Конспект по курсу В. Меласа «Дополнительные главы оптимального планирования эксперимента»

Собрано 1 декабря 2015 г. в 11:00

Содержание

1. Асимптотические свойства нелинейного метода наименьших квадратов

Изложение материала данного вопроса имеется в разделе 1.2 Учебного Пособия: «Локально Оптимлаьные Планы Эксперимента». Для данного вопроса необходимо понимать, как устроена нелинейная регрессионная модель (вопрос 2).

Устройство нелинейной модели и основные понятия. Заданы $N \in \mathbb{N}$ (объем выборки), $m \in \mathbb{N}$, $\Theta \in \mathbb{R}^m$ (неизвестный многомерный параметр), $\mathcal{X}$ — некоторое множество[1]. Пусть происходит «эксперимент», в котором наблюдаются (одномерные) «результаты эксперимента» $y_1, y_2, \dots, y_N \in \mathbb{R}^1$. Рассмотрим отображение $\eta : \mathcal{X} \times \mathbb{R}^m \mapsto \mathbb{R}^1$. Аналитическое задание отображения η как функции двух аргументов нам известно.

Модель эксперимента задается следующим образом: для всех $j \in 1 : N$

$$y_j = \eta(x_j, \Theta) + \varepsilon_j, \tag{1}$$

где $x_1, x_2, \dots, x_N \in \mathcal{X}$ — «условия эксперимента», $\Theta = (\theta_1, \theta_2, \dots, \theta_m)^{\mathrm{T}} \in \mathbb{R}^m$, а $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ — некоррелированные, центрированные, гомоскедастичные случайные величины, т.е. $\mathbb{E}\varepsilon_j = 0$ и $\mathbb{D}\varepsilon_j = \sigma^2$ для всех $j \in 1 : N$.[2]

Задача: оценить параметр Θ. Ясно, что задача является регрессионной, причем функция η является регрессией.

Нужно формально объяснить, что значит «нелинейная модель», то есть чем эта модель отличается от «линейной». Будем говорить, что параметр θ_j, где $j \in 1 : m$, входит нелинейно в модель (1), если при фиксированном x

$$\frac{\partial \eta(x, \cdot)}{\partial \theta_j}(\theta_j)$$

существует и не является постоянной. Если же указанная функция является постоянной, то говорим, что параметр θ_j входит в модель линейно. Если есть хотя бы один параметр θ_j, который входит в модель нелинейно, то модель (1) называют нелинейной. Регрессию η в таком случае тоже называют нелинейной (по параметрам).

Для того, чтобы определить неизвестный многомерный параметр Θ, нужно выбрать эксперементальные условия $x_1, x_2, \dots, x_N$ и метод оценивания параметров. Определимся сначала с первым вопросом.

Определение. *Любой набор из (не обязательно различных) N элементов множества $\mathcal{X}$ будем называть точным планом эксперимента.*

Определение. *Пусть n — фиксированное натуральное число. Приближенным планом эксперимента называют дискретную вероятностную[3] меру, задаваемую таблицей*

$$\xi = \{x_1, x_2, \dots, x_n; \mu_1, \mu_2, \dots, \mu_n\}, \tag{2}$$

где x_j различные, $\mu_i > 0$ для всех i, а $\sum_{i=1}^n \mu_i = 1$[4].

Заметим, что все условия, наложенные на меру являются простыми (необременняющими) и естественными.

[1] В самом общем описании, никакие условия на это множество не накладываются.

[2] Естественно ε_j играют роль ошибок измерения, шума.

[3] *то есть нормированную на единицу.*

[4] *Подразумевается, что $\xi(x_i) = \mu_i$ для всех i*

Выбор «наилучшего» плана является отдельной задачей. Пусть план фиксирован, тогда в качестве метода оценивания параметров рассмотрим (нелинейный) метод наименьших квадратов. Будем обозначать $\hat{\Theta}$ — решение экстремальной задачи МНК:

$$\sum_{j=1}^{N}(\eta(x_j,\Theta)-y_j)^2 \to \min_{\Theta\in\mathbb{R}^m}.$$

Оценки $\hat{\Theta}$ обладают хорошими асимптотическими свойствами.

Асимптотические свойства МНК-оценок. В данном разделе мы начинаем вводить ограничения на множества Ω и $\mathcal{X}$. Пусть Ω — ограниченное замкнутое множество в $\mathbb{R}^m$, $\mathcal{X}$ — ограниченное замкнутое множество в $\mathbb{R}^k$, где $k\in\mathbb{N}$.

Пусть функция регрессии $\eta_(x,\Theta)$ нелинейна по параметрм и определена при всех $x\in\mathcal{X}$, $\Theta\in\Omega$. Через Θ_u будем обозначать истинное значение вектора параметров, т.е. такое значение Θ, при котором верна модель (1).

Под планом в дальнейшем всегда подразумеваем приближенный. Для дискретных мер $\xi=\{x_1,x_2,\dots,x_n;\mu_1,\mu_2,\dots,\mu_n\}$ используем стандартную запись (интеграл по мере, 2 курс):

$$\int_{\mathcal{X}} g(x)\,d\xi(x)=\sum_{i=1}^{n} g(x_i)\mu_i,$$

где g — произвольная функция, определенная на $\mathcal{X}$[5].

Введем предположения:

1. регрессия $\eta(x,\Theta)$ непрерывна на множестве $\mathcal{X}\times\Omega$;
2. имеется слабая сходимость распределений $\mathcal{L}(\xi_N)\Rightarrow\mathcal{L}(\xi)$, где ξ — некоторый план, то есть для любой функции $g\in \mathrm{C}(\mathcal{X})$ имеет место сходимость
$$\int_{\mathcal{X}} g(x)\,d\xi_N(x)=\int_{\mathcal{X}} g(x)\,d\xi(x);$$
3. для $\Theta,\overline{\Theta}\in\Omega$ величина
$$\int_{\mathcal{X}}\left(\eta(x,\Theta)-\eta(x,\overline{\Theta})\right)^2 d\xi(x)$$
равна нулю только при $\Theta=\overline{\Theta}$[6];
4. Частные производные первого и второго порядка регрессии η по параметру существуют и непрерывны на $\mathcal{X}\times\Omega$, то есть $\eta\in \mathrm{C}^2_\Theta(\mathcal{X}\times\Omega)$.
5. Истинное значение параметра Θ_u является внутренней точкой Ω[7].
6. Матрица[8]
$$M(\xi,\Theta)=\int_{\mathcal{X}} f(x,\Theta)f^{\mathrm{T}}(x,\Theta)\,d\xi(x), \tag{3}$$

[5]На самом деле тут должна быть измеримость по мере, почему мы ее не требуем?
[6]тогда и только тогда, правда?
[7]то есть не принадлежит $\mathrm{frac}(\Omega)$. Это существенно, так как множество Ω является замкнутым.
[8]Убедитесь, что понимаете, что это, действительно, матрица.

где

$$f(x,\Theta) = \left(\frac{\partial\eta(x,\Theta)}{\partial\theta_1}, \frac{\partial\eta(x,\Theta)}{\partial\theta_2}, \ldots, \frac{\partial\eta(x,\Theta)}{\partial\theta_m}\right)^{\mathrm{T}}$$

не вырождена при $\Theta = \Theta_u$.

Теперь пусть

$$\xi_N = \{x_1, x_2, \ldots, x_N; 1/N, 1/N, \ldots 1/N\},$$

где x_i — необязательно различные точки,

$$\hat{\Theta}_N = \underset{\Theta\in\Omega}{\arg\min} \sum_{i=1}^{N} (\eta(x_i,\Theta) - y_i)^2. \tag{4}$$

Теорема (без доказательства). *Если случайные ошибки $\{\varepsilon_i\}_{i=1}^N$ некоррелированы, одинаково распределены и являются центрированными и гомоскедастичными, результаты экспериментов описываются уравнением* (1) *и выполнены предположения 1–3, то последовательность МНК-оценок сильно состоятельна, т. е. при $N \to \infty$*

$$\hat{\Theta}_N \to \Theta_u$$

с вероятностью 1, где $\hat{\Theta}_N$ определено формулой (4).

Если дополнительно выполняются предположения 4–6, то последовательность случайных векторов $\sqrt{N}(\hat{\Theta}_N - \Theta_u)$ имеет асимптотически нормальное распределение с нулевым вектором средних и ковариационной матрицей $\sigma^2 M^{-1}(\xi, \Theta_u)$. [9]

Матрицу $M(\xi, \Theta_u)$ называют информационной матрицей для нелинейных по параметрам регрессионных моделей.

2. Постановка задачи оптимального планирования для нелинейных моделей. Теорема эквивалентности для локально оптимальных планов.

Пусть $N \in \mathbb{N}$, $y_1, \ldots, y_N \in \mathbb{R}$, $x_1, \ldots, x_N \in \mathbb{X}$, где $\mathbb{X}$ некоторое множество, обычно $\mathbb{R}^k$, а $y_1, \ldots, y_N, x_1, \ldots, x_N$ — наши «наблюдения», которые мы будем называть результатами эксперимента.

Введем множество параметров Θ и предположим, что наблюдения описываются следующей моделью:

$$y_i = \eta(x_j, \theta) + \varepsilon_j, \tag{5}$$

где $\theta \in \Theta$ — параметр, значения которого мы и будем пытаться в дальнейшем оценить, а ε_j — случайный шум, про который мы предположим, что

$$E\varepsilon = 0,\ E\varepsilon^2 = \sigma^2$$

Будем предполагать, что $\Theta \subset \mathbb{R}^m$.

---

[9] *Вспомните, откуда тут σ.*

Определение. *Будем говорить, что параметр* θ_j *входит в* (5) *нелинейным образом, если для фиксированного* $x \in \mathbb{X}$ *существует и не является постоянной функция*

$$\phi_{j,x}(\theta) = \frac{\partial \eta(x,\theta)}{\theta_j}$$

Если $\phi_{j,x}(\theta) = const$, *то* θ_j *входит в модель линейным образом.*

Определение. *Под точным планом эксперимента будем понимать* N *точек* $x_1, \ldots, x_N \in \mathbb{X}$

Определение. *Под приближенным планом эксперимента будем понимать* $n \in \mathbb{N}$ *пар* (x_i, μ_i), *где*

$$x_i \in \mathbb{X}, x_i \neq x_j, i \neq j,$$

$$\mu_i > 0, \ \sum_{i=1}^{n} \mu_i = 1,$$

Пусть N — доступное число «ресурсов» (кол-во экспериментов, которое можно провести). Тогда при использование приближенного плана рекомендуется в точке x_j провести $\mu_j N$ экспериментов. В итоге получится точный план, как работать с которым уже ясно.

Определение. *При фиксированном плане для оценки* θ *будем использовать метод наименьших квадратов:*

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{j=1}^{N} \left(\eta(x_j, \theta) - y_j)^2\right)$$

Наша задача — выбрать некоторым образом точки $x_1, \ldots, x_N$, чтобы МНК-оценка была в некотором смысле оптимальной.

Введем еще несколько обозначений:

Определение. *Пусть* ξ *— дискретная вероятностная мера с носителем* $x_1, \ldots, x_n$. *Тогда*

$$\int_{\mathbb{X}} g(x) d\xi(x) = \sum_{i=1}^{n} g(x_i) \xi_i$$

Определение. *Пусть* $f(x,\theta)^{\mathrm{T}} = \left(\frac{\partial \eta(x,\theta)}{\partial \theta_1}, \ldots, \frac{\partial \eta(x,\theta)}{\partial \theta_1}\right)$. *Пусть* θ^u *— истинной значение оцениваемого параметра. Тогда информационной матрицей будем называть*

$$M(\xi, \theta_u) = \int_{\mathbb{X}} f(x,\theta) f(x,\theta)^{\mathrm{T}} d\xi(x)$$

Заметим, что $M(\xi, \theta_u)$ в случае, когда все параметры входят линейно, не зависит от θ_u и т.к. обратная к информационной матрице — «нижняя оценка» на дисперсию оцениваемого параметра (в многомерном случае под дисперсией понимается ковариационная матрица), то можно естественным образом ввести различные понятия оптимальности, опираясь на собственные числа информационной матрицы. Например, D-критерий предлагает выбирать планы, максимизирующие определитель информационной матрицы.

В нелинейном случае все сложнее. Информационная матрица зависит от «истинного» значения параметра, которое неизвестно. Предположим, что у нас есть некоторое приближение θ^0 «истинного» параметра. Тогда будем называть план, максимизирующий определитель матрицы $M(\xi, \theta^0)$ локально D-оптимальным.

Разложим η в ряд Тейлора в окресности $\theta^0 \in \Theta \subset \mathbb{R}^m$:

$$\eta(x,\theta) = \eta(x,\theta^0) + (\theta-\theta^0)^{\mathrm{T}} f(x,\theta^0) + r(x,\theta)$$

Введем следующие обозначения:

$$f(x)^{\mathrm{T}} = f(x,\theta^0)^{\mathrm{T}} = \left(\frac{\partial(x,\theta^0)}{\partial(\theta_i}, \dots, \frac{\partial x, \theta_0}{\theta_m}\right)$$

$$M(\xi) = M(\xi,\theta^0) = \int_{\mathbb{X}} f(x) f(x) d\xi(x)^{\mathrm{T}}$$

$$d(x,\xi) = f(x)^{\mathrm{T}} M^{-1}(\xi) f(x)$$

Для данных обозначение будет верна следующая теорема:

Теорема (Эквивалентности). *План ξ^* является локально D-оптимальным для модели* (5) *тогда и только тогда, когда*

$$m = \max_{x\in\mathbb{X}} d(x,\theta^*)$$

Кроме того,

$$\max_{x\in\mathbb{X}} d(x,\theta^*) = \inf_{\xi} \max_{x\in\mathbb{X}} d(x,\theta)$$

Функция $d(x,\xi^*)$ достигает максимального значение во всех точках любого локального D-оптимального плана. Информационные матрицы всех локально D-оптимального планов совпадают.

Доказательство. Без доказательства. Является переформулировкой теоремы Кифера-Вольфовица (которая видимо была раньше). □

3. Системы Чебышева. Два эквивалентных определения.

Определение (Конструктивное). *Пусть $u_0, \dots, u_n$ — заданные непрерывные вещественные функции на $[a,b]$. Система называется системой функций Чебышева, если определители*

$$U\begin{pmatrix} 0 & 1 & \dots & n \\ t_0 & t_1 & \dots & t_n \end{pmatrix} = \det\begin{pmatrix} u_0(t_0) & \dots & u_0(t_n) \\ u_1(t_0) & \dots & u_1(t_n) \\ \dots & \dots & \dots \\ u_n(t_0) & \dots & u_n(t_n) \end{pmatrix}$$

строго положительны для $\forall a \le t_0 < t_1 < \dots < t_n \le b$.[10]

Здесь нужно рассказать о (по всей видимости) естественности такой штуки через определитель Вандермонда, но я пока сам не понимаю.

Определение. *Обобщенным многочленом называется функция $u(t) = \sum_{i=0}^{n} a_i u_i(t)$, $a_i \in \mathbb{R}$.* [11]

Определение. *Многочлен называется нетривиальным, если $\sum_{i=0}^{n} a_i^2 \neq 0$.* ***Придирка****: Это условие глядится странновато. На u_i задана упорядоченность или нет? Если да, значит обобщенные многочлены не просто так названы многочленами. У любого нормального многочлена есть степень! Тут она тоже должна быть, иначе термин обобщенный многочлен слишком натянут. А если есть степень, то разумно требовать, чтобы коэффициент при старшем члене был не 0.*

[10] *На самом деле, ничего ведь страшного, если все определители будут строго отрицательны? Это используется в теореме этого билета, обратите на это внимание.*

[11] *Здесь не накладывается никаких дополнительных ограничений! Просто произвольная линейная комбинация.*

Количество нулей обобщенного многочлена u обозначим $Z(u)$.

Определение (Аксиоматическое). *Система вещественных, непрерывных функций $\{u_i\}_{i=0}^n$, определенных на отрезке $[a,b]$ называется системой Чебышева если $Z(u) \leqslant n$ для любого нетривиального обобщенного многочлена u, построенного по этой системе.*

Теорема. *Пусть $\{u_i\}_{i=0}^n$ — система вещественных непрерывных функций, определенных на отрезке $[a,b]$. СУР:*

1. *Система $\{u_i\}_{i=0}^n$ с точностью до знака некоторых из u_i* [12] *образует систему Чебышева 3.*
2. *Система $\{u_i\}_{i=0}^n$ образует систему Чебышева 3.*

Доказательство. Пусть $a = (a_0, \dots, a_n)^{\mathrm{T}} \in \mathbb{R}^{n+1}$ такой, что $\sum_{i=1}^n a_i^2 \neq 0$.. Рассмотрим обобщенный многочлен $u(t) = \sum_{i=0}^n a_i u_i(t)$. Для произвольного набора точек $\{t_i\}_{i=0}^n \subset [a,b]$ введем матрицу

$$\mathbf{U}(t_0, t_1, \dots, t_n) = \begin{pmatrix} u_0(t_0) & \dots & u_0(t_n) \\ u_1(t_0) & \dots & u_1(t_n) \\ \dots & \dots & \dots \\ u_n(t_0) & \dots & u_n(t_n) \end{pmatrix}.$$

$1 \to 2$. Пусть $Z(u) \geqslant n+1$ и $t_0, t_1, \dots, t_n$ — первые $n+1$ нулей многочлена u. Тогда $\mathbf{U}(t_0, t_1, \dots, t_n)a = \mathbf{0}$[13], что противоречит невырожденности $\mathbf{U}$.

$2 \to 1$. Пусть система $\{u_i\}_{i=0}^n$ — не чебышевская в смысле определения 3. Тогда найдется такой набор точек $t_0, t_1, \dots, t_n$, матрица $\mathbf{U} = \mathbf{U}(t_0, t_1, \dots, t_n)$ и вектор $a = (a_0, a_1 \dots, a_n)^{\mathrm{T}} \in \mathbb{R}^{n+1}$, что $\mathbf{U}a = \mathbf{0}$. То есть существует обобщенный многочлен, имеющий не менее $n+1$ нуля. Противоречие. □

4. Системы Чебышева. Метод проверки, основанный на последовательном дифференцировании. Примеры применения (экспоненциальные модели.)

Пусть $u_0, u_2, \dots, u_k$ — некоторая система функций. Мы хотим проверить, что она является Чебышевской. Рассмотрим следующий набор функций:

$$F_{00}(t) = u_0(t), \dots, F_{0k}(t) = u_n(t)$$

$$F_{11}(t) = \left(\frac{F_{01}}{F_{00}}\right)', \dots, F_{1k}(t) = \left(\frac{F_{0k}}{F_{00}}\right)'$$

$$F_{22}(t) = \left(\frac{F_{1k}}{F_{11}}\right)', \dots, F_{2k}(t) = \left(\frac{F_{1k}}{F_{11}}\right)'$$

$$\dots$$

$$F_{kk} = \left(\frac{F_{k-1,k}}{F_{k-1,k-1}}\right)'$$

Теорема. *Если существуют все функции F_{ij} и $F_{ii} > 0$, то система $u_0, \dots, u_k$ является системой Чебышева.*

[12] *Наверное это нужно написать формально, но мне не приходят в голову изящные способы*

[13] Здесь временненый шрифт.

Доказательство. Пусть это не так. Тогда $\exists u(t) = \sum\limits_{i=0}^{k} a_i u_i$, обращающийся в 0 в $k+1$ точках. Не умаляя общности будем считать, что все $a_i > 0$. Тогда

$$f_0(t) = a_0 u_0(t) \left(1 + \frac{a_1}{a_0}\frac{u_1(t)}{u_0(t)} + \dots \frac{a_k}{a_0}\frac{u_k(t)}{u_0(t)} \right)$$

По условию, $u_0(t) > 0$, а значит вторая скобка обращатся в 0 в $k+1$ точках. Вспоминаем теорему Ролля — между двумя корнями непрерывной функции есть корень ее производной. Отсюда следует, что функция $f_1(t) = \left(1 + \frac{a_1}{a_0}\frac{u_1(t)}{u_0(t)} + \dots \frac{a_k}{a_0}\frac{u_k(t)}{u_0(t)} \right)'$ — обращется в ноль в k точках. Заметим, что количество слагаемых уменьшилось на 1. Итерируя процесс, получим последовательность функций $f_0(t), f_1(t), \dots, f_k(t)$. В $f_i(t)$ будет $k-i+1$ ненулевых слагаемых и $k-i$ нулей. Таким образом, $f_k(t) = \alpha F_{kk}$, где α — некоторое ненулевое число — имеет хотя бы один ноль. Противоречие, т.к. по предположению $f_k(t) = F_{kk}(t) > 0 \forall\, t$ □

4.1. Пример: Экспоненциальная регрессия

Пусть $\eta(t,\theta) = \sum\limits_{i=1}^{k} b_i e^{\lambda_i t}$, $b_i \in \mathbb{R}, \lambda_i \in \mathbb{R}, \lambda_i \neq \lambda_j\ i \neq j$. В данной модели параметрами являются b_i и λ_i и они входят нелинейно. Рассмотрим систему функций $\left\{ \frac{\partial\eta(t,\theta)}{\partial\lambda_i}, \frac{\partial\eta(t,\theta)}{\partial b_i} \right\}_{i=1}^{k}$. Оказывается, данная система является системой Чебышева. Для доказательства достаточно повторить рассуждение, легшее в основу доказательства прошлой теоремы (4) и воспользоваться тем, что $e^{\lambda t} > 0 \forall \lambda \in R$.

4.2. Пример: модель Михаэлина-Менте

[14] $\eta(t,\theta) = \frac{at}{t+b}$ на $[a,b]$, $a > 0$. Производные $\left\{ \frac{\partial\theta}{\theta_i} \right\}$ также образуют систему Чебышева. Действительно,

$$\frac{\partial\eta}{a} = u_0(t) = \frac{t}{t+b}$$
$$\frac{\partial\eta}{b} = u_1(t) = \frac{-at}{(t+b)^2}$$

Пусть имеется $u(t) = \alpha_0 u_0(t) - \alpha_1 u_1(t)$. Вынесем $-u_1(t)$ за скобку и получим

$$u(t) = \frac{at}{(t+b)^2}\left(\alpha_0(t+b) + \alpha_1\right)$$

Вспомним, что $a > 0$, а значит $t > 0$ и $\frac{at}{(t+b)^2} > 0$. Второе слагаемое — линейная функция, про которую мы и без дифференцирования знаем, что у нее имеется не более одного нуля.

5. Расширенные системы Чебышева.

6. Неотрицательные многочлены с заданными нулями

6.1. Теорема о числе нулей

Определение. *Пусть* u *— некоторая функция (непрерывная) на* $[a,b]$. *Тогда* $Z(u)$ *— число нулей* u *на* $[a,b]$.

[14] я наверно не правильно распарсил имена, надо поправить

Определение. *Ноль называется узловым, если*

- *Он совпадает с граничной точкой (либо a, либо b)*
- *Функция меняет знак, проходя через этот ноль*

В противном случае ноль называется неузловым.

Определение. *$\overline{Z(u)}$ — число нулей функции u, где неузловые нули засчитываются дважды.*

Теорема. *Если система функций $\{u_i\}_{i=0}^n$ - Чебышевская на $[a,b]$, то для любого нетривиального многочлена $\overline{Z}(u) \le n$.*

Доказательство. Пусть $\bar{(Z)}(u) \ge n+1$ для некоторого нетривиального u. Обозначим различные нули u через $t_1, \ldots, t_k$. Добавим для первого неузлового нуля точки $t_i - \varepsilon, t_i + \varepsilon$, а для остальных неузловых нулей точки $t_i - \varepsilon$. Выбрав ε достаточно маленьким, можно получить, что все точки будут содержаться в $[a,b]$. Пусть у нас было m_1 узловых и m_2 неузловых нулей. Тогда после проделанной операции мы получили $m_1+2m_2+1 \ge n+2$ точек ($m_1+2m_2 \ge n+1$). Переобозначим получившиеся точки за s_i и возьмем первые $n+2$ из них. Не умаляя общности, можем считать, что $u(s_i) \ge 0$ для четных i, $u(s_i) \le 0$ для нечетных i[15]. Отсюда получаем, что следующий определитель равен 0 (т.к. первая стручка — линейная комбинация следующих):

$$\begin{vmatrix} u(s_0) & u(s_1) & \ldots & u(s_{n+1}) \\ u_0(s_0) & u_0(s_1) & \ldots & u_0(s_{n+1}) \\ \ldots & \ldots & \ldots & \ldots \\ u_n(s_0) & u_n(s_1) & \ldots & u_n(s_{n+1}) \end{vmatrix} = 0 \tag{6}$$

Далее $\{u_i\}$ — система Чебышева, а значит

$$\begin{vmatrix} u_0(t_0) & u_0(s_1) & \ldots & u_0(t_n) \\ \ldots & \ldots & \ldots & \ldots \\ u_n(t_0) & u_n(t_1) & \ldots & u_n(t_n) \end{vmatrix} > 0$$

для любых $t_0 < t_1 < \ldots < t_n$. Поэтому, разложив определитель (6) по первой сточке, получим[16], что

$$\sum_{i=0}^{n+1} \alpha_i u(s_i) = 0,$$

где α_i строго чередуются в знаке. Кроме того, $u(s_i)$ совпадают по знаку с α_i. Таким образом, суммируются неотрицательные слагаемые. Значит $\forall i \; u(s_i) = 0$. Получили противоречие с одним из определений системы Чебышева(3) □

Теорема. *Обратно, если для любого нетривиального многочлена $u(t)$ верно, что $\overline{Z}(u) \le n$, то система является Чебышевской*

Доказательство. Следует из второго определения чебышевской системы[17] 3:

$$Z(u) \le \overline{Z}(u) \le n$$

□

[15]Проверьте это. Достаточно нарисовать рисунок и все станет ясно.

[16]как мы все помним, при разложение определителя знаки перед минорами чередуются, а сами миноры у нас положительны

[17]$Z(u)$ ведь количество нулей многочлена

6.2. Неотрицательные многочлены с заданными нулями

Задача: построить неотрицательный многочлен, имеющий нули в точках $t_1 < t_2 < \ldots < t_k$. Многочлен неотрицательный, поэтому все внутренние нули должны быть неузловыми. Введем функцию ω:

$$\omega(t) = \begin{cases} \omega(a) = 1 \\ \omega(b) = 1 \\ \omega(t_i) = 2, i \in (a,b) \end{cases}$$

Теорема. *Пусть $t_1, \ldots, t_k$ — различные и такие, что $\sum\limits_{i=1}^{k} \omega(t_i) \leq n$. Пусть $\{u_i\}_{i=0}^{n}$ — чебышевская. Тогда $\exists u(t)$, который обращается в ноль в этих и только этих точках, за исключением случая, когда $n = 2m$ и одна из точек совпадает с граничной точкой*[18]

Доказательство. Докажем для $n = 2m + 1$ и $a < t_1 < \ldots < t_k < b$[19]. □

[18] *Исключение получается по следующей простой причине: до этого мы доказали теорему о том, что число нулей $\overline{Z} \leq n$. Если $n = 2m$, и одна точка совпадает с граничной, то $k < m$ и $2k + 1 < 2m$, а значит возможна ситуация, что во второй граничной точке также будет ноль.*

[19] Остальные случаи получаются аналогично с небольшими модификациями.