Оценивание вещественного параметра (продолжение)

Малов Сергей Васильевич

Санкт-Петербургский государственный электротехнический университет

10 октября 2020 г.

План

1. Минимаксный и байесовский подходы
2. Информационное неравенство
3. Асимптотическая нормальность оценок максимального правдоподобия

Минимаксный и байесовский подходы

Определим следующие функционалы риска:

- Максимальный риск: $R_M(\delta) = \sup\limits_{\theta \in \Theta} R_\delta(\theta)$
- Байесовский риск по отношению к априорному распределению Q:
$$R_Q(\delta) = \int_\Theta R_\delta(\theta)\, dQ(\theta)$$
  - Q – распределение (априорное) на борелевских подмножествах Θ.

Определение

Оценка $\delta(X)$ параметра θ называется минимаксной, если она минимизирует максимальный риск, т. е. $R_M(\delta) \le R_M(\delta^*)$ для любой оценки $\delta^*(X)$.

Определение

Оценка $\delta(X)$ параметра θ называется байесовской по отношению к априорному распределению Q, если она минимизирует байесовский риск, т. е. $R_Q(\delta) \le R_Q(\delta^*)$ для любой оценки $\delta^*(X)$.

Байесовское оценивание

Байесовский подход позволяет говорить о совместном распределении наблюдений и параметра на измеримом пространстве $\mathfrak{X} \times \Theta$.

- $f(x, \theta) = q(\theta) p_\theta(x)$ – плотность (дискретная плотность, плотность относительно некоторой доминирующей меры μ^*: $\mu^*(dx; d\theta) = \mu(dx)\,\mu'(d\theta)$)
- По формуле Байеса получаем апостериорное распределение с плотностью
$$f(\theta \,|\, X) = \frac{q(\theta) p_\theta(X)}{\int_\Theta q(\theta) p_\theta(X)\,\mu'(d\theta)}.$$
- Байесовский риск (используем функцию потерь Гаусса) имеет вид
$$R_Q(\delta) = \int_\Theta \mathbb{E}_\theta (\delta(X) - \theta)^2\, Q(d\theta) = \int_\Theta \mathbb{E}_\theta (\delta(X) - \theta)^2\, q(\theta) d\mu'(\theta).$$
- Путем дифференцирования по δ находим $\delta_*(X)$, минимизирующую байесовский риск
$$\delta_*(X) = \frac{\int_\Theta \theta f(X;\, \theta)\, Q(d\theta)}{\int_\Theta f(X;\, \theta)\, Q(d\theta)} = \int_\Theta \theta\, f(\theta \,|\, X)\, Q(d\theta)$$
- Байесовская оценка – условное (апостериорное) среднее параметра θ при условии X.

Байесовское оценивание

Упражнение

Пусть $X_1, \ldots, X_n$ – выборка из $\mathcal{N}(\theta, \sigma^2)$ с известным $\sigma = \sigma_0$. Найти байесовскую оценку θ в предположении, что параметр θ имеет априорное нормальное распределение $N(\mu, b^2)$ (μ, b^2 – известны).

Решение. Совместная плотность θ и $X = (X_1, \ldots, X_n)$ имеет вид

$$q(\theta)p_\theta(x) = \frac{1}{2\pi\sigma_0 b}\exp\Big(-\frac{1}{2\sigma_0^2}\textstyle\sum_{i=1}^n (x_i-\theta)^2\Big)\exp\Big(-\frac{1}{2b^2}(\theta-\mu)^2\Big).$$

Тогда апостериорная плотность имеет вид

$$U(X)\exp\left(-\frac{1}{2}\theta^2\left(\frac{n}{\sigma_0^2}+\frac{1}{b^2}\right)+\theta\left(\frac{n\overline{X}}{\sigma_0^2}+\frac{\mu}{b^2}\right)\right)=U(X)\exp\left(-\frac{1}{2}\left(\frac{n}{\sigma_0^2}+\frac{1}{b^2}\right)\left(\theta^2-2\theta\frac{n\overline{X}+\mu\sigma_0^2/b^2}{n+\sigma_0^2/b^2}\right)\right),$$

где $U(X) = \dfrac{\exp\left(-\frac{1}{2}\left(\frac{\sum_i X_i^2}{\sigma_0^2}+\frac{\mu^2}{b^2}\right)\right)}{\int_{-\infty}^{\infty} p(\theta)f(X;\theta)\,d\theta}$ – нормирующий множитель. Получили нормальную плотность со средним $\mathbb{E}(\theta|X) = \dfrac{n\overline{X}+\mu\sigma_0^2/b^2}{n+\sigma_0^2/b^2}$ и дисперсией $\mathbb{D}(\theta\,|\,X) = (n+\sigma_0^2/b^2)^{-1}$. Следовательно, байесовская оценка имеет вид

$$\delta(X) = \mathbb{E}(\theta\,|\,X) = \frac{n}{n+\sigma_0^2/b^2}\overline{X} + \frac{\mu}{nb^2/\sigma_0^2+1} \qquad \blacksquare$$

Минимаксное оценивание

Теорема (Леман)

Пусть $\{\delta_k\}_{k\in\mathbb{N}}$– последовательность байесовских оценок по отношению к априорным распределениям $\{Q_k\}$ соответственно; оценка δ:

$$\sup_\theta R_\delta(\theta) \leq \varlimsup_{k\to\infty} \int_\Theta R_{\delta_k}(\theta)\, dQ_k.$$

Тогда δ – минимаксна.

Доказательство. Пусть δ^* – произвольная оценка. Тогда, поскольку Q_k – вероятностная мера и поскольку δ_k – байесовская:

$$\sup_\theta R_\theta(\delta^*) \geq \int_\Theta R_{\delta^*}(\theta)\, Q_k(d\theta) \geq \int_\Theta R_{\delta_k}(\theta)\, Q_k(d\theta).$$

Переходим к пределу

$$\sup_\theta R_{\delta^*}(\theta) \geq \varlimsup_{k\to\infty} \int_\Theta R_{\delta_k}(\theta)\, Q_k(d\theta) \geq \sup_\theta R_\delta(\theta).$$

Следовательно, δ – минимаксна. ∎

Минимаксное оценивание

Упражнение

Пусть $X_1, \ldots, X_n$ – выборка из $\mathcal{N}(\theta, \sigma^2)$ с известным $\sigma = \sigma_0$. Найти минимаксную оценку θ.

Решение. Предположим, что параметр θ имеет нормальное распределение $\mathcal{N}(0, k)$, $k \in \mathbb{N}$. Получаем последовательность байесовских оценок:

$$\delta_k(X) = \frac{n\overline{X}}{n + \sigma_0^2/k}$$

и соответствующую последовательность байесовских рисков:

$$R(\delta_k) = \int\limits_{-\infty}^{\infty} \mathbb{E}_\theta \left(\frac{n\overline{X}}{n + \sigma_0^2/k} - \theta \right)^2 Q_k(d\theta) = \int\limits_{-\infty}^{\infty} \mathbb{E}_\theta \left(\frac{n(\overline{X} - \theta) - \theta\sigma_0^2/k}{n + \sigma_0^2/k} \right)^2 Q_k(d\theta) =$$

$$= \frac{1}{(n + \sigma_0^2/k)^2} \left(n\sigma_0^2 - 2\frac{\theta\sigma_0^2}{k} \mathbb{E}_\theta n(\overline{X} - \theta) + \int\limits_{-\infty}^{\infty} \theta^2\, Q(d\theta)/k^2 \right) \xrightarrow[k \to \infty]{} \sigma_0^2/n.$$

Далее отметим, что

$$\mathbb{E}_\theta(\overline{X} - \theta)^2 = \sigma_0^2/n.$$

Следовательно, по теореме Лемана $\overline{X}$ – минимаксная оценка θ. ∎

План

1. Минимаксный и байесовский подходы

2. Информационное неравенство

3. Асимптотическая нормальность оценок максимального правдоподобия

Информация по Фишеру

Эвристические предпосылки:

- Параметры тем легче различать, чем больше различаются соответствующие распределения
- Информация, содержащаяся в независимых экспериментах, равна сумме информации, содержащейся в каждом из них

Базовые предположения:

- $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ – статистический эксперимент
  - $\Theta \subseteq \mathbb{R}$ – параметр распределения $\theta \in \Theta$ вещественный
- Семейство $\mathcal{P}$ доминировано мерой μ.
  - обычно μ – либо мера Лебега (абсолютно непрерывный случай), либо считающая мера (дискретный случай)
  - по теореме Радона–Никодима существует соответствующее семейство плотностей $\{p_\theta\}_{\theta \in \Theta}$: $p_\theta = \dfrac{dP_\theta}{d\mu}$, $\theta \in \Theta$

Информация по Фишеру

Определение

Будем называть эксперимент регулярным, если при каждом $\theta \in \Theta$:

(i) $L(x;\theta)$ непрерывна и непрерывно дифференцируема по θ;

(ii) допустимо дифференцирование под знаком интеграла

$$\int_{\mathfrak{X}} \frac{\partial}{\partial\theta} p_\theta(x)\,\mu(dx) = \frac{\partial}{\partial\theta}\int_{\mathfrak{X}} p_\theta(x)\,\mu(dx) = 0;$$

(iii) существует и отличен от нуля интеграл

$$0 < I(\theta) = \mathbb{E}_\theta(U(X;\theta))^2 = \int_{\mathfrak{X}} (U(x;\theta))^2 L(x;\theta)\,\mu(dx),$$

- $U(x;\theta) = \frac{\partial}{\partial\theta}\log L(x;\theta)$.

- Величина $I(\theta)$ называется информацией Фишера, содержащейся в исходном наборе наблюдений.

Свойства информации Фишера

- Последнее равенство в условии (ii) выполнено всегда:
$$\int_{\mathfrak{X}} p_\theta(x)\,\mu(dx) = 1 \text{ при любом } \theta.$$
- Условие (ii) обычно нарушается, если параметр выходит на границы интеграла
  - в этом случае появляется производная интеграла с переменным пределом
  - в частности, это происходит, если носитель распределения $A = \{x : p_\theta(x) > 0\}$ зависит от параметра
- Условие (ii) может быть переписано в виде: $\frac{\partial}{\partial\theta}\mathbb{E}U(x;\theta) = 0$
- Если
$$\int_{\mathfrak{X}} \frac{\partial^2}{\partial\theta^2} p_\theta(x)\,\mu(dx) = \frac{\partial^2}{\partial\theta^2}\int_{\mathfrak{X}} p_\theta(x)\,\mu(dx) = 0;$$
(функция правдоподобия дважды непрерывно дифференцируема под знаком интеграла по θ), то
$$I(\theta) = -\mathbb{E}\Big(\frac{\partial^2}{\partial\theta^2}\log L(X;\theta)\Big).$$

Информация по Фишеру

Утверждение

Пусть $(\mathfrak{X}_1, \mathfrak{F}_1, \mathcal{P}_1)$ и $(\mathfrak{X}_2, \mathfrak{F}_2, \mathcal{P}_2)$ – независимые регулярные эксперименты с информацией Фишера $I_1(\theta)$ и $I_2(\theta)$ соответственно. Тогда эксперимент $(\mathfrak{X}_1 \times \mathfrak{X}_2, \sigma(\mathfrak{F}_1 \times \mathfrak{F}_2), \mathcal{P})$: $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$, где $P_\theta(dx) = P_{1\theta}(dx_1)P_{2\theta}(dx_2), x = (x_1, x_2)$ (т.е. $p_\theta(x) = p_{1\theta}(x_1)p_{2\theta}(x_2), x_1, x_2 \in \mathbb{R}$), регулярен и $I(\theta) = I_1(\theta) + I_2(\theta)$.

- Доказательство состоит в непосредственной проверке условий регулярности
- Эксперимент, состоящий в проведении набора независимых регулярных экспериментов, регулярен, а информация Фишера равна сумме информаций Фишера, составляющих его независимых экспериментов
- Информация Фишера I_θ, содержащаяся в выборке $X_1, X_2, \ldots, X_n$, в n раз больше информации, содержащейся в каждом наблюдении, т. е. $I(\theta) = nI_1(\theta)$.

Неравенство Рао–Крамера

Определение

Оценка δ называется разрешенной, если

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta\delta(X) = \mathbb{E}_\theta\Big(\delta(X)\frac{\partial}{\partial\theta}\log L(X;\theta)\Big).$$

- Оценка разрешенная, если допускается дифференцирование под знаком интеграла

$$\frac{\partial}{\partial\theta}\int_{\mathfrak{X}}\delta(x)L(x;\theta)\,\mu(dx) = \int_{\mathfrak{X}}\delta(x)\frac{\partial}{\partial\theta}L(x;\theta)\,\mu(dx).$$

Теорема (Неравенство Рао–Крамера)

Пусть эксперимент регулярен, δ – разрешенная оценка параметра θ. Тогда,

$$\mathbb{E}_\theta(\delta-\theta)^2 \geq \frac{(1+b'(\theta))^2}{I(\theta)} + b^2(\theta)$$

или

$$\mathbb{D}_\theta\delta \geq \frac{(1+b'(\theta))^2}{I(\theta)},$$

где $b(\delta) = \mathbb{E}\delta - \theta$ – смещение.

Неравенство Рао–Крамера

Доказательство. По определению смещения $\mathbb{E}_\theta\delta = \theta + b(\theta)$. Используя свойство разрешенности оценки, после дифференцирования получаем

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta\delta = \int_{\mathfrak{X}} \delta(x) f'(x,\theta)\,\mu(dx) = \mathbb{E}_\theta(\delta(X)\,U(X,\theta)).$$

Тогда, с учетом условия (ii) регулярности эксперимента, получаем равенство

$$\mathbb{E}_\theta\big((\delta(X) - \mathbb{E}_\theta\delta(X))U(X,\theta)\big) = 1 + b'(\theta).$$

Применяем неравенство Коши – Буняковского

$$(1 + b'(\theta))^2 \le \mathbb{E}_\theta(\delta(X) - \mathbb{E}_\theta\delta(X))^2\mathbb{E}_\theta U^2(X,\theta) = \mathbb{D}_\theta\delta(X)\ I(\theta),$$

из которого получаем второе неравенство. Отсюда первое неравенство получается тривиальным образом, так как $\mathbb{D}_\theta\delta(X) = \mathbb{E}_\theta(\delta-\theta)^2 - b_g^2(\theta)$. ∎

- Если δ – несмещенная оценка, то неравенство Рао – Крамера примет вид $$\mathbb{E}_\theta(\delta(X) - \theta)^2 \ge 1/I(\theta).$$

Эффективные по Фишеру оценки

Определение

Несмещенная оценка δ параметра θ, для которой достигается равенство в неравенстве Рао – Крамера, называется эффективной по Фишеру, или R-эффективной.

Условия существования R-эффективной оценки:

- Должна существовать несмещенная оценка
- Равенство в неравенстве Рао – Крамера достигается, если достигается равенство в неравенстве Коши–Буняковского
$$\mathbb{E}_\theta\Big((\delta(X) - \mathbb{E}_\theta\delta(X))U(X,\theta)\Big) = \mathbb{E}_\theta(\delta(X) - \mathbb{E}_\theta\delta(X))^2\mathbb{E}_\theta U^2(X,\theta).$$
- Последнее равенство выполнено только если
$$a_*(\theta)(\delta(X) - \mathbb{E}_\theta\delta(X)) = U(X;\,\theta) = \frac{\partial}{\partial\theta}\ln L(X;\,\theta).$$
- В этом случае,
$$L(X,\,\theta) = h(X)\,\exp(a(\theta)\delta(X) + r(\theta)).$$

Эффективные по Фишеру оценки

Определение

Несмещенная оценка δ параметра θ, для которой достигается равенство в неравенстве Рао – Крамера, называется эффективной по Фишеру, или R-эффективной.

Условия существования R-эффективной оценки:

- Должна существовать несмещенная оценка
- Равенство в неравенстве Рао – Крамера достигается, если достигается равенство в неравенстве Коши–Буняковского
$$\mathbb{E}_\theta\Big((\delta(X) - \mathbb{E}_\theta\delta(X))U(X,\theta)\Big) = \mathbb{E}_\theta(\delta(X) - \mathbb{E}_\theta\delta(X))^2\mathbb{E}_\theta U^2(X,\theta).$$
- Последнее равенство выполнено только если
$$a_*(\theta)(\delta(X) - \mathbb{E}_\theta\delta(X)) = U(X;\,\theta) = \frac{\partial}{\partial\theta}\ln L(X;\,\theta).$$
- В этом случае,
$$L(X,\theta) = h(X)\exp(a(\theta)\delta(X) + r(\theta)).$$

Условия существования R-эффективной оценки

Выводы:

- Эффективная по Фишеру оценка существует только если плотности распределения семейства распределений $\mathcal{P}$ представимы в виде
$$p_\theta(x_1,\ldots,x_n) = h(x_1,\ldots,x_n)\exp(a(\theta)\delta(x_1,\ldots,x_n) + r(\theta))$$
  - в частном случае выборки данное представление означает, что плотность распределения X_1
$$p_\theta(x) = h_*(x)\exp(a(\theta)\delta_*(x) + r(\theta)).$$
  - множество распределений такого вида – однопараметрическое экспоненциальное семейство
- При замене параметризации $\theta^* = g(\theta)$: $p_\theta(x) = p^*_{\theta^*}(x)$ свойства несмещености и R-эффективности оценки не сохраняются
  - очевидно, что $b^*(\theta^*)' = b'(\theta)/g'(\theta)$ и $I^*(\theta^*) = I(\theta)/g(\theta)^2$.
  - неравенство Рао–Крамера: $\mathbb{D}_{\theta^*}\delta \geq \frac{(g'(\theta)+b'(\theta))^2}{I(\theta)} + b^2(\theta)$, $\theta = g^{-1}(\theta^*)$
  - поскольку $0 = \mathbb{E}_\theta U(X,\theta) = a'(\theta)\mathbb{E}_\theta\delta(X) + r'(\theta)$, для каждого однопараметрического экспоненциального семейства существует единственная параметризация $\theta^* = -r'(\theta)/a'(\theta)$, в которой допускается R-эффективное оценивание параметра θ.

Условия существования R-эффективной оценки

Теорема

Пусть эксперимент регулярен; δ – R-эффективная оценка θ. Тогда δ является оценкой максимального правдоподобия.

Доказательство. Поскольку $U(X,\ \theta) = a^*(\theta)(\delta(X) - \theta)$, из несмещенности и R-эффективности следует, что

$$a^*(\theta) = \sqrt{I(\theta)/\mathbb{D}_\theta\delta(X)} = 1/\mathbb{D}_\theta\delta(X) > 0.$$

Тогда $U(X, \delta(X)) = 0$. Следовательно, δ – точка локального максимума функции $L(X;\ \theta)$ по θ. ■

- Если δ – R-эффективная оценка θ, то она является НРМД-оценкой

Многомерный случай

В случае d-мерного параметра $\theta = (\theta_1, \ldots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$, определение регулярного эксперимента вводится аналогично

Определение

Эксперимент регулярный, если при каждом $\theta \in \Theta$:

(i) $L(x; \theta)$ непрерывна и имеет непрерывные частные производные по каждому аргументу θ_s, $s = 1, \ldots, d$;

(ii) допустимо дифференцирование под знаком интеграла по каждому аргументу, и $\mathbb{E}U(X, \theta) = 0$

- $U(X, \theta) = \frac{\partial}{\partial \theta} \log L(X, \theta)$ – градиент (вектор-столбец) прологарифмированной функции правдоподобия

(iii) существует положительно определенная матрица информации Фишера

$$\mathbb{I}(\theta) = \mathbb{E}_\theta U(X; \theta)' U(X; \theta)$$

Многомерный случай

- В предположениях регулярности, элементы информационной матрицы $I_{i,j}(\theta) = \mathbb{E}\left(\frac{\partial}{\partial\theta_i}\log L(X;\theta)\frac{\partial}{\partial\theta_j}\log L(X;\theta)\right)$ – ковариации компонент градиента функции правдоподобия $U(X,\theta)$.
- Если плотности распределения непрерывно дважды дифференцируемы, и допускается дифференцирование плотности распределения под знаком интеграла дважды, то $\mathbb{I}(\theta) = -\mathbb{E}H(X,\theta)$
  - $H(X,\theta) = \|h_{ij}(X,\theta)\|_{ij}$ – матрица Гёссе прологарифмированной функции правдоподобия
  - $h_{ij}(X,\theta) = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L(X,\theta)$, $i,j = 1,\ldots,d$.

Многомерный случай

Для случая многопараметрического семейства может быть получено похожее неравенство. Условия регулярности заключаются в наличии частных производных под знаком интеграла по каждому параметру и невырожденности информационной матрицы каждого наблюдения $\mathbb{I}(\theta) = \|I_{i,j}(\theta)\|$, где

$$I_{i,j}(\theta) = \mathbb{E}\left(\frac{\partial}{\partial\theta_i}\ln L(x;\,\theta)\frac{\partial}{\partial\theta_j}\ln L(x;\,\theta)\right).$$

Тогда для любой разрешенной оценки $\delta = (\delta_1, \ldots, \delta_k)$ параметра θ справедливо неравенство

$$\mathrm{Var}(\delta) \geq (\mathbb{E} + \mathcal{B}'(\theta))\mathbb{I}^{-1}(\theta)(\mathbb{E} + \mathcal{B}'(\theta))^T,$$

где $\mathbb{E}$ – единичная матрица; $\mathcal{B}'(\theta)$ – матрица частных производных компонент вектора смещений по параметрам, и, в частности, если δ несмещенная, то

Примеры

Упражнение

Пусть $X_1, \ldots, X_n$– выборка из нормального распределения $\mathcal{N}(a, \sigma^2)$. Будет ли R-эффективной НРМД оценка $\tilde{\theta} = (\overline{X}, s'^2)$, $s'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$.

Решение. Ранее было показано, что ОМП имеет вид $\hat{\theta} = (\overline{X}, s^2)$. Поскольку $\tilde{\theta} \neq \hat{\theta}$, она не является R-эффективной. ■

- Информационная матрица $\mathbb{I}(a, \sigma^2)$ и $\mathbb{I}(a, \sigma^2)^{-1}$ имеют вид соответственно
$$\begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^4) \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}.$$
- Ковариационная матрица оценки $\tilde{\theta}$ есть
$$\begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/(n-1) \end{pmatrix}.$$
- Поскольку $\tilde{\theta}$ – НРМД-оценка, R-эффективной оценки (для σ^2) не существует.

План

1. Минимаксный и байесовский подходы
2. Информационное неравенство
3. Асимптотическая нормальность оценок максимального правдоподобия

Асимптотическая нормальность ОМП

Теорема (Асимптотическая нормальность ОМП)

Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения, принадлежащего регулярному семейству распределений

(i) выполнено условие дважды дифференцируемости плотности распределения под знаком интеграла

(ii) $\overline{\mathbb{I}}(\theta) = n\,I(\theta)$ – информационная матрица

(iii) оценка максимального правдоподобия $\hat{\theta}$ состоятельна

Тогда

$$n^{1/2}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, I(\theta)^{-1}).$$

- В условиях теоремы, оценка максимального правдоподобия является асимптотически эффективной

Асимптотическая нормальность ОМП

Доказательство. Отметим, что $U(X;\theta)$ представляет собой сумму независимых и одинаково распределенных случайных величин с нулевым средним и матрицей ковариаций

$$\mathbb{I}(\theta) = nI(\theta).$$

С использованием центральной предельной теоремы Леви получаем

$$n^{-1/2}U(X;\theta) \Rightarrow \mathcal{N}(0, I(\theta)).$$

Пусть $\hat{\theta}$ — оценка максимального правдоподобия параметра θ. Используя формулу Тейлора с остаточным членом в форме Лагранжа запишем

$$U(X;\theta) - U(X,\hat{\theta}) = -H(X;\theta^*)(\hat{\theta} - \theta),$$

где θ^* — некоторая точка, лежащая на отрезке, соединяющем точки θ и $\hat{\theta}$. Поскольку $\hat{\theta}$ — оценка максимального правдоподобия,

$$U(X;\hat{\theta}) = 0.$$

Таким образом,

$$n^{-1/2}U(X;\theta) = -n^{-1}H(X;\theta^*)\, n^{1/2}(\hat{\theta} - \theta) = -\overline{H}(X;\theta^*)\, n^{1/2}(\hat{\theta} - \theta).$$

Поскольку элементы матрицы Гессе представляют собой суммы независимых и одинаково распределенных случайных величин, $\mathbb{E}_\theta(H(X;\theta)) = -nI(\theta)$ и наблюдается сходимость по вероятности

$$-n^{-1}H(X;\theta) \to_{P_\theta} I(\theta).$$

Асимптотическая нормальность ОМП

Доказательство асимптотической нормальности ОМП (продолжение).
Предположим, что оценка максимального правдоподобия $\hat{\theta}$ является состоятельной оценкой параметра $\theta \in \Theta$. В силу непрерывности функции $H(X;\theta)$ по θ при каждом фиксированном значении X, с использованием закона больших чисел получаем, что

$$\overline{H}(X;\theta^*) = \overline{H}(X;\theta) + \mathrm{O}_P(1) = -I(\theta) + \mathrm{O}_P(1).$$

Тогда, в силу положительной определенности матрицы $I(\theta)$, с вероятностью, стремящейся к единице, матрица $\overline{H}$ также является положительно-определенной, а следовательно, обратимой. Таким образом,

$$n^{1/2}(\hat{\theta} - \theta) = -\overline{H}(X;\theta)^{-1}\, n^{-1/2} U(X;\theta) = I(\theta)^{-1}\, n^{-1/2} U(X;\theta) + \mathrm{O}_P(1).$$

Следовательно,

$$n^{1/2}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \Gamma),$$

где $\Gamma = I(\theta)^{-1} I(\theta) I(\theta)^{-1} = I(\theta)^{-1}$ ■