

Использование экстраполяции Ричардсона для повышения точности обработки и анализа эмпирических данных

О. А. Попова

Сибирский федеральный университет, Красноярск, Россия, e-mail: OlgaArc@yandex.ru

Исследованы свойства эмпирических данных в условиях случайной неопределённости. Предложен новый подход для повышения точности в задачах построения функции плотности вероятности и оценки её погрешности. Метод основан на использовании экстраполяции Ричардсона и правила Рунге для ядерных оценок с различными параметрами сглаживания. Показано, что применение правила Рунге позволяет оценить погрешность ядерных оценок для функции плотности вероятности и значение её второй производной.

Ключевые слова: ядерные оценки, повышение точности, экстраполяция Ричардсона, правило Рунге, оценки производных функции плотности вероятности.

При исследовании эмпирических данных со случайной неопределённостью существенное место занимает построение оценок функции плотности вероятности измеряемого показателя. Оценка функции плотности вероятности и её производных на эмпирических данных – один из наиболее важных вопросов в области прикладных исследований [1–4]. В данной статье рассмотрен новый подход к повышению точности в задачах построения аппроксимации функции плотности вероятности и оценке её погрешности. Подход основан на комбинации ядерных оценок с различными параметрами сглаживания. Автор применяет экстраполяцию Ричардсона и правило Рунге для вычисления второй производной оценки функции плотности вероятности [5]. В отличие от известных методов, предложенный подход не требует дифференцирования ядерных оценок или вычисления разностных производных эмпирической функции плотности вероятности. Подробный обзор существующих методов оценки производных и библиография представлены в работе [6].

Использование оценок вторых производных даёт возможность получить реалистичные оценки математических ожиданий в норме l_2 погрешности аппроксимации функции плотности вероятности. Это позволяет рассчитать оптимальный параметр сглаживания для ядерных оценок функции плотности вероятности.

Экстраполяция Ричардсона – мощный подход, эффективно применяемый для повышения производительности численных расчётов математических моделей и получения результатов высокой точности по формулам низкого порядка [7, 8]. Он был представлен Ричардсоном в начале XX века, после чего многие учёные и инженеры стали его использовать для решения разнообразных практических проблем. Например, этот метод применяется при моделировании задач математической физики. Экстраполяция Ричардсона используется при решении численными методами дифференциальных уравнений в частных производных, в разностных методах решения задач Коши в системах обыкновенных дифференциальных уравнений (ОДУ) и краевых задач для дифференциальных

уравнений в частных производных [7–9]. Экстраполяцию Ричардсона можно также применять для оценки и проверки вычислительных ошибок [10].

Экстраполяция Ричардсона основана на разложении по степеням приближённого решения уравнения u^h как суммы вида

$$u^h = u + h^k v + O(h^{k+m}), \quad (1)$$

где u – искомое точное решение; h – малый параметр дискретизации, т. е. шаг разностной сетки; v – неизвестная функция; O – некоторая функция, имеющая порядок малости $k+m$; k – порядок точности приближённого решения; $m > 0$ – порядок точности приближённого решения с учётом слагаемого $h^k v$.

Поскольку параметры u , v не зависят от h , то для параметра $h/2$ справедливо разложение

$$u^{h/2} = u + (h/2)^k v + O(h^{k+m}). \quad (2)$$

Объединим эти разложения так, чтобы исключить ошибку порядка h^k . Умножим выражение (2) на 2^k , вычтем полученный результат из уравнения (1) и получим приближение точного решения уравнения (1) с более высоким порядком точности:

$$u = [2^k / (2^k - 1)] u^{h/2} - [1 / (2^k - 1)] u^h + O(h^{k+m}).$$

Одно из первых правил для практической оценки погрешности было предложено К. Рунге в начале XX века. Сначала его широко применяли в области квадратурных вычислений, а затем – в методе конечных элементов и разностных методах [6].

Для приближённой оценки погрешности вычтем выражение (2) из уравнения (1), чтобы избавиться от неизвестного значения точного решения u , и определим главный член погрешности:

$$\begin{aligned} u^h - u^{h/2} &= v(h/2)^k (2^k - 1) + O(h^{k+m}); \\ u^{h/2} - u &\approx (u^h - u^{h/2}) / (2^k - 1). \end{aligned} \quad (3)$$

Поскольку в формуле (3) отброшен остаточный член порядка $O(h^{k+m})$, то это выражение не приведёт к гарантированной оценке погрешности, но при достаточно малых значениях h даст представление о погрешности численного решения [11]. Особенности ядерных оценок функций плотности вероятности заключаются в том, что параметр h нельзя выбрать произвольно малым [6]. Следовательно, в случаях произвольных h надо увеличивать число слагаемых разложения вида (2).

Теоретические аспекты экстраполяции Ричардсона. Рассмотрим некоторый процесс вычисления $y(h)$, зависящий от параметра $h>0$. Для этого процесса справедливо разложение в ряд по параметру $h>0$:

$$y(h) = y(0) + c_1 h^{p_1} + c_2 h^{p_2} + c_3 h^{p_3} + \dots, \quad (4)$$

где $y(0)$ – точное решение; c_i – константы, независимые от h ; p_i – целые числа, $i=1, 2, \dots$

Решим задачу (4), подставив параметр h_i , $i=1, 2, \dots, l$:

$$\left. \begin{aligned} y(h) &= y(0) + c_1 h_1^{p_1} + c_2 h_1^{p_2} + c_3 h_1^{p_3} + \dots; \\ y(h) &= y(0) + c_1 h_2^{p_1} + c_2 h_2^{p_2} + c_3 h_2^{p_3} + \dots; \\ &\dots \\ y(h) &= y(0) + c_1 h_l^{p_1} + c_2 h_l^{p_2} + c_3 h_l^{p_3} + \dots + c_l h_l^{p_l}. \end{aligned} \right\} \quad (5)$$

Отбросим слагаемые порядка выше p_l и запишем систему линейных алгебраических уравнений (СЛАУ), которую можно разрешить относительно $y(0)$ и c_i , $i=1, \dots, l-1$ через известные значения $y(h_i)$, $i=1, \dots, l$. Для исследования подобных систем рассмотрим частный случай для $l=3$, $h_i = h t^{i-1}$, где $t>0$ параметр и $i=1, 2, 3$:

$$\left. \begin{aligned} s(0) + c_1 h^2 + c_2 h^4 &= s(h); \\ s(0) + c_1 h^2 t^2 + c_2 h^4 t^4 &= s(ht); \\ &\dots \\ s(0) + c_1 h^2 t^4 + c_2 h^4 t^8 &= s(ht^2). \end{aligned} \right\} \quad (6)$$

Решим систему (6) и запишем

$$\begin{aligned} s(0) &= [s(h)t^6 - s(ht)t^4 + s(ht^2)] / (t^6 + t^4 + t^2 + 1); \\ c_1 &= \{[s(ht) - s(h)]t^4 - s(ht^2) + s(h)\} / (t^6 - 2t^4 + t^2); \\ c_2 &= -\{[s(ht) - s(h)]t^2 - s(ht^2) + s(h)\} / (t^8 - t^6 - t^4 + t^2). \end{aligned}$$

Рассмотрим частный случай для $t = \sqrt{2}$ и получим

$$s(0) = [8s(h) - 6s(\sqrt{2}h) + s(2h)] / 3.$$

Если использовать только два разложения, то можно записать решение для h и $2h$ в виде:

$$s(0) = [4s(h) - s(2h)] / 3.$$

Таким образом, получено решение СЛАУ с $h, \sqrt{2}h$ и $2h$. Предложенный подход уменьшает погрешность решения СЛАУ и позволяет применять экстраполяцию Ричардсона для повышения точности ядерных оценок.

Разложение ядерных оценок для экстраполяции Ричардсона. Для использования экстраполяции Ричардсона в ядерных оценках функции плотности вероятности необходимо получить разложения общего вида для системы (5). Для этого рассмотрим свойства ядерных оценок.

1. Ядерную оценку можно записать компактно [1, 2]:

$$\begin{aligned} \hat{f}(x) &= 1/Nh \sum_{i=1}^N K[(x - \xi_i)/h] = 1/N \sum_{i=1}^N K_h(x, \xi_i); \\ K_h(x, \xi_i) &= K[(x - \xi_i)/h], \end{aligned}$$

где $K(x)$ – ядерная функция; ξ – случайная величина с плотностью вероятности $f(x)$; ξ_i – повторная выборка случайной величины ξ размерностью N .

Тогда справедливо

$$\begin{aligned} E[\hat{f}(x)] &= E[K_h(x, \xi)]; \\ \sigma_N = D[\hat{f}(x)] &= 1/ND[K_h(x, \xi)], \end{aligned}$$

где σ_N – константа.

2. Математическое ожидание можно представить в виде

$$\begin{aligned} E[K_h(x, \xi)] &= 1/h \int_{-\infty}^{\infty} K[(x-t)/h] f(t) dt = \\ &= \int_{-\infty}^{\infty} K(\eta) f(x - h\eta) d\eta. \end{aligned}$$

Отметим, что

$$\begin{aligned} f(x - h\eta) &= f(x) - hf'(x)\eta + (h^2/2)f''(x)\eta^2 + \\ &+ (h^3/6)f^{(3)}(x)\eta^3 + (h^4/24)f^{(4)}(x)\eta^4 + O(h^5); \\ E[K_h(x, \xi)]f(x) &= \int_{-A}^{\infty} K(\eta) d\eta - hf'(x) \int_{-\infty}^{\infty} \eta K(\eta) d\eta + \\ &+ (h^2/2)f''(x) \int_{-\infty}^{\infty} \eta^2 K(\eta) d\eta + (h^3/6)f^{(3)}(x) \int_{-\infty}^{\infty} \eta^3 K(\eta) d\eta + \\ &+ (h^4/24)f^{(4)}(x) \int_{-\infty}^{\infty} \eta^4 K(\eta) d\eta + O(h^5). \end{aligned}$$

Предположим, что $f \in C^{2n+2}$, а ядро K подчиняется требованиям

$$\int_{-\infty}^{\infty} K(\eta) d\eta = 1.$$

Ядра $K(\eta)=K(-\eta)$ симметричные, тогда для всех нечётных α выполнено соотношение

$$\int_{-\infty}^{\infty} \eta^{\alpha} K(\eta) d\eta = 0.$$

Обозначим

$$\int_{-\infty}^{\infty} \eta^i K(\eta) d\eta = \sigma_f^2.$$

Тогда

$$\bar{f} = E[K_h(x, \xi)] = f(x) + \sum_{i=1}^n \sigma_{2i}^2 h^{2i} f^{(2i)}(x) / (2i!) + O(h^{2n+2});$$

$$E[\hat{f}(x) - f(x)] = \sum_{i=1}^n \sigma_{2i}^2 h^{2i} f^{(2i)}(x) / 2 + O(h^{2n+2}).$$

Определим $f^h(x)$, $f^{th}(x)$:

$$f^h(x) = E[\hat{f}(x)] = f(x) + \sum_{i=1}^n \sigma_{2i}^2 h^{2i} f^{(2i)}(x) / 2 + O(h^{2n+2});$$

$$f^{th}(x) = E[\hat{f}^t(x)] = f(x) + \sum_{i=1}^n \sigma_{2i}^2 t^{2i} h^{2i} f^{(2i)}(x) / 2 + O(h^{2n+2}).$$

Применим экстраполяцию Ричардсона для $f^h(x)$ [4]. Решим СЛАУ и запишем для $t=2$ и значений h , $2h$ следующее представление:

$$f(x) = 4/3 f^h(x) - 1/3 f^{2h}(x) + O(h^4).$$

Заметим, что получена аппроксимация $f(x)$ с точностью $O(h^4)$:

$$f_{\text{cor}}^h(x) = 4/3 \hat{f}^h(x) - 1/3 \hat{f}^{2h}(x).$$

Используем значения $h, \sqrt{2}h, 2h$ и аппроксимируем $f(x)$ с точностью $O(h^6)$.

$$f_{\text{cor}}^h(x) = [8\hat{f}^h(x) - 6\hat{f}^{\sqrt{2}h}(x) + \hat{f}^{2h}(x)] / 3.$$

Оценим дисперсию $f_{\text{cor}}^h(x)$. Для этого применим формулу для суммы случайных величин

$$D\left(\sum_{i=1}^n \alpha_i f_i\right) = \sum_{i=1}^n \alpha_i^2 D(f_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(f_i, f_j).$$

Ковариацию можно оценить следующим образом:

$$|\text{cov}(f_i, f_j)| \leq \sqrt{D(f_i)D(f_j)}.$$

Дисперсии ядерных оценок $\hat{f}^h(x)$ известна из работы [2]:

$$D[\hat{f}^h(x)] = f(x) / Nh \|K\|_2^2 + [f(x)^2 / N] + O(h/N).$$

Тогда положим, что

$$D[\hat{f}^h(x)] \geq D[\hat{f}^{\sqrt{2}h}(x)];$$

$$D[\hat{f}^h(x)] < D[\hat{f}^{2h}(x)].$$

Окончательно запишем выражение

$$D[f_{\text{cor}}^h(x)] \leq CD[\hat{f}^h(x)],$$

где $C \leq (101/9+6)$.

Оценку константы C следует учитывать при выборе параметров h , N .

Правило Рунге и оценки погрешности. Используя правило Рунге, можно получить реалистичные оценки погрешности. Если применить две ядерные оценки с параметрами h , $2h$, то погрешность будет иметь точность порядка $O(h^4)$ или асимптотически точные оценки при малых значениях h . В случае достаточно гладких функций f для повышения точности оценок следует учитывать большее число членов в разложении и, соответственно, большее число ядерных оценок с параметрами h . Таким образом, с учётом ядерных оценок для трёх различных параметров $h, \sqrt{2}h, 2h$, можно вывести оценку погрешности в виде $c_1 + c_2$ с точностью порядка $O(h^6)$:

$$c_1 = \{4[\hat{f}^{\sqrt{2}h}(x) - \hat{f}^h(x)] - \hat{f}^{2h}(x) + \hat{f}^{\sqrt{2}h}(x)\} / 2;$$

$$c_2 = -\{2[\hat{f}^{\sqrt{2}h}(x) - \hat{f}^h(x)] - \hat{f}^{2h}(x) + \hat{f}^{\sqrt{2}h}(x)\} / 6.$$

Оценка производных. Пусть

$$c_i = \sigma_{2i}^2 h^{2i} f^{(2i)}(x) / (2i!), \quad i = 1, 2, \dots$$

Тогда зная, например, параметры c_i , σ_i , можно определить значение второй производной $f''(x)$

$$f''(x) = E[2c_1(x) / (\sigma_2 h^2)].$$

Численные примеры. В качестве примера рассмотрим функцию плотности вероятности для случайной величины, определяемой как сумма числа n независимых случайных величин, каждая из которых имеет равномерное распределение. Такое распределение известно в статистике и теории вероятностей как распределение Ирвин–Холла [12]. Плот-

ность вероятности суммы n равномерно распределённых случайных величин можно представить как

$$p_n(x) = 1/(n-1)! [x^{n-1} - C_n^1(x-1)^{n-1} + C_n^2(x-2)^{n-1} - \dots],$$

где C_n^k – биномиальные коэффициенты.

Для каждого фиксированного аргумента x сумма в скобках относится только к тем переменным, для которых значение $(x-k)$, $k=1, 2, \dots$ неотрицательно.

Рассмотрим пример, когда $n=6$. Тогда функция плотности вероятности соответственно при $x \in [0,1]; [1, 2]; [2, 3]; [3, 4]; [4, 5]; [5, 6]$ имеет аналитическое представление в виде

$$p(x) = \begin{cases} x^5/120; \\ \frac{-5x^5+30x^4-60x^3+60x^2-30x+6}{120}; \\ \frac{10x^5-120x^4+540x^3-1140x^2+1170x-474}{120}; \\ \frac{-10x^5+180x^4-1260x^3+4260x^2-6930x+4386}{120}; \\ \frac{5x^5-120x^4+1140x^3-5340x^2+12270x-10974}{120}; \\ \frac{-x^5+30x^4-360x^3+2160x^2-6480x+7776}{120}. \end{cases}$$

На рис. 1 показано уточнение ядерной оценки распределения Ирвина–Холла шестого порядка на интервале (2, 4). Для этого примера приведём погрешность ядерной оценки кубическими ядрами с $h=1$ и её аппроксимацию:

$$Err^2 = \int_2^4 [p(x) - \hat{f}^h(x)]^2 dx;$$

$$Err_{app}^2 = \int_2^4 [c_1(x) + c_2(x)]^2 dx.$$

Для проверки теоретических положений были проведены численные эксперименты при $N=10^6$ и получены следующие оценки математических ожиданий $E[Err]=0,0369$, $E[Err_{app}]=0,0358$. Указанные значения представляют реалистичные оценки погрешности, позволяющие судить об истинной погрешности ядерных оценок.

На рис. 2 показана оценка второй производной (кривая 1) распределения Ирвина–Холла (кривая 2) на интервале (2, 4), размерность выборки $N=10^6$.

Заключение. Представленные результаты демонстрируют новый подход для задач, связанных с проблемой повышения точности оценки функции плотности вероятности по эмпирическим данным в условиях случайной неопределённости. Это достигается за счёт численных процедур, основанных на идее экстраполяции

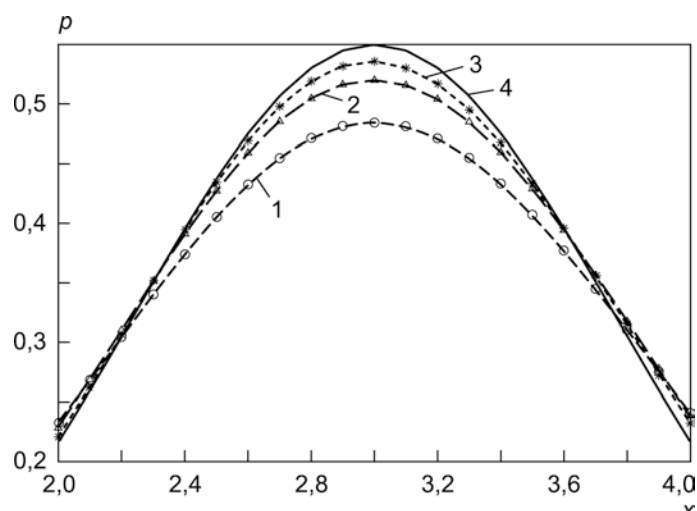


Рис. 1. Уточнение ядерной оценки:
1 – оценка кубическими ядрами с $h=1$, размерность выборки 10000; 2 – экстраполяция Ричардсона, использующая две ядерные оценки кубическими ядрами с $h=1$; 3 – уточнение, использующее три ядерные оценки кубическими ядрами с $h=1$, $t=\sqrt{2}$; 4 – точное представление функции плотности вероятности для распределения Ирвина–Холла при $n=6$

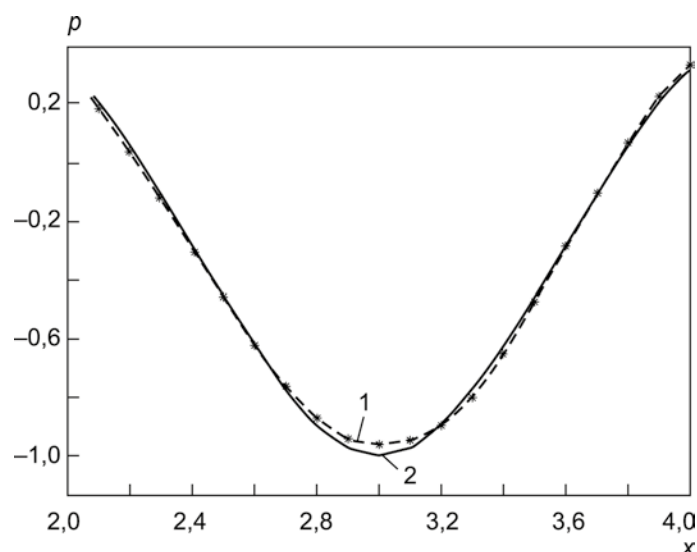


Рис. 2. Оценка второй производной:
1 – экстраполяция Ричардсона, 2 – вторая производная распределения Ирвина–Холла

Ричардсона. Для оценивания погрешностей ядерных оценок и значений вторых производных функции плотности вероятности используется правило Рунге. Поскольку рассмотренные подходы не требуют значительного увеличения вычислительных затрат и повышают точность оценки функции плотности вероятности, то они будут весьма полезны для практических задач обработки и анализа эмпирических данных.

Литература

1. **Тарасенко Ф. П.** Непараметрика. Томск: ТГУ, 1976.
2. **Scott R. W.** Multivariate density estimation: theory, practice, and visualization. N. Y.: John Wiley & Sons, 2015.
3. **Dobronets B. S., Popova O. A.** Numerical probabilistic analysis under aleatory and epistemic uncertainty // *Reliable Computing*. 2014. Т 19. С. 274–289.
4. **Dobronets B. S., Popova O. A.** Improving the accuracy of the probability density function estimation // *Journal of Siberian Federal University, Mathematics and Physics*. 2017. Т. 10 (1). С. 16–21.
5. **Dobronets B. S., Popova O. A.** Improving reliability of aggregation, numerical simulation and analysis of complex systems by empirical data // *IOP Conf. Series: Materials Science and Engineering* 2018. Т. 354. P. 012006. DOI:10.1088/1757-899X/354/1/012006.
6. **Березин И. С., Жидков Н. П.** Методы вычислений. Т. 1, 2. М.: Наука, 1966.
7. **Marchuk G. I., Shaidurov V. V.** Difference methods and their extrapolations. N. Y.: Springer–Verlag, 1983.
8. **Zlatev Z., Dimov I., Farago I., Havasi A.** Richardson Extrapolation. Practical Aspects and Applications. Walter de Gruyter GmbH, Berlin–Boston, 2018.
9. **Asadzadeh M., Schatz A. H., Wendland W. A.** New Approach to Richardson Extrapolation in the Finite Element Method for Second Order Elliptic Problems // *Mathematics Of Computation*. 2009. V. 78. No. 268. P. 1951–1973.
10. **Добронец Б. С., Шайдуров В. В.** Двусторонние численные методы. Новосибирск: Наука, 1990.
11. **Sidi A.** Practical Extrapolation Methods. Theory and Applications. N. Y.: Cambridge University Press, 2003.
12. **Johnson N. L., Kotz S., Balakrishnan N.** Continuous Univariate Distributions, V. 2. 2nd Edition. N. Y.: Wiley, 1995.

Дата принятия: 11.12.2018.

Using Richardson extrapolation to improve the accuracy of processing and analyzing empirical data

O. A. Popova

Siberian Federal University, Krasnoyarsk, Russia, e-mail: OlgaArc@yandex.ru

The article studies the properties of empirical data under random uncertainty. A new approach is proposed to improve the accuracy in problems of constructing the probability density function and estimating its error. The method is based on the application of Richardson extrapolation and the Runge rule for kernel estimates with different smoothing parameters. It is shown that the application of the Runge rule makes it possible to estimate the error of kernel estimates for the probability density function and the values of its second derivative.

Key words: kernel estimates, improving accuracy Richardson extrapolation, Runge rule, estimates of the derivatives of the probability density function.