Здравствуйте!

Математическая статистика

Задачи математической статистики в некотором смысле прямо противоположны задачам теории вероятностей. Если в теории вероятностей основной задачей является изучение вероятностных свойств случайных событий и случайных величин в предположении, что их характеристики частично или полностью известны, то в статистике производятся опыты и наблюдаются те значения, которые принимает случайная величина. Задачей же является получение сведений о параметрах, характеризующих распределение случайной величины или о характере самих распределений.

Условно математическую статистику можно разделить на теорию оценок и теорию проверки статистических гипотез.

Теория оценок. Постановка задачи

Пусть имеется случайная величина ξ (одномерная или многомерная), с плотностью вероятностей $p_\xi(x \mid \theta)$, которая известна нам **с точностью до параметров** $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$. Другими словами, нам известен вид $p_\xi(x \mid \theta)$, а вот входящие в неё параметры – нет. Наша задача – оценить эти неизвестные нам параметры.

Естественно, что для оценки этих параметров нужна какая-то информация, которую может дать только опыт. Пусть мы ставим опыт, в результате которого случайная величина ξ приняла значение x_1; ставим второй опыт, в результате которого случайная величина ξ приняла значение x_2; ….; ставим n-й опыт, в котором случайная величина ξ приняла значение x_n. Опытные данные

$$(x_1, x_2, \dots, x_n) = \vec{x}$$

носят название **выборки**, а число проделанных опытов n – **объёма выборки**.

Пусть неизвестный параметр θ всего лишь один. Что же мы должны получить от выборки? По выборке нам нужно построить одно число $\hat{\theta}$, которое бы было близко к истинному значению параметра θ. Это число называется **точечной оценкой** или просто **оценкой** параметра θ.

Итак, нам нужно сделать соответствие

$$(x_1, x_2, \ldots, x_n) \Rightarrow \text{число}\,\hat{\theta}.$$

В математике такое соответствие называется функцией, и поэтому $\hat{\theta}$ есть функция от $(x_1, x_2, \ldots, x_n)$, то есть

$$\hat{\theta} = T(x_1, x_2, \ldots, x_n).$$

Требования к оценкам

Для нахождения вида функции $T(x_1, x_2, \ldots, x_n)$ сначала сформулируем требования, предъявляемые к оценкам.

1. **Несмещённость оценки.**

Это требование заключается в том, чтобы в среднем оценка совпадала с истинным значением параметра θ, то есть чтобы

$$\mathrm{M}\{\hat{\theta}\} = \mathrm{M}\{T(x_1, x_2, \ldots, x_n)\} = \theta.$$

Такие оценки называются **несмещёнными**.

Пусть отдельные опыты независимы. Тогда

$$p_\xi(\vec{x} \mid \theta) = p_\xi(x_1, x_2, ..., x_n \mid \theta) = \prod_{i=1}^{n} p_\xi(x_i \mid \theta)$$

и условие несмещённости в явном виде выглядит так

$$\mathrm{M}\{\hat{\theta}\} = \int\limits_{-\infty}^{\infty} T(\vec{x}) p_\xi(\vec{x} \mid \theta) dx_1 dx_2 ... dx_n = \theta.$$

К сожалению, выполнить это условие удаётся не всегда. Величина

$$b(\theta, n) = \mathrm{M}\{\hat{\theta}\} - \theta$$

называется **смещением** оценки. Это смещение зависит от значения неизвестного параметра θ и объёма выборки n.

Если построить несмещённую оценку не удаётся, то обычно требуют, чтобы выполнялось условие

$$\lim_{n\to\infty} b(\theta, n) = 0.$$

Оценки, удовлетворяющие этому условию, называются **асимптотически несмещёнными**.

Итак, оценка должна быть несмещённой или, на худой конец, асимптотически несмещённой.

Чтобы сформулировать второе требование, мы должны рассмотреть вопрос о точности оценки. Естественно, что разные функции $T(x_1, x_2, ..., x_n)$ дают оценки различной точности, но чем измеряется точность оценки?

Эффективность оценки.

Естественно, что оценка, как функция от случайных значений $(x_1, x_2, ..., x_n)$, сама является случайной величиной и отличается от истинного значения параметра θ. Это отличие принято измерять так называемой **вариацией** оценки

$$V(\hat{\theta}) = \mathrm{M}\{(\hat{\theta} - \theta)^2\}.$$

В явном виде

$$V = \int_{-\infty}^{\infty} (T(\vec{x}) - \theta)^2 p_{\xi}(\vec{x} \mid \theta) dx_1 dx_2 ... dx_n.$$

Обратите внимание на отличие вариации оценки от её дисперсии

$$\mathrm{D}\{\hat{\theta}\} = \mathrm{M}\{(\hat{\theta} - \mathrm{M}\{\hat{\theta}\})^2\},$$

и на то, что для **несмещенной** оценки вариация совпадает с дисперсией.

Естественно, нам хотелось бы сделать вариацию оценки как можно меньше. В идеале хотелось бы получить $V(\hat{\theta}) = 0$. Но возможно ли это?

Оказывается, что нет. Существует некоторая нижняя граница для вариации оценки, меньше которой вариацию сделать нельзя. Эта нижняя граница даётся так называемом **неравенством информации**, которое мы сейчас и выведем.

Пусть наша оценка $\hat{\theta}$ несмещённая, то есть

$$\int_{-\infty}^{\infty} T(\vec{x}) p_{\xi}(\vec{x} \mid \theta) d\vec{x} = \theta.$$

(Под $d\vec{x}$ понимается комбинация $dx_1 dx_2 \ldots dx_n$). Дифференцируя это соотношение по θ, получим

$$\int_{-\infty}^{\infty} T(\vec{x}) \frac{\partial p_{\xi}(\vec{x} \mid \theta)}{\partial \theta} d\vec{x} = 1. \qquad (1)$$

С другой стороны мы имеем условие нормировки

$$\int_{-\infty}^{\infty} p_{\xi}(\vec{x} \mid \theta) d\vec{x} = 1,$$

дифференцируя которое по θ, получим

$$\int_{-\infty}^{\infty} \frac{\partial p_{\xi}(\vec{x} \mid \theta)}{\partial \theta} d\vec{x} = 0 \qquad (2)$$

Умножим (2) на θ и вычтем из (1):

$$\int_{-\infty}^{\infty} (T(\vec{x})-\theta)\frac{\partial p_{\xi}(\vec{x}\,|\,\theta)}{\partial\theta}d\vec{x}=1. \tag{3}$$

Преобразуем это выражение. Имеем

$$\frac{\partial \ln p_{\xi}(\vec{x}\,|\,\theta)}{\partial\theta}=\frac{1}{p_{\xi}(\vec{x}\,|\,\theta)}\cdot\frac{\partial p_{\xi}(\vec{x}\,|\,\theta)}{\partial\theta}.$$

Тогда (3) можно записать так

$$\int\limits_{-\infty}^{\infty}(T(\vec{x})-\theta)\frac{\partial p_{\xi}(\vec{x}\,|\,\theta)}{\partial\theta}d\vec{x}=\int\limits_{-\infty}^{\infty}(T(\vec{x})-\theta)\frac{\partial\ln p_{\xi}(\vec{x}\,|\,\theta)}{\partial\theta}\,p_{\xi}(\vec{x}\,|\,\theta)d\vec{x}=$$

$$=\mathrm{M}\left\{(T(\vec{x})-\theta)\frac{\partial\ln p_{\xi}(\vec{x}\,|\,\theta)}{\partial\theta}\right\}=1.$$

В отношении математического ожидания имеет место одно очень важное неравенство

$$\left(\mathrm{M}\{\xi\eta\}\right)^2\le\mathrm{M}\{\xi^2\}\mathrm{M}\{\eta^2\}.$$

Считая $T(\vec{x})-\theta$ за ξ и $\partial \ln p_\xi(\vec{x}\,|\,\theta)/\partial\theta$ за η, получим

$$\mathrm{M}\left\{(T(\vec{x})-\theta)\frac{\partial \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta}\right\}^2 = 1 \le$$

$$\le \mathrm{M}\{(T(\vec{x})-\theta)^2\}\cdot \mathrm{M}\left\{\left(\frac{\partial \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta}\right)^2\right\}. \qquad (4)$$

Но $\mathrm{M}\{(T(\vec{x})-\theta)^2\}$ есть вариация оценки $\hat{\theta}$. Величина

$$I = \mathrm{M}\left\{\left(\frac{\partial \ln p_{\xi}(\vec{x}\,|\,\theta)}{\partial\theta}\right)^2\right\}$$

называется **количеством информации** по Фишеру. Неравенство (4) можно записать так

$$V(\hat{\theta})I \geq 1,$$

откуда

$$V(\hat{\theta}) \geq \frac{1}{I}.$$

Это неравенство и называют **неравенством информации**.

Таким образом, вариацию несмещенной оценки нельзя сделать меньше величины

$$V_{\min} = \frac{1}{I}.$$

В связи с этим и появляется понятие **эффективности оценки**. Эффективностью оценки называется величина

$$eff(\hat{\theta}) = \frac{V_{\min}}{V(\hat{\theta})} = \frac{1}{IV(\hat{\theta})}.$$

Ясно, что $0 < eff(\hat{\theta}) \le 1$. Если $eff(\hat{\theta}) = 1$, то оценка называется **эффективной**. Если

$$\lim_{n \to \infty} eff(\hat{\theta}) = 1,$$

то оценка называется **асимптотически эффективной**.

Отсюда следует второе требование, предъявляемое к оценке: она должна быть эффективной или, по крайней мере, асимптотически эффективной.

Количество информации по Фишеру

Выше у нас появилась величина I – количество информации по Фишеру:

$$I = \mathrm{M}\left\{\left(\frac{\partial \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta}\right)^2\right\}.$$

Сейчас мы получим для неё другие выражения.

1. Так как

$$\frac{\partial \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta} = \frac{1}{p_\xi(\vec{x}\,|\,\theta)}\cdot\frac{\partial p_\xi(\vec{x}\,|\,\theta)}{\partial\theta},$$

то

$$\frac{\partial^2 \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2} = \frac{1}{p_\xi(\vec{x}\,|\,\theta)}\cdot\frac{\partial^2 p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2} - \frac{1}{p^2{}_\xi(\vec{x}\,|\,\theta)}\left(\frac{\partial p_\xi(\vec{x}\,|\,\theta)}{\partial\theta}\right)^2 =$$

$$= \frac{1}{p_\xi(\vec{x}\,|\,\theta)}\cdot\frac{\partial^2 p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2} - \left(\frac{\partial \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta}\right)^2.$$

Но

$$\mathrm{M}\left\{\frac{1}{p_\xi(\vec{x}\,|\,\theta)}\cdot\frac{\partial^2 p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2}\right\}=\int\limits_{-\infty}^{\infty}\frac{1}{p_\xi(\vec{x}\,|\,\theta)}\cdot\frac{\partial^2 p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2}\,p_\xi(\vec{x}\,|\,\theta)\,d\vec{x}=$$

$$=\int\limits_{-\infty}^{\infty}\frac{\partial^2 p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2}\,d\vec{x}=\frac{\partial^2}{\partial\theta^2}\left(\int\limits_{-\infty}^{\infty}p_\xi(\vec{x}\,|\,\theta)\,d\vec{x}\right)=0,$$

так как по условию нормировки внутренний интеграл равен 1. Итак

$$I=\mathrm{M}\left\{\left(\frac{\partial\ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta}\right)^2\right\}=-\mathrm{M}\left\{\frac{\partial^2\ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2}\right\}.$$

2. Пусть наши опыты независимы. Тогда

$$p_\xi(\vec{x}\,|\,\theta) = \prod_{i=1}^{n} p_\xi(x_i\,|\,\theta),$$

$$\ln p_\xi(\vec{x}\,|\,\theta) = \sum_{i=1}^{n} \ln p_\xi(x_i\,|\,\theta),$$

$$\frac{\partial^2 \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln p_\xi(x_i\,|\,\theta)}{\partial\theta^2}.$$

Усредняя и учитывая, что все x_i распределены одинаково, получим

$$I = -\mathrm{M}\left\{\frac{\partial^2 \ln p_\xi(\vec{x}\,|\,\theta)}{\partial\theta^2}\right\} = -n\,\mathrm{M}\left\{\frac{\partial^2 \ln p_\xi(x_i\,|\,\theta)}{\partial\theta^2}\right\} = nI_0,$$

$$I_0 = -\mathrm{M}\left\{\frac{\partial^2 \ln p_\xi(x_i\,|\,\theta)}{\partial\theta^2}\right\} = -\int\limits_{-\infty}^{\infty} \frac{\partial^2 \ln p_\xi(x\,|\,\theta)}{\partial\theta^2} p_\xi(x\,|\,\theta)\,dx.$$

I_0 – это количество информации, содержащееся в одном опыте (в одном измерении). Таким образом, $I = nI_0$, $V_{\min} = 1\,/\,nI_0$.

Случай многих параметров

Пусть теперь $\theta = (\theta_1, \theta_2, ..., \theta_s)$, то есть $p_\xi(x \mid \theta)$ зависит от s неизвестных параметров. Что изменится в этом случае?

Пусть мы снова имеем выборку $(x_1, x_2, ..., x_n)$ объёма n.

1. По ней нам придется строить s функций

$$T_i(x_1, x_2, ..., x_n) = \hat{\theta}_i, \quad i = \overline{1, s},$$

дающих оценки $\hat{\theta}_i$ неизвестных параметров θ_i. От них снова требуют несмещенности $\mathrm{M}\{\hat{\theta}_i\} = \theta_i$, или хотя бы асимптотической несмещенности.

2. Качество оценки измеряется теперь **матрицей вариаций V** размерности $s \times s$ с элементами

$$V_{ij} = \mathrm{M}\{(T_i(\vec{x}) - \theta_i)(T_j(\vec{x}) - \theta_j)\}.$$

Величины V_{ii} представляют собой вариации отдельных оценок.

Вместо количества информации I появляется информационная матрица **I** с элементами

$$I_{ij} = -\mathrm{M}\left\{\frac{\partial^2 \ln p_\xi(\vec{x} \mid \theta)}{\partial\theta_i \partial\theta_j}\right\} = nI_{ij}^{(0)},$$

$$I_{ij}^{(0)} = -\mathrm{M}\left\{\frac{\partial^2 \ln p_\xi(x \mid \theta)}{\partial\theta_i \partial\theta_j}\right\},$$

и нижняя граница для вариаций отдельных оценок даётся неравенством

$$V_{ii} \geq I_{ii}^{(-1)},$$

где $I_{ii}^{(-1)}$ – элемент матрицы, <u>**обратной матрице I**</u> (а не $1/\,I_{ii}$!).

Доверительные интервалы

То, о чём шла речь выше, называют точечными оценками, так как оценка $\hat{\theta}$ параметра θ представляла собой **число**. В статистике приняты ещё так называемые **интервальные оценки**, когда кроме $\hat{\theta}$ указывается ещё некоторый интервал $[\theta_1, \theta_2]$ и утверждается, что неизвестный параметр θ принадлежит этому интервалу.

Но можем ли мы, имея дело со случайными величинами, делать какие-то заключения, которые были бы верны с вероятностью 1? Разумеется, нет. Поэтому, указывая интервал $[\theta_1, \theta_2]$, в котором лежит неизвестный нам параметр, мы всегда должны указывать вероятность того, что мы говорим правду, или вероятность того, что мы ошиблись и на самом деле $\theta \notin [\theta_1, \theta_2]$.

Терминология. Интервал $[\theta_1, \theta_2]$, в котором, как мы утверждаем, лежит неизвестный параметр, называется **доверительным интервалом**; вероятность того, что мы не ошиблись – **доверительной вероятностью**; вероятность того, что мы ошибаемся – **доверительным уровнем.**

В принципе возможны два варианта:

1. задать интервал $[\theta_1, \theta_2]$ произвольно, но тогда мы обязаны сосчитать и указать доверительную вероятность;

2. задаться доверительной вероятностью (или доверительным уровнем) и найти доверительный интервал $[\theta_1, \theta_2]$.

В статистике принят второй вариант – задаются доверительным уровнем (его обычно обозначают 2α; доверительная вероятность равна, естественно, $1 - 2\alpha$) и ищут доверительный интервал $[\theta_1, \theta_2]$. Величина доверительного уровня стандартизована: обычно таблицы составлены для $2\alpha = 0{,}05$ (5%); 0,01 (1%) и 0,001 (0,1%). Реже используют $2\alpha = 0{,}1$ (10%). Выбор того или иного доверительного уровня связан с важностью проводимых исследований, точнее – с теми потерями, которые понесёт общество, если мы ошиблись в своих выводах.

Построение доверительных интервалов основано на так называемом принципе «практической невозможности маловероятных событий», который гласит, что если вероятность наступления события мала, то можно считать, что оно не наступит, и вести себя так, как будто это событие было бы невозможным событием. С этих позиций величина 2α и есть та граница, которая отделяет события практически невозможные от событий практически возможных. Если вероятность наступления события меньше 2α, то это событие можно считать практически невозможным и вести себя так, как если бы это событие никогда не наступит.

Пусть мы имеем оценку $\hat{\theta}$, естественно, зависящую от выборки $\vec{x} = (x_1, x_2, ..., x_n)$. В принципе, можно найти $p(\hat{\theta} \mid \theta)$, пользуясь методиками вычисления плотности вероятностей функций от случайных величин.

Возьмём доверительный уровень 2α и разделим его на две равные части $2\alpha = \alpha + \alpha$. Найдём $\hat{\theta}_{\text{н}}$ из условия

$$P\{\hat{\theta} < \hat{\theta}_{\text{н}}\} = \int_{-\infty}^{\hat{\theta}_{\text{н}}} p(\hat{\theta} \mid \theta) d\theta = \alpha$$

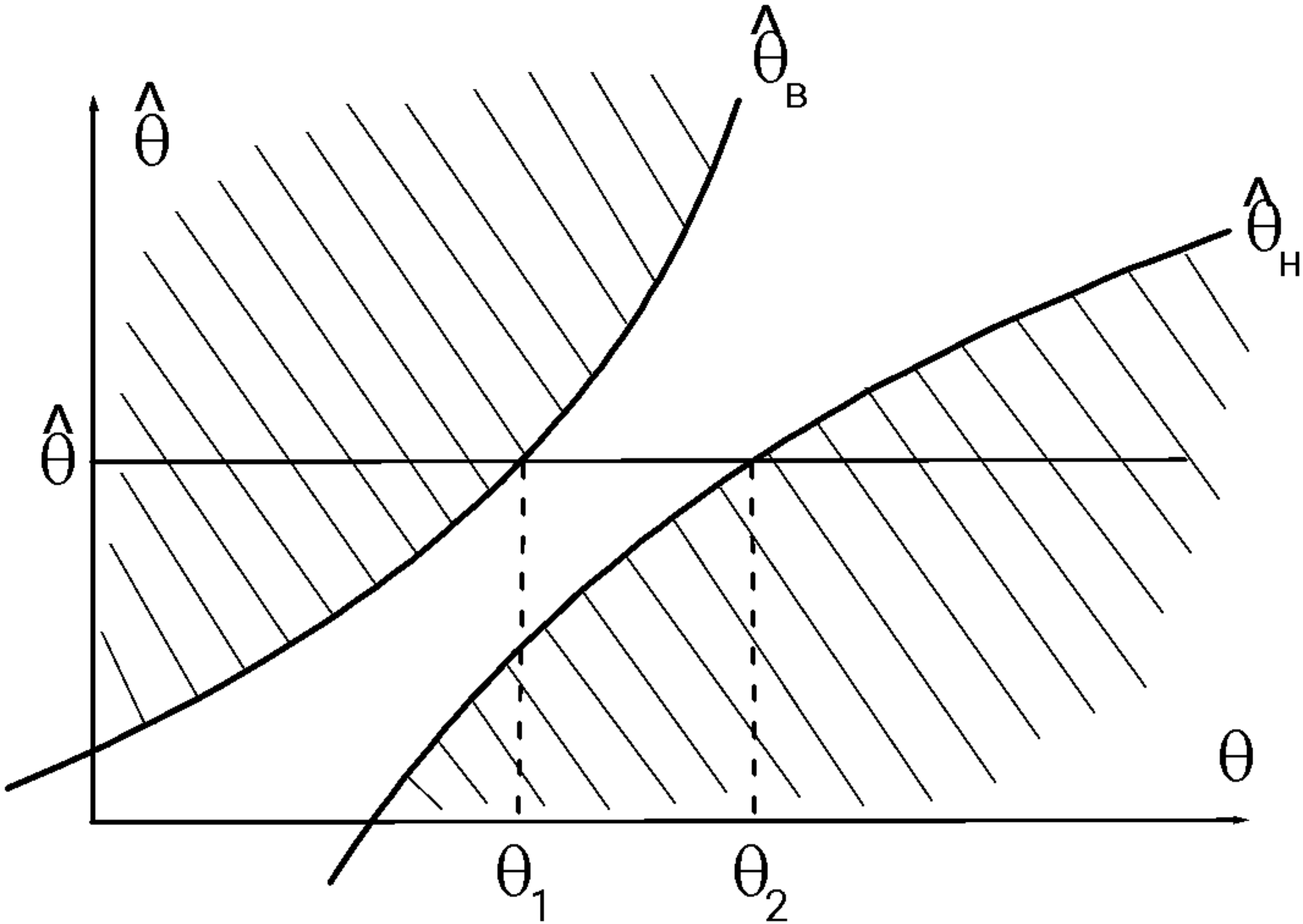
Тогда это будет событие практически невозможное.

Аналогично, найдём $\hat{\theta}_{\text{в}}$ из условия

$$P\{\hat{\theta} > \hat{\theta}_{\text{в}}\} = \int_{\hat{\theta}_{\text{в}}}^{\infty} p(\hat{\theta} \mid \theta) d\theta = \alpha .$$

Это будет также практически невозможное событие.

Событие, заключающееся в том, что $(\hat{\theta} < \hat{\theta}_{\text{н}}) \cup (\hat{\theta} > \hat{\theta}_{\text{в}})$ имеет вероятность 2α и поэтому также является практически невозможным событием.



$\hat{\theta}_B$
$\hat{\theta}$
$\hat{\theta}_H$
$\hat{\theta}$
θ
θ_1
θ_2

На рисунке заштрихована область практически невозможных событий.

Пусть теперь мы по выборке $\vec{x} = (x_1, x_2, \ldots, x_n)$ нашли оценку $\hat{\theta}$. Каким значениям параметра θ она может соответствовать?

Легко догадаться, глядя на рисунок, что если считать события, имеющие вероятность наступления 2α, практически невозможными и, поэтому, не наступающими, то значения параметра θ, при которых наступившее событие будет практически возможным, будут лежать в интервале $[\theta_1, \theta_2]$. Это и будет доверительный интервал для параметра θ, соответствующий доверительному уровню 2α.