Тема 8. Методы получения оценок

8.1. Асимптотическая нормальность

Познакомимся с ещё одним важным свойством статистических оценок, усиливающим понятие состоятельности и характеризующим скорость сходимости оценки к оцениваемой величине.

Определение. Оценка $\hat{\theta}(X_1, \ldots, X_n)$ называется *асимптотически нормальной с асимптотической дисперсией* σ^2, если

$$\mathbf{P}((\hat{\theta}-\theta)\sqrt{n}/\sigma \le x) \to \Phi(x) \quad \text{при } n \to \infty, \tag{1}$$

где $\Phi(x)$ — функция распределения стандартного нормального закона $N(0,1)$ (функция $\Phi(x)$ определяется в формуле (4) из темы 6). Заметим, что сходимость (1) равносильна сходимости

$$\mathbf{P}((\hat{\theta}-\theta)\sqrt{n} \le x) \to \Phi(x/\sigma) \text{ при } n \to \infty,$$

где предельной функцией является функция распределения нормального закона $N(0,\sigma^2)$. Это объясняет название «асимптотическая дисперсия» для величины σ^2. Оценки, обладающие свойством асимптотической нормальности, можно сравнивать на основе их асимптотических дисперсий, характеризующих степень «разброса» оценок для выборки большого размера n.

Например, центральная предельная теорема (ЦПТ) из раздела 6.4 в свете данного определения является утверждением об асимптотической нормальности выборочного среднего $\overline{X}$ как оценки для неизвестного среднего значения $\mu = \mathrm{M}X_1$:

$$\mathbf{P}((\overline{X}-\mu)\sqrt{n}/\sigma \le x) \to \Phi(x) \quad \text{при } n \to \infty, \tag{2}$$

где асимптотической дисперсией σ^2 служит дисперсия отдельного наблюдения $\mathrm{D}X_1$.

Определение. *Относительной асимптотической эффективностью* оценок $\hat{\theta}_1$ и $\hat{\theta}_2$ называется отношение их асимптотических дисперсий $e_{\hat{\theta}_1, \hat{\theta}_2} = \sigma_2^2/\sigma_1^2$.

Почему в качестве показателя относительной эффективности разумно использовать отношение асимптотических дисперсий, а не отношение σ_2/σ_1? Пусть требуется оценить параметр θ с заданной точностью Δ, причём за каждое наблюдение X_i приходится платить цену C. Тогда размеры выборок n_1 и n_2, обеспечивающие заданную точность первой и второй оценок соответственно, определяются из соотношения

$$\Delta = \sigma_1/\sqrt{n_1} = \sigma_2/\sqrt{n_2}.$$

Таким образом, $\sigma_2^2/\sigma_1^2 = n_2/n_1 = (Cn_2)/(Cn_1)$, т. е. относительная асимптотическая эффективность — отношение затрат при использовании второй оценки к затратам при использовании первой.

Примером асимптотически нормальных оценок служат так называемые выборочные моменты.

Определение. *Выборочным моментом порядка* k называется случайная величина $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.

Определение. *Теоретическим моментом порядка* k называется число $\alpha_k = \mathrm{M}X^k$.

Покажем, что выборочные моменты A_k служат асимптотически нормальными оценками для соответствующих теоретических моментов α_k. Действительно, применяя ЦПТ к независимым и одинаково распределённым случайным величинам $Y_i = X_i^k$, устанавливаем сходимость

$$\mathbf{P}((A_k - \alpha_k)\sqrt{n}/\sigma \le x) \to \Phi(x) \quad \text{при } n \to \infty, \tag{3}$$

где асимптотическая дисперсия $\sigma^2 = \mathsf{D}Y_1 = \mathsf{M}Y_1^2 - (\mathsf{M}Y_1)^2 = \mathsf{M}X_1^{2k} - (\mathsf{M}X_1^k)^2 = \alpha_{2k} - \alpha_k^2$.

Замечание. Существуют состоятельные оценки, не обладающие свойством асимптотической нормальности (1). Например, в задаче 8.6 для оценки $\hat{\theta}_1 = \max\{X_1, \ldots, X_n\}$ в модели равномерного распределения на отрезке $[0, \theta]$ предлагается для произвольного числа $x > 0$ установить сходимость

$$\mathbf{P}(n(\theta - \hat{\theta}_1) \le x) \to 1 - e^{-x/\theta} \quad \text{при } n \to \infty.$$

Таким образом, погрешность оценки $\hat{\theta}_1$ имеет порядок малости $const/n$, и предельным законом служит не нормальное, а экспоненциальное распределение.

8.2. Медиана, квартили, процентили

Наряду с математическим ожиданием практический интерес представляют другие характеристики распределения признака: медиана, квартили, p-квантили. Определим эти понятия. Ради простоты рассмотрим непрерывную и строго возрастающую функцию распределения $F_X(x)$.

Определение. Значение аргумента функции $F_X(x)$, при котором она достигает уровня $0 < p < 1$, называется *p-квантилью* [1] и обозначается через x_p (рис. 1). Короче говоря, $x_p = F_X^{-1}(p)$.

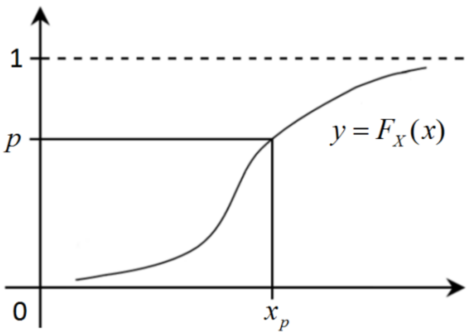

Рис. 1

Определение. *Нижней квартилью*, *медианой* и *верхней квартилью* называются частные случаи p-квантилей при $p = 1/4$, 1/2, 3/4 соответственно.

Медиана, наряду с математическим ожиданием $\mathsf{M}X$, является характеристикой «центра» распределения. У перекошенного распределения, имеющего сильно асимметричную плотность, медиана заметно отличается от математического ожидания.

Согласно данным Росстата средняя зарплата в 2016 году в Москве составила 59 тыс. руб., а медианная зарплата — 44 тыс. руб. (74% от средней). Во всей Российской Федерации аналогичные показатели составляли, соответственно, 31 тыс. руб. и 23 тыс. руб. (75% от средней).[2] Отметим, что в 2013 году отношение медианной и средней зарплат составляло 66% в Москве и 74% в России, так что за 3 года в Москве произошло уменьшение асимметрии распределения до общероссийского уровня.

---

[1] Также p-квантиль называют $(100 \cdot p)$-й *процентилью* или *персентилью*.

[2] http://www.gks.ru/free_doc/new_site/population/bednost/tabl/tab-bed1-2-6.htm

Для каких признаков медиана представляет интерес? Прежде всего, для неаддитивных (от additio — (лат.) *добавление*), т. е. таких признаков, для которых не имеет смысла суммирование значений. Бессмысленно складывать, скажем, следующие показатели: продолжительность жизни человека, нагрузка разрушения при испытании некоторого материала на прочность, температура больного. В противном случае появляется такая анекдотическая характеристика, как «средняя температура по больнице, включая морг».

Напротив, аддитивными признаками являются, например, надой молока от коровы, запас древесины у заготавливаемого дерева, ежемесячный доход гражданина.

В свою очередь, *межквартильный интервал* $(x_{1/4}, x_{3/4})$ на практике нередко служит альтернативой для области типичных значений $(\mathrm{M}X - \sqrt{\mathrm{D}X},\ \mathrm{M}X + \sqrt{\mathrm{D}X})$ (см. рис. 9 в теме 4). Действительно, $\mathbf{P}(x_{1/4} < X < x_{3/4}) = 1/2$, т. е. вероятность попадания значений признака X в интервал $(x_{1/4}, x_{3/4})$ равна 1/2, поэтому примерно 50% наблюдений оказывается внутри этого интервала.

Децильный коэффициент. Как пример социологического показателя, строящегося на основе квантилей $x_{0,1}$ и $x_{0,9}$ распределения дохода граждан, рассмотрим *децильный коэффициент* — отношение доходов 10% самых богатых граждан страны к доходам 10% самых бедных.[3] Этот коэффициент служит важным индикатором экономической стабильности государства.

По данным Росстата в 2017 году децильный коэффициент в России был равен 15,5.[4] Для сравнения приведём его значения, упорядоченные по возрастанию, в ряде стран за 2012-2015 годы.[5]

Название страны	Доля дохода 10% богатых	Доля дохода 10% бедных	Децильный коэффициент	Год
Казахстан	22,3	4,3	5,2	2015
Белоруссия	22,2	4,0	5,6	2015
Финляндия	22,0	3,9	5,6	2014
Словения	21,1	3,8	5,6	2014
Чехия	22,1	3,9	5,7	2014
Норвегия	21,6	3,5	6,2	2014
Германия	24,9	3,3	7,5	2013
Великобритания	26,2	2,9	9,0	2014
Китай	31,4	2,0	15,7	2012
США	30,2	1,7	17,8	2013
Гондурас	37,8	1,2	31,5	2015
Колумбия	39,6	1,2	33,0	2015
Бразилия	40,5	1,1	36,8	2015
Замбия	45,2	1,0	45,2	2015

«Как только децильный коэффициент достигает 10, в стране создаются условия для социальных беспорядков, — пояснил "Известиям" в 2007 году глава Института экономики РАН Руслан Гринберг. — Это правило не действует разве что в Америке, где коэффициент держится на уровне 10-12.[6] Но там это считается нормальным, поскольку философия американцев отличается от нашей. Там считается: если ты бедный, то сам виноват».[7]

---

[3] Иногда его называют *коэффициентом фондов*, а децильным коэффициентом — отношение $x_{0,9}/x_{0,1}$.

[4] http://ac.gov.ru/events/016271.html

[5] По результатам опросов Всемирного банка, см. http://data.worldbank.org/indicator, раздел «Poverty»

[6] По данным Всемирного банка децильный коэффициент в США составлял 14,8 в 1991 г. и 17,8 в 2013 г.

[7] https://archive.is/20120907172509/www.izvestia.ru/economic/article3107047/

Каким образом можно оценить p-квантиль x_p (в частности, медиану и квартили) по выборке? В Excel и R используется оценка $\hat{x}_p$, определяемая следующим образом. Упорядочим элементы выборки $X_1, \ldots, X_n$ по возрастанию значений и получим так называемый *вариационный ряд*:

$$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}.$$

В частности, $X_{(1)}$ — минимум выборки, $X_{(n)}$ — максимум выборки. Отметим на плоскости точки с координатами $(X_{(i)}, (i-1)/(n-1))$, где $i = 1, \ldots, n$. Последовательно соединив их отрезками, получим ломаную, задающую график функции $\hat{F}(x)$ (рис. 2). Для произвольного числа p из интервала (0, 1) положим $\hat{x}_p = \hat{F}^{-1}(p)$.

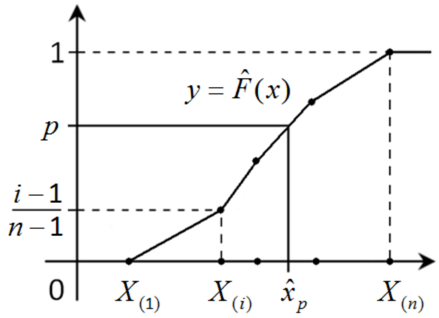

Рис. 2

В частности, оценка $\hat{x}_{1/2}$ называется *выборочной медианой*. Её также обозначают через MED (от medio — (лат.) *середина*). Она явно выражается через члены вариационного ряда:

$$MED = \begin{cases} X_{((n+1)/2)} & \text{при нечётном } n, \\ (X_{(n/2)} + X_{(n/2+1)})/2 & \text{при чётном } n. \end{cases}$$

Возвращаясь к случаю произвольного $0 < p < 1$, приведём условия, обеспечивающие асимптотическую нормальность оценки $\hat{x}_p$.

Теорема. Если существует плотность распределения $f_X(x)$ и $f_X(p) > 0$, то

$$\mathbf{P}((\hat{x}_p - x_p)\sqrt{n}/\sigma \le x) \to \Phi(x) \quad \text{при } n \to \infty, \tag{4}$$

где асимптотическая дисперсия $\sigma^2 = p(1-p)/f_X^2(p)$.

Индекс массы тела. В антропологии в качестве показателя степени ожирения (похудения) человека используется *индекс массы тела* (ИМТ):

$$\textbf{ИМТ} = 10^6 \cdot \textbf{ВЕС} / \textbf{РОСТ}^2.$$

Здесь предполагается, что ВЕС измеряется в килограммах, а РОСТ — в сантиметрах. На рис. 3 приведены оценки процентилей уровней 10%, 25%, 50%, 75%, 90% индекса массы тела мальчиков возраста от 7 до 17 лет, проживавших в 2011 году в Москве (на основе выборки из 500 человек). Для визуального уменьшения влияния случайности процентили сглажены кубическими многочленами. Норму веса задает полоса между 25% и 75% уровнями. Если ИМТ мальчика превышает процентиль уровня 75%, то его вес является избыточным. Превышение же процентили уровня 90% свидетельствует об ожирении, представляющем опасность для здоровья.

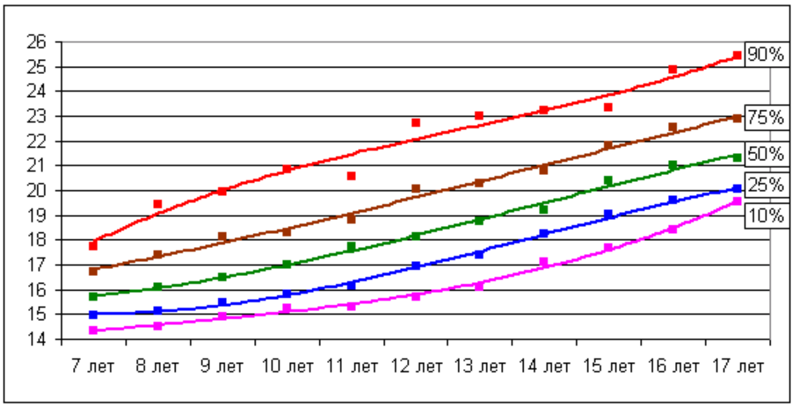

Рис. 3

8.3. Метод моментов

Излагаемый ниже метод моментов является одним из универсальных способов получения состоятельных оценок параметров в многопараметрических статистических моделях.

Согласно формуле (3) выборочные моменты A_k служат асимптотически нормальными (следовательно, состоятельными) оценками для теоретических моментов α_k. Поэтому для выборки большого размера n будут выполняться приближённые равенства

$$A_k \approx \alpha_k, \ k = 1, 2, \ldots .$$

На этом соображении и основан метод моментов.

Пусть статистическая модель задана семейством Ψ функций распределения $F(x, \theta_1, \ldots, \theta_m)$, которые зависят от m параметров $\theta_1, \ldots, \theta_m$. Тогда теоретические моменты α_k также будут зависеть от этих параметров. Составим систему из m уравнений (вообще говоря, нелинейных)

$$\begin{cases} \alpha_1(\theta_1, \ldots, \theta_m) = A_1 \\ \qquad\qquad \ldots \\ \alpha_m(\theta_1, \ldots, \theta_m) = A_m \end{cases}$$

Отметим, что в правой части системы стоят случайные величины A_k. Поэтому решение системы (если, конечно, оно существует) является случайным вектором $(\hat{\theta}_1, \ldots, \hat{\theta}_m)$. Компоненты этого вектора и называются *оценками по методу моментов* неизвестных параметров $\theta_1, \ldots, \theta_m$.

Пример 1. Рассмотрим модель сдвига экспоненциального закона с плотностью распределения $f(x, \mu) = e^{-(x-\mu)}) I_{\{x>\mu\}}$ (рис. 4).

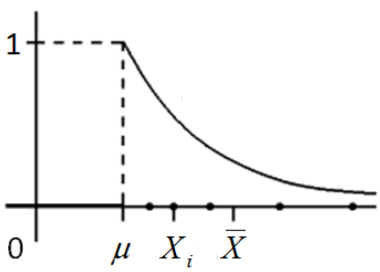

Рис. 4

Здесь $m=1$, $\theta_1=\mu$. Система состоит из единственного уравнения $\alpha_1(\mu)=A_1=\overline{X}\equiv(X_1+\ldots+X_n)/n$. Интегрируя по частям, нетрудно подсчитать, что

$$\alpha_1(\mu)=\mathrm{M}X=\mu+\int_0^{\infty}xe^{-x}dx=\mu+1.$$

Отсюда находим, что оценкой по методу моментов в этой модели является $\hat{\mu}=\overline{X}-1$.

8.4. Метод максимального правдоподобия

Следует поставить перед собой цель изыскать способ
решения всех задач одним и притом простым методом.

Ж. Даламбер

Одним из наиболее часто используемых общих методов получения оценок является метод максимального правдоподобия. Ниже будет показано, что оценки, полученные этим методом, в определённом смысле являются самыми точными.

Сначала предположим для простоты, что элементы выборки $X_1,\ldots,X_n$ имеют дискретное распределение, зависящее от скалярного параметра θ. Введём обозначение

$$f(x,\theta)\equiv\mathbf{P}(X=x).$$

Ввиду независимости элементов выборки совместная вероятность представляется в виде

$$\mathbf{P}(X_1=x_1,\ \ldots,X_n=x_n)=\prod_{i=1}^{n}\mathbf{P}(X_i=x_i)=\prod_{i=1}^{n}f(x_i,\theta). \tag{5}$$

В правой части формулы (5) стоит функция от аргументов $x_1,\ldots,x_n$ и θ. Рассматриваемая как функция от единственного аргумента θ при фиксированных значениях аргументов $x_1,\ldots,x_n$, она называется ***функцией правдоподобия*** и обозначается через $L(\theta)$.[8]

Значение $L(\theta_0)$ в фиксированной точке θ_0 можно считать мерой правдоподобия параметра θ_0 для имеющейся реализации выборки $x_1,\ldots,x_n$. Представляется разумным в качестве оценки для неизвестного параметра θ взять наиболее правдоподобное значение $\tilde{\theta}=\tilde{\theta}(x_1,\ldots,x_n)$, которое получается при максимизации функции $L(\theta)$ по переменной θ. Обычно проще найти точку максимума функции $\ln L(\theta)$, совпадающую с $\tilde{\theta}$ ввиду монотонности логарифма.

В 1778 г. швейцарским физиком, механиком и математиком Даниилом Бернулли (племянником Якоба Бернулли, установившего в 1713 году справедливость закона больших чисел для частоты «успехов» в независимых испытаниях) была опубликована в изданиях Петербургской Академии наук работа «Наиболее вероятное определение по нескольким расходящимся между собой наблюдениям и устанавливаемое отсюда наиболее правдоподобное заключение», где впервые был высказан и использован для оценки неизвестного параметра принцип максимального правдоподобия.[9]

Определение. Оценка $\tilde{\theta}(X_1,\ldots,X_n)$ называется *оценкой максимального правдоподобия* (ОМП).

Обратите внимание на то, что в качестве аргументов в функцию $\tilde{\theta}(x_1,\ldots,x_n)$ подставляются элементы выборки $X_1,\ldots,X_n$. Поэтому ОМП — случайная величина.

---

[8] Likelihood — (англ.) *правдоподобие*.

[9] Широкое распространение метод получил после публикации статьи Рональда Фишера в 1912 году.

Пример 2. Рассмотрим модель испытаний Бернулли с неизвестной вероятностью «успеха» θ (см. пример 4 в теме 6). Для отдельного испытания X_i при $x = 0$ или 1 верно представление

$$f(x, \theta) = \mathbf{P}(X = x) = \theta^x (1-\theta)^{1-x}.$$

Тогда из формулы (5) выводим, что $L(\theta) = \theta^s (1-\theta)^{n-s}$, где $s = x_1 + \ldots + x_n$. Таким образом, функция правдоподобия представляет собой многочлен n-й степени от переменной θ. График функции $L(\theta)$ для случая $s > 1$ и $n - s > 1$ имеет вид, изображённый на рис. 5.

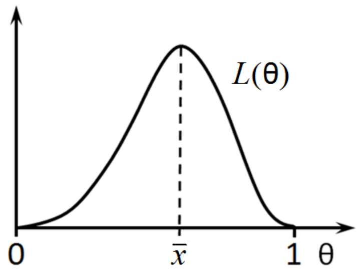

Рис. 5

Удобнее искать точку максимума функции не функции $L(\theta)$, а её логарифма

$$\ln L(\theta) = s \ln\theta + (n-s)\ln(1-\theta).$$

Продифференцировав по θ, получим уравнение $s/\theta - (n-s)/(1-\theta) = 0$. Решив его, найдём корень $\widetilde{\theta} = s/n = \bar{x}$. Таким образом, ОМП для неизвестной вероятности «успеха» θ есть не что иное, как частота «успехов» $\overline{X}$.

В непрерывном случае будем использовать обозначение $f(x, \theta)$ для плотности распределения отдельного элемента X_i выборки $X_1, \ldots, X_n$. Функцию правдоподобия $L(\theta)$ определим как правую часть формулы (5).

Пример 3. Рассмотрим модель сдвига показательного закона с плотностью $f(x, \theta) = e^{-(x-\theta)} I_{\{x \geq \theta\}}$. В этой модели функция правдоподобия имеет вид

$$L(\theta) = \prod_{i=1}^{n} e^{-(x_i - \theta)} I_{\{x_i \geq \theta\}} = e^{-\sum x_i} e^{n\theta} \prod_{i=1}^{n} I_{\{x_i \geq \theta\}} = e^{-\sum x_i} e^{n\theta} I_{\{x_{(1)} \geq \theta\}}, \tag{6}$$

где $x_{(1)} = \min\{x_1, \ldots, x_n\}$. Функция в правой части формулы (6) экспоненциально быстро растёт по переменной θ до значения $x_{(1)}$, после которого становится равной нулю из-за индикатора $I_{\{x_{(1)} \geq \theta\}}$ (рис. 6). Из-за разрыва при $\theta = x_{(1)}$ правдоподобие $L(\theta)$ не является дифференцируемой функцией, и ОМП нельзя находить путём приравнивания нулю производной функции $L(\theta)$ по переменной θ. Из графика на рис. 6 заключаем, что ОМП для параметра сдвига θ служит $\widetilde{\theta} = X_{(1)} = \min\{X_1, \ldots, X_n\}$.

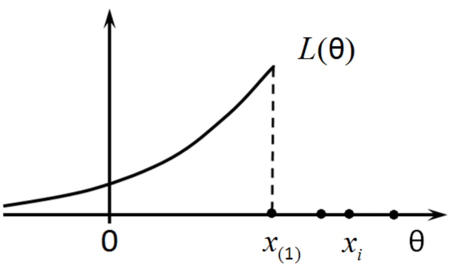

Рис. 6

Метод максимального правдоподобия легко обобщается на случай моделей, зависящих от m параметров $\theta_1, \ldots, \theta_m$. В случае, когда правдоподобие $L(\theta_1, \ldots, \theta_m)$ является дифференцируемой функцией от аргументов $\theta_1, \ldots, \theta_m$, оценки максимального правдоподобия $\tilde{\theta}_1, \ldots, \tilde{\theta}_m$ находятся путём решения системы из m (вообще говоря, нелинейных) уравнений:

$$\frac{\partial L(\theta_1, \ldots, \theta_m)}{\partial \theta_j} = 0, \quad j = 1, \ldots, m. \tag{7}$$

Иногда решение системы можно найти явно (задача 8.7), но обычно приходится вычислять его приближённо с помощью итерационных методов (например, метода Ньютона).

Какими свойствами обладают оценки максимального правдоподобия? Ради простоты ограничимся случаем скалярного параметра.[10] Оказывается, что ОМП являются асимптотически нормальными при выполнении для семейства распределений $\Psi = \{F(x, \theta)\}$ определённых условий гладкости и интегрируемости (так называемых *условий регулярности*).[11]

Теорема 1. В условиях регулярности для достаточно большого размера выборки n с вероятностью 1 существует решение уравнения правдоподобия $L'(\theta) = 0$, дающее состоятельную оценку $\tilde{\theta}$, причём

$$\mathbf{P}((\tilde{\theta} - \theta)\sqrt{n}/\sigma \le x) \to \Phi(x) \quad \text{при } n \to \infty, \tag{8}$$

где асимптотическая дисперсия задаётся формулой $\sigma^2 = 1/I(\theta)$, в которой величина

$$I(\theta) = \mathrm{D}\frac{\partial \ln f(X_1, \theta)}{\partial \theta} \tag{9}$$

называется *информацией Фишера*.

Теорема 2. Пусть выполнены условия регулярности, $\hat{\theta}$ — любая асимптотически нормальная оценка с асимптотической дисперсией σ^2, непрерывно зависящей от параметра θ. Тогда $\sigma^2(\theta) \ge 1/I(\theta)$.

Таким образом, при выполнении для модели условий регулярности ОМП обладает наименьшей возможной асимптотической дисперсией $1/I(\theta)$. Это важнейшее свойство ОМП называется *асимптотической эффективностью*.

---

[10] Векторный случай рассматривается, например, в книге Леман Э. «Теория точечного оценивания», с. 379.
[11] Один из возможных вариантов условий регулярности приведён в указанной выше книге Лемана на с. 361.