НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАРШИНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ

ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ
им. В.А. ТРАПЕЗНИКОВА РАН

ИНСТИТУТ МАТЕМАТИКИ
им. В.И. РОМАНОВСКОГО АН РУЗ

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ (ИТММ-2022)

**МАТЕРИАЛЫ**
**XXI Международной конференции**
**имени А. Ф. Терпугова**
**25–29 октября 2022 г.**

ТОМСК
Издательство Томского
государственного университета
2023

УДК 519
ББК 22.17
И74
Информационные технологии и математическое моделирование (ИТММ-2022): Материалы XXI Международной конференции имени А. Ф. Терпугова (25–29 октября 2022 г.). — Томск: Издательство Томского государственного университета, 2023. — 442 с.

ISBN 978–5–907572–98–0

Сборник содержит избранные материалы XXI Международной конференции имени А.Ф. Терпугова по следующим направлениям: теория массового обслуживания и ее приложения, интеллектуальный анализ данных и визуализация, информационные технологии и программная инженерия, математическое и компьютерное моделирование технологических процессов. Также в сборник вошли материалы международного симпозиума "Современные стохастические модели и проблемы актуарной математики"(МАМОНТ-2022).

Для специалистов в области информационных технологий и математического моделирования.

**УДК 519**
**ББК 22.17**



ISBN 978–5–907572–98–0

NATIONAL RESEARCH TOMSK STATE UNIVERSITY

KARSHI STATE UNIVERSITY

PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA

V.A. TRAPEZNIKOV INSTITUTE OF CONTROL
SCIENCES OF RUSSIAN ACADEMY OF SCIENCES

V.I. ROMANOVSKY INSTITUTE OF MATHEMATICS
OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF UZBEKISTAN

# INFORMATIONAL TECHNOLOGIES AND MATHEMATICAL MODELLING (ITMM-2022)

**PROCEEDINGS**
**of the 21th International Conference**
**named after A. F. Terpugov**
**2022 October, 25–29**

TOMSK
Tomsk State
University Publishing
2023

UDC 519
LBC 22.17
I60
Informational technologies and mathematical modelling (ITMM-2022): Proceedings of the 21th International Conference named after A. F. Terpugov (2022 October, 25–29). — Tomsk: Tomsk State University Publishing, 2022. — 442 p.

ISBN 978–5–907572–98–0

This volume presents selected papers from the XXI International Conference named after A.F. Terpugov. The papers are devoted to new results in the following areas: queuing theory and its applications, data mining and visualization, information technology and software engineering, mathematical and computer modeling of technological processes. The collection also presents the proceedings of symposium "Modern Stochastic Models and Problem of Actuarial Mathematics"(MAMMOTH-2022).

**UDC 519**
**LBC 22.17**

E d i t o r s:
**A.A. Nazarov**, Doctor of Technical Sciences, Professor,
**S.P. Moiseeva**, Doctor of Physical and Mathematical Sciences, Professor,
**A.N. Moiseev**, Doctor of Physical and Mathematical Sciences, Associate Professor.
**D.V. Semenova**, Candidate of Physical and Mathematical Sciences, Associate Professor.

ISBN 978–5–907572–98–0

# ЛОКАЛЬНЫЕ КОМБИНИРОВАННЫЕ ОЦЕНКИ ДОЛИ

М. А. Бахчаева, Ю. Г. Дмитриев

*Национальный исследовательский Томский государственный университет, г. Томск, Россия*

В работе приводятся комбинированные оценки доли, являющиеся взвешенной суммой обычной эмпирической оценки и априорной догадки, которая представляется в виде некоторого заданного значения искомой доли, высказываемой исследователем или экспертом на основании своего опыта и знаний. Рассматривается вопрос оценки оптимального весового коэффициента на основании локальных аппроксимаций и построения локальных адаптивных комбинированных оценок доли. Анализируются свойства оценок при конечном объеме наблюдений.

**Ключевые слова:** *априорная догадка, локальная оценка, комбинированная оценка, адаптивная оценка.*

## Введение

Статистическая обработка экспериментальных данных включает в себя ряд проблем, одна из которых это оценка неизвестной доли объектов в генеральной совокупности с заданным значением признака. Широкое практическое применение данная проблема имеет место в задачах оценки и контроле надежности, оценки и управления качеством продукции и т.д. Для сокращения объема дорогостоящих экспериментальных данных или повышения точности оценивания при фиксированном объеме наблюдений, разумно привлекать дополнительную информацию, которая может быть разнообразной и иметь различные источники поступления.

В данной работе предлагаются локальные адаптивные комбинированные оценки доли, учитывающие совместно обычную эмпирическую оценку и априорную догадку в форме некоторого заданного значения искомой доли, которое высказывается исследователем или экспертом на основании своего опыта и знаний. Анализируются среднеквадратические ошибки таких оценок при конечном объеме наблюдений. Указываются условия, при которых локальные комбинированные оценки предпочтительнее обычной эмпирической оценки.

## 1. Постановка задачи и структура оценки

Пусть имеется конечная генеральная совокупность объема $N$ и случайным образом производится выборка без возвращения объема $n$ из этой генеральной совокупности. Под $P(B)$ будем понимать долю объектов в генеральной совокупности с заданным значением $B$ признака. Пусть $p_a$ – априорная догадка, которая выступает в качестве возможного значения доли $P = P(B)$. Требуется оценить долю объектов в генеральной совокупности с заданным значением признака, учитывая совместно эмпирическую оценку $\hat{P} = \sum\limits_{i=1}^{n} I_B(X_i)$ и $p_a$, здесь $I_B(X_i)$ – индикаторная функция, равная 1, если у объекта $X_i$ признак принимает значение $B$.

Следуя работам [1, 2, 4, 4, 5], рассмотрим комбинированную оценку доли

$$\hat{P}_\lambda = (1-\lambda)\hat{P} + \lambda p_a, \tag{1}$$

где оптимальный весовой коэффициент $\lambda$ выбран из условия минимума среднеквадратической ошибки (СКО) $S^2(\lambda) = M[\hat{P}_\lambda - P]^2$ и определяется выражением

$$\lambda = \lambda(P) = \left(1 + n\frac{(P-p_a)^2}{P(1-P)\frac{N-n}{N-1}}\right)^{-1} = \left(1 + n\frac{\Delta^2}{\sigma^2}\right)^{-1}, \tag{2}$$

Здесь $\sigma^2 = P(1-P)\dfrac{N-n}{N-1}$ , $\Delta = P - p_a$ – величина отклонения априорной догадки от истинного значения искомой доли. Оптимальный весовой коэффициент $\lambda$ изменятся в пределах $0 < \lambda \leq 1$ и показывает, какое влияние оказывает каждое из слагаемых в комбинированной оценке.

В соответствии с (1) минимум СКО представляется в виде:

$$S^2 = M[\hat{P}-P]^2 - \frac{[M(\hat{P}-P)^2]^2}{D\hat{P} + (P-p_a)^2} = D\hat{P} - \frac{(D\hat{P})^2}{D\hat{P} + \Delta^2} = (1-\lambda)\sigma^2/n, \tag{3}$$

где дисперсия $D\hat{P} = P(1-P)(N-n)/(N-1)n$ характеризует точность оценки $\hat{P}$, а коэффициент $(1-\lambda)$ показывает, во сколько раз уменьшается СКО комбинированной оценки (1) по сравнению с $\hat{P}$.

Величина $(D\hat{P})^2/(D\hat{P} + \Delta^2)$ в (3) задает выигрыш в точности оценивания за счет привлечения априорной догадки $p_a$ при оптимальном $\lambda$, если исходной оценкой является $\hat{P}$.

## 2. Локальные аппроксимации весового коэффициента

Незнание оптимального коэффициента $\lambda$ в (2) затрудняет практическое использование комбинированной оценки (1). Выходом из этого положения является нахождение той или иной оценки весового коэффициента и построения адаптивной комбинированной оценки. Рассмотрим подход, основанный на локальной аппроксимации $\lambda$, и изучим вопрос, при каких условиях локальные адаптивные комбинированные оценки являются предпочтительнее обычной оценки $\hat{P}$ по величине СКО.

Рассмотрим первую аппроксимацию весового коэффициента (2) в окрестности априорной догадки $p_a$. Используя разложение Тейлора, получим

$$\lambda_T(P) = 1 - n\frac{N-1}{N-n}\frac{(P-p_a)^2}{p_a(1-p_a)} + R_n,$$

где $R_n$ - остаточный член. Пренебрегая $R_n$, получим первый локальный весовой коэффициент

$$\lambda_1 = 1 - n\frac{N-1}{N-n}\frac{(P-p_a)^2}{p_a(1-p_a)}. \tag{4}$$

Из (4) следует, что значения $\lambda_1$ изменяются в пределах $0 < \lambda_1 \leq 1$, если выполнено условие

$$P \in \left(p_a - \sqrt{\frac{p_a(1-p_a)(N-n)}{n(N-1)}}; p_a + \sqrt{\frac{p_a(1-p_a)(N-n)}{n(N-1)}}\right)$$

Данному коэффициенту соответствует локальная комбинированная оценка $\hat{P}_{\lambda_1} = \hat{P} - \lambda_1(\hat{P} - p_a)$ с СКО в виде

$$S^2_{\lambda_1} = (1-\lambda_1)^2\sigma^2/n + {\lambda_1}^2(P-p_a)^2. \tag{5}$$

Рассмотрим оценку $\tilde{P}$, которая отличается от $\hat{P}$ только тем, что $\tilde{P} = p_a$ при $\hat{P} = 0$ или 1. Заменяя в (4) неизвестное $P$ на $\tilde{P}$, получим оценку локального весового коэффициента $\lambda_1$

$$\hat{\lambda_1} = 1 - n\frac{N-1}{N-n}\frac{(\tilde{P}-p_a)^2}{p_a(1-p_a)}$$

и соответствующую ему первую локальную адаптивную комбинированную оценку $\hat{P}_1 = \hat{P} - \hat{\lambda_1}(\hat{P} - p_a)$. СКО этой оценки можно вычислить, используя гипергеометрический закон, по которому распределено $n\hat{P}$:

$$\pi_k = P\{n\hat{P} = k\} = \frac{(NP)![N(1-P)]!n!(N-n)!}{k!(NP-k)!(n-k)![N(1-P)-(n-k)]!N!}, k = \overline{0,n}.$$

$$S_1^2 = M[\hat{P}_1 - P]^2 = \sum_{k=0}^{n} [\Psi_1(k,n,p_a,N) - P]^2 \pi_k, \tag{6}$$

где $\Psi_1(k,n,p_a,N) = p_a + n\dfrac{N-1}{N-n}\dfrac{(k/n-p_a)^3}{p_a(1-p_a)}$.

Рассмотрим второй локальный весовой коэффициент в виде

$$\lambda_2 = 1 - n\frac{N-1}{N-n}(P-p_a)^2 p_a(1-p_a)$$

и соответствующую ему вторую локальную комбинированную оценку $\hat{P}_{\lambda_2} = \hat{P} - \lambda_2(\hat{P} - p_a)$. СКО данной оценки представляется в следующем виде

$$S_{\lambda_2}^2 = (1-\lambda_2)^2\sigma^2/n + {\lambda_2}^2(P-p_a)^2. \tag{7}$$

Заменяя неизвестное $P$ на $\tilde{P}$, получаем оценку $\hat{\lambda_2} = 1 - n\dfrac{N-1}{N-n}(\tilde{P} - p_a)^2 p_a(1-p_a)$ весового коэффициента $\lambda_2$ и вторую локальную адаптивную комбинированную оценку $\hat{P}_2 = \hat{P} - \hat{\lambda_2}(\hat{P} - p_a)$. СКО представляется в следующем виде

$$S_2^2 = M[\hat{P}_2 - P]^2 = \sum_{k=0}^{n} [\Psi_2(k,n,p_a,N) - P]^2 \pi_k, \tag{8}$$

где $\Psi_2(k,n,p_a,N) = p_a + n\dfrac{N-1}{N-n}(k/n-p_a)^3 p_a(1-p_a)$.

Для сравнения СКО оценок введем отношения

$$E_{\lambda_1} = \frac{S_{\lambda_1}^2}{D\hat{P}}, E_1 = \frac{S_1^2}{D\hat{P}}, E_{\lambda_2} = \frac{S_{\lambda_2}^2}{D\hat{P}}, E_2 = \frac{S_2^2}{D\hat{P}}.$$

Формулы (5)–(8) позволяют рассчитать значения отношений СКО в зависимости от значений $P, n, p_a, N$. Данные величины позволяют выделить интервалы значений $P$, при которых эти отношения меньше единицы. Наличие таких интервалов для $P$ совместно с объемом генеральной совокупности $N$, объемом выборки $n$ и значениями $p_a$ указывают условия, при которых локальные адаптивные комбинированные оценки предпочтительнее обычной оценки $\hat{P}$ по величине СКО. Результат для различных локальных комбинированных оценок и их адаптаций представлен в виде графиков на рис.1-4.

Рис.1 и рис.2 показывают сравнение СКО первой локальной оценки с эмпирической оценкой при различном объеме наблюдений $n = 1, 2, 3, 4,$

$N = 40$ и $p_a = 0.5$ (комбинированная оценка – слева, адаптивная – справа, оптимальная – пунктирная линия)

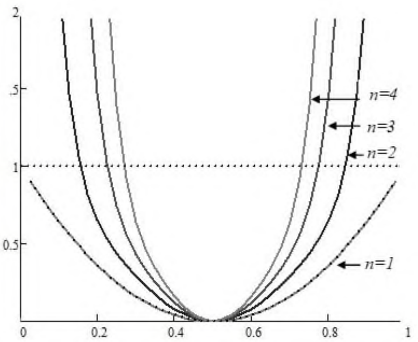


Рис. 1. Сравнение СКО первой локальной оценки с эмпирической оценкой

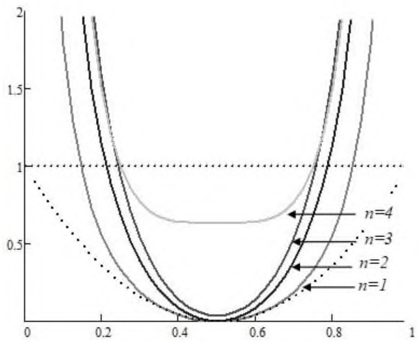


Рис. 2. Сравнение СКО первой локальной адаптивной оценки с эмпирической

Рис.3 и рис.4 показывают сравнение СКО второй локальной оценки с эмпирической оценкой при различном объеме наблюдений $n = 1, 4, 8, 12, 16$, $N = 40$ и $p_a = 0.5$ (комбинированная оценка – слева, адаптивная – справа, оптимальная – пунктирная линия)

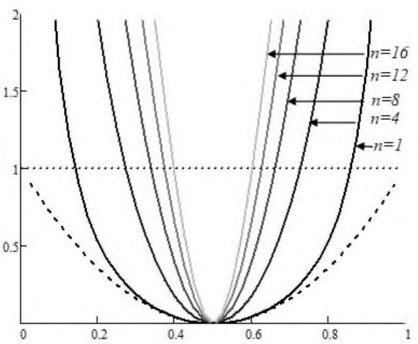


Рис. 3. Сравнение СКО второй локальной оценки с эмпирической оценкой

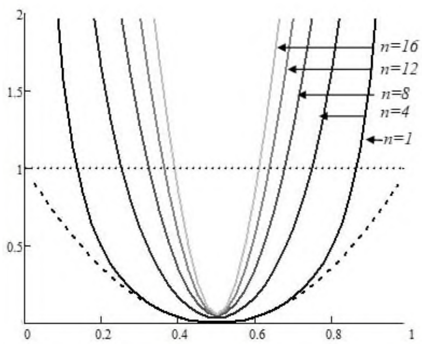


Рис. 4. Сравнение СКО второй локальной адаптивной оценки с эмпирической

### Заключение

В работе были рассмотрены локальные комбинированные оценки доли, проанализированы их точности по величине СКО при конечном объеме наблюдений. В частности, локальная аппроксимация в окрестности априорной догадки с применением формулы Тейлора дает хороший результат для теоретических значений весового коэффициента, в то время как локальная адаптивная оценка дает результат только при небольшом объеме наблюдений, а при возрастании объема наблюдений имеет большую погрешность из-за нарушения условия применимости весового коэффициента.

Вторая локальная аппроксимация показывает приемлемые результаты как для теоретических значений, так и для выборочных, оказываясь достаточно близко к оптимальной оценке. Следует отметить, что с ростом объема наблюдений сужается интервал значений $P$, где данные оценки предпочтительнее по величине СКО обычной эмпирической оценки.

### СПИСОК ЛИТЕРАТУРЫ

1. *Dmitriev Yu., Tarassenko P., Ustinov Yu.* On Estimation of Linear Functional by Utilizing a Prior Guess // ITMM2014, CCIS 487. 2014 P. 82–90.
2. *Дмитриев Ю. Г., Тарасенко П. Ф.* Использование априорной информации в статистической обработке экспериментальных данных // Известия вузов. Физика. 1992. № 9. С. 136–142.
3. *Дмитриев Ю. Г., Кошевая Т. О.* Комбинированные оценки доли // Информационные технологии и математическое моделирование (ИТММ-2015): Материалы XIV Международной конференции им. А. Ф. Терпугова (18-22 ноября 2015). 2015. Т. 1. С. 50–55.
4. *Дмитриев Ю. Г., Кошевая Т. О.* О комбинированных оценках вероятности // Известия вузов. Физика. 2015. Т. 58. № 11/2. С. 242–247.
5. *Дмитриев Ю. Г., Кошевая Т. О.* Оценки вероятности с учетом априорных догадок // Известия вузов. Физика. 2016. Т. 59. № 8/2. С. 25–28.

**Бахчаева Мария Андреевна** — аспирант, ассистент каф. системного анализа и математического моделирования ИПМКН ТГУ. E-mail: *m.bakhchaeva@mail.ru*

**Дмитриев Юрий Глебович** — д.ф.-м.н., доц., проф. каф. системного анализа и математического моделирования ИПМКН ТГУ. E-mail: *dmit70@mail.ru*