УДК 004.4:517.962

ИНТЕРВАЛЬНАЯ ОЦЕНКА ПАРАМЕТРОВ БЕТА-РАСПРЕДЕЛЕНИЯ ПРИ ОПРЕДЕЛЕНИИ ДОВЕРИТЕЛЬНОЙ ТРУДОЕМКОСТИ АЛГОРИТМОВ

Кривенцов А.С., аспирант,

Московский государственный университет приборостроения и информатики,

Ульянов М.В., д-р. техн. наук, проф.,

НИУ «Высшая Школа Экономики», Московский государственный университет печати им. Ивана Федорова.

Аннотация

В статье рассматриваются классические методы определения трудоемкости алгоритмов и обосновывается необходимость оценки по критерию доверительной трудоемкости. Данная методика определения доверительной трудоёмкости состоит в аппроксимации наблюдаемого в эксперименте распределения частот известным законом распределения. В рамках этой методики возникает задача определения параметров распределения, и их доверительных интервалов. Показано, что погрешность, обусловленная использованием точечных оценок при определении параметров бета-распределения не является значимой.

1. Введение

Задача любой науки состоит в выявлении и исследовании закономерностей, которым подчиняются реальные процессы. Найденные закономерности имеют не только теоретическую ценность, они широко применяются на практике — в планировании, управлении и прогнозировании. Задача оценки искомых характеристик, а также степени точности получаемых результатов является основной задачей математической статистики. Статистические методы применяются в самых различных областях знания, в частности, в теории алгоритмов. Одной из интересных задач данной области является исследование и сравнение алгоритмов по критерию трудоемкости. Под трудоемкостью алгоритма понимается число заданных им базовых операций в принятой модели вычислений. Наибольший практический интерес представляют результаты исследования частотной встречаемости значений трудоемкости, поскольку существующие в настоящее время методики позволяют определить трудоемкость в лучшем, худшем и средних случая, не оперируя вероятностями наблюдения таких значений трудоёмкости. Известные методики оценки трудоемкости в худшем случае [1] дают возможность сравнения алгоритмов по критерию временной сложности, что позволяет решить задачу о выборе рационального алгоритма в условиях строгих гарантированных ограничений, например, при определении максимальной нагрузки на сервер, определении ресурсов компьютера для поддержки различных приложений и так далее. В условиях «мягких» требований к ресурсам сравнение алгоритмов может быть проведено по критерию трудоемкости «в среднем». Однако для большинства алгоритмов возможно лишь ограничение данного вида оценок «сверху», а для некоторых алгоритмов получение оценок такого рода в теории является достаточно сложной задачей. Следовательно, универсальной является только оценка трудоемкости в худшем случае, однако она зачастую не является приемлемой, поскольку значения трудоемкости, близкие к худшему случаю имеют, как правило, незначительную частотную встречаемость, из чего следует, что подобная оценка является сильно завышенной. Таким образом, представляет интерес оценка трудоемкости алгоритма, позволяющая отбросить маловероятные значения, близкие к худшему случаю, задав пороговое значение вероятности, и тем самым улучшить результаты оценки трудоемкости по сравнению с худшим случаем.

2. Доверительная трудоёмкость и особенности её определения

Одно из возможных решений этой задачи, предложенное авторами статьи и В.Н. Петрушиным в [2], связано с рассмотрением трудоёмкости алгоритма при фиксированной длине входа как дискретной ограниченной случайной величины, имеющей некоторое неизвестное распределение. Подход авторов состоит в построении доверительного интервала трудоёмкости на основе аппроксимации неизвестного дискретного распределения значений трудоёмкости непрерывным распределением с ограниченной вариацией, в качестве которого предлагается использовать бета-распределение. Получаемое решение позволяет с заданной доверительной вероятностью указать более реальную правую границу трудоемкости алгоритма при фиксированной длине входа. Применённый подход был назван авторами доверительной трудоёмкостью — f_γ, поскольку оценка представляет собой значение трудоемкости, которое не будет превышено для единичного входа с заданной доверительной вероятностью γ [2]. Иными словами, для некоторого единичного входа алгоритма трудоемкость будет заключена между лучшим случаем и доверительной трудоемкостью, то есть в сегменте $[f^\vee, f_\gamma]$ с вероятностью γ, где $f^\vee$ — теоретическое значение трудоемкости в «лучшем» случае.

Предложенное решение основано на использовании функции плотности некоторого распределения вероятности для аппроксимации гистограммы распределения относительных частот значений трудоемкости исследуемого алгоритма. Основной практической ценностью данного метода является возможность работать с экспериментальными относительными частотами как с математической функцией. Как следствие, данная методика представляет собой инструмент для решения класса задач, связанных с анализом алгоритмов по экспериментальным данным. Другой важной особенностью предложенного метода является возможность построения прогнозов относительно числа базовых операций, которое не будет превышено с заданной доверительной вероятностью при других длинах входа. Результаты данных прогнозов позволяют ввести новую оценку качества алгоритмов по критерию доверительной трудоемкости, что улучшает существующую на настоящее время методику оценки трудоемкости алгоритмов и, в частности, дает новые подходы к решению задачи выбора рационального алгоритма из рассматриваемого множества.

В рамках данной методики возникает задача определения закона распределения и его параметров [2]. Часто применяемый на практике для реконструкции параметров распределения по экспериментальным данным метод моментов оперирует с точечными оценками математического ожидания, дисперсии, и, возможно, центральных моментов более высоких порядков. Очевидно, что более корректно оперировать доверительным интервалом для оценки неизвестных параметров распределения, что приводит либо к интервальному определению параметров, либо к оценке погрешности результатов, получаемых на основе точечных оценок.

Целью данной статьи является оценка погрешности определения доверительной трудоёмкости при использовании точечных значений выборочного среднего и выборочной исправленной дисперсии в качестве оценок математического ожидания и дисперсии при определении параметров используемого распределения.

3. Выбор аппроксимирующей функции плотности распределения вероятностей

Очевидно, оценка погрешности при определении параметров зависит от типа распределения, следовательно, первоочередным является вопрос выбора аппроксимирующего закона распределения. Выбор этот, в значительной мере, является достаточно неопределенной, и во многом субъективной процедурой, при этом многое зависит от априорных знаний об объекте и его свойствах, следовательно, выбор закона распределения носит характер принятия той или иной гипотезы, которая должна быть определённым образом проверена. Проверка предположения о том, что исследуемая случайная величина, в данном случае наблюдаемое в эксперименте распределение частот значений трудоемкости, подчиняется выбранному закону распределения осуществляется по некоторому статистическому критерию. Статистическим критерием является строгое математическое правило, по которому с известным уровнем значимости принимается или отвергается та или иная статистическая гипотеза. Критерий представляет собой некоторую функцию от значений результатов наблюдений (ряда эмпирически полученных значений признака), которая служит для выявления меры расхождения между эмпирическими и гипотетическими значениями. Критерии проверки гипотезы о законе распределения называются критериями согласия. В настоящее время применяются критерии Пирсона, Колмогорова-Смирнова, Андерсона-Дарлинга, Жака-Бера и Шапиро-Вилка и др. [3].

Опираясь на ранее полученные результаты [2] авторы рекомендуют использовать в качестве аппроксимирующей функции плотности бета-распредление. Бета-распределение относится к семейству одномерных абсолютно непрерывных распределений, плотность вероятности которого зависит от двух параметров α и β, и имеет следующий вид [3]:

$$b(x,\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},\quad x\in[0,1],$$

где $\Gamma(\cdot)$ — гамма функция Эйлера, α,β — параметры функции плотности бета-распределения.

Бета-распределение в ряде случаев является предпочтительным, поскольку его функция плотности имеет ограниченный носитель $[0,1]$, а также обладает гибкой формой с произвольной асимметрией (см. рис. 1.). Как видно из рис. 1., функция плотности распределения вероятности бета-распределения может быть выпуклой, вогнутой, монотонно возрастающей и монотонно убывающей, что удобно для аппроксимации гистограммы относительных частот значений функции трудоемкости. Рассмотрим частные случаи значений параметров α и β, и соответствующий им вид функции $b(x,\alpha,\beta)$:

— $\alpha<1,\ \beta<1$ — график выпуклый вниз и уходит в бесконечность на границах;

— $\alpha<1,\beta\geq 1$ или $\alpha=1,\beta>1$ — график строго убывающий;

— $\alpha=1,\ \beta=1$ — график совпадает с графиком плотности стандартного непрерывного равномерного распределения;

— $\alpha=1,\ \beta<1$ или $\alpha>1,\ \beta\leq 1$ — график строго возрастающий;

— $\alpha>2,\ \beta=1$ — график строго выпуклый;

— $1<\alpha<2,\ \beta=1$ — график строго вогнутый;

— $\alpha>1,\ \beta>1$— график унимодальный;

— $\alpha=\beta$, плотность вероятности симметрична относительно центра носителя, то есть:

$$b(x-1/2,\alpha,\alpha)=b(x+1/2,\alpha,\alpha),\ x\in[0,1/2].$$

Особо отметим, что в унимодальном случае, т. е. когда $\alpha>1,\ \beta>1$ соотношение $\alpha<\beta$ приводит к левой асимметрии распределения, а $\alpha>\beta$ — к правой.

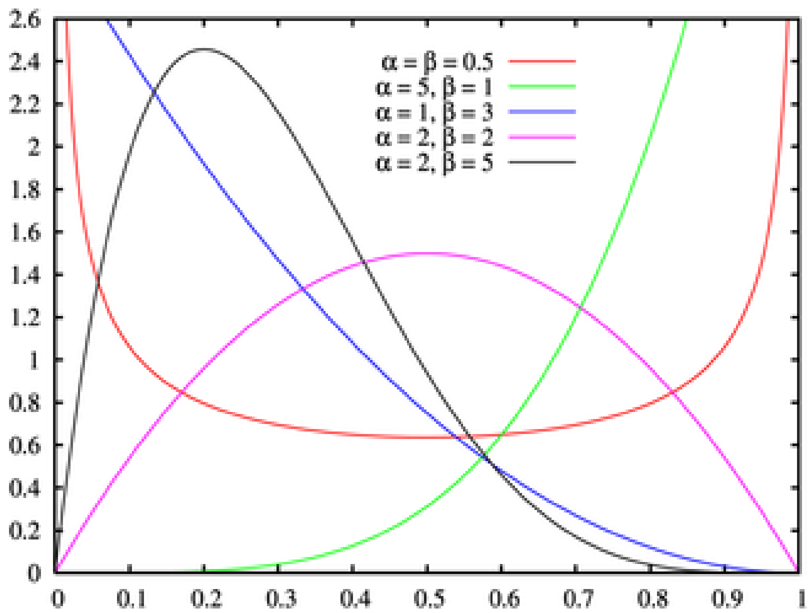

Рис. 1. Плотность вероятности бета-распределения.

Математическое ожидание M и дисперсия D бета-распределения определяются его параметрами α и β, и задаются следующими формулами [3]:

$$M(\alpha,\beta)=\frac{\alpha}{\alpha+\beta},\quad D(\alpha,\beta)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Отметим, что в носителе $[0,1]$ при любых допустимых α и β значение M ограничено носителем $[0,1]$, а значение D не превышает $1/4$ [4]. Обращением этих формул получаются формулы для определения коэффициентов α и β по известным математическому ожиданию и дисперсии. Такой метод восстановления параметров распределения называется методом моментов:

$$\alpha(M,D)=\frac{(1-M)M^2}{D}-M,\ \beta(M,D)=\frac{M(M-1)^2}{D}+M-1. \tag{1}$$

Параметры указанного распределения зависят от математического ожидания и дисперсии, которые, как правило, неизвестны. Для них по экспериментальным данным могут быть получены или точечные или интервальные оценки. При этом задача оценки состоит в получении наилучших (несмещённых и состоятельных) оценок параметров распределения генеральной совокупности на основании выборочных данных.

4. Интервальные оценки параметров распределения

Точечные оценки параметров выборочной совокупности могут быть приняты в качестве ориентировочных, первоначальных оценок при обработке выборочных данных. В ряде случаев требуется найти не только численные значения этих параметров, но и оценить их надежность и точность. Требуется знать — к каким ошибкам может привести замена параметров генеральной совокупности их точечными оценками по данным выборки, и с какой уверенностью можно ожидать, что ошибки не выйдут за известные пределы. Надежностью (доверительной вероятностью) оценки Θ^* неизвестного параметра Θ называется вероятность γ того, что выполняется неравенство $|\Theta^*-\Theta|<\delta$, то есть

$$P(\Theta^*-\delta<\Theta<\Theta^*+\delta)=\gamma.$$

Таким образом, γ есть вероятность того, что значение Θ попадает в интервал $\left(\Theta^*-\delta,\Theta^*+\delta\right)$, который и называется доверительным интервалом с надежностью γ [3].

Для решения задачи построения доверительного интервала используются центральные предельные теоремы — класс теорем теории вероятностей, утверждающих, что сумма достаточно большого количества слабо зависимых случайных величин, имеющих примерно одинаковые масштабы (ни одно из слагаемых не доминирует, не вносит в сумму определяющего вклада), имеет распределение, близкое к нормальному. В частности, используется теорема Ляпунова [5], согласно которой сумма достаточно большого количества случайных величин подчиняется нормальному закону распределения при выполнении следующих условий [5] — случайные величины имеют конечное математическое ожидание и дисперсию, и все значения случайных величин равномерно ограничены относительно математических ожиданий.

Пусть X — случайная величина, распределенная по нормальному закону с неизвестным математическим ожиданием M_X, тогда случайная величина

$$R = \frac{(\overline{X} - M_X)\sqrt{n}}{\mathrm{S}}$$

имеет распределение Стьюдента с $n-1$ степенями свободы [6], где $\overline{X}$ — выборочное среднее, S^2 — выборочная исправленная дисперсия, n — объем выборки. Обозначим значение M_X через a. Задача построения доверительного интервала для математического ожидания заключается в том, чтобы по заданной надежности γ и по числу степеней свободы $n-1$ найти такое число r_γ, чтобы выполнялось равенство

$$P\left(\left|\frac{(\overline{X} - a)\sqrt{n}}{S}\right| < r_\gamma\right) = \gamma,$$

или эквивалентное равенство

$$P(\overline{X} - \delta < a < \overline{X} + \delta) = \gamma, \ \delta = r_\gamma \frac{S}{\sqrt{n}}, \tag{2}$$

таким образом, мы получаем доверительный интервал для $a = M_X$. Значение r_γ вычисляется стандартными статистическими пакетами по вероятности γ и числу степеней свободы $n-1$. Отметим, что при $n > 50$ значение r_γ слабо зависит от n и можно считать, что для $\gamma = 0{,}95$ и при $n > 50$ значение $r_\gamma \approx 1{,}96$ [6].

Для построения доверительного интервала для дисперсии при неизвестном математическом ожидании используется теорема Фишера [6]. Пусть $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ — независимая выборка из нормального распределения с параметрами μ, σ^2, тогда случайная величина

$$H = \frac{(n-1)S^2}{\sigma^2},$$

имеет распределение χ^2 с $(n-1)$ степенями свободы. Данный закон распределения не симметричен, следовательно, левую и правую границы нужно определять отдельно. Тогда имеем [6]:

$$P\left(\chi^2_{\frac{1-\xi}{2}, n-1} \le H \le \chi^2_{\frac{1+\xi}{2}, n-1}\right) = \xi,$$

где ξ — доверительная вероятность, n — число степеней свободы. Критическое значение χ^2 вычисляется стандартным статистическим пакетом. После подстановки выражения для H и несложных алгебраических преобразований получаем:

$$P\left(\frac{(n-1)S^2}{\chi^2_{\frac{1+\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\frac{1-\alpha}{2},n-1}}\right) = \xi \,. \tag{3}$$

Таким образом, мы имеем доверительные интервалы для математического ожидания и дисперсии.

Будем использовать далее следующие обозначения — $M_{\min} = \overline{X} - \delta_M$, $M_{\max} = \overline{X} + \delta_M$ для границ интервала, в который с заданной доверительной вероятностью $\gamma_M = \gamma$ попадает значение M математического ожидания, а обозначения $D_{\min}$ и $D_{\max}$ — для границ доверительного интервала для значения дисперсии с доверительной вероятностью $\gamma_S = \xi$.

5. Интервальная оценка параметров бета-распределения

D соответствии с методикой, описанной авторами и В. Н. Петрушиным в [2] на основании формулы (1) заменой значений математического ожидания и дисперсии на их выборочные точечные оценки $\overline{X}$ и S^2 определяются точечные оценки параметров бета-распределения, аппроксимирующего наблюдаемую в эксперименте гистограмму распределения относительных частот значений трудоёмкости исследуемого алгоритма:

$$\alpha\left(\overline{X},S^2\right) = \frac{\left(1-\overline{X}\right)\overline{X}^2}{S^2} - \overline{X}\,,\ \beta\left(\overline{X},S^2\right) = \frac{\overline{X}(\overline{X}-1)^2}{S^2} + \overline{X} - 1\,.$$

Далее, определяется значение доверительной трудоёмкости на основе вычисления значения x_γ — γ-квантиля бета-распределения, имеющего идентифицированные параметры.

Содержательный интерес представляет исследование чувствительности этого γ-квантиля по изменению значений M и D в соответствующих доверительных интервалах. При этом значение γ-квантиля бета распределения однозначно определяется его параметрами α и β. Таким образом, задача состоит в исследовании экстремальных значений параметров бета-распределения при изменении возможных значений M и D в границах своих доверительных интервалов.

Вначале исследуем зависимость параметров бета-распределения от изменения дисперсии. Дифференцируя зависимости $\alpha(M,D)$ и $\beta(M,D)$, полученные в (1), по D, имеем:

$$\frac{\partial\alpha}{\partial D} = -\frac{M^2 - M^3}{D^2} = \frac{M^3 - M^2}{D^2}\,, \tag{4}$$

$$\frac{\partial \beta}{\partial D} = -\frac{M(M-1)^2}{D^2}, \tag{5}$$

при этом возможные значения M и D заключены в доверительные интервалы (2), (3) с соответствующими доверительными вероятностями

$$P(M_{\min} \le M \le M_{\max}) = \gamma_M, \; P(D_{\min} \le D \le D_{\min}) = \gamma_S.$$

При $M \in (0,1)$ обе производные (4), (5) монотонны и отрицательны. Из этого следует, что при $M = const$ функции $\alpha(D)$ и $\beta(D)$ убывают на всей области определения: $0 < D < 1/4$. На основе этого анализа найдем $\alpha_{\max}, \beta_{\max}$, $\alpha_{\min}$, $\beta_{\min}$ в доверительном интервале дисперсии, а именно:

$$\begin{aligned} \min\alpha &= \alpha(D_{\max}),\ \max\alpha = \alpha(D_{\min}) \\ \min\beta &= \beta(D_{\max}),\ \max\beta = \beta(D_{\min}). \end{aligned} \tag{6}$$

Для исследования зависимости параметров от математического ожидания продифференцируем зависимости $\alpha(M,D)$ и $\beta(M,D)$ по M:

$$\frac{\partial \alpha}{\partial M} = \frac{2M - 3M^2}{D} - 1, \; \frac{\partial \beta}{\partial M} = \frac{3M^2 - 4M + 1}{D} + 1. \tag{7}$$

Приравняв нулю полученные производные находим точки экстремума, а также минимальные и максимальные значения функций $\alpha(M)$ и $\beta(M)$ на границе доверительных интервалов. В этом случае мы имеем классическую задачу поиска экстремума в области, поскольку экстремальные значения могут лежать как на границах сегмента, так и внутри него.

Таким образом, могут быть найдены границы доверительных интервалов параметров α и β в зависимости от изменения математического ожидания в границах доверительного интервала. Для определения глобальных границ доверительных интервалов указанных параметров возможно использование численных значений границ доверительного интервала дисперсии.

Таким образом, могут быть определены значения $\alpha_{\min}$, $\alpha_{\max}$ и $\beta_{\min}$, $\beta_{\max}$, т. е. границы доверительных интервалов параметров α и β бета-распределения, соответствующие изменениям M и D в границах своих доверительных интервалов.

6. Чувствительность доверительной трудоёмкости

На основе полученных доверительных интервалов параметров бета-распределения можно получить и доверительный интервал для его γ-квантиля. Однако проблема состоит в том, что обратная интегральная функция бета-распределения не выражается в аналитических функциях, и, следовательно, мы не можем получить аналитическое решение для границ доверительного интервала γ-квантиля. Наиболее простое решение состоит в построении фрагмента функции $x_\gamma = x_\gamma(\alpha, \beta, \gamma)$, для которого значения α и β изменяются в полученных доверительных интервалах, а значение доверительной вероятности для функции трудоёмкости алгоритма — γ является наперёд заданным. Экспериментальные данные, полученные авторами при исследовании доверительной трудоёмкости ряда алгоритмов свидетельствуют о том, что наблюдаемые значения α и β удовлетворяют условию $\alpha >> 1$, $\beta >> 1$ (унимодальный случай с малой дисперсией), при этом, как правило, дисперсия падает с ростом длины входа алгоритма. Поскольку известно, что соотношение $\alpha < \beta$ приводит к левой асимметрии распределения, а $\alpha > \beta$ — к правой асимметрии, мы можем говорить о монотонности функции $x_\gamma = x_\gamma(\alpha, \beta, \gamma)$ в области таких значений аргументов. Следовательно область поиска экстремальных значений γ-квантиля может быть сужена до углов исследуемой области, т. е. точек с координатами границ доверительных интервалов для α и β.

7. Пример интервальной оценки

Модельным примером данного исследования был алгоритм сортировки методом вставки при длине входа $n = 100$. Для этого алгоритма при объёме выборки $k = 20000$ (массивы в 100 элементов генерировались стандартным генератором псевдосучайных чисел) авторами были получены следующие результаты, приведённые в таблице 1.

Табл. 1.

Параметр	Значение	Описание
$\overline{X}$	0,4988571945	Выборочное среднее.
S^2	0,0011619236	Выборочная исправленная дисперсия.
$M_{\min} = \overline{X} - \delta_M$	0,4983847727	Левая граница доверительного интервала математического ожидания при $\gamma_M = 0{,}95$.
$M_{\max} = \overline{X} + \delta_M$	0,4993296162	Правая граница доверительного интервала математического ожидания при $\gamma_M = 0{,}95$.
$D_{\min} = S^2 - \delta_s$	0,0011394748	Левая граница доверительного интервала дисперсии при $\gamma_s = 0{,}95$.
$D_{\max} = S^2 + \delta_s$	0,00118503289	Правая граница доверительного интервала дисперсии при $\gamma_s = 0{,}95$.

В данном примере точки глобального экстремума параметров α и β по M находятся вне границ доверительного интервала математического ожидания, и, следовательно, в этом интервале функции $\alpha(M)$ и $\beta(M)$ монотонны. Поскольку поведение α и β по D глобально монотонно, то экстремальные точки находятся на границах области доверительных интервалов для M и D. Путем несложных вычислений получаем:

$$\alpha_{\min} = 104{,}642065,\ \alpha_{\max} = 109{,}053051.$$

$$\beta_{\min} = 105{,}122870,\ \beta_{\max} = 109{,}551275.$$

Отметим, что относительная погрешность при использовании точечной оценки вместо интервальной на данном этапе составляет: $\delta_\alpha = 2\%$, $\delta_\beta = 2\%$.

Поскольку полученные значения α и β достаточно велики, то монотонность функции $x_\gamma = x_\gamma(\alpha, \beta, \gamma)$ при $\gamma = 0{,}95$ должна соблюдаться. Построенная поверхность $x_\gamma = x_\gamma(\alpha, \beta, \gamma)$ (см. рисунок 2.) в области доверительных интервалов для α и β подтверждает это предположение.

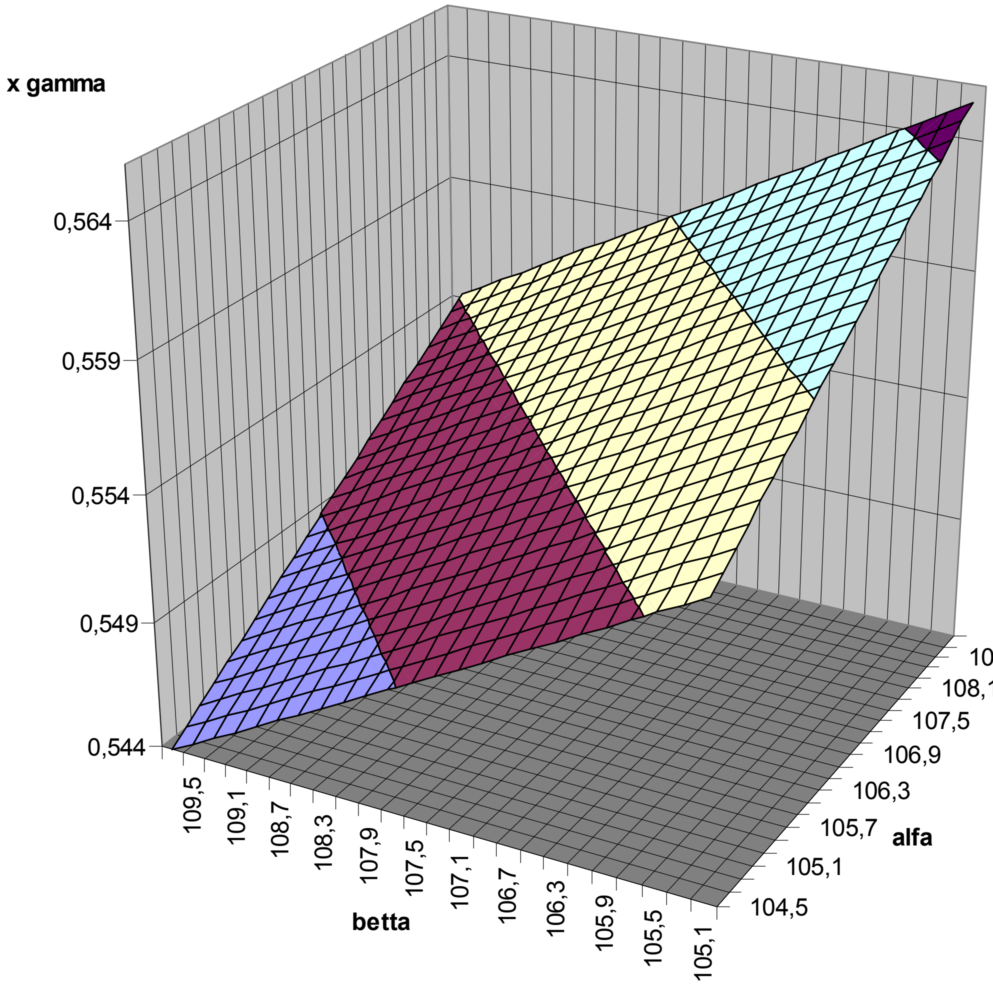

Рис. 2. Зависимость $x_\gamma = x_\gamma(\alpha, \beta, \gamma)$ при $\gamma = 0{,}95$ в области доверительных интервалов параметров бета-распределения.

Из свойства монотонности функции $x_\gamma = x_\gamma(\alpha, \beta, \gamma)$ в данной области и поведения асимметрии бета-распределения (с увеличением значения β и с уменьшением значения α растёт левая асимметрия, тем самым значение γ-квантиля уменьшается, и, наоборот) следует, что:

$$\max x_\gamma = x_\gamma(\alpha_{\max}, \beta_{\min}),\ \min x_\gamma = x_\gamma(\alpha_{\min}, \beta_{\max}).$$

Полученное решение подтверждается фрагментом поверхности $x_\gamma = x_\gamma(\alpha, \beta, \gamma)$ на рис. 2.

Для алгоритма сортировки методом вставок, результаты тестирования которого приведены в табл. 1 получаем

$$\min x_\gamma = 0{,}544668,\ \max x_\gamma = 0{,}565201,$$

и погрешность точечной оценки доверительной вероятности при доверительных вероятностях $\gamma_M = 0{,}95$ и $\gamma_s = 0{,}95$ относительно границ полученного интервала составляет всего 1,84%, что является вполне приемлемым.

8. Заключение

Таким образом, в статье описан подход к определению погрешности доверительной трудоемкости алгоритмов на основе доверительных интервалов для точечных оценок центральных моментов аппроксимирующей функции плотности распределения вероятности. Показано, что в диапазонах значений, актуальных при исследовании компьютерных алгоритмов, погрешность при использовании точечных значений выборочного среднего и выборочной исправленной дисперсии вместо их интервальных оценок, при определении доверительной трудоемкости не является значимой.

9. Библиографический список

1. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ — М.: «Вильямс», 2006. — ISBN 0-07-013151-1.
2. Ульянов М.В., Петрушин В.Н., Кривенцов А.С. Доверительная трудоемкость — новая оценка качества алгоритмов // Информационные технологии и вычислительные системы. 2009. №2. С. 23 – 37.
3. Королюк В.С., Портенко Н.И., Скороход А.В. Справочник по теории вероятностей и математической статистике — М.: Наука, 1985.
4. Петрушин В.Н, Ульянов М.В. Информационная чувствительность компьютерных алгоритмов. — М.: ФИЗМАТЛИТ, 2010. — 224 с. ISBN: 978-5-9221-1264-2
5. Прохоров Ю. В., Розанов Ю. А. Теория вероятностей (Основные понятия. Предельные теоремы. Случайные процессы). — М.: Главная редакция физико-математической литературы издательства «Наука», 1973. — 494 с.
6. Воронин В. Ф., Жильцова Ю. В. Статистика — М.: «Юнити-Дана», 2012. ISBN 978-5-238-02244-4.