

Пакет для численного обращения характеристических функций `cfinversion`

Имамутдинова Л. Р., Национальный исследовательский университет «Высшая школа
экономики», Санкт-Петербург irimamutdinova@edu.hse.ru,
Михайлов М.Д., Санкт-Петербургский государственный университет
mikhailovmikhaild@yandex.ru¹

Аннотация

Данная работа посвящена пакету `cfinversion`, предоставляющему набор различных методов численного обращения характеристических функций для абсолютно непрерывных распределений, а также анализу этих методов применительно к некоторым распределениям.

Введение

Характеристические функции (х.ф.) являются важным инструментом для работы с вероятностными распределениями. Для случайной величины X с функцией распределения F_X её характеристическая функция определяется как

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{+\infty} e^{itx} dF_X(x).$$

Характеристическая функция всегда существует и однозначно задает распределение случайной величины. В случае, когда X имеет абсолютно непрерывное распределение с плотностью $f_X(x)$, характеристическая функция равна преобразованию Фурье плотности $f_X(x)$.

В ряде задач необходимо уметь восстанавливать плотность или функцию распределения по заданной характеристической функции. Примерами таких задач служат вычисление распределений времени ожидания в различных моделях теории очередей [1], расчет хвостовых вероятностей в распределениях из финансовой математики [3] и расчет *p*-value в некоторых статистических тестах [4].

В абсолютно непрерывном случае существует несколько вариаций формулы обращения характеристической функции. Общая формула обращения для восстановления плотности $f_X(x)$ по характеристической функции $\varphi_X(t)$ приведена в [6]. Для достаточно широкого класса абсолютно непрерывных

¹Работа выполнена при поддержке СПбГУ (Pure ID 116636233)

распределений имеют место формулы обращения (1) и (2):

$$F_X(x) = \frac{1}{2} - \frac{1}{2\pi} \cdot \text{v.p.} \int_{-\infty}^{+\infty} \frac{e^{-itx} \varphi_X(t)}{it} dt, \quad (1)$$

$$f_X(x) = \frac{1}{2\pi} \cdot \text{v.p.} \int_{-\infty}^{+\infty} e^{-itx} \varphi_X(t) dt. \quad (2)$$

Формула (1) верна для всех а.н. распределений [8], формула (2) верна только когда соответствующее главное значение существует.²

Не всегда возможно провести обращение х.ф. аналитически, поэтому возникает потребность в применении численных методов. На данный момент существует несколько специализированных инструментов с открытым исходным кодом, которые позволяют решать задачу обращения х.ф.:

- библиотека `CharFunTool`, написанная на R. Также имеет реализацию на MATLAB. В этой библиотеке реализованы только два метода для обращений характеристических функций абсолютно непрерывных распределений: на основе обычного численного интегрирования и на основе DFT;
- библиотеки для дискретного преобразования Фурье (DFT) или его вариантов, которые можно использовать для аппроксимации интеграла в формуле обращения;
- библиотеки для численного интегрирования, которые содержат специальные алгоритмы для вычисления интегралов возникающих в формуле обращения (например QAWF в QUADPACK).

Обращение преобразования Фурье доступно также в проприетарных математических пакетах MATLAB и Wolfram, а библиотеки ISML и NAG предлагают специальные квадратуры для вычисления интегралов вида (1) и (2). В пакете SAS обращение х.ф. недоступно в качестве самостоятельной процедуры, но используется, например, при проверке временных рядов на стационарность.

Несмотря на то, что процедура QAWF и её аналоги подходят для численного обращения характеристической функции, более продвинутые методы обращения х.ф., предложенные в работах [1] и [2], на данный момент не имеют реализации. Следует также отметить, что процедура QAWF и её аналоги

²В общей формуле обращения из [6] интеграл в (2) понимается в смысле суммирования по Чезаро.

не оптимизированы для восстановления значений плотности/функции распределения сразу в нескольких точках. Поэтому разработка инструмента для обращения характеристических функций является актуальной задачей.

Описание пакета `cfinversion`

Пакет `cfinversion` написан на Python в рамках проекта PySATL. В пакете реализованы методы из пакета `CharFunTool` в соответствии с [1] («наивное» численное интегрирование) и [5] (на основе DFT). Также реализованы пять методов расчета функции распределения, предложенные в работе Г. Бохмана [2], которые базируются на техниках из теории аппроксимации. Для вычисления плотности с помощью этих методов в пакете используется численное дифференцирование.

Работу методов можно понимать как вычисление интегралов в (1) и (2) с помощью численных методов, заменяя интегралы по \mathbb{R} на интегралы по отрезку $[-T; T]$. Метод «наивного» численного интегрирования применяет правило трапеции к этим формулам с $2N + 1$ точкой и шагом δ . В [1] отмечается, что для интегралов такого вида правило трапеции подходит лучше, чем другие правила.

Методы Бохмана используются для вычисления функции распределения, плотность вычисляется через численное дифференцирование функции распределения. Все реализованные методы предполагают, что $\varphi''(0)$ существует. Для простоты реализации внутренних вычислений характеристические функции стандартизируются с помощью преобразования

$$\varphi_X(t) \mapsto \exp\left(-it\frac{\mu}{\sigma}\right) \varphi_X\left(\frac{t}{\sigma}\right),$$

где $\mu = i\varphi'(0)$ и $\sigma = \sqrt{-\varphi''(0) - \mu^2}$. Величины μ и σ вычисляются с помощью численного дифференцирования на основе приближений, представленных в работе [7] или могут быть заданы пользователем при инициализации метода. Заметим, что хотя формально для корректной работы методов требуется существование $\varphi''(0)$, в [2] указывается, что эти методы также могут работать с распределениями с более тяжелыми хвостами.

В предположении, что $\varphi'_X(0) = 0$ и $\varphi''_X(0) = 1$, методы Бохмана работают следующим образом (запись \sum' означает, что слагаемое, отвечающее нулевому индексу, отсутствует).

А. Вычисление суммы Римана (3) для оценки интеграла в формуле (1)

$$F_X(x) \approx \frac{1}{2} + \frac{\delta x}{2\pi} - \sum_{n=-N}'^N \frac{\varphi_X(\delta n)}{2\pi i n} e^{-i\delta n x}. \quad (3)$$

В. Обращение х.ф. $\varphi_X(t)C(t/T)$, где $C(t)$ — специальная «срезающая» характеристическая функция с носителем на $[-1; 1]$:

$$F_X(x) \approx \frac{1}{2} + \frac{\delta x}{2\pi} - \sum_{n=-N}'^N C\left(\frac{n}{N}\right) \frac{\varphi_X(\delta n)}{2\pi i n} e^{-i\delta n x}.$$

С. Вычислять разность функции распределения $F_X(x)$ и функции распределения стандартного нормального закона $\Phi(x)$. В этом методе функция распределения вычисляется как $F_X(x) \approx \Phi(x) + H_N(x, \delta)$, где

$$H_N(x, \delta) = \sum_{n=-N}'^N \frac{\exp(-\frac{1}{2}\delta^2 n^2) - \varphi_X(\delta n)}{2\pi i n} e^{-i\delta n x}.$$

Д. Использовать метод С с дополнительной поправкой на алиасинг:

$$F_X(x) \approx \Phi(x) + H_N(x, \delta) - \sum_{k=1}^{K-1} H_N(x + kd_1, \delta_1),$$

где $K, N : K$ — свободный параметр и величины d_1, δ_1 определяются через N, K и δ .

Е. Комбинация методов В и Д:

$$F_X(x) \approx \Phi(x) + G_N(x, \delta) - \sum_{k=1}^{K-1} G_N(x + kd_1, \delta_1),$$

где, как и ранее, $K, N : K$ — свободный параметр, числа d_1, δ_1 определяются через N, K и δ , и

$$G_N(x, \delta) = \sum_{n=-N}'^N C\left(\frac{n}{N}\right) \frac{\exp(-\frac{1}{2}\delta^2 n^2) - \varphi_X(\delta n)}{2\pi i n} e^{-i\delta n x}.$$

DFT-метод ускоряет вычисления в методе A, используя прием, описанный например в [5]. Полагая $\delta' = \frac{\delta}{2\pi}$, $x_m = \frac{m}{\delta'(2N+1)}$, $m = -N \dots, N$ в (3), получаем следующее:

$$F_X(x_m) \approx \frac{1}{2} + x\delta' - \sum_{n=-N}^{n=N} \frac{\varphi_X(2\pi\delta n)}{2\pi i n} e^{-2\pi i \frac{nm}{2N+1}}. \quad (4)$$

Для вычисления суммы (4) в точках x_m используется быстрое преобразование Фурье, а в остальных точках значение интерполируется. В отличие от реализации в CharFunTool, для DFT-метода в cfinversion используется кусочно-линейная интерполяция вместо РНСР-интерполяции при вычислении функции распределения. Процедура вычисления плотности полностью аналогична процедуре вычисления функции распределения.

Для возможности добавления новых методов в будущем, реализации методов следуют общему интерфейсу ContinuousInverter, который имеет методы fit, pdf и cdf. Метод fit принимает на вход характеристическую функцию и выполняет необходимые предварительные вычисления. Методы pdf и cdf возвращают значения плотности и функции распределения соответственно в заданных точках.

Асимптотическая сложность реализованных методов представлена в таблице 1. В таблице C_{cf} обозначает сложность вычисления х.ф. в одной точке, асимптотики методов cdf и pdf приведены также для вычисления в одной точке.

Метод	fit(φ)	cdf(x) / pdf(x)
Методы Бохмана A-C	$\mathcal{O}(N \cdot C_{cf})$	$\mathcal{O}(N)$
Методы Бохмана D, E	$\mathcal{O}(K \cdot N \cdot C_{cf})$	$\mathcal{O}(K \cdot N)$
Naive Integration	$\mathcal{O}(N \cdot C_{cf})$	$\mathcal{O}(N)$
DFT-метод	$\mathcal{O}(N(\log N + C_{cf}))$	$\mathcal{O}(\log N)$

Таблица 1: Асимптотическая сложность методов

Численное исследование реализованных методов

Для исследования точности методов было проведено два численных эксперимента на четырех распределениях: стандартное нормальное $\mathcal{N}(0, 1)$, Лапласа $\text{Laplace}(0, 1)$, равномерное на отрезке $U(0, 1)$ и распределение квадра-

та величины, имеющей равномерное распределение $U^2(0, 1)$. Эти распределения были выбраны, так как обладают рядом свойств, влияющими на поведение методов: быстро убывающая плотность, плотность с разрывом производной в 0, плотность с разрывами и плотность с сингулярностью соответственно.

Дизайн численных экспериментов

Первый эксперимент изучает, какие значения NLRE (negative log relative error) получаются в результате применения реализованных методов. NLRE определяется как

$$\text{NLRE}(y_{\text{estim}}(x_i), y_{\text{true}}(x_i)) = -\frac{\log_{10} |y_{\text{estim}}(x_i) - y_{\text{true}}(x_i)|}{\log_{10} |y_{\text{true}}(x_i)|},$$

где x_1, \dots, x_m — точки, в которых вычисляется плотность (функция распределения), $y_{\text{estim}}(x)$ — значение плотности (функции распределения) в точке x , вычисленное через численное обращение х.ф., $y_{\text{true}}(x)$ — теоретическое значение плотности (функции распределения) в точке x . Целая часть NLRE равна количеству корректно вычисленных значащих цифр. Для изучения поведения NLRE были выбраны следующие параметры методов: размер сетки дискретизации $N = 10^5$, шаг дискретизации $\delta = 10^{-2}$.

Во втором эксперименте изучается поведение ℓ^∞ -нормы ошибки с увеличением размера N сетки дискретизации, где под ℓ^∞ -нормой понимается величина

$$\|y_{\text{estim}} - y_{\text{true}}\| = \max_{i=1, \dots, m} |y_{\text{estim}}(x_i) - y_{\text{true}}(x_i)|.$$

Ожидается, что ℓ^∞ -норма ошибки будет уменьшаться, так как она мажорируется \sup -нормой функции ошибки $\text{err}(x) = y_{\text{true}}(x) - y_{\text{estim}}(x)$. Отметим, что в общем случае \sup -норма ошибки может сходиться к 0, в то время как минимум NLRE будет оставаться постоянным.

В этом эксперименте размер N сетки дискретизации последовательно принимает значения:

$$10^3; 2.5 \cdot 10^3; 5 \cdot 10^3; 7.5 \cdot 10^3; 10^4; 2.5 \cdot 10^4; 5 \cdot 10^4; 7.5 \cdot 10^4; 10^5.$$

При всех N шаг дискретизации составляет $\delta = 10^{-2}$.

В обоих экспериментах $m = 500$ и точки x_i образуют равномерную сетку либо на отрезке $[-2; 2]$ в случае распределений $\mathcal{N}(0, 1)$ и $\text{Laplace}(0, 1)$, либо на отрезке $[0.01; 0.99]$ в случае распределений $U(0; 1)$ и $U^2(0; 1)$.

Результаты численного исследования

На рис. 1, 2, 3 и 4 для каждого метода изображены значения NLRE при вычислении плотности и функции распределения в точках x_i .

Для нормального распределения (рис. 1), наивное численное интегрирование позволяет достичь точности порядка 16 значащих цифр, что соответствует погрешности машинной арифметики. При этом методы Бохмана дают такой же результат для функции распределения, но для плотности распределения значительно уступают наивному интегрированию, видимо, из-за численного дифференцирования. Для распределения Лапласа (рис. 2) при вычислении плотности разница между методами незначительна, однако для вычисления функции распределения метод на основе DFT, а также методы Бохмана В и Е оказываются хуже на несколько значащих цифр. С другой стороны для равномерного распределения и его квадрата (рис. 3 и рис. 4 соответственно) методы В и Е, наоборот, демонстрируют более высокую точность. При этом отсутствие сильных осцилляций на графиках NLRE для методов В и Е указывает на то, что эти методы выдают менее осциллирующие результаты. Отметим также, что точность метода на основе DFT при вычислении функции распределения для $U^2(0, 1)$ сравнима с точностью методов В и Е, а для распределения $U(0, 1)$ метод на основе DFT уступает методу Е на одну значащую цифру в среднем.

Поведение ℓ^∞ -норм ошибок при росте количества точек дискретизации изображено на рис. 5, 6, 7 и 8 в логарифмических осях. Результаты согласуются с первым экспериментом: для восстановления плотности распределения, в случае когда она непрерывна, методы Бохмана В и Е, которые демонстрируют меньшую точность. С другой стороны, для распределений с разрывами и сингулярностями в плотности методы В и Е могут давать выигрыш на один-два порядка при вычислении плотности при достаточно большом размере сетки дискретизации (см. рис. 7 и 8).

В целом можно заключить, что метод на основе DFT, предложенный в [5], хоть и асимптотически быстрее, но может показывать меньшую точность по сравнению с методами Бохмана. С другой стороны, метод на основе прямого численного интегрирования, рассмотренный в [1], проще в реализации и дает наилучшие результаты в случае, когда плотность непрерывна. Однако если в плотности есть разрывы или сингулярности, то наилучшую точность можно получить, используя методы Бохмана В и Е для вычисления плотности. Для вычисления же функции распределения лучше также использовать эти методы или метод на основе DFT.

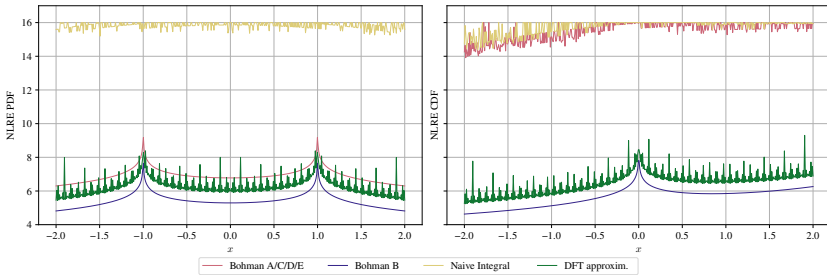


Рис. 1: Разброс NLRE при вычислении плотности и ф.р. для $\mathcal{N}(0, 1)$ на $(-2; 2)$.

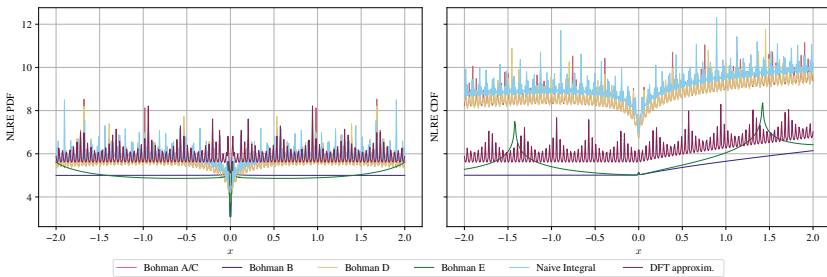


Рис. 2: Разброс NLRE при вычислении плотности и ф.р. для $\text{Laplace}(0, 1)$ на $(-2; 2)$.

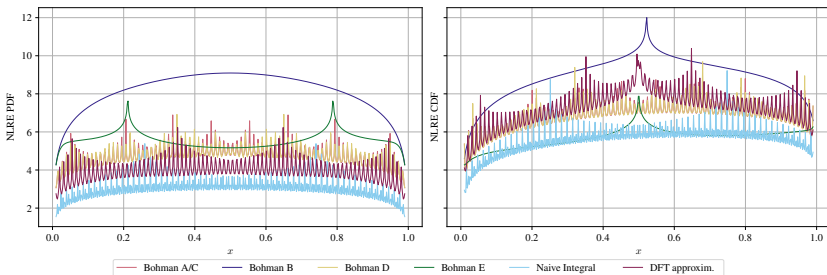


Рис. 3: Разброс NLRE при вычислении плотности и ф.р. для $U(0, 1)$ на $(0.01; 0.99)$.

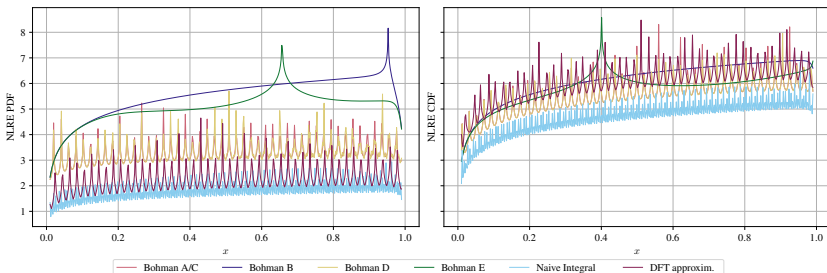


Рис. 4: Разброс NLRE при вычислении плотности и ф.р. для $U^2(0, 1)$ на $(0.01; 0.99)$.

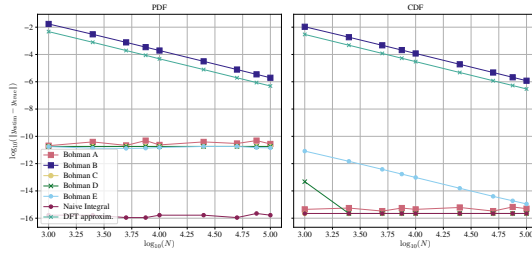


Рис. 5: Поведение ℓ^∞ -нормы ошибки при увеличении N для $\mathcal{N}(0, 1)$.

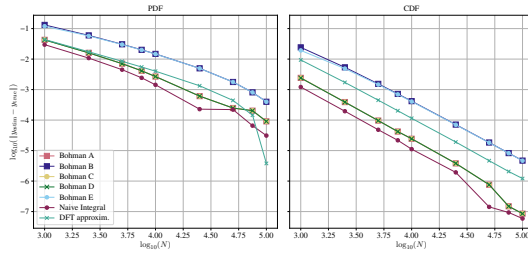


Рис. 6: Поведение ℓ^∞ -нормы ошибки при увеличении N для $\text{Laplace}(0, 1)$.

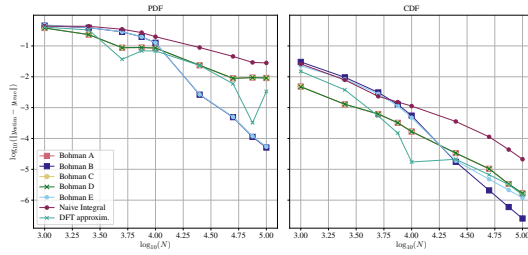


Рис. 7: Поведение ℓ^∞ -нормы ошибки при увеличении N для $U(0, 1)$.

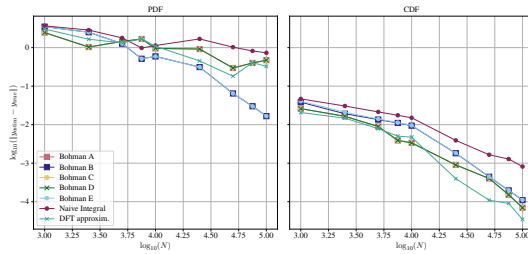


Рис. 8: Поведение ℓ^∞ -нормы ошибки при увеличении N для $U^2(0, 1)$.

Заключение

В рамках работы был реализован и опубликован³ пакет `cfinversion`, включающий в себя методы для численного обращения характеристических функций абсолютно непрерывных распределений, предложенные в [2] и [5]. Для реализованных методов проведены численные эксперименты, показавшие, что методы В и Е, предложенные в [2], дают более точные результаты по сравнению с методами на основе дискретного преобразования Фурье и простым численным интегрированием при вычислении значений плотности, имеющей разрывы или точки сингулярности.

Список литературы

- [1] Abate J., Whitt W. The Fourier-series method for inverting transforms of probability distributions // *Queueing Syst.* 1992. Т. 10. № 1–2. С. 5–87.
- [2] Bohman H. Numerical inversions of characteristic functions // *Scandinavian Actuarial Journal*. Informa UK Limited, 1975. Т. 1975, № 2. С. 121–124.
- [3] Feng L., Lin X. Inverting Analytic Characteristic Functions and Financial Applications // *SIAM J. Finan. Math.* 2013. Т. 4. № 1. С. 372–398.
- [4] Witkovský V. Computing the exact distribution of the Bartlett's test statistic by numerical inversion of its characteristic function // *Journal of Applied Statistics*. 2019. Т. 47. № 13–15. С. 2749–2764.
- [5] Witkovsky V. Numerical inversion of a characteristic function: An alternative tool to form the probability distribution of output quantity in linear measurement models // *Acta IMEKO*. – 2016. – Т. 5. – №. 3. – С. 32–44.
- [6] Lukacs E. Inversion formulae for characteristic functions of absolutely continuous distributions // *The American Mathematical Monthly*. – 1964. – Т. 71. – №. 1. – С. 44–47.
- [7] Witkovský V. On the exact computation of the density and of the quantiles of linear combinations of t and F random variables // *Journal of Statistical Planning and Inference*. – 2001. – Т. 94. – №. 1. – С. 1–13.
- [8] Gil-Pelaez J. Note on the inversion theorem // *Biometrika*. – 1951. – Т. 38. – №. 3–4. – С. 481–482.

³<https://github.com/PySATL/pysatl-cfinversion>, дата обращения 2025-06-05