Записки научных
семинаров ПОМИ
Том 431, 2014 г.

М. С. Ермаков

АСИМПТОТИЧЕСКИ ЭФФЕКТИВНЫЕ ПРОЦЕДУРЫ МЕТОДА СУЩЕСТВЕННОЙ ВЫБОРКИ ДЛЯ БУТСТРАПА

§1. Введение. Постановка задачи.

Метод существенной выборки является одним из наиболее распространенных методов вычисления малых вероятностей. Задачи нахождения малых вероятностей событий часто интерпретируются как задачи теории вероятностей больших и умеренных уклонений. Эффективные процедуры метода существенной выборки для вычисления вероятностей больших уклонений рассматривались в большом числе работ (см. [1, 2, 8, 9] и ссылки в них).

Для бутстрапа эффективные процедуры метода существенной выборки были предложены Джонсом [6]. Метод существенной выборки для бутстрапа изучался Джонсом в зоне действия нормальной аппроксимации статистических распределений. В настоящей работе мы доказываем асимптотическую эффективность бутстрап-процедур метода Джонса в зоне вероятностей умеренных уклонений для широкого класса дифференцируемых статистических функционалов. Данные результаты базируются на условных принципах больших уклонений для вероятностей умеренных уклонений взвешенных эмпирических бутстрап-мер при условии, что эмпирическая мера фиксирована. Они являются модификацией аналогичных условных принципов больших уклонений для эмпирических бутстрап-мер, доказаных в [4]. Для задач оценивания вероятностей умеренных уклонений статистических функционалов асимптотическая эффективность аналогов процедуры Джонса была доказана в [3].

В работах [3] и [5] результаты о слабой сходимости распределений дифференцируемых статистических функционалов (см. [11]) были распространены на зону вероятностей умеренных уклонений. Мы

Ключевые слова: метод существенной выборки, принцип больших уклонений, вероятности умеренных уклонений, эмпирическая мера, бутстрап.

Работа поддержана грантами РФФИ 14-01-00856-а и 14-01-00271-а.

применяем этот подход к задачам о вероятностях умеренных уклонений для дифференцируемых статистических функционалов, зависящих от взвешенных эмпирических бутстрап-мер, что позволяет доказать асимптотическую эффективность соответствующих процедур метода существенной выборки.

Пусть $X_1, \dots, X_n$ – независимые одинаково распределенные случайные величины, имеющие вероятностную меру $\mathbf{P}_0$, заданную на σ-алгебре $\Im$ хаусдорфова топологического пространства S. Обозначим $\widehat{\mathbf{P}}_n$ эмпирическую меру $X_1, \dots, X_n$. Обозначим Λ множество всех вероятностных мер на $(S, \Im)$.

Пусть функционал $T : \Lambda \to R^1$ дифференцируем по Адамару. Для статистики $T(\widehat{\mathbf{P}}_n)$ мы хотим оценить вероятность $\omega_n \doteq \mathbf{P}_0(T(\widehat{\mathbf{P}}_n) - T(\mathbf{P}_0) > a_n)$ в зоне вероятностей умеренных уклонений, т.е., когда $a_n > 0, a_n \to 0, na_n^2 \to \infty$ при $n \to \infty$. Поскольку вероятностная мера $\mathbf{P}_0$ считается неизвестной, оценка ω_n будет построена на основе метода существенной выборки, соответствующего взвешенному бутстрапу.

Стандартная процедура бутстрапа состоит в следующем. В качестве оценки ω_n берется

$$W_n^* = \widehat{\mathbf{P}}_n(T(\mathbf{P}_n^*) - T(\widehat{\mathbf{P}}_n) > a_n), \tag{1.1}$$

где $\mathbf{P}_n^*$ – эмпирическая мера независимых случайных величин $X_1^*, \dots, X_n^*$, имеющих вероятностную меру $\widehat{\mathbf{P}}_n$. Оценка W_n^* находится в свою очередь по k-выборкам $X_{i1}^*, \dots, X_{in}^*, 1 \leqslant i \leqslant k$, имеющим вероятностную меру $\widehat{\mathbf{P}}_n$. В качестве оценки W_n^* берется

$$\widehat{W}_n = \frac{1}{n} \sum_{i=1}^{k} \chi(T(\mathbf{P}_{ni}^*) - T(\widehat{\mathbf{P}}_n) > a_n),$$

где $\mathbf{P}_{ni}^*$ – эмпирическая мера независимых случайных величин $X_{i1}^*, \dots, X_{in}^*$. Здесь и в дальнейшем $\chi(A)$ обозначает индикатор события A.

Когда W_n^* мало, его вычисление трудоемко. Поэтому W_n^* обычно оценивается на основе метода существенной выборки, используя взвешенный бутстрап.

Опишем стандартную процедуру метода существенной выборки, когда вероятностная мера $\mathbf{P}_0$ известна. Она состоит в следующем. Предположим, что при моделировании мы используем вероятностную меру $\mathbf{Q}_n$. Предположим, что $\mathbf{Q}_n$ абсолютно непрерывна относительно

$\mathbf{P}_0$ и имеет плотность $q_n = d\mathbf{Q}_n/d\mathbf{P}_0$. Тогда мы моделируем k независимых выборок $Y_1^{(i)}, \dots, Y_n^{(i)}, 1 \leqslant i \leqslant k$, имеющих вероятностную меру $\mathbf{Q}_n$. После этого в качестве оценки ω_n берется

$$\hat{\omega}_n = \sum_{i=1}^{n} \chi(T(\widehat{\mathbf{Q}}_n^{(i)}) - T(\mathbf{P}_0) > a_n) \prod_{j=1}^{n} q_n^{-1}(Y_j^{(i)}).$$

Здесь $\widehat{\mathbf{Q}}_n^{(i)}$ обозначает эмпирическую меру $Y_1^{(i)}, \dots, Y_n^{(i)}$.

Легко видеть, что

$$\mathbf{E}[\widehat{\omega}_n] = \omega_n$$

и

$$\mathbf{Var}[\widehat{\omega}_n] = \mathbf{E}_{\mathbf{Q}_n}\left[\chi(T(\widehat{\mathbf{Q}}_n^{(1)}) - T(\mathbf{P}) > b_n) \prod_{j=1}^{n} q_n^{-1}(Y_j^{(i)})\right] \doteq U_n - \omega_n^2. \quad (1.2)$$

Решение задачи будет осуществлено в два этапа. Сначала мы предположим, что вероятностная мера $\mathbf{P}_0$, а следовательно и функция влияния, известны. При этой априорной информации будет указана асимптотически эффективная процедура метода существенной выборки, основанная на взвешенном бутстрапе. После этого будут указаны условия, которым должна удовлетворять оценка функции влияния, чтобы рассматриваемый метод существенной выборки оставался асимптотически эффективным при неизвестной вероятностной мере $\mathbf{P}_0$.

Результаты базируются на принципе больших уклонений для условных вероятностей умеренных уклонений для взвешенной эмпирической бутстрап-меры при условии, что эмпирическая мера известна. В дальнейшем мы будем называть такой принцип больших уклонений условным принципом умеренных уклонений.

Во взвешенном бутстрапе моделирование бутстрап-выборки на основе эмпирической меры $\widehat{\mathbf{P}}_n$ заменяется моделированием случайных величин, распределенных в соответствии со взвешенной эмпирической мерой $\widetilde{\mathbf{P}}_n$. Случайная величина Y, имеющая меру $\widetilde{\mathbf{P}}_n$, задается следующими вероятностями:

$$\widetilde{\mathbf{P}}_n(Y = X_i) = \frac{1}{n}(1 + a_n(h_n(X_i) - \bar{h}_n)), \quad 1 \leqslant i \leqslant n, \quad (1.3)$$

где $h_n \in L_1(\mathbf{P}_0) \cap L_2(\mathbf{P}_0)$ и $h_n(x) > -b_n^{-1}$ для всех $x \in S$. Здесь $\bar{h}_n = \frac{1}{n}\sum_{i=1}^{n} h_n(X_i)$ и $b_n > 0, b_n \to 0, b_n a_n^{-1} \to \infty$ при $n \to \infty$.

Пусть $Z_1, \dots, Z_{k_n}$ – независимые случайные величины, имеющие вероятностную меру $\widetilde{\mathbf{P}}_n$. Обозначим $\widetilde{\mathbf{P}}^*_{k_n}$ эмпирическую меру $Z_1, \dots, Z_{k_n}$. Назовем $\widetilde{\mathbf{P}}^*_{k_n}$ взвешенной эмпирической бутстрап-мерой.

Обозначим $\mathbf{H}_n, \mathbf{H}_n(S) = 0$, последовательность зарядов, таких что $\mathbf{H}_n$ абсолютно непрерывна относительно $\mathbf{P}_0$ и $h_n = \frac{d\mathbf{H}_n}{d\mathbf{P}_0}$.

Условимся обозначать

– Λ_0 – множество всех зарядов $\mathbf{H}$ на $(S, \Im)$, таких что $\mathbf{H}(S) = 0$;

– $\mathbf{Q} \ll \mathbf{P}$, если $\mathbf{Q} \in \Lambda$ абсолютно непрерывна относительно $\mathbf{P} \in \Lambda$;

– C, c – положительные постоянные;

– $\int$ всегда обозначает $\int\limits_S$.

Для каждого заряда $\mathbf{H} \in \Lambda_0$ определим меру $|\mathbf{H}|$, такую что

$$|\mathbf{H}|(A) = \sup\big\{|\mathbf{H}(B) - \mathbf{H}(D)| : D, B \in \Im; D, B \subset A\big\}$$

для любого измеримого множества $A \in \Im$.

§2. Основные результаты

2.1. Основные определения. Результаты будут даны в терминах τ_Θ-топологий. Для $t > 3$ обозначим $\Theta = \Theta_t$ множество измеримых функций $f : S \to R^1$, таких что $\mathbf{E}[|f(X)|^t] < \infty$.

Обозначим

$$\Lambda_{0\Theta} = \left\{\mathbf{H} : \int |f|\, d|\mathbf{H}| < \infty,\ \mathbf{H} \in \Lambda_0,\ f \in \Theta\right\}.$$

Топология слабой сходимости τ_Θ в $\Lambda_{0\Theta}$ порождается всеми отображениями

$$\mathbf{Q} \Rightarrow \int f\, d\mathbf{Q}, \quad \mathbf{Q} \in \Lambda_{0\Theta}, \quad f \in \Theta$$

(в дальнейшем все топологические свойства будут рассматриваться относительно τ_Θ-топологии). Для произвольного множества $\Omega \subset \Lambda_{0\Theta}$ обозначим соответственно $\mathfrak{clo}(\Omega)$ и $\mathfrak{int}(\Omega)$ замыкание и внутренность множества Ω в τ_Θ-топологии. Обозначим σ_Θ наименьшую σ-алгебру, порожденную τ_Θ-топологией.

Определим функционал действия ρ_0. Для $\mathbf{G} \in \Lambda_{0\Theta}$ функционал действия равен

$$\rho_0^2(\mathbf{G}) = \begin{cases} \frac{1}{2}\int \left(\frac{d\mathbf{G}}{d\mathbf{P}_0}\right)^2 d\mathbf{P}_0, & \mathbf{G} \ll \mathbf{P}_0; \\ \infty, & \text{в противном случае.} \end{cases}$$

Для множества $\Omega \subset \Lambda_{0\Theta}$ обозначим $\rho_0^2(\Omega) = \inf\big\{\rho_0(\mathbf{G}) : \mathbf{G} \in \Omega\big\}$.

2.2. Внешняя и внутренняя вероятности. Статистические функционалы, зависящие от эмпирических мер (см. [5, 11]), часто не являются измеримыми. Поэтому результаты будут даны в терминах внутренних и внешних вероятностей. Пусть $(\Psi, \Im, \mathbf{P})$ является вероятностным пространством. Внешняя вероятность множества $B \subset \Psi$ равна

$$(\mathbf{P})^*(B) = \inf\{\mathbf{P}(A); B \subseteq A, A \in \Im\},$$

а его внутренняя вероятность равна $(\mathbf{P})_*(B) = 1 - (\mathbf{P})^*(\Psi \setminus B)$.

Для последовательности случайных величин $Z_n : \Psi \to R^1$ (Z_n не обязательно измеримы) скажем, что $\liminf_{n\to\infty} Z_n \geqslant c$ внутренне почти наверное ($a.s_*$), если существует измеримая случайная величина Δ_n, такая что $\Delta_n \leqslant Z_n$, и $\mathbf{P}(\liminf_{n\to\infty} \Delta_n \geqslant c) = 1$.

Скажем, что $\limsup_{n\to\infty} Z_n \leqslant c$ внешне почти наверное ($a.s^*$), если $\liminf_{n\to\infty} -Z_n \geqslant -c$ $a.s_*$. Скажем что $\limsup_{n\to\infty} Z_n = -\infty$ внутренне почти наверное ($a.s_*$), если $\liminf_{n\to\infty} -Z_n \geqslant -c$ $a.s_*$ для любого $c > 0$.

2.3. Условный принцип умеренных уклонений для взвешенной эмпирической бутстрап-меры. Введем следующие условия на последовательность зарядов $\mathbf{H}_n \ll \mathbf{P}_0$.

A. Найдется такой заряд $\mathbf{H} \in \Lambda_{0\Theta}, \mathbf{H} \ll \mathbf{P}_0$, что

$$\lim_{n\to\infty} \int \left(\frac{d(\mathbf{H}_n - \mathbf{H})}{d\mathbf{P}_0}\right)^t d\mathbf{P}_0 = 0$$

и $\int |h|^t \, d\mathbf{P}_0 < \infty$, где $h = \frac{d\mathbf{H}}{d\mathbf{P}_0}$.

B. Заряды $\mathbf{H}_n$ имеют плотности

$$\frac{d\mathbf{H}_n}{d\mathbf{P}_0}(x) = h(x)\,\chi(|h(x)| > -b_n^{-1}) - \int h(x)\,\chi(h(x) > -b_n^{-1})\, d\mathbf{P}_0.$$

Теорема 1. *Пусть выполнены условия* **A, B** *и* $t > 3$. *Пусть* $a_n > 0$ – *такая убывающая последовательность, что*

$$na_n^t \to 0, \quad a_{n+1}/a_n \to 1, k_n a_n^2 \to \infty$$

при $n \to \infty$. *Пусть* $\Omega \subset \Lambda_{0\Theta}$. *Тогда*

$$\liminf_{n\to\infty} (k_n a_n^2)^{-1} \log(\widetilde{\mathbf{P}}_n)_*(\widetilde{\mathbf{P}}^*_{k_n} \in \widehat{\mathbf{P}}_n + a_n\Omega) \geqslant -\rho_0^2(\mathsf{int}(\Omega) - \mathbf{H}) \quad a.\, s_* \tag{2.1}$$

и

$$\limsup_{n\to\infty}(k_n a_n^2)^{-1}\log(\widetilde{\mathbf{P}}_n)^*(\widetilde{\mathbf{P}}^*_{k_n}\in\widehat{\mathbf{P}}_n+a_n\Omega)\leqslant-\rho_0^2(\mathfrak{cl}(\Omega)-\mathbf{H})\quad a.\,s^*, \tag{2.2}$$

где внешняя вероятностная мера $(\widetilde{\mathbf{P}}_n)^*$ *и внутренняя вероятностная мера* $(\widetilde{\mathbf{P}}_n)_*$ *рассматриваются относительно* σ_Θ*-алгебры.*

Приводимая ниже теорема 2 дает оценку скорости сходимости в условном принципе умеренных уклонений.

Теорема 2. *Пусть выполнено условие* $\mathbf{A}$ *и* $t>3$. *Пусть* $a_n>0$ – *такая убывающая последовательность, что*

$$na_n^t\to0,\quad a_{n+1}/a_n\to1, k_na_n^2\to\infty$$

при $n\to\infty$. *Пусть* $\Omega\subset\Lambda_{0\Theta}$. *Тогда для любых*

$$\epsilon>0\quad и\quad n>n_0(\epsilon,\{k_i\}_{i=1}^\infty,\Omega)$$

имеет место

$$(k_na_n^2)^{-1}\log(\widetilde{\mathbf{P}}_n)_*(\widetilde{\mathbf{P}}^*_{k_n}\in\widehat{\mathbf{P}}_n+a_n\Omega)\geqslant-\rho_0^2(\mathfrak{int}(\Omega)-\mathbf{H})-\epsilon \tag{2.3}$$

и, если $\rho_0^2(\mathfrak{cl}_\Theta(\Omega))<\infty$ *дополнительно, то*

$$(k_na_n^2)^{-1}\log(\widetilde{\mathbf{P}}_n)^*(\widetilde{\mathbf{P}}^*_{k_n}\in\widehat{\mathbf{P}}_n+a_n\Omega)\leqslant-\rho_0^2(\mathfrak{cl}(\Omega)-\mathbf{H})+\epsilon \tag{2.4}$$

на множествах событий, имеющих внутреннюю вероятность больше $\kappa_n(\epsilon,\Omega)=1-C(\epsilon,\Omega)(n^{1-t/3}+nb_n^t)$.

Если $\rho_0^2(\mathfrak{cl}(\Omega))=\infty$, *то для любого* $L>0$

$$(k_na_n^2)^{-1}\log(\widetilde{\mathbf{P}}_n)^*(\mathbf{P}^*_{k_n}\in\widehat{\mathbf{P}}_n+a_n\Omega)\leqslant-L \tag{2.5}$$

на множестве событий, имеющем внутреннюю вероятность больше $\kappa_{1n}=1-C(L,\Omega)(n^{1-t/3}+nb_n^t)$.

2.4. Принцип сжатия для статистических функционалов, дифференцируемых по Адамару. В [3] и [5] для статистических функционалов, дифференцируемых по Фреше и Адамару соответственно, было показано, что методы доказательств слабой сходимости их распределений применимы и для изучения их вероятностей умеренных уклонений. В [5] предполагалось, что статистические функционалы зависят от эмпирических процессов, заданных на некотором метрическом пространстве. Таким образом, чтобы применить методы [5], необходимо ввести в $\Lambda_{0\Theta}$ топологию метрического пространства.

В [4] было показано, что расстояние Колмогорова–Смирнова, снабженное определенными весовыми функциями, непрерывно в τ_Θ-топологии. Используя это утверждение, мы можем применить теоремы 1 и 2, и на основе принципа сжатия получить условные принципы вероятностей умеренных уклонений для статистических функционалов. Ниже мы укажем аналог теоремы 1 для этой постановки задачи.

Приведем теорему 4.1, доказанную в [4]. Предположим, что

$$\int |x|^{t\kappa}\, d\mathbf{P}_0 < \infty$$

для $t > 3$ и $\kappa > 0$.

Определим множество Υ измеримых функций $f : R^d \to R^1$, такое что

$$|f(x)| \leqslant C(1 + |x|^\kappa), \quad x \in R^d.$$

Определим множество $\Lambda_{0\kappa}$ зарядов G, таких что

$$\int |x|^\kappa\, d|\mathbf{G}| < \infty.$$

Множество Λ_κ определяется аналогично. Это множество всех вероятностных мер $\mathbf{P} \in \Lambda$, таких что

$$\int |x|^\kappa\, d\mathbf{P} < \infty.$$

Для всех $\mathbf{G} \in \Lambda_{0\kappa}$ определим функции $F_{\mathbf{G}}(x) = \mathbf{G}(I_x)$, где $I_x = (-\infty, x_1) \times \ldots \times (-\infty, x_d), x = (x_1, \ldots, x_d) \in R^d$. Для любых $\mathbf{G}$ и $\mathbf{R}$ из Λ_0 определим расстояние

$$\rho_\kappa(\mathbf{G}, \mathbf{R}) = \sup_{x \in R^d} |F_{\mathbf{G}}(x) - F_{\mathbf{R}}(x)|(1 + |x|^\kappa).$$

Определим ρ_κ-топологию в Λ_κ, порожденную расстоянием ρ_κ.

Теорема 3. *ρ_κ-топология слабее τ_Υ-топологии.*

Скажем, что функционал $T : \Lambda_\kappa \to R^1$ дифференцируем по Адамару в точке $\mathbf{P}_0$, если существует такая функция $q \in \Upsilon$, что для всякой $\mathbf{G} \in \Lambda_{0\kappa}, \mathbf{G} \ll \mathbf{P}_0, \int (d\mathbf{G}/d\mathbf{P}_0)^2 d\mathbf{P}_0 < \infty$, имеет место

$$\lim_{n \to \infty} \frac{T(\mathbf{P}_0 + t_n \mathbf{G}_n) - T(\mathbf{P}_0)}{t_n} = \int q\, d\mathbf{G}$$

для всех $t_n \to 0$ при $n \to \infty$ и $\mathbf{G}_n \in \Lambda_{0\kappa}$, таких что $\mathbf{P}_0 + t_n \mathbf{G}_n \in \Lambda$ и $\rho_\kappa(\mathbf{G}_n, \mathbf{G}) \to 0$ при $n \to \infty$. Функция q называется функцией влияния функционала T.

Теорема 4. *Пусть выполнены условия теоремы* 1. *Пусть функционал* $T : \Lambda_\kappa \to R^1$ *непрерывно дифференцируем по Адамару в точке* $\mathbf{P}_0$ *и имеет функцию влияния* q. *Тогда*

$$\begin{aligned} &\liminf_{n\to\infty}(k_n a_n^2)^{-1}\log(\widetilde{\mathbf{P}}_n)_*(T(\widetilde{\mathbf{P}}^*_{k_n}) - T(\widehat{\mathbf{P}}_n) > a_n) \\ &= \limsup_{n\to\infty}(k_n a_n^2)^{-1}\log(\widetilde{\mathbf{P}}_n)^*(T(\widetilde{\mathbf{P}}^*_{k_n}) - T(\widehat{\mathbf{P}}_n) > a_n) \\ &= -\inf\left\{\frac{1}{2}\int\left(\frac{d\mathbf{R}}{d\mathbf{P}_0}\right)^2, \int q\, d(\mathbf{R}-\mathbf{H}) = 1, \mathbf{R}\in\Lambda_0\Upsilon\right\} \qquad (2.6) \\ &= -\frac{\left(1+\int qh\, d\mathbf{P}_0\right)^2}{2\int q^2\, d\mathbf{P}_0} \quad a.\, s_*. \end{aligned}$$

Доказательство теоремы 4 получается по существу повторением доказательства теоремы 3.1 в [5] с добавлением в ключевых моментах рассуждений фразы: "внутренне почти наверное". При этом в качестве принципа больших уклонений, на котором базируется принцип сжатия, выступает теорема 1. Мы опустим соответствующие рассуждения.

2.5. Асимптотически эффективные процедуры метода существенной выборки. Следуя [1, 2, 3, 8], назовем процедуру метода существенной выборки асимптотически эффективной, если для рассматриваемой постановки задачи

$$\lim_{n\to\infty}\frac{\log U_n}{2\log\omega_n} = 1 \quad a.s_*.$$

Теорема 5. *Пусть выполнены условия теорем* 1 *и* 4. *Тогда для того, чтобы процедура метода существенной выборки была асимптотически эффективной, необходимо и достаточно, чтобы* $h = \sigma_q^{-2}q$, *где* $\sigma_q^2 = \int q^2\, d\mathbf{P}_0$.

Доказательство теоремы 5 по существу повторяет рассуждения доказательства теоремы 2.2 в [3] и будет опущено. Достаточно заметить, что оценки (3.1)–(3.4) и (3.11)–(3.19) в доказательстве теоремы 2.2 в [3] справедливы и для этой постановки задачи.

Замечание. Пусть $\hat{q}_n(x)$ – оценка функции влияния $q(x)$, построенная по реализациям независимых случайных величин $Y_1,\dots,Y_{l_n}$ с вероятностной мерой $\mathbf{P}_0$. Пусть $l_n \leqslant n$ и $l_n\to\infty$ при $n\to\infty$. Пусть $Y_1, Y_2, \dots$

не зависят от $X_1, X_2, \ldots$. Предположим, что

$$\mathbf{E}|\hat{q}_n(X) - q(X)|^t < Cl_n^{-\gamma} \quad \text{и} \quad \sum_{n=1}^{\infty} l_n^{-\gamma} < \infty.$$

Тогда все рассуждения доказательств теорем 1, 4 и 5 справедливы для $\hat{q}_n$, и мы можем использовать $\hat{q}_n$ вместо q при вычислении вероятностей умеренных уклонений ω_n. Таким образом мы можем разбить выборку $X_1, \ldots, X_n$ на две части: $X_1, \ldots, X_{l_n}$ и $X_{l_n}, \ldots, X_n$. По первой из них мы строим оценку функции влияния, а по второй оцениваем малые вероятности на основе метода существенной выборки.

§3. Доказательство теорем 1 и 2

Для $r > 0$ обозначим

$$\Gamma_r = \{\mathbf{G} : \mathbf{G} \ll \mathbf{P}_0, \int (d\mathbf{G}/d\mathbf{P}_0)^2 \, d\mathbf{P}_0 \leqslant r, \mathbf{G} \in \Lambda_0\}.$$

Доказательство теорем базируется на следующей лемме 3.1, доказанной в [3].

Лемма 3.1. *Имеет место*

i. $\Gamma_r \subset \Lambda_{0\Theta}$,

ii. Γ_r *является* τ_Θ*-компактным и секвенциально* τ_Θ*-компактным множеством в* $\Lambda_{0\Theta}$.

Мы дадим только доказательство верхней границы (2.4) в теореме 2. Используя приведенную ниже оценку (3.4), доказательство верхней границы (2.4) будет сведено к доказательству аналогичной верхней границы для эмпирических бутстрап-мер, установленной в теореме 2.2 в [4]. Доказательства нижней границы в теореме 2 и теоремы 1 также выводятся из аналогичных утверждений для эмпирических бутстрап-мер в [4] на основе оценки (3.4). Это нижняя граница в теореме 2.2 и теорема 2.1 в [4] соответственно. Мы опустим эти доказательства.

Обозначим $\eta = \rho_0^2(\mathfrak{cl}(\Omega) - \mathbf{H})$ и зафиксируем $\delta, 0 < 2\delta < \eta$. Ясно, что $\Gamma_{\eta-\delta} \subset \Lambda_{0\Theta} \setminus (\Omega - \mathbf{H})$.

Для любых $f_1, \ldots, f_l \in \Theta$, $\mathbf{G} \in \Lambda_{0\Theta}$ и $\gamma > 0$ обозначим

$$U(f_1, \ldots, f_l, \mathbf{G}, \gamma) = \left\{ \mathbf{R} : \left| \int f_i d(\mathbf{R} - \mathbf{G}) \right| < \gamma, \mathbf{R} \in \Lambda_{0\Theta}, 1 \leqslant i \leqslant l \right\}.$$

Так как множество $\Gamma_{\eta-\delta}$ компактно, то существует такое конечное покрытие множества $\Gamma_{\eta-\delta}$ множествами

$$U_1 = U(f_{11}, \dots, f_{1l_1}, \mathbf{G}_1, c_1), \dots, U_m = U(f_{m1}, \dots, f_{ml_m}, \mathbf{G}_m, c_m),$$

что $U_i \cap (\Omega - \mathbf{H}) = \varnothing, 1 \leqslant i \leqslant m$.

Здесь $f_{ij} \in \Theta, \mathbf{G}_i \in \Lambda_{0\Theta}$ для $1 \leqslant j \leqslant l_i, 1 \leqslant i \leqslant m$. Обозначим $U = \cup_{i=1}^m U_i$.

Для доказательства верхней границы достаточно оценить правую часть

$$(\widetilde{\mathbf{P}}_n)^*(\mathbf{P}^*_{k_n} \in \widehat{\mathbf{P}}_n + a_n\Omega) \leqslant \widetilde{\mathbf{P}}_n(\mathbf{P}^*_{k_n} \notin \widehat{\mathbf{P}}_n + a_n U).$$

Рассуждения, аналогичные доказательству теоремы 2.2 в [3], позволяют заменить множество U на множество $V + \mathbf{H}$, имеющее более простую геометрию. Множество V строится по конечному числу функций $q_1, \dots, q_l \in \Gamma_{\eta-2\delta}$, таких что

$$\mathbf{E}[q_i(X)] = 0, \mathbf{E}[q_i^2(X)] = 2(\eta - 2\delta), \quad 1 \leqslant i \leqslant l,$$

и каждая функция $q_s, 1 \leqslant s \leqslant l$, является конечной линейной комбинацией функций $f_{ij}, 1 \leqslant j \leqslant l_i, 1 \leqslant i \leqslant m$.

Множество V задается следующим образом:

$$V = \cap_{i=1}^k V_i,$$

где

$$V_i = V(q_i) = \left\{ \mathbf{G} : \left| \int q_i d\mathbf{G} \right| < 2(\eta - 2\delta), \mathbf{G} \in \Lambda_{0\Theta} \right\}.$$

Таким образом, задача свелась к оценке сверху слагаемых правой части

$$\begin{aligned} \widetilde{\mathbf{P}}_n(\mathbf{P}^*_{k_n} \in \widehat{\mathbf{P}}_n + a_n V + a_n\mathbf{H}) &\leqslant \sum_{i=1}^k \widetilde{\mathbf{P}}_n(\mathbf{P}^*_{k_n} \notin \widehat{\mathbf{P}}_n + a_n V_i + a_n \mathbf{H}) \\ = \sum_{i=1}^k \widetilde{\mathbf{P}}_n \left(\int q_i \, d(\mathbf{P}^*_{k_n} - \widehat{\mathbf{P}}_n - a_n\mathbf{H}) - 2a_n(\eta - 2\delta) > 0 \right). \end{aligned} \tag{3.1}$$

Остается доказать, что для каждого $f \in \Theta, \mathbf{E}[f(X)] = 0, \mathbf{E}[f^2(X)] = \eta - 2\delta$, и всех $n > n_0(\epsilon, f)$ имеет место

$$\begin{aligned} &(k_n a_n^2)^{-1} \log \widetilde{\mathbf{P}}_n \left(\int f d(\widetilde{\mathbf{P}}^*_{k_n} - \widehat{\mathbf{P}}_n - a_n\mathbf{H}_n) > 2a_n(\eta - 2\delta) \right) \\ &\leqslant -2\frac{(\eta - 2\delta)^2}{\mathbf{Var}[f(X_1)]}(1 - \epsilon) = -2(\eta - 2\delta)(1 - \epsilon) \end{aligned} \tag{3.2}$$

с вероятностью $\kappa_n(\epsilon, U(f,\mathbf{H}))$.

Аналогичное утверждение (см. неравенство (5.3) в [4]) было доказано для эмпирических бутстрап-мер

$$\begin{aligned}&(k_n a_n^2)^{-1}\log\widehat{\mathbf{P}}_n\left(\int f d(\mathbf{P}^*_{k_n}-\widehat{\mathbf{P}}_n) > 2a_n(\eta-2\delta)\right)\\ &\leqslant -2\frac{(\eta-2\delta)^2}{\mathbf{Var}[f(X_1)]}(1-\epsilon) = -2(\eta-2\delta)(1-\epsilon).\end{aligned} \tag{3.3}$$

Определим множество событий

$$A_{nfh} = \{X_1,\dots,X_n : \max_{1\leqslant s\leqslant n}|f(X_s)| < b_n^{-1}, |h_n(X_s)| < b_n^{-1}\}.$$

Имеем

$$\mathbf{P}_0(A_{nfh}) \geqslant 1 - n\mathbf{P}_0(|f(X_1)| > b_n^{-1}) - n\mathbf{P}_0(|h_n(X_1)| > b_n^{-1}) \geqslant 1 - Cnb_n^t.$$

Так что в дальнейших оценках мы будем предполагать, что событие A_{nfh} имеет место.

Используем (3.3) для доказательства (3.2).

Зададим случайный вектор (Z_1, X_1^*) на множестве значений (X_i, X_j), $1\leqslant i,j\leqslant n$ так, что $Z_1 = X_1^*$ с вероятностью

$$1 - n^{-1}a_n\sum_{h(X_i)<0}|h(X_i)|,$$

и случайные величины Z_1 и X_1^* имеют маргинальные распределения $\widetilde{\mathbf{P}}_n$ и $\widehat{\mathbf{P}}_n$ соответственно. Обозначим $\bar{\mathbf{P}}$ вероятностную меру (Z_1, X_1^*).

Определим сначала множество событий, где $Z_1 = X_1^*$. Для наблюдений $X_i, 1\leqslant i\leqslant n$, для которых $h(X_i) > 0$, имеет место $\bar{\mathbf{P}}(Z_1 = X_1^* = X_i) = 1/n$. Для наблюдений $X_i, 1\leqslant i\leqslant n$, для которых $h(X_i) < 0$, имеет место $\bar{\mathbf{P}}(Z_1 = X_1^* = X_i) = 1/n + n^{-1}a_n h(X_i)$. На множестве событий вероятности $n^{-1}a_n\sum\limits_{h(X_i)<0}|h(X_i)|$ зададим случайный вектор (Z_1, X_1^*), принимающий значения $(X_i, X_j), 1\leqslant i,j\leqslant n$, такие что $h(X_i) > 0$ и $h(X_j) < 0$. При этом маргинальные вероятности следующие:

$Z_1 = X_i$ с вероятностью $n^{-1}a_n h(X_i)$,

а $X_1^* = X_j$ с вероятностью $n^{-1}a_n|h(X_j)|$.

Ясно, что выбор вероятностной меры на этом множестве событий неоднозначен, но это не повлияет на дальнейшие рассуждения.

Докажем, что для выборки независимых случайных векторов (Z_i, X_i^*), $1 \leqslant i \leqslant k_n$, для любого $C_0 > 0$, для $n > n_0(C_0)$, имеет место

$$I = \bar{\mathbf{P}}\left(\left|\int f(d\widetilde{\mathbf{P}}^*_{k_n} - d\mathbf{P}^*_{k_n} - a_n d\mathbf{H})\right| > \epsilon a_n\right) \leqslant \exp\{-C_0 k_n a_n^2\} \qquad (3.4)$$

с вероятностью $1 - C_2 n b_n^t - C n^{1-t/3}$. Тогда (3.2) будет следовать из (3.3).

Имеем

$$\begin{aligned} &\mathbf{Var}\left[\int f\, d(\widetilde{\mathbf{P}}^*_{k_n} - \mathbf{P}^*_{k_n} - a_n\mathbf{H})\right] \\ &= \frac{a_n^2}{n}\sum_{i=1}^n f^2(X_i)|h_n(X_i)| - a_n^2\left(\int f\, d\mathbf{H}\right)^2 \doteq J_n. \end{aligned} \qquad (3.5)$$

Обозначим

$$\tilde{f}_n = \frac{1}{n}\sum_{i=1}^n h(X_i)f(X_i) - \int f\, d\mathbf{H}$$

и

$$\tilde{h}_n = \frac{1}{n}\sum_{i=1}^n |h(X_i)| - \mathbf{E}|h(X_1)|.$$

Используя условие **A**, получаем

$$\left|\int f\, d(\mathbf{H}_n - \mathbf{H})\right| \leqslant \left(\int f^2\, d\mathbf{P}_0\right)^{1/2}\left(\int\left(\frac{d(\mathbf{H}_n - \mathbf{H})}{d\mathbf{P}_0}\right)^2 d\mathbf{P}_0\right)^{1/2} = o(1).$$

Отсюда, по теореме 28, гл. 9 в [7], для любого $\epsilon > 0$ получаем

$$\mathbf{P}_0(|\tilde{f}_n| > \epsilon) < C_\epsilon n^{1-t/2}. \qquad (3.6)$$

Рассуждая аналогично, получаем

$$\mathbf{P}_0(|\tilde{h}_n| > \epsilon) < C_\epsilon n^{1-t}. \qquad (3.7)$$

Заметим, что теорема 28, гл. 9 в [7], была сформулирована для фиксированной вероятностной меры $\mathbf{P}_0$. Однако анализ доказательств предыдущей теоремы 27 и теоремы 28 показывает, что оценка скорости сходимости выборочных средних, доказанная в них, имеет место равномерно по всем вероятностным мерам, у которых моменты порядка t ограничены константой.

По теореме 28, гл. 9 в [7], получаем

$$\mathbf{P}_0(J_n > C a_n^2) < C n^{1-t/2}. \qquad (3.8)$$

В дальнейших оценках мы будем считать, что имеют место события

$$|\tilde{f}_n| < \epsilon, \quad |\tilde{h}_n| < \epsilon, \quad \text{и} \quad J_n < Ca_n^2. \tag{3.9}$$

Применяя неравенство Чернова, получаем

$$I \leqslant \mathbf{E}_{\bar{\mathbf{P}}_n} \exp\left\{-\epsilon a_n t + t\int f d(\widetilde{\mathbf{P}}_{k_n} - \mathbf{P}^*_{k_n} - a_n\mathbf{H})\right\}. \tag{3.10}$$

Обозначим δ_x дельта-меру в точке $x \in S$. Обозначим

$$M_n = \sum_{h(X_i)>0} h(X_i)\Big(\exp\Big\{\frac{t}{k_n} f(X_i)\Big\} - 1\Big) + \sum_{h(X_i)<0} h(X_i)\Big(\exp\Big\{-\frac{t}{k_n} f(X_i)\Big\} - 1\Big).$$

Имеем

$$\begin{aligned}
&\mathbf{E}_{\bar{\mathbf{P}}} \exp\left\{t\int f d(\widetilde{\mathbf{P}}_{k_n} - \mathbf{P}^*_{k_n} - a_n\mathbf{H})\right\}\\
&= \mathbf{E}_{\bar{\mathbf{P}}} \exp\left\{\frac{t}{n}\sum_{i=1}^{k_n}\int f(x)(d\delta_{\tilde{X}_i^*} - \delta_{X_i^*}) - a_n t\int f\,d\mathbf{H}\right\}\\
&= \Big(1 + \frac{a_n}{n}M_n\Big)^{k_n} \exp\Big\{-a_n t\int f\,d\mathbf{H}\Big\}\\
&= \exp\left\{k_n \log\Big(1 + \frac{a_n}{n}M_n\Big) - a_n t\int f\,d\mathbf{H}\right\}\\
&= \exp\left\{\frac{k_n a_n}{n}M_n - \frac{k_n a_n^2}{2n^2}M_n^2(1+o(1)) - a_n t\int f\,d\mathbf{H}\right\}\\
&= \exp\left\{\frac{a_n t}{n}\sum_{i=1}^n h(X_i)f(X_i) - a_n t\int f\,d\mathbf{H}\right.\\
&\quad\left. - \frac{a_n t^2}{2nk_n}\sum_{i=1}^n |h(X_i)| f^2(X_i) - \frac{a_n^2 k_n}{2n^2}M_n^2(1+o(1))\right\}\\
&= \exp\left\{a_n t\tilde{f}_n - \frac{a_n t^2}{2nk_n}\sum_{i=1}^n |h(X_i)| f^2(X_i) - \frac{a_n^2 k_n}{2n^2}M_n^2(1+o(1))\right\}.
\end{aligned} \tag{3.11}$$

Возьмем $t = C_1\epsilon^{-1}k_n a_n$.

Так как имеет место событие A_{nfh}, то

$$\left|\exp\left\{-\frac{t}{k_n}f(X_i)\right\}-1\right| < Ct(k_nb_n)^{-1} = CC_1\epsilon^{-1}a_nb_n^{-1}.$$

Следовательно,

$$\begin{aligned}\frac{a_n^2k_n}{2n^2}M_n^2 &\leqslant \frac{Ca_n^2k_nt^2}{2k_n^2b_n^2n^2}\left(\sum_{i=1}^n|h(X_i)|\right)\\ &< CC_1^2\epsilon^{-2}a_n^2k_na_n^2b_n^{-2} = o(a_n^2k_n).\end{aligned} \tag{3.12}$$

Поскольку имеет место событие A_{nfh}, то

$$\frac{a_nt^2}{2nk_n}\sum_{i=1}^n|h(X_i)|f^2(X_i) \leqslant Ck_n^{-3}a_n^{-1} \tag{3.13}$$

с вероятностью $1-Cn^{1-t/3}$.

Из (3.9)–(3.13) получаем (3.4), что завершает доказательство верхней границы (2.4) в теореме 2.

Литература

1. J. A. Bucklew, *Large deviations techniques in decision, simulation and estimation.* John Wiley and Sons, NY, 1990.
2. J. A. Bucklew, P. Ney, J. S. Sadowsky, *Monte-Carlo simulation and large deviation theory for uniformly recurrent Markov chains.* — J. Appl. Probab. **27** (1990), 44–59.
3. M. S. Ermakov, *Importance sampling for simulation of moderate deviation probabilities of statistics.* — Statist. Decision **25** (2007), 265–284.
4. М. С. Ермаков, *Принцип больших уклонений для вероятностей умеренных уклонений эмпирических бутстрап мер.* — Зап. научн. семин. ПОМИ **412** (2013), 135–177.
5. F. Gao, X. Zhao, *Delta method in large deviations and moderate deviations for estimators.* — Ann. Statist. **39** (2011), 1211–1240.
6. M. Johns, *Importance sampling for bootstrap confidence intervals.* — Journ. Amer. Statist. Assoc. **83**(403), (1988), 709–714.
7. V. V. Petrov, *Sums of Independent Random Variables.* Springer Verlag, NY, 1975.
8. J. S. Sadowsky, *On Monte-Carlo estimation of large deviation probabilities.* — Ann. Appl. Probab. **6** (1996), 399–422.
9. J. S. Sadowsky, J. A. Bucklew, *On large deviation theory and asymptotically efficient Monte-Carlo estimation.* — IEEE Trans. Inf. Theory **36** (1990), 579–588.
10. Л. Саулис, В. Статулявичус, *Предельные теоремы для больших уклонений.* Мокслас, Вильнюс, 1989.
11. A. W. van der Vaart, J. A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, NY, 1996.

Ermakov M. S. Asymptotically efficient importance sampling for the bootstrap.

We establish Large Deviation Principle for conditional moderate deviation probabilities of weighted bootstrap empirical measures given empirical measures. On the base of this result, for the problem of estimation of moderate deviation probabilities of statistics having Hadamard derivatives, we prove asymptotic efficiency of importance sampling based on influence functions.

Институт Проблем Машиноведения РАН
Большой пр. В.О. 61
С.-Петербург 199178 и
С-Петербургский
государственный университет
Университетский пр., 28, Петродворец
Санкт-Петербург 198504, Россия
E-mail: erm2512@mail.ru

Поступило 10 ноября 2014 г.