

Вычисления обратных функций распределений: алгоритмы и программы

Л.В. Агамиров^{1,2}, В.Л. Агамиров^{1,2}✉, В.А. Вестяк¹

¹ Московский авиационный институт (Национальный исследовательский университет), г. Москва, 125993, Россия

² Московский технический университет связи и информатики (МТУСИ), г. Москва, 111024, Россия

Ссылка для цитирования

Агамиров Л.В., Агамиров В.Л., Вестяк В.А. Вычисления обратных функций распределений: алгоритмы и программы // Программные продукты и системы. 2024. Т. 37. № 2. С. 137–145. doi: 10.15827/0236-235X.142.137-145

Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 06.10.2023

После доработки: 13.11.2023

Принята к публикации: 21.11.2023

Аннотация. Точный расчет ряда функций распределения случайных переменных в прикладных задачах математической статистики вызывает существенные вычислительные трудности, обусловленные наличием бесконечных пределов интегрирования, необходимостью минимизации целевых функций и отсутствием удовлетворительных аппроксимаций. Для решения поставленных задач авторами данной статьи получены точные аналитические соотношения, позволяющие производить численное интегрирование функций распределения коэффициента вариации и нецентрального распределения Стьюдента, сводящееся к вычислению однократных интегралов. Обратные функции распределения определяются минимизацией на основе симплекс-метода Нелдера–Мида. Аналогично решается задача точного вычисления числовых характеристик порядковых статистик. Разработаны алгоритмы и программы на JavaScript с открытым кодом для реализации указанных вычислительных задач. Расчеты иллюстрируются графиками и таблицами, в которых представлены результаты вычисления квантилей нецентрального t -распределения Стьюдента в диапазоне объемов выборки от 3 до 50, вероятностей от 0,01 до 0,99 и доверительных вероятностей 0,9, 0,95 и 0,99. Время расчета в полном диапазоне всех параметров составляет не более 10–15 секунд на компьютере средней производительности, точность расчета – порядка 10^{-5} . Отмечается, что основные временные затраты составляет численное интегрирование в связи с наличием бесконечных пределов интегрирования, в то время как минимизация осуществляется довольно быстро (не более 20–30 итераций). В статье также представлены результаты вычисления квантилей относительных коэффициентов вариации для объемов выборки 3–10, генеральных коэффициентов вариации 0,05, 0,3, 0,5 и вероятностей в диапазоне от 0,01 до 0,99. Выполнены сравнительные расчеты числовых характеристик нормальных и weibullовских порядковых статистик, полученных прямым интегрированием и соответствующими аппроксимациями. В рассматриваемых программах применяются лишь простейшие и достаточно точные аппроксимации стандартных распределений: нормального распределения, гамма-функции, неполной гамма-функции. Разработанные алгоритмы пригодны для широкого класса непрерывных распределений, обратные функции которых не имеют приемлемых аппроксимаций.

Ключевые слова: алгоритмы и программы, JavaScript, нецентральное распределение Стьюдента, точное распределение коэффициента вариации, распределение порядковых статистик, нормальное распределение, распределение Вейбулла

Введение. В прикладных задачах математической статистики существует ряд случайных переменных, точный расчет функций распределения которых вызывает значительные вычислительные трудности из-за наличия бесконечных пределов интегрирования, необходимости минимизации целевых функций, отсутствия удовлетворительных аппроксимаций и т.п. К таким распределениям относятся нецентральное распределение Стьюдента, распределения коэффициента вариации, порядковых статистик и другие. Кроме задач вычисления мощности статистических критериев, значения квантилей нецентрального распределения Стьюдента необходимы в задачах надежности для обоснования толерантных интервалов и гарантированных ресурсных характеристик техни-

ческих систем [1]. Область применения коэффициента вариации в инженерных задачах рассматривается в работах [1, 2]. Доказательство точного распределения коэффициента вариации для нормальной выборки содержится в классической книге Лемана [3], все последующие работы – это различной точности аппроксимации этого распределения. Математические ожидания и ковариации порядковых статистик широко применяются в прикладных задачах математической статистики.

Анализ литературных источников по теме данного исследования, связанного с разработкой алгоритмов и программ для точного вычисления функций статистических распределений, показал, что их количество весьма ограничено. Так, например, в [4, 5] рассматривается

наиболее распространенное применение обратных функций – моделирование случайных величин методом обратных функций, а также разработка аналитических приближений для обращения функций вероятностных распределений [6]. Многие исследования посвящены новым методам представления аналитических аппроксимаций для прямых и обратных функций сложных нецентральных статистических распределений [7–9]. В то же время непосредственно связанные с темой данной работы источники практически отсутствуют. По мнению авторов, причина в том, что в настоящее время специалисты в области прикладной статистики используют в основном аппроксимации, созданные в математических пакетах с закрытым кодом типа Boost [10], Matlab, Statistica, Mathcad и в других. В свою очередь, как показал анализ документации по этим пакетам, они в значительной степени основаны на алгоритмах и программах прикладной статистики в рамках проекта Royal Statistical Society “Applied Statistics algorithms” (<http://lib.stat.cmu.edu/apstat/>), содержащего около 250 алгоритмов начиная с 1968 г. (проект завершился в 1997 г.), переведенных с языка Algol на Fortran, а затем на C++ (https://people.sc.fsu.edu/~jburkardt/cpp_src/cpp_src.html). Точные вычисления содержатся в уникальных статистических таблицах [11] и их русскоязычных аналогах (особенно в части нецентральных распределений) [12], которые, возможно, выполнены методами численного интегрирования в специализированных математических институтах, но недоступны для анализа кода и использования в задачах компьютерного моделирования. Кроме того, приведенные в табличном виде процентные точки обладают дискретностью, что вызывает необходимость интерполяции или экстраполяции, а это, в свою очередь, снижает точность расчетов. Очевидно, что при современном уровне развития информационных технологий использование таблиц является анахронизмом и может служить лишь для контроля точности численных расчетов, а также в учебных целях. Следует также упомянуть статистические программы, встроенные в популярный язык Python, которые полностью базируются на динамических библиотеках Boost с закрытым кодом. В отличие от Python эти библиотеки в стандарт C++ пока не встроены и требуют отдельной, часто трудоемкой установки. Тем не менее анализ некоторых открытых кодов статистических функций Boost показывает, что они основаны на многоуровневых аппроксима-

циях, выполненных многими исследователями в разное время с множеством условий. Преимуществами аппроксимаций Boost являются их высокая точность и практически мгновенное действие. Аппроксимации, встроенные в специализированный язык R (The R Project for Statistical Computing, <https://www.r-project.org/>), уступают по точности функциям Boost. Все сказанное касается, разумеется, сложных распределений, стандартные распределения в данной работе не рассматриваются.

Таким образом, исследования, направленные на разработку ПО для реализации точных алгоритмов вычисления сложных функций статистических распределений, являются актуальной задачей. При этом точность решений определяется точностью не аппроксимаций, а численного интегрирования и методов минимизации, которые можно неограниченно увеличивать с учетом лишь временных затрат.

Следует также отметить, что приемлемых по точности расчетов параметров порядковых статистик авторы не обнаружили, поэтому преимущество предлагаемых методов расчета прямым интегрированием сложных функций с последующей минимизацией целевых функций является не только точность вычислений (зачастую аппроксимации дают не меньшую точность), но и сам подход. Он позволяет распространять данную методику практически на любую схожую задачу с небольшими ограничениями типа непрерывности и дифференцируемости исходных функций, в то время как аппроксимацию следует полностью модифицировать для каждой конкретной задачи с тщательным последующим тестированием. Исключения, разумеется, встречаются, например, при численном интегрировании двойных интегралов с бесконечными пределами. В этих случаях альтернативы аппроксимациям нет из-за неоправданно больших временных затрат.

Целью настоящего исследования является разработка алгоритмов и программ расчета обратных функций сложных статистических распределений. Описываемые алгоритмы пригодны для широкого класса непрерывных распределений. В рассматриваемых программах применяются лишь простейшие и достаточно точные аппроксимации стандартных распределений. В настоящей работе это аппроксимации нормального распределения, гамма-функции, неполной гамма-функции, точность которых оказалась выше алгоритмов на Фортране (библиотека SSP) [13]. Авторы предлагают алгоритмы и программы на языке Javascript [14], ко-

торые выложены по адресу <http://g2.plzvpn.ru:3000/Rank/rank.html> (исходный программный код реализации по ссылке https://github.com/AVL095/Inverse_distribution_functions). Выбор языка обусловлен его общедоступностью и быстродействием.

Вышеуказанные задачи основаны на распределении отношения независимых случайных величин $x = \zeta_1/\zeta_2$:

$$F(x) = \int_0^\infty F_1(tx)f_2(t)dt \tag{1}$$

или

$$F(x) = 1 - \int_0^\infty f_1(t)F_2\left(\frac{t}{x}\right)dt, \tag{2}$$

где $f_1(t)$, $F_1(t)$ – функция плотности и функция распределения случайной величины ζ_1 ; $f_2(t)$, $F_2(t)$ – функция плотности и функция распределения случайной величины ζ_2 .

В соответствии с этим в выборке объема n из нормального распределения $N(a, \sigma)$ с выборочными средним $\hat{a} = \bar{x} = \sum_{i=1}^n x_i / n$ и дисперсией

$\hat{\sigma}^2 = s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$ случайная величина $t' = (\bar{x} - a + \delta)\sqrt{n} / s$ имеет нецентральное распределение Стьюдента с функцией распределения [3]

$$F(t') = \beta = \int_0^\infty f(x)\Phi(t'x, \delta, 1)dx, \tag{3}$$

откуда определяется квантиль распределения уровня β . Здесь

$$\Phi(x, a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left[-\frac{(t-a)^2}{2\sigma^2}\right]dt \tag{4}$$

функция распределения нормального закона; δ – параметр нецентральности; $f = n - 1$ – число степеней свободы. Распределение выборочного среднего подчиняется нормальному закону с параметрами $N(a, \delta/n^{0.5})$. Переменная $y = s^2(n - 1)/\delta^2$ имеет χ^2 распределение с f степенями свободы и плотностью

$$\varphi(y) = \frac{2^{-f/2}}{\Gamma(f/2)} y^{f/2-1} \exp(-y/2),$$

где $\Gamma(x)$ – гамма-функция (см. модуль *stat.js*). Плотность выборочного распределения стандартного отклонения $x = (s^2/f)^{0.5}$ в формуле (3) определяется на основании теоремы о плотности монотонной функции случайной величины:

$$f(x) = \varphi(fx^2)(fx^2)' = \frac{2(f/2)^{f/2}}{\Gamma(f/2)} x^{f-1} \exp(-fx^2/2). \tag{5}$$

Таким образом, распределение коэффициента вариации $\gamma = \sigma/a$ подчиняется нецентраль-

ному распределению Стьюдента с параметром нецентральности $\delta = (n)^{0.5}/\gamma$ и числом степеней свободы f [3], то есть $v_\beta = \delta/t'_\beta = \sqrt{n}/(t'_\beta \gamma)$.

Поэтому для вычисления квантилей коэффициента вариации достаточно иметь квантили нецентрального распределения Стьюдента. Прямое вычисление функции распределения выборочного коэффициента вариации сводится к уравнению

$$F(v) = \beta = 1 - \int_0^\infty f(x)\Phi\left(x\frac{v}{\gamma}, 1, \gamma/\sqrt{n}\right)dx, \tag{6}$$

откуда определяется квантиль распределения коэффициента вариации уровня β , γ – генеральное значение коэффициента вариации.

В настоящей работе анализировались оба подхода, по точности и быстродействию они показали идентичные результаты.

Для вычисления функций нецентрального распределения Стьюдента и коэффициента вариации необходимо численное интегрирование уравнений (3) или (6). Квантили распределения, соответствующие заданной вероятности, определялись последовательными приближениями. С этой целью в настоящей работе использовался симплекс-метод Нелдера–Мида (деформируемого многогранника).

Алгоритм расчета обратной функции нецентрального распределения Стьюдента реализован путем интегрирования уравнения (3) с последующей минимизацией квадратичной функции $q = (\beta - \beta_0)^2$, представляющей собой квадрат разности заданной β_0 и расчетной β доверительных вероятностей. Параметры функции *simpl* (модуль *stat.js*): $x[nx]$ – вектор размерности nx , на входе содержащий начальные приближения, на выходе – точки минимума; nx – число переменных минимизируемой функции (в рассматриваемом случае $nx = 1$); *step* – начальный шаг минимизации; *eps* – относительная точность выхода; *lim* – максимальное число итераций, функция *simpl* возвращает число выполненных итераций *iter*; *funx* – имя минимизируемой функции. Вызов функции: *iter = simpl(x, nx, stepx, eps, lim, funx)*.

Численное интегрирование уравнения (3) осуществлялось методом Буля (модуль *stat.js*). Параметры функции *IntegrateFunction*: k – номер выбранного метода, соответствующий размерности аппроксимирующего полинома (3, 4 – метод Симпсона, 5 – метод Буля); *nstep* – количество шагов интегрирования; *xl*, *xu* – нижний и верхний пределы интегрирования; *fpolinom* – имя интегрируемой функции, функция возвращает значение интеграла. Вызов функции: *beta =*

= *IntegrateFunction(k, nstep, xl, xu, fpolinom)*. Вспомогательные программы содержатся в модуле *stat.js*. С целью исключения возможных ошибок в программах используются лишь простейшие и проверенные конструкции языка JavaScript, включая ввод данных и вывод результатов на экран.

Аналогичный алгоритм реализован и для распределения коэффициента вариации, он отличается только типом интегрируемой функции (в данном случае (6)).

Как показали расчеты, быстрый и успешный поиск точек минимума в программе *simpl* существенно зависит от заданных на входе начальных приближений. С этой целью использовалась следующая нормальная аппроксимация квантиля [1] (функция *invnontap* в модуле *stat.js*):

$$t'(\beta, \delta, f) = \frac{\left(1 - \frac{1}{4f}\right)\delta + z_{\beta}\sqrt{\left(1 - \frac{1}{4f}\right)^2 - \frac{z_{\beta}^2}{2f} + \frac{\delta^2}{2f}}}{\left(1 - \frac{1}{4f}\right)^2 - \frac{z_{\beta}^2}{2f}}, \quad (7)$$

где δ – параметр нецентральности; f – число степеней свободы; z_{β} – квантиль нормированного нормального распределения уровня β ; β – доверительная вероятность.

Применительно к распределению коэффициента вариации в качестве первого приближения использовалась аппроксимация

$$\frac{v_p}{\gamma} = \sqrt{\frac{\chi_{p,f}^2}{(1+1/\gamma^2)f - \chi_{p,f}^2/(f+1)}}, \quad (8)$$

где v_p – квантиль распределения коэффициента вариации уровня вероятности p ; γ – генеральное значение коэффициента вариации; $\chi_{p,f}^2$ – квантиль распределения хи-квадрат [1].

Аппроксимация (8) дает существенные погрешности при значениях γ , больших 0,3.

Менее сложной с точки зрения программирования представляется задача точного вычисления числовых характеристик (математических ожиданий $E(x_r)$ и дисперсий $D(x_r)$) порядковых статистик x_r [2, 4, 5] в выборке объема n прямым интегрированием:

$$E(x_r) = \frac{\int_{-\infty}^{\infty} \varphi(x)x[1-F(x)]^{n-r}[F(x)]^{r-1} dx}{B(r, n-r+1)}, \quad (9)$$

$$D(x_r) = \frac{\int_{-\infty}^{\infty} \varphi(x)x^2[1-F(x)]^{n-r}[F(x)]^{r-1} dx}{B(r, n-r+1)} -$$

$$-E^2(x_r), \quad (10)$$

где $r = 1 - n$ – номер порядковой статистики; $\varphi(x)$ – плотность распределения; $F(x)$ – функция распределения; $B(a, b)$ – бета-функция. Численное интегрирование осуществляется с помощью функции *dataorder* в модуле *cvar.js*.

В таблице 1 показаны результаты вычисления квантилей нецентрального распределения Стьюдента. В программе варьировались объем испытаний n от 3 до 50, вероятности p от 0,01 до 0,99 и доверительные вероятности 0,9, 0,95 и 0,99. В таблице представлен фрагмент расчетов для n , равных от 3 до 10. Время расчета в полном диапазоне всех параметров составляет не более 10–15 секунд на компьютере невысокой производительности (индекс производительности – 1,0, ОЗУ – 4Гб, процессор – Intel® Pentium® 1,9 GHz). Точность расчета – порядка 10^{-5} . Сопоставления по быстродействию с аналогичными расчетами, выполняемыми приближенными методами, например, в библиотеках Boost, не анализируются, так как вычисления аппроксимаций составляют доли секунды.

Необходимо отметить, что изменение требуемой точности расчетов (*eps*), то есть относительной точности выхода в программе минимизации, практически не влияет на машинное время, поскольку при надлежащем задании начальных приближений программа очень быстро выходит на минимум. В результате вариация этого параметра в широком диапазоне значений приводит к одним и тем же точечным оценкам с незначительным увеличением числа итераций. Несколько большее значение имеют задаваемые предельное число итераций (*lim*) и шаг минимизации (*step*), но только тогда, когда программа вообще не находит минимума, что считается неудовлетворительным и подлежит исключению. В этом случае необходимо уточнить начальные приближения. Применительно к задачам, рассматриваемым в данной работе, такого не наблюдалось. Следует также отметить, что скорость вычислений на языке C++ в десятки раз выше, чем на Javascript и Python.

Основные временные затраты приходятся именно на численное интегрирование, а программа минимизации работает достаточно быстро (не более 20–30 итераций), что вполне объяснимо, так как в связи с наличием бесконечных пределов интегрирования для достижения требуемой точности необходимо охватить максимальный диапазон переменной независимо от применяемого метода. В рассматриваемых программах верхний предел интегрирования задавался равным 5, но может изменяться пользователем в зависимости от требуемой точности расчетов.

Таблица 1

Квантили нецентрального распределения Стьюдента для объемов выборки $n = 3-10$, доверительной вероятности β и вероятностей p (параметр нецентральности $\delta = z_p n^{0.5}$)

Table 1

Quantiles of non-central Student distribution for sample sizes $n = 3-10$, confidence probability β , probabilities p (noncentrality parameter $\delta = z_p n^{0.5}$)

$\beta = 0,99$									
n	$P = 0,01$	$0,05$	$0,1$	$0,3$	$0,5$	$0,7$	$0,9$	$0,95$	$0,99$
3	-1,3536792	-0,5105654	0,1245945	2,8786357	6,9645544	13,0712125	24,2407483	30,0860653	41,3883239
4	-1,8472004	-0,8857286	-0,2451913	1,8985941	4,5407048	8,223549	14,7597818	18,1669008	24,7745658
5	-2,296835	-1,214473	-0,5315038	1,4986883	3,7469478	6,7518861	11,9891732	14,7096064	19,9882608
6	-2,7131195	-1,5144931	-0,7810481	1,2576921	3,3649308	6,1001286	10,8048983	13,2408281	17,9659398
7	-3,1028759	-1,7934666	-1,0080186	1,0830341	3,1426678	5,7580231	10,210297	12,5087421	16,9644092
8	-3,4708345	-2,0558281	-1,2188556	0,9431022	2,9979522	5,5626035	9,8915944	12,1205258	16,43825
9	-3,8204406	-2,3045162	-1,4171848	0,8241254	2,8964601	5,4469441	9,7212346	11,9167803	16,166635
10	-4,1543004	-2,5416339	-1,6053317	0,7190867	2,8214376	5,3788426	9,6383305	11,8215914	16,0445284
$\beta = 0,95$									
3	-1,9566387	-1,1070312	-0,5793325	1,0051572	2,919986	5,6762818	10,6612588	13,2604081	18,2778621
4	-2,4923255	-1,4866088	-0,8877785	0,7109377	2,3533635	4,5308156	8,3238654	10,2877502	14,0847225
5	-2,9760095	-1,8286089	-1,1600269	0,5300989	2,1318469	4,1607664	7,6174653	9,3974797	12,8374608
6	-3,4205601	-2,1427448	-1,4080225	0,3909637	2,0150488	4,0125171	7,3637937	9,0819335	12,3992868
7	-3,8343448	-2,4350796	-1,6378358	0,2733724	1,9431808	3,9544098	7,2901881	8,9941472	12,2808368
8	-4,2231165	-2,7097237	-1,8532134	0,169149	1,8945787	3,9403276	7,3027406	9,0150293	12,3145682
9	-4,5910212	-2,9696243	-2,0567122	0,0742025	1,8595475	3,9502649	7,3612672	9,093709	12,4290667
10	-4,9411662	-3,2169811	-2,2501852	-0,0138155	1,8331132	3,974285	7,4460279	9,205277	12,5893981
$\beta = 0,90$									
3	-2,3570178	-1,4550122	-0,9262696	0,4518083	1,8856184	3,8590338	7,3753589	9,1997526	12,7140114
4	-2,9103929	-1,844764	-1,2341415	0,2596394	1,6377444	3,3860116	6,3756891	7,9131314	10,8764665
5	-3,4089776	-2,196217	-1,5099056	0,1169958	1,5332068	3,2551829	6,1320779	7,6022592	10,4334531
6	-3,8660398	-2,5186127	-1,7621781	-0,0031124	1,4758845	3,2272904	6,1082676	7,5735243	10,3920429
7	-4,2904886	-2,8181408	-1,996231	-0,1097249	1,4397552	3,2413693	6,1716037	7,6562707	10,5089838
8	-4,6884925	-3,0990964	-2,215594	-0,2070261	1,414924	3,2754144	6,2751304	7,7902943	10,6986604
9	-5,0645002	-3,3645879	-2,4227766	-0,2973447	1,3968153	3,3197823	6,3986238	7,9497059	10,9243219
10	-5,4218393	-3,6169423	-2,6196388	-0,3821411	1,3830284	3,369699	6,5322147	8,1219114	11,1680831

В таблице 2 представлены результаты вычислений квантилей относительных коэффициентов вариации v_p/γ для объемов выборки $n = 3-10$, генеральных коэффициентов вариации $\gamma = 0,05, 0,3$ и $0,5$ и вероятностей p в диапазоне от $0,01$ до $0,99$. Быстродействие и точность примерно эквивалентны приведенным в предыдущем примере.

На рисунке отображены графики функций распределения коэффициента вариации при $\gamma = 0,3$ для $n = 5, 10$ и 20 , построенные в соответствии с приведенной методикой.

Сравнение результатов расчета математических ожиданий и дисперсий нормальных порядковых статистик, полученных аппроксима-

цией Дэйвида–Джонсона (функция *ordern*, модуль *stat.js*), с порядком разложения в ряд $(n + 2)^3$ с табличными значениями показывают расхождение лишь в 4-м, 5-м знаке после запятой. В функции *ordern* содержатся производные функции распределения вплоть до 6-й, а также комментарии. Быстродействие настолько высокое, что в данной работе не анализируется.

Предлагаемое авторами прямое интегрирование функций распределения позволяет вычислять также числовые характеристики порядковых статистик для законов распределения, отличных от нормального. Функция распределения Вейбулла, например, с параметрами b и c имеет следующий вид:

Таблица 2

**Квантили выборочных коэффициентов вариации v_p/γ для объемов выборки $n = 3-10$,
коэффициентов вариации γ и вероятностей p**

Table 2

**Quantiles of sample variation coefficients v_p/γ for sample sizes $n = 3-10$,
variation coefficients γ and probabilities p**

$\gamma = 0,05$									
n	$P = 0,01$	0,05	0,1	0,3	0,5	0,7	0,9	0,95	0,99
3	0,1002102	0,2263956	0,3244862	0,5971515	0,8326885	1,0979006	1,51971	1,7344276	2,1533418
4	0,1955307	0,3422886	0,4411568	0,6887522	0,8881656	1,1058458	1,4454562	1,616912	1,9503398
5	0,2723549	0,421281	0,5154112	0,7405768	0,9161456	1,1048769	1,3962501	1,5426097	1,8265554
6	0,3327173	0,4783546	0,5672227	0,7744235	0,9329624	1,1017842	1,3606271	1,4901799	1,7410467
7	0,3809654	0,521763	0,6057963	0,7985351	0,9441737	1,0982481	1,3332994	1,4506201	1,6774501
8	0,4204085	0,5561038	0,6358691	0,816735	0,9521778	1,0947712	1,3114685	1,4193896	1,6277746
9	0,4533285	0,5841021	0,6601271	0,8310509	0,9581766	1,0915088	1,2935028	1,3939159	1,587587
10	0,4812971	0,6074751	0,6802103	0,8426637	0,9628401	1,0884974	1,278376	1,3726192	1,5542125
$\gamma = 0,3$									
3	0,0988091	0,2234869	0,3208023	0,5945781	0,8371579	1,1205464	1,6048203	1,8737959	2,4638789
4	0,1916642	0,3363389	0,4345316	0,6843881	0,8914472	1,1257984	1,5156774	1,7275832	2,1801959
5	0,266242	0,413217	0,507012	0,7353899	0,9187356	1,122733	1,4564944	1,6353082	2,0107654
6	0,3248317	0,468889	0,5577623	0,7688484	0,935102	1,1180122	1,4137562	1,5705593	1,8958781
7	0,3717281	0,5113646	0,5956947	0,7927851	0,9459959	1,1131818	1,3810865	1,5220013	1,8117507
8	0,4101458	0,5450793	0,6253817	0,8109232	0,9537644	1,1086514	1,3550862	1,4838882	1,7468857
9	0,442285	0,5726575	0,649414	0,8252398	0,959582	1,1045128	1,3337646	1,4529646	1,6949849
10	0,4696559	0,5957507	0,6693761	0,8368895	0,9641011	1,1007598	1,3158731	1,427234	1,6522852
$\gamma = 0,5$									
3	0,0963902	0,2184256	0,3143173	0,5895348	0,844215	1,1621726	1,7900129	2,2137372	3,5256371
4	0,1853166	0,3264466	0,4233633	0,6764026	0,8962664	1,1613484	1,6601829	1,9763649	2,8335787
5	0,2564419	0,4000839	0,4931283	0,7261626	0,9223899	1,1540322	1,5763381	1,8337191	2,4843834
6	0,312353	0,4536427	0,5422898	0,7590871	0,9380435	1,1461675	1,5170257	1,7369937	2,2692031
7	0,3572225	0,4947248	0,5792782	0,7828211	0,9484576	1,138908	1,4724104	1,6662827	2,1213024
8	0,3941057	0,527509	0,6084066	0,800924	0,9558808	1,132438	1,4373606	1,6118752	2,0123565
9	0,4250757	0,5544631	0,6321201	0,8152957	0,9614377	1,1267084	1,4089261	1,5684337	1,9281692
10	0,4515489	0,5771415	0,6519189	0,8270491	0,9657532	1,1216241	1,3852797	1,5327653	1,8607923

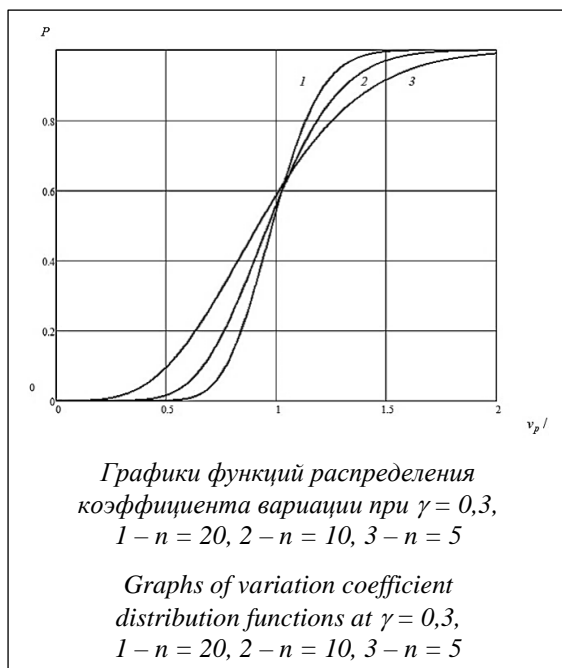
$$F(x) = 1 - \exp\left(-\frac{x}{c}\right)^b. \tag{11}$$

Приводя распределение к виду с параметрами сдвига $a_w = lnc$ и масштаба $\sigma_w = 1/b$, $z = (\ln x - a_w)/\sigma_w$, получим соотношения для функции и плотности распределения:

$$F(z) = 1 - e^{-e^z}; \varphi(z) = \frac{\partial F(z)}{\partial z} = ze^{z-e^z}, \tag{12}$$

которые подставляются в (9) и (10). Сравнительные расчеты числовых характеристик по-

рядковых статистик распределения Вейбулла по интегралам (9), (10) (функция *dataorder* в модуле *cvar.js*) и приближенных, основанных на аппроксимации Дэйвида–Джонсона (функция *orderw*, модуль *stat.js*), показали расхождение в 3-м, 4-м знаке после запятой. Пользователям рекомендуется самостоятельно принимать решение о применении той или иной модели в зависимости от требуемой точности расчетов. Предлагаемая аппроксимация позволяет также вычислять ковариации порядковых статистик.



Точное решение для ковариаций здесь не рассматривается, так как требует вычисления двойных интегралов. Поскольку распределение Вейбулла является несимметричным, следует вычислять все порядковые статистики от 1 до n в отличие от нормального закона.

Заключение

Разработанные алгоритмы и программы вычисления обратных функций нецентрального распределения Стьюдента и распределения коэффициента вариации основаны на сочетании численного интегрирования и симплекс-метода минимизации, позволяющих получать быстрые

и точные решения для применения данных функций в задачах прикладной статистики.

Приведенные в таблицах сравнительные расчеты показали высокую точность вычислений и особенности существующих аппроксимаций. Так, например, алгоритмы аппроксимации функции нецентрального распределения Стьюдента дают весьма точные приближения во всем диапазоне вероятностей, объемов выборки и параметров нецентральности и вполне пригодны для последующей минимизации с целью получения процентных точек. В то же время известная аппроксимация МакКау для коэффициента вариации дает погрешности до 15 % при значениях коэффициента вариации, больших 0,3.

Представленные в полном объеме и в открытом доступе модифицированные программы и аппроксимации на языке Javascript позволяют без труда распространять полученные решения для других статистических задач, связанных с точным распределением отношений двух независимых случайных переменных.

Разработанные программы интегрирования функций с бесконечными пределами для получения числовых характеристик порядковых статистик позволяют производить необходимые вычисления для большого класса непрерывных функций распределения. В работе рассмотрены нормированное нормальное и Вейбулла распределения, а также представлены аппроксимации методом Дэвида–Джонсона. Расчеты показали достаточно высокую точность разложения по сравнению с интегральными решениями (расхождения наблюдаются в 3-м, 4-м знаке после запятой).

Список литературы

1. Агамиров Л.В., Вестяк В.А. Вероятностные методы расчета показателей надежности авиационных конструкций при переменных нагрузках. М.: МАИ, 2022. 256 с.
2. Tothfalusi L., Endrenyi L. An exact procedure for the evaluation of reference-scaled average bioequivalence. AAPS J., 2016, vol. 18, no. 2, pp. 476–489. doi: 10.1208/s12248-016-9873-6.
3. Lehmann E.L. Testing Statistical Hypothesis. NY, John Wiley & Sons Publ., 1986, 388 p.
4. Холкина А.В. Моделирование случайных величин методом обратных функций // Математическое и программное обеспечение информационных, технических и экономических систем: матер. VII Междунар. молодежн. научн. конф. 2019. С. 73–75.
5. Царев А.Д. Программные генераторы случайных блужданий // Вестн. КРАУНЦ. Физ.-мат. науки. 2020. Т. 31. № 2. С. 226–235. doi: 10.26117/2079-6641-2020-31-2-226-235.
6. Попов Г.А. Формула обращения для рациональных характеристических функций вероятностных распределений // Вестн. АГТУ. 2018. Т. 2018. № 2. doi: 10.24143/1812-9498-2018-2-7-22.
7. Gil A., Segura J., Temme N.M. New asymptotic representations of the noncentral t-distribution. Studies in Applied Math., 2023, vol. 151, no. 3, pp. 857–882. doi: 10.1111/sapm.12609.
8. Gil A., Segura J., Temme N.M. On the computation and inversion of the cumulative noncentral beta distribution. Appl. Math. Comput., 2019, vol. 361, pp. 74–86.
9. Gil A., Segura J., Temme N.M. A new asymptotic representation and inversion method for the Student's t distribution. Integral Transforms and Special Functions, 2022, vol. 33, no. 8, pp. 597–608. doi: 10.1080/10652469.2021.2007906.

10. Полухин А. Разработка приложений на C++ с использованием Boost. М.: ДМК Пресс, 2020. 346 с.
11. Pearson E.S., Hartley H.O. Biometrika Tables for Statisticians. Cambridge University Press, 1976, 286 p.
12. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.
13. 1130 Sci. Subroutine Package. Programmer's Manual, Program Number 1130-CM-02X. IBM Corporation, 1967, 191 p.
14. Агамиров Л.В., Вестяк В.А. Программа вычисления обратных функций сложных статистических распределений: Свид. о регистр. ПрЭВМ № 2022612358. Рос. Федерация, 2022.

Software & Systems

doi: 10.15827/0236-235X.142.137-145

2024, 37(2), pp. 137–145

Calculating inverse distribution functions: Algorithms and programs

Levon V. Agamirov ^{1,2}, Vladimir L. Agamirov ^{1,2}✉, Vladimir A. Vestyak ¹¹ Moscow Aviation Institute (National Research University),
Moscow, 125993, Russian Federation² Moscow Technical University of Communications and Informatics,
Moscow, 111024, Russian Federation

For citation

Agamirov, L.V., Agamirov, V.L., Vestyak, V.A. (2024) 'Calculating inverse distribution functions: Algorithms and programs', *Software & Systems*, 37(2), pp. 137–145 (in Russ.). doi: 10.15827/0236-235X.142.137-145

Article info

Received: 06.10.2023

After revision: 13.11.2023

Accepted: 21.11.2023

Abstract. In applied problems of mathematical statistics, the exact calculation of some distribution functions of random variables causes significant computational difficulties due to infinite integration limits, the need to minimize objective functions, and the lack of satisfactory approximations. To solve these problems, the authors of the paper have obtained exact analytical relations that allow numerical integration of the distribution functions of a variation coefficient and the non-central Student distribution, which are reduced to calculating single integrals. Inverse distribution functions are determined by minimization based on the Nelder–Mead simplex method. The problem of accurately calculating the numerical characteristics of order statistics is solved in a similar way. The paper describes the developed algorithms and open source JavaScript programs to implement these computational tasks. The calculations are illustrated by graphs and tables that present the results of calculating non-central Student t-distribution quantiles in a range of sample sizes from 3 to 50, probabilities from 0.01 to 0.99 and confidence probabilities of 0.9, 0.95 and 0.99. The calculation time in the full range of all parameters is no more than 10–15 seconds on an average-performance computer. The calculation accuracy is about 10^{-5} . It is noted that the main time expenditure is numerical integration due existing infinite integration limits, while minimization is quick (no more than 20–30 iterations). The paper also presents calculations results of relative variation coefficient quantiles for sample sizes of 3–10, general variation coefficients of 0.05, 0.3 and 0.5 and probabilities in the range from 0.01 to 0.99. There are also comparative calculations of the numerical characteristics of normal and Weibull order statistics obtained by direct integration and corresponding approximations. The programs under consideration use only the simplest and fairly accurate approximations of standard distributions: normal distribution, gamma function, incomplete gamma function. It is noted that the developed algorithms are suitable for a wide class of continuous distributions with their inverse functions that have no acceptable approximations.

Keywords: algorithms and programs, JavaScript, non-central Student distribution, variation coefficient exact distribution, order statistics distribution, normal distribution, Weibull distribution

References

1. Agamirov, L.V., Vestyak, V.A. (2023) *Probabilistic Methods for Calculating Reliability Indicators of Aircraft Structures under Variable Loads*. Moscow, 256 p. (in Russ.).
2. Tothfalusi, L., Endrenyi, L. (2016) 'An exact procedure for the evaluation of reference-scaled average bioequivalence', *AAPS J.*, 18(2), pp. 476–489. doi: 10.1208/s12248-016-9873-6.
3. Lehmann, E.L. (1986) *Testing Statistical Hypothesis*. NY: John Wiley & Sons Publ., 388 p.
4. Kholkina, A.V. (2019) 'Modeling random variables using the inverse function method', *Proc. VII Int. Youth Sci. Conf. Mathematical Support and Software for Information, Technical and Economic Systems*, pp. 73–75 (in Russ.).
5. Tsarev, A.D. (2020) 'Random walk software generation', *Bull. KRASEC. Phys. and Math. Sci.*, 31(2), pp. 226–235 (in Russ.). doi: 10.26117/2079-6641-2020-31-2-226-235.

6. Popov, G.A. (2018) 'Formula of inversion for rational characteristic functions of probability distributions', *Bull. of ASTU*, 2018(2), pp. 226–235 (in Russ.). doi: 10.26117/2079-6641-2020-31-2-226-235.
7. Gil, A., Segura, J., Temme, N.M. (2023) 'New asymptotic representations of the noncentral t-distribution', *Studies in Applied Math.*, 151(3), pp. 857–882. doi: 10.1111/sapm.12609.
8. Gil, A., Segura, J., Temme, N.M. (2019) 'On the computation and inversion of the cumulative noncentral beta distribution', *Appl. Math. Comput.*, 361, pp. 74–86.
9. Gil, A., Segura, J., Temme, N.M. (2022) 'A new asymptotic representation and inversion method for the Student's t distribution', *Integral Transforms and Special Functions*, 33(8), pp. 597–608. doi: 10.1080/10652469.2021.2007906.
10. Polukhin, A. (2020) *Development of Applications in C++ Using Boost*. Moscow, 346 p. (in Russ.).
11. Pearson, E.S., Hartley, H.O. (1976) *Biometrika Tables for Statisticians*. Cambridge University Press, 286 p.
12. Bolshev, L.N., Smirnov, N.V. (1983) *Tables of Mathematical Statistics*. Moscow, 416 p. (in Russ.).
13. *1130 Sci. Subroutine Package. Programmer's Manual, Program Number 1130-CM-02X* (1967) IBM Corporation, 191 p.
14. Agamirov, L.V., Vestyak, V.A. (2022) *A Program for Calculating Inverse Functions of Complex Statistical Distributions*. Pat. RF, № 2022612358.

Авторы

Агамиров Левон Владимирович^{1,2}, д.т.н., профессор, itno_agamirov@mail.ru
Агамиров Владимир Леонович^{1,2}, к.т.н., доцент, avl095@mail.ru
Вестяк Владимир Анатольевич¹, д.ф.-м.н., доцент, заведующий кафедрой, kaf311@yandex.ru

¹ Московский авиационный институт (Национальный исследовательский университет), г. Москва, 125993, Россия

² Московский технический университет связи и информатики (МТУСИ), г. Москва, 111024, Россия

Authors

Levon V. Agamirov^{1,2}, Dr.Sc. (Engineering), Professor, itno_agamirov@mail.ru
Vladimir L. Agamirov^{1,2}, Cand. of Sci. (Engineering), Associate Professor, avl095@mail.ru
Vladimir A. Vestyak¹, Dr.Sc. (Physics and Mathematics), Associate Professor, Head of Chair, kaf311@yandex.ru

¹ Moscow Aviation Institute (National Research University), Moscow, 125993, Russian Federation

² Moscow Technical University of Communications and Informatics, Moscow, 111024, Russian Federation