

МЕТОДИКА ИСПОЛЬЗОВАНИЯ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ В ЗАДАЧЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ С ОГРАНИЧЕННОЙ ОБЛАСТЬЮ РАССЕЙЯНИЯ

¹Поршнеv С.В., ¹Копосов А.С.

¹ФГАОУ ВПО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Институт радиоэлектроники и информационных технологий – РТФ, Екатеринбург, Россия (620000, Екатеринбург, ул. Мира 32), e-mail: alexkopas@gmail.com

В статье обсуждаются результаты использования генетических алгоритмов в задаче оценки параметров распределений с ограниченной областью рассеяния. В качестве изучаемого закона распределения использовался обобщенный нормальный закон распределения в ограниченной области рассеяния. Модель данного закона основана на концепции мнимых источников. Показано, что нормальное распределение с ограниченной областью рассеяния является 5-параметрическим. Приведено описание методики применения генетических алгоритмов в задаче оценивания параметров 5-ти параметрической функции распределения по случайной выборке. Определены параметры генетического алгоритма, обоснование которых необходимо для решения поставленной задачи. Приведено обоснование выбора области поиска параметров нормального распределения с ограниченной областью рассеяния. Исследовано влияние некоторых наиболее распространенных настроек генетических алгоритмов на точность оценивания параметров изучаемого распределения. В результате исследования была предложена методика нахождения параметров данного распределения, основанная на использовании генетических алгоритмов, и получено подтверждение ее работоспособности. Получены оценки точности оценивания параметров нормального распределения с ограниченной областью рассеяния и интегрального показателя, характеризующего качество оценки изучаемого распределения в целом. Обоснован выбор настроек генетических алгоритмов, обеспечивающих наилучшую точность оценок параметров нормального распределения с ограниченной областью рассеяния.

Ключевые слова: функция распределения, плотность распределения, ограниченная область рассеяния, аппроксимация, метод мнимых источников, генетические алгоритмы, функция приспособленности

TECHNIQUE OF GENETIC ALGORITHMS USAGE IN AN ESTIMATION OF THE PARAMETERS OF DISTRIBUTION WITH A BOUNDED SCATTERING REGION PROBLEM

¹Porshnev S.V., ¹Koposov A.S.

¹Ural Federal University named after the first President of Russia B.N. Yeltsin, Institute of Radioelectronics and Information Technologies – RTF, Ekaterinburg, Russia (620000, Ekaterinburg, street Mira, 32), e-mail: alexkopas@gmail.com

In this article are considered the results of usage of genetic algorithms in an estimation of the parameters of distribution with a bounded scattering region problem. The generalized normal distribution law with a bounded scattering region was studied. This distribution law model is based on the imaginary sources model. It is shown that normal distribution with a bounded scattering region is a five-parameter distribution. The description of a technique of genetic algorithms usage in an estimation of the parameters of five-parameter probability distribution function problem is given. In this article are defined the parameters of genetic algorithm which are necessary for a problem solving. The argumentation of a choice of a distribution parameters searching area is given. This article gives an analysis of an influence of a most commonly used genetic algorithm tunes upon distribution parameters estimation accuracy. As the result of this research the technique of the parameters estimation of a generalized normal distribution is proposed. The technique is based on genetic algorithms. The confirmation of technique efficiency is received. This article provides estimation of calculation accuracy of the distribution parameters and integral index which characterizes estimation accuracy in common.

Keywords: cumulative distribution function, probability distribution function, bounded scattering region, approximation, imaginary sources method, genetic algorithms, fitness function

Введение

В работе [4] рассмотрена задача аппроксимации распределения содержания углерода в коксующихся углях, которое изменяется в диапазоне от ρ_{\min} до некоторого максимального значения ρ_{\max} . Рассматриваемая задача имеет важное практическое значение, поскольку параметры той или иной партии углей, оцениваемые по функции распределения углерода по конечному набору независимых тестовых выборок, например:

- размах возможных содержаний углерода, вероятности содержания углерода в возможном диапазоне;
 - вероятности содержания углерода, находящиеся в возможном диапазоне, в частности, вероятность нормативных пределов (обеспеченность);
 - фактические значения, отвечающие вероятности 95-процентного попадания измеренных величин в эти пределы (нормативная обеспеченность);
- используются при торгово-промышленной маркировке угля и, следовательно, от точности их оценок зависит отпускная стоимость данной партии угля.

В [4] обосновано, что адекватной моделью подобных распределений является не обычное гауссово (нормальное) распределение, но обобщенный нормальный закон с ограниченной областью рассеяния, который возникает, например, при описании случайных блужданий броуновской частицы в ограниченной области рассеяния. Плотность вероятности случайной величины с нормальным законом распределения с ограниченной областью рассеяния описывается следующим выражением [3]:

$$f_{LAD}(x; \mu, \sigma, l) = A \left[\varphi(x; \mu, \sigma, l) + \sum_{n=0}^{\infty} \varphi_{2n+1}^{\pm}(x; \mu, \sigma, l) + \sum_{n=1}^{\infty} \varphi_{2n}^{\pm}(x; \mu, \sigma, l) \right], \quad (1)$$

где A – нормировочный коэффициент, определяемый из условия

$$\int_a^b f_{LAD}(\xi; \mu, \sigma, l) d\xi = 1, \quad (2)$$

$$\varphi(x; \mu, \sigma, l) = \exp\left[-(x - \mu)^2 / 2\sigma^2\right],$$

$$\varphi_{2n+1}^{\pm}(x; \mu, \sigma, l) = \exp\left[-(x - x_{2n+1}^{\pm})^2 / 2\sigma^2\right],$$

$$\varphi_{2n}^{\pm}(x; \mu, \sigma, l) = \exp\left[-(x - x_{2n}^{\pm})^2 / 2\sigma^2\right],$$

$$x_{2n}^{\pm} = \pm 4nl + \mu, \quad x_{2n+1}^{\pm} = \pm(4n + 2)l - \mu - \text{координаты центров фиктивных источников}$$

l – размер области рассеяния,

μ – математическое ожидание случайной величины при отсутствии области ограничения,

σ – СКО случайной величины при отсутствии области ограничения.

Отметим, что, формально, распределение, является трех параметрическим (оно зависит от следующих параметров: μ – положение центра рассеяния, σ – СКО при отсутствии ограничения, l – размаха области рассеяния). Однако, на самом деле, если принять во внимание, что $l = x_{\max} - x_{\min}$, где x_{\min}, x_{\max} – координаты, левой и правой границ области рассеяния, знание значений которых оказывается весьма важным в практических приложениях. Также стоит отметить, что, хотя в выражении (1) присутствует бесконечная система фиктивных источников, для практических вычислений достаточно некоторого конечного числа N . Таким образом, распределение становится 5-параметрическим:

$$f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N) = A \left[\begin{array}{c} \varphi(x; \mu, \sigma, x_{\min}, x_{\max}) + \\ + \sum_{n=0}^N \varphi_{2n+1}^{\pm}(x; \mu, \sigma, x_{\min}, x_{\max}) \\ + \sum_{n=1}^N \varphi_{2n}^{\pm}(x; \mu, \sigma, x_{\min}, x_{\max}) \end{array} \right] \quad (3)$$

$$F_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N) = \int_a^x f_{LAD}(\xi; \mu, \sigma, x_{\min}, x_{\max}, N) d\xi. \quad (4)$$

Таким образом, задача состоит в следующем: априорно известно, что случайная величина описывается нормальным законом распределения с ограниченной областью рассеяния (4), требуется на основе экспериментальной выборки определить параметры $\mu, \sigma, x_{\min}, x_{\max}, N$ обобщенного нормального распределения с ограниченной областью рассеяния $f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N)$. Традиционно вектор параметров находят как решение задачи:

$$\bar{a} = \arg \min \left(\Phi \left(f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N) - \tilde{f}_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N) \right) \right) \quad (5)$$

где Φ – тот или иной функционал, вид которого зависит от выбранного критерия близости функций.

В общем случае, при любой выбранной мере приближения, нахождение ее глобального минимума в (5) приводит к необходимости решения той или иной системы нелинейных уравнений, у которой в рассматриваемом случае не удастся найти аналитические решения. Как следствие, здесь приходится использовать соответствующие численные методы (например, известен удачный опыт использования метода нелинейного

программирования при выборе в качестве оптимальной меры приближения метода наименьших квадратов [4]). Однако, сходимость итерационных методов к истинному решению оказывается существенно зависящей от выбора начального приближения вектора параметров, который в большинстве случаев является, вообще говоря, самостоятельной нетривиальной задачей. (С нашей точки зрения, данное обстоятельство явилось причиной того, что на практике модель обобщенного нормального распределения с ограниченной областью рассеяния не получила такого широкого распространения, как нормальный закон распределения или усеченный нормальный закон распределения [3]).

Таким образом, оказывается актуальным расширение круга методов, используемых для нахождения значений аргументов в (5), соответствующих глобальному минимуму функции $\Phi(f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N) - \tilde{f}_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, N))$. В этой связи, была изучена возможность использования в рассматриваемой задаче генетических алгоритмов, относящихся к эвристическим методам случайного поиска, которые, как ожидается [5], позволяют находить искомые решения нелинейных систем практически при любых начальных условиях.

В статье описана методика использования генетических алгоритмов для нахождения параметров нормального распределения с ограниченной областью рассеяния.

Методика использования генетических алгоритмов в задаче оценивания параметров 5-ти параметрической функции распределения по случайной выборке

Обобщенная блок-схема работы классического генетического алгоритма представлена на рисунке 1а [5].

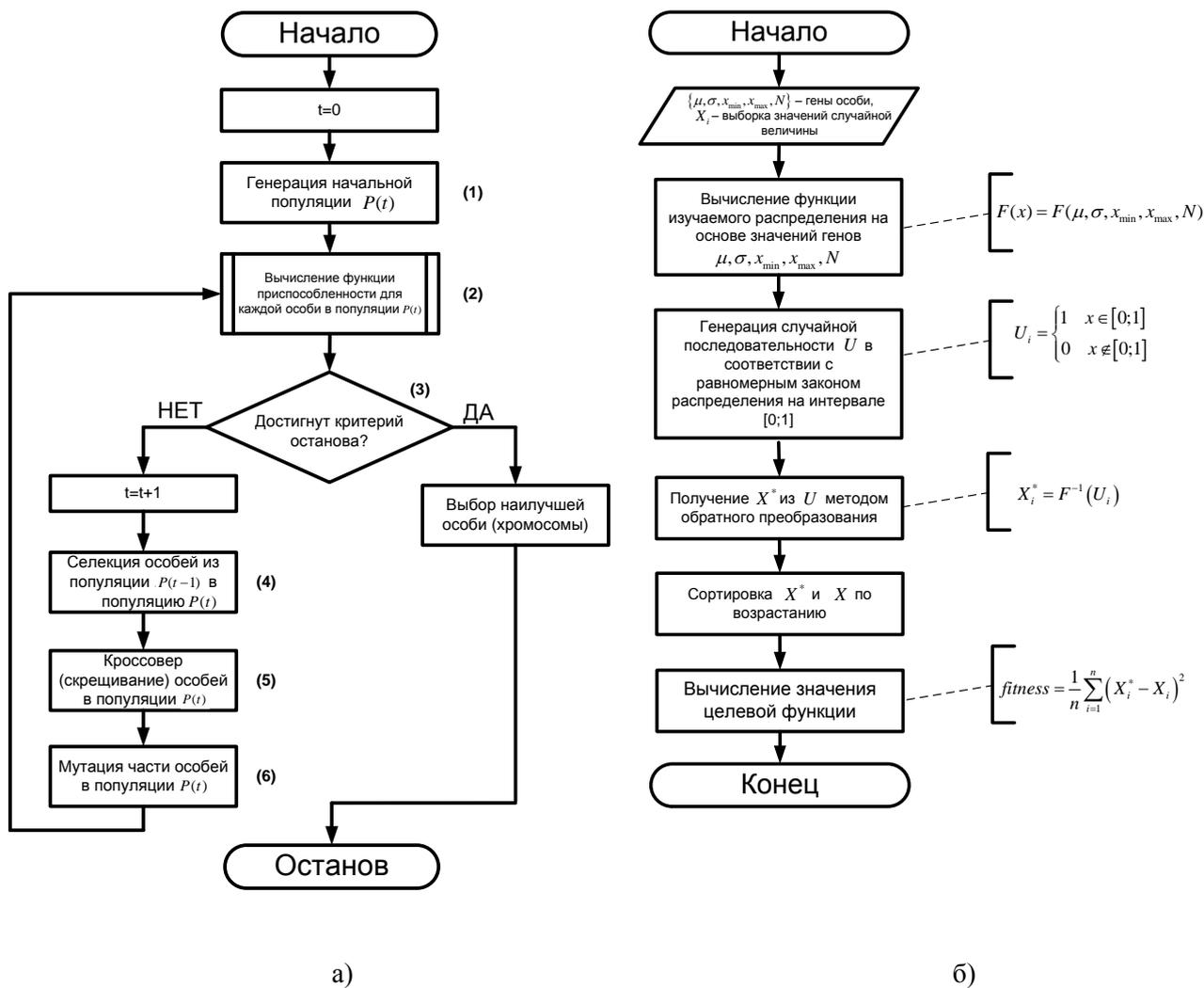


Рис. 1. Блок схемы классического генетического алгоритма (а) и алгоритма вычисления функции приспособленности (б): X - анализируемая случайная последовательность

Из рисунка 1а видно, что для применения ГА необходимо установить взаимно однозначное соответствие между основными структурными элементами: видом элемента популяции (особи), оператор кроссовера (скрещивания), мутации, селекции (отбора), видом функции приспособленности (фитнес-функции) и параметрами изучаемого распределения $\mu, \sigma, x_{\min}, x_{\max}, N$.

Напомним, что в ГА используется следующая иерархия понятий.

Популяция, которая состоит из **особей**. Каждая особь состоит из одной или нескольких **хромосом**, содержащих закодированное текущее решение рассматриваемой задачи (в данном случае, значения параметров распределения). Хромосома состоит из последовательности генов, в дискретном случае имеющих значения 0 или 1.

Рассматриваемая задача относится к классу оптимизационных задач в непрерывной области поиска, поэтому каждая хромосома представлялась 5-ти мерным вектором, координаты которого (гены) имели вещественные значения, соответствующие параметрам плотности распределения (3). Соответственно, отдельная особь в популяции состояла из одной хромосомы и пяти генов (рисунок 2а).

Далее решение данной задачи находилось выполнением следующей последовательности действий.

1. Формирование начальной популяции, состоящей из $popSize$ особей (блок 1 на рисунке 1а). (Число особей определяет производительность генетического алгоритма.) С математической точки зрения популяция есть матрица вещественных чисел P , размерности $popSize \times 5$, в которой каждая строка-вектор – отдельная особь (рисунок 2б).

Начальная популяция в ГА, как правило, инициализируется случайным образом: каждому гену присваивается случайное число, выбираемое из генеральной совокупности с известной функцией распределения. На практике обычно используются случайные величины с равномерным распределением в ограниченной области. Следовательно, в выбранном методе определение области ограничения значений генов также является важной задачей.

2. Вычисление для каждой особи в популяции значений **функции приспособленности** особей (блок 2 на Рисунке 1а)

$$\Phi(P_i) = \Phi(\{\mu_i, \sigma_i, x_{\min i}, x_{\max i}, N_i\}) \quad (6)$$

где P_i - i -ая особь в популяции P .

Блок-схема алгоритма вычисления функции приспособленности для особи $P_i = \{\mu_i, \sigma_i, x_{\min i}, x_{\max i}, N_i\}$ представлена на рисунке 1а.

Значения функции приспособленности используются для количественной оценки пригодности данного решения: если $\Phi(P_i) < \Phi(P_j)$, то особь P_i более приспособлена (лучше) особи P_j .

3. Проведение **селекции** (отбора) популяции (Блок 4 на рисунке 1а). **Селекция** – процесс, посредством которого более приспособленные особи получают большую возможность для воспроизводства потомков.

Основные виды селекции:

- *Селекция на основе рулетки.* При его реализации каждой особи в популяции соответствует зона на колесе рулетки, пропорционально соразмерная с величиной функции

приспособленности. Тогда при повороте колеса рулетки каждая особь имеет некоторую вероятность выбора для селекции

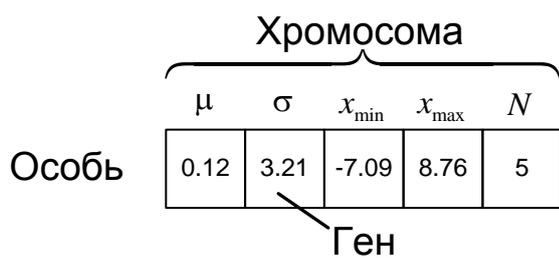
- *Элитная селекция.* В этом случае выбираются лучшие (элитные) особи на основе сравнения значений функции приспособленности.
- *Турнирная селекция.* Согласно размеру «турнира» случайно выбирается некоторое число особей, и лучшие особи в этой группе выбираются для воспроизведения.
- *Равномерная селекция.* Все особи имеют равные шансы для воспроизведения

В результате выполнения селекции из популяции P размером $popSize$ особей формируется временная «родительская» популяция P' в количестве $2 \cdot popSize$ особей, сгруппированных по парам. При этом родительскую популяцию образуют выделенные более приспособленные особи (рисунок 2в).

4. Применение к каждой паре родителей оператора **кроссовера** (блок 5 на Рисунке 1а). **Кроссовер (скрещивание)** – процесс, позволяющий на основе родительских особей получать потомков.

Основными видами кроссовера являются:

- *Одноточечный кроссовер.* Случайным образом выбирается точка в двух случайных хромосомах, где они должны быть «разрезаны». Меняя элементы после точки кроссовера, можно получить двух новых потомков (рисунок 2г).
- *Двухточечный кроссовер.* В каждой хромосоме выбираются случайным образом две точки кроссовера и хромосомы обмениваются генами.
- *Многоточечный кроссовер.* Выполняется аналогично.



а)



б)

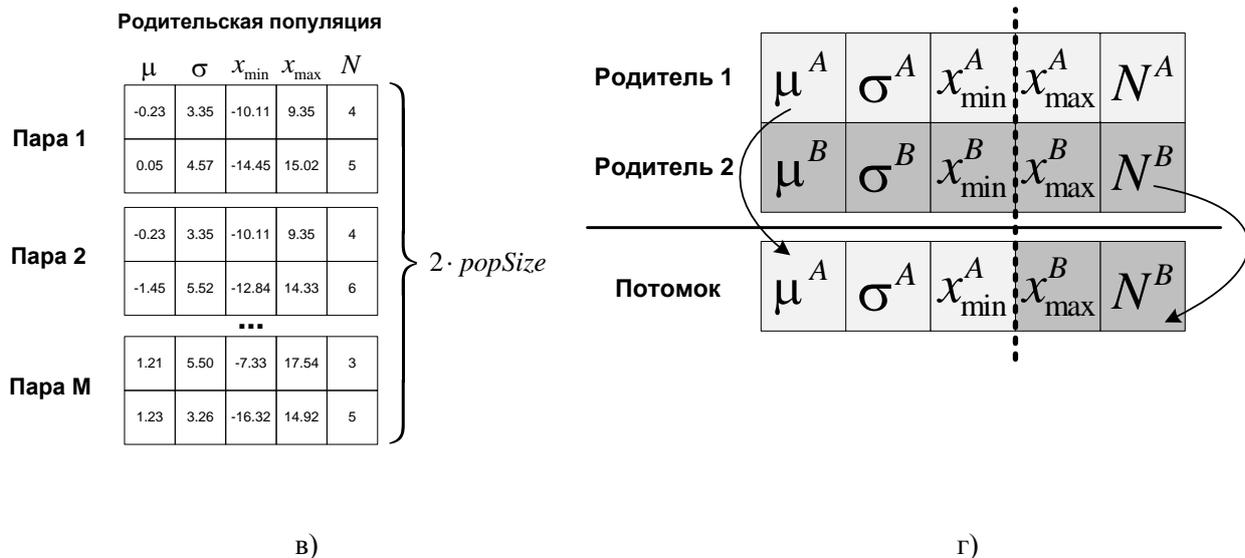


Рис. 2. К установлению соответствия между параметрами распределения и параметрами ГА: а – представление особи, б – пример популяции, в – родительская популяция, полученная в результате селекции, г – пример одноточечного кроссовера

Каждая родительская пара родителя производит одного потомка (Рисунок 2г). В результате из родительской популяции P' размером $2 \cdot popSize$ особей получают популяцию следующего поколения размером $popSize$.

5. Выполнение **мутации** (этап б на рисунок 1а). **Мутация** – процесс внесения случайных изменений в хромосомы, приводящий к генетическому разнообразию популяции и позволяющий генетическому алгоритму «выбираться» из локальных экстремумов и исследовать более обширное пространство параметров.

Здесь задается число особей $mutSize < popSize$, которое будет подвержено мутации. На каждой итерации генетического алгоритма после кроссовера случайным образом выбирается $mutSize$ особей из популяции P . Далее, для каждого гена выбранных особей задаются новые значения, выбираемые из генеральной совокупности случайных величин с заданным законом распределения.

6. Проверка выполнения критерия останова. За критерий останова можно принять появление в популяции особи с функцией приспособленности меньше заданного значения Φ_{\min} или выполнения условия

$$\bar{\Phi}(P_i) - \max_i(\Phi(P_i)) < \varepsilon, \quad (7)$$

где $\bar{\Phi}(P_i)$ - среднее значение функции приспособленности в популяции, ε - заданная граница колебания функции приспособленности.

Если критерий останова достигнут, завершение работы алгоритма, иначе повторение действий, описанных в пп. 2–5.

Из приведенного выше описания методики использования ГА для оценки параметров распределения (3) по случайной выборке видно, что для решения поставленной задачи требуется обоснование выбора следующих параметров ГА:

- границ области изменения параметров-генов;
- размера популяции.
- способа формирования начальной популяции;
- функции приспособленности особи;
- количества особей, подвергаемых процедуре мутации;
- оператора мутации;
- оператор селекции;
- оператора кроссовера.

Обоснование выбора области поиска параметров нормального распределения с ограниченной областью рассеяния

Рекомендации по выбору областей поиска параметров основаны на результатах моделирования броуновских блужданий в ограниченной области с абсолютно упруго отражающими стенками [2], согласующимися с результатами аналитического анализа рассматриваемого движения [1], состоят в следующем:

1. В качестве значений границ области рассеяния случайной величины X целесообразно использовать значения:

$$\tilde{x}_{\min} = \min(X_i),$$

$$\tilde{x}_{\max} = \max(X_i).$$

2. Нулевое приближение параметра μ нужно выбирать из отрезка $[\tilde{x}_{\min}, \tilde{x}_{\max}]$, где $\tilde{x}_{\min}, \tilde{x}_{\max}$ – минимальное и максимальное значения случайной последовательности $X_i, i = 1, M$.

3. Область поиска истинного значения параметра μ следует ограничить отрезком $[\tilde{x}_{\min}, \tilde{x}_{\max}]$.

4. Оценка нулевого приближения параметра σ следует получать на основе анализа зависимости

$$\langle x^2 \rangle = f(n), \quad (8)$$

где $\langle x^2 \rangle$ усредненное по ансамблю реализаций среднеквадратическое отклонение броуновской частицы, n – число шагов броуновской частицы, получаемой с помощью

статистического моделирования броуновского движения в соответствии с методикой, описанной в [2]. (Типичная зависимость $\langle x^2 \rangle$ от числа шагов броуновских блужданий представлена на рисунке 3.)

5. Область поиска истинного значения параметра σ следует ограничить отрезком $[0, \sigma_{\max}]$,

где $\sigma_{\max} = (\max f(n))^{0.5}$.

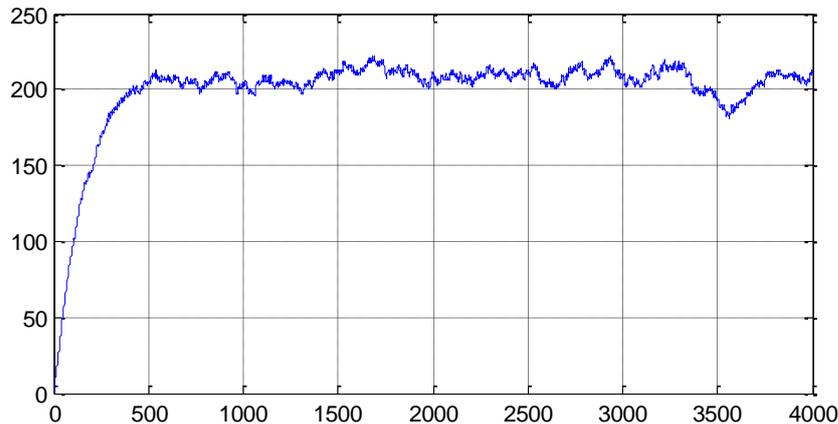


Рис. 3. Пример зависимость $\langle x^2 \rangle_n$ броуновского блуждания в ограниченной области

рассеяния: $x_0 = 0$, $x_2 = -x_1$, $|x_1| = |x_2| = 2l = 50$, $N = 1000$, $n = \overline{1, 4000}$

Необходимо отметить, что плотность распределения нормальной случайной величины (3), вообще говоря, формируется бесконечной системой фиктивных источников. Однако, расчеты показывают, что достаточная точность вычисления плотности вероятности в соответствие с (3) достигается при $N = 5$ пар мнимых источников (соответственно, число мнимых источников – $2N = 10$), а потому, указанную величину можно не оценивать с помощью ГА. Для иллюстрации данного утверждения на рисунке 4 представлена зависимость

$$\varepsilon(n, \mu, \sigma, x_{\min}, x_{\max}) = \ln \left[\max \left(\left| f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, n) - f_{LAD}(x; \mu, \sigma, x_{\min}, x_{\max}, 100) \right| \right) \right],$$

вычисленная для следующих значений параметров: $\mu = 0, \sigma = 0, x_{\min} = -5, x_{\max} = 5$.

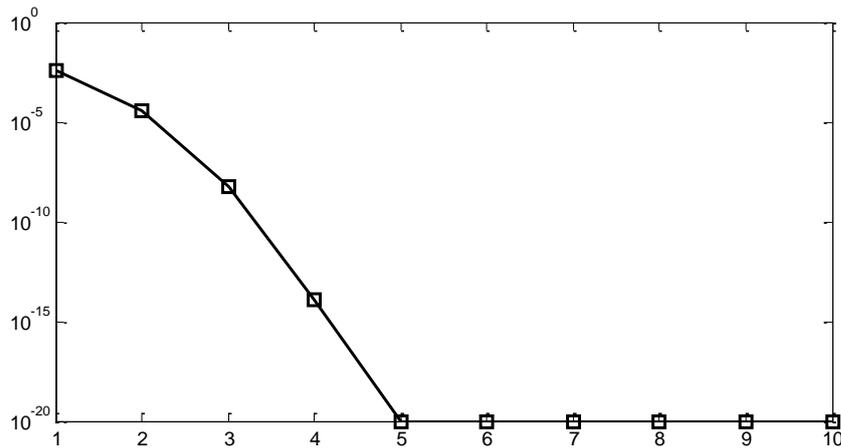


Рис. 4. График зависимости $\varepsilon(n, 0, 10, -5, 5)$

Таким образом, методика нахождения параметров нормального распределения случайной последовательности с ограниченной областью рассеяния по случайной выборке $X_i, i = \overline{1, N}$ реализуется выполнением следующей последовательности действий:

1. Вычисление оценки левой границы области рассеяния распределения (9)
 $\tilde{x}_{\min} = \min_i (X_i)$.
2. Вычисление оценки правой границы области рассеяния распределения (9)
 $\tilde{x}_{\max} = \max_i (X_i)$.
3. Задание количества пар фиктивных источников $N = 5$.
4. Задание интервала ограничения для параметра $\mu : [\tilde{x}_{\min}, \tilde{x}_{\max}]$.
5. Задать интервал ограничения для параметра $\sigma : [0, \sigma_{\max}]$, где σ_{\max} вычисляется по зависимости (8).
6. Вычислить, используя ГА, значения параметров μ, σ .

Анализ перечня действий, реализующих методику решения рассматриваемой задачи с помощью ГА, показывает, что использованные подходы к оценке ряда значений параметров распределения (3) позволяют, фактически, уменьшить размерность пространства параметров до двух – μ, σ .

Методика исследования влияния настроек ГА на точность оценивания параметров нормального распределения с ограниченной областью рассеяния

В ходе проведенных исследований рассмотрены были изучены нормальные распределения с ограниченной областью рассеяния со следующими параметрами: $n = 5$, $\sigma = 10$, $x_{\min} = -L/2$, $x_{\max} = L/2$, где L – размер области рассеяния, для $numL = 4$ различных

значений интервала рассеяния – $L \in [2\sigma \ 3\sigma \ 5\sigma \ 10\sigma]$, и $numMu = 3$ различных положений центра распределения (3) относительно границ интервала: $\mu \in \left[0 \ \frac{L}{4} \ \frac{L}{2}\right]$. Таким образом, было изучено $numSeq = numMu \times numL = 4 \times 3 = 12$ различных распределений нормальных случайных последовательностей с ограниченной областью рассеяния.

В исследовании использовались следующие настройки ГА:

- Селекция: $S = \langle \text{равномерная; турнирная; на основе рулетки} \rangle$;
- Мутация: $M = \langle \text{адаптивная} \rangle$ (т.к. в задаче присутствуют ограничения)
- Кроссовер: $C = \langle \text{одноточечный; двухточечный; усредненный; разбросанный, эвристический} \rangle$;
- Доля кроссовера: $Cfr = \langle 0,3; 0,6; 0,9 \rangle$;
- Доля мутации $1 - Cfr$;
- Размер популяции: $P = \langle 5, 10, 25 \rangle$.

Таким образом, каждый эксперимент характеризовался следующим кортежем:

$$\langle S_s, M_m, C_c, Cfr_f, P_p \rangle,$$

поэтому всего было рассмотрено

$$numT = numS \times numM \times numC \times numCfr \times numP = 3 \times 1 \times 5 \times 3 \times 3 = 135$$

различных комбинаций настроек.

Для каждого набора настроек было выполнено $numExp = 10$ независимых статистических испытаний, в ходе которых генерировались случайные последовательности $\{x_i\}^{(k)}, i = \overline{1, Np}$, ($Np = 500$) в соответствии с заданным законом распределения и, далее, следуя описанной выше методике, оценивались параметры данного распределения. Кроме того по каждому k -ому независимому испытанию оценивалась величина, характеризующее отличие между теоретической и экспериментальной функциями распределения:

$$\left[\Delta_{integr} \right]^{(k)} = \frac{\sum_i^{Np} \left(F_{teor} (x_i, \mu, \sigma, x_{\min}, x_{\max}, n) - \left[F_{pract} (x_i, \tilde{\mu}, \tilde{\sigma}, \tilde{x}_{\min}, \tilde{x}_{\max}, n) \right]^{(k)} \right)^2}{\sum_i^{Np} F_{teor} (x_i, \mu, \sigma, x_{\min}, x_{\max}, n)^2}.$$

Таким образом, общее число запусков ГА составило

$$numTotal = numT \times numSeq \times numExp = 40 \times 135 \times 10 = 54000.$$

Анализ результатов вычислительных экспериментов

Наилучшие результаты работы ГА для каждого набора параметров распределения выбранных из результатов k независимых испытаний представлены в табл. 1, 2.

Таблица 1

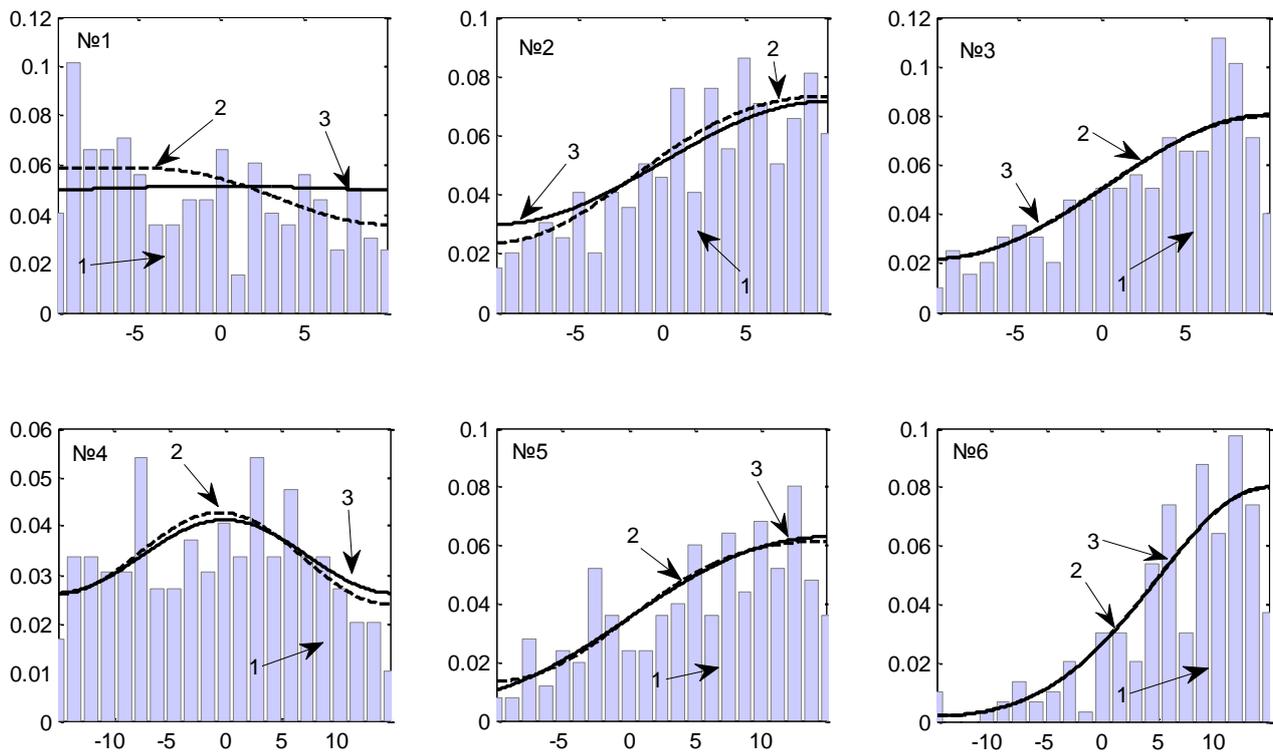
Результаты оценки параметров нормального распределения с ограниченной областью рассеяния

№ набора параметров	Параметры	μ	σ	x_{\min}	\tilde{x}_{\max}	Δ_{integr}
1	<i>Теоретические значения</i>	0.000	10.000	-10.000	10.000	–
	Наилучший результат	-1.473	7.968	-9.730	9.990	2.692·10⁻²
2	<i>Теоретические значения</i>	5.000	10.000	-10.000	10.000	–
	Наилучший результат	3.486	7.788	-9.910	9.870	3.920·10⁻³
3	<i>Теоретические значения</i>	10.000	10.000	-10.000	10.000	–
	Наилучший результат	5.166	8.547	-9.790	9.990	4.000·10⁻⁵
4	<i>Теоретические значения</i>	0.000	10.000	-15.000	15.000	–
	Наилучший результат	-0.371	9.481	-14.925	14.685	1.590·10⁻³
5	<i>Теоретические значения</i>	7.500	10.000	-15.000	15.000	–
	Наилучший результат	7.144	8.693	-9.915	14.955	4.600·10⁻⁴
6	<i>Теоретические значения</i>	15.000	10.000	-15.000	15.000	–
	Наилучший результат	11.281	9.258	-14.775	14.985	1.000·10⁻⁵
7	<i>Теоретические значения</i>	0.000	10.000	-25.000	25.000	–
	Наилучший результат	0.553	10.253	-20.775	24.775	3.320·10⁻³
8	<i>Теоретические значения</i>	12.500	10.000	-25.000	25.000	–
	Наилучший результат	12.370	9.945	-14.025	24.675	9.000·10⁻⁵
9	<i>Теоретические значения</i>	25.000	10.000	-25.000	25.000	–
	Наилучший результат	20.186	8.562	-2.625	24.875	9.000·10⁻⁵
10	<i>Теоретические значения</i>	0.000	10.000	-50.000	50.000	–
	Наилучший результат	0.072	9.658	-25.650	25.450	6.000·10⁻⁴
11	<i>Теоретические значения</i>	25.000	10.000	-50.000	50.000	–
	Наилучший результат	24.647	10.339	-0.650	45.350	4.590·10⁻³
12	<i>Теоретические значения</i>	50.000	10.000	-50.000	50.000	–
	Наилучший результат	47.359	9.736	13.850	49.950	1.000·10⁻⁴

Настройки генетического алгоритма для каждого набора параметров распределения

№	Селекция	Мутация	Кроссовер	Доля кроссовера	Размер популяции
1	равномерная	адаптивная	одноточечный	0.9	25
2	рулетка	адаптивная	двухточечный	0.6	25
3	равномерная	адаптивная	одноточечный	0.3	25
4	равномерная	адаптивная	усредненный	0.6	25
5	турнирная	адаптивная	разбросанный	0.6	25
6	турнирная	адаптивная	эвристический	0.9	25
7	равномерная	адаптивная	эвристический	0.3	25
8	турнирная	адаптивная	эвристический	0.3	25
9	равномерная	адаптивная	одноточечный	0.3	25
10	равномерная	адаптивная	одноточечный	0.3	25
11	равномерная	адаптивная	эвристический	0.6	25
12	турнирная	адаптивная	эвристический	0.6	25

Гистограммы случайных последовательностей, теоретические функции плотности распределения и наилучшие аппроксимированные функции плотности распределения, соответствующие наборам параметров, представленным в табл. 1, представлены на рисунке 5.



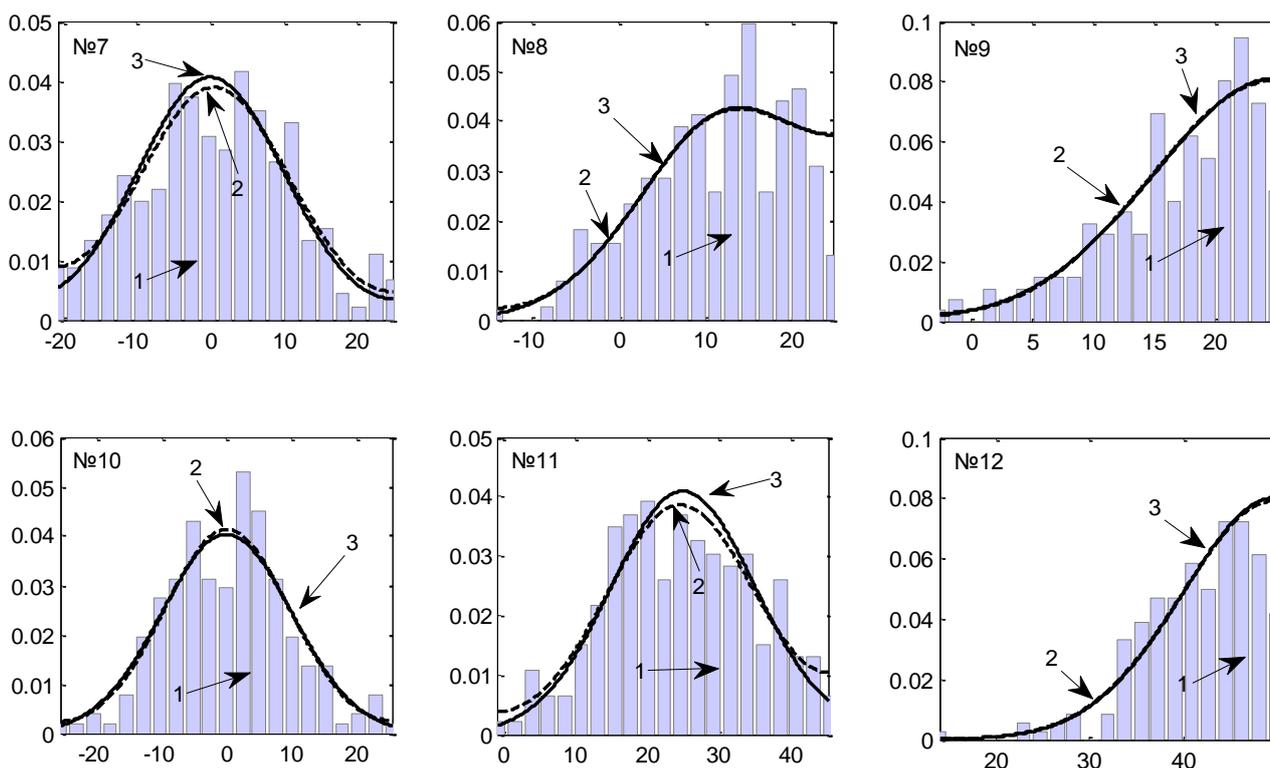


Рис. 5. Функции плотности распределения (3), при вычислении которых использованы значения параметров, представленные в табл. 1 и настройки ГА, представленные в табл. 2:

1 – гистограмма случайной последовательности, 2 – экспериментальная функция плотности распределения, 3 – теоретическая функция плотности распределения

Из таблиц 1, 2 и рисунка 5 видно, что наилучшие результаты получаются при следующих настройках ГА:

- селекция: **равномерная** или **турнирная**;
- мутация: **адаптивная**;
- кроссовер – **одноточечный** или **эвристический**;
- доля кроссовера находится в интервале от **0.2** до **0.7**;
- размер популяции: **25–50**.

Отметим, что увеличение размера популяции, увеличивает разнообразие входящих в нее особей, что, в свою очередь, увеличивает вероятность нахождения оптимального решения. Однако увеличение размера популяции пропорционально увеличивает объем, и, соответственно, время вычислений. В этой связи размер популяции приходится для каждой конкретной задачи выбирать эмпирически, исходя из соотношения между временем расчетов и точностью оценки параметров плотности нормального распределения с ограниченной областью рассеяния случайных последовательностей.

Выводы

Анализ результатов применения ГА в задаче нахождения параметров случайных последовательностей с нормальным распределением с ограниченной областью рассеяния позволяет сделать следующие выводы:

1. Предложена методика нахождения параметров данного распределения, основанная на использовании ГА, и получено подтверждение ее работоспособности.
2. Обоснованы рекомендации выбора областей поиска значений параметров нормального распределения с ограниченной областью рассеяния.
3. Получены оценки точности оценки параметров нормального распределения с ограниченной областью рассеяния и интегрального показателя, характеризующего качество оценки изучаемой плотности распределения в целом.
4. Обоснован выбор настроек ГА, обеспечивающих наилучшую точность оценок параметров нормального распределения с ограниченной областью рассеяния.

Работа выполнена в рамках договора № 02.G25.31.0055 (проект 2012-218-03-167).

Список литературы

1. Поршнева С.В., Копосов А.С. Аналитическое исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния // *Фундаментальные исследования*. – 2013. – № 4 (часть 1). – стр. 57-64.
2. Поршнева С.В., Копосов А.С. Исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния на основе статистического моделирования // *Фундаментальные исследования*. – 2013. – № 6 (часть 2) - стр. 284-290.
3. Поршнева С.В., Копосов А.С. О выборе математических моделей распределений ограниченных случайных // *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]*. – Краснодар: КубГАУ, 2012. – №10(84). – Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/53.pdf>, 1,000 у.п.л.
4. Поршнева С.В. Теория и алгоритмы аппроксимации эмпирических зависимостей и распределений / Е. В. Овечкина, В.Е. Каплан // Екатеринбург: УрО РАН, 2006. - 166 с.
5. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы / Под ред. В.М. Курейчика. 2-е изд., исправл. и доп. М.: ФИЗМАТЛИТ, 2010. - 368 с.

Рецензенты:

Кубланов В.С., д.т.н., доцент, профессор кафедры радиоэлектроники информационных систем ГАОУ ВПО «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина», г. Екатеринбург.

Доросинский Л.Г., д.т.н., профессор, заведующий кафедрой информационных технологий ГАОУ ВПО «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина», г. Екатеринбург.

Криштоп В.В., д.ф.-м.н., профессор, заведующий кафедрой «Физика», Дальневосточный государственный университет путей сообщения, г. Хабаровск, профессор Университета Kwangwoon University, Korea.