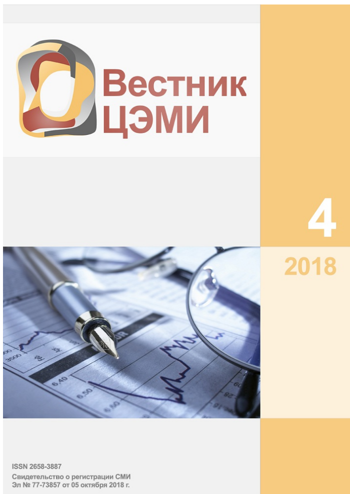

Herald of CEMI. 2013-2020

ISSN 2658--3887

URL - http://cemi.jes.su

Issue 4 Volume . 2018

Estimation of the integral of an unknown function from random observations of its values

Y. Pastukhova
CEMI RAS
Moscow, Nakhimovky prospect 47

Abstract

Let random observations of the function at some points in time are known. Statistical estimates of the integral of the function are useful in solving some economic problems. In this paper we consider the case of estimation the linear functional of a function, which is assumed only that it belongs to a known compact. It is assumed that observation errors of the additive model have a known distribution density. The sample design is given, i.e. the points at which observations are made are independent and equally distributed with a given density distribution density. A nonparametric estimator based on the maximum likelihood method, asymptotically reaching the minimax lower bounds, is proposed.

Keywords list (en): Nonparametric estimation of the functional, asymptotically efficient estimator

Date of publication: 03.02.2019

Citation link:

Pastukhova Y. Estimation of the integral of an unknown function from random observations of its values // Herald of CEMI. 2018. Issue 4 [Electronic resource]. Access for registered users. URL: https://cemi.jes.su/s265838870000164-3-1/ (circulation date: 05.06.2020). DOI: 10.33276/S0000164-3-1

1 **Введение.**

2 Будем предполагать, что известны наблюдения $(X_1, X_2, \dots, X_n)$, произведенные в точках $(t_1, t_2, \dots, t_n)$ множества T над неизвестной функцией *f(t)*. Точки $t_1, t_2, \dots, t_n$ являются случайными величинами, плотность распределения которых *p(t)* известна. В этом случае говорят, что план

наблюдений $t^{(n)} = (t_1, t_2, \dots, t_n)$ задан. Наблюдения X_i могут быть представлены в виде

3

$$X_i(t_i) = f(t_i) + \varepsilon_i(t_i), \; i = (1, \dots, n).$$

4 Ошибки ε_i условно независимы при заданном плане $t^{(n)}$, условное распределение которых *q*(*x*|*t)* известно, **E**$\{\varepsilon_i(t_i)|t_i\} = 0$.

5 Решается задача оценивания функционалов вида

6 L(f) =$\int_T f(t)dt; \; L_1(f) = \int_T f(t)l(t)dt$,

7 где *l(t)* –известная функция аргумента *t.*

8 Оценивание интеграла от неизвестной функции по случайным наблюдениям ее значений может оказаться полезным при решении ряда экономических задач. Интерпретируя множество *T* как временной интервал (0,*T*), функцию *f(t)* как денежный поток в момент времени *t*, а известную функцию *l(t)* как дисконтирующий множитель, можно считать функционал $L_1(f)$ дисконтируемой стоимостью денежного потока через временной интервал *T*. Если считать, что функция *f(t)* описывает изменение производительности некоторого производства с течением времени, то функционал *L(f)* определяет объем продукции, произведенной за промежуток времени *T=(a,b).* Если предполагать, что функция *f(t)* - интенсивность потребления некоторого продукта в момент времени *t* из временного отрезка *(a,b)*, то *L(f)* - суммарный объем потребления на данном отрезке времени. Еще одно приложение задачи оценивания функционала *L(f)* связано с проблемой определения объема нефти, добытой из месторождения за определенный период. Наблюдениями в этом случае является количество нефти, в нефтегазосодержащей жидкости определенного объема, полученной в результате исследования в некоторые (можно считать случайные) моменты времени. Содержание примесей в виде воды и газа в исследуемых пробах можно считать случайными ошибками измерений.

9 **Описание построения оценки.** Будем предполагать также, что функция плотности *p(t)* строго ограничена и положительна:

10

$$0 < c \le p(t) \le C < \infty.$$

11 Плотность условного распределения *q*(*x*|*t)* абсолютно непрерывна по *x* относительно меры Лебега в R^1 имеет конечное информационное количество Фишера

12

$$I(t) = \int_{R^1} \frac{(q_x'(x|t))^2}{q(x|t)} dx$$

13 и удовлетворяет некоторым условиям регулярности (см. [1]).

14 Пусть на множестве T задана конечная мера υ, функция f принадлежит некоторому известному компакту K в пространстве суммируемых с квадратом по мере υ на T функций. В этом случае можно построить непараметрическую оценку функционала

15 $$L(f) = \int_T f(t)d\upsilon(t),$$

16 асимптотически достигающую своих нижних минимаксных границ рисков для квадратической функции потерь (см.[1]).

17 *Определение 1*. Оценка $\hat{L}_n$, для которой справедливо равенство

18 $$\lim_{n\to\infty}\left[n \sup_K \mathbf{E}\left\{\hat{L}_n - L(f)\right\}^2\right] = \int_T \frac{d\upsilon(t)}{I(t)p(t)},$$

19 называется асимптотически эффективной в K непараметрической оценкой (АЭНО) линейного функционала $L(f)$ при известном условном распределении ошибок и заданном плане наблюдений.

20 Основная идея построения АЭНО состоит в следующем. Множество T разбивается н а N непересекающихся соизмеримых подмножеств $\Delta_1, \dots, \Delta_N$ так, чтобы выполнялись соотношения

21 $$1/2 \le \upsilon(\Delta_i)/\upsilon(\Delta_j) \le 2,$$

22 $$\sup_K\left[\int_T f^2(t)p(t)d\upsilon(t) - \sum_{k=1}^{N}\left(\int_{\Delta_k} f(t)p(t)d\upsilon(t)\right)^2 \Big/ \int_{\Delta_k} p(t)d\upsilon(t)\right] \to 0, \quad N\to\infty.$$

23 Заметим, что в простом случае, когда множество T – конечный интервал числовой оси, а мера υ- мера Лебега, то в силу теоремы Рисса достаточно разбить интервал T на N равных частей.

24 Если исследователь может планировать наблюдения (см. [2]), то ему достаточно на каждом из подмножеств Δ_k выбрать точку t_j, в которой он произведет достаточное число наблюдений над функцией $f(t)$. Это дает возможность построить на каждом из Δ_k оценку максимального правдоподобия. В рассматриваемом случае заданного плана наблюдений задача усложняется. Применение метода максимального правдоподобия обеспечивается за счет подбора параметров разбиения множества T и обоснования сходимости на каждом из подмножеств случайной величины $f(t_j)$ к постоянной величине в некотором вероятностном смысле.

25 Выбор числа подмножеств N можно осуществить разными способами. Этот выбор должен обеспечить выполнение соотношения:

26 $$\lim_{n\to\infty}[n(\mathbf{P}(\delta_k \le \gamma\upsilon(\Delta_k))] = 0.$$

27 Здесь γ -постоянная величина, которая подбирается вместе с N так, чтобы данное соотношение выполнялось. Случайная величина δ_k представляет собой сумму независимых бернуллиевых случайных величин $\delta_k = x_1 + x_2 + \ldots + x_n$. Каждая из x_i принимает значение 0 или 1 в зависимости от попадания точки наблюдения t_i в интервал Δ_k. Таким образом, случайная

величина δ_k представляет собой число точек плана $(t_1, t_2, \dots, t_n)$, попавших при выбранном способе разбиения множества *T* в подмножество Δ_k.

28 Например, выбор $N = \left[\frac{\gamma n}{\ln n}\right]$, где γ удовлетворяет неравенству $0 < \gamma < (1{,}5 - \sqrt{2})c\upsilon(T)$, обеспечивает справедливость равенства $\mathbf{P}(\delta_k \le \ln n) = o(1/n)$ при $n \to \infty$.

29 На каждом из множеств Δ_k применим (см. [2]) несмещенный вариант оценки максимального правдоподобия для параметра сдвига. Оценки максимального правдоподобия будут строиться только на тех интервалах, в которые в соответствии с планом оценивания попало достаточное число наблюдений. Обозначим A_k случайное событие, состоящее в том, что число наблюдений удовлетворяет условию $\delta_k > \gamma\upsilon(\Delta_k)$, χ (A_k) – индикатор события A_k . Пусть $\hat{\theta}_k$ - величина, максимизирующая по θ выражение

30 $\prod_{t_i \in \Delta_k} q(X_i(t_i) - \theta | t_i)\, p(t_i)$.

31 **Вывод.**

32 Оценка вида

33

$$\hat{L}_n = \sum_{k=1}^{N} \upsilon(\Delta_k)\, \hat{\theta}_k \chi(A_k)$$

34 является асимптотически эффективной в *K* непараметрической оценкой линейного функционала $L(f) = \int_T f(t)\, d\upsilon(t)$ по наблюдениям $X_1, X_2, \dots, X_n$ в смысле определения 1 . В [3] предложена АЭНО функционала $L_1(f)$ относительно Лебеговой меры на [0,1].

References:

1. Pastukhova Yu.I., Khas'minskij R.Z. Neparametricheskoe otsenivanie linejnogo funktsionala ot regressii pri zadannom plane nablyudenij. Problemy peredachi informatsii. 1988. T. 24. S. 55-65

2. Khas'minskij R.Z. O neparametricheskom otsenivanii linejnogo funktsionala ot regressii pri planirovanii nablyudenij. Problemy peredachi informatsii. 1986. T. 22. S. 43-61

3. Pastukhova Yu.I. Applying the maximum likelihood method for constructing asymptotically effective nonparametrical estimators of functionals from the regression function. Analytical and Computation Methods in Probability Theory and its Applications. Proceedings of International Scientific Conference. 2017, Moscow, Russia. P/ 239-243

Оценивание интеграла от неизвестной функции по случайным наблюдениям ее значений

Пастухова Ю. И.
Центральный экономико-математический институт РАН
Москва, Нахимовский проспект, 47

Аннотация

Статистические оценки интеграла от функции, известной лишь по случайным наблюдениям, произведенным в некоторые моменты времени, могут оказаться полезными при решении ряда экономических задач. В настоящей работе рассмотрен случай оценивания линейного функционала от функции, о которой предполагается лишь, что она принадлежит известному компакту. Предполагается также, что ошибки наблюдений аддитивной модели имеют известную плотность распределения, а точки, в которых производятся наблюдения (план наблюдений), независимы и одинаково распределены с заданной плотностью распределения. Предложена непараметрическая оценка, основанная на методе максимального правдоподобия, асимптотически достигающая нижних минимаксных границ.

Ключевые слова: Непараметрическое оценивание функционала, асимптотически эффективная оценка

Дата публикации: 03.02.2019

Ссылка для цитирования:

Пастухова Ю. И. Оценивание интеграла от неизвестной функции по случайным наблюдениям ее значений // Вестник ЦЭМИ РАН. 2018. Выпуск 4 [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: https://cemi.jes.su/s265838870000164-3-1/ (дата обращения: 05.06.2020). DOI: 10.33276/S0000164-3-1
