УДК 519.233.2

Использование бутстрепа для построения равномерных доверительных интервалов в оценке плотности

А. С. Циглер

Московский физико-технический институт (государственный университет)

Институт проблем передачи информации им. А. А. Харкевича РАН

Аннотация

В данной работе рассматривается модель оценки плотности распределения случайных векторов, предложенная Лоадером в 1996 году. Даётся процедура построения равномерных по конечному множеству точек доверительных интервалов для параметров модели с помощью бутстрепа. Все результаты данной статьи неасимптотические, все константы даны явно.

1 Модель

Пусть $X_1, X_2, \dots X_n$ - независимые одинаково распределенные в $\mathbb{R}^d$ случайные векторы, распределение которых имеет гладкую плотность.

Лоадер в работе [1] показал, что для любой гладкой функции $g(x)$ истинная функция плотности ρ максимизирует следующий функционал:

$$F(\rho) = \mathbb{E}g(X_1)\log(\rho(X_1)) - \int g(x)\rho(x)dx \tag{1}$$

Значит, для оценки значения функции плотности в h-окрестности точки η целесообразно искать максимум функционала $F(\hat{\rho})$ для функции $g(x) = K(\frac{x-\eta}{h})$ по некоторому семейству $\{\hat{\rho}(x, \theta)\}$, где K - некоторая ядерная функция с носителем в единичном шаре.

В данной работе для локальной оценки плотности в окрестности точки η рассматривается семейство $\log(\hat{\rho}(x, \theta)) = \Psi(\frac{x-\eta}{h})^T\theta$, где $\Psi(x)$ - некоторый вектор из p базисных функций.

Так мы пришли к максимизации функционала

$$\mathbb{E}K(\frac{X_1-\eta}{h})\Psi(\frac{X_1-\eta}{h})^T\theta - \int K(\frac{x-\eta}{h})\exp\bigl(\Psi(\frac{X_1-\eta}{h})^T\theta\bigr) \tag{2}$$

Онако, ожидание в функционале нам неизвестно. Поэтому мы можем лишь заменить его на эмпирическое, получив при этом функционал

$$\frac{1}{n}\sum_{i=1}^{n} K(\frac{X_i-\eta}{h})\Psi(\frac{X_i-\eta}{h})^T\theta - \int K(\frac{x-\eta}{h})\exp\bigl(\Psi(\frac{x-\eta}{h})^T\theta\bigr)\,dx \tag{3}$$

Введем

$$S_\eta = \frac{1}{nh^d}\sum \Psi(\frac{X_i-\eta}{h})K(\frac{X_i-\eta}{h}) \quad A(\theta)=\int K(z)\exp(\theta^T\Psi(z))dz \tag{4}$$

Мы пришли к функционалу

$$L(\theta,\eta,h) = \theta^T S_\eta - A(\theta) \tag{5}$$

Обозначим

$$\widetilde{\theta}_\eta = \operatorname{argmax}_\theta L(\theta,\eta,h),\ \theta^*_\eta = \operatorname{argmax}_\theta \mathbb{E}L(\theta,\eta,h) \tag{6}$$

Для дальнейшего исследования нам понадобится ввести

$$D^2(\theta) = -\nabla^2_\theta L(\theta,\eta,h) = \nabla^2_\theta A(\theta) = \int K(z)\Psi(z)\Psi(z)^T\exp\big(\theta^T\Psi(z)\big)dz \tag{7}$$

Введем также обозначение $D^2_\eta = D^2(\theta^*_\eta)$.

Обозначим

$$C_\eta = \sup_{z\in B_1(0)} \|D_\eta^{-1}\Psi(z)\|_2. \tag{8}$$

2 Концентрация, теорема Вилкса

В данной работе мы будем строить доверительные интервалы для параметра θ^*_η в форме

$$\big\{\theta : L(\widetilde{\theta}_\eta,\eta,h) - L(\theta,\eta,h) < \mathtt{z}\big\}, \tag{9}$$

где $\mathtt{z}$ - некоторое число, зависящее от целевой вероятности попадания в доверительный интервал.

Обозначим

$$r_\eta = \|D_\eta(\widetilde{\theta}_\eta - \theta^*_\eta)\|_2.$$

Для того, чтобы не работать напрямую с функцией $L(\theta,\eta,h)$, нам понадобятся следующие теоремы, которые доказаны для широкого класса моделей в работе [3].

Теорема 1 (о концентрации).

$$\|D_\eta^{-1}(S_\eta - \mathbb{E}S_\eta)\|2 \geq \sup_{r<r_\eta}\frac{1}{2}r\exp\big(-rC_\eta\big) \tag{10}$$

Теорема 2 (Вилкс).

$$\Big|2L(\widetilde{\theta}_\eta,\eta,h) - 2L(\theta^*_\eta,\eta,h) - \|D_\eta^{-1}(S_\eta-\mathbb{E}S_\eta)\|_2^2\Big| \leq \\ \leq r_\eta^2\exp(r_\eta C_\eta)(\exp(r_\eta C_\eta)-1) \tag{11}$$

Из теоремы Вилкса из предыдущей главы легко видеть, что задача свелась к аппроксимации распределения случайной величины $\|D_\eta^{-1}(S_\eta - \mathbb{E}S_\eta)\|^2$

Пусть $\Xi = \{\eta_i\}_{i=1}^M$ - множество точек, в которых мы хотим построить доверительные интервалы.

Теперь можно считать $S \in \mathbb{R}^{pM}$. То есть S - вектор, у которого есть M блоков по p компонент, каждый из которых соответствует одной из точек η_i. Также будем обозначать за D^2 матрицу состоящую из M^2 блоков, соответствующих парам (η_i, η_j), причем блок, соответствующий (η_i, η_j), равен $D_{\eta_i}\delta_{i,j}$.

3 Гауссовская аппроксимация

S - среднее независимых одинаково распределенных векторов. Естественно, возникает желание заменить вектор S на гауссовский.

Обозначим
$$S_{i,\eta} = \frac{1}{h^d}\Psi(\frac{X_i - \eta}{h})K(\frac{X_i - \eta}{h}) \tag{12}$$
-независимые одинаково распределенные векторы.

За S_i обозначим вектор $(S_{\eta_1}^T, \ldots, S_{\eta_M}^T)^T$ (аналогично вектору S).

Введем также
$$\widetilde{S}_i \in \mathcal{N}(\mathbb{E}S_i, \operatorname{Var} S_i) \tag{13}$$
- независимые гауссовские векторы, независимые также и от $\{S_i\}$.

Обозначим
$$\widetilde{S} = \frac{1}{n}\sum_i \widetilde{S}_i \tag{14}$$
.

Возьмем
$$f = g \circ F_\beta, \tag{15}$$
где
$$F_\beta(x) = \beta^{-1}\log\Big(\sum_j \exp\big(\beta(\|x_{\eta_j}\|_2^2 - q)\big)\Big), \tag{16}$$
$g(x) = g_0(\phi^{-1}x)$, а $g_0 : \mathbb{R} \to \mathbb{R}$ - некоторая монотонная гладкая функция, такая что
$$g_0(x) = \begin{cases} 1, & x \le 0, \\ 0, & x \ge 1. \end{cases} \tag{17}$$

Обозначим за G_1, G_2, G_3 соответственно максимумы первой, второй и третьей производных функции g_0.

Обозначим $L = \sup_{x,j}\|D_{\eta_j}^{-1}\Psi(\frac{\eta_j - x}{h})K(\frac{\eta_j - x}{h})\|_2$.

Введем

$$\chi_1(q) = \frac{4}{3(nh^d)^3} L^3 \left(1 + 2\sqrt{\frac{2}{\pi}}\right) \big(\phi^{-1} G_1 (24\beta q^{\frac{1}{2}} + 32\beta^2 q^{\frac{3}{2}} + 8\beta q^{\frac{3}{2}}) + \\ + \phi^{-2} G_2 (12 q^{\frac{1}{2}} + 48\beta q^{\frac{3}{2}}) + \phi^{-3} G_3 (8 q^{\frac{3}{2}})\big). \quad (18)$$

Тогда выполено следующее утверждение:

$$\left| \mathbb{E}\Big[f\big(D^{-1}(S - \mathbb{E}S)\big) - f\big(D^{-1}(\widetilde{S} - \mathbb{E}S)\big) \Big] \right| \le \chi_1(q + \phi) \quad (19)$$

4 Сравнение гауссовских распределений

Для бутстрепа мы будем генерировать векторы

$$S^\flat_{\eta_j} = \frac{1}{n} \sum_i S_i w_i^\flat = \frac{1}{nh^d} \sum_1^n \Psi\Big(\frac{X_i - \eta_j}{h}\Big) K\Big(\frac{X_i - \eta_j}{h}\Big) w_i^\flat, \quad (20)$$

где w_i - веса из $\mathcal{N}(1,1)$.

Тогда

$$D^{-1}(\widetilde{S} - \mathbb{E}S) \in \mathcal{N}\Big(0, \frac{1}{n} D^{-1}\big(\mathbb{E}S_1 S_1^T - \mathbb{E}S_1 \mathbb{E}S_1^T\big) D^{-1}\Big) \quad (21)$$

$$D^{-1}(S^\flat - S) \in \mathcal{N}\Big(0, \frac{1}{n} D^{-1}\big(\frac{1}{n} \sum S_i S_i^T\big) D^{-1}\Big) \quad (22)$$

Введем

$$\chi_2(q) = \\ = \frac{L^2}{2n}\Big(\frac{2\sqrt{2}(\tau + 2\log M)}{nh^{2d}} + \Upsilon_d^2 \max_x \rho(x)^2\Big)\Big(G_1\big(2p^2 + 8\beta q\big) + 4G_2 q\Big), \quad (23)$$

где Υ_d - объем единичного шара в d-мерном пространстве.

Верен следующий факт:

Для любого положительного τ с вероятностью $\ge 1 - 2\exp(-\tau)$

$$\left| \mathbb{E} f\Big(D^{-1}\big(\widetilde{S} - \mathbb{E}S\big)\Big) - \mathbb{E} f\Big(D^{-1}\big(S^\flat - S\big)\Big) \right| \le \chi_2(q + \phi) \quad (24)$$

5 Совмещение результатов

Совместим результаты предыдущих двух секций и следующий факт:

$$I\{\max_j(\|x_{\eta_j}\|^2 - q) + \beta^{-1}\log(M) < 0\} \le f(x) \le I\{\max_j(\|x_{\eta_j}\|^2 - q) - \phi < 0\}. \quad (25)$$

С вероятностью $\geq 1-2\exp(-\tau)$

$$\mathbb{P}\big\{\max_j \|D_{\eta_j}^{-1}(S_{\eta_j}-\mathbb{E}S_{\eta_j})\|_2^2 < q^2\big\} \geq$$
$$\geq \mathbb{P}\big\{\max_j \|D_{\eta_j}^{-1}(S^{\flat}_{\eta_j}-\mathbb{E}S^{\flat}_{\eta_j})\|_2^2 < q^2-\phi-\beta^{-1}\log(M)\big\} - \chi_2(q)-\chi_1(q) \tag{26}$$

Осталось лишь вернуться к функции $L(\theta,\eta,h)$.

Обозначим $C=\max_i C_{\eta_i}$.

Совмещая результаты теорем 1 и 2 с полученными в предыдущих двух главах, можно показать основной результат данной работы:

Теорема 3. Пусть $q<\frac{1}{eC}$, . Тогда для любых положительных ϕ,β выполнено следующее:

$$\mathbb{P}\Big\{\max_j\big[2L(\widetilde{\theta}_\eta,\eta,h)-2L(\theta^*_\eta,\eta,h)\big]\leq q^2\big(1+4\exp(3eqC)(\exp(eqC)-1)\big)\Big\}\geq$$
$$\geq \mathbb{P}\Big\{\max_j\big[2L(\widetilde{\theta}^{\flat}_{\eta_j},\eta,h)-2L(\widetilde{\theta}_{\eta_j},\eta,h)\big]\leq (q^2-\phi-\beta^{-1}\log(M))\times$$
$$\times\big(1-4\exp(3eC\sqrt{q^2-\phi-\beta^{-1}\log(M)})\times$$
$$\times(\exp(eC\sqrt{q^2-\phi-\beta^{-1}\log(M)})-1)\big)\Big\}-\chi_2(q)-\chi_1(q) \tag{27}$$

Для каждого значения q эта теорема даёт возможность оценить вероятность события

$$\Big\{\max_j\big[2L(\widetilde{\theta}_\eta,\eta,h)-2L(\theta^*_\eta,\eta,h)\big]\leq q^2\big(1+4\exp(3eqC)(\exp(eqC)-1)\big)\Big\}, \tag{28}$$

а значит и построить доверительный интервал.

Список литературы

[1] Loader, C.: Local Likelihood Density Estimation. Ann. Statist. Volume 24, Number 4, 1602-1618 (1996)

[2] Chernozhukov, V., Chetverikov, D. and Kato, K.: Gaussian approximations and multiplier bootstrap for maxima of sums of highdimensional random vectors. Ann. Statist. 41 2786-2819 (2013).

[3] Spokoiny, V.: Parametric estimation. Finite sample theory. Ann. Statist. Volume 40, Number 6, 2877-2909 (2012) http://arxiv.org/abs/1111.3029