Записки научных
семинаров ПОМИ
Том 320, 2004 г.

В. Н. Солев, Ф. Хагиги

ОЦЕНКА ПАРАМЕТРА В ОДНОЙ МОДЕЛИ С БЕСКОНЕЧНОМЕРНЫМ МЕШАЮЩИМ ПАРАМЕТРОМ

Введение

Пусть X случайная величина с функцией распределения $F(t)$ и плотностью $f(t)$, $\Phi(t)$ возрастающая абсолютно непрерывная функция, $\Psi(t)$ обратная функция: $\Psi = \Phi^{-1}$. Обозначим $Y = \Phi(X)$. Функцию распределения $G(t)$ и плотность $g(t)$ случайной величины Y легко сосчитать

$$G(t) = F(\Psi(t)), \quad g(t) = f(\Psi(t))\psi(t),$$
$$\text{где} \quad \psi(t) = \frac{d}{dt}\,\Psi(t),\ \phi(t) = \frac{d}{dt}\,\Phi(t). \tag{1}$$

Рассмотрим случай, когда наблюдаются две выборки $X, X_1, \cdots, X_{n_1}$ и $Y, Y_1, \cdots, Y_{n_2}$ – первая с функцией распределения $F(t)$ и плотностью $f(t)$, вторая с функцией распределения $G(t) = F(\Psi(t))$ и плотностью $g(t) = f(\Psi(t))\psi(t)$.

Предположим, что требуется оценить функцию ψ, в ситуации, когда функция f неизвестна и выступает как мешающий параметр. Для неотрицательных функций ψ_1, ψ_2 примем обозначение $h(\psi_1, \psi_2)$ для величины

$$h^2(\psi_1, \psi_2) = \frac{1}{2}\int \left(\sqrt{\psi_1(x)} - \sqrt{\psi_1(x)}\right)^2 dx.$$

Мы предположим, что $f \in \mathcal{F}, \psi \in \mathcal{B} \subset L^1$, причем для любого $\varepsilon > 0$

$$\delta(\varepsilon) = \inf_{\substack{h(\psi_1,\psi_2)\geq\varepsilon \\ f\in\mathcal{F},\ \psi_1,\psi_2\in\mathcal{B}}} h(f(\Psi_1(x))\psi_1(x), f(\Psi_2(x))\psi_2(x)) > 0. \tag{2}$$

Исследования первого автора частично поддержаны грантами РФФИ 02-01-00262, РФФИ-ННИО 04-01-04000, НШ 2258.2003.1.

Мы покажем, при условии (2) и надлежащих условиях на параметрические множества $\mathcal{F}$ и $\mathcal{B}$, что оценка максимума правдоподобия $\widehat{\psi}$ для неизвестной функции ψ является состоятельной.

В дальнейшем для простоты обозначений мы будем предполагать, что $n_1 = n_2 = n$. В этом случае мы можем считать, что мы имеем дело с независимыми одинаково распределенными случайными векторами $Z, Z_1, \ldots, Z_n$, где $Z_i = (X_i, Y_i)$. Очевидно, распределение случайного вектора Z имеет плотность

$$p(z_1, z_2) = f(z_1) f(\Psi(z_2)) \psi(z_2).$$

Энтропия параметрического множества в метрике Хеллингера

Пусть $(\mathcal{Y}, \mathcal{B}, \mu)$ – пространство с мерой и $Y_1, \ldots, Y_n$ независимые одинаково распределенные случайные элементы со значениями из $\mathcal{Y}$ с общим распределением $P \in \mathcal{P}$ и плотностью f,

$$f(y) = \frac{dP}{d\mu}(y), \quad f \in \mathcal{F} = \left\{ f : \, f = \frac{dP}{d\mu}, \text{ для некоторых } \; P \in \mathcal{P} \right\}.$$

Для неотрицательных f, g пусть $h(f, g)$ расстояния Хеллингера,

$$h^2(f, g) = \frac{1}{2} \int\limits_{\mathcal{Y}} \left(\sqrt{f} - \sqrt{g} \right)^2 d\mu.$$

Для пары неотрицательных функций $g^L \leq g^R$ обозначим $V(g^L, g^R)$ множество

$$V(g^L, g^R) = \left\{ g : \, g^L \leq g \leq g^R \right\}.$$

Обозначим через $N(\varepsilon, \mathcal{F})$ наименьшее m такое, что

$$\mathcal{F} \subset \bigcup_{j=1}^{m} V(g_j^L, g_j^R), \text{ где } h(g_j^L, g_j^R) \leq \varepsilon, \; j = 1, \ldots, m.$$

Назовем $\varepsilon-$энтропией величину $H(\varepsilon, \mathcal{F}) = \ln N(\varepsilon, \mathcal{F})$.

Мы предположим, что величина $N(\varepsilon, \mathcal{F})$ конечна для любого $\varepsilon > 0$ и для некоторого $c > 0$

$$\int\limits_{\varepsilon^2}^{\varepsilon} H^{1/2}(s, \mathcal{F}) \, ds \leq c\varepsilon^2 \sqrt{n} \qquad (3).$$

Теорема (W. H. Wong and X. Shen). *Предположим, что $f(x)$ истинная плотность, $f \in \mathcal{F}$. Тогда при условии (3) существуют такие неотрицательные константы c_1, c_2, C, что*

$$P\left\{\sup_{\substack{h(g,f)\geq\varepsilon,\\ g\in\mathcal{F}}} \prod_{j=1}^{n} \frac{g(Y_j)}{f(Y_j)} \geq \exp\{-c_1 n\varepsilon^2\}\right\} \leq C\exp\{-c_2 n\varepsilon^2\}.$$

Предположим теперь, что $q_n = q_n(y;x)$ неотрицательное ядро $q_n : \mathcal{Y}^n \times \mathcal{Y} \to \mathbb{R}^1$, такая, что

$$\int\limits_{\mathcal{Y}} q_n(y;x)\,\mu(dx) = 1 \ \ (\mu^n\text{–почти всюду по } y).$$

Здесь $y = (y_1 \dots, y_n) \in \mathcal{Y}^n, x \in \mathcal{Y},\ \ \mu^n = \underbrace{\mu \times \ldots \times \mu}_{n-times}$. Предположим,что

$$q_n(y;\cdot) \in \mathcal{F} \ \ (\mu^n\text{–почти всюду по } y).$$

Так что случайная величина $\tilde{f}_n(x) = q_n(Y_1, \dots, Y_n; x)$ может рассматриваться в качестве оценки для f.

Предположим, что $f(x)$ истинная плотность, а функция $q_n(y;\cdot)$ удовлетворяет условию

$$\prod_{j=1}^{n} \frac{q_n(y;y_j)}{f(y_j)} \geq \exp\left\{-c_1 n\varepsilon^2\right\} \ \ (\mu^n\text{–почти всюду по } y). \qquad (4)$$

Здесь $y = (y_1, \dots, y_n)$.

Лемма 1. *Пусть $f(x)$ истинная плотность, $\tilde{f}_n(x)$ оценка для f со значениями из $\mathcal{F}$. Тогда при условиях (3) и (4) для некоторых положительных c, C*

$$P\left\{h(\tilde{f}_n, f) \geq \varepsilon\right\} \leq C\exp\{-cn\varepsilon^2\}.$$

Доказательство. Пусть $\tilde{f}_n(\cdot) = q_n(Y_1, \dots, Y_n; \cdot)$. Мы предположим, что $q_n(y;\cdot) \in \mathcal{Y}$. Обозначим $d(y)$ расстояние Хеллингера между $q_n(y;\cdot)$ и $f(\cdot)$,

$$d^2(y) = \frac{1}{2}\int\limits_{\mathcal{Y}} \left(\sqrt{q_n(y;x)} - \sqrt{f(x)}\right)^2 \mu(dx).$$

Ясно, что

$$\{y:\ d(y)\geq\varepsilon,\ \}\ \subset\ \left\{ y:\ \sup_{\substack{h(g,f)\geq\varepsilon,\\ g\in\mathcal{F}}} \prod_{j=1}^{n}\frac{g(y_j)}{f(y_j)}\geq\exp\left\{-c_1 n\varepsilon^2\right\}\right\}.$$

Поэтому

$$P\left\{h(\tilde{f}_n,f)\geq\varepsilon\right\}\leq P\left\{ \sup_{\substack{h(g,f)\geq\varepsilon,\\ g\in\mathcal{F}}} \prod_{j=1}^{n}\frac{g(Y_j)}{f(Y_j)}\geq\exp\left\{-c_1 n\varepsilon^2\right\}\right\}.$$

Утверждение леммы следует из Теоремы (W. H. Wong and X. Shen).

Оценка максимума правдоподобия

Пусть наблюдаются независимые одинаково распределенными случайные векторы $Z_1,\ldots,Z_n$, где $Z_i=(X_i,Y_i)$ с неизвестной плотностью

$$p(z_1,z_2)=f(z_1)f(\Psi(z_2))\psi(z_2).$$

Обозначим

$$\mathcal{L}=\{p:\ p(z_1,z_2)=f(z_1)f(\Psi(z_2))\psi(z_2)\ \text{для некоторых}\ f\in\mathcal{F},\psi\in\mathcal{B}\ \}.$$

Обозначим через P_n эмпирическую меру,

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

Рассмотрим оценку максимума правдоподобия $\widehat{p}_n$ для неизвестной $p\in\mathcal{L}$,

$$\int\ln\widehat{p}_n\,dP_n=\max_{q\in\mathcal{L}}\int\ln q\,dP_n.$$

Примем обозначение $\widehat{\psi}_n$ для соответствующей оценки функции ψ, которая восстанавливается по функции $\widehat{p}_n$, если выполнено условие (2).

Мы предположим, что для любого $\varepsilon>0$ существуют такой конечный набор неотрицательных функций $\left\{f_i^L\leq f_i^R;\ i=1,\ldots,m_1\right\}$

и такой конечный набор неотрицательных функций $\left\{\psi_i^L \leq \psi_i^R;\ i = 1, \ldots, m_2\right\}$, что

$$\mathcal{F} \subset \bigcup_{j=1}^{m_1} V(f_j^L, f_j^R),\ h(f_j^L, f_j^R) \leq \varepsilon, \tag{5}$$

$$\mathcal{B} \subset \bigcup_{j=1}^{m_2} V(\psi_j^L, \psi_j^R), \tag{6}$$

причем для любых фиксированных $i = 1, \ldots, m_1,\ j = 1, \ldots, m_2$ и $\psi, \psi_* \in V(\psi_j^L, \psi_j^R)$

$$\int \left(\sqrt{f_i^R(\Psi(x))\psi(x)} - \sqrt{f_i^L(\Psi_*(x))\psi_*(x)}\right)^2 dx \leq \varepsilon^2. \tag{7}$$

В этом случае мы для краткости будем говорить, что для пары параметрических множеств $\mathcal{F}$ и $\mathcal{B}$ выполнено условие $(*)$.

Теорема 1. *Пусть для пары параметрических множеств $\mathcal{F}$ и $\mathcal{B}$ выполнено условие $(*)$ и для любого $\varepsilon > 0$ величина $\delta(\varepsilon > 0)$. Тогда для любого $\varepsilon > 0$ при $n \to \infty$*

$$P\left\{h(\psi, \widehat{\psi}_n) > \varepsilon\right\} \to 0$$

равномерно по $\psi \in \mathcal{B}, f \in \mathcal{F}$.

Доказательство. Очевидно, при условии $(*)$ выполнена посылка леммы 1 в применении к задаче оценивания функции p. Поэтому для любого $\varepsilon > 0$ при $n \to \infty$

$$P\left\{h(p, \widehat{p}_n) > \varepsilon\right\} \to 0$$

равномерно по $\psi \in \mathcal{B}, f \in \mathcal{F}$.

Отсюда и из условия (2) следует утверждение теоремы.

Литература

1. Wing Hung Wong,d N. Xiatong Shen, *Probability inequalities and convergence rates of sieve mles.* — The Annals of Statistics, **2** (1995), 2555–2591.

Solev V. N., Haghighi F. Estimation in a model with infinite dimensional nuisance parameter.

Let X_1 be a random variable with density function $f(t)$, $\Psi(t)$ be an increasing absolutely continuous function, $\Phi(t)$ be the inverse function, random variable X_2 be defined by $X_2 = \Phi(X_1)$. We consider the maximum likelihood estimator for density ψ of function Ψ as we observe two independent samples from the distribution of X_1 and X_2. Under appropriate conditions on the involved distributions, we prove the consistency of maximum likelihood estimator.

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

Поступило 24 декабря 2004 г.