Теоремы Фишера и Уилкса для оценки лог-плотности правдоподобия

С. А. Довгаль

Кафедра проблем передачи информации и анализа данных,
Московский физико-технический институт (ГУ)

Научный руководитель к.ф.-м.н., проф. В. Г. Спокойный

Москва, 27 июня 2016

План

План

Постановка задачи

Задано: Выборка из независимых одинаково распределённых случайных величин

$$X_1, X_2, \ldots, X_n \in \mathbb{R}^d, \quad X_i \sim f(x)$$

$f(x)$ — любая «достаточно гладкая» (определение ниже).

Цель:

- Оценить производные $\log f(x)$ в точке x_0.
- Оценка, минимизирующая «несмещённый ℓ_2-риск»
$$\|S \cdot (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\bullet})\|^2 \to \min$$

Известный оптимальный результат:

- $h(n) = \arg\min_h \big(O(h^p) + O_p\big((nh^d)^{-1/2}\big)\big)$

Модель

Функция квази-лог-правдоподобия:

$$\mathrm{L}(\boldsymbol{\theta}; \mathrm{X}, \mathrm{x}_0, \mathrm{h}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{K}_\mathrm{i} \boldsymbol{\Psi}_\mathrm{i}^\mathrm{T} \boldsymbol{\theta} - \mathrm{n} \int \mathrm{K} \exp(\boldsymbol{\Psi}^\mathrm{T} \boldsymbol{\theta}) \mathrm{dx} \;, \tag{1}$$

- $\boldsymbol{\Psi}_\mathrm{i} = (\psi_0(\mathrm{T}_\mathrm{i}), \psi_1(\mathrm{T}_\mathrm{i}), \ldots, \psi_{\mathrm{p}-1}(\mathrm{T}_\mathrm{i}))^\mathrm{T}$ — базис
- $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{\mathrm{p}-1})$ — целевой параметр
- $\mathrm{K}_\mathrm{i} = \mathrm{K}(\mathrm{T}_\mathrm{i})$ — ядро: $\mathrm{supp}\, \mathrm{K} \subseteq [-1, 1]$, $\mathrm{K}(\mathrm{t}) \leq 1$,
- $\mathrm{T}_\mathrm{i} = \dfrac{\mathrm{X}_\mathrm{i} - \mathrm{x}_0}{\mathrm{h}}$ — нормированная замена переменных

$$\widetilde{\boldsymbol{\theta}} \;=\; \arg\max_{\boldsymbol{\theta}} \mathrm{L}(\boldsymbol{\theta}), \quad \boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathrm{EL}(\boldsymbol{\theta}) \;, \tag{2}$$

Ключевые объекты

• Матрица информации (кривизны) D_n^2,
• квазигауссовский вектор $\boldsymbol{\xi}_n$, • матрица ковариации V_n^2:

$$\begin{aligned} D_n^2 &= -\nabla^2 \mathrm{E}L(\boldsymbol{\theta}^*) \ , & D_n^2(\boldsymbol{\theta}) &= -\nabla^2 \mathrm{E}L(\boldsymbol{\theta}) \ , \\ \boldsymbol{\xi}_n &= D_n^{-1} \nabla L(\boldsymbol{\theta}^*) \ , & V_n^2 &= \mathrm{Var}\left(\nabla L(\boldsymbol{\theta}^*)\right) \ . \end{aligned} \tag{3}$$

• Окрестность концентрации для $\widetilde{\theta}$:

$$\Theta_n(z) = \left\{ \boldsymbol{\theta} \colon \|D_n(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \le r_0(z) \right\} \ . \tag{4}$$

Несмещённый параметр

- Несмещённый параметр $\boldsymbol{\theta}^\bullet$:

$$\varphi(\mathrm{x}) = \log \mathrm{f}(\mathrm{x}).\ \exists \boldsymbol{\theta}^\bullet:\ \forall \mathrm{t} \in [-1, 1]$$

$$\varphi(\mathrm{x}_0 + \mathrm{th}) - \boldsymbol{\Psi}^{\mathrm{T}}(\mathrm{t})\boldsymbol{\theta}^\bullet \leq \mathrm{M}_{\mathrm{p,h}} \asymp \mathrm{O}(\mathrm{h}^{\mathrm{p}})\ ,$$

- Пример: одномерный полиномиальный базис

$$\log \mathrm{f}(\mathrm{x}) = \theta_0^\bullet + \theta_1^\bullet \frac{\mathrm{x} - \mathrm{x}_0}{\mathrm{h}} + \ldots + \theta_{\mathrm{p}-1}^\bullet \left(\frac{\mathrm{x} - \mathrm{x}_0}{\mathrm{h}}\right)^{\mathrm{p}-1} + \mathrm{R}_{\mathrm{h}}(\mathrm{x})$$

$$\theta_{\mathrm{j}}^\bullet(\mathrm{h}) = \frac{\mathrm{h}^{\mathrm{j}}}{\mathrm{j}!} \frac{\partial^{\mathrm{j}} \log \mathrm{f}(\mathrm{x})}{\partial \mathrm{x}^{\mathrm{j}}}\bigg|_{\mathrm{x}=\mathrm{x}_0}\ . \qquad (5)$$

План

Теорема о концентрации

Теорема (Теорема о концентрации)

Пусть условия $(\boldsymbol{\mathcal{I}})$, $(\boldsymbol{\mathcal{L}}_0)$, $(\boldsymbol{\mathcal{C}})$, (ED_0) выполнены с некоторыми константами $\mathfrak{a}$, δ, n_0, ν_0. Пусть

$$\Theta_{\mathrm{n}}(\mathrm{z}) = \left\{\boldsymbol{\theta} \colon \|\mathrm{D}_{\mathrm{n}}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \le \mathrm{r}_0(\mathrm{z})\right\}, \tag{6}$$

$$\mathrm{r}_0(\mathrm{z}) = 4\mathfrak{a} \cdot \nu_0(\sqrt{\mathrm{p}} + \sqrt{2\mathrm{z}}), \qquad \mathrm{z} \le \mathfrak{g}^2/4 . \tag{7}$$

Тогда

$$\mathrm{P}\left(\widetilde{\boldsymbol{\theta}} \notin \Theta_{\mathrm{n}}(\mathrm{z})\right) \le 2\mathrm{e}^{-\mathrm{z}} + 8.4\mathrm{e}^{-\mathfrak{g}^2/4} . \tag{8}$$

Асимптотический аналог: $\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \mathrm{O}((\mathrm{nh}^{\mathrm{d}})^{-1/2})$.

Теорема Фишера

Теорема (Роберт Фишер)

Пусть выполнены условия $(\boldsymbol{\mathcal{C}})$, $(\boldsymbol{\mathcal{I}})$, (ED_0), $(\boldsymbol{\mathcal{L}}_0)$.
Тогда для $\widetilde{\boldsymbol{\theta}} \in \Theta_n(z)$ с вероятностью

$$\mathrm{P}\left(\widetilde{\boldsymbol{\theta}} \in \Theta_n(z)\right) \geq 1 - \left(2e^{-z} + 8.4e^{-\mathfrak{g}^2/4}\right) \tag{9}$$

выполнено

$$\|D_n(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) - \boldsymbol{\xi}_n\| \leq r_0(z) \cdot \delta_n(r_0) \ , \quad z \leq \mathfrak{g}^2/4 \ , \tag{10}$$

<u>Асимптотический аналог:</u> вместо $O((nh^d)^{-1/2})$ в ЦПТ:

$$\left\| d_0(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) - (nh^d)^{-1/2}\boldsymbol{\xi}_n \right\| = O((nh^d)^{-1})$$

Теорема Уилкса

Теорема (Уилкс)

Пусть выполнены условия $(\boldsymbol{\mathcal{C}})$, $(\boldsymbol{\mathcal{I}})$, $(\mathbb{E}D_0)$, $(\boldsymbol{\mathcal{L}}_0)$.
Тогда для $\widetilde{\boldsymbol{\theta}} \in \Theta_n(z)$ с вероятностью

$$\mathrm{P}\left(\widetilde{\boldsymbol{\theta}} \in \Theta_n(z)\right) \geq 1 - (2e^{-z} + 8.4e^{-\mathfrak{g}^2/4}) \tag{11}$$

$$\left|\sqrt{2L(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)} - \|D_n(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\circ)\|\right| \leq 2r_0(z) \cdot \delta_n(r_0) \tag{12}$$

$$\left|\sqrt{2L(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)} - \|\boldsymbol{\xi}\|\right| \leq 3r_0(z) \cdot \delta_n(r_0) \tag{13}$$

Точное малое смещение

Теорема (Точное малое смещение)

- Пусть «h мало», «$\mathrm{nh^d}$ велико».
- Тогда $\phi_1, \phi_2, \varepsilon \asymp \mathrm{O(h)}$, $|\mathrm{B_{p,h}} - 1| \asymp \mathrm{O(h^p)}$, и т.д.

$$\|\mathrm{d}_0(\boldsymbol{\theta}^\circ)(\boldsymbol{\theta}^* - \boldsymbol{\theta}^\bullet)\| \lesssim \sqrt{\mathrm{p}}\sqrt{\mathrm{I_K}}(1-\varepsilon)^{-1}(1+\mathrm{c_{f,h}})\cdot \mathrm{f}(\mathrm{x}_0)\cdot|\mathrm{B_{p,h}}-1| \ .$$

<u>Асимптотический аналог:</u>

$$\|\mathrm{d}_0(\boldsymbol{\theta}^\circ)(\boldsymbol{\theta}^* - \boldsymbol{\theta}^\bullet)\| \lesssim \sqrt{2^{\mathrm{d}}}\mathrm{f}(\mathrm{x}_0)\mathrm{O}(\mathrm{h^p}) \ .$$

Следствие: оптимальная оценка

$$\|\mathrm{D_n}(\boldsymbol{\theta}^\circ)(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|^2 \to \min$$

Теорема о концентрации + Точное малое смещение:

$$\mathrm{h(n)} = \arg\min_{\mathrm{h}} \left(\underbrace{\mathrm{O(h^p)}}_{\text{Th. 4}} + \underbrace{\mathrm{O_p}\big((\mathrm{nh^d})^{-1/2}}_{\text{Th. 1}}\big) \right)$$

Выражение становится неасимптотическим.

План

1. Постановка задачи и модель
2. Основные результаты
   - Теорема о концентрации
   - Теорема Фишера и Уилкса
   - Точное малое смещение
3. Неасимптотические условия
   - Условия из техники Спокойного
   - Оракульные условия и константы
   - Другие результаты

Условия из техники Спокойного

$(\boldsymbol{\mathcal{I}})$ $\exists \mathfrak{a} > 0$, такая, что

$$\mathfrak{a}^2 \mathrm{D}_{\mathrm{n}}^2 \succeq \mathrm{V}_{\mathrm{n}}^2 \ . \tag{14}$$

$(\boldsymbol{\mathcal{L}}_0)$ $\exists \delta_{\mathrm{n}}(\mathrm{r}_0)\colon\ \forall \boldsymbol{\theta} \in \Theta_{\mathrm{n}}(\mathrm{z})$:

$$\|\mathrm{I}_{\mathrm{p}} - \mathrm{D}_{\mathrm{n}}^{-1} \mathrm{D}_{\mathrm{n}}^2(\boldsymbol{\theta}) \mathrm{D}_{\mathrm{n}}^{-1}\| \le \delta_{\mathrm{n}}(\mathrm{r}_0) \ . \tag{15}$$

(ED_0) $\boldsymbol{\zeta} = \mathrm{V}_{\mathrm{n}}^{-1} \nabla \mathrm{L}$. $\exists \mathfrak{g} > 0$, $\nu_0 > 0\colon\ \forall \boldsymbol{\gamma} \in \mathbb{R}^{\mathrm{p}}$

$$\log \mathrm{E} \exp\left(\lambda \frac{\boldsymbol{\gamma}^{\mathrm{T}}(\boldsymbol{\zeta} - \mathrm{E}\boldsymbol{\zeta})}{\|\boldsymbol{\gamma}\|}\right) \le \frac{\nu_0^2 \lambda^2}{2} \qquad \forall \lambda\colon |\lambda| \le \mathfrak{g} \tag{16}$$

$(\boldsymbol{\mathcal{C}})$ $\mathrm{r}_0(1 - \delta_{\mathrm{n}}(\mathrm{r}_0)) \ge 2\zeta(\mathrm{p}, \mathrm{z})$.

Оракульные условия

- Условие малой осцилляции:

$$\forall \mathrm{x}\colon |\mathrm{x}-\mathrm{x}_0| \le \mathrm{h} \to \left|1-\frac{\mathrm{f}(\mathrm{x})}{\mathrm{f}(\mathrm{x}_0)}\right| \le \mathrm{c}_{\mathrm{f},\mathrm{h}}$$

- Условие малого смещения: $\exists \boldsymbol{\theta}^{\bullet}$, $\exists \mathrm{B}_{\mathrm{p},\mathrm{h}}\colon \forall \mathrm{t} \in [-1,1]$

$$\mathrm{B}_{\mathrm{p},\mathrm{h}} \ge \exp\left(\varphi(\mathrm{x}_0+\mathrm{th}) - \boldsymbol{\Psi}^{\mathrm{T}}(\mathrm{t})\boldsymbol{\theta}^{\bullet}\right) \quad ,$$

Условия на модель

Условие оптимизации кривой:

$$\mathfrak{c}_1^2 = \sup_{\mathrm{t}\in[-1,1]} \boldsymbol{\Psi}^{\mathrm{T}}(\mathrm{t}) \left[\int_{-1}^{1} \mathrm{K}(\tau)\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}^{\mathrm{T}}(\tau)\mathrm{d}\tau \right]^{-1} \boldsymbol{\Psi}(\mathrm{t}) \ . \tag{17}$$

- Одномерный полиномиальный случай: $\mathfrak{c}_1^2 = \mathrm{p}^2/2$
- Двумерный квадратичный: $\mathfrak{c}_1^2 = 13/2$.

Условия на ширину ядра

$$\mathrm{f_0 = f(x_0), \quad I_k = \int_{-1}^{1} K(\tau) d\tau \le 2^d}$$

$$\begin{aligned}
\phi_1^2 &= 2\mathrm{I_k} \cdot \max\Big\{ |\mathrm{c_{f,h}}(\log \mathrm{f_0} - 1) + (1 + \mathrm{c_{f,h}}) \log(1 + \mathrm{c_{f,h}})|, \\
&\qquad |-\mathrm{c_{f,h}}(\log \mathrm{f_0} - 1) + (1 - \mathrm{c_{f,h}}) \log(1 - \mathrm{c_{f,h}})| \Big\} \ , \\
\phi_2^2 &= \mathrm{I_K d\tau \cdot f_0^3 (c_{f,h} - \log B_{p,h})^2} \ .
\end{aligned}$$

- $\mathfrak{c}_1 \phi_1 < \frac{\sqrt{5} - 1}{2}$
- $\mathfrak{c}_1 \phi_2 < 1$

Условие на эффективный размер выборки

$$\sqrt{\mathrm{nh^d}} \geq \mathrm{f}(\mathrm{x}_0)\frac{4\mathfrak{c}_1\zeta(\mathrm{p},\mathrm{z})}{\log 3/2\sqrt{1-\mathfrak{c}_1\phi_1}} \quad , \quad \zeta(\mathrm{p},\mathrm{z}) = \mathfrak{a}\nu_0(\sqrt{\mathrm{p}}+\sqrt{2\mathrm{z}}) \tag{18}$$

• Опасный поворот! Области с низкой плотностью нужно рассматривать по-другому.

Другие результаты

Лемма (Аналог Коши-Буняковского для матриц и векторов)
Пусть $\boldsymbol{\psi}(\mathrm{t})\colon [-1,1] \to \mathbb{R}^{\mathrm{p}}$ — некоторая векторнозначная интегрируемая функция, $\delta(\mathrm{t})\colon [-1,1] \to \mathbb{R}$ — интегрируемая функция. Тогда выполнено следующее матричное неравенство:

$$\int_{-1}^{1} \boldsymbol{\psi}(\mathrm{t})\delta(\mathrm{t})\mathrm{dt} \int_{-1}^{1} \boldsymbol{\psi}^{\mathrm{T}}(\mathrm{t})\delta(\mathrm{t})\mathrm{dt} \leq \int_{-1}^{1} \boldsymbol{\psi}(\mathrm{t})\boldsymbol{\psi}^{\mathrm{T}}(\mathrm{t})\mathrm{dt} \int_{-1}^{1} \delta^{2}(\mathrm{t})\mathrm{dt} \tag{19}$$

Другие результаты

Лемма (Оценка на константу $\mathfrak{c}_1$, полиномиальный случай)
Пусть $\mathbf{\Psi}(\mathrm{t}) = (1, \mathrm{t}, \mathrm{t}^2, \ldots, \mathrm{t}^{\mathrm{p}-1})^{\mathrm{T}}$. Рассмотрим матрицу

$$\mathrm{A}^2 = \int_{-1}^{1} \mathbf{\Psi}(\mathrm{t})\mathbf{\Psi}^{\mathrm{T}}(\mathrm{t})\mathrm{d}\mathrm{t} \ . \tag{20}$$

Тогда полином, определённый формулой

$$\mathrm{P}(\mathrm{t}) = \mathbf{\Psi}^{\mathrm{T}}(\mathrm{t})\mathrm{A}^{-2}\mathbf{\Psi}(\mathrm{t}) \tag{21}$$

на отрезке $[-1, 1]$ достигает максимума в точках $\mathrm{t} = \pm 1$, и это значение равно p^2.

Лемма
Пусть $\mathrm{P}(\mathrm{t}) = \boldsymbol{\Psi}^{\mathrm{T}}(\mathrm{t})\mathrm{A}^{-2}\boldsymbol{\Psi}(\mathrm{t})$, $\mathrm{Q}(\mathrm{t}) = \boldsymbol{\Psi}^{\mathrm{T}}(\mathrm{t})\mathrm{B}^{-2}\boldsymbol{\Psi}(\mathrm{t})$, где $\mathrm{A}^2 \asymp \mathrm{D}_{\mathrm{n}}^2$, $\mathrm{B}^2 \asymp \mathrm{V}_{\mathrm{n}}^2$. Тогда

$$\sup_{\mathrm{t}\in[-1,1]} \mathrm{Q}(\mathrm{t}) \lesssim \sup_{\mathrm{t}\in[-1,1]} \mathrm{P}(\mathrm{t}) + \mathrm{O}(\mathrm{h}^{\mathrm{p}}) \qquad (22)$$

• <u>Идея:</u> одноранговое преобразование Шермана–Моррисона.

Заключение

- Неасимптотическая оценка, согласуется с известным ответом
- Основа для решения практических задач: доверительные интервалы, полосы
- Факты из линейной алгебры

Спасибо за внимание