УЧЕНЫЕ ЗАПИСКИ КАЗАНСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
*Том 152, кн. 1* Физико-математические науки *2010*

УДК 517.5+517.9

# АСИМПТОТИЧЕСКИ НАИБОЛЕЕ ТОЧНЫЕ ДВУСТОРОННИЕ ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ СРЕДНЕГО В НОРМАЛЬНО-НОРМАЛЬНОЙ МОДЕЛИ

*Р.Ф. Салимов, С.В. Симушкин*

**Аннотация**

В статье приводится вид наиболее точного семейства доверительных интервалов для нормально-нормальной модели в $d$-апостериорном подходе. Устанавливается его асимптотическое поведение с ростом объема выборки. Предлагается способ построения асимптотически наиболее точного доверительного семейства.

**Ключевые слова:** нормально-нормальная модель, $d$-апостериорные вероятности ошибок, доверительные двусторонние интервалы.

---

## Введение

Рассматривается задача построения «доверительного множества» для неизвестного параметра $\theta$ в ситуации, когда значение $\theta$ есть реализация случайной величины $\vartheta$ с полностью известным априорным распределением. Обычно в такой ситуации строят так называемые «байесовские доверительные множества» $\mathcal{B}$, то есть множества, апостериорная вероятность которых больше некоторой фиксированной надежности $1-\alpha$. Если $\mathcal{B}$ используется в дальнейшем для проверки гипотез, что является обычной практикой, то можно гарантировать, что априорная средняя ошибка будет не превосходить $\alpha$. В большинстве практических приложений эта характеристика является слишком «обобщающей» и зачастую не может удовлетворить исследователя. Один из возможных способов решения этой проблемы основан на классическом (ортодоксальном) подходе к определению рисков статистических процедур. В этом подходе доверительное множество устроено так, что критерий, отвергающий гипотезу, когда $(1-\alpha)$-доверительное множество попадает в область альтернативы, будет автоматически удовлетворять заданным ограничениям $\alpha$ на вероятность ошибки первого рода.

В работах [1, 2] была развита методология $d$-апостериорного подхода, в котором вместо классических вероятностей ошибок первого и второго рода рассматривались вероятности справедливости альтернативы (гипотезы) при условии, что статистический эксперимент закончился принятием (отвержением) проверяемой гипотезы. В рамках этого подхода в статье [2] был предложен способ построения наиболее точных доверительных множеств (называемых Б-доверительными), согласованный с задачей проверки гипотезы. Другими словами, если гипотеза принимается, когда Б-доверительное множество надежности $1-\alpha$ полностью попадает в область гипотезы, то вероятность $d$-апостериорной ошибки первого рода (средняя доля ошибки среди экспериментов, закончившихся принятием гипотезы) такого критерия не будет превосходить $\alpha$. В этой же статье было установлено, что для вероятностных моделей с монотонным отношением правдоподобия верхние Б-доверительные границы, как и классические границы, могут быть описаны посредством единственной статистики $\overline{\theta}$.

В настоящей статье рассматривается задача построения семейств двусторонних доверительных интервалов в рамках $d$-апостериорного подхода. Отличительной особенностью этой задачи является то, что семейство таких интервалов принципиально не может быть описано парой статистик $(\underline{\theta};\overline{\theta})$. Аналогичная проблема возникает и при построении байесовских доверительных интервалов, однако в этом случае среди множества всех байесовских интервалов часто можно выбрать в некотором смысле самый узкий, надежность которого будет также равна $1-\alpha$. При $d$-апостериорном подходе доверительные интервалы вида $(\underline{\theta};\overline{\theta})$ обладают избыточной надежностью (в частности, как будет показано ниже, с ростом объема выборки надежность таких интервалов будет приближаться к 1).

Здесь мы рассматриваем задачу построения семейств двусторонних доверительных интервалов для нормальной вероятностной модели наблюдений с неизвестным параметром среднего $\theta$ и известной дисперсией $\sigma^2$. Распределение параметра $\vartheta$ также предполагается нормальным с известными средним $\mu$ и дисперсией $\tau^2$. Для этой модели находится вид Б-доверительного семейства интервалов, минимизирующего точность семейства – $d$-апостериорную вероятность ошибки 2-го рода. Показывается, что с ростом объема выборки точность этого семейства стремится к нулю, и устанавливается её скорость сходимости. Кроме того, находится асимптотический вид наиболее точного доверительного семейства, предлагается вариант построения асимптотически наиболее точных семейств. Аналогичное исследование было проведено в статье [2] для верхней Б-доверительной границы.

## 1. Определения и вспомогательные результаты

Определения и результаты, относящиеся к общей модели статистического вывода, приведены в [1]. Здесь мы переформулируем эти утверждения применительно к заявленной во введении нормально-нормальной модели.

В эксперименте наблюдается выборка $(X_1,\dots,X_n)$ из распределения с нормальной плотностью

$$\frac{1}{\sigma}\varphi\left(\frac{x-\theta}{\sigma}\right)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\},\quad x\in\mathcal{R}^1.$$

Параметр $\theta$ есть реализация случайной величины $\vartheta$ с нормальной плотностью $\tau^{-1}\varphi\left((\theta-\mu)/\tau\right)$, среднее значение $\mu$ и дисперсия $\tau^2$ которой считаются известными. Дисперсия выборки $\sigma^2$ также известна. Поскольку в настоящей работе параметры $\mu,\ \sigma^2,\ \tau^2$ не играют существенной роли, то в целях сокращения записи мы будем излагать все результаты в частном случае $\mu=0,\ \sigma^2=1,\ \tau^2=1$.

В рассматриваемой модели выборочный вектор может быть редуцирован до значения достаточной статистики $T(X_1,\dots,X_n)=\overline{X}=\sum\limits_{i=1}^{n}X_i/n$. Как известно, условное (при фиксированном $\theta$) распределение этой статистики также нормально $(\theta,n^{-1})$, а безусловное (априорное) распределение нормально $(\mu,(n+1)/n)$. Совместное распределение вектора $(T,\vartheta)$ будем обозначать символом $\mathbf{P}$.

С каждым выборочным значением статистики $T=t$ свяжем семейство двусторонних интервалов $\mathcal{D}(t)=\{[\theta_1,\theta_2],\ \theta_1,\theta_2\in\mathcal{R}^1\}$, интерпретируемое как семейство интервалов, которым исследователь может доверять. Другими словами, для любых фиксированных значений $A<B$ гипотеза $\mathbf{H}:\theta\in[A,B]$ принимается только тогда, когда интервал $[A,B]$ попадает в семейство $\mathcal{D}$.

**Определение 1.** Условную вероятность

$$\mathcal{Q}(\mathcal{D}\,|\,A,B):=\mathbf{P}\left\{A\leqslant\vartheta\leqslant B\,/\,\mathcal{D}(T)\ni[A,B]\right\}$$

как функцию $A,\ B$ назовем *надежностью* семейства $\mathcal{D}$, а условную вероятность

$$\mathcal{A}(\mathcal{D}\,|\,A,B) := \mathbf{P}\left\{A \leqslant \vartheta \leqslant B\,/\,\mathcal{D}(T) \not\ni [A,B]\right\}$$

назовем *точностью* этого семейства.

В случае, если вероятность $\mathbf{P}\left\{\mathcal{D}(T) \ni [A,B]\right\} = 0$, мы полагаем надежность равной 1. Соответственно, точность полагается равной 0 при нулевой вероятности условия.

**Определение 2.** Семейство интервалов $\mathcal{D}$ назовем *доверительным* (уровня $1-\alpha$), если для всех $A<B$ его надежность

$$\mathcal{Q}(\mathcal{D}\,|\,A,B) \geqslant 1-\alpha.$$

Доверительное семейство $\mathcal{D}^*$ называется *наиболее точным*, если его точность $\mathcal{A}(\mathcal{D}^*|A,B)$ минимальна при всех $A<B$ среди всех доверительных семейств $\mathcal{D}$ таких, что вероятность $\mathbf{P}\left\{\mathcal{D}(T) \ni [A,B]\right\} > 0$.

Ранее (см. [1]) одним из авторов было установлено, что оптимальный (с точки зрения $d$-апостериорных ошибок 1-го и 2-го рода) критерий проверки гипотезы $\mathbf{H}:\theta\in\Theta_0$ основан на статистике

$$L(\Theta_0\,|\,T) = \mathbf{P}\left\{\vartheta\in\Theta_0\,/\,T\right\},$$

апостериорной вероятности справедливости гипотезы $\mathbf{H}$ при фиксированном значении вектора наблюдений $T$. Основываясь на этом утверждении, предлагается следующий способ построения наиболее точных доверительных семейств.

Если выборочное значение статистики $T=t$, то интервал $[A,B]$ включается в $\mathcal{D}_n^*(t)$, если

$$\mathbf{P}\left\{\vartheta\in[A,B]\,/\,L(\,[A,B]\,|\,T) \geqslant L(\,[A,B]\,|\,t)\right\} \geqslant 1-\alpha.$$

Для того чтобы описать это семейство более конструктивно, введем функцию (называемую в дальнейшем *апостериорной надежностью*)

$$\mathfrak{I}_n(A,B\,|\,t) := \frac{\int\limits_A^B \Big[\Phi\Big(\sqrt{n}\big(2\,\mathcal{Z}_n - t - \theta\big)\Big) - \Phi\Big(\sqrt{n}\big(t-\theta\big)\Big)\Big]\,\varphi\,(\theta)\;d\theta}{\Phi\Big((2\mathcal{Z}_n - t)\zeta_n\Big) - \Phi\Big(t\,\zeta_n\Big)}, \tag{1}$$

где $\zeta_n^2 = n/(n+1)$, $\mathcal{Z}_n = (A+B)/2\zeta_n^2$, $\Phi(\cdot)$ – стандартная нормальная функция распределения. Чтобы устранить неопределенность в точке $t=\mathcal{Z}_n$, можно положить по непрерывности (см. ниже лемму 2)

$$\mathfrak{I}_n(A,B\,|\,\mathcal{Z}_n) := 1 - 2\Phi\Big(-\sqrt{n+1}\,\delta\Big).$$

**Теорема 1.** *Наиболее точное семейство имеет вид*

$$\mathcal{D}_n^*(t) = \{[A;B]:\ \mathfrak{I}_n(A,B\,|\,t)\ \geqslant 1-\alpha\}.$$

*Надежность этого семейства для любых* $A<B$ *равна*

$$\mathcal{Q}(\mathcal{D}_n^*\,|\,A,B) = 1-\alpha.$$

Доказательство этой теоремы, а также следующих теорем 2, 3 приведено в [2].

Наиболее точное семейство может быть описано посредством задания «граничного» семейства интервалов

$$\{[A;B]:\, \mathfrak{J}_n(A,B\,|\,t)=1-\alpha\}. \tag{2}$$

Достаточность такого описания вытекает из следующего утверждения.

**Теорема 2.** *Пусть* $\theta_c=(A+B)/2$ *– центр интервала* $[A,B]$, *константа* $\delta^*=\delta^*(\theta_c)$ *такая, что*

$$\mathfrak{J}_n(\theta_c-\delta^*,\theta_c+\delta^*\,|\,t)=1-\alpha.$$

*Тогда интервал* $[A,B]$ *входит в наиболее точное семейство* $\mathcal{D}_n^*$, *если его ширина* $(B-A)>2\delta^*$.

Существование константы $\delta^*$ следует из утверждения $(iii)$ приводимой ниже леммы 1.

Функцию апостериорной надежности как функцию середины интервала $\theta_c = (A+B)/2$ и его половинной ширины $\delta=(B-A)/2$ будем обозначать тем же символом:

$$\mathfrak{J}_n(\theta_c,\delta\,|\,t)=\mathfrak{J}_n(\theta_c-\delta,\theta_c+\delta\,|\,t).$$

## 2. Свойства функции апостериорной надежности

Для дальнейшего нам понадобятся некоторые свойства функции апостериорной надежности $\mathfrak{J}_n$ (доказательства утверждений настоящего и следующего пунктов приведены в п. 4).

Сначала рассмотрим её как функцию ширины интервала $\delta$.

**Лемма 1.** *При любых фиксированных* $n,\ \theta_c,\ t$
(i) $\mathfrak{J}_n(\theta_c,\delta\,|\,t)$ *есть непрерывная, возрастающая функция* $\delta>0$;
(ii) $\sup_\delta \mathfrak{J}_n(\theta_c,\delta\,|\,t)=1,\ \inf_\delta \mathfrak{J}_n(\theta_c,\delta\,|\,t)=0$;
(iii) *существует единственная точка* $\delta_n^*=\delta_n^*(\theta_c,t)$ *такая, что*

$$\mathfrak{J}_n(\theta_c,\delta_n^*\,|\,t)=1-\alpha.$$

В следующей лемме устанавливаются свойства $\mathfrak{J}_n$ как функции результата статистического эксперимента $t$.

**Лемма 2.** *При любых фиксированных* $n,\theta_c,\delta$ *функция* $\mathfrak{J}_n(\theta_c,\delta\,|\,t)$
(i) *симметрична по* $t$ *относительно точки* $\mathcal{Z}_n=(1+1/n)\,\theta_c$:

$$\mathfrak{J}_n(\theta_c,\delta\,|\,t)=\mathfrak{J}_n(\theta_c,\delta\,|\,2\mathcal{Z}_n-t);$$

(ii) *непрерывна по* $t$;
(iii) *возрастает при* $t<\mathcal{Z}_n$ *и убывает при* $t>\mathcal{Z}_n$;
(iv) *достигает максимума при* $t=\mathcal{Z}_n$:

$$\max_t \mathfrak{J}_n(\theta_c,\delta\,|\,t)=\mathbf{P}\left\{\theta_c-\delta\leqslant\vartheta\leqslant\theta_c+\delta\,/\,T=\mathcal{Z}_n\right\}=1-2\Phi\Big(-\sqrt{n+1}\,\delta\Big);$$

(v) $\min\limits_t \mathfrak{J}_n(\theta_c,\delta\,|\,t)=\mathbf{P}\left\{\theta_c-\delta\leqslant\vartheta\leqslant\theta_c+\delta\right\}=\Phi(\theta_c+\delta)-\Phi(\theta_c-\delta).$

Заметим, что интервалы $[A, B]$, вероятность которых не больше $1-\alpha$:

$$\mathbf{P}\{A \leqslant \vartheta \leqslant B\} = \Phi(B) - \Phi(A) \geqslant 1-\alpha,$$

всегда (независимо от результатов наблюдений) будут включаться в наиболее точное доверительное семейство. Поэтому мы исключим такие интервалы из нашего рассмотрения.

**Теорема 3.** *Если $\theta_c$, $\delta$ такие, что*

$$\Phi(\theta_c+\delta) - \Phi(\theta_c-\delta) < 1-\alpha < 1 - 2\Phi\Big(-\sqrt{n+1}\,\delta\Big),$$

*то:*

(i) *существует единственная точка $\overline{t}_n = \overline{t}_n(\theta_c, \delta) > \mathcal{Z}_n$ такая, что*

$$\mathfrak{I}_n(\theta_c, \delta \,|\, \overline{t}_n) = 1-\alpha;$$

(ii) *интервал $[\theta_c-\delta, \theta_c-\delta]$ включается в наиболее точное доверительное семейство $\mathcal{D}_n^*$ в том и только в том случае, когда выборочное значение статистики $T=t$ удовлетворяет неравенствам*

$$2\mathcal{Z}_n - \overline{t}_n \leqslant t \leqslant \overline{t}_n.$$

Таким образом, изучение асимптотического поведения наиболее точного доверительного семейства сводится к изучению асимптотик для $\delta_n^*$ и $\overline{t}_n$.

**2.1. Асимптотическое поведение апостериорной надежности.** Как показывают следующие утверждения, в отличие от классических доверительных интервалов, наиболее точное доверительное семейство в $d$-апостериорном подходе не вырождается даже при $n\to\infty$.

Следующее утверждение устанавливается простым предельным переходом под знаком интеграла.

**Лемма 3.** *При любых фиксированных $\delta > 0$, $\theta_c$, $t$*

$$\lim_{n\to\infty} \mathfrak{I}_n(\theta_c, \delta \,|\, t) = \mathfrak{I}_\infty(\theta_c, \delta \,|\, t) := \begin{cases} \dfrac{\Phi(\theta_c+\delta) - \Phi(\theta_c-\delta)}{|\Phi(2\theta_c - t) - \Phi(t)|}, & \textit{если}\ \ t \notin [\theta_c-\delta; \theta_c+\delta], \\ 1, & \textit{если}\ \ t \in [\theta_c-\delta; \theta_c+\delta]. \end{cases}$$

Определим константу $0 < \delta^* = \delta^*(\theta_c, t) < |\theta_c - t|$ (очевидно, всегда существующую и единственную) как решение уравнения

$$\frac{\Phi(\theta_c+\delta^*) - \Phi(\theta_c-\delta^*)}{|\Phi(2\theta_c - t) - \Phi(t)|} = 1-\alpha,$$

полагая $\delta^* = 0$ при $t = \theta_c$. Семейство

$$\mathcal{B}^* = \mathcal{B}^*(t) = \big\{[\underline{\theta}, \overline{\theta}] = [\theta_c-\delta, \theta_c+\delta] : -\infty < \theta_c < \infty,\ \delta \geqslant \delta^*(\theta_c, t)\big\}$$

образует асимптотическое семейство доверительных интервалов при выборочном значении $T=t$.

Ниже (см. рис. 1) приведён пример 95%-доверительного семейства при $t=2$.

На этом рисунке наклонная прямая, проходящая через начало координат, есть линия середин интервалов $\theta_c$. Окаймляющие штрих-пунктирные линии – границы

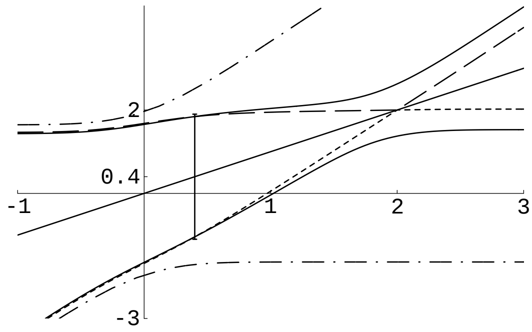


Рис. 1. Семейства 95%-доверительных интервалов ($T = 2,\ n = 10,\ n = \infty$)

интервалов, которые могут быть включены в доверительное семейство без проведения наблюдений ($n = 0$). Две сплошные линии – семейство верхних и нижних границ интервалов при объеме выборки $n = 10$, внутренние пунктирные линии – верхняя и нижняя границы асимптотических интервалов ($n = \infty$). В качестве примера показан интервал $[A, B] = [-1.1, 1.9]$, попадающий в оба семейства, поскольку половинная ширина этого интервала $\delta = 1.5$ превышает соответствующие значения для наиболее точного и асимптотически наиболее точного интервалов с той же серединой $\theta_c = 0.4$:

$$\delta_n^* = 1.434, \quad \delta^* = 1.444.$$

Отметим, что наиболее точный интервал при конечном объеме выборки не всегда шире асимптотического интервала.

Примечательно также, что при бесконечном объеме выборки, когда значение статистики $T$, по-существу, совпадает с истинным значением параметра $\theta$, в доверительное семейство попадают интервалы, не содержащие это истинное значение. Этот факт становится вполне естественным, если взглянуть на него с точки зрения точности и надежности семейства. Действительно, можно показать, что если интервал $[A, B]$ не попадает в «предельное» наиболее точное семейство, то этот интервал автоматически не будет содержать наблюдаемое значение $t = \theta$. Поэтому

$$\mathbf{P}\left\{\vartheta \in [A, B] / \mathcal{B}^* \not\ni [A, B]\right\} \leqslant \frac{\mathbf{P}\left\{(\vartheta \in [A, B]) \cap (\vartheta \notin [A, B])\right\}}{\mathbf{P}\left\{\mathcal{B}^* \not\ni [A, B]\right\}} = 0,$$

что ожидаемо при $n = \infty$. С другой стороны, если интервал включается в доверительное семейство только тогда, когда этот интервал содержит истинное значение $\theta$, то надежность такого семейства будет равна 1, то есть семейство излишне надежно.

## 3. Асимптотические свойства наиболее точного доверительного семейства

Изучение асимптотических свойств наиболее точного доверительного семейства начнем с установления скорости сходимости оптимальной ширины доверительного интервала $\delta_n^*$ к ее предельному значению $\delta^*$.

**Лемма 4.** *Для любых* $\theta_c,\ t$

$$\lim_{n \to \infty} \delta_n^* = \delta^*.$$

**Теорема 4.**

(I) *Для любых* $\theta_c \neq t$ *таких, что*

$$Q(\theta_c, t) := \frac{(1-\alpha)|(2\theta_c + t)\varphi(2\theta_c - t) + t\varphi(t)|}{2(\varphi(\theta_c + \delta^*) + \varphi(\theta_c - \delta^*))} \neq 0,$$

*справедливо асимптотическое равенство*

$$|\delta_n^* - \delta^*| \asymp \frac{Q(\theta_c, t)}{n}, \ n \to \infty.$$

(II) *Пусть* $\theta_c = t$, *тогда*

$$\delta_n^* \asymp \frac{z_\alpha}{\sqrt{n}}, \ n \to \infty,$$

*где* $z_\alpha = \Phi^{-1}(1-\alpha/2)$ *– верхняя* $\dfrac{\alpha}{2}$*-квантиль стандартного нормального закона.*

Заметим, что константа $Q(\theta_c, t)$ принимает нулевые значения не более чем в двух точках $\theta_c$ для каждого значения $t$. Специальное исследование для этих точек мы здесь не проводим.

Как показано в теореме 3, наиболее точное семейство содержит интервал $[\theta_c - \delta, \theta_c + \delta]$, если выборочное значение статистики $T = t$ удовлетворяет неравенствам $2\mathcal{Z}_n - \overline{t}_n \leqslant t \leqslant \overline{t}_n$. Аналогичное утверждение справедливо и для асимптотического семейства. А именно, пусть $t^* > \theta_c + \delta$ – единственное решение уравнения

$$\frac{\Phi(\theta_c + \delta) - \Phi(\theta_c - \delta)}{\Phi(t^*) - \Phi(2\theta_c - t^*)} = 1 - \alpha.$$

Тогда $[\theta_c - \delta, \theta_c + \delta] \in \mathcal{B}^*(t)$ в том и только в том случае, когда $2\theta_c - t^* \leqslant t \leqslant t^*$.

**Теорема 5.** *Для любых* $\theta_c$ *и* $\delta > 0$ *справедливо асимптотическое равенство*

$$|\overline{t}_n - t^*| \asymp \frac{\tau(\theta_c, \delta)}{n}, \quad n \to \infty,$$

*с константой* $\tau(\theta_c, \delta) = [t^*\varphi(t^*) - (2\theta_c - t^*)\varphi(2\theta_c - t^*)] \cdot [\varphi(t^*) + \varphi(2\theta_c - t^*)]^{-1}/2$.

Рассмотрим теперь произвольную числовую последовательность $q_n \to 0$ при $n \to \infty$ и определим семейство

$$\mathcal{B}_n = \{[\underline{\theta}, \overline{\theta}] = [\theta_c - \delta, \theta_c + \delta] : -\infty < \theta_c < \infty, \, \delta \geqslant \delta^* + q_n\}.$$

Как показано в следующей теореме, это семейство асимптотически надежно при любом выборе последовательности $q_n$. Точность этого семейства будет по порядку совпадать с точностью наиболее точного семейства лишь при $q_n = c_n/n$, где последовательность $c_n \to c_0$, причем константа $c_0$ может равняться нулю.

**Теорема 6.**

(I) *Семейство* $\mathcal{B}_n$ *асимптотически надежно:*

$$\lim_{n\to\infty} \mathcal{Q}(\mathcal{B}_n \,|\, [A, B]) = 1 - \alpha \quad (\forall A < B).$$

(II) *Если* $q_n = O_c\left(\dfrac{1}{n}\right)$, *то для любых* $A < B$

$$0 < \lim_{n\to\infty} \frac{\mathcal{A}\big(\mathcal{B}_n \,|\, [A; B]\big)}{\mathcal{A}\big(\mathcal{D}_n^* \,|\, [A; B]\big)} < \infty.$$

Заметим, что если последовательность $q_n$ сходится к нулю медленнее, чем $\dfrac{1}{n}$, то точность семейства повышается, однако достигается это, конечно, за счёт уменьшения надежности.

Доказательство теоремы 6 опирается на следующее утверждение относительно скорости сходимости к нулю интегралов, определяющих вероятности отвержения гипотезы $\theta \in [A, B]$ при использовании доверительных семейств интервалов, подобных наиболее точному или асимптотически наиболее точному семействам.

**Лемма 5.** *Если* $t > B$ *и последовательность* $r_n = c_n/n$, *где* $c_n \to c_0$, *то*

$$\int\limits_A^B \Phi\big(\sqrt{n}(\theta - t - r_n)\big)\varphi(\theta)\, d\theta \asymp \frac{\varphi(B)e^{-c_0(t-B)}}{\sqrt{2\pi}(t-B)^2}\, \frac{e^{-(t-B)^2 n/2}}{n\sqrt{n}}, \quad n \to \infty.$$

Утверждение леммы остается верным и для $r_n \equiv c_0 = 0$.

## 4. Доказательства

Утверждения леммы 1 вполне очевидны кроме, может быть, свойства (ii), для доказательства которого достаточно заметить, что числитель $\mathfrak{J}_n$ есть вероятность

$$\mathbf{P}\left\{(\vartheta \in [A, B]) \cap (t \leqslant T \leqslant 2\mathcal{Z}_n - t)\right\},$$

а знаменатель отличается от него лишь тем, что в качестве интервала $[A, B]$ взята вся числовая прямая. Поэтому если ширина интервала $\delta \nearrow \infty$, то функция апостериорной надежности $\mathfrak{J}_n \nearrow 1$.

**Доказательство Леммы 2.**

Симметрия и непрерывность функции $\mathfrak{J}_n$ сразу следуют из её представления.

Для доказательства остальных утверждений заметим сначала, что функция $\mathfrak{J}_n$ есть ни что иное, как условная вероятность (с $\mathcal{L}(T) = \mathbf{P}\left\{\vartheta \in [A, B]/T\right\}$)

$$\mathfrak{J}_n(A, B\,|\,t) = \mathbf{P}\left\{\vartheta \in [A, B]/\mathcal{L}(T) \geqslant p\right\} \tag{3}$$

с константой $p$, равной экспериментальному значению апостериорной вероятности

$$\mathcal{L}(t) = \mathbf{P}\left\{\vartheta \in [A, B]/T = t\right\} = \Phi\Big(\frac{B - \mathfrak{M}_n}{\mathfrak{S}_n}\Big) - \Phi\Big(\frac{A - \mathfrak{M}_n}{\mathfrak{S}_n}\Big),$$

где апостериорные среднее и дисперсия (см., например, [3, с. 341]) имеют вид

$$\mathfrak{M}_n = \frac{t\,\tau^2 + \mu\sigma^2 n^{-1}}{\tau^2 + \sigma^2 n^{-1}} = \frac{t}{1 + n^{-1}}, \quad \mathfrak{S}_n^2 = \frac{\tau^2\sigma^2 n^{-1}}{\tau^2 + \sigma^2 n^{-1}} = \frac{1}{n+1}.$$

Хорошо известно, что для нормального закона вероятность попадания в конечный интервал есть убывающая функция расстояния между центром распределения $\mathfrak{M}_n$ и серединой интервала $\theta_c$. В работе [1] показано, что вероятность (3) есть возрастающая функция $p$, поэтому при $t > (1 + \dfrac{1}{n})\theta_c = \mathcal{Z}_n$ значение константы $p = L([A, B]\,|\,t)$ будет уменьшаться с ростом $t$, а вместе с ней будет уменьшаться и $\mathfrak{J}_n$.

По свойству условного математического ожидания

$$\mathbf{P}\left\{\vartheta \in [A, B]/\mathcal{L}(T) \geqslant p\right\} = \mathbf{E}\left\{\,\mathbf{P}\left\{\vartheta \in [A, B]/T\right\}/\mathcal{L}(T) \geqslant p\,\right\} \;\geqslant\; p.$$

Следовательно, справедливы неравенства (с $p = L(\,[A,B]\,|\,t)\,)$

$$\mathcal{L}(t) \leqslant \mathfrak{J}_n(A,B;t) \leqslant \max_t \mathcal{L}(t).$$

Утверждение (iv) леммы будет следовать из этих неравенств, если выбрать $t$ так, чтобы апостериорное среднее совпало с центром интервала $\theta_c$.

Утверждение (v) следует из того, что при $t \to \pm\infty$ функция $\mathcal{L}(t) \to 0$, а событие $\{\mathcal{L}(T)) \geqslant \mathcal{L}(t)\}$ стремится к достоверному событию, поэтому условная вероятность

$$\mathfrak{J}_n(A,B\,|\,t) = \mathbf{P}\left\{\vartheta \in [A,B]\,/\,\mathcal{L}(T) \geqslant \mathcal{L}(t)\right\} \stackrel{t\to\infty}{\longrightarrow} \mathbf{P}\left\{\vartheta \in [A,B]\right\}.$$

□

**Доказательство Леммы 3.**
Очевидно, знаменатель функции $\mathfrak{J}_n$

$$\Phi\left(\frac{2\mathcal{Z}_n - t}{\sqrt{1+n^{-1}}}\right) - \Phi\left(\frac{t}{\sqrt{1+n^{-1}}}\right) \to \Phi(2\theta_c - t) - \Phi(t).$$

По теореме Лебега об ограниченной сходимости (подынтегральная функция не превосходит функции $\varphi(x)$) законен переход к пределу по $n \to \infty$ в интеграле, определяющем функцию $\mathfrak{J}_n$.

Учитывая симметрию функции $\mathfrak{J}_n$, рассмотрим только случай $t < \theta_c$. Пусть $t < \theta_c - \delta$, тогда $2\mathcal{Z}_n - t > \theta_c + \delta$ (начиная с некоторого $n$, так как $\mathcal{Z}_n \to \theta_c$), поэтому при всех $\theta \in [\theta_c - \delta, \theta_c + \delta]$

$$\Phi(\sqrt{n}(2\mathcal{Z}_n - t - \theta)) \to 1, \quad \Phi(\sqrt{n}(t-\theta)) \to 0\,.$$

Следовательно, числитель $\mathfrak{J}_n$,

$$\int\limits_{\theta_c-\delta}^{\theta_c+\delta} [\Phi(\sqrt{n}(2\mathcal{Z}_n - t - \theta)) - \Phi(\sqrt{n}(t-\theta))]\varphi(\theta)\,d\theta \to \Phi(\theta_c+\delta) - \Phi(\theta_c-\delta).$$

Пусть теперь $\theta_c - \delta < t < \theta_c$, тогда аналогично предыдущему

$$\int\limits_{\theta_c-\delta}^{\theta_c+\delta} [\Phi(\sqrt{n}(2\mathcal{Z}_n - t - \theta)) - \Phi(\sqrt{n}(t-\theta))]\varphi(\theta)\,d\theta \to \Phi(2\theta_c - t) - \Phi(t),$$

что совпадает с предельным значением для знаменателя $\mathfrak{J}_n$.

□

**Доказательство Леммы 4.**

Предположим, что для некоторого $\varepsilon > 0$ нижний предел $\varliminf\limits_{n\to\infty} \delta_n^* < \delta^* - \varepsilon$. Тогда в силу строгой монотонности функций $\mathfrak{J}_n$ и $\mathfrak{J}_\infty$ по параметру $\delta$ справедливо соотношение

$$1 - \alpha = \varliminf_{n\to\infty} \mathfrak{J}_n(\theta_c, \delta_n^*\,|\,t) \leqslant \mathfrak{J}_\infty(\theta_c, \delta^* - \varepsilon\,|\,t) < 1 - \alpha\,.$$

Полученное противоречие доказывает утверждение леммы. □

**Доказательство Теоремы 4.**

(I) Обозначим через $R_n(\delta)$ и $D_n$ числитель и знаменатель функции $\mathfrak{J}_n$ соответственно (см.(1)). Числитель и знаменатель функции $\mathfrak{J}_\infty$ будем обозначать через $R(\delta)$ и $D$. Тогда константы $\delta_n^*$ и $\delta^*$ можно определить как решения уравнений

$$R_n(\delta_n^*) - (1-\alpha)D_n = 0, \quad R(\delta^*) - (1-\alpha)D = 0.$$

Формула конечных приращений, примененная к функции $R_n(\delta)$, позволяет представить расхождение $\delta^* - \delta_n^*$ в виде

$$\delta^* - \delta_n^* = \frac{R_n(\delta^*) - (1-\alpha)D_n}{R_n'(\tilde{\delta}_n)} = \frac{(R_n(\delta^*) - R(\delta^*)) - (1-\alpha)(D_n - D)}{R_n'(\tilde{\delta}_n)},$$

где производная $R_n'(\tilde{\delta}_n)$ функции $R_n$ вычислена в некоторой средней точке между $\delta_n^*$ и $\delta^*$.

С учётом симметрии функций $\mathfrak{J}_n$ и $\mathfrak{J}_\infty$ достаточно рассмотреть только случай $t < \theta_c$. Из леммы 4 следует, что $\tilde{\delta}_n \to \delta^*$, а по построению $\theta_c - \delta^* > t$, поэтому при $n \to \infty$

$$\sqrt{n}(2\mathcal{Z}_n - t - \theta_c \pm \tilde{\delta}_n) = \sqrt{n}(\theta_c \pm \tilde{\delta}_n - t) + \frac{2\theta_c}{\sqrt{n}} \to +\infty,$$

$$\sqrt{n}\big(t - \theta_c \pm \tilde{\delta}_n\big) \to -\infty.$$

Таким образом,

$$R_n'(\tilde{\delta}_n) = \Big[\Phi\Big(\sqrt{n}\big(2\,\mathcal{Z}_n - t - \theta_c - \tilde{\delta}_n\big)\Big) - \Phi\Big(\sqrt{n}\big(t - \theta_c - \tilde{\delta}_n\big)\Big)\Big]\,\varphi\left(\theta_c + \tilde{\delta}_n\right) +$$
$$+ \Big[\Phi\Big(\sqrt{n}\big(2\,\mathcal{Z}_n - t - \theta_c + \tilde{\delta}_n\big)\Big) - \Phi\Big(\sqrt{n}\big(t - \theta_c + \tilde{\delta}_n\big)\Big)\Big]\,\varphi\left(\theta_c - \tilde{\delta}_n\right) \to$$
$$\to \varphi(\theta_c + \delta^*) + \varphi(\theta_c - \delta^*) > 0.$$

Применяя хорошо известные неравенства (см., например, [4, с. 196])

$$\frac{\varphi\,(x)}{x}\left(1 - \frac{1}{x^2}\right) \leqslant 1 - \Phi(x) = \Phi(-x) \leqslant \frac{\varphi\,(x)}{x}, \quad x > 0, \tag{4}$$

можно легко установить, что разность числителей функций $R_n(\delta^*)$ и $R(\delta^*)$ имеет экспоненциальный порядок малости при $n \to \infty$. Поэтому исследуемая нами скорость сходимости полностью определяется разностью

$$D_n - D = \left[\Phi\left(\frac{2\mathcal{Z}_n - t}{\sqrt{1+n^{-1}}}\right) - \Phi(2\theta_c - t)\right] - \left[\Phi\left(\frac{t}{\sqrt{1+n^{-1}}}\right) - \Phi(t)\right].$$

Воспользовавшись снова формулой конечных приращений, а также разложением в ряд Тейлора функции $(1+x)^{-1/2}$, получаем, что

$$D_n - D \asymp \frac{\Delta}{n}, \quad n \to \infty,$$

с константой $\Delta = \Delta(\theta_c, t) = (2\theta_c + t)\varphi(2\theta_c - t) + t\varphi(t)$.

(II) В случае, когда $\theta_c = t$, константа $\delta^* = 0$, и знаменатель $\mathfrak{J}_n$ есть

$$D_n = \Phi\left(\frac{t + 2tn^{-1}}{\sqrt{1+n^{-1}}}\right) - \Phi\left(\frac{t}{\sqrt{1+n^{-1}}}\right) \asymp \frac{2t}{n}\,\varphi(t), \quad n \to \infty.$$

Числитель $\mathfrak{J}_n$ после замены $\sqrt{n}(t-\theta)=y$ можно представить в виде

$$R_n(\delta)=\frac{1}{\sqrt{n}}\int\limits_{-\sqrt{n}\delta}^{\sqrt{n}\delta}\Big[\Phi\Big(y+\frac{2t}{\sqrt{n}}\Big)-\Phi(y)\Big]\varphi\Big(t-\frac{y}{\sqrt{n}}\Big)\,dy.$$

Пусть $\sqrt{n}\delta_n\to C$, тогда в силу теоремы Лебега об ограниченной сходимости можно перейти к пределу под знаком интеграла, что дает при $n\to\infty$ представление

$$R_n(\delta_n)\asymp\frac{1}{\sqrt{n}}\frac{2t}{\sqrt{n}}\,\varphi(t)\int\limits_{-C}^{C}\varphi(y)\,dy=\frac{2t}{n}\,\varphi(t)(2\Phi(C)-1).$$

Таким образом, если взять $C=z^{\alpha/2}$, то

$$\lim_{n\to\infty}\mathfrak{J}_n(\theta_c,\delta_n\,|\,t)=\lim_{n\to\infty}\frac{R_n(\delta_n)}{D_n}=(1-\alpha)\,.$$

С другой стороны, если, например, $\overline{\lim\limits_{n\to\infty}}\,\sqrt{n}\delta_n^*>z^{\alpha/2}(1+\varepsilon)$, то

$$(1-\alpha)=\overline{\lim_{n\to\infty}}\,\mathfrak{J}_n\left(\theta_c,\delta_n^*\,|\,t\right)\geqslant(2\Phi(z^{\alpha/2}(1+\varepsilon))-1)>(1-\alpha)\,,$$

что доказывает утверждение теоремы. □

**Доказательство Теоремы 5.**

Заметим сначала, что аналогично лемме 4 легко доказывается сходимость $\overline{t}_n\to\to t^*$ при $n\to\infty$. Отсюда, в частности, следует, что для некоторого $\varepsilon>0$, начиная с некоторого $N_0$, константы $\overline{t}_n>\theta_c+\delta+\varepsilon$ для любого $n>N_0$.

Запишем уравнения, определяющие константы $\overline{t}_n$ и $t^*$, в виде

$$\frac{1}{1-\alpha}\int\limits_{\theta_c-\delta}^{\theta_c+\delta}\Big[\Phi(\sqrt{n}(\overline{t}_n-\theta))-\Phi\Big(\sqrt{n}\Big(2\Big(1+\frac{1}{n}\Big)\theta_c-\overline{t}_n-\theta\Big)\Big)\Big]\varphi(\theta)\,d\theta=$$
$$=\Phi\left(\frac{\overline{t}_n}{\sqrt{1+n^{-1}}}\right)-\Phi\left(\frac{2(1+n^{-1})\theta_c-\overline{t}_n}{\sqrt{1+n^{-1}}}\right),$$

$$\frac{1}{1-\alpha}\int\limits_{\theta_c-\delta}^{\theta_c+\delta}\varphi(\theta)\,d\theta=\Phi(t^*)-\Phi(2\theta_c-t^*).$$

Неравенства (4), а также сделанное выше замечание, обеспечивают экспоненциальную скорость сходимости к нулю разности левых частей этих уравнений. Таким образом, разность уравнений (при соответствующем выборе точек $t_n'$, $t_{cn}$ и $q>0$) может быть записана как

$$o(e^{-qn})=\left(\frac{\overline{t}_n}{\sqrt{1+n^{-1}}}-t^*\right)(\varphi(t_n')+\varphi(t_{cn}))+2\left(\frac{\theta_c+\theta_c n^{-1}}{\sqrt{1+n^{-1}}}-\theta_c\right)\varphi(t_{cn}),$$

причем последовательности точек $t_n'\to t^*$, $t_{cn}\to2\theta_c-t^*$. Утверждение теоремы сразу следует из этого представления, если воспользоваться разложением в ряд Тейлора:

$$\left(1+\frac{1}{n}\right)^{-1/2}=1-\frac{1}{2n}+o\left(\frac{1}{n^2}\right),\quad n\to\infty.$$

□

**Доказательство Леммы 5.**

После замены переменных $\theta \mapsto B - \dfrac{y}{n}$ искомое интегральное выражение может быть представлено в виде

$$I_n = \frac{1}{n} \int\limits_0^{n(B-A)} \Phi\left(\sqrt{n}\left(-(\frac{y}{n} + (t-B) + r_n)\right)\right) \varphi\Big(B - \frac{y}{n}\Big)\, dy.$$

Двусторонние неравенства (4) обеспечивают возможность эквивалентной подстановки $\Phi(-x) \asymp x^{-1}\varphi(x)$ (при $x \to \infty$) в полученное выражение

$$I_n \asymp \frac{1}{\sqrt{2\pi}\, n\sqrt{n}} \int\limits_0^{n(B-A)} \frac{\exp\{-n(yn^{-1} + (t-B) + r_n)^2/2\}}{(yn^{-1} + (t-B) + r_n)} \varphi\Big(B - \frac{y}{n}\Big)\, dy =$$

$$= \frac{\exp\{-n(t-B)^2/2 - (t-B)\, n\, r_n - n\, r_n^2/2\}}{\sqrt{2\pi}\, n\sqrt{n}} \int\limits_0^{n(B-A)} e^{-(t-B)y} H_n(y)\, dy,$$

где последовательность функций

$$H_n(y) = \frac{\exp\left(-y^2 n^{-1}/2\right) \exp\left(-y\, r_n\right)}{(yn^{-1} + (t-B) + r_n)} \varphi\Big(B - \frac{y}{n}\Big) \ \to\ \frac{\varphi(B)}{t-B} > 0.$$

Доказательство леммы завершается теперь простым предельным переходом под знаком интеграла, если заметить, что последовательность $H_n(y)$ ограничена.

$\square$

**Доказательство Теоремы 6.**

(I) Поскольку включение интервала $[\theta_c - \delta, \theta_c + \delta]$ в семейство $\mathcal{B}_n$ эквивалентно включению интервала $[\theta_c - \delta + q_n, \theta_c + \delta - q_n]$ в семейство $\mathcal{B}^*$, то, как было замечено перед формулировкой теоремы 5, событие $\{\mathcal{B}_n \ni [\theta_c - \delta, \theta_c + \delta]\}$ выполняется тогда и только тогда, когда наблюдаемое значение статистики $T = t$ принадлежит области $2\theta_c - t_n^* \leqslant t \leqslant t_n^*$, где $t_n^* = t^*(\theta_c, \delta - q_n)$. Непрерывность и монотонность функции $\mathfrak{J}_\infty$ по обоим параметрам $t$ и $\delta$ обеспечивает сходимость $t_n^* \to t^*$, откуда, воспользовавшись схемой доказательства леммы 3, находим, что при $n \to \infty$ надежность

$$\mathcal{Q}(\mathcal{B}_n \,|\, \theta_c - \delta, \theta_c + \delta) \to \frac{\Phi(\theta_c + \delta) - \Phi(\theta_c - \delta)}{\Phi(t^*) - \Phi(2\theta_c - t^*)} = 1 - \alpha.$$

(II) Запишем равенства, определяющие константы $t^*$ и $t_n^*$, в виде

$$(\Phi(t_n^*) - \Phi(2\theta_c - t_n^*))(1-\alpha) = \Phi(\theta_c + \delta - q_n) - \Phi(\theta_c - \delta + q_n),$$
$$(\Phi(t^*) - \Phi(2\theta_c - t^*))(1-\alpha) = \Phi(\theta_c + \delta) - \Phi(\theta_c - \delta).$$

Вычитая из первого равенства второе и применяя формулу конечных приращений, устанавливаем, что при $n \to \infty$

$$t_n^* - t^* \asymp -\frac{\varphi(\theta_c + \delta) + \varphi(\theta_c - \delta)}{(1-\alpha)(\varphi(t^*) + \varphi(2\theta_c - t^*))}\, q_n.$$

Таким образом, как точное семейство $\mathcal{D}_n^*$ (см. теорему 5), так и семейство $\mathcal{B}_n$ содержат интервал $[\theta_c - \delta, \theta_c + \delta]$ только, когда значение статистики $T$ удовлетворяет неравенствам вида $2\theta_c - t^* - r_n \leqslant T \leqslant t^* + r_n$, причем для обоих семейств последовательность $r_n = O_c(n^{-1})$. Точность этих семейств для фиксированного интервала $[\theta_c - \delta, \theta_c + \delta]$ равна

$$\mathbf{P}\left\{\theta_c - \delta \leqslant \vartheta \leqslant \theta_c + \delta / T \notin [2\theta_c - t^* - r_n,\, t^* + r_n]\right\} =$$

$$= \frac{\int\limits_{\theta_c-\delta}^{\theta_c+\delta} \Big[\Phi\Big(\sqrt{n}(\theta - t^* - r_n)\Big) + \Phi\Big(\sqrt{n}(2\theta_c - \theta - t^* - r_n)\Big)\Big]\, \varphi(\theta)\, d\theta}{1 - \Big[\Phi\Big((t^* + r_n)\zeta_n\Big) - \Phi\Big((2\theta_c - t^* - r_n)\zeta_n\Big)\Big]} =$$

$$= \frac{\int\limits_{\theta_c-\delta}^{\theta_c+\delta} \Phi\Big(\sqrt{n}(\theta - t^* - r_n)\Big)\Big(\varphi(\theta) + \varphi(2\theta_c - \theta)\Big)\, d\theta}{1 - \Big[\Phi\Big((t^* + r_n)\zeta_n\Big) - \Phi\Big((2\theta_c - t^* - r_n)\zeta_n\Big)\Big]},$$

где $\zeta_n = 1/\sqrt{1 + n^{-1}}$.

Применяя утверждение леммы 5 (с заменой функции $\varphi(\theta)$ на $\varphi(\theta) + \varphi(2\theta_c - \theta)$) получаем, что точность обоих рассматриваемых семейств асимптотически эквивалентна (при $n \to \infty$) последовательности

$$\frac{(\varphi(\theta_c + \delta) + \varphi(\theta_c - \delta))\exp\left(-c_0(t^* - \theta_c - \delta)\right)}{(\Phi(-t^*) + \Phi(2\theta_c - t^*))\,(t^* - \theta_c - \delta)^2}\; \frac{\exp\left(-(t^* - \theta_c - \delta)^2 n/2\right)}{n\sqrt{n}},$$

возможно, с различными константами $c_0$. Доказательство теоремы завершено. $\square$

**Summary**

*R.F. Salimov, S.V. Simushkin.* Asymptotically Most Accurate Double-Side Confidence Intervals in Normal-Normal Model.

This article presents a form of the most precise class of confidence intervals for the normal-normal model in $d$-posterior approach. It is possible to state its asymptotical behavior with the increase of the sample size. A way of the most asymptotically accurate confidence family construction is also suggested in this article.

**Key words:** normal-normal model, $d$-posteriory probability of errors, double-side confidence intervals.

**Литература**

1. *Володин И.Н, Новиков Ан.А., Симушкин С.В.* Апостериорный подход к проблеме гарантийности статистического контроля качества // Обозрение прикл. и промышл. матем. – 1994. – № 2. – С. 148–178.

2. *Володин И.Н., Симушкин С.В.* Доверительное оценивание в d-апостериорном подходе // Теория вероятностей и её применения. – 1990. – № 2. – С. 242–254.

3. *Закс Ш.* Теория статистических выводов. – М.: Мир, 1975. – 776 с.

4. *Феллер В.* Введение в теорию вероятностей и её приложения. – М.: Мир, 1967. – 499 с.

Поступила в редакцию
09.09.09

**Салимов Рустем Фаритович** – аспирант кафедры математической статистики Казанского государственного университета.

E-mail: *rust1k@mail.ru*

**Симушкин Сергей Владимирович** – кандидат физико-математических наук, доцент кафедры математической статистики Казанского государственного университета.

E-mail: *smshkn@gmail.com*