

УДК 519.25

МАТЕМАТИКА

И. А. ИБРАГИМОВ, Р. З. ХАСЬМИНСКИЙ

ОБ АППРОКСИМАЦИИ СТАТИСТИЧЕСКИХ ОЦЕНОК СУММАМИ НЕЗАВИСИМЫХ ВЕЛИЧИН

(Представлено академиком А. И. Колмогоровым 9 X 1972)

1. Пусть $X_1, \dots, X_n, X_i \in \mathcal{X}$, — независимые наблюдения из генеральной совокупности с плотностью $f(x, \theta)$, $x \in \mathcal{X}$, $\theta \in \Theta$, относительно σ -конечной меры ν , определенной на σ -алгебре $\mathcal{A}$ множеств из $\mathcal{X}$. Предположим, что параметр θ неизвестен наблюдателю, а параметрическое множество Θ — открытое подмножество прямой R^1 . В большом числе работ изучались свойства статистических оценок параметра θ . Мы сосредоточим внимание на аппроксимации с вероятностью 1 оценок θ_n , $n = 1, 2, \dots$ (нижний индекс n означает, что оценка строится по наблюдениям $X_1, X_2, \dots, X_n$), суммами соответствующего числа независимых и одинаково распределенных случайных величин.

В заметке рассматриваются такие употребительные статистические оценки, как оценка максимального правдоподобия (о.м.п.) и байесовские оценки при данной априорной плотности $\pi(\theta)$ и функции потерь $\psi(t - \theta)$ (определение этих оценок см., например, в (1)). Будем обозначать E_θ, P_θ соответственно математическое ожидание и вероятность, вычисленные при условии, что значение параметра равно θ . В последующем будут наложены некоторые ограничения гладкости на функцию $\ln f(x, \theta)$. Некоторые из них не обусловлены существом дела, но упрощают изложение.

2. Теорема 1. Пусть выполнены условия I и III работы (2), а функция $l(x, \theta) = \ln f(x, \theta)$ дважды дифференцируема по θ , причем равенство

$$\int f(x, \theta) \nu(dx) = 1$$

можно два раза дифференцировать по θ под знаком интеграла. Пусть кроме того, функции $E|l''_{\theta\theta}(X_1, \theta)|^{1+\delta}$ и $E_\theta|l'_\theta(X_1, \theta)|^{2+\delta}$ при некотором $\delta > 0$ ограничены в каждом компакте из Θ . Предположим, наконец, что для некоторого $\gamma > 0$ справедливо неравенство

$$|l''_{\theta\theta}(x, \theta_2) - l''_{\theta\theta}(x, \theta_1)| < k(x)|\theta_2 - \theta_1|^\gamma, \quad (1)$$

причем

$$E_\theta k(X_1) < \infty. \quad (2)$$

Тогда для о.м.п. $\hat{\theta}_n$ справедливо соотношение

$$\sqrt{n}(\hat{\theta}_n - \theta)I(\theta) - \frac{1}{\sqrt{n}} \sum_{j=1}^n l'_\theta(X_j, \theta) \xrightarrow[n \rightarrow \infty]{} 0 \quad (п.н.), \quad (3)$$

где $I(\theta)$ — информационное количество Фишера плотности f . Более того, для некоторого $\varepsilon > 0$ справедливо и соотношение

$$\sup_n \left\{ \left| \sqrt{n}(\hat{\theta}_n - \theta)I(\theta) - \frac{1}{\sqrt{n}} \sum_{j=1}^n l'_\theta(X_j, \theta) \right| n^\varepsilon \right\} < \infty \quad (P_\theta - п.н.). \quad (4)$$

Теорема 2. Утверждения теоремы 1 справедливы и для байесовских оценок t_n параметра θ , если априорная плотность $\pi(\theta)$ непрерывна, поло-

жительна в Θ и растет не быстрее степенной функции при $|\theta| \rightarrow \infty$, а выпуклая функция потерь $\psi(y)$ при некоторых $a \geq 1$, $b > 0$, $c > 0$, $\delta > 0$ удовлетворяет условиям

$$\begin{aligned}\psi(y) &= |y|^a + O(|y|^{a+\delta}), \quad y \rightarrow 0, \\ \psi(y) &< c(1 + |y|^b).\end{aligned}$$

Доказательство теоремы 1 и некоторые соображения по поводу теоремы 2 приведены в п. 4.

3. Теоремы 1 и 2 позволяют свести изучение асимптотических свойств последовательности оценок θ_n к изучению свойств последовательности сумм независимых случайных величин $\xi_j = l_{\theta'}(X_j, \theta)$, т. е. к хорошо изученной задаче. В частности, простое следствие теорем 1, 2 и закона повторного логарифма (см. (3), стр. 381) представляет следующая

Теорема 3. Если выполнены условия теорем 1, 2, то для о.м.п. и байесовских оценок справедливы соотношения

$$\overline{\lim}_{n \rightarrow \infty} \left\{ (\theta_n - \theta) \left[\frac{nI(\theta)}{2 \log \log n} \right]^{1/2} \right\} = - \lim_{n \rightarrow \infty} \left\{ (\theta_n - \theta) \left[\frac{nI(\theta)}{2 \log \log n} \right]^{1/2} \right\} = 1 \quad (n, n).$$

Последовательность случайных величин $\eta(n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j$ образует цепь Маркова, причем, очевидно,

$$\eta(n+1) - \eta(n) = \left(-\frac{1}{2(n+1)} + O\left(\frac{1}{(n+1)^2}\right) \right) \eta(n) + \frac{1}{\sqrt{n+1}} \xi_{n+1}. \quad (5)$$

Заметим также, что при выполнении условий теоремы 1

$$E_0 \xi_j = 0, \quad E_0 \xi_j^2 = I(\theta), \quad E_0 |\xi_j|^{2+\delta} < \infty. \quad (6)$$

Определим непрерывный случайный процесс $Y_n(s)$, $s \geq 0$ (случайную ломаную) так, чтобы в точках $s_{kn} = \ln(1 + k/n)$, $k = 0, 1, 2, \dots$, он принимал значения $\eta(n+k)$: $Y_n(s_{kn}) = \eta(n+k)$ и был линейен в каждом из интервалов $s_{kn} < s < s_{k+1n}$. Для этого процесса с помощью (5), (6) без труда получим соотношения

$$\Delta s_{kn} = s_{k+1n} - s_{kn} = o(1),$$

$$\begin{aligned}a_n(s_{kn}, x) &= \frac{1}{\Delta s_{kn}} E_0 \{Y_n(s_{k+1n}) - Y_n(s_{kn}) / Y_n(s_{kn}) = x\} = -\frac{x}{2} \left(1 + O\left(\frac{1}{n}\right) \right) \\ \sigma_n^2(s_{kn}, x) &= \frac{1}{\Delta s_{kn}} E_0 \{[Y_n(s_{k+1n}) - Y_n(s_{kn})]^2 / Y_n(s_{kn}) = x\} - \\ &\quad - \Delta s_{kn} a_n^2(s_{kn}, x) = I(\theta) + x^2 O(1/n^2),\end{aligned}$$

где соотношения o и O выполнены при $n \rightarrow \infty$ равномерно относительно k . Отсюда и из (5), (6) вытекает, что для процесса $Y_n(s)$ выполнены условия теоремы 1 ((4), стр. 595) и теоремы 2 ((4), стр. 604). Применение этих теорем приводит к следующему утверждению (несколько более слабый результат по аналогичному поводу приведен в (5), гл. 6).

Лемма. Распределение в $C[0, T]$, порожденное процессом $Y_n(s)$, слабо сходится при $n \rightarrow \infty$ к распределению, порожденному стационарным диффузионным процессом $Y(s)$, определяемым стохастическим уравнением

$$dY(s) = -\frac{1}{2} Y(s) ds + \frac{1}{\sqrt{I(\theta)}} d\xi(s), \quad (7)$$

где $\xi(s)$ — стандартный винеровский процесс.

Приведенная лемма в сочетании с теоремами 1 и 2 позволяет, в частности, строить интервалы, которым с заданной доверительной вероятностью

принадлежит неизвестный параметр при всех n из некоторого временного интервала. Действительно, из упомянутых утверждений сразу вытекает, что при достаточно гладкой функции $\varepsilon(t)$, $0 \leq t \leq T$, для рассмотренных выше оценок

$$P_0 \left\{ \sqrt{n} |\theta_n - \theta| < \varepsilon \left(\ln \frac{n}{N} \right), n \in [N, Ne^T] \right\} \xrightarrow{N \rightarrow \infty} \rightarrow P \{ |Y(t)| < \varepsilon(t), t \in [0, T] \}. \quad (8)$$

Вероятность в правой части последнего соотношения можно вычислить, решив краевую задачу для параболического уравнения

$$\frac{\partial u}{\partial t} = \frac{1}{2I(\theta)} \frac{\partial^2 u}{\partial y^2} - \frac{1}{2} y \frac{\partial u}{\partial y}. \quad (9)$$

Действительно, обозначим $u(y, t)$ решение уравнения (9), удовлетворяющее краевым условиям

$$u(y, 0) = 1, \quad u(\pm \varepsilon(T - t), t) = 0. \quad (10)$$

Тогда, как известно, справедливо равенство (см., например, ⁽⁶⁾)

$$u(y, t) = P \{ |Y^y(t)| < \varepsilon(t), t \in [0, T] \},$$

где $Y^y(t)$ — решение уравнения (7), удовлетворяющее начальному условию $Y(0) = y$. Так как нас интересует стационарное решение уравнения (7), то

$$P \{ |Y(t)| < \varepsilon(t), t \in [0, T] \} = \int_{-\varepsilon(T)}^{\varepsilon(T)} \sqrt{\frac{I(\theta)}{2\pi}} \exp \left\{ -\frac{I(\theta)y^2}{2} \right\} u(y, T) dy. \quad (11)$$

Рассматривая семейство функций $\varepsilon(t) = \varepsilon(t, C)$, зависящее от параметра C , можно, решая численно задачу (9), (10), найти затем с помощью формулы (11) для данного доверительного уровня ε_0 и данного T постоянную $c_0 = c_0(\varepsilon_0, T)$ так, чтобы

$$P \{ |Y(t)| < \varepsilon(c_0, t), t \in [0, T] \} = 1 - \varepsilon_0.$$

В силу (8) при больших N этот выбор c_0 обеспечит и справедливость приближенного равенства

$$P_0 \left\{ \theta \in \left[\theta_n - \frac{\varepsilon(c_0, \ln n/N)}{\sqrt{n}}, \theta_n + \frac{\varepsilon(c_0, \ln n/N)}{\sqrt{n}} \right], n \in [N, Ne^T] \right\} \approx 1 - \varepsilon_0. \quad (12)$$

Если $\varepsilon(T, C) = c$ или $\varepsilon(T, C) = c \exp \{ (T - t) / 2 \}$, то вычисление функции $c_0 = c_0(\varepsilon_0, T)$ сравнительно просто, так как задача (9), (10) может быть решена методом Фурье (во втором случае после замены переменных $\tau = 1 - e^{-t}$, $z = e^{-t/2}y$, приводящей уравнение (9) к уравнению теплопроводности).

Заметим, что построение доверительных интервалов, аналогичных (12), рассматривалось ранее в работе ⁽⁷⁾ с других позиций.

4. Докажем теорему 1. С помощью тождества

$$\sum_{j=1}^n l'_0(X_j, \hat{\theta}_n) = 0$$

и формулы Тейлора, учитывая также равенство $E_0 l''_{0\theta}(X_1, \theta) = -I(\theta)$,

получим

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^n l'_0(X_j, \theta) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n [l'_0(X_j, \theta) - l'(X_j, \hat{\theta}_n)] = \\ &= \sqrt{n}(\hat{\theta}_n - \theta) I(\theta) + \frac{1}{\sqrt{n}}(\theta - \hat{\theta}_n) \sum_{j=1}^n [l''_{00}(X_j, \theta) - E_\theta l''_{00}(X_j, \theta)] + \\ &+ \frac{1}{\sqrt{n}}(\theta - \hat{\theta}_n) \sum_{j=1}^n [l''_{00}(X_j, \theta_j) - l''_{00}(X_j, \theta)], \end{aligned} \quad (13)$$

где числа θ_j таковы, что $|\theta_j - \theta| < |\hat{\theta}_n - \theta|$. Обозначим J_1 и J_2 второе и третье слагаемые в правой части (13) и займемся их оценкой. Очевидно

$$J_1 = n^{1/2-2\varepsilon}(\theta - \hat{\theta}_n)n^{2\varepsilon-1} \sum_{j=1}^n [l''_{00}(X_j, \theta) - E_\theta l''_{00}(X_j, \theta)].$$

Так как случайные величины $l''_{00}(X_j, \theta)$ имеют момент степени $1 + \delta$, то при $2\varepsilon < \delta(1 + \delta)^{-1}$

$$n^{2\varepsilon-1} \sum_{j=1}^n [l''_{00}(X_j, \theta) - E_\theta l''_{00}(X_j, \theta)] \xrightarrow{n \rightarrow \infty} 0 \quad (\text{п. н.}) \quad (14)$$

по теореме Марцинкевича (см. (8), стр. 257).

Далее, в (2) показано, что в условиях теоремы случайные величины $\sqrt{n}(\hat{\theta}_n - \theta)$ имеют ограниченные моменты всех степеней, начиная с достаточно больших n . Поэтому

$$P_\theta \{n^{1/2-2\varepsilon}|\hat{\theta}_n - \theta| > n^{-\varepsilon}\} = P_\theta \{\sqrt{n}|\hat{\theta}_n - \theta| > n^\varepsilon\} \leq n^{-1/2k\varepsilon} E_\theta |\hat{\theta}_n - \theta| \sqrt{n}^k. \quad (15)$$

Выберем $k > 2\varepsilon^{-1}$. Тогда из леммы Бореля — Кателли, примененной к (15), и из (14) вытекает равенство $J_1 = o(n^{-\varepsilon})$ (п.н.).

С помощью (1) находим

$$|J_2| < \frac{1}{\sqrt{n}} |\hat{\theta}_n - \theta|^{1+\gamma} \sum_{j=1}^n K(X_j).$$

Применяя усиленный закон больших чисел к сумме $K(X_j)$ и снова пользуясь соотношениями типа (15), при $\varepsilon < \gamma/2$ получим отсюда, что и $J_2 = o(n^{-\varepsilon})$ (п.н.). Соотношение (5), а с ним и теорема 1, доказаны.

Для байесовских оценок соотношение (4) (при несколько более ограничительных условиях) вытекает из теоремы 1 и основной теоремы работы (9), утверждающей, что $\sqrt{n}|\hat{\theta}_n - t_n| \rightarrow 0$ (п.н.), если t_n — байесовская оценка. Справедливо следующее усиление этого результата.

Теорема 4. Если выполнены условия теоремы 2, то $|\hat{\theta}_n - t_n| n^{1/2+\varepsilon} \rightarrow 0$ (п.н.) при $n \rightarrow \infty$ для некоторого $\varepsilon > 0$.

Из теорем 1 и 4 вытекает теорема 2.

В заключение отметим, что все результаты переносятся и на случай многомерного параметрического множества Θ .

Ленинградский государственный университет
им. А. А. Жданова

Поступило
6 IX 1972

Институт проблем передачи информации
Академии наук СССР
Москва

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- ¹ Г. Крамер, Математические методы статистики, ИЛ, 1948. ² И. А. Ибрагимов, Р. З. Хасьяминский, Теор. вероятн. и ее применен., 17, 3, 4 (1972). ³ В. В. Петров, Суммы независимых случайных величин, «Наука», 1972. ⁴ И. И. Гихман, А. В. Скороход, Введение в теорию случайных процессов, «Наука», 1965. ⁵ М. Б. Невельсон, Р. З. Хасьяминский, Стохастическая аппроксимация и рекуррентное оценивание, «Наука», 1972. ⁶ Е. Б. Дынкин, Марковские процессы, М., 1963. ⁷ H. Robbins, Ann. Math. Statist., 41, 5 (1970). ⁸ М. Лозэв, Теория вероятностей, ИЛ, 1962. ⁹ M. Chao, Ann. Math. Statist., 41, 2 (1970).