Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Барсуков Егор Вячеславович

Оптимальные планы для оценивания производных в полиномиальной регрессионной модели без свободного члена

Отчет о научно-исследовательской работе

Научный руководитель:
д. ф.-м. н., профессор В. Б. Мелас

Санкт-Петербург

2019

Оглавление

Введение

Рассмотрим регрессионную модель

$$y_j = \theta^\top f_j(x_j) + \varepsilon_j, \quad j = 1 \dots N,\, x_j \in \mathcal{X}, \tag{1}$$

где N — количество экспериментов, $\mathcal{X} \subset \mathbb{R}$, $f(x) = (f_1(x), \dots, f_n(x))^\top$ — регрессионная функция, $\theta = (\theta_1, \dots, \theta_n)^\top$ — неизвестные параметры, ε_i — некоррелированные ошибки наблюдения. При этом $\mathrm{E}[\varepsilon_i] = 0$, $\mathrm{D}[\varepsilon_i] = \sigma^2$.

Для улучшения в каком-либо смысле некоторых оценок по данной модели при минимизации N строят *планы эксперимента*, т.е. наборы точек $x_i \in \mathcal{X}$ в каждой из которых должно быть произведено n_i экспериментов так, что $\sum_i^N n_i = N$. В каком именно смысле будет улучшена оценка зависит от одного из множества критериев оптимальности. В этой работе будут рассматриваться c-оптимальные планы эксперимента, определение которым будет дано далее.

В работе будут рассматриваться полиномиальные регрессионные модели, т.е. $f_i(x) = x^{k_i}$. Такие модели были хорошо изучены и для многих случаев были построены явные аналитические решения. Множество работ были посвящены нахождению D-оптимальных планов [?, ?, ?, ?]. Также существуют для такой модели явные решения для нахождения E-оптимальных планов [?, ?, ?, ?].

C-оптимальным планом эксперимента является план минимизирующий дисперсию значения скалярного произведения θ и c для заданного $c \in \mathbb{R}^n$. В общем случае нахождение c-оптимальных планов может быть достаточно сложно. Некоторые решения с очень маленьким количеством неизвестных параметров описаны опираясь на использование теоремы Элвинга [?].

Часто рассматриваются задача нахождения оптимального плана при $c = f(z)$ для произвольного z — задача экстраполяции и $c = f'(z)$ — задача оценки производной, они были решены в общем виде для обычной полиномиальной регрессии вида $f(x) = (1, x, \dots, x^n)^\top$ [?, ?].

В этой работе рассмотрен случай нахождения нахождения плана для оценивания производной при полиномиальной модели без свободного члена при $\mathcal{X} = [0, d]$. Этот случай существенно отличается от $\mathcal{X} = [-1, 1]$, который был рассмотрен в [?]. Также был значительно улучшен алгоритм численного нахождения c-оптимальных планов эксперимента для произвольного c.

Глава 1

Общие сведения

1.1. Определения

Согласно [?] непрерывным планом эксперимента в регрессионной модели (1) будем называть дискретную вероятностную меру

$$\xi = \begin{pmatrix} x_1 & \dots & x_m \\ \omega_1 & \dots & \omega_m \end{pmatrix}, \quad x_i \in \mathcal{X}.$$

Для проведения N измерений с таким планом эксперимента необходимо провести $n_i \approx N\omega_i$ измерений в точке x_i таким образом, чтобы $\sum_i^m n_i = N$.

Для непрерывного плана эксперимента определим информационную матрицу следующим образом

$$M(\xi) = \int_{\mathcal{X}} f(x) f^\top(x) \xi(dt).$$

Определение 1. *c-оптимальным планом* для некоторого вектора c называется план эксперимента ξ минимизирующий следующую функцию

$$\Phi(\xi) = \begin{cases} c^\top M(\xi)^- c, & \text{если существует } v, \text{ такой, что } \; c = M(\xi)v \\ +\infty, & \text{иначе} \end{cases},$$

где $M(\xi)^-$ — матрица, обобщенно обратная к информационной матрице плана ξ. План называется допустимым, если существует такое v, что $c = M(\xi)v$.

Как было отмечено в введении, c-оптимальный план минимизирует дисперсию несмещенной МНК оценки $c^\top \hat{\theta}$ линейной комбинации $c^\top \theta$ [?].

Определение 2. Если $c = f(z)$ для некоторого $z \in \mathbb{R}$, то соответствующий c-оптимальный план называется *оптимальном планом экстраполяции* в точке z.

Определение 3. Если $c = f'(z)$ для некоторого $z \in \mathbb{R}$, то соответствующий c-оптимальный план называется *оптимальным планом для оценки производной* в точке z.

1.2. Теорема Элвинга

Для решения задачи нахождения c-оптимальных планов в множестве случаев (в том числе в данной работе) используется теорема Элвинга, являющаяся геометрически интерпретируемым критерием c-оптимальности плана эксперимента.

Теорема 1. (Элвинга) [?] Допустимый план $\xi^\star$ с носителем $x_1, \dots, x_m \in \mathcal{X}$ и весами $\omega_1, \dots, \omega_m$ является c-оптимальным тогда и только тогда, когда существует $p \in \mathbb{R}^k$ и константа h такие, что выполняются следующие условия:

$$\left|p^\top f(x_i)\right| = 1 \qquad i = 1..m \leqslant n \tag{1.1a}$$

$$\left|p^\top f(x)\right| \leqslant 1 \qquad x \in \mathcal{X} \tag{1.1b}$$

$$c = h \sum_{i=1}^{m} \omega_i f(x_i) p^\top f(x_i). \tag{1.1c}$$

Кроме того

$$h^2 = c^\top M^-(\xi^*) c$$

Функция $p^\top f(x_i)$ в определениях теоремы Элвинга называют *экстремальным многочленом.*

1.3. Явная формула для весов оптимального плана

Теорема 2. Оптимальный план для оценивания производной полиномиальной модели без свободного члена с опорными точками $t_1^*, \dots, t_m^*$, где $m = n$ или $m = n - 1$ имеет веса вычисленные по следующей формуле:

$$\omega_i = \frac{|L_i'(z)|}{\sum_{j=1}^m \left|L_j'(z)\right|}, \tag{1.2}$$

где L_i задается следующим образом

$$L_i(x) = \frac{x \prod_{l=1}^n (x - t_l^*)}{t_i^* \prod_{l \neq i}^n (t_i^* - t_l^*)},$$

то есть является i-ым базисным многочленом Лагранжа без нулевого члена построенным по точкам $t_1^*, \dots, t_m^*$.

Глава 2

План для нахождения производной на промежутке $[0, d]$

Построим полином без свободного члена $S_n(x)$ степени n, не превосходящий по модули единицу на промежутке $[0, 1]$, и достигающий её в n точках. Пусть $T_n(x)$ — многочлен Чебышёва степени n. Тогда по свойствам многочленов Чебышёва T_n не превосходит по модулю единицу на промежутке $[-1, 1]$ и достигает её в $n+1$ точках, в том числе в точках -1 и 1. Известно, что корни T_n имеют следующий вид

$$x_i = \cos\left(\frac{\pi(i + 1/2)}{n}\right), \quad i = 0, \ldots, n-1.$$

Если мы возьмём самый маленький корень $x_{\min} = -\cos\frac{\pi}{2n}$ и положим $\widehat{S_n}(x) = T_n(x + x_{\min})$, то $\widehat{S_n}$ будет являться полиномом степени n с нулевым свободным членом, так как $\widehat{S_n}(0) = 0$ по построению. При этом $\left|\widehat{S_n}(x)\right| \leqslant 1$ для x на промежутке $[0, 1 + \cos\left(\frac{\pi}{2n}\right)]$, при этом в этом промежутке абсолютная величина достигает единицы n раз в силу того, что левый край не равен 1.

Для того, чтобы привести промежуток к виду $[0, 1]$ достаточно добавить множитель $1 + \cos\frac{\pi}{2n}$ к x в левой части определения $\widehat{S_n}$. После этого получается многочлен, удовлетворяющий всем требуемым свойствам

$$S_n(x) = T_n\left(x\left(1 + \cos\frac{\pi}{2n}\right) - \cos\frac{\pi}{2n}\right).$$

Для более общего случая в виде промежутка $[0, d]$ требуемый многочлен (обозначим его S_n^d) можно выразить из $S_n(x)$ как $S_n^d(x) = S_n\left(\frac{x}{d}\right)$.

Исходя из данного построения и известных экстремальных точек многочлена Чебышёва, можно легко выразить экстремальные точки S_n. Обозначим их как $s_{i,n}$, тогда

$$s_{i,n} = \frac{\cos\frac{(n-i)\pi}{n} + cos\frac{\pi}{2n}}{1 + \cos\frac{\pi}{2n}}, \quad i = 1, \ldots, n$$

при этом

$$0 < s_{1,n} < \ldots < s_{n,n}$$

и

$$S_n(s_{i,n}) = (-1)^{n+i}, \quad i = 1, \ldots, n$$

Построим базисные полиномы Лагранжа степени n без нулевого члена по точкам $\{s_{i,n}\}_{i=1}^n$

$$L_i(x) = \frac{x \prod_{l=1}^n (x - s_{l,n})}{s_{i,n} \prod_{l \neq i}^n (s_{i,n} - s_{l,n})}$$

Так как теорема 4 о весах оптимального плана в случае полиномиальной модели работает для любых промежутков, то для нахождения весов можно использовать её.

$$\omega_i = \frac{|L_i'(z)|}{\sum_{j=1}^n \left| L_j'(z) \right|}$$

В силу свойств многочленов Чебышёва $S_n(x)$ и $-S_n(x)$ — единственные многочлены степени n без нулевого члена, которые удовлетворяют свойствам 1-2 теоремы Элвинга,поэтому осталось проверить для только свойство 3. Для этого введем обозначения $F = \left(s_{j,n}^i\right)_{i,j=1}^n$, $h = \sum_{j=1}^n \left| L_j'(z) \right|$ и $\beta = \left(|L_i'(z)| \, (-1)^{i+n} \right)_{i=1}^n$. Так как $s_{j,n}$ при $i = 1, \ldots, n$ являются экстремальными точками многочлена S_n и при этом $S_n(s_{j,n}) = (-1)^{j+n}$, то выполнение равенства

$$f'(z) = hF\beta \tag{2.1}$$

при $\omega_i \geqslant 0$, $i = 1, \ldots, n$ и $\sum_{i=1}^n \omega_i = 1$ эквивалентно выполнению условия 3 теоремы Элвинга для нахождения оптимального плана оценки производной в точке z.

Так как равенство $F^{-1}F = I_n$, где I_n — единичная матрица размера n, можно переписать, как систему равенств

$$e_i^\top F^{-1} f(s_{j,n}) = \delta_{ij}\,, \quad i, j = 1, \ldots, n, \tag{2.2}$$

где δ_{ij} — дельта Кронекера, а e_i — i–ый единичный вектор. Поскольку в левой части равенств (2.2) содержатся многочлены без нулевого коэффициента степени не больше n вычисленные в точках $s_{j,n}$, $j = 1, \ldots, n$, а для каждого i существует только одно j, такое, что $\delta_{ij} \neq 0$, то они определяют все базисные многочлены Лагранжа без нулевого члена степени n вычисленные в точках $s_{j,n}$, $j = 1, \ldots, n$, таким образом

$$e_i^\top F^{-1} f(z) = L_i(z)\,, \quad i, j = 1, \ldots, n. \tag{2.3}$$

Если в предыдущем выражении вычислить производную по z и переписать полученное выражение в векторной форме получим

$$f'(z) = F \left(L_1'(z), \ldots, L_n'(z) \right)^\top. \tag{2.4}$$

Приравняв правые части (2.5) и (2.1) и домножив равенство на F^{-1} слева, получаем, что

$$h\beta = \left(L_1'(z), \ldots, L_n'(z)\right)^\top, \tag{2.5}$$

что с учетом введенных ранее обозначений влечет, что $\operatorname{sign}(L_i'(z)) = \operatorname{sign}((-1)^{i+n})$, $i = 1, \ldots, n$ или, вспомнив, что экстремальным многочленом также может быть $-S(x)$, $\operatorname{sign}(L_i'(z)) = \operatorname{sign}((-1)^{i+n+1})$, $i = 1, \ldots, n$.

Таким образом для того, чтобы доказать, что оптимальный план находится в точках $(s_{i,n})_{i=1}^n$, $i = 1, \ldots, n$ с указными ранее весами, осталось доказать равенство знаков $L_i'(z)$ и $\pm S_n(s_{i,n})$. Но так как знаки экстремальных точек многочлена S_n чередуются, достаточно показать при каких z выражения $(-1)^i L_i'(z)$ для имеет одинаковый знак для $i = 1, \ldots, n$.

Обозначим корни многочлена L_i' как $u_{i,1}, \ldots, u_{i,n-1}$, $i = 1, \ldots, n$. Так как для L_i и L_j выполняются требования леммы 2 для любых i и j таких что $i < j$, то последовательно применяя ее для всех базисных многочленов получаем, что

$$u_{n,1} < u_{n-1,1} < \ldots < u_{1,1} < u_{n,2} < u_{n-1,2} < \ldots < u_{1,2} < \ldots < u_{1,n-1}. \tag{2.6}$$

Можно видеть, что, так как все узловые точки больше нуля, знак многочлена $L_i(z)$ при $z \to -\infty$ будет равен $(-1)^{n+i+1}$. В то же время знак $L_i'(z)$ будет противоположным $L_i(z)$ так как меняется четность многочлена и при этом не меняется знак при старшем коэффициенте, то есть $\operatorname{sign}(L_i'(z)) = \operatorname{sign}((-1)^{n+i})$ при $z \to -\infty$. И, следовательно, $\operatorname{sign}((-1)^i L_i'(z)) = \operatorname{sign}((-1)^{n+2i}) = \operatorname{sign}((-1)^n)$ при $z \to -\infty$, то есть $\operatorname{sign}((-1)^i L_i'(z))$ не зависит от i и имеет постоянный знак для любых i, что означает, что при $z \in (-\infty, u_{n,1})$ третье условие теоремы Элвинга выполняется и план является оптимальным.

Осталось изучить как ведут себя знаки $\operatorname{sign}((-1)^i L_i'(z))$ на остальных промежутках. На промежутках $[u_{j,1}, u_{j,n-1}]$ каждый базисный многочлен меняет свой знак ровно 1 раз и на этих промежутках знаки производных не совпадают со знаками экстремального многочлена, а на промежутках $(u_{j,n-1}, u_{j-1,1})$ нет ни одного корня и поэтому $\operatorname{sign}((-1)^i L_i'(z)) = \operatorname{sign}((-1)^{n+j}$ при $z \in (u_{j,n-1}, u_{j-1,1})$, что также подтверждает третье условие теоремы Элвинга и показывает, что показанный план оптимален для $j = 1, \ldots n-1$.

На промежутке $(-\infty, u_{1,n-1}$ каждый базисный многочлен поменял свой знак одинаковое количество раз, а так как при $z \to -\infty$ условие выполнялось, то при $z \in (u_{1,n-1}, +\infty)$ план также является оптимальным.

Глава 3

Численное нахождение оптимальных планов