# Точечное оценивание параметров: ММП и МНК

Лекция №6

Статистические методы в ядерном эксперименте, ИЯФ, 2024 г.

1. Метод максимального правдоподобия (ММП)
2. Метод наименьших квадратов (МНК)

# Метод максимального правдоподобия (I)

ММП представляет собой технику оценки значений параметров распределений на основе конечного набора данных.
ММП также позволяет найти ошибки полученных оценок.

Идея метода состоит в том, что в качестве оценки параметра $\theta$ нужно взять такое его значение, которое максимизирует вероятность получения уже имеющейся выборки $\{X_k\}_n$.

Функцией правдоподобия (ФП) $F(\vec{X};\theta)$ для выборки дискретной/непрерывной с.в. называется функция:

$$F(\vec{X};\theta) = \begin{cases} P(x = X_1;\theta)\cdot \ldots \cdot P(x = X_n;\theta) \\ f(X_1;\theta)\cdot \ldots \cdot f(X_n;\theta) \end{cases} = \begin{cases} \prod\limits_{i=1}^{n} P(x = X_i;\theta) \\ \prod\limits_{i=1}^{n} f(X_i;\theta) \end{cases}$$

где $f(x;\theta)$ – ф.п.в. непрерывной с.в.

Таким образом, ФП в дискретном случае есть вероятность реализации наблюдаемой выборки $\{X_k\}_n$. В непрерывном случае ФП есть совместная плотность вероятности $n-$мерного случайного вектора выборки.

# Метод максимального правдоподобия (II)

Логарифмической функцией правдоподобия (ЛФП) $\mathcal{L}(\vec{X};\theta)$ называется функция:

$$\mathcal{L}(\vec{X};\theta) = \ln F(\vec{X},\theta) = \sum_{k=1}^{n} f(X_k;\theta)$$

Оценкой максимального правдоподобия $\theta^*$ для параметра $\theta$ называется такое значение $\theta$, при котором достигается максимум (Л)ФП.

Замечание: Так как функция $y = \ln x$ строго монотонна, то экстремум функции $\mathcal{L}(\vec{X};\theta)$ достигается в той же точке $\theta^*$, что и для $F(\vec{X},\theta)$:

$$\left.\frac{\partial\mathcal{L}}{\partial\theta}\right|_{\theta=\theta^*} = 0 \text{ – условие экстремума}$$

$$\left.\frac{\partial^2\mathcal{L}}{\partial\theta^2}\right|_{\theta=\theta^*} < 0 \text{ – условие максимума}$$

# Метод максимального правдоподобия (III)

Замечание: Статистика $f_\theta(X_1, ..., X_n)$ называется достаточной относительно параметра $\theta$, если она содержит в себе всю информацию о параметре $\theta$, которая может быть получена на основе заданной выборки $\{X_k\}_n$.

Пример 1: Пусть $\{X_k\}_n$ выборка из распределения Бернулли с параметром $p$. Найдём оценку $p^*$ параметра $p$ методом МП.

Пример 2: Пусть $\{X_k\}_n$ выборка из распределения Пуассона с параметром $\mu$. Найдём оценку $\mu^*$ параметра $\mu$ методом МП.

Пример 3: Пусть $\{X_k\}_n$ выборка из распределения $\gamma(a, b)$. Найдём оценки параметров $a$ и $b$ методом МП.

Пример 4: Пусть $\{X_k\}_n$ выборка из распределения $N(\mu; \sigma)$ . Найдём оценки параметров $\mu$ и $\sigma$ методом МП.

# Метод максимального правдоподобия (IV)

## Свойства оценок методом МП:

- Если существует достаточная оценка, то ММП дает именно её и тем самым исчерпывает всю имеющуюся информацию. Таким образом, оценка ММП является эффективной в своем классе.

**Эффективность**

Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется эффективной оценкой параметра $\theta$ в некотором классе состоятельных оценок, если она не хуже в среднеквадратичном смысле $E((\theta^* - \theta)^2) \leq E((\theta_1^* - \theta)^2)$.

- Оценка $\theta^*$ ММП является асимптотически нормальной со средним $\theta$ (т.е. асимптотически несмещённой) и дисперсией $D(\theta) = -1/E(\frac{\partial^2 \mathcal{L}(\vec{X},\theta)}{\partial\theta^2})$.
- Оценка ММП инвариантна относительно преобразований оцениваемого параметра. То есть, если $g = g(\theta)$, то $g^* = g(\theta^*)$. Так, если $s^2 = \frac{1}{n}\sum\limits_{k=1}^{n}(X_k - \langle X\rangle)^2$ есть оценка ММП для дисперсии $\sigma^2$, то оценкой ММП для среднеквадратичного отклонения $\sigma = \sqrt{\sigma^2}$ будет просто $s = \sqrt{s^2}$.
- Фактором, ограничивающим применимость ММП, является необходимость знания ф.п.в. для построения функции правдоподобия. Поэтому ММП дает лишь лучшее значение параметра уже выбранной модели (ф.п.в.).

# Метод максимального правдоподобия (V)

Пример 5: Пусть имеется $n$ независимых и нормально распределенных результатов измерений $Y_k$ некоторой величины $y$, истинное (неизвестное) значение которой есть $\theta$, а $\sigma_k^2$ – соответствующие известные дисперсии. Провести усреднение результатов этих измерений, используя ММП.

Функция правдоподобия для статистики $\{Y_k\}_n$ будет:

$$F(\vec{Y};\theta) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(Y_k-\theta)^2}{2\sigma_k^2}}$$

$$\mathcal{L}(\vec{Y};\theta) = \ln F(\vec{Y};\theta) = -\left[\frac{n}{2}\ln 2\pi + \sum_{k=1}^{n} \ln \sigma_k\right] - \frac{1}{2}\sum_{k=1}^{n} \frac{(Y_k-\theta)^2}{\sigma_k^2}$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{k=1}^{n} \frac{(Y_k-\theta)}{\sigma_k^2} = 0 \Rightarrow \theta^* = \sum_{k=1}^{n} \frac{Y_k}{\sigma_k^2} \Big/ \sum_{k=1}^{n} \frac{1}{\sigma_k^2}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \theta^2} = -\sum_{k=1}^{n} \frac{1}{\sigma_k^2} \Rightarrow D(\theta^*) = \left(\sum_{k=1}^{n} \frac{1}{\sigma_k^2}\right)^{-1}$$

Если $\sigma_k = \sigma$ (одинаковая точность), то получаем ожидаемый результат: $\theta^* = \langle y \rangle$, $\sigma_\theta = \sigma/\sqrt{n}$.

# Метод наименьших квадратов (I)

В предыдущем примере фактически минимизировалась функция $\tilde{\mathcal{L}}(\vec{Y};\theta) = \sum\limits_{k=1}^{n} \frac{(Y_k-\theta)^2}{\sigma_k^2}$, равная сумме квадратов отклонений измеренных значений от истинных значений функции, но отклонения берутся в единицах стандартных отклонений. Это одна из реализаций т.н. метода наименьших квадратов (МНК).
Первой работой, в которой использовался МНК, является работа Лежандра 1805 г., в которой метод применялся для определения параметров орбит комет.
Возможны и другие критерии оптимизации, в результате получим другие оценки параметров.
Так, в 1792 г. Лаплас применял критерий, согласно которому наименьшее значение должна принимать сумма абсолютных значений, а не квадратов отклонений.
В 1831 г. Коши предложил метод, в котором требовалось, чтобы наибольшее из всех отклонений (по абсолютной величине) было минимальным. Этот принцип минимакса применяется и по сей день.

# Метод наименьших квадратов (II)

Пусть наблюдаемая случайная величина $Y$ зависит от случайной величины $X$, $Y = Y(X; \vec{\theta})$. Значения величины $X$ мы наблюдаем или задаем. Обозначим через $f(t; \vec{\theta})$ функцию, отражающую зависимость среднего значения величины $Y$ от значения $X$, $E(Y(X = t; \theta)) = f(t; \vec{\theta})$. Функция $f(t; \vec{\theta})$ называется линией регрессии $Y$ на $X$.

В результате $n$ измерений, в которых $X$ принимает набор значений $t_1$, $t_2$, ..., $t_n$, получим значения $y_1$, $y_2$, ..., $y_n$. Рассмотрим случай линейной зависимости $Y$ от оцениваемых параметров $\vec{\theta}$:

$$f(t; \vec{\theta}) = \theta_1 g_1(t) + \ldots + \theta_m g_m(t)$$

Тогда в общем случае можно записать: $y_k = f(t_k; \vec{\theta}) + \varepsilon_k$, где $\varepsilon_k$ – случайные отклонения, такие что $E(\varepsilon_k) = 0$ и $D(\varepsilon_k) = \sigma_k^2$.

Найдём оценку вектора параметров $\vec{\theta}$ из условия минимума квадратичной формы $S$, которая в общем случае наличия корреляций между величинами $y_k$, имеет вид:

$$S = \sum_{i,j=1}^{n} (y_i - f(t_i; \vec{\theta})) D_{ij}^{-1} (y_j - f(t_j; \vec{\theta})),$$

где $D_{ij}$ – ковариационная матрица случайной величины $y_k$.

Используем обозначения:

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \vec{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_m \end{pmatrix} \hat{\mathcal{G}} = \begin{pmatrix} g_1(t_1) & \cdots & g_m(t_1) \\ \vdots & \ddots & \vdots \\ g_1(t_n) & \cdots & g_m(t_n) \end{pmatrix}$$

Тогда:

$$\vec{y} = \hat{\mathcal{G}}\vec{\theta} + \vec{\varepsilon}, \;\; \vec{\varepsilon} = \vec{y} - \hat{\mathcal{G}}\vec{\theta}, \;\; S = \vec{\varepsilon}^{\dagger} D^{-1}\vec{\varepsilon} = (\vec{y} - \hat{\mathcal{G}}\vec{\theta})^{\dagger} D^{-1}(\vec{y} - \hat{\mathcal{G}}\vec{\theta})$$

$$\delta S = 2\delta\vec{\theta}^{\dagger}(\hat{\mathcal{G}}^{\dagger} D^{-1}\hat{\mathcal{G}}\vec{\theta} - \hat{\mathcal{G}}^{\dagger} D^{-1}\vec{y}) = 0, \; \hat{\mathcal{G}}^{\dagger} D^{-1}\hat{\mathcal{G}}\vec{\theta} = \hat{\mathcal{G}}^{\dagger} D^{-1}\vec{y},$$

$$\hat{\mathcal{B}} = \hat{\mathcal{G}}^{\dagger} D^{-1}\hat{\mathcal{G}}, \; \hat{\mathcal{B}} = \hat{\mathcal{B}}^{\dagger} \;\; \Rightarrow \vec{\theta}^{*} = \hat{\mathcal{B}}^{-1}\hat{\mathcal{G}}^{\dagger} D^{-1}\vec{y}$$

$$D(\vec{\theta}^{*}) = E((\vec{\theta}^{*} - E(\vec{\theta}^{*}))(\vec{\theta}^{*} - E(\vec{\theta}^{*}))^{\dagger}) =$$

$$E(\hat{\mathcal{B}}^{-1}\hat{\mathcal{G}}^{\dagger} D^{-1}(\vec{y} - E(\vec{y}))(\vec{y} - E(\vec{y}))^{\dagger} D^{-1}\hat{\mathcal{G}}\hat{\mathcal{B}}^{-1}) =$$

$$= \hat{\mathcal{B}}^{-1}\hat{\mathcal{G}}^{\dagger} D^{-1} E((\vec{y} - E(\vec{y}))(\vec{y} - E(\vec{y}))^{\dagger}) D^{-1}\hat{\mathcal{G}}\hat{\mathcal{B}}^{-1} = \hat{\mathcal{B}}^{-1}\hat{\mathcal{G}}^{\dagger} D^{-1} D D^{-1}\hat{\mathcal{G}}\hat{\mathcal{B}}^{-1} =$$

$$= \hat{\mathcal{B}}^{-1} \;\; \Rightarrow \;\; D(\vec{\theta}^{*}) = \hat{\mathcal{B}}^{-1}$$

# Метод наименьших квадратов (IV)

Пример 1: Пусть $f(t;\vec{\theta}) = \theta_1 + \theta_2 t$ и $\{y_k\}$ независимы. Найдём $\theta_1^*$ и $\theta_2^*$ методом наименьших квадратов.

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \hat{\mathcal{G}} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}, D = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ . & . & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

$$D^{-1} = \begin{pmatrix} \sigma_1^{-2} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-2} & \cdots & 0 \\ . & . & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n^{-2} \end{pmatrix}, \hat{\mathcal{B}} = \hat{\mathcal{G}}^\dagger D^{-1} \hat{\mathcal{G}} = \begin{pmatrix} \sum\limits_{i=1}^n \frac{1}{\sigma_i^2} & \sum\limits_{i=1}^n \frac{t_i}{\sigma_i^2} \\ \sum\limits_{i=1}^n \frac{t_i}{\sigma_i^2} & \sum\limits_{i=1}^n \frac{t_i^2}{\sigma_i^2} \end{pmatrix}$$

$$\hat{\mathcal{B}}^{-1} = \frac{1}{\det\hat{\mathcal{B}}} \begin{pmatrix} \sum\limits_{i=1}^n \frac{t_i^2}{\sigma_i^2} & -\sum\limits_{i=1}^n \frac{t_i}{\sigma_i^2} \\ -\sum\limits_{i=1}^n \frac{t_i}{\sigma_i^2} & \sum\limits_{i=1}^n \frac{1}{\sigma_i^2} \end{pmatrix}, \det\hat{\mathcal{B}} = \left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)\left(\sum_{i=1}^n \frac{t_i^2}{\sigma_i^2}\right) - \left(\sum_{i=1}^n \frac{t_i}{\sigma_i^2}\right)^2$$

$$\det\hat{\mathcal{B}} = \left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)^2 (\langle t^2\rangle - \langle t\rangle^2) = \left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)\left(\sum_{i=1}^n \frac{(t_i-\langle t\rangle)^2}{\sigma_i^2}\right), \ \langle t\rangle = \sum_{i=1}^n \frac{t_i}{\sigma_i^2} \Big/ \sum_{i=1}^n \frac{1}{\sigma_i^2}$$

$$\theta_1^* = \frac{1}{\det\hat{\mathcal{B}}}\left(\left(\sum_{i=1}^n \frac{y_i}{\sigma_i^2}\right)\left(\sum_{i=1}^n \frac{t_i^2}{\sigma_i^2}\right) - \left(\sum_{i=1}^n \frac{t_i}{\sigma_i^2}\right)\left(\sum_{i=1}^n \frac{t_i y_i}{\sigma_i^2}\right)\right) = \frac{\langle y\rangle\langle t^2\rangle - \langle t\rangle\langle ty\rangle}{\langle t^2\rangle - \langle t\rangle^2}$$

$$\theta_2^* = \frac{1}{\det\hat{\mathcal{B}}}\left(\left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)\left(\sum_{i=1}^n \frac{t_i y_i}{\sigma_i^2}\right) - \left(\sum_{i=1}^n \frac{t_i}{\sigma_i^2}\right)\left(\sum_{i=1}^n \frac{y_i}{\sigma_i^2}\right)\right) = \frac{\langle ty\rangle - \langle t\rangle\langle y\rangle}{\langle t^2\rangle - \langle t\rangle^2}$$

# Метод наименьших квадратов (V)

Пример 2: В предыдущем примере $\theta_1^*$ и $\theta_2^*$ оказались коррелированными ($(\hat{\mathcal{B}}^{-1})_{12} \neq 0$). Если ввести $f(t;\vec{\omega}) = \omega_1 + \omega_2(t - \langle t\rangle)$, где $\omega_1 = \theta_1 + \theta_2\langle t\rangle$ и $\omega_2 = \theta_2$, $\langle t\rangle = \sum\limits_{i=1}^{n} \frac{t_i}{\sigma_i^2} \Big/ \sum\limits_{i=1}^{n} \frac{1}{\sigma_i^2}$, то:

$$\vec{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} 1 & \langle t\rangle \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \hat{\mathcal{A}}\vec{\theta} \;\Rightarrow\; \vec{\omega}^* = \hat{\mathcal{A}}\vec{\theta}^*$$

$$\omega_1^* = \langle y\rangle = \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} \Big/ \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$$

$$\omega_2^* = \frac{\langle ty\rangle - \langle t\rangle\langle y\rangle}{\langle t^2\rangle - \langle t\rangle^2} = \frac{\langle y(t-\langle t\rangle)\rangle}{\langle (t-\langle t\rangle)^2\rangle} = \sum_{i=1}^{n} \frac{y_i(t_i-\langle t\rangle)}{\sigma_i^2} \Big/ \sum_{i=1}^{n} \frac{(t_i-\langle t\rangle)^2}{\sigma_i^2}$$

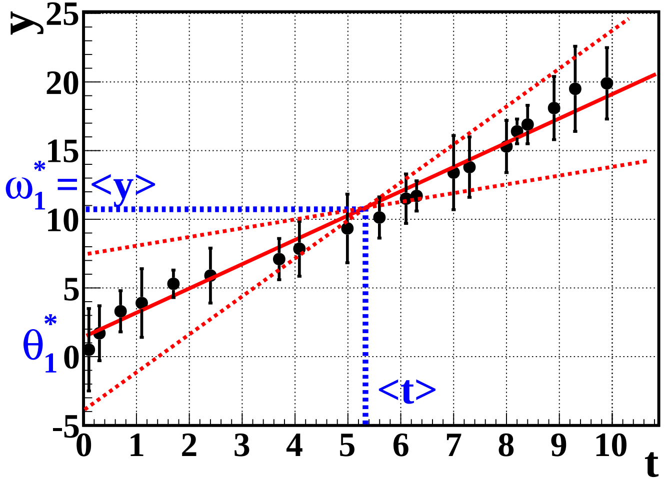
$$D(\vec{\omega}^*) = E((\vec{\omega} - \vec{\omega}^*)(\vec{\omega} - \vec{\omega}^*)^\dagger) = \hat{\mathcal{A}}E((\vec{\theta} - \vec{\theta}^*)(\vec{\theta} - \vec{\theta}^*)^\dagger)\hat{\mathcal{A}}^\dagger = \hat{\mathcal{A}}\hat{\mathcal{B}}^{-1}\hat{\mathcal{A}}^\dagger$$

$$\hat{\mathcal{B}}^{-1} = \frac{\Sigma^2}{\Delta_t^2} \begin{pmatrix} \langle t^2\rangle & -\langle t\rangle \\ -\langle t\rangle & 1 \end{pmatrix}, \quad \Sigma^2 = \left(\sum_{i=1}^{n} 1/\sigma_i^2\right)^{-1}, \quad \Delta_t^2 = \langle t^2\rangle - \langle t\rangle^2$$

$$D(\vec{\omega}^*) = \frac{\Sigma^2}{\Delta_t^2} \begin{pmatrix} 1 & \langle t\rangle \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \langle t^2\rangle & -\langle t\rangle \\ -\langle t\rangle & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \langle t\rangle & 1 \end{pmatrix} = \begin{pmatrix} \Sigma^2 & 0 \\ 0 & \frac{\Sigma^2}{\Delta_t^2} \end{pmatrix}$$

Параметры $\omega_1^*$ и $\omega_2^*$ некоррелированы.

Таким образом при линейной аппроксимации искомая прямая линия всегда проходит через точку $(\langle t \rangle, \langle y \rangle)$. Причём параметр $\omega_1^* = \langle y \rangle$ не меняется при вариации параметра $\omega_2^*$ (наклона прямой).



# Метод наименьших квадратов (VII)

Рассмотрим случай, когда величина $y = f(t; \vec{\theta})$ есть произвольная функция параметров $\vec{\theta}$. Тогда, если $\vec{\theta}'$ есть некоторое достаточно хорошее начальное приближение к истинному значению $\vec{\theta}^*$, то:

$$f(t_i; \vec{\theta}) = f(t_i|\vec{\theta}') + \nabla f(t_i|\vec{\theta}')(\vec{\theta} - \vec{\theta}') + \mathcal{O}((\vec{\theta} - \vec{\theta}')^2)$$

Таким образом, линеаризуя зависимость $y$ от искомых параметров $\vec{\theta}$ становится возможным применить МНК. При этом оптимальное значение $\vec{\theta}^*_1$, найденное после первой итерации МНК, используются снова в качестве нового, более точного, приближения $\vec{\theta}'$ и процедура повторяется.
На некотором шаге отличие новых оптимальных значений $\vec{\theta}^*$ от входных значений $\vec{\theta}'$ становится меньше наперёд заданного малого $\varepsilon$, и процедура определения $\vec{\theta}^*$ считается завершенной.

# Метод наименьших квадратов (VIII)

Оценки параметров, полученные МНК, обладают следующими свойствами, <u>независящими от вида распределения</u> величин $\vec{\varepsilon} = \vec{y} - \hat{\mathcal{G}}\vec{\theta}$, при условиях, что $E(\varepsilon_k) = 0$, $D(\varepsilon_k) < \infty$ и $\varepsilon_k$ независимы:

- $\vec{\theta}^*$ является <u>несмещенной</u> оценкой вектора параметров $\vec{\theta}$.
- В классе несмещённых оценок, линейных по $\vec{y}$, $\vec{\theta}^*$ является <u>эффективной</u> оценкой.
- Если $D(\varepsilon_k) = \sigma^2\ \forall k$, то для остаточной суммы квадратов $R$ имеем:

$$R = (\vec{y} - \hat{\mathcal{G}}\vec{\theta^*})^\dagger(\vec{y} - \hat{\mathcal{G}}\vec{\theta^*}), \quad E(R) = \sigma^2(n-m),$$

  т.е. остаточную сумму квадратов $R$ можно использовать для оценки дисперсии $\varepsilon_k$: $\sigma^{2*} = R/(n-m)$, где $m$–число параметров $\vec{\theta}^*$, определяемых из минимизации.
- Если же $\sigma^2$ известна, то величина $R/(\sigma^2(n-m))$ (ее близость к 1) может быть использована как характеристика "качества" аппроксимации (с помощью критерия согласия $\chi^2$).