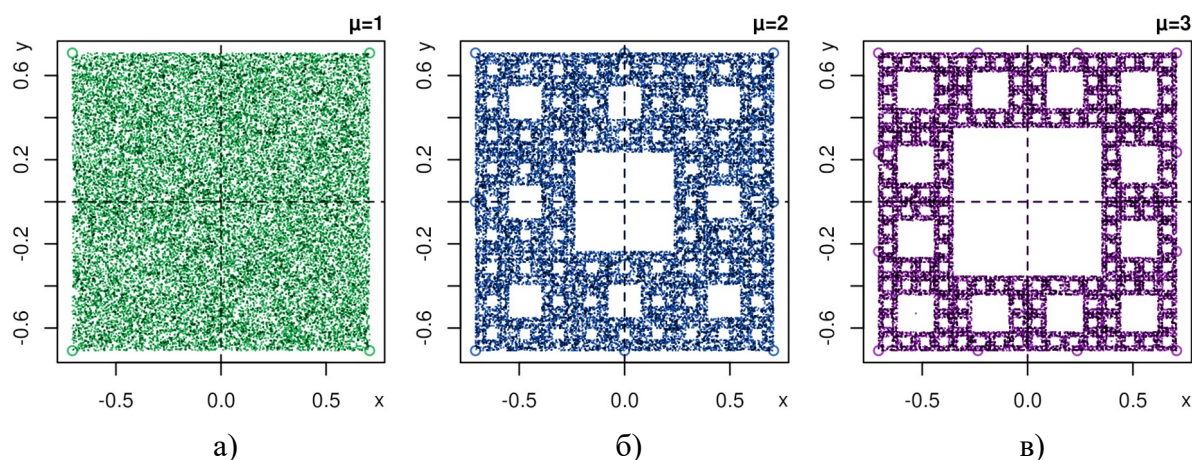


Рис. 1. Выборочные реализации X объемом $N = 5 \cdot 10^4$ точек для различных множеств Z и значений параметра μ : а) $X_{41}(Z_{41} | \mu = 1)$; б) $X_{42}(Z_{42} | \mu = 2)$; в) $X_{43}(Z_{43} | \mu = 3)$

Все показанные на рис. 1 реализации X_{km} построены с помощью функции «pre-RIFS()» из пакета «RIFS», выпущенного авторами для системы R на условиях свободной лицензии GNU GPL-3 [3]. Указанные реализации X_{km} соответствуют выборочным множествам из $N = 5 \cdot 10^4$ точек, равномерно заполняющих квадраты Серпинского первого, второго и третьего порядков, размерности подобия которых d_S для реализаций X_{km} определяются как [4]:

$$d_S = -\frac{\ln km}{\ln s} = -\frac{\ln k + \ln m}{\ln \frac{1}{1+\mu}}, \quad (2)$$

где km – число точек порождающего множества Z ; s – коэффициент подобия, связанный с параметром RIFS соотношением $s = 1/(1 + \mu)$.

Нетрудно видеть, что определение (2) дает априорную оценку фрактальной размерности множеств, аппроксимируемых выборочными реализациями X_{km} , которые могут иметь различный объем N . В связи с этим возникает вопрос об оценке объема выборки N , оптимального для достижения заданной точности аппроксимации фрактального множества. В качестве метрики качества воспользуемся апостериорными оценками фрактальной размерности множеств X_{km} по методу клеточной размерности d_B [5].

Все анализируемые реализации множеств X_{km} сохранялись в виде графических файлов формата PNG в цветовом пространстве оттенков серого с разрешением 8 бит на пиксель и пространственным разрешением 1024×1024 пикселей. Для получения оценок d_B использовалось программное обеспечение ImageJ, распространяемое на условиях свободной лицензии GNU GPL-3 [6].

Корреляционные поля оценок клеточной размерности $d_B(N|X_{km})$ в зависимости от объема выборки $N = \{1, 2, 3, 5, 8, 13, 22, 36, 60, 100\} \times 10^4$ точек для различных реализаций РСИФ показаны на рис. 2 следующими символами: X_{31} – «красный круг»; X_{32} – «желтый квадрат»; X_{41} – «зеленый ромб»; X_{42} – «синий треугольник вверх»; X_{43} – «фиолетовый треугольник вниз». Тонкими сплошными горизонтальными линиями соответствующих цветов показаны априорные оценки размерности подобия $d_S(X_{km})$, сделанные в соответствии с (2). Штриховыми горизонтальными линиями соответствующих цветов показаны уровни $0,975 d_S(X_{km})$, используемые для оценки нижних значений объемов выборок, оптимальных для получения апостериорных оценок $d_B(N|X_{km})$.

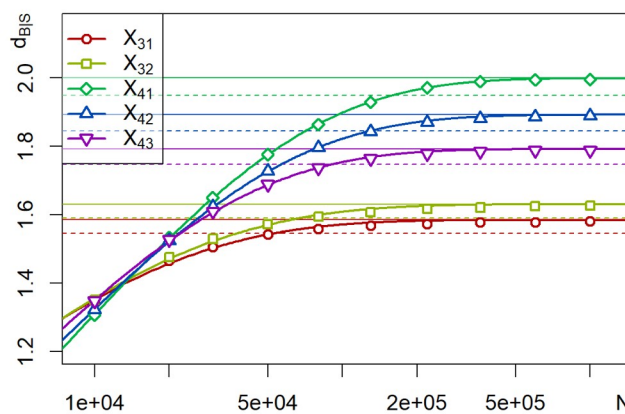


Рис. 2. Сходимость апостериорных оценок $d_B(N|X_{km})$ к априорным $d_S(X_{km})$ в зависимости от объема выборки N для реализаций RIFS: X_{31} , X_{32} , X_{41} , X_{42} , X_{43}

Все показанные реализации с ростом числа точек N демонстрируют сходимость апостериорных оценок $d_B(N|X_{km})$ к априорным $d_S(X_{km})$, что позволяет сформулировать первую эмпирическую гипотезу:

$$\lim_{N \rightarrow +\infty} d_B(N) = d_S - . \quad (3)$$

Предельное соотношение (3), с учетом проведенных вычислительных экспериментов, также позволяет сформулировать вторую эмпирическую гипотезу о применимости показательно-степенной аппроксимации оценок $d_B(N|X_{km})$ в виде:

$$d_{Bi,j} = d_{Sj} \left[1 - \exp \left(-\frac{N_i^{b_2}}{b_1} \right) \right] + e_i, \quad (4)$$

где $b = (b_1, b_2)$ – вектор масштабных коэффициентов; e_i – вектор аддитивных ошибок показательно-степенной модели (4), минимизируемый нелинейным методом наименьших квадратов. Нетрудно проверить, что лежащая в основе модели (4) функция проходит через начало координат, монотонно возрастает при $N > 0$ и ограничена сверху горизонтальной асимптотой $d_B \rightarrow d_S$ при $N \rightarrow +\infty$.

Аппроксимации вида (4) для различных реализаций RIFS X_{km} показаны на рис. 2 утолщенными сплошными линиями соответствующих цветов. Для оценки параметров регрессионной модели использовалась функция «`gsl_nls()`» из пакета «`gslnls`», выпущенного под лицензией GNU GPL-3 для системы R [7]. Пример результата, полученного при построении аппроксимации вида (4) для оценок клеточной размерности d_{Bi} показанной на рис. 1б реализации квадратного множества Серпинского X_{42} при различных объемах выборки N_i , приведены в следующем листинге.

```
Formula: db42 ~ ds42 * (1 - exp(-N^b2/b1))

Parameters:
  Estimate Std. Error t value Pr(>|t|)
b1 46.557341   1.994288   23.34 1.20e-08 ***
b2  0.437154   0.004275  102.27 9.34e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.003599 on 8 degrees of freedom

Number of iterations to convergence: 23
Achieved convergence tolerance: 1.368e-16
```

В целом, качество построенной аппроксимации можно охарактеризовать как хорошее. Если использовать в качестве метрики качества остаточное стандартное отклонение (RSE), то можно показать, что радиусы доверительных интервалов для прогнозируемых значений $d_B(N = 5 \cdot 10^4 | \{X_{41}, X_{42}, X_{43}\})$ не будут превышать 1,5%.

Анализируя представленные результаты можно увидеть, что для получения апостериорных оценок клеточной фрактальной размерности $d_B(N|X_{km})$ для квадратных множеств Серпинского X_{41}, X_{42}, X_{43} с погрешностью, не превышающей 2,5%, необходимо использовать объем выборки $N \geq 2 \cdot 10^5$ точек. В то же время для получения апостериорных оценок клеточной фрактальной размерности $d_B(N|X_{km})$ для треугольных множеств Серпинского X_{31}, X_{32} с погрешностью, не превышающей 2,5%, вполне можно ограничиться вдвое меньшим объемом выборки $N \geq 1 \cdot 10^5$ точек.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (проект № FSFS-2024-0007).

Список литературы

1. Barnsley M. Fractals Everywhere. – New York: Academic Press, 1993. – 534 p. – DOI: [10.1016/C2013-0-10335-2](https://doi.org/10.1016/C2013-0-10335-2).
2. Bukhovets A. G., Bukhovets E. A. Modeling of fractal data structures // Automation and Remote Control. – 2012. – V. 73. – P. 381-385. – DOI: [10.1134/S0005117912020154](https://doi.org/10.1134/S0005117912020154).

3. Moskalev P. V., Bukhovets A. G., Biruchinskay T. Ya. RIFS: Random Iterated Function System. – CRAN: Contributed Packages, 2022. – DOI: [10.32614/CRAN.package.RIFS](https://doi.org/10.32614/CRAN.package.RIFS).
4. Москалев П. В., Буховец А. Г. О размерности подобию рандомизированной системы итеративных функций // Компьютерные исследования и моделирование. – 2012. – Т. 4, № 4. – С. 681-691. – DOI: [10.20537/2076-7633-2012-4-4-681-691](https://doi.org/10.20537/2076-7633-2012-4-4-681-691).
5. Москалев П. В. Перколяционное моделирование пористых структур. – М.: URSS, 2018. – 240 с. – EDN: [ZRJSWD](https://www.edn.ru/ZRJSWD).
6. ImageJ: Image Processing and Analysis in Java. – URL: <https://imagej.net/ij/> (online: 29.11.2024).
7. Chau J. gslnls: GSL Multi-Start Nonlinear Least-Squares Fitting. – CRAN: Contributed Packages, 2024. – DOI: [10.32614/CRAN.package.gslnls](https://doi.org/10.32614/CRAN.package.gslnls).