Записки научных
семинаров ПОМИ
Том 341, 2007 г.

К. Хубер, В. Н. Солев, И. Вонта

ОЦЕНКИ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ: НЕПАРАМЕТРИЧЕСКИЙ ПОДХОД

Введение

Пусть $W_1, \dots, W_n$ – независимые одинаково распределенные элементы измеримого пространства $(\mathcal{X}, \mathcal{B})$ с общим распределением $L \in \mathcal{L}$. Задача состоит в оценивании по наблюдениям $W_1, \dots, W_n$ неизвестной плотности $p = p(L)$,

$$p = p(L) = \frac{dL}{d\mu}. \tag{1}$$

Здесь $\mathcal{L}$ известное множество, мы предполагаем, что все распределения $L \in \mathcal{L}$ абсолютно непрерывны по отношению к σ-конечной мере μ и используем обозначение

$$\mathcal{K} = \{p : \ p = p(L) \quad \text{для некоторой} \quad L \in \mathcal{L}\}. \tag{2}$$

Расстояние Кульбака–Лейблера для пары вероятностных распределений L_1 и L_2 определяется соотношением

$$K(L_1, L_2) = \int \ln \frac{d\,L_1}{d\,L_2}\, dL_1.$$

Здесь предполагается, что мера L_1 абсолютно непрерывна по отношению к мере L_2. В противном случае величина $K(L_1, L_2)$ полагается равной ∞. Пусть меры L_1 и L_2 имеют плотности p_1 и p_2 по отношению к σ-конечной мере μ. Мы по определению полагаем, что

$$K(p_1, p_2) := K(L_1, L_2) = \int \ln \frac{p_1}{p_2}\, p_1\, d\mu,$$

с понятным предположением о том, что величина $\frac{p_1}{p_2}$ считается равной единице вне носителя меры L_2.

Работа второго автора выполнена при поддержке грантов РФФИ 05-01-00920, НШ–4222.2006.1, РФФИ–ННИО 04-01-04000.

Пусть P_n – эмпирическое распределение, построенное по выборке $W_1, \ldots, W_n$,

$$P_n^W\{A\} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_A(W_j).$$

Здесь $\mathbb{I}_A(x)$ – индикатор множества A.

Эмпирический вариант $K_n(p_1, p_2)$ величины $K(p_1, p_2)$, в случае, когда p_1 – “истинная” плотность распределения величин $W_1, \ldots, W_n$, определяется соотношением

$$K_n(p_1, p_2) = \int \ln\frac{p_1}{p_2}\; dP_n^W.$$

Р. Фишер предложил минимизировать значение функционала $J_n(q) = K_n(p_1, q)$ на параметрическом множестве $\mathcal{K}$, а найденную точку минимума $\widehat{p}_n$ рассматривать в качестве оценки неизвестной плотности p. Эта оценка по известным причинам называется оценкой максимального правдоподобия и определяется соотношением

$$\int \ln\widehat{p}_n\, dP_n^W \geq \int \ln g\, dP_n^W \quad \text{для всех} \quad g \in \mathcal{K}. \tag{3}$$

У. Гренандер предложил рассматривать так называемые “огрубленные” оценки максимального правдоподобия, определяемые для подходящим образом выбранных неотрицательных величин υ_n уравнением (см. N. Xiatong Shen (1997))

$$\int \ln\widetilde{f}_n\, dP_n^W \geq \int \ln g\, dP_n^W - \upsilon_n \quad \text{для всех} \quad g \in \mathcal{K}, \tag{4}$$

каковые мы здесь и будем рассматривать, сохраняя здесь за ними для краткости название “огрубленные” оценки максимального правдоподобия.

Введем в рассмотрение расстояние Хеллингера $h(P, Q)$,

$$h^2(P, Q) := h^2(f, q) = \frac{1}{2}\int \left(\sqrt{f} - \sqrt{q}\right)^2\, d\mu. \tag{5}$$

Здесь f и q – плотности мер P и Q по отношению к мере μ. Разумеется, величина $h^2(P, Q)$ не зависит от выбора меры μ. Удобно следующее представление

$$h^2(f, q) = 1 - \int \sqrt{fq}\; d\mu. \tag{6}$$

Качество процедуры оценивания, использующей оценку $\widehat{p}_n$, мы будем измерять функцией риска

$$R(\widehat{p}_n) = R(\widehat{p}_n; \mathcal{K}) = \sup_{p \in \mathcal{K}} \mathbb{E}_p\, h(\widehat{p}_n, p).$$

Цель настоящей заметки состоит в применении некоторых известных результатов, касающихся оценок максимального правдоподобия, к задаче оценивания неизвестной плотности из выборки, прошедшей цензурирование.

Именно, пусть $X_1, \ldots, X_n$ – независимые одинаково распределенные случайные величины с общим распределением $P \in \mathcal{P}$ и плотностью $f = f(P) \in \mathcal{F}$, $Y_1, \ldots, Y_n$ – независимые одинаково распределенные случайные величины ц общим распределением $Q \in \mathcal{Q}$ и плотностью $q = q(Q) \in \mathcal{G}$, независимые от $X_1, \ldots, X_n$. Будем обозначать $F = F^f$, $G = G^q$ функции распределения с плотностями f и q.

Цензурирование состоит в следующем: выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$ недоступны наблюдению. Вместо них мы наблюдаем $W_1 = (Z_1, \delta_1), \ldots, W_n = (Z_n, \delta_n)$, где

$$Z_j = X_j \wedge Y_j, \quad \delta_j = \mathbb{I}_{X_j \le Y_j}. \tag{7}$$

Случайные векторы $W_1, \ldots, W_n$ имеют общую плотность $p(z,t)$,

$$p(z,t) = \left[f(z)(1 - G(z))\right]^t \times \left[g(z)(1 - F(z))\right]^{1-t}, \tag{8}$$

по отношению к мере $\mu(dz \times dt) = dz \times \nu(dt)$. Здесь $\nu(dt)$ – считающая мера на $\{0, 1\}$. Задача состоит в оценивании неизвестных плотностей $f \in \mathcal{F}$ и $g \in \mathcal{G}$ по выборке $W_1 = (Z_1, \delta_1), \ldots, W_n = (Z_n, \delta_n)$.

1. Энтропия и точность оценивания

По определению “bracket” $[\underline{f}, \overline{f}] = \left\{f : \underline{f} \le f \le \overline{f}\right\}$. Bracket $[\underline{f}, \overline{f}]$ называется ε-bracket'ом, если $h(\underline{f}, \overline{f}) \le \varepsilon$. Предположим, что для любого $\varepsilon > 0$ множество $\mathcal{F}$ может быть покрыто конечным числом ε-bracket'ов:

$$\mathcal{F} = \bigcup_{j=1}^{N} [\underline{f}_j, \overline{f}_j].$$

Наименьшее из таких N обозначим $N(\varepsilon, \mathcal{F})$. Функция $H(\varepsilon, \mathcal{F}) = \ln N(\varepsilon, \mathcal{F})$ называется bracketing-энтропией множества $\mathcal{F}$ (см. L. Devroye and G. Lugosi (2001)).

Ниже будет предполагаться, что размер выборки n и величина ε соотнесены следующим образом

$$\int_{\varepsilon^2/2^8}^{\sqrt{2}\varepsilon} H^{1/2}(d_1 t, \mathcal{K}) \le d_2 \sqrt{n}\varepsilon^2. \tag{9}$$

Теорема 1 (Wing Hung Wong, N. Xiatong Shen (1995)). *Предположим, что для любого $\varepsilon > 0$ множество $\mathcal{K}$ может быть покрыто конечным числом ε-bracket'ов, а размер выборки n связан с величиной $\varepsilon = \varepsilon_n$ соотношением* (9). *Положительные константы d_1, d_2, c_1, c_2 могут быть выбраны так, чтобы*

$$\mathbb{P}_p\Big\{ \sup_{h(g,p)\ge\varepsilon, g\in\mathcal{K}} \prod_{j=1}^{n} g(W_j)/p(W_j) \ge \exp\{-c_1 n\varepsilon^2\}\Big\} \le 4\exp\{-c_2 n\varepsilon^2\}. \tag{10}$$

Теорема 1 позволяет исследовать "огрубленные" оценки максимального правдоподобия.

Теорема 2 (Wing Hung Wong, N. Xiatong Shen (1995)). *В условиях теоремы* 1 *предположим, что*

$$\upsilon_n \le c_1 \varepsilon_n^2. \tag{11}$$

Тогда для оценки $\widetilde{p}_n$, определенной соотношением

$$\int \ln \widetilde{p}_n \, dP_n^W \ge \int \ln g \, dP_n^W - \upsilon_n \quad \text{для всех} \quad g \in \mathcal{K}, \tag{12}$$

$$\mathbb{P}_p\{h(\widetilde{p}_n), p) \ge \varepsilon\} \le 5\exp\{-c_2 n\varepsilon^2\}. \tag{13}$$

2. Основной результат

Вернемся к задаче оценивания плотности по косвенно наблюдаемой выборке $(X_1, Y_1), \ldots, (X_n, Y_n)$. Напомним, что случайные величины X_i, Y_i предполагались независимыми. Мы обозначали через f, F – плотность и функцию распределения величин X_i, через g, G – плотность и функцию распределения величин Y_i. Предполагалось, что неизвестные плотности f и g черпаются из параметрических множеств $\mathcal{F}$ и $\mathcal{G}$: $f \in \mathcal{F}$, $g \in \mathcal{G}$. Нам потребуются некоторые условия на носители соответствующих мер.

Положим

$$\begin{aligned} \tau_X &= \sup_{f\in\mathcal{F}} \sup\left\{x:\ F^f(x)<1\right\}, \\ \tau_Y &= \inf_{g\in\mathcal{G}} \sup\left\{x:\ G^g(x)<1\right\}. \end{aligned} \tag{14}$$

Мы будем предполагать, что

$$\tau_X \le \tau_Y. \tag{15}$$

Другое наше предположение будет касаться скорости роста соответствующих энтропий: при некотором $\beta > 1$

$$H(\varepsilon,\mathcal{F}) \le C_1\left(\frac{1}{\varepsilon}\right)^{\frac{1}{\beta}}, \quad H(\varepsilon,\mathcal{G}) \le C_2\left(\frac{1}{\varepsilon}\right)^{\frac{1}{\beta}}. \tag{16}$$

Хорошо известно следующее достаточное условие для (16). Для $\beta > 0$ обозначим $r = r(\beta)$ наибольшее целое, меньшее β и $\alpha = \alpha(\beta) = \beta - r(\beta)$. Так что

$$\beta = r + \alpha, \quad 0 < \alpha \le 1.$$

Для вектора $m = (m_1, \ldots, m_d)$ размерности d обозначим $|m| = m_1 + \ldots + m_d$, и рассмотрим дифференциальный оператор

$$D^m = \frac{\partial^{|m|}}{\partial^{m_1}\ldots\partial^{m_d}}.$$

Пусть $G \subset \mathbb{R}^d$. Примем обозначение

$$\|h\|_{(\beta)} = \max_{|m|\le r}\sup_x D^m\, h(x) + \max_{|m|\le r}\sup_{x\ne y}\frac{|D^m\, h(x) - D^m\, h(y)|}{\|x-y\|^{\alpha}}, \tag{17}$$

где супремум берется по всем внутренним точкам x, y из G, $x \ne y$. Пусть $\mathbf{C}_C^{\beta} = \mathbf{C}_C^{\beta}(G)$ множество всех непрерывных h с полунормой $\|h\|_{(\beta)} \le C$.

Тхеорем 3 (Колмогоров А. Н., Тихомиров В. М. (1959)). *Пусть G ограниченное выпуклое подмножество $\mathbb{R}^d$ с непустой внутренностью. Тогда для некоторой константы $K = K(G,\beta,d)$*

$$H\left(\varepsilon, \mathbf{C}_1^{\beta}(G), \|\cdot\|_{\infty}\right) \le K(G,\beta,d)\left(\frac{1}{\varepsilon}\right)^{\frac{d}{\beta}}.$$

Лемма 1. *При условии* (16)

$$H(\varepsilon, \mathcal{K}) \le C_3 \left(\frac{1}{\varepsilon}\right)^{\frac{1}{\beta}}. \tag{18}$$

Утверждение леммы очевидным образом следует из вида плотностей $p(z,t)$ (см. (30).

Положим

$$\widehat{p}_n(z,t) = \left[\widehat{f}_n(z)(1 - \widehat{G}_n(z))\right]^t \times \left[\widehat{g}_n(z)(1 - \widehat{F}_n(z))\right]^{1-t}. \tag{19}$$

Здесь мы обозначили $\widehat{f}_n, \widehat{g}_n$ υ_n – огрубленные оценки максимума правдоподобия для неизвестных f, g по наблюдениям $W_1, \ldots, W_n$ и взяли

$$\widehat{F}_n = F^{\widehat{f}_n}, \quad \widehat{G}_n = G^{\widehat{g}_n}. \tag{20}$$

Очевидно, $\widehat{p}_n$ – υ_n – огрубленная оценка максимума правдоподобия для неизвестной плотности p по наблюдениям $W_1, \ldots, W_n$. Выбирая величины $\varepsilon = \varepsilon_n$, следующим образом:

$$\varepsilon_n \asymp \left(\frac{1}{n}\right)^{\frac{\beta}{2\beta+1}}, \quad \upsilon_n \le c_1 \varepsilon_n^2, \tag{21}$$

мы из теоремы 1 выводим справедливость следующей леммы.

Лемма 2. *При условиях* (16), (21) *при некоторых положительных* c_1, c_2, C

$$\mathbb{P}_p \left\{ \sup_{h(g,p)\ge\varepsilon, g\in\mathcal{K}} \prod_{j=1}^{n} g(W_j)/p(W_j) \ge \exp\{-c_1 n\varepsilon^2\} \right\} \le C \exp\{-c_2 n \varepsilon^2\}. \tag{22}$$

Заметим, что если энтропия параметрического множества растет не слишком быстро, например, так, как это указано в (16), то взамен условия (9) может быть взято условие (21).

Следующая лемма выводится из теоремы 2.

Лемма 3. *При условиях* (16), (21) *при некоторых положительных* c_1, c_2, C

$$\mathbb{P}_p \left\{ h(\widehat{p}_n), p) \ge c_1 \varepsilon \right\} \le C \exp\{-c_2 n \varepsilon^2\}. \tag{23}$$

Теорема 4. *Предположим, что выполнены следующие условия:* (15), (16), (21).*Тогда при некоторых положительных* c_1, c_2, C

$$\mathbb{P}_p\left\{\int \left(1-\widehat{G}_n\right)\left(\sqrt{\widehat{f}_n}-\sqrt{f}\right)^2 dx \geq c_1\left(\frac{1}{n}\right)^{\frac{2\beta}{2\beta+1}}\right\} \leq$$
$$\leq C\exp\{-c_2 n^{\frac{1}{2\beta+1}}\}, \quad (24)$$

$$\mathbb{P}_p\left\{\int \left(1-\widehat{F}_n\right)\left(\sqrt{\widehat{g}_n}-\sqrt{g}\right)^2 dx \geq c_1\left(\frac{1}{n}\right)^{\frac{2\beta}{2\beta+1}}\right\} \leq$$
$$\leq C\exp\{-c_2 n^{\frac{1}{2\beta+1}}\}, \quad (25)$$

$$\mathbb{E}_p \int \left(1-\widehat{G}_n\right)\left(\sqrt{\widehat{f}_n}-\sqrt{f}\right)^2 dx \leq C\left(\frac{1}{n}\right)^{\frac{2\beta}{2\beta+1}}, \qquad (26)$$

$$\mathbb{E}_p \int \left(1-\widehat{F}_n\right)\left(\sqrt{\widehat{g}_n}-\sqrt{g}\right)^2 dx \leq C\left(\frac{1}{n}\right)^{\frac{2\beta}{2\beta+1}}, \qquad (27)$$

Доказательство. Поскольку

$$\int \left(1-\widehat{F}_n\right)\left(\sqrt{\widehat{g}_n}-\sqrt{g}\right)^2 dx \leq 2,$$

$$\int \left(1-\widehat{F}_n\right)\left(\sqrt{\widehat{g}_n}-\sqrt{g}\right)^2 dx \leq 2,$$

соотношения (26) и (27) следуют из (24) и (25). Далее, заметим, что

$$h^2(\widehat{p}_n, p) = 1 - \int \sqrt{\widehat{p}_n p}\, d\mu, \qquad (28)$$

где $\mu(dz \times dt) = dz \times \nu(dt)$, $\nu(dt)$ – считающая мера на $\{0,1\}$. При этом

$$p(z,t) = [f(z)(1-G(z))]^t \times [g(z)(1-F(z))]^{1-t}, \qquad (29)$$

$$\widehat{p}_n(z,t) = \left[\widehat{f}_n(z)(1-\widehat{G}_n(z))\right]^t \times \left[\widehat{g}_n(z)(1-\widehat{F}_n(z))\right]^{1-t}. \qquad (30)$$

Так как

$$\sqrt{\widehat{p}_n p} \le \left[\sqrt{f(z)\widehat{f}_n(z)}\,\frac{(1-G(z))+(1-\widehat{G}_n(z))}{2}\right]^t \times$$
$$\times\left[\sqrt{(1-F(z))(1-\widehat{F}_n(z))}\,\frac{g(z)+\widehat{g}_n(z)}{2}\right]^{1-t}$$

и функция $g^*(z)=\frac{g(z)+\widehat{g}_n(z)}{2}$ является плотностью вероятностного распределения с функцией распределения $G^*(z), 1-G^*(z)=\frac{(1-G(z))+(1-\widehat{G}_n(z))}{2}$, то

$$\sqrt{\widehat{p}_n p} \le \sqrt{[f(z)(1-G^*(z)]^t\ [g^*(z)(1-F(z)]^{1-t}} \times$$
$$\times\sqrt{\left[\widehat{f}_n(z)(1-G^*(z)\right]^t\ \left[g^*(z)(1-\widehat{F}_n(z)\right]^{1-t}}.$$

Отсюда, принимая обозначения

$$p_1(z,t)=[f(z)(1-G^*(z)]^t\ [g^*(z)(1-F(z)]^{1-t},$$
$$p_2(z,t)=\left[\widehat{f}_n(z)(1-G^*(z)\right]^t\ \left[g^*(z)(1-\widehat{F}_n(z)\right]^{1-t},$$

получаем

$$\sqrt{\widehat{p}_n p} \le \sqrt{p_1 p_2}.$$

Поскольку p_1, p_2 – вероятностные плотности, из (28) мы выводим, что

$$h^2(\widehat{p}_n,p) \ge h^2(p_1,p_2) \ge \frac{1}{2}\int (1-G^*)\left(\sqrt{\widehat{f}_n}-\sqrt{f}\right)^2 dz, \qquad (31)$$

а потому

$$h^2(\widehat{p}_n,p) \ge \frac{1}{4}\int (1-\widehat{G}_n)\left(\sqrt{\widehat{f}_n}-\sqrt{f}\right)^2 dz. \qquad (32)$$

Точно так же устанавливается, что

$$h^2(\widehat{p}_n,p) \ge \frac{1}{4}\int (1-\widehat{F}_n)\left(\sqrt{\widehat{g}_n}-\sqrt{g}\right)^2 dz. \qquad (33)$$

Теперь видно, что неравенства (24), (25) являются следствием леммы 3.

Литература

1. L. Devroye and G. Lugosi, *Combinatorial methods in density estimation.* — Springer-Verlag (2001).
2. Wing Hung Wong and N. Xiatong Shen. *Probability inequalities and convergence rates of sieve mles.* — Ann. Statist. **2:6** (1995), 339–362.
3. N. Xiatong Shen, *On methods sieves and penalization.* — Ann. Statist. **2** 1997, 2555–2591.
4. А. Н. Колмогоров, В. М. Тихомиров, *Эпсилон-энтропия и эпсилон-емкость множеств в функциональных пространствах.* — Успехи мат. наук, **86**, (2), 3–86 (1959).

Huber K., Solev V. N., Vonta I. Maximum likelihood estimator: the nonparametric approach.

In the paper the maximum likelihood estimator for unknown density is investigated . We deal with censored observations and construct a confidence region for this density.

Université René
Descartes - Paris 5

Поступило 6 декабря 2006 г.

С.-Петербургское отделение
Математического института
им. В. А. Стеклова РАН

Deparment of Mathematics
and Statistics, University of Cyprus,
Nicosia, Cyprus