

Тема 7. Достаточные статистики и эффективные оценки.

1. Среднеквадратический подход к сравнению точечных оценок.

Пусть совокупность наблюдаемых величин (ξ_1, \dots, ξ_n) является выборкой из распределения $F_\xi(x|\theta)$, где θ – скалярный параметр, который принимает значения из множества допустимых значений Θ .

В общем случае для определения неизвестного значения параметра θ могут быть предложены несколько различных оценок, и в этом случае неизбежно возникает ряд вопросов о сравнении оценок и выборе одной наилучшей оценки.

Один из подходов к сравнению оценок заключается в вычислении среднеквадратических отклонений.

Определение 7.1.

Оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ лучше оценки $\tilde{\theta}(\xi_1, \dots, \xi_n)$ в смысле среднеквадратического подхода, если:

$$1) \forall \theta \in \Theta : M_\theta[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)^2] \leq M_\theta[(\tilde{\theta}(\xi_1, \dots, \xi_n) - \theta)^2],$$

$$2) \exists \theta \in \Theta : M_\theta[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)^2] < M_\theta[(\tilde{\theta}(\xi_1, \dots, \xi_n) - \theta)^2].$$

Определение 7.1 означает, что оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ равномерно лучше оценки $\tilde{\theta}(\xi_1, \dots, \xi_n)$, то есть $\hat{\theta}(\xi_1, \dots, \xi_n)$ не хуже оценки $\tilde{\theta}(\xi_1, \dots, \xi_n)$ для всех допустимых значений из множества Θ и хотя бы при одном значении параметра лучше.

В соответствии с определением 7.1 оценкам $\hat{\theta}(\xi_1, \dots, \xi_n)$ и $\tilde{\theta}(\xi_1, \dots, \xi_n)$ сопоставляются функции $\hat{d}(\theta)$ и $\tilde{d}(\theta)$:

$$\hat{d}(\theta) = M_\theta[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)^2],$$

$$\tilde{d}(\theta) = M_\theta[(\tilde{\theta}(\xi_1, \dots, \xi_n) - \theta)^2],$$

которые не всегда оказываются сравнимыми, поскольку вполне возможно найдутся два различных значения $\theta_1 \in \Theta$ и $\theta_2 \in \Theta$ таких, что:

$$\hat{d}(\theta_1) < \tilde{d}(\theta_2),$$

$$\hat{d}(\theta_1) > \tilde{d}(\theta_2),$$

и в этом случае из двух оценок $\hat{\theta}(\xi_1, \dots, \xi_n)$ и $\tilde{\theta}(\xi_1, \dots, \xi_n)$ не возможно выбрать наилучшую в среднеквадратическом смысле. Таким образом, отношение «лучше в среднеквадратическом смысле» является лишь отношением частичного порядка на множестве всех возможных оценок.

Кроме того, легко видеть, что среди всех возможных оценок может вообще не существовать наилучшей в среднеквадратическом смысле оценки. Действительно, предположим, что оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является наилучшей в среднеквадратическом смысле среди всех возможных оценок параметра θ . Какова функция $\hat{d}(\theta)$ такой наилучшей оценки?

Рассмотрим произвольное допустимое значение параметра $\theta_1 \in \Theta$ и «вырожденную» оценку $\tilde{\theta}(\xi_1, \dots, \xi_n) = \theta_1$. Оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является наилучшей, поэтому согласно определению 7.1 для любой оценки, в том числе и для $\tilde{\theta}(\xi_1, \dots, \xi_n)$:

$$\forall \theta \in \Theta : M_\theta[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)^2] \leq M_\theta[(\tilde{\theta}(\xi_1, \dots, \xi_n) - \theta)^2].$$

Откуда для $\theta = \theta_1$:

$$\hat{d}(\theta_1) = M_{\theta_1}[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta_1)^2] \leq M_{\theta_1}[(\tilde{\theta}(\xi_1, \dots, \xi_n) - \theta_1)^2] = M_{\theta_1}[(\theta_1 - \theta_1)^2] = 0,$$

$$\hat{d}(\theta_1) = 0.$$

Поскольку значение θ_1 было выбрано произвольно, то для любого допустимого значения параметра $\theta_1 \in \Theta$ значение $\tilde{d}(\theta_1) = 0$. Следовательно, функция $\tilde{d}(\theta)$ тождественно равна нулю на множестве Θ :

$$\begin{aligned}\forall \theta \in \Theta : \hat{d}(\theta) &= 0, \\ \forall \theta \in \Theta : M_{\theta}[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)^2] &= 0.\end{aligned}$$

Последнее равенство возможно только в том случае, когда:

$$\forall \theta \in \Theta : \hat{\theta}(\xi_1, \dots, \xi_n) = \theta.$$

Последнее равенство означает, что наилучшая оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ всегда точно определяет значение параметра θ по выборке (ξ_1, \dots, ξ_n) из распределения $F_{\xi}(x | \theta)$.

В частных случаях это возможно, например, для случая выборки из распределения $R[\theta, \theta + 0.5]$, где параметр θ принимает значения из множества натуральных чисел (в этом случае $\hat{\theta}(\xi_1, \dots, \xi_n) = [\xi_1]$, где $[\cdot]$ означает целую часть, будет наилучшей оценкой). Однако указанный случай и другие подобные ему являются тривиальными, а в общем случае на существование такой наилучшей оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ рассчитывать не приходится.

Поскольку в классе всех возможных оценок наилучшей оценки может не существовать, приходится выделять более узкие классы оценок, например, на основе смещений оценок.

Определение 7.2.

Смещением оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ называется функция $b(\theta)$:

$$b(\theta) = M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)] - \theta.$$

Согласно известному соотношению:

$$\begin{aligned}M_{\theta}[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)^2] &= D_{\theta}[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)] + (M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n) - \theta])^2 = \\ &= D_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)] + b^2(\theta)\end{aligned}\quad (7.1)$$

поэтому в рамках среднеквадратического подхода сравнение оценок из класса оценок K_b , имеющих заданное смещение $b(\theta)$:

$$K_b = \{\hat{\theta}(\xi_1, \dots, \xi_n) : M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)] - \theta = b(\theta)\},$$

эквивалентно сравнению дисперсий оценок $D_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)]$, поскольку в (7.1) смещение $b(\theta)$ оказывается одинаковым для всех оценок.

Определение 7.3.

Оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ называется *эффективной в классе K_b* , если для любой оценки $\tilde{\theta}(\xi_1, \dots, \xi_n) \in K_b$:

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}(\xi_1, \dots, \xi_n) - \theta)^2] \leq M_{\theta}[(\tilde{\theta}(\xi_1, \dots, \xi_n) - \theta)^2].$$

Из всех классов оценок K_b , как правило, особо выделяют класс оценок K_0 , имеющих нулевое смещение $b(\theta) = 0$, $\theta \in \Theta$ (класс несмещенных оценок).

Определение 7.4.

Эффективная оценка в классе K_0 называется *эффективной*.

Всегда ли существуют оценка, имеющая заданное смещение $b(\theta)$? Ответ на этот вопрос отрицательный. Согласно определению (7.1) если оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ имеет смещение $b(\theta)$ то:

$$\begin{aligned}M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)] &= \theta + b(\theta), \\ \int_{R^n} \hat{\theta}(x_1, \dots, x_n) dF_{\xi}(x_1, \dots, x_n) &= \theta + b(\theta)\end{aligned}\quad (7.2)$$

где $F_{\xi}(x_1, \dots, x_n) = \prod_{i=1}^n F_{\xi}(x_i)$ – функция распределения выборки (ξ_1, \dots, ξ_n) . Таким образом,

функция $\hat{\theta}(x_1, \dots, x_n)$ является решением интегрального уравнения (7.2), которое в некоторых случаях не всегда имеет решение для заданной функции $b(\theta)$. Отсюда следует, что и существование несмещенных оценок представляет собой вопрос, который в частных случаях решается по-разному (в некоторых случаях несмещенных оценок не существует).

Существование эффективной оценки в классе K_b также представляет собой вопрос, не имеющий единого ответа для всех случаев, тем не менее, всегда можно быть уверенным в том, что в каждом классе K_b не может существовать двух различных эффективных оценок.

Утверждение 7.5.

Если в классе K_b существует эффективная оценка, то она единственна.

Доказательство:

Докажем утверждение от противного: предположим, что существуют две различные эффективные оценки $\hat{\theta}_1(\xi_1, \dots, \xi_n) \in K_b$ и $\hat{\theta}_2(\xi_1, \dots, \xi_n) \in K_b$.

1) Поскольку оценка $\hat{\theta}_1(\xi_1, \dots, \xi_n)$ эффективная в K_b , то согласно определению 7.3:

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_1 - \theta)^2] \leq M_{\theta}[(\hat{\theta}_2 - \theta)^2]. \quad (7.3)$$

Аналогично, в силу того, что $\hat{\theta}_2(\xi_1, \dots, \xi_n)$ также эффективная оценка в K_b :

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_1 - \theta)^2] \geq M_{\theta}[(\hat{\theta}_2 - \theta)^2] \quad (7.4)$$

Из неравенств (7.3) и (7.4) следует равенство среднеквадратических отклонений:

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_1 - \theta)^2] = M_{\theta}[(\hat{\theta}_2 - \theta)^2] \quad (7.5)$$

2) Рассмотрим оценку $\tilde{\theta}(\xi_1, \dots, \xi_n)$:

$$\tilde{\theta}(\xi_1, \dots, \xi_n) = \frac{\hat{\theta}_1(\xi_1, \dots, \xi_n) + \hat{\theta}_2(\xi_1, \dots, \xi_n)}{2}.$$

Легко видеть, что оценка $\tilde{\theta}(\xi_1, \dots, \xi_n) \in K_b$, действительно:

$$\begin{aligned} M_{\theta}[\tilde{\theta}] - \theta &= M_{\theta}\left[\frac{\hat{\theta}_1 + \hat{\theta}_2}{2}\right] - \theta = \frac{M_{\theta}[\hat{\theta}_1] + M_{\theta}[\hat{\theta}_2]}{2} - \theta = \\ &= \frac{\theta + b(\theta) + \theta + b(\theta)}{2} - \theta = \theta + b(\theta) - \theta = b(\theta). \end{aligned}$$

3) Вычислим среднеквадратическое отклонение оценки $\tilde{\theta}(\xi_1, \dots, \xi_n)$:

$$\begin{aligned} M_{\theta}[(\tilde{\theta} - \theta)^2] &= M_{\theta}\left[\left(\frac{\hat{\theta}_1 + \hat{\theta}_2}{2} - \theta\right)^2\right] = M_{\theta}\left[\left(\frac{\hat{\theta}_1 - \theta + \hat{\theta}_2 - \theta}{2}\right)^2\right] = \\ &= \frac{1}{4} M_{\theta}[(\hat{\theta}_1 - \theta + \hat{\theta}_2 - \theta)^2] = \frac{1}{4} M_{\theta}[(\hat{\theta}_1 - \theta)^2 + 2(\hat{\theta}_1 - \theta)(\hat{\theta}_2 - \theta) + (\hat{\theta}_2 - \theta)^2] = \\ &= \frac{1}{4} M_{\theta}[2(\hat{\theta}_1 - \theta)^2 - (\hat{\theta}_1 - \theta)^2 + 2(\hat{\theta}_1 - \theta)(\hat{\theta}_2 - \theta) + 2(\hat{\theta}_2 - \theta)^2 - (\hat{\theta}_2 - \theta)^2] = \\ &= \frac{1}{4} M_{\theta}[2(\hat{\theta}_1 - \theta)^2 + 2(\hat{\theta}_2 - \theta)^2 - (\hat{\theta}_1 - \theta)^2 + 2(\hat{\theta}_1 - \theta)(\hat{\theta}_2 - \theta) - (\hat{\theta}_2 - \theta)^2] = \\ &= \frac{1}{4} M_{\theta}[2(\hat{\theta}_1 - \theta)^2 + 2(\hat{\theta}_2 - \theta)^2 - ((\hat{\theta}_1 - \theta) - (\hat{\theta}_2 - \theta))^2] = \\ &= \frac{1}{4} M_{\theta}[2(\hat{\theta}_1 - \theta)^2 + 2(\hat{\theta}_2 - \theta)^2 - (\hat{\theta}_1 - \hat{\theta}_2)^2] = \end{aligned}$$

$$= \frac{1}{2} M_{\theta}[(\hat{\theta}_1 - \theta)^2] + \frac{1}{2} M_{\theta}[(\hat{\theta}_2 - \theta)^2] - \frac{1}{4} M_{\theta}[(\hat{\theta}_1 - \hat{\theta}_2)^2].$$

Таким образом,

$$M_{\theta}[(\tilde{\theta} - \theta)^2] = \frac{1}{2} M_{\theta}[(\hat{\theta}_1 - \theta)^2] + \frac{1}{2} M_{\theta}[(\hat{\theta}_2 - \theta)^2] - \frac{1}{4} M_{\theta}[(\hat{\theta}_1 - \hat{\theta}_2)^2].$$

Откуда в силу (7.5):

$$\begin{aligned} M_{\theta}[(\tilde{\theta} - \theta)^2] &= M_{\theta}[(\hat{\theta}_1 - \theta)^2] - \frac{1}{4} M_{\theta}[(\hat{\theta}_1 - \hat{\theta}_2)^2], \\ M_{\theta}[(\hat{\theta}_1 - \hat{\theta}_2)^2] &= 4(M_{\theta}[(\hat{\theta}_1 - \theta)^2] - M_{\theta}[(\tilde{\theta} - \theta)^2]) \end{aligned} \quad (7.6)$$

Оценка $\hat{\theta}_1(\xi_1, \dots, \xi_n)$ по предположению является эффективной в K_b , и в пункте 2 было доказано, что $\tilde{\theta}(\xi_1, \dots, \xi_n) \in K_b$, тогда из определения (7.3):

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_1 - \theta)^2] \leq M_{\theta}[(\tilde{\theta} - \theta)^2],$$

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_1 - \theta)^2] - M_{\theta}[(\tilde{\theta} - \theta)^2] \leq 0.$$

Таким образом, из равенства (7.6):

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_1 - \hat{\theta}_2)^2] = 4(M_{\theta}[(\hat{\theta}_1 - \theta)^2] - M_{\theta}[(\tilde{\theta} - \theta)^2]) \leq 0,$$

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_1 - \hat{\theta}_2)^2] \leq 0$$

Из последнего неравенства следует, что $\hat{\theta}_1(\xi_1, \dots, \xi_n) = \hat{\theta}_2(\xi_1, \dots, \xi_n)$, что противоречит исходному утверждению о том, что оценки $\hat{\theta}_1(\xi_1, \dots, \xi_n)$ и $\hat{\theta}_2(\xi_1, \dots, \xi_n)$ являются различными.

Утверждение доказано.

2. Достаточные статистики.

Пусть, как и ранее, совокупность наблюдаемых величин (ξ_1, \dots, ξ_n) является выборкой из распределения $F_{\xi}(x | \theta)$, где θ – скалярный параметр, который принимает значения из множества допустимых значений Θ .

Поскольку распределение величин (ξ_1, \dots, ξ_n) зависит от параметра θ , то значение параметра определяет вероятности появления реализаций величин (ξ_1, \dots, ξ_n) в эксперименте. Отсюда следует, что на основе появления той или иной реализации можно сделать вывод о неизвестном значении параметра.

Аналогичные рассуждения можно применить и к любой статистике $S(\xi_1, \dots, \xi_n) = (S_1(\xi_1, \dots, \xi_n), \dots, S_k(\xi_1, \dots, \xi_n))$ (в общем случае многомерной), распределение которой в некоторой степени зависит от параметра: значения статистики $S(\xi_1, \dots, \xi_n)$ также позволяют делать выводы о неизвестном значении параметра.

Класс всевозможных статистик чрезвычайно богат и разнообразен, поскольку включает в себя огромное количество статистик с самыми разными свойствами. Например, нетрудно представить себе статистику, распределение которой вовсе не зависит от параметра. На основе значений такой статистики невозможно сделать никакого содержательного вывода о неизвестном значении параметра. Напротив, следует ожидать существование статистик, распределения которых таким образом зависят от параметра, что значения статистик сообщают столько же информации о параметре, сколько и реализации исходных наблюдаемых величин (ξ_1, \dots, ξ_n) . Такие статистики «вбирают в себя» полностью всю информацию о параметре, содержащуюся в совокупности величин (ξ_1, \dots, ξ_n) , и потому называются статистиками достаточными для параметра.

Определение 7.6.

Статистика $S(\xi_1, \dots, \xi_n) = (S_1(\xi_1, \dots, \xi_n), \dots, S_k(\xi_1, \dots, \xi_n))$ называется *достаточной* для

параметра θ , если условное распределение $L_\theta(x_1, \dots, x_n | S)$ величин (ξ_1, \dots, ξ_n) относительно статистики $S(\xi_1, \dots, \xi_n)$ не зависит от параметра θ .

Достаточные статистики имеют весьма широкое применение в самых различных задачах. В частности, если размерность достаточной статистики k меньше количества наблюдаемых случайных величин n , то достаточная статистика может быть использована для «сжатия» информации о параметре без потерь в статистическом смысле: достаточно по заданной реализации (x_1, \dots, x_n) наблюдаемых величин (ξ_1, \dots, ξ_n) вычислить значение достаточной статистики $(s_1, \dots, s_k) = (S_1(x_1, \dots, x_n), \dots, S_k(x_1, \dots, x_n))$ после чего реализацию (x_1, \dots, x_n) можно отбросить. Полученное значение достаточной статистики (s_1, \dots, s_k) при условии $k < n$ требует меньший объем памяти при хранении и меньшее время при передаче по каналу связи по сравнению с исходной реализацией (x_1, \dots, x_n) .

Использование значения (s_1, \dots, s_k) в статистических процедурах может быть организовано не менее эффективно, чем использование исходной реализации (x_1, \dots, x_n) : для всякой статистической процедуры $\delta(\xi_1, \dots, \xi_n)$, основанной на использовании величин (ξ_1, \dots, ξ_n) может быть построена «эквивалентная» процедура $\delta'(S_1, \dots, S_k)$, основанная на достаточной статистике (S_1, \dots, S_k) . Действительно, вычислим значение достаточной статистики $(s_1, \dots, s_k) = (S_1(x_1, \dots, x_n), \dots, S_k(x_1, \dots, x_n))$ и вектор (x_1, \dots, x_n) отбросим (он более не потребуются). Далее, поскольку условное распределение $L_\theta(x_1, \dots, x_n | s_1, \dots, s_k)$ случайного вектора (ξ_1, \dots, ξ_n) относительно достаточной статистики (S_1, \dots, S_k) не зависит от неизвестного значения параметра θ , то может быть построен генератор реализаций случайного вектора (ξ_1, \dots, ξ_n) , на вход которого подается значение достаточной статистики (s_1, \dots, s_k) , а на выходе которого будет получена реализация (x'_1, \dots, x'_n) величин (ξ_1, \dots, ξ_n) . Вектор (x'_1, \dots, x'_n) , вообще говоря, может отличаться от исходного вектора (x_1, \dots, x_n) , но значения достаточной статистики (S_1, \dots, S_k) для вектора (x'_1, \dots, x'_n) и для вектора (x_1, \dots, x_n) одинаковы:

$$\begin{aligned} S_1(x'_1, \dots, x'_n) &= S_1(x_1, \dots, x_n), \\ &\dots, \\ S_k(x'_1, \dots, x'_n) &= S_k(x_1, \dots, x_n) \end{aligned}$$

и в этом смысле реализация (x'_1, \dots, x'_n) «равноценна» исходной реализации (x_1, \dots, x_n) . В завершении остается лишь применить статистическую процедуру $\delta(\xi_1, \dots, \xi_n)$ к вектору (x'_1, \dots, x'_n) .

Нахождение достаточных статистик на основе определения во многих случаях является затруднительным, поскольку требует нахождения условного распределения. Более простой способ нахождения достаточных статистик дает приводимый далее критерий факторизации.

Теорема 7.7. (Неймана – Фишера) (критерий факторизации)

Пусть (ξ_1, \dots, ξ_n) – выборка и $L(x_1, \dots, x_n | \theta)$ – функция правдоподобия вектора (ξ_1, \dots, ξ_n) . Статистика $S(\xi_1, \dots, \xi_n) = (S_1(\xi_1, \dots, \xi_n), \dots, S_k(\xi_1, \dots, \xi_n))$ является достаточной для параметра θ тогда и только тогда, когда функция правдоподобия $L(x_1, \dots, x_n | \theta)$ имеет вид:

$$L(x_1, \dots, x_n | \theta) = g(S(x_1, \dots, x_n), \theta)h(x_1, \dots, x_n),$$

где g и h некоторые функции.

Доказательство:

Рассмотрим доказательство только для случая, когда все случайные величины ξ_i ($i = \overline{1, n}$) дискретны.

1) Пусть статистика $S(\xi_1, \dots, \xi_n)$ является достаточной для параметра θ , покажем, что:

$$L(x_1, \dots, x_n | \theta) = g(S(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n).$$

Функция правдоподобия $L(x_1, \dots, x_n | \theta)$ равна вероятности события

$$A(x_1, \dots, x_n) = \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n\} :$$

$$L(x_1, \dots, x_n | \theta) = P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n\}.$$

Рассмотрим событие $B(x_1, \dots, x_n) = \{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}$. Легко видеть, что если при некотором ω выполняются равенства $\xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n$, то при этом же ω выполняется равенство $S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)$, поэтому, очевидно:

$$A(x_1, \dots, x_n) \subseteq B(x_1, \dots, x_n)$$

откуда следует, что совместное наступление событий A и B есть событие A :

$$A(x_1, \dots, x_n) = A(x_1, \dots, x_n) \cap B(x_1, \dots, x_n)$$

то есть,

$$\{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n\} =$$

$$= A(x_1, \dots, x_n) = A(x_1, \dots, x_n) \cap B(x_1, \dots, x_n) =$$

$$= \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n, S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}.$$

Отсюда следует равенство для вероятностей событий:

$$P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n\} =$$

$$= P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n, S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}$$

Вероятность справа представим по формуле умножения как произведение условной и безусловной вероятностей:

$$P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n, S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\} =$$

$$= P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n | S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\} \cdot$$

$$\cdot P_\theta \{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}$$

Условная вероятность $L_\theta(x_1, \dots, x_n | S(x_1, \dots, x_n))$:

$$L_\theta(x_1, \dots, x_n | S(x_1, \dots, x_n)) = P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n | S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}$$

не зависит от параметра θ , поскольку статистика $S(\xi_1, \dots, \xi_n)$ является достаточной для параметра θ , и может зависеть только от x_1, \dots, x_n и $S(x_1, \dots, x_n)$. Таким образом, условная вероятность $L_\theta(x_1, \dots, x_n | S(x_1, \dots, x_n))$ является функцией только x_1, \dots, x_n :

$$L_\theta(x_1, \dots, x_n | S(x_1, \dots, x_n)) = h(x_1, \dots, x_n).$$

Безусловная вероятность $P_\theta \{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}$ очевидно зависит от величины $S(x_1, \dots, x_n)$ и, возможно, от параметра θ :

$$P_\theta \{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\} = g(S(x_1, \dots, x_n), \theta).$$

Таким образом, для функции правдоподобия $L(x_1, \dots, x_n | \theta)$ получим:

$$L(x_1, \dots, x_n | \theta) = P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n\} =$$

$$= P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n, S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\} =$$

$$= P_\theta \{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n | S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\} \cdot$$

$$\cdot P_\theta \{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\} =$$

$$= h(x_1, \dots, x_n) \cdot g(S(x_1, \dots, x_n), \theta).$$

2) Пусть имеет место разложение для функции правдоподобия $L(x_1, \dots, x_n | \theta) = g(S(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n)$, покажем, что в этом случае статистика $S(\xi_1, \dots, \xi_n)$ является достаточной для параметра θ , то есть условная вероятность $L_\theta(x_1, \dots, x_n | S)$ не зависит от параметра θ . По определению условной вероятности:

$$L_{\theta}(x_1, \dots, x_n | s) = P_{\theta}\{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n | S(\xi_1(\omega), \dots, \xi_n(\omega)) = s\} =$$

$$= \frac{P_{\theta}\{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n, S(\xi_1(\omega), \dots, \xi_n(\omega)) = s\}}{P_{\theta}\{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = s\}}$$

Если $s \neq S(x_1, \dots, x_n)$, то вероятность, стоящая в числителе равна нулю независимо от значения параметра θ .

Если $s = S(x_1, \dots, x_n)$, тогда:

$$\frac{P_{\theta}\{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n, S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}}{P_{\theta}\{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}} =$$

$$= \frac{P_{\theta}\{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n\}}{P_{\theta}\{\omega : S(\xi_1(\omega), \dots, \xi_n(\omega)) = S(x_1, \dots, x_n)\}} = \frac{P_{\theta}\{\omega : \xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n\}}{\sum_{\substack{(y_1, \dots, y_n): \\ S(y_1, \dots, y_n) = S(x_1, \dots, x_n)}} P_{\theta}\{\omega : \xi_1(\omega) = y_1, \dots, \xi_n(\omega) = y_n\}} =$$

$$= \frac{L(x_1, \dots, x_n | \theta)}{\sum_{\substack{(y_1, \dots, y_n): \\ S(y_1, \dots, y_n) = S(x_1, \dots, x_n)}} L(y_1, \dots, y_n | \theta)} =$$

$$= \frac{g(S(x_1, \dots, x_n), \theta) \cdot h(x_1, \dots, x_n)}{\sum_{\substack{(y_1, \dots, y_n): \\ S(y_1, \dots, y_n) = S(x_1, \dots, x_n)}} g(S(y_1, \dots, y_n), \theta) \cdot h(y_1, \dots, y_n)} = \frac{g(S(x_1, \dots, x_n), \theta) \cdot h(x_1, \dots, x_n)}{\sum_{\substack{(y_1, \dots, y_n): \\ S(y_1, \dots, y_n) = S(x_1, \dots, x_n)}} g(S(x_1, \dots, x_n), \theta) \cdot h(y_1, \dots, y_n)} =$$

$$= \frac{g(T(x_1, \dots, x_n), \theta) \cdot h(x_1, \dots, x_n)}{g(T(x_1, \dots, x_n), \theta) \cdot \sum_{\substack{(y_1, \dots, y_n): \\ S(y_1, \dots, y_n) = S(x_1, \dots, x_n)}} h(y_1, \dots, y_n)} = \frac{h(x_1, \dots, x_n)}{\sum_{\substack{(y_1, \dots, y_n): \\ S(y_1, \dots, y_n) = S(x_1, \dots, x_n)}} h(y_1, \dots, y_n)}.$$

Таким образом,

$$L_{\theta}(x_1, \dots, x_n | s) = \begin{cases} 0 & , s \neq S(x_1, \dots, x_n) \\ \frac{h(x_1, \dots, x_n)}{\sum_{\substack{(y_1, \dots, y_n): \\ S(y_1, \dots, y_n) = S(x_1, \dots, x_n)}} h(y_1, \dots, y_n)} & , s = S(x_1, \dots, x_n) \end{cases}.$$

Выражение, стоящее справа, очевидно, не зависит от параметра θ , поэтому условная вероятность $L_{\theta}(x_1, \dots, x_n | S)$ не зависит от параметра θ , и следовательно статистика $S(\xi_1, \dots, \xi_n)$ является достаточной для параметра θ .

Теорема доказана.

Следствие 7.8.

Пусть $S(\xi_1, \dots, \xi_n)$ — статистика достаточная для параметра θ , тогда МП-оценка параметра $\theta^*(\xi_1, \dots, \xi_n)$ является функцией достаточной статистики $S(\xi_1, \dots, \xi_n)$.

Доказательство:

Поскольку $S(\xi_1, \dots, \xi_n)$ — статистика достаточная для параметра θ , то по критерию факторизации для функции правдоподобия $L(x_1, \dots, x_n | \theta)$ имеется разложение:

$$L(x_1, \dots, x_n | \theta) = g(S(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n)$$

В соответствии с определением МП-оценки оценка $\theta^*(\xi_1, \dots, \xi_n)$ доставляет наибольшее значение функции правдоподобия $L(x_1, \dots, x_n | \theta)$:

$$L(\xi_1, \dots, \xi_n | \theta^*) = \sup_{\theta} L(\xi_1, \dots, \xi_n | \theta) = \sup_{\theta} g(S(\xi_1, \dots, \xi_n), \theta) h(\xi_1, \dots, \xi_n) =$$

$$= h(\xi_1, \dots, \xi_n) \sup_{\theta} g(S(\xi_1, \dots, \xi_n), \theta).$$

Значение параметра θ , соответствующее наибольшему значению функции $g(S(\xi_1, \dots, \xi_n), \theta)$, будет зависеть от ξ_1, \dots, ξ_n только через значения статистики $S(\xi_1, \dots, \xi_n)$, поэтому МП-оценка $\theta^*(\xi_1, \dots, \xi_n)$ будет функцией статистики $S(\xi_1, \dots, \xi_n)$:

$$\theta^*(\xi_1, \dots, \xi_n) = G(S(\xi_1, \dots, \xi_n)).$$

Следствие доказано.

Следствие 7.9.

Пусть $S(\xi_1, \dots, \xi_n)$ – статистика, достаточная для параметра θ и $S^*(\xi_1, \dots, \xi_n)$ – статистика, через которую можно выразить статистику $S(\xi_1, \dots, \xi_n)$, то есть,

$$S(\xi_1, \dots, \xi_n) = G(S^*(\xi_1, \dots, \xi_n)),$$

где $G(s)$ – некоторая функция, тогда статистика $S^*(\xi_1, \dots, \xi_n)$ тоже является достаточной для параметра θ .

Доказательство:

Действительно, если $S(\xi_1, \dots, \xi_n)$ – статистика, достаточная для параметра θ , то по теореме (7.7) для функции правдоподобия получим факторизацию:

$$L(x_1, \dots, x_n | \theta) = g(S(x_1, \dots, x_n), \theta)h(x_1, \dots, x_n),$$

подставляя сюда выражение статистики S , через статистику S^* ,

$$S(\xi_1, \dots, \xi_n) = G(S^*(\xi_1, \dots, \xi_n)),$$

получим факторизацию:

$$L(x_1, \dots, x_n | \theta) = g(G(S^*(x_1, \dots, x_n)), \theta)h(x_1, \dots, x_n) = g^*(S^*(x_1, \dots, x_n), \theta)h(x_1, \dots, x_n),$$

откуда по теореме 7.7 (в обратную сторону) статистика $S^*(\xi_1, \dots, \xi_n)$ является достаточной для параметра θ .

Следствие доказано.

В общем случае достаточных статистик может быть несколько. В частности, из определения легко видеть, что достаточной статистикой является исходная совокупность наблюдаемых случайных величин (ξ_1, \dots, ξ_n) , которую называют *тривиальной достаточной статистикой*. Разумеется, тривиальная достаточная статистика не представляет большого интереса, напротив крайне желательно располагать достаточными статистиками $S(\xi_1, \dots, \xi_n)$ размерностями k меньше n .

Как показывает следствие 7.9 класс достаточных статистик может быть достаточно обширным, любая статистика $S^*(\xi_1, \dots, \xi_n)$ через которую можно выразить достаточную статистику $S(\xi_1, \dots, \xi_n)$ также становится достаточной. Заметим, однако, что если статистика $S(\xi_1, \dots, \xi_n)$ выражается через статистику $S^*(\xi_1, \dots, \xi_n)$, то это означает, что статистика $S(\xi_1, \dots, \xi_n)$ «не сложнее» статистики $S^*(\xi_1, \dots, \xi_n)$ (речь здесь не обязательно касается размерностей статистик). Таким образом, на множестве достаточных статистик может быть введено, по крайней мере, отношение частичного порядка путем задания подчиненности статистик.

Определение 7.10.

Статистика $\hat{T}(\xi_1, \dots, \xi_n)$ подчинена статистике $\tilde{T}(\xi_1, \dots, \xi_n)$, если:

$$\hat{T}(\xi_1, \dots, \xi_n) = G(\tilde{T}(\xi_1, \dots, \xi_n)),$$

для некоторой функции G .

Определение 7.11.

Если $\hat{T}(\xi_1, \dots, \xi_n)$ подчинена статистике $\tilde{T}(\xi_1, \dots, \xi_n)$ и наоборот $\tilde{T}(\xi_1, \dots, \xi_n)$ подчинена статистике $\hat{T}(\xi_1, \dots, \xi_n)$, то статистики $\hat{T}(\xi_1, \dots, \xi_n)$ и $\tilde{T}(\xi_1, \dots, \xi_n)$ называются *эквивалентными*.

Поскольку отношение подчиненности фактически означает отношение «не сложнее чем», то наибольший интерес представляют достаточные статистики, которые подчинены всем другим достаточным статистикам.

Определение 7.12.

Достаточная статистика называется *минимальной*, если она подчинена любой другой достаточной статистике.

Согласно определению минимальная достаточная статистика является самой «простой» из всех достаточных статистик и её дальнейшее упрощение без потери достаточности оказывается невозможным.

Теорема 7.13.

Если функция правдоподобия $L(x_1, \dots, x_n | \theta)$ при всех (x_1, \dots, x_n) является функцией непрерывной справа (или слева) по θ на множестве Θ , и МП-оценка $\theta^*(\xi_1, \dots, \xi_n)$ единственна и является достаточной статистикой, тогда $\theta^*(\xi_1, \dots, \xi_n)$ является минимальной достаточной статистикой.

Без доказательства.

Большое значение имеет приводимая далее теорема, которая указывает на возможность улучшения произвольных оценок с помощью достаточных статистик.

Теорема 7.14. (Блекуэлл, Рао, Колмогоров)

Пусть оценка $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$, $S(\xi_1, \dots, \xi_n)$ – статистика достаточная для параметра θ , и случайная величина $\hat{\theta}_S$ является условным математическим ожиданием оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ относительно статистики $S(\xi_1, \dots, \xi_n)$:

$$\hat{\theta}_S = M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n) | S(\xi_1, \dots, \xi_n)],$$

тогда

1) случайная величина $\hat{\theta}_S$, зависит от величин (ξ_1, \dots, ξ_n) только через $S(\xi_1, \dots, \xi_n)$, то есть $\hat{\theta}_S = \hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ и, следовательно, $\hat{\theta}_S$ является статистикой;

2) статистика $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ является оценкой из класса K_b :

$$\hat{\theta}_S(S(\xi_1, \dots, \xi_n)) \in K_b;$$

3) оценка $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ не хуже оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ в среднеквадратическом смысле:

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_S - \theta)^2] \leq M_{\theta}[(\hat{\theta} - \theta)^2].$$

Доказательство:

1) Согласно определению условного математического ожидания величина $\hat{\theta}_S = M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n) | S(\xi_1, \dots, \xi_n)]$ является функцией статистики $S(\xi_1, \dots, \xi_n)$:

$$\hat{\theta}_S = \hat{\theta}_S(S(\xi_1, \dots, \xi_n)).$$

Условное математическое ожидание $\hat{\theta}_S = \hat{\theta}_S(S(\xi_1, \dots, \xi_n))$, вообще говоря, могло бы зависеть от параметра θ , однако, такой зависимости нет, поскольку условное распределение (ξ_1, \dots, ξ_n) относительно статистики $S(\xi_1, \dots, \xi_n)$ не зависит от параметра в силу того, что статистика $S(\xi_1, \dots, \xi_n)$ является достаточной.

2) Вычислим математическое ожидание $M_{\theta}[\hat{\theta}_S(S(\xi_1, \dots, \xi_n))]$, воспользовавшись свойством условного математического ожидания:

$$M_{\theta}[\hat{\theta}_S(S(\xi_1, \dots, \xi_n))] = M_{\theta}[M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n) | S(\xi_1, \dots, \xi_n)]] = M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)] = \theta + b(\theta),$$

где последнее равенство получено в силу условия теоремы о том, что оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ принадлежит классу K_b и стало быть её смещение равно функции $b(\theta)$:

$$b(\theta) = M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)] - \theta.$$

3) Рассмотрим среднеквадратическое отклонение оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$:

$$\begin{aligned} M_{\theta}[(\hat{\theta} - \theta)^2] &= M_{\theta}[(\hat{\theta} - \hat{\theta}_S + \hat{\theta}_S - \theta)^2] = \\ &= M_{\theta}[(\hat{\theta} - \hat{\theta}_S)^2] + 2M_{\theta}[(\hat{\theta} - \hat{\theta}_S)(\hat{\theta}_S - \theta)] + M_{\theta}[(\hat{\theta}_S - \theta)^2]. \end{aligned} \quad (7.7)$$

Преобразуем средний множитель, используя свойства условного математического ожидания:

$$\begin{aligned} M_{\theta}[(\hat{\theta} - \hat{\theta}_S)(\hat{\theta}_S - \theta)] &= M_{\theta}[M_{\theta}[(\hat{\theta} - \hat{\theta}_S)(\hat{\theta}_S - \theta) | S]] = M_{\theta}[(\hat{\theta}_S - \theta)M_{\theta}[(\hat{\theta} - \hat{\theta}_S) | S]] = \\ &= M_{\theta}[(\hat{\theta}_S - \theta)(M_{\theta}[\hat{\theta} | S] - M_{\theta}[\hat{\theta}_S | S])] = M_{\theta}[(\hat{\theta}_S - \theta)(\hat{\theta}_S - \hat{\theta}_S)] = M_{\theta}[(\hat{\theta}_S - \theta) \cdot 0] = 0 \end{aligned}$$

Таким образом, из (7.7):

$$M_{\theta}[(\hat{\theta} - \theta)^2] = M_{\theta}[(\hat{\theta} - \hat{\theta}_S)^2] + M_{\theta}[(\hat{\theta}_S - \theta)^2]. \quad (7.8)$$

Поскольку $(\hat{\theta} - \hat{\theta}_S)^2$ является неотрицательной случайной величиной, то $M_{\theta}[(\hat{\theta} - \hat{\theta}_S)^2] \geq 0$, тогда из (7.8):

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta} - \theta)^2] \geq M_{\theta}[(\hat{\theta}_S - \theta)^2].$$

Теорема доказана.

Заметим, что в теореме 7.14 оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ должна лишь принадлежать классу K_b и никаких других условий на оценку $\hat{\theta}(\xi_1, \dots, \xi_n)$ не накладывается. Это означает, что в качестве оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ можно использовать даже «плохие» оценки из класса K_b , в том числе и не обладающие свойством состоятельности.

Получаемая в теореме 7.14 оценка $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ принадлежит классу K_b , поэтому к оценке $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ теорема 7.14 может применяться повторно. Пусть $S_1(\xi_1, \dots, \xi_n)$ и $S_2(\xi_1, \dots, \xi_n)$ достаточные статистики и $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$ — некоторая оценка, тогда, применяя теорему 7.14 к оценке $\hat{\theta}(\xi_1, \dots, \xi_n)$ и статистике $S_1(\xi_1, \dots, \xi_n)$, получим оценку $\hat{\theta}_{S_1}(S_1(\xi_1, \dots, \xi_n))$, которая не хуже $\hat{\theta}(\xi_1, \dots, \xi_n)$ в среднеквадратическом смысле, и, применяя теорему 7.14 к оценке $\hat{\theta}_{S_1}(S_1(\xi_1, \dots, \xi_n))$ и статистике $S_2(\xi_1, \dots, \xi_n)$, получим оценку $\hat{\theta}_{S_2}(S_2(\xi_1, \dots, \xi_n))$ не хуже $\hat{\theta}_{S_1}(S_1(\xi_1, \dots, \xi_n))$.

Таким образом, последовательное применение теоремы 7.14 может приводить к последовательному улучшению оценок. Заметим, однако, что теорема 7.14 утверждает лишь то, что оценка $\hat{\theta}_{S_2}(S_2(\xi_1, \dots, \xi_n))$ не хуже исходной оценки $\hat{\theta}_{S_1}(S_1(\xi_1, \dots, \xi_n))$, поэтому последовательное применение теоремы 7.14 не обязательно приводит к существенному улучшению оценки. Интуитивно ясно, что существенное улучшение оценки следует ожидать, например, в том случае, когда статистика $S_2(\xi_1, \dots, \xi_n)$ существенно «проще» статистики $S_1(\xi_1, \dots, \xi_n)$, другими словами $S_2(\xi_1, \dots, \xi_n)$ подчинена и не эквивалентна $S_1(\xi_1, \dots, \xi_n)$. Отсюда следует что процесс последовательного улучшения оценок должен останавливаться при использовании достаточных статистик, подчиненных всем остальным достаточным статистикам, то есть при использовании минимальных достаточных статистик.

Таким образом, использование в теореме 7.14 минимальных достаточных статистик приводит к наилучшим оценкам.

Вполне возможно, что наилучшие оценки, вычисляемые как условное математическое ожидание относительно минимальных достаточных статистик, окажутся наилучшими во всем классе K_b , то есть эффективными в классе K_b . Оказывается, что эффективность таких оценок определяется свойством полноты достаточных статистик.

Определение 7.15.

Статистика $T(\xi_1, \dots, \xi_n)$ называется *полной*, если для любой функции $g(t)$:

$$\{\forall \theta \in \Theta : M_{\theta}[g(T)] = 0\} \Rightarrow \{g(t) \equiv 0\}.$$

Утверждение 7.16.

Если статистика $T(\xi_1, \dots, \xi_n)$ является полной, тогда в каждом классе K_b существует не более одной оценки $\hat{\theta}(T(\xi_1, \dots, \xi_n))$, являющейся функцией $T(\xi_1, \dots, \xi_n)$.

Если есть класс K_b , в котором существует единственная оценка $\hat{\theta}(T(\xi_1, \dots, \xi_n))$, являющейся функцией $T(\xi_1, \dots, \xi_n)$, тогда $T(\xi_1, \dots, \xi_n)$ является полной статистикой.

Доказательство:

1) Пусть $T(\xi_1, \dots, \xi_n)$ — полная статистика.

Вполне возможно, что во всех классах K_b нет оценок, являющихся функциями от $T(\xi_1, \dots, \xi_n)$, в этом случае утверждение, очевидно, выполнено: в каждом классе не более одной оценки, являющейся функцией $T(\xi_1, \dots, \xi_n)$. Пусть все же существуют классы, в которых содержатся оценки, являющиеся функциями статистики $T(\xi_1, \dots, \xi_n)$, и K_b любой из этих классов. Докажем утверждение для класса K_b от противного: предположим, что в классе K_b существуют несколько различных оценок и две из них это $\hat{\theta}_1(T(\xi_1, \dots, \xi_n))$ и $\hat{\theta}_2(T(\xi_1, \dots, \xi_n))$, каждая из которых является некоторой функцией статистики $T(\xi_1, \dots, \xi_n)$.

Поскольку оценки $\hat{\theta}_1(T)$ и $\hat{\theta}_2(T)$ принадлежат классу K_b , то:

$$M_{\theta}[\hat{\theta}_1(T)] = \theta + b(\theta),$$

$$M_{\theta}[\hat{\theta}_2(T)] = \theta + b(\theta).$$

Откуда следует, что математическое ожидание разности $\hat{\theta}_1(T) - \hat{\theta}_2(T)$:

$$\forall \theta \in \Theta : M_{\theta}[\hat{\theta}_1(T) - \hat{\theta}_2(T)] = M_{\theta}[\hat{\theta}_1(T)] - M_{\theta}[\hat{\theta}_2(T)] = \theta + b(\theta) - \theta - b(\theta) = 0. \quad (7.9)$$

Поскольку $T(\xi_1, \dots, \xi_n)$ является полной статистикой, то из определения 7.15 и (7.9) следует:

$$\hat{\theta}_1(T) - \hat{\theta}_2(T) = 0,$$

$$\hat{\theta}_1(T) = \hat{\theta}_2(T).$$

Полученное равенство противоречит исходному предположению о том, что $\hat{\theta}_1(T(\xi_1, \dots, \xi_n))$ и $\hat{\theta}_2(T(\xi_1, \dots, \xi_n))$ являются различными.

2) Пусть K_b — класс, в котором существует единственная оценка $\hat{\theta}(T(\xi_1, \dots, \xi_n))$, являющейся функцией $T(\xi_1, \dots, \xi_n)$.

Рассмотрим любую функцию $g(t)$ такую, что:

$$\forall \theta \in \Theta : M_{\theta}[g(T)] = 0, \quad (7.10)$$

и покажем, что отсюда следует равенство $g(t) = 0$.

Построим оценку $\tilde{\theta}(T(\xi_1, \dots, \xi_n))$:

$$\tilde{\theta}(T(\xi_1, \dots, \xi_n)) = \hat{\theta}(T(\xi_1, \dots, \xi_n)) + g(T(\xi_1, \dots, \xi_n)). \quad (7.11)$$

Легко видеть, что в силу (7.10) оценка $\tilde{\theta}(T(\xi_1, \dots, \xi_n))$ также принадлежит классу K_b , действительно:

$$M_{\theta}[\tilde{\theta}(T)] = M_{\theta}[\hat{\theta}(T) + g(T)] = M_{\theta}[\hat{\theta}(T)] + M[g(T)] = \theta + b(\theta) + 0,$$

поскольку $\hat{\theta}(T(\xi_1, \dots, \xi_n)) \in K_b$.

Таким образом, построенная оценка $\tilde{\theta}(T(\xi_1, \dots, \xi_n))$ и исходная оценка $\hat{\theta}(T(\xi_1, \dots, \xi_n))$ принадлежат классу K_b , и обе оценки $\tilde{\theta}(T(\xi_1, \dots, \xi_n))$ и $\hat{\theta}(T(\xi_1, \dots, \xi_n))$ являются функциями

статистики $T(\xi_1, \dots, \xi_n)$, но по условию в классе K_b есть только одна оценка, являющаяся функцией статистики $T(\xi_1, \dots, \xi_n)$, отсюда следует, что оценки $\tilde{\theta}(T(\xi_1, \dots, \xi_n))$ и $\hat{\theta}(T(\xi_1, \dots, \xi_n))$ совпадают:

$$\tilde{\theta}(T(\xi_1, \dots, \xi_n)) = \hat{\theta}(T(\xi_1, \dots, \xi_n)),$$

тогда из (7.11):

$$g(T(\xi_1, \dots, \xi_n)) = \tilde{\theta}(T(\xi_1, \dots, \xi_n)) - \hat{\theta}(T(\xi_1, \dots, \xi_n)) = 0$$

Поскольку функция $g(t)$, удовлетворяющая (7.10), была выбрана произвольно, то условие определения 7.15 выполняется для любой функции $g(t)$, следовательно статистика $T(\xi_1, \dots, \xi_n)$ является полной.

Утверждение доказано.

Утверждение 7.17.

Пусть оценка $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$ и $S(\xi_1, \dots, \xi_n)$ – достаточная для параметра θ и полная статистика, тогда оценка $\hat{\theta}_S$:

$$\hat{\theta}_S(S(\xi_1, \dots, \xi_n)) = M_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n) | S(\xi_1, \dots, \xi_n)],$$

является единственной эффективной оценкой в классе K_b .

Доказательство:

1) Согласно теореме 7.14 оценка $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ принадлежит классу K_b и является функцией $S(\xi_1, \dots, \xi_n)$.

Поскольку $S(\xi_1, \dots, \xi_n)$ по условию является полной, то согласно утверждению 7.16 оценка $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ – единственная оценка в K_b , которая является функцией $S(\xi_1, \dots, \xi_n)$.

2) Покажем, что оценка $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ является эффективной.

Пусть $\tilde{\theta}(\xi_1, \dots, \xi_n)$ – любая оценка из класса K_b и $\tilde{\theta}_S(S(\xi_1, \dots, \xi_n))$ – условное математическое ожидание оценки $\tilde{\theta}(\xi_1, \dots, \xi_n)$ относительно достаточной статистики $S(\xi_1, \dots, \xi_n)$:

$$\tilde{\theta}_S(S(\xi_1, \dots, \xi_n)) = M_{\theta}[\tilde{\theta}(\xi_1, \dots, \xi_n) | S(\xi_1, \dots, \xi_n)].$$

Согласно пунктам 1 и 2 теоремы 7.14 оценка $\tilde{\theta}_S(S(\xi_1, \dots, \xi_n))$ является функцией $S(\xi_1, \dots, \xi_n)$ и принадлежит классу K_b . Как уже было доказано в пункте 1 утверждения, в классе K_b существует только одна оценка, являющаяся функцией $S(\xi_1, \dots, \xi_n)$, откуда следует, что оценки $\tilde{\theta}_S(S(\xi_1, \dots, \xi_n))$ и $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ совпадают:

$$\tilde{\theta}_S(S(\xi_1, \dots, \xi_n)) = \hat{\theta}_S(S(\xi_1, \dots, \xi_n)).$$

Из совпадения оценок следует равенство среднеквадратических отклонений:

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_S - \theta)^2] = M_{\theta}[(\tilde{\theta}_S - \theta)^2]. \quad (7.12)$$

В соответствии с пунктом 3 теоремы 7.14 для оценки $\tilde{\theta}_S(S(\xi_1, \dots, \xi_n))$ справедливо неравенство:

$$\forall \theta \in \Theta : M_{\theta}[(\tilde{\theta}_S - \theta)^2] \leq M_{\theta}[(\tilde{\theta} - \theta)^2],$$

из которого в силу (7.12) следует:

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_S - \theta)^2] = M_{\theta}[(\tilde{\theta}_S - \theta)^2] \leq M_{\theta}[(\tilde{\theta} - \theta)^2],$$

$$\forall \theta \in \Theta : M_{\theta}[(\hat{\theta}_S - \theta)^2] \leq M_{\theta}[(\tilde{\theta} - \theta)^2].$$

Поскольку оценка $\tilde{\theta}(\xi_1, \dots, \xi_n)$ была выбрана произвольным образом, то последнее неравенство справедливо для всех оценок из класса K_b и следовательно по определению 7.3

оценка $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ является эффективной в классе K_b .

3) Согласно пунктам 1 и 2 утверждения в классе K_b существует эффективная оценка $\tilde{\theta}_S(S(\xi_1, \dots, \xi_n))$, тогда из утверждения 7.5 оценка $\tilde{\theta}_S(S(\xi_1, \dots, \xi_n))$ является единственной эффективной оценкой.

Утверждение доказано.

Следствие 7.18.

Если $\hat{\theta}(\xi_1, \dots, \xi_n)$ является несмещенной оценкой, то $\hat{\theta}_S(S(\xi_1, \dots, \xi_n))$ является единственной эффективной оценкой.

Утверждение 7.19.

Если $S(\xi_1, \dots, \xi_n)$ — полная достаточная статистика, тогда $S(\xi_1, \dots, \xi_n)$ — минимальная статистика.

Без доказательства.

3. Неравенство Рао-Крамера.

Пусть, как и ранее, (ξ_1, \dots, ξ_n) является выборкой из распределения $F_\xi(x|\theta)$, которое некоторым образом зависит от скалярного параметра θ , принимающего значения из допустимого множества значений Θ .

В общем случае в совокупности величин (ξ_1, \dots, ξ_n) содержится лишь ограниченное количество информации о параметре θ , и поскольку величины (ξ_1, \dots, ξ_n) являются основой для построения оценок параметра θ , то получаемые оценки не могут быть сколь угодно «хорошими» и не могут иметь сколь угодно малые значения среднеквадратических отклонений. Отсюда возникает вопрос о том, насколько малым может быть среднеквадратическое отклонение оценки из некоторого класса. Ответ на этот вопрос в так называемых регулярных случаях дает приводимое далее неравенство Рао-Крамера.

Определение 7.20.

Пусть (ξ_1, \dots, ξ_n) — вектор случайных величин и $L(x_1, \dots, x_n|\theta)$ функция плотности вероятности (или вероятность) вектора (ξ_1, \dots, ξ_n) . Функция $L(\xi_1, \dots, \xi_n|\theta)$ рассматриваемая как функция параметра θ называется *функцией правдоподобия*.

Определение 7.21.

Случайная функция $U(\xi_1, \dots, \xi_n|\theta)$:

$$U(\xi_1, \dots, \xi_n|\theta) = \frac{\partial \ln L(\xi_1, \dots, \xi_n|\theta)}{\partial \theta}$$

называется *функцией вклада*.

Определение 7.22.

Функция $I_n(\theta)$:

$$I_n(\theta) = M_\theta[U^2(\xi_1, \dots, \xi_n|\theta)]$$

называется *информацией Фишера* о параметре θ , содержащейся в совокупности (ξ_1, \dots, ξ_n) .

Везде далее будем считать, что выполнены следующие условия, которые носят название *условий регулярности*:

R1) Существует множество X такое, что:

$$\forall \theta \in \Theta : P\{(\xi_1, \dots, \xi_n) \in X\} = 1,$$

$$\forall \theta \in \Theta \forall (x_1, \dots, x_n) \in X : L(x_1, \dots, x_n|\theta) > 0$$

R2) При всех $(x_1, \dots, x_n) \in X$ функция $\sqrt{L(x_1, \dots, x_n|\theta)}$ непрерывно дифференцируема по параметру θ для всех допустимых значений $\theta \in \Theta$.

R3) При всех $\theta \in \Theta$ интеграл $I_n(\theta)$:

$$0 < \int_x \left(\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \right)^2 L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n < \infty$$

положителен, существует и непрерывен по θ .

Заметим, что при каждом значении θ функция правдоподобия $L(x_1, \dots, x_n | \theta)$ является функцией плотности вероятности (или функцией вероятности), поэтому, как и для всякой другой функции плотности вероятности, для функции $L(x_1, \dots, x_n | \theta)$ справедливо равенство:

$$\int_x L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 1.$$

Продифференцируем левую и правую часть по θ :

$$\frac{\partial}{\partial \theta} \int_x L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = \frac{\partial}{\partial \theta} 1 = 0$$

и, пользуясь условиями регулярности, внесем дифференцирование под знак интеграла:

$$\int_x \frac{\partial L(x_1, \dots, x_n | \theta)}{\partial \theta} dx_1 \dots dx_n = 0$$

В силу условия R1 преобразуем подынтегральную функцию:

$$\begin{aligned} \frac{\partial L(x_1, \dots, x_n | \theta)}{\partial \theta} &= \frac{\partial L(x_1, \dots, x_n | \theta)}{\partial \theta} \frac{L(x_1, \dots, x_n | \theta)}{L(x_1, \dots, x_n | \theta)} = \\ &= \left(\frac{1}{L(x_1, \dots, x_n | \theta)} \frac{\partial L(x_1, \dots, x_n | \theta)}{\partial \theta} \right) L(x_1, \dots, x_n | \theta) = \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} L(x_1, \dots, x_n | \theta). \end{aligned}$$

Таким образом,

$$\int_x \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 0$$

Заметим, что слева от знака равенства стоит в точности $M_\theta[U(\xi_1, \dots, \xi_n | \theta)]$, таким образом, при выполнении условий регулярности:

$$M_\theta[U(\xi_1, \dots, \xi_n | \theta)] = 0. \quad (7.13)$$

Кроме того, при выполнении условий регулярности из (7.13):

$$\begin{aligned} I_n(\theta) &= M_\theta[U^2(\xi_1, \dots, \xi_n | \theta)] = D_\theta[U^2(\xi_1, \dots, \xi_n | \theta)] + (M_\theta[U(\xi_1, \dots, \xi_n | \theta)])^2, \\ I_n(\theta) &= D_\theta[U^2(\xi_1, \dots, \xi_n | \theta)]. \end{aligned} \quad (7.14)$$

Теорема 7.23. (неравенство Рао-Крамера)

Пусть выполнены условия регулярности и оценка $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$ такова, что:

- 1) $\forall \theta \in \Theta : M_\theta[\hat{\theta}^2(\xi_1, \dots, \xi_n)] < c < \infty$,
- 2) функция $b(\theta)$ дифференцируема по параметру θ при всех $\theta \in \Theta$,

тогда:

$$D_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)},$$

где $I_n(\theta)$ информация Фишера о параметре θ , содержащаяся в совокупности величин (ξ_1, \dots, ξ_n) .

Доказательство:

- 3) Поскольку $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$, то:

$$M_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] = \theta + b(\theta),$$

$$\int_x \hat{\theta}(x_1, \dots, x_n) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = \theta + b(\theta)$$

Продифференцируем левую и правую часть по θ :

$$\frac{\partial}{\partial \theta} \int_x \hat{\theta}(x_1, \dots, x_n) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = \frac{\partial}{\partial \theta} (\theta + b(\theta)).$$

В силу условий регулярности в левой части можно внести дифференцирование под знак интеграла:

$$\int_x \hat{\theta}(x_1, \dots, x_n) \frac{\partial L(x_1, \dots, x_n | \theta)}{\partial \theta} dx_1 \dots dx_n = 1 + b'(\theta)$$

Преобразуем правую часть в силу условия R1 (также как при выводе 7.13):

$$\begin{aligned} \int_x \hat{\theta}(x_1, \dots, x_n) \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n &= 1 + b'(\theta) \\ \int_x \hat{\theta}(x_1, \dots, x_n) U(x_1, \dots, x_n | \theta) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n &= 1 + b'(\theta) \end{aligned} \quad (7.15)$$

4) При выполнении условий регулярности справедливо соотношение (7.13):

$$\int_x U(x_1, \dots, x_n | \theta) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 0$$

Умножим левую и правую часть на $\theta + b(\theta)$ и внесем в правой части $\theta + b(\theta)$ как множитель, не зависящий от переменных интегрирования x_1, \dots, x_n :

$$\int_x (\theta + b(\theta)) U(x_1, \dots, x_n | \theta) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 0 \cdot (\theta + b(\theta)) = 0 \quad (7.16)$$

5) Вычтем (7.16) из (7.15):

$$\int_x (\hat{\theta}(x_1, \dots, x_n) - (\theta + b(\theta))) U(x_1, \dots, x_n | \theta) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 1 + b'(\theta)$$

Поскольку $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$, то $M_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] = \theta + b(\theta)$, тогда:

$$\int_x (\hat{\theta}(x_1, \dots, x_n) - M_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)]) U(x_1, \dots, x_n | \theta) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 1 + b'(\theta)$$

В условиях регулярности математическое ожидание функции вклада $U(x_1, \dots, x_n | \theta)$ равно нулю (соотношение (7.13)), тогда:

$$\begin{aligned} \int_x (\hat{\theta}(x_1, \dots, x_n) - M_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)]) (U(x_1, \dots, x_n | \theta) - M_\theta[U(\xi_1, \dots, \xi_n | \theta)]) L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n &= \\ &= 1 + b'(\theta), \\ \text{cov}_\theta(\hat{\theta}(\xi_1, \dots, \xi_n), U(\xi_1, \dots, \xi_n | \theta)) &= 1 + b'(\theta). \end{aligned} \quad (7.17)$$

В соответствии со свойством ковариации:

$$\left| \text{cov}_\theta(\hat{\theta}(\xi_1, \dots, \xi_n), U(\xi_1, \dots, \xi_n | \theta)) \right|^2 \leq D_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] \cdot D_\theta[U(\xi_1, \dots, \xi_n | \theta)].$$

Таким образом,

$$[1 + b'(\theta)]^2 = \left| \text{cov}_\theta(\hat{\theta}(\xi_1, \dots, \xi_n), U(\xi_1, \dots, \xi_n | \theta)) \right|^2 \leq D_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] \cdot D_\theta[U(\xi_1, \dots, \xi_n | \theta)].$$

Отсюда,

$$D_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] \geq \frac{[1 + b'(\theta)]^2}{D_\theta[U(\xi_1, \dots, \xi_n | \theta)]},$$

и в силу (7.14):

$$D_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)}$$

Теорема доказана.

Следствие 7.24.

В условиях теоремы

$$M_{\theta}[(\hat{\theta} - \theta)^2] \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)} + b^2(\theta).$$

Доказательство:

$$M_{\theta}[(\hat{\theta} - \theta)^2] = D_{\theta}[\hat{\theta} - \theta] + (M_{\theta}[\hat{\theta} - \theta])^2 = D_{\theta}[\hat{\theta}] + b^2(\theta) \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)} + b^2(\theta).$$

Следствие доказано.

Определение 7.25.

Оценка $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$ называется *R-эффективной (регулярно-эффективной)* в классе K_b , если:

$$M_{\theta}[(\hat{\theta} - \theta)^2] = \frac{[1 + b'(\theta)]^2}{I_n(\theta)} + b^2(\theta).$$

Определение 7.26.

Оценка R-эффективная в классе K_0 называется *R-эффективной*.

Легко видеть, что всякая R-эффективная в классе K_b оценка является эффективной в классе K_b . Действительно, пусть $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$ является R-эффективной в классе K_b оценкой. Рассмотрим любую оценку $\tilde{\theta}(\xi_1, \dots, \xi_n) \in K_b$, согласно теореме 7.23 и следствию 7.24:

$$M_{\theta}[(\tilde{\theta} - \theta)^2] \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)} + b^2(\theta),$$

тогда из определения 7.25:

$$M_{\theta}[(\tilde{\theta} - \theta)^2] \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)} + b^2(\theta) = M_{\theta}[(\hat{\theta} - \theta)^2]$$

при всяком значении $\theta \in \Theta$. Откуда по определению 7.3 оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является эффективной в классе K_b .

Обратное утверждение не всегда верно, эффективная в классе K_b оценка не обязательно является R-эффективной. Другими словами, если оценка обладает наименьшим среднеквадратическим отклонением, оно не обязательно совпадает с нижней границей, утверждаемой теоремой 7.23. Причина кроется не в том, что эффективная оценка является недостаточно хорошей (в своем классе эффективная оценка является наилучшей), а в том, что нижняя граница в неравенстве Рао-Крамера является неточной.

4. Экспоненциальные семейства распределений.

Пусть (ξ_1, \dots, ξ_n) является выборкой из распределения $F_{\xi}(x | \theta)$ с функцией плотности вероятности $f_{\xi}(x | \theta)$, где θ — скалярный параметр, который принимает значения из множества допустимых значений Θ .

Существуют случаи, в которых построение R-эффективных оценок становится особенно простым, и может быть выполнено на основе вида функции правдоподобия.

Определение 7.27.

Семейство распределений с функцией плотности вероятности $f_{\xi}(x | \theta)$, имеющей вид:

$$f_{\xi}(x | \theta) = h(x) \cdot e^{\sum_{j=1}^k a_j(\theta) b_j(x) + c(\theta)}$$

называется *экспоненциальным семейством распределений*.

Легко видеть, что для экспоненциального семейства распределений функция правдоподобия $L(\xi_1, \dots, \xi_n | \theta)$ выборки (ξ_1, \dots, ξ_n) имеет вид:

$$\begin{aligned}
L(\xi_1, \dots, \xi_n | \theta) &= \prod_{i=1}^n f_{\xi}(\xi_i | \theta) = \prod_{i=1}^n h(\xi_i) \cdot e^{\sum_{j=1}^k a_j(\theta) r_j(\xi_i) + v(\theta)} = \prod_{i=1}^n h(\xi_i) \cdot e^{\sum_{j=1}^k \sum_{i=1}^n a_j(\theta) r_j(\xi_i) + nv(\theta)} = \\
&= \prod_{i=1}^n h(\xi_i) \cdot e^{\sum_{j=1}^k a_j(\theta) \sum_{i=1}^n r_j(\xi_i) + nv(\theta)} = \prod_{i=1}^n h(\xi_i) \cdot e^{\sum_{j=1}^k a_j(\theta) S_j(\xi_1, \dots, \xi_n) + nv(\theta)},
\end{aligned}$$

$$\text{где } S_j(\xi_1, \dots, \xi_n) = \sum_{i=1}^n r_j(\xi_i).$$

Откуда непосредственно по критерию факторизации следует, что статистика $S(\xi_1, \dots, \xi_n)$:

$$S(\xi_1, \dots, \xi_n) = \begin{pmatrix} S_1(\xi_1, \dots, \xi_n) \\ \dots \\ S_k(\xi_1, \dots, \xi_n) \end{pmatrix}$$

является достаточной для параметра θ . Более того, можно показать, что статистика $S(\xi_1, \dots, \xi_n)$ является минимальной достаточной статистикой и в некоторых случаях даже полной статистикой.

Теорема 7.28.

Пусть выполнены условия регулярности и оценка $\hat{\theta}(\xi_1, \dots, \xi_n) \in K_b$ такова, что:

- 1) $\forall \theta \in \Theta : M_{\theta}[\hat{\theta}^2(\xi_1, \dots, \xi_n)] < c < \infty$,
- 2) функция $b(\theta)$ дифференцируема по параметру θ при всех $\theta \in \Theta$.

Оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является R-эффективной в классе K_b тогда и только тогда, когда функция правдоподобия $L(\xi_1, \dots, \xi_n | \theta)$ имеет вид:

$$L(\xi_1, \dots, \xi_n | \theta) = H(\xi_1, \dots, \xi_n) e^{A(\theta) \cdot \hat{\theta}(\xi_1, \dots, \xi_n) + V(\theta)}.$$

Схема доказательства:

Согласно определению 7.25 оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является R-эффективной в классе K_b тогда и только тогда, когда:

$$M_{\theta}[(\hat{\theta} - \theta)^2] = \frac{[1 + b'(\theta)]^2}{I_n(\theta)} + b^2(\theta) \quad (7.18)$$

Среднеквадратическое отклонение $M_{\theta}[(\hat{\theta} - \theta)^2]$ и дисперсия $D_{\theta}[\hat{\theta}(\xi_1, \dots, \xi_n)]$ оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ связаны соотношением:

$$\begin{aligned}
D_{\theta}[\hat{\theta}] &= D_{\theta}[\hat{\theta} - \theta] = M_{\theta}[(\hat{\theta} - \theta)^2] - (M_{\theta}[\hat{\theta} - \theta])^2, \\
D_{\theta}[\hat{\theta}] &= M_{\theta}[(\hat{\theta} - \theta)^2] - b^2(\theta)
\end{aligned} \quad (7.19)$$

поэтому из (7.18) (с учетом (7.19)) немедленно следует:

$$D_{\theta}[\hat{\theta}] = \frac{[1 + b'(\theta)]^2}{I_n(\theta)} + b^2(\theta) - b^2(\theta) = \frac{[1 + b'(\theta)]^2}{I_n(\theta)} \quad (7.20)$$

и наоборот, из (7.20) следует (7.18) (благодаря (7.19)). Таким образом, соотношения (7.18) и (7.20) эквивалентны. Далее, (7.20) выполняется тогда и только тогда, когда:

$$\begin{aligned}
D_{\theta}[\hat{\theta}] I_n(\theta) &= [1 + b'(\theta)]^2, \\
D_{\theta}[\hat{\theta}] D_{\theta}[U] &= [1 + b'(\theta)]^2
\end{aligned}$$

При доказательстве теоремы 7.23 было доказано, что $1 + b'(\theta) = \text{cov}_{\theta}(\hat{\theta}, U)$ (равенство (7.17)), тогда:

$$D_{\theta}[\hat{\theta}] D_{\theta}[U] = [\text{cov}_{\theta}(\hat{\theta}, U)]^2,$$

$$\sqrt{D_{\theta}[\hat{\theta}]D_{\theta}[U]} = |\text{cov}_{\theta}(\hat{\theta}, U)|. \quad (7.21)$$

Таким образом, равенства (7.18) и (7.21) эквивалентны. В левой части (7.21):

$$D_{\theta}[\hat{\theta}] = \int_{\mathbf{x}} (\hat{\theta}(x_1, \dots, x_n) - (\theta + b(\theta)))^2 L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n \quad (7.22)$$

$$\begin{aligned} D_{\theta}[U] &= \int_{\mathbf{x}} U(x_1, \dots, x_n)^2 L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = \\ &= \int_{\mathbf{x}} \left(\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \right)^2 L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n. \end{aligned} \quad (7.23)$$

поскольку в условиях регулярности $M_{\theta}[U] = 0$.

В силу условия регулярности R3 $D_{\theta}[U] > 0$, поэтому равенство (7.21) возможно тогда и только тогда, когда функции в интегралах (7.22) и (7.23) связаны линейной зависимостью:

$$\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} = c(\theta)(\hat{\theta}(x_1, \dots, x_n) - (\theta + b(\theta))). \quad (7.24)$$

с некоторой функцией $c(\theta)$. При некоторых условиях соотношение (7.24) эквивалентно следующему равенству:

$$\ln L(x_1, \dots, x_n | \theta) = \hat{\theta}(x_1, \dots, x_n) \int c(\theta) d\theta + \int c(\theta)(\theta + b(\theta)) d\theta + G(x_1, \dots, x_n), \quad (7.25)$$

которое получается интегрированием (7.24) по параметру θ . Если ввести обозначения:

$$A(\theta) = \int c(\theta) d\theta, \quad V(\theta) = \int c(\theta)(\theta + b(\theta)) d\theta,$$

тогда в результате потенцирования придем к эквивалентному равенству:

$$\begin{aligned} L(x_1, \dots, x_n | \theta) &= e^{G(x_1, \dots, x_n)} e^{A(\theta)\hat{\theta}(x_1, \dots, x_n) + V(\theta)}, \\ L(x_1, \dots, x_n | \theta) &= H(x_1, \dots, x_n) e^{A(\theta)\hat{\theta}(x_1, \dots, x_n) + V(\theta)}. \end{aligned} \quad (7.26)$$

Поскольку все соотношения от (7.18) до (7.26) являются эквивалентными, то повторение соотношений в обратную сторону от (7.26) к (7.18) доказывает теорему в сторону достаточности.

Теорема доказана.

Следствие 7.29.

В условиях теоремы 7.28 R-эффективная в классе K_b оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является достаточной для параметра θ статистикой.

Доказательство:

При выполнении условий теоремы 7.28 имеет место следующая факторизация функции правдоподобия:

$$L(x_1, \dots, x_n | \theta) = H(x_1, \dots, x_n) e^{A(\theta)\hat{\theta}(x_1, \dots, x_n) + V(\theta)}.$$

Положим $h(x_1, \dots, x_n) = H(x_1, \dots, x_n)$ и $g(\hat{\theta}(x_1, \dots, x_n), \theta) = e^{A(\theta)\hat{\theta}(x_1, \dots, x_n) + V(\theta)}$ тогда:

$$L(x_1, \dots, x_n | \theta) = h(x_1, \dots, x_n) g(\hat{\theta}(x_1, \dots, x_n), \theta),$$

и по критерию факторизации (теорема 7.7) оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является достаточной для параметра θ .

Следствие доказано.

Следствие 7.30.

В условиях теоремы 7.28 R-эффективная оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является МП-оценкой параметра θ .

Доказательство:

1) Оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является R-эффективной, следовательно, принадлежит классу K_0 (классу оценок с нулевым смещением), тогда функция смещения $b(\theta)$ оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ на

множестве Θ равна нулю:

$$b(\theta) = 0.$$

В условиях теоремы из R-эффективности в классе K_0 следует соотношение (7.23), в котором функция смещения $b(\theta) = 0$:

$$\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} = c(\theta)(\hat{\theta}(x_1, \dots, x_n) - \theta). \quad (7.27)$$

2) Покажем, что в (7.27) функция $c(\theta)$ на множестве Θ принимает только положительные значения. Действительно, условия теоремы 7.28 совпадают с условиями теоремы 7.23, следовательно, выполняется равенство (7.17), в котором $b'(\theta) = 0$, поскольку $b(\theta) = 0$:

$$\begin{aligned} \text{cov}_\theta(\hat{\theta}, U) &= 1, \\ \text{cov}_\theta\left(\hat{\theta}, \frac{\partial \ln L}{\partial \theta}\right) &= 1, \end{aligned}$$

и в силу (7.27):

$$\text{cov}_\theta(\hat{\theta}, c(\theta)(\hat{\theta} - \theta)) = 1.$$

Функция $c(\theta)$ не является случайной, поэтому функцию $c(\theta)$ можно вынести из ковариации как постоянный множитель:

$$c(\theta) \text{cov}_\theta(\hat{\theta}, \hat{\theta} - \theta) = 1.$$

Параметр θ также не является случайной величиной, поэтому по свойству ковариации удаление постоянного смещения не изменяет ковариации:

$$\begin{aligned} c(\theta) \text{cov}_\theta(\hat{\theta}, \hat{\theta}) &= 1, \\ c(\theta) D_\theta[\hat{\theta}] &= 1. \end{aligned} \quad (7.28)$$

Оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является R-эффективной, поэтому дисперсия $D_\theta[\hat{\theta}]$ оценки совпадает с нижней границей в неравенстве Рао-Крамера (в которой $b'(\theta) = 0$):

$$D_\theta[\hat{\theta}] = \frac{1}{I_n(\theta)},$$

где информация Фишера $I_n(\theta) > 0$ в силу условия регулярности R3:

$$D_\theta[\hat{\theta}] = \frac{1}{I_n(\theta)} > 0.$$

Таким образом, из (7.28):

$$c(\theta) = \frac{1}{D_\theta[\hat{\theta}]} > 0. \quad (7.29)$$

3) Теперь из (7.27) с учетом (7.29) следует, что если в качестве значений параметра θ выбирать величины меньше значения оценки $\hat{\theta}(x_1, \dots, x_n)$, то производная слева положительна:

$$\hat{\theta}(x_1, \dots, x_n) > \theta : \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} = c(\theta)(\hat{\theta}(x_1, \dots, x_n) - \theta) > 0.$$

Если же наоборот в качестве значений параметра θ выбирать величины больше значения оценки $\hat{\theta}(x_1, \dots, x_n)$, то производная слева отрицательна:

$$\hat{\theta}(x_1, \dots, x_n) < \theta : \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} = c(\theta)(\hat{\theta}(x_1, \dots, x_n) - \theta) < 0.$$

Таким образом, в точке, соответствующей значению оценки $\hat{\theta}(x_1, \dots, x_n)$, функция $\ln L(x_1, \dots, x_n | \theta)$ (как функция параметра θ) имеет локальный максимум. Следовательно, для

всякой реализации величин (ξ_1, \dots, ξ_n) оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ доставляет наибольшее значение логарифмической функции правдоподобия $\ln L(\xi_1, \dots, \xi_n | \theta)$ и функции правдоподобия $L(\xi_1, \dots, \xi_n | \theta)$:

$$L(\xi_1, \dots, \xi_n | \hat{\theta}(\xi_1, \dots, \xi_n)) = \sup_{\theta \in \Theta} L(\xi_1, \dots, \xi_n | \theta).$$

Отсюда R-эффективная оценка $\hat{\theta}(\xi_1, \dots, \xi_n)$ является оценкой максимального правдоподобия.

Следствие доказано.

Следствие 7.29 показывает, что множество R-эффективных в классе K_b оценок содержится во множестве достаточных статистик. Другими словами, если найдена достаточная статистика, то вполне возможно она же является и R-эффективной оценкой в некотором классе K_b .

Следствие 7.30 аналогичным образом утверждает, что R-эффективные оценки содержатся в множестве МП-оценок. Следовательно, если найдена оценка по методу максимального правдоподобия, и она имеет нулевое смещение (принадлежит классу K_0), то возможно эта оценка является R-эффективной.

5. Информация Фишера.

Вычисление информации Фишера $I_n(\theta)$ непосредственно из определения 7.25 может вызывать существенные трудности, связанные необходимостью вычисления интеграла:

$$I_n(\theta) = M_\theta[U^2(\xi_1, \dots, \xi_n | \theta)] = \int_x \left(\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \right)^2 L(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Тем не менее, в условиях регулярности и при некоторых дополнительных условиях вычисление информации Фишера может быть выполнено иначе.

Утверждение 7.31.

Пусть выполнены условия регулярности и функция правдоподобия $L(\xi_1, \dots, \xi_n | \theta)$ дважды дифференцируема по θ , тогда:

$$I_n(\theta) = -M_\theta \left[\frac{\partial^2 \ln L(\xi_1, \dots, \xi_n | \theta)}{\partial \theta^2} \right].$$

Доказательство:

При выполнении условий регулярности справедливо равенство (7.13):

$$M_\theta[U(\xi_1, \dots, \xi_n | \theta)] = 0,$$

$$\int_x \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 0.$$

Продифференцировав левую и правую часть по θ получим:

$$\frac{\partial}{\partial \theta} \left(\int_x \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n \right) = 0,$$

$$\int_x \frac{\partial}{\partial \theta} \left(\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} L(x_1, \dots, x_n | \theta) \right) dx_1 \dots dx_n = 0,$$

$$\int_x \left(\frac{\partial^2 \ln L(x_1, \dots, x_n | \theta)}{\partial \theta^2} L(x_1, \dots, x_n | \theta) + \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \frac{\partial L(x_1, \dots, x_n | \theta)}{\partial \theta} \right) dx_1 \dots dx_n = 0,$$

$$\int_x \frac{\partial^2 \ln L(x_1, \dots, x_n | \theta)}{\partial \theta^2} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n + \int_x \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \frac{\partial L(x_1, \dots, x_n | \theta)}{\partial \theta} dx_1 \dots dx_n = 0.$$

Проведя в силу условия R1 во втором интеграле преобразование такое же как и при

выводе (7.13) получим:

$$\begin{aligned} \int_x \frac{\partial^2 \ln L(x_1, \dots, x_n | \theta)}{\partial \theta^2} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n + \\ + \int_x \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 0, \\ \int_x \frac{\partial^2 \ln L(x_1, \dots, x_n | \theta)}{\partial \theta^2} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n + \int_x \left(\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \right)^2 L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = 0, \\ \int_x \left(\frac{\partial \ln L(x_1, \dots, x_n | \theta)}{\partial \theta} \right)^2 L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n = - \int_x \frac{\partial^2 \ln L(x_1, \dots, x_n | \theta)}{\partial \theta^2} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n. \end{aligned}$$

Слева от знака равенства располагается информация Фишера $I_n(\theta) = M_\theta[U^2(\xi_1, \dots, \xi_n | \theta)]$, поэтому, окончательно:

$$\begin{aligned} I_n(\theta) &= - \int_x \frac{\partial^2 \ln L(x_1, \dots, x_n | \theta)}{\partial \theta^2} L(x_1, \dots, x_n | \theta) dx_1 \dots dx_n, \\ I_n(\theta) &= -M_\theta \left[\frac{\partial^2 \ln L(\xi_1, \dots, \xi_n | \theta)}{\partial \theta^2} \right]. \end{aligned}$$

Утверждение доказано.

Утверждение 7.32. (аддитивность информации Фишера)

Пусть в совокупности величин (ξ_1, \dots, ξ_n) случайные величины ξ_i ($i = \overline{1, n}$) совместно независимы и имеют плотности вероятности $f_i(x | \theta)$ соответственно.

Если выполнены условия утверждения 7.31, тогда информация Фишера $I_n(\theta)$:

$$I_n(\theta) = \sum_{i=1}^n J_i(\theta),$$

где $J_i(\theta)$ информация Фишера, содержащаяся в одной случайной величине ξ_i .

Доказательство:

Поскольку случайные величины ξ_i ($i = \overline{1, n}$) совместно независимы и имеют функции плотности вероятности $f_i(x | \theta)$ ($i = \overline{1, n}$), то функция правдоподобия $L(\xi_1, \dots, \xi_n | \theta)$:

$$L(\xi_1, \dots, \xi_n | \theta) = \prod_{i=1}^n f_i(\xi_i | \theta),$$

тогда,

$$\frac{\partial^2 \ln L(\xi_1, \dots, \xi_n | \theta)}{\partial \theta^2} = \frac{\partial^2}{\partial \theta^2} \sum_{i=1}^n \ln f_i(\xi_i | \theta) = \sum_{i=1}^n \frac{\partial^2 \ln f_i(\xi_i | \theta)}{\partial \theta^2}.$$

В условиях утверждения справедливо утверждение 7.31:

$$I_n(\theta) = -M_\theta \left[\frac{\partial^2 \ln L(\xi_1, \dots, \xi_n | \theta)}{\partial \theta^2} \right],$$

откуда

$$I_n(\theta) = -M_\theta \left[\sum_{i=1}^n \frac{\partial^2 \ln f_i(\xi_i | \theta)}{\partial \theta^2} \right] = \sum_{i=1}^n \left(-M_\theta \left[\frac{\partial^2 \ln f_i(\xi_i | \theta)}{\partial \theta^2} \right] \right) = \sum_{i=1}^n J_i(\theta),$$

где $J_i(\theta) = -M_\theta \left[\frac{\partial^2 \ln f_i(\xi_i | \theta)}{\partial \theta^2} \right]$ - информация Фишера, содержащаяся в одной случайной величине ξ_i .

Утверждение доказано.

Утверждение 7.33. (информация Фишера в случае выборки)

Пусть совокупность (ξ_1, \dots, ξ_n) является выборкой из распределения с плотностью вероятности $f_\xi(x|\theta)$. Если выполнены условия утверждения 7.32, тогда информация Фишера:

$$I_n(\theta) = n \cdot I_1(\theta),$$

где $I_1(\theta) = -M_\theta \left[\frac{\partial^2 \ln f_\xi(\xi_1|\theta)}{\partial \theta^2} \right]$ – информация Фишера, содержащаяся в одной из случайных величин ξ_1, \dots, ξ_n , например, в случайной величине ξ_1 .

Доказательство:

В условиях утверждения справедливо утверждение 7.32, согласно которому информация Фишера:

$$I_n(\theta) = \sum_{i=1}^n J_i(\theta),$$

$$J_i(\theta) = -M_\theta \left[\frac{\partial^2 \ln f_i(\xi_i|\theta)}{\partial \theta^2} \right].$$

Поскольку случайные величины ξ_i ($i = \overline{1, n}$) образуют выборку, то все плотности вероятности одинаковы, $f_i(x|\theta) = f_\xi(x|\theta)$, тогда:

$$J_i(\theta) = -M_\theta \left[\frac{\partial^2 \ln f_i(\xi_i|\theta)}{\partial \theta^2} \right] = -M_\theta \left[\frac{\partial^2 \ln f_\xi(\xi_i|\theta)}{\partial \theta^2} \right] = -M_\theta \left[\frac{\partial^2 \ln f_\xi(\xi_1|\theta)}{\partial \theta^2} \right] = I_1(\theta),$$

и следовательно,

$$I_n(\theta) = \sum_{i=1}^n J_i(\theta) = \sum_{i=1}^n I_1(\theta) = n \cdot I_1(\theta).$$

Утверждение доказано.

Замечание 7.34.

Пусть выполнены условия теоремы 7.23 и утверждения 7.33, тогда для всякой оценки $\hat{\theta}(\xi_1, \dots, \xi_n)$ из класса K_b , согласно неравенству Рао-Крамера (теорема 7.23):

$$D_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)}.$$

В силу утверждения 7.33 информация Фишера $I_n(\theta) = n \cdot I_1(\theta)$, тогда:

$$D_\theta[\hat{\theta}(\xi_1, \dots, \xi_n)] \geq \frac{[1 + b'(\theta)]^2}{I_n(\theta)} = \frac{[1 + b'(\theta)]^2}{n \cdot I_1(\theta)} = \frac{C(\theta)}{n},$$

где $C(\theta) = \frac{[1 + b'(\theta)]^2}{I_1(\theta)}$ – величина, зависящая только от θ не зависящая от n . Полученное неравенство показывает, что дисперсия всякой оценки в случае выборки не может убывать быстрее чем $1/n$.

Можно показать, что количества информации Фишера о параметре θ , содержащейся в исходной совокупности величин (ξ_1, \dots, ξ_n) и в достаточной статистике $S(\xi_1, \dots, \xi_n) = (S_1(\xi_1, \dots, \xi_n), \dots, S_k(\xi_1, \dots, \xi_n))$ одинаково. Действительно, представим функцию плотности вероятности (или вероятности) $L_\xi(x_1, \dots, x_n|\theta)$ по формуле умножения через условную плотность вероятности (или вероятность) $L_\theta(x_1, \dots, x_n|s_1, \dots, s_k)$ величин (ξ_1, \dots, ξ_n) относительно достаточной статистики $S = (S_1, \dots, S_k)$ и безусловной плотности вероятности

(или вероятности) $L_S(s_1, \dots, s_k | \theta)$ статистики (S_1, \dots, S_k) :

$$L_\xi(\xi_1, \dots, \xi_n | \theta) = L_\theta(\xi_1, \dots, \xi_n | S_1, \dots, S_k) \cdot L_S(S_1, \dots, S_k | \theta).$$

Прологарифмируем левую и правую части и вычислим производные левой и правой частей по θ :

$$\frac{\partial \ln L_\xi(\xi_1, \dots, \xi_n | \theta)}{\partial \theta} = \frac{\partial \ln L_\theta(\xi_1, \dots, \xi_n | S_1, \dots, S_k)}{\partial \theta} + \frac{\partial \ln L_S(S_1, \dots, S_k | \theta)}{\partial \theta}.$$

Поскольку статистика (S_1, \dots, S_k) является достаточной для параметра θ , то согласно определению условное распределение $L_\theta(x_1, \dots, x_n | s_1, \dots, s_k)$ не зависит от параметра θ и поэтому производная от логарифма $\ln L_\theta(x_1, \dots, x_n | s_1, \dots, s_k)$ равна нулю, тогда:

$$\frac{\partial \ln L_\xi(\xi_1, \dots, \xi_n | \theta)}{\partial \theta} = \frac{\partial \ln L_S(S_1, \dots, S_k | \theta)}{\partial \theta}.$$

Вычисляя вторые начальные моменты от случайных величин слева и справа, в итоге приходим к равенству:

$$M_\theta \left[\left(\frac{\partial \ln L_\xi(\xi_1, \dots, \xi_n | \theta)}{\partial \theta} \right)^2 \right] = D_\theta \left[\left(\frac{\partial \ln L_S(S_1, \dots, S_k | \theta)}{\partial \theta} \right)^2 \right],$$

в котором величина слева есть информация Фишера о параметре θ , содержащаяся в совокупности величин (ξ_1, \dots, ξ_n) , а величина справа есть информация Фишера о параметре θ , содержащаяся в достаточной статистике (S_1, \dots, S_k) .