Лекция 3.

Применение теоремы эквивалентности. С-оптимальные планы

Яковлев Д. М.

д. ф.-м. н.,
профессор Мелас В. Б.

3 октября 2024 г.

Содержание

1 Полиномиальная регрессионная модель

Рассматривается полиномиальная регрессионная модель

$$y = \eta(x, \theta) + \epsilon, \ x \in [-1, 1],$$

где

$$\eta(x, \theta) = \theta^{\mathrm{T}} f(x) - \text{линейная функция регрессии}$$

и

$$f(x) = (f_1(x), \dots, f_m(x))^{\mathrm{T}} = (1, x, \dots, x^{m-1})^T.$$

Случайная величины $\epsilon_i, \ i = 1, \dots, N$ — ошибки наблюдений, для которой предполагаем:

1. $E\epsilon_i = 0$ — несмещённость;

2. $E\epsilon_i\epsilon_j = 0, \ i \neq j$ — некоррелированность;

3. $D\epsilon = E\epsilon_i^2 = \sigma^2 > 0$ — равномощность.

Общей задачей является нахождение оптимальных планов, или планов, доставляющих экстремальное значение некоторой выпуклой (или вогнутой) функции, заданной на множестве информационных матриц. Планы $\xi \in \Xi$, где $\Xi = \cup_{n=1}^{\infty} \Xi_n$, где Ξ_n — множество приближённых планов, заданных в n точках (с ненулевыми коэффициентами).

Тогда для такой модели верна следующая теорема:

Теорема.
Для полиномиальной регрессии на отрезке $[-1, 1]$ D-оптимальный план существует, единственен и сосредоточен с равными весами в m точках, которые являются корнями $(x^2 - 1)P'_{m-1}(x)$, где $P_{m-1}(x)$ — полином Лежандра порядка $m - 1$.

Доказательство. Из непрерывности регрессионных функций $f(x)$ и компактности множества информационных матриц $\mathcal{M}$ получаем, по обобщённой теореме Вейерштрасса (так как критерий — непрерывная функция) существование D-оптимального плана. Поскольку $\mathcal{M}$ — компактно, то можно воспользоваться теоремой Кифера-Вольфовица

Теорема.
(Кифер-Вольфовец)
Пусть множество информационных матриц компактно. Тогда следующие условия эквивалентны

1. *план ξ^* — D-оптимальный;*

2. *план ξ^* — G-оптимальный;*

3. $\max\limits_{x \in \mathcal{X}} d(x, \xi^*) = m,$

где m — число параметров модели. Причём, если план ξ^* сосредоточен в конечном числе точек, то последнее равенство достигается в точках x_i^* оптимального плана ξ^*. Кроме того, информационные матрицы всех оптимальных планов совпадают.

Тогда нахождение D-оптимального плана ξ^* можно заменить эквивалентной задачей — нахождением плана $\xi^* : \max\limits_{x\in\mathcal{X}} d(x,\xi^*) = m$.

$$\max_{x\in\mathcal{X}} d(x,\xi^*) = \max_{x\in\mathcal{X}} f^{\mathrm{T}}(x)\mathbf{D}(\xi^*)f(x) = \max_{x\in\mathcal{X}} \sum_{i,j=1}^{m} d_{ij}(\xi^*)x^{i+j-2} = m,$$

где $\mathbf{D}(\xi^*) = (d_{ij}(\xi^*))_{i,j=1}^m$. $d(x,\xi^*)$ — многочлен степени $2m-2$. Заметим, что из положительной определённости матрицы $\mathbf{D}(\xi^*)$ выходит, что её диагональные элементы положительны. По определению,

$$\forall x \in \mathbb{R}^m \backslash \{0\} \quad (\mathbf{D}(\xi^*)x, x) > 0.$$

Подбирая в качестве x орты $e_i,\ i = 1,\dots,m$ получается, что $d_{ii} > 0,\ i = 1,\dots,m$.
С одной стороны, $d(x,\xi^*)$ имеет не более m точек локального максимума на отрезке $[-1,1]$, включая её концы. Это связано с чередованием знаков производной. Если предположить, что все корни — вещественны, то получится $2m-1$ сегментов, откуда будет либо $m-1$, либо m точек локального максимума.
С другой стороны, по теореме Кифера-Вольфовица локальный максимум для конечного оптимального плана ξ^* достигается во всех его точках, а из свойств информационных матриц план невырожден, если число его точек не меньше m. Отсюда, план ξ^* имеет вид

$$\xi^* = \begin{pmatrix} x_1 & x_2 & \dots & x_{m-1} & x_m \\ \omega_1 & \omega_2 & \dots & \omega_{m-1} & \omega_m \end{pmatrix}$$

где $-1 = x_1 < x_2 < \dots < x_{m-1} < x_m = 1$. [1]
Тогда перепишем информационную матрицу плана ξ^* в матричный вид:

$$\mathbf{M}(\xi^*) = \sum_{i=1}^{m} f(x_i)f^{\mathrm{T}}(x_i)\omega_i$$
$$= \underbrace{\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \\ \vdots & \ddots & \vdots \\ x_1^{m-1} & \dots & x_m^{m-1} \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_m \end{pmatrix}}_{\mathbf{W}} \underbrace{\begin{pmatrix} 1 & x_1 & \dots & x_1^{m-1} \\ 1 & x_2 & \dots & x_2^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & \dots & x_m^{m-1} \end{pmatrix}}_{\mathbf{F}^{\mathrm{T}}}$$

Посчитаем определитель

$$\det(\mathbf{M}(\xi^*)) = \det(\mathbf{F}\mathbf{W}\mathbf{F}^{\mathrm{T}}) = \det(\mathbf{F}^{\mathrm{T}})^2 \prod_{i=1}^{n} \omega_i, \tag{1}$$

где $\det(\mathbf{F}^{\mathrm{T}})$ — определитель Вандермонда. Обозначим $\triangle_{m-1} = \det(\mathbf{F}^{\mathrm{T}})$ и получим

$$\triangle_{m-1} = \prod_{1\leqslant j<i\leqslant m} (x_i - x_j).$$

[1] Кажется, что мы ищем полином наилучшего равномерного приближения (ПНРП). Тогда задачу с поиском оптимального плана можно свести к поиску минимума разности норм ПНРП степени $2m-2$ и ещё какой-то функции.

Чтобы теперь найти D-оптимальный план ξ^*:

$$\xi^* = \arg\sup_{\xi\in\Xi} \det(\mathbf{M}(\xi)),$$

необходимо максимизировать определитель Вандермонда и $\prod_{i=1}^m \omega_i$, для которого максимум достигается при $\omega_1 = \dots = \omega_m = \frac{1}{m}$.
Максимизируем $\triangle_{m-1}$, найдя точку экстремума. Так как $x_1 = -1, x_2 = 1$ уже фиксированы, то получаем для $x_i,\ i = 2, \dots, m-1$ систему уравнений:

$$\frac{\partial\triangle_{m-1}}{\partial x_i} = 0 \Leftrightarrow \frac{1}{x_i - x_1} + \dots + \frac{1}{x_i - x_{i-1}} + \frac{1}{x_i - x_{i+1}} + \dots + \frac{1}{x_i - x_m} = 0.$$

Для проверки достаточно посчитать производную и разделить на $\triangle_{m-1}$. Если взять функцию $\varphi(x) = (x - x_2)\dots(x - x_{m-1})$, то систему можно переписать как

$$\begin{aligned}&\frac{1}{x_i+1} + \frac{1}{x_i-1} + \frac{\varphi''(x_i)}{2\varphi'(x_i)} = 0,\ i = 2, \dots, m-1 \Leftrightarrow\\ &(x_i^2 - 1)\varphi''(x_i) + 4x_i\varphi'(x_i) = 0,\ i = 2, \dots, m-1.\end{aligned}$$

Из равенства их корней и степеней с многочленом $\varphi(x)$ следует, что эти многочлены подобные, тогда

$$(x^2 - 1)\varphi''(x) + 4x\varphi'(x) = const\varphi(x)$$

Приравняв коэффициенты при x^{m-2} находим $const = (m-2)(m+1) = m(m-1) - 2$ и получаем ОДУ второго порядка

$$(x^2 - 1)\varphi''(x) + 4x\varphi'(x) - (m(m-1) - 2)\varphi(x) = 0.$$

$\varphi(x)$ можно связать с полиномом Лагранжа $P_{m-1}(x)$, выписав его характеристическое дифференциальное уравнение и взяв производную

$$\begin{aligned}&((x^2 - 1)P''_{m-1}(x) + 2xP'_{m-1}(x) - m(m-1)P_{m-1}(x))' = 0 \Leftrightarrow\\ &(x^2 - 1)P'''_{m-1}(x) + 4xP''_{m-1}(x) - (m(m-1) - 2)P'_{m-1}(x) = 0\end{aligned}$$

Откуда $\varphi(x) = P'_{m-1}(x)$.[2]. □

2 Тригонометрическая регрессионная модель

Рассматривается тригонометрическая регрессионная модель

$$y = \eta(x, \theta) + \epsilon,\ x \in [-\pi, \pi],$$

где

$$\eta(x, \theta) = \theta^{\mathrm{T}} f(x) - \text{линейная функция регрессии}$$

и

$$\begin{aligned}f(x) &= (f_1(x), \dots, f_{2m+1}(x))^{\mathrm{T}}\\ &= (1, \sin(x), \cos(x), \dots, \sin(x(m-1)), \cos(x(m-1)), \sin(xm), \cos(xm))^T.\end{aligned}$$

Ошибки наблюдений ϵ_i предполагаются такими же, как и в случае полиномиальной регрессионной модели (стандартное предположение). Сформулируем теорему для D-оптимального плана.

[2]Если успею, то напишу подробнее про многочлен Лагранжа

Теорема.
Пусть $\mathcal{X} = [-\pi, \pi]$. Непрерывным D-оптимальным планом для тригонометрической регрессионной модели является любой план

$$\xi_N^* = \begin{pmatrix} x_1^* & \dots & x_N^* \\ 1/N & \dots & 1/N \end{pmatrix},$$

$$x_i^* = \frac{i-1}{N} 2\pi - \pi, \; i = 1, \dots, N, \; N \geqslant 2m+1, \; m - \textit{порядок регрессионной модели.}$$

Также D-оптимальным является равномерный план

$$\xi^* \sim U(-\pi, \pi).$$

Доказательство. В идеях воспользоваться 3-им условием эквивалентности из теоремы Кифера-Вольфовица.
Знаем, что в $L_2[-\pi, \pi]$ базис можно образовать ортогональной тригонометрической системой. Тогда

$$\|1\|^2 = 1, \|\sin(kx)\|^2 = \|\cos(kx)\|^2 = \frac{1}{2}, \; k = 1, 2, \dots, N, N+1, \dots.$$

По этой причине, для равномерного плана информационная матрица диагонализируется и получится

$$\mathbf{M}(\xi^*) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) f^{\mathrm{T}} \xi(dx) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \frac{1}{2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{2} \end{pmatrix}$$

Понятно, какая матрица $\mathbf{D}(\xi^*)$ получится. Пользуясь теоремой Кифера-Вольфовица, покажем, что $d(x, \xi^*) = 1 + 2m$. Тогда

$$d(x, \xi^*) = f^{\mathrm{T}}(x) \mathbf{D}(\xi^*) f(x) = 1 + 2(\sin^2 x + \cos^2 x + \dots + \sin^2 mx + \cos^2 mx) = 1 + 2m.$$

Таким образом, показали результат для равномерного плана ξ^*.
Для оптимального конечного плана ξ_N^* из условий теоремы:

- На диагонали получим $diag\,\mathbf{M}(\xi_N^*) = (1, \frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$;
- На не-диагонали получим, например, для $a_{2i,2j+1}, \; i \neq j$:

$$a_{2i,2j+1} = \frac{1}{N} \sum_{k=1}^{N} \sin(i x_k^*) \cos(j x_k^*)$$

Получаются суммы из произведений косинусов или синусов. Их можно разложить на суммы или разности косинусов или синусов, и показать, Что их суммы равны нулю. Тогда достигается $d(x, \xi_N^*)$ и ξ_N^* — D-оптимальный план. □

3 С-критерий оптимальности

Итак, рассматривали до этого задачи, построенные на нахождении D-оптимальных планов, пользовались определителем. Теперь мы хотим оценить величину типа $c^{\mathrm{T}}\theta$, где c — заданный вектор в стандартной линейной по параметрам регрессионной модели. Заметим, что первый символ в названии критерия определяет, чем или относительно чего будет искаться оптимальный план. Вектор c может быть произвольным, например, быть ортой, производной от регрессионной функции $f'(x)$ и так далее.

Определение.
План ξ допустим в задаче оценивания величины $c^{\mathrm{T}}\theta$, если величина оцениваема по результатам эксперимента.

Определение.
c-оптимальный план ξ^* — допустимый план, который минимизирует дисперсию оценки $c^{\mathrm{T}}\theta$.
Таким образом, рассматривается не минимизация дисперсии ξ, а оценки $c^{\mathrm{T}}\theta$, что уместно для линейных регрессионных моделей (про нелинейные, наверно, всё не так однозначно).
Когда C-критерий оптимальности может быть полезен? Например, когда рассматривается задача экстраполяции. Пусть имеется линейная регрессионная модель

$$y = \eta(x, \theta) + \epsilon, \ x \in \mathcal{X},$$

где $\eta(x, \theta) = f(x)\theta^{\mathrm{T}}$. Для $z \notin \mathcal{X}$ хотим подобрать такой план ξ^*, чтобы минимизировать оценку $f(z)\theta^{\mathrm{T}}$, где $c = f(z)$.

4 Теорема Гаусса-Маркова в случае матриц неполного ранга

Сформулируем теорему Гаусса-Маркова в случае матриц неполного ранга. Для чего она нужна? До этого была теорема Гаусса-Маркова для невырожденных матриц. Эта теорема рассматривает условия оцениваемости для вырожденных матриц.

Определение.
Векторная параметрическая функция $\tau = \mathbf{T}\theta, \ (\mathbf{T} \in \mathbb{R}^{k\times m}, \ k \in \{1, \dots, m\})$ — оцениваема, если для неё $\exists$ несмещённая оценка вида $\hat{\tau} = \mathbf{A}Y, \ (\mathbf{A} \in \mathbb{R}^{k\times N})$.

Теорема.
Для линейной (классической) регрессионной модели

- *Параметрическая функция* $\tau = \mathbf{T}\theta, \ (\mathbf{T} \in \mathbb{R}^{k\times m}, \ k \in \{1, \dots, m\})$ *оцениваема тогда и только тогда, когда*
$$\mathbf{T}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{T};$$

- *если выполнено это условие, то оценка для τ имеет вид*

$$\hat{\tau} = \mathbf{T}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X}^{\mathrm{T}}Y,$$

определена однозначно и является наилучшей линейной несмещённой оценкой. Ковариационная матрица оценки $\hat{\tau}$ имеет вид

$$\mathbf{D}_{\hat{\tau}} = \sigma^2\mathbf{D},\ \mathbf{D} = \mathbf{T}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{T}^{\mathrm{T}}.$$

Доказательство — обобщение теоремы Гаусса-Маркова для невырожденных матриц и несколько лемм об оцениваемых параметрических функциях.

5 Теорема эквивалентности для C-критерия

Если теорема Гаусса-Марков для вырожденных матриц не используется здесь, то можно переместить тему ближе к c-критерию оптимальности.
Сначала будем рассматривать невырожденный c-оптимальный план, а затем вырожденный план.
Случай невырожденного c-оптимального плана. Введём функцию

$$\varphi(x, \xi) = (f(x)^{\mathrm{T}}\mathbf{M}^{-1}(\xi)c)^2.$$

Почему же используется именно такая функция? Прежде всего, для нашего критерия хотим учитывать c-вектор, затем хотим среди доступных планов найти c-оптимальный, а это подразумевает минимизацию дисперсионной матрицы для оценки $c^{\mathrm{T}}\theta$.

Теорема.
Если существует невырожденный c-оптимальный план, то невырожденный план ξ^* является c-оптимальным

$$\Leftrightarrow \max_{x\in\mathcal{X}} \varphi(x, \xi^*) = c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c.$$

Кроме того, в опорных точках оптимального плана функция $\varphi(x, \xi^*)$ достигает своего максимального значения.

Вывод в виде достижения максимального значения в точках оптимального плана напоминает теорему Кифера-Вольфовица. Заметим, что схема доказательства теоремы эквивалентности для C-критерия аналогична схеме доказательства теоремы Кифера-Вольфовица. Это связано с тем, что сам C-критерий отличается от D-критерия наличием c векторов при минимизации $\mathbf{D}(\xi)$.

Лемма 1. *Для произвольного $k \in \mathbb{N}$ и фиксированного вектора $c \in \mathbb{R}^k$ функция $\psi(\mathbf{A}) = c^{\mathrm{T}}\mathbf{A}c$ — выпуклая на множестве положительно определённых матриц $k \times k$.*

Доказательство. Проверяем, что

$$\psi((1-\alpha)\mathbf{A} + \alpha\mathbf{B}) \leqslant (1-\alpha)\psi(\mathbf{A}) + \alpha\psi(\mathbf{B}).$$

□

Лемма 2. *Для произвольного фиксированного вектора* $c \in \mathbb{R}^k$ *и обратимой матрицы* $\mathbf{A}$, *элементы которой являются дифференцируемыми функциями параметра* α *имеет место формула*

$$\frac{\partial}{\partial \alpha} c^{\mathrm{T}} \mathbf{A}(\alpha)^{-1} c = c^{\mathrm{T}} \mathbf{A}(\alpha)^{-1} \mathbf{A}'(\alpha) \mathbf{A}(\alpha)^{-1} c.$$

Доказательство. Рассмотрим $\mathbf{I}_k = \mathbf{A}(\alpha)^{-1}\mathbf{A}(\alpha)$. Тогда

$$\begin{aligned}&\mathbf{0} = (\mathbf{I}_k)'_\alpha = (\mathbf{A}(\alpha)^{-1}\mathbf{A}(\alpha))' = \mathbf{A}^{-1}(\alpha)'\mathbf{A}(\alpha) + \mathbf{A}^{-1}(\alpha)\mathbf{A}'(\alpha) \Leftrightarrow \\ &(\mathbf{A}^{-1}(\alpha))' = -\mathbf{A}^{-1}(\alpha)\mathbf{A}'(\alpha)\mathbf{A}^{-1}(\alpha)\end{aligned}$$

Домножая на c, c^{T} где необходимо, получаем нашу формулу. □

Лемма 3. *Для любого невырожденного плана* ξ *справедливо неравенство*

$$\max_{x \in \mathcal{X}} \varphi(x, \xi) \geqslant c^{\mathrm{T}} \mathbf{M}^{-1}(\xi) c.$$

Доказательство.

$$\begin{aligned}&\max_{x \in \mathcal{X}} \varphi(x, \xi) \geqslant \int_{\mathcal{X}} \varphi(x, \xi)\xi(dx) = \int_{\mathcal{X}} (f^{\mathrm{T}}(x)\mathbf{M}^{-1}(\xi)c)^2 \xi(dx) \\ &= \int_{\mathcal{X}} \mathrm{tr} f^{\mathrm{T}}(x)\mathbf{M}^{-1}(\xi) c f^{\mathrm{T}}(x) \mathbf{M}^{-1}(\xi) c \xi(dx) = \int_{\mathcal{X}} \mathrm{tr} f^{\mathrm{T}}(x) \mathbf{M}^{-1}(\xi) f(x) \xi(dx)\ c^{\mathrm{T}} \mathbf{M}^{-1}(\xi) c \\ &= c^{\mathrm{T}} \mathbf{M}^{-1}(\xi) c,\end{aligned}$$

так как для следа матрицы выполняется:

$$\mathrm{tr}\mathbf{AB} = \mathrm{tr}\mathbf{BA},\ \mathrm{tr}\mathbf{A} = \mathrm{tr}\mathbf{A}^{\mathrm{T}}.$$

□

Теперь докажем саму теорему.

Доказательство. **Необходимость.** Пусть ξ^* — c-оптимальный план. Рассмотрим план $\xi_\alpha = (1-\alpha)\xi^* + \alpha\xi_x$, где $\xi_x = \binom{x}{1}$ — план, сосредоточенный в точке x, $0 < \alpha < 1$. В силу C-оптимальности плана ξ^* имеет место неравенство:

$$\begin{aligned}c^{\mathrm{T}} \mathbf{M}^{-1}(\xi_\alpha) c \geqslant c^{\mathrm{T}} \mathbf{M}^{-1}(\xi^*) c \Leftrightarrow \\ c^{\mathrm{T}} \frac{\mathbf{M}^{-1}(\xi_\alpha) - \mathbf{M}^{-1}(\xi^*)}{\alpha} c \geqslant 0\end{aligned}$$

В пределе $\alpha \to 0_+$:

$$\begin{aligned}&\frac{\partial}{\partial \alpha} c^{\mathrm{T}} \mathbf{M}^{-1}(\xi_\alpha) c|_{\alpha=0_+} \\ &= -c^{\mathrm{T}} \mathbf{M}^{-1}(\xi^*) \frac{\partial}{\partial \alpha} \left(\mathbf{M}(\xi_\alpha)\right)|_{\alpha=0_+} \mathbf{M}^{-1}(\xi^*) c \geqslant 0.\end{aligned}$$

Учитывая, что $\xi_\alpha = (1-\alpha)\xi^* + \alpha\xi_x$

$$\frac{\partial}{\partial \alpha} \left(\mathbf{M}(\xi_\alpha)\right)|_{\alpha=0_+} = -\mathbf{M}(\xi^*) + \mathbf{M}(\xi_x).$$

Подставляя это в неравенство, получим

$$
\begin{aligned}
&-c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)\left(-\mathbf{M}(\xi^*)+\mathbf{M}(\xi_x)\right)\mathbf{M}^{-1}(\xi^*)c \geqslant 0 \Leftrightarrow \\
&c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c \geqslant c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)\mathbf{M}(\xi_x)\mathbf{M}^{-1}(\xi^*)c = c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)f(x)f^{\mathrm{T}}(x)\mathbf{M}^{-1}(\xi^*)c \\
&= (f^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c)^2 = \varphi(x,\xi^*)
\end{aligned}
$$

Отсюда, $\max\varphi(x,\xi^*) \leqslant c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c$. Используя лемму 3 (отметить), приходим к равенству.

Достаточность. От противного. Пусть существует план

$$\xi^* :\ \max_x(x,\xi^*) = c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c,$$

который не является c-оптимальным планом. Пусть ξ_c — c-оптимальный план. Для плана $\xi_\alpha = (1-\alpha)\xi^* + \alpha\xi_c$ в силу леммы 1 (о вогнутости на множестве положительно определённых матриц) имеем

$$c^{\mathrm{T}}\mathbf{M}^{-1}(\xi_\alpha)c \leqslant (1-\alpha)c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c + \alpha c^{\mathrm{T}}\mathbf{M}^{-1}(\xi_c)c.$$

Вычитая $c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c$, получаем

$$c^{\mathrm{T}}\mathbf{M}^{-1}(\xi_\alpha)c - c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c \leqslant \alpha c^{\mathrm{T}}(\mathbf{M}^{-1}(\xi_c) - \mathbf{M}^{-1}(\xi^*))c < 0,$$

так как план ξ^* не c-оптимальный. Из полученного неравенства следует, что

$$c^{\mathrm{T}}\frac{\mathbf{M}^{-1}(\xi_\alpha) - \mathbf{M}^{-1}(\xi^*)}{\alpha}c < 0,$$

и значит,

$$\frac{\partial}{\partial\alpha}c^{\mathrm{T}}\mathbf{M}^{-1}(\xi_\alpha)c|_{\alpha=0_+} = -c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)\left(-\mathbf{M}(\xi^*)+\mathbf{M}(\xi_c)\right)\mathbf{M}^{-1}(\xi^*)c < 0,$$

откуда

$$c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c < c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)\mathbf{M}(\xi_c)\mathbf{M}^{-1}(\xi^*)c.$$

С другой стороны, план $\xi^* :\ \max_x\varphi(x,\xi^*) = c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c$. Это означает, что

$$c^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)\mathbf{M}(\xi_c)\mathbf{M}^{-1}(\xi^*)c = \int(f^{\mathrm{T}}\mathbf{M}^{-1}(\xi^*)c)^2\xi_c(dx) = \int\varphi(x,\xi^*)\xi_c(dx) \leqslant \max_x\varphi(x,\xi^*)$$

Получили противоречие. Следовательно, план ξ^* — c-оптимальный план. □

6 Определения

Определение.
$\mathcal{X}$ — некоторое заданное множество планов, предполагается компактным.

Определение.
Точный план эксперимента ξ_N — это дискретная вероятностная мера

$$\xi_N \sim \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ 1/N & 1/N & \dots & 1/N \end{pmatrix},\ x_i \in \mathcal{X},\ i = 1,\dots,N.$$

Определение.
Приближённый план эксперимента ξ — это вероятностная мера

$$\xi \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \omega_1 & \omega_2 & \dots & \omega_n \end{pmatrix},\ x_i \in \mathcal{X},\ i = 1, \dots, n;\ \omega_i > 0,\ \sum_{i=1}^{n} \omega_i = 1.$$

Определение.
Информационная матрица $\mathbf{M}(\xi)$— это

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} f(x) f(x)^{\mathrm{T}} \xi(dx),$$

где $f(x) = (f_0(x), \dots, f_m(x))^{\mathrm{T}}$, где $f_i(x),\ i = 0, \dots, m$ — заданные функции.

Определение.
Ξ_n — множество приближённых планов, заданных в n точках (с ненулевыми коэффициентами). $\Xi = \cup_{n=1}^{\infty} \Xi_n$

Определение.

$$\mathcal{M} = \{\mathbf{M} : \mathbf{M} = \mathbf{M}(\xi), \xi \in \Xi\}$$

это множество всех информационных матриц. Если рассматривается линейная регрессионная модель, то $\mathcal{M}$ — компактное множество в пр–ве $\mathbb{R}^{m(m+1)/2}$, где m — размерность вектора $f(x)$.

Определение.
Если для некоторого плана ξ информационная матрица $\mathbf{M}(\xi)$ невырожденная, то ξ — невырожденный план.

Определение.
Дисперсионная матрица невырожденного плана ξ — это матрица

$$\mathbf{D}(\xi) = \mathbf{M}(\xi)^{-1}.$$

Теорема Каратеодори обосновывает применимость приближённых планов эксперимента.
Не очень понятно, для чего нужно высказывание: “Как правило, не существует такого невырожденного плана ξ^*, что

$$\mathbf{D}(\xi^*) \leqslant \mathbf{D}(\xi)$$

для любого невырожденного плана $\xi \in \Xi$.”
Тогда придумаем пример. Пусть $f(x) = x, \mathcal{X} = \mathbb{R}$, тогда

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} x^2 \xi(dx) = \sum_{i=1}^{n} w_i x_i^2.$$

Понятно, что

$$\mathbf{D}(\xi) = \mathbf{M}(\xi)^{-1} = \left(\sum_{i=1}^{n} w_i x_i^2 \right)^{-1}.$$

Допустим,

$$\xi = \begin{pmatrix} x_1 \\ 1 \end{pmatrix}.$$

Если $x_1 \to \infty$, то $\mathbf{D}(\xi) \to 0$, из чего следует, что $\arg\inf\limits_{\mathcal{X}} \mathbf{D}(\xi) \notin \mathcal{X}$.
Следующий вопросы: для чего нужны критерии оптимальности и что подразумевается под "вогнутыми функциями на множестве информационных матриц"?

- Для чего нужны критерии оптимальности? Поскольку не существует универсального способа определения наилучшего невырожденного плана, предлагается ввести критерии оптимальности по которым можно найти оптимальный план — решение критерия. Каждый критерий ставит свою задачу, для которой мы и находим оптимальный план.

- Вогнутые функции на множестве информационных матриц/Выпуклые функции на множестве дисперсионных матриц — на самом деле критерии можно интерпретировать как класс функций с определёнными свойствами, которые сами по себе интерпретируют информационную матрицу, сопоставляя ей какое-либо скалярное значением.
  Тем самым, критерий — скалярная функция от матрицы, для которой оптимальный план — план $\xi \in \Xi$, на котором достигается $\inf,\ \sup$.

Далее возникает вопрос, считать ли inf критерия от $\mathbf{M}^{-}(\xi)$ ($= \mathbf{D}(\xi)$, если $\mathbf{M}^{-}(\xi)$ — обратима) или от $\mathbf{M}(\xi)$? Наверно, хотим найти $\arg\inf\limits_{\xi} \mathbf{D}(\xi)$. В статистике оценка $\hat{\theta}$ называется эффективной (для несмещённых оценок), если на ней достигается $\inf \mathbf{D}\hat{\theta}$. Приведём несколько примеров критериев:

Определение.
D-критерий (D — determinant).

$$\det \mathbf{M}(\xi) \to \sup_{\xi \in \Xi}$$

или

$$\det \mathbf{D}(\xi) \to \inf_{\xi \in \Xi}.$$

Определение.
G-критерий. (G — general variance).

$$\max_{x \in \mathcal{X}} d(x, \xi) \to \inf_{\xi \in \Xi_H},$$

где $d(x,\xi) = f^{\mathrm{T}}(x)\mathbf{D}(\xi)f(x)$. Эти критерии уместны для нахождения оптимального плана в классе линейных несмещённых оценок.