

УДК 004.942

UDC 004.942

О ВЫБОРЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ РАСПРЕДЕЛЕНИЙ ОГРАНИЧЕННЫХ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

ABOUT MAKING DECISIONS IN MATHEMATICAL MODELS OF DISTRIBUTIONS OF BOUNDED RANDOM SEQUENCES

Поршнеv Сергей Владимирович
д.т.н., профессор

Porshnev Sergey Vladimirovich
Dr.Sci.Tech., professor

Копосов Александр Сергеевич
аспирант
Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, Екатеринбург, Россия

Koposov Alexander Sergeevich
postgraduate student
Ural Federal University named after the first President of Russia B.N. Yeltsin, Ekaterinburg, Russia

В работе обоснована целесообразность использования в задаче построения математических моделей распределений ограниченных случайных последовательностей решений уравнения Фоккера-Планка, описывающих одномерное броуновское блуждание в ограниченной области

In this article we found out the usage reasonability of the solution of the Fokker-Planck equation for the problem of building of mathematical models of distributions of bounded random sequences

Ключевые слова: НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, УСЕЧЕННОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ С ОГРАНИЧЕННОЙ ОБЛАСТЬЮ РАССЕЙЯНИЯ

Keywords: NORMAL DISTRIBUTION, TRUNCATED NORMAL DISTRIBUTION, NORMAL RANDOM VARIABLE WITH LIMITED RANGE OF SCATTERING

1. Введение

Нормальное распределение случайной величины имеет неограниченную область рассеяния. В тоже время область значений случайных величин, анализ которых приходится проводить в естествознании, технике и экономике, оказывается конечной. Как следствие, распределения данных случайных величин отличаются от идеализированной модели с неограниченной областью рассеяния.

Наиболее очевидной физической моделью случайного процесса с ограниченной областью рассеяния служат песочные часы, в которых песок из «точечного» источника высыпается на горизонтальную плоскость, ограниченную непроницаемыми абсолютно упругими вертикальными стенками. При достаточном удалении стенок от источника они не будут оказывать влияния на формирование кучи песка, поэтому ее форма будет

симметричной. В противоположном случае песчинки будут отражаться от стенок, что приведет к отличию формы кучи от «нормальной».

В качестве примеров физических характеристик с ограниченной областью рассеяния можно привести плотность углей, изменяющейся в диапазоне от ρ_{\min} до некоторого максимального значения ρ_{\max} , время безотказной работы группы однотипных приборов, изменяющегося в диапазоне от некоторого минимального значения T_{\min} до некоторого максимального T_{\max} и т.д.

Следует отметить, что существует два альтернативных подхода к построению математических моделей обсуждаемых случайных величин. Первый подход основан на описании распределения случайной величины с помощью усеченных функций распределения [1,2]. Во втором подходе, базирующемся на работах А. Эйнштейна и М. Смолуховского по теории броуновского движения [3], модель статистического распределения случайной последовательности строится как распределение конечного состояния некоторого случайного процесса без последствия с ограниченной областью рассеяния [4]. Следует отметить, что сегодня в подавляющем большинстве учебников по теории надежности технических систем используется первый подход, как правило, без каких-либо обоснований подобного выбора [1,2]. В этой связи разработка сравнение обоих подходов к построению математических моделей распределений случайных последовательностей и выбор из них наилучшего является актуальной.

2. Классическая модель нормального закона

Эквивалентность решения уравнения Фоккера-Планка, описывающего процесс диффузии,

$$\frac{\partial}{\partial t} f(t, x) = D \frac{\partial^2}{\partial x^2} f(t, x), \quad (1)$$

где $f(t, x)$ – плотность распределения частиц $f(t, x) = \frac{\partial}{\partial x} P(t, x)$, $P(t, x)$ – вероятность нахождения частицы в момент времени t в точке x ; D – коэффициент диффузии; для неограниченной области рассеивания броуновских частиц было найдено А. Эйнштейном [5]

$$f(t, x) = \frac{n}{\sqrt{4\pi D}} \frac{e^{-\frac{x^2}{4Dt}}}{\sqrt{t}}, \quad (2)$$

здесь n – нормировочный коэффициент, выбираемый из условия $\int_{-\infty}^{\infty} f(t, x) dx = 1$. Действительно, трактуя величину $2Dt$ как дисперсию случайного процесса для конечного времени t , получаем математическую модель нормального закона распределения

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}. \quad (3)$$

Отметим, что аналогичный результат можно получить, рассмотрев броуновское движение точечной частицы на бесконечной прямой под действием случайных толчков [6], вызывающих с вероятностью p смещение в одну сторону, и с вероятностью q смещение в другую сторону. Вероятность $P_n(m)$ того, что после n случайных толчков частица окажется в m -ой точке рассчитывается по формуле

$$P_n(m) = \frac{n!}{\left(\frac{n+m}{2}\right)! \left(\frac{n-m}{2}\right)!} p^{\frac{m+n}{2}} q^{\frac{n-m}{2}}, \quad (4)$$

где $n+m$, $n-m$ – число толчков и, следовательно, элементарных смещений (шагов) вправо и влево, соответственно. При $p = q = \frac{1}{2}$

$$P_n(m) = C_n^{n+m} \left(\frac{1}{2}\right)^2. \quad (5)$$

Формула (5) с физической точки зрения может быть истолкована следующим образом [6]. Предположим, что в начальный момент времени в точке $x = 0$ имелось большое количество частиц. Далее частицы начинают перемещаться под действием случайных толчков независимо друг от друга. Тогда (5) дает долю частиц, оказавшихся на расстоянии m от начала отсчета.

При больших n для расчета вероятности $P_n(m)$, как следует из теоремы Муавра-Лапласа, можно использовать асимптотическую формулу

$$P_n(m) = \sqrt{\frac{2}{\pi n}} e^{-\left(\frac{m^2}{2n}\right)}. \quad (6)$$

Можно показать [7, С. 78], используя следующее порождающее уравнение

$$P_n(i) = \frac{P_{n+1}(i+1)}{2} + \frac{P_{n-1}(i-1)}{2}, \quad (7)$$

и полагая в нем $t = n\tau$, $x = ia$ и $P_n(i) = aP(t, x)$, где τ – время между шагами, a – величина единичного шага, что в непрерывном пределе $\tau \rightarrow 0$, $a \rightarrow 0$ при конечном $D = a^2/\tau$ получается уравнение Фоккера-Планка (1).

3. Распределение случайного броуновского движения с ограниченной областью рассеяния

На практике встречается целый ряд задач, в которых область изменения случайной величины оказывается ограниченной. Подобные ситуации возникают, например, в задачах: анализа случайных блужданий броуновской частицы в ограниченной области [5]; анализа содержания химических элементов в углях; анализа прочностных характеристик углей [4]. Функцию плотности вероятности случайной величины в данном случае

можно получить, проведя аналогию между рассматриваемой случайной величиной и одномерным броуновским движением частицы в ограниченной области. Особенности статистических свойств данного движения были изучены М. Смолуховским [5].

Наиболее наглядное решение обсуждаемой задачи для области рассеяния, ограниченной с правой стороны отражающей стенкой, расположенной в точке x_{\max} , изложено в [6, с. 103–106]. Искомая вероятность нахождения броуновской частицы в точке с координатой m в рассматриваемом случае вычисляется по формуле

$$P_n(m) \sim \frac{1}{\sqrt{2\pi n}} \left\{ e^{-\frac{m^2}{2n}} + e^{-\frac{(2x_{\max}-m)^2}{2n}} \right\}. \quad (8)$$

Из (8) видно, что вероятность нахождения броуновской частицы в произвольной точке m , может быть интерпретирована, как сложение действий двух источников (действительного и фиктивного), расположенных в точках $x_1 = 0$, $x_2 = 2x_{\max} - m$, соответственно.

Рассуждая аналогично, можно показать, что для броуновского движения с ограниченной областью рассеяния слева в точке x_{\min} вероятность нахождения броуновской частицы в точке с координатой m в данном случае вычисляется по формуле

$$P_n(m) \sim \frac{1}{\sqrt{2\pi n}} \left\{ e^{-\frac{m^2}{2n}} + e^{-\frac{(m-2x_{\min})^2}{2n}} \right\}. \quad (9)$$

Из (9) видно, что вероятность нахождения броуновской частицы в произвольной точке m , может быть интерпретирована, как сложение действий двух источников (действительного и фиктивного), расположенных в точках $x_1 = 0$, $x_2 = m - 2x_{\min}$, соответственно.

В связи с тем, что при практическом использовании модели броуновского движения в ограниченной области одним из основных оказывается вопрос о вычислении координат точек расположения фиктивных источников, рассмотрим его более подробно. Выберем систему координат с началом в середине отрезка области рассеяния (рис. 1).

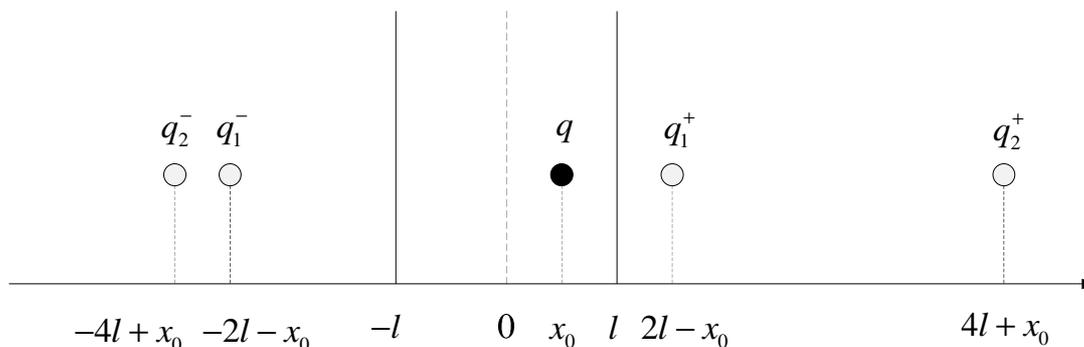


Рис. 1 – К вычислению координат фиктивных источников

Из рис. 1 видно, что наличие двух отражающих поверхностей порождает бесконечную систему фиктивных источников. Действительно, фиктивный источник q_n^+ , полученный отражением относительно плоскости $x=+l$ и находящийся справа от нее, в свою очередь отражается относительно плоскости $x=-l$, формируя фиктивный источник $q_{n+1}^- = q_n^+$. И наоборот, фиктивный источник q_n^- , расположенный слева от левой плоскости $x=-l$, отражается относительно правой обкладки в фиктивный источник $q_{n+1}^+ = q_n^-$. Дополним эти соотношения правилами преобразования координат фиктивных источников q_n^\pm :

$$\begin{aligned} q_{n+1}^+ &= q_n^-, & x_{n+1}^+ - l &= -(x_n^- - l), \\ q_{n+1}^- &= q_n^+, & x_{n+1}^- + l &= -(x_n^+ + l). \end{aligned} \tag{10}$$

Принимая во внимание «начальные условия» $q_0^+ = q_0^- = q$, $x_0^+ = x_0^- = x_0$, получаем из (10) следующие простые формулы:

$$x_{2n}^\pm = \pm 4nl + x_0, \quad x_{2n+1}^\pm = \pm(4n+2)l - x_0, \tag{11}$$

где $n=0,1,2,\dots$, по которым можно вычислить координаты любого из мнимых источников.

Таким образом, используя отдельно суммирование по четным и нечетным индексам, можно записать следующее выражение для плотности вероятности:

$$f_{LAD}(x; x_0, s, l) = A \left[j(x; x_0, s, l) + \sum_{n=0}^{\infty} j_{2n+1}^{\pm}(x; x_0, s, l) + \sum_{n=1}^{\infty} j_{2n}^{\pm}(x; x_0, s, l) \right], \quad (12)$$

где A – нормировочный коэффициент, определяемый из условия

$$\int_a^b f_{LAD}(x; x_0, s, l) dx = 1, \quad (13)$$

$$j(x; x_0, s, l) = \exp\left[-(x - x_0)^2 / 2s^2\right],$$

$$j_{2n+1}^{\pm}(x; x_0, s, l) = \exp\left[-(x - x_{2n+1}^{\pm})^2 / 2s^2\right],$$

$$j_{2n}^{\pm}(x; x_0, s, l) = \exp\left[-(x - x_{2n}^{\pm})^2 / 2s^2\right],$$

здесь $x_{2n+1}^{\pm}, x_{2n}^{\pm}$ вычисляются в соответствии с (11).

Соответственно, функция распределения вычисляется по формуле

$$F_{LAD}(x; x_0, s, l) = \int_a^x f_{LAD}(x; x_0, s, l) dx. \quad (14)$$

Отметим, что, формально, распределение (14), является трех параметрическим (оно зависит от следующих параметров: x_0 – положение центра рассеяния, σ – СКО при отсутствии ограничения, l – размаха области рассеяния). Однако, на самом деле, если принять во внимание, что $l = b - a$, где a, b – координаты, левой и правой границ области рассеяния, знание значений которых оказывается весьма важным в практических приложениях, распределение (14) следует отнести к классу 4-х параметрических распределений.

4. Усеченное нормальное распределение и его свойства

Усеченное нормальное распределение (УНР) традиционно используется в задачах оценки надежности технических систем [1] и точности производства [2]. Напомним, что УНР $N'(m, s)$ имеет случайная

подпоследовательность $\{x'_k\}$, извлеченная из случайной последовательности $\{x\}$, имеющей нормальное распределение $N(m,s)$, каждый элемент которой удовлетворяет следующим условиям:

$$x_{\min} \leq x'_k \leq x_{\max},$$

где x_{\min}, x_{\max} – точки усечения.

Основные характеристики УНР представлены в таблице

Таблица – Характеристики усеченного нормального распределения [2]

Название характеристики	Формула или числовое значение характеристики
Обозначение	$N'(m,s,q,x_{\min},x_{\max})$
Параметры	m,s,x_{\min},x_{\max}
Плотность распределения	$j'_N(x;m,s,a_1,a_2) = \frac{1}{s} \frac{j\left(\frac{x-m}{s}\right)}{F\left(\frac{x_{\max}-m}{s}\right) - F\left(\frac{x_{\min}-m}{s}\right)},$ <p>где $j(x)$ – плотность стандартного нормального распределения $N(0,1)$, $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$ – функция Лапласа, m,s – параметры исходного нормального распределения, $x_{\min} < x_{\max}$ – точки усечения</p>
Функция распределения	$F'_N(x;m,s,x_{\min},x_{\max}) = \int_{a_1}^x j'(x;m,s,x_{\min},x_{\max}) dx =$ $\frac{1}{s} \frac{1}{F\left(\frac{x_{\max}-m}{s}\right) - F\left(\frac{x_{\min}-m}{s}\right)} \int_{a_1}^x j\left(\frac{x-m}{s}\right) dx =$ $\frac{F\left(\frac{x-m}{s}\right) - F\left(\frac{a_1-m}{s}\right)}{F\left(\frac{x_{\max}-m}{s}\right) - F\left(\frac{x_{\min}-m}{s}\right)}$
Среднее значение	$M[x] = m - (I_2 - I_1)s,$ <p>где $I_1 = \frac{j(x_1)}{F(x_2) - F(x_1)}$, $I_2 = \frac{j(x_2)}{F(x_2) - F(x_1)}$, $x_1 = \frac{x_{\min} - m}{s}$, $x_2 = \frac{x_{\max} - m}{s}$</p>
Дисперсия	$D[x] = (1 + I_1x_1 - I_2x_2 + (I_1 + I_2)^2) s^2$

Из таблицы видно, что усеченное нормальное распределение $N'(m, s, q, a_1, a_2)$ также относится к классу четырех параметрических распределений.

5. Сравнение плотностей и функций распределения броуновского процесса с ограниченной областью рассеяния и усеченного нормального распределения

Из изложенного в предыдущих разделах понятно, что для случайных последовательностей, область значений которых из физических соображений является конечной, следует использовать, соответственно, модели (12) и (14). (Для нахождения их численного значения может быть, например, использован метод максимального правдоподобия, приводящий к поиску решения сложной системы нелинейных уравнений.)

Однако при нахождении параметров распределений (12) и (14) для реальных случайных последовательностей с неизбежностью приходится использовать не бесконечное, но некоторое конечное число мнимых источников, т.е. модели (12), (14), фактически, являются не четырех, а пяти параметрическими. Следовательно, с вычислительной точки зрения 4-х параметрическая модель усеченного нормального распределения является более предпочтительной. В этой связи представляется целесообразным сравнить плотности и функции распределения броуновского процесса с ограниченной областью рассеяния (количество источников $N = 200$) и усеченного нормального распределения, представленных на рис. 2–4.

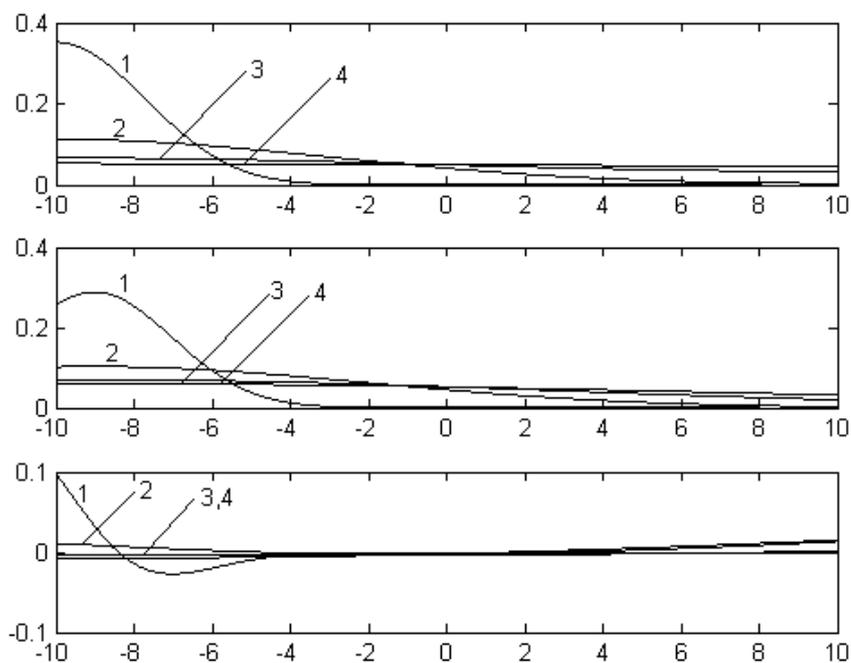


Рис. 2а – Вверху: плотность распределения, описываемая моделью (12); в середине: плотность усеченного нормального распределения; внизу: разность плотностей распределения (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = -9$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

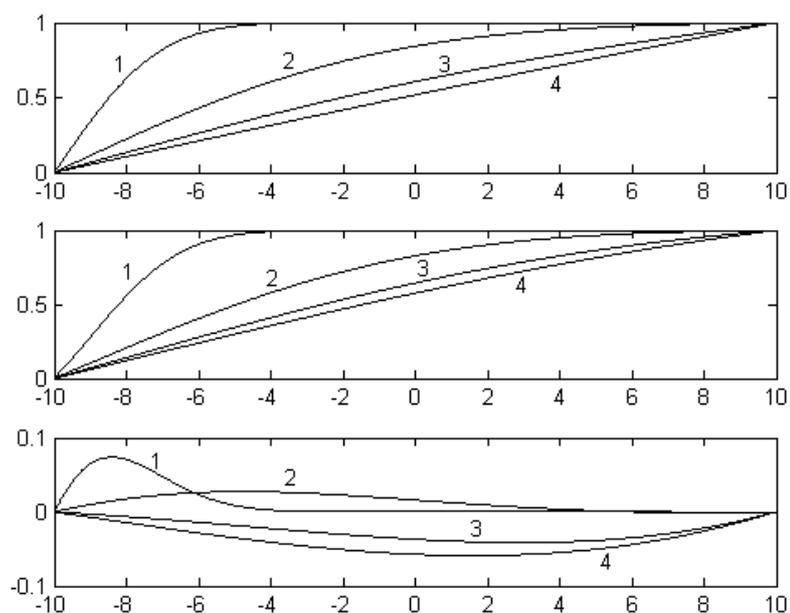


Рис. 2б – Вверху: распределение, описываемое моделью (14); в
 середине: усеченное нормальное распределение; внизу: разность
 распределений (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = -9$,
 $1 - \sigma = 2; 2 - \sigma = 7; 3 - \sigma = 12; 4 - \sigma = 17$)

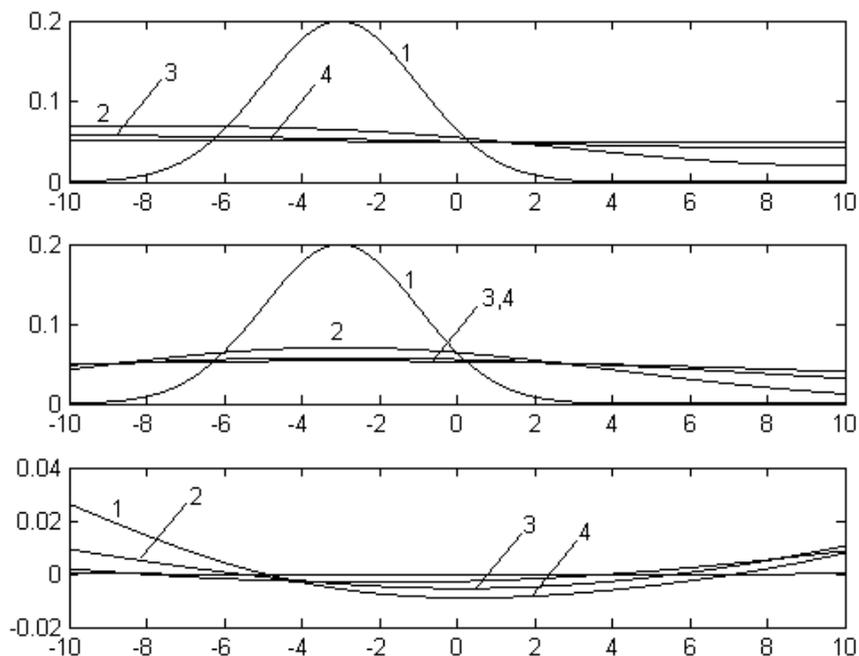


Рис. 3а – Вверху: плотность распределения, описываемая моделью
 (12); в середине: плотность усеченного нормального распределения; внизу:
 разность плотностей распределения (параметры функций: $x_{\min} = -10$,
 $x_{\max} = 10$, $\mu = -3$, $1 - \sigma = 2; 2 - \sigma = 7; 3 - \sigma = 12; 4 - \sigma = 17$)

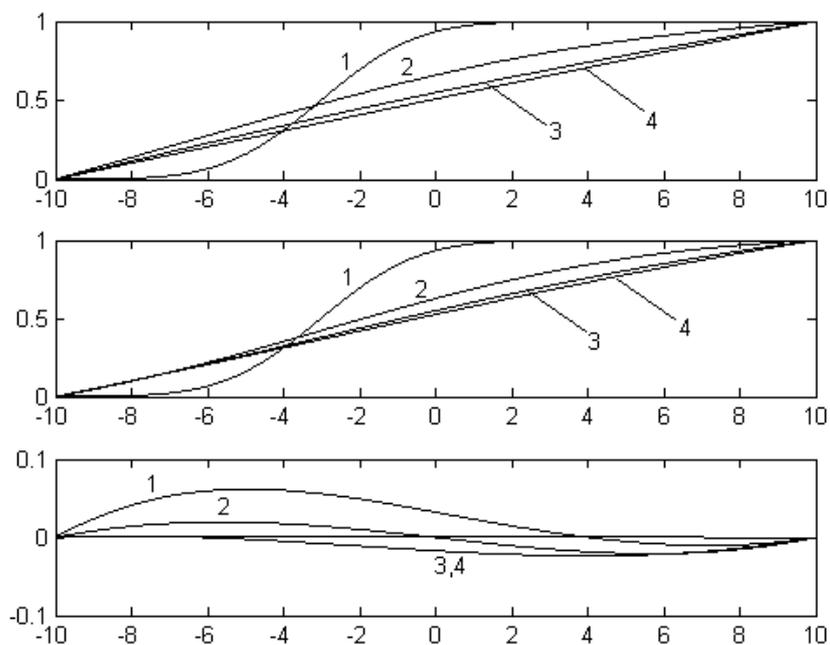


Рис. 3б – Вверху: распределение, описываемое моделью (14); в середине: усеченное нормальное распределение; внизу: разность распределений (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = -3$, $1 - \sigma = 2$; $2 - \sigma = 7$; $3 - \sigma = 12$; $4 - \sigma = 17$)

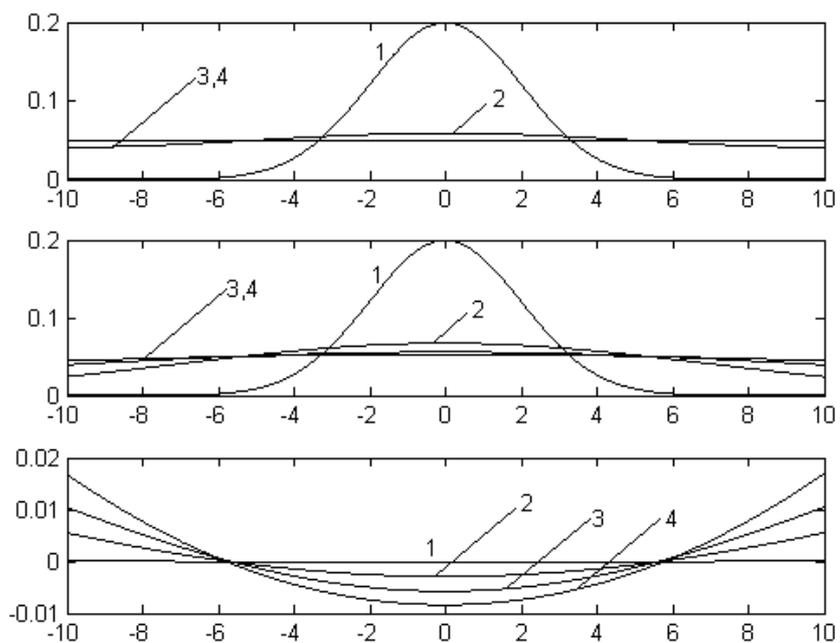


Рис. 4а – Вверху: плотность распределения, описываемая моделью (12); в середине: плотность усеченного нормального распределения; внизу: разность плотностей распределения (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = 0$, $1 - \sigma = 2; 2 - \sigma = 7; 3 - \sigma = 12; 4 - \sigma = 17$)

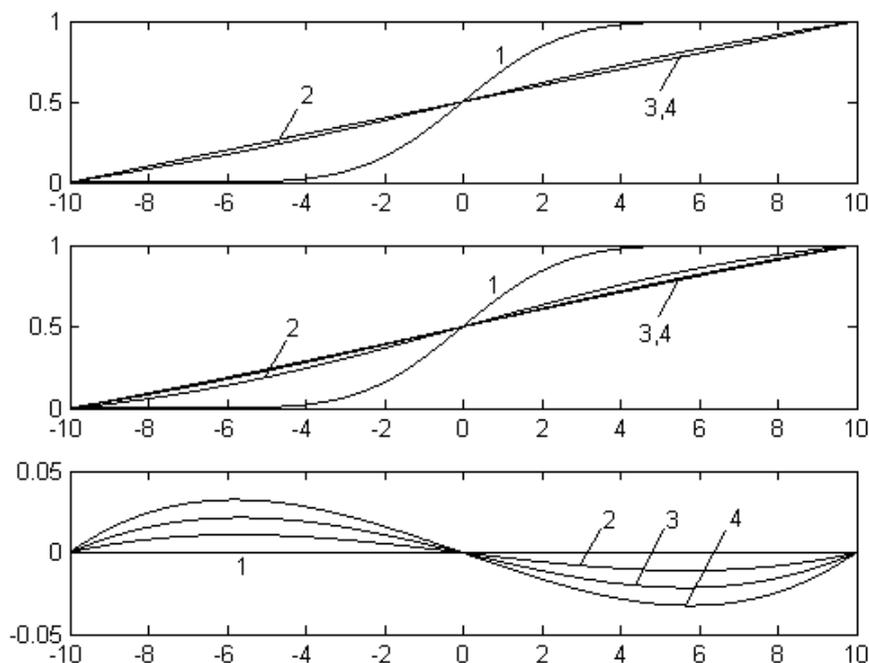


Рис. 4б – Вверху: распределение, описываемое моделью (13); в середине: усеченное нормальное распределение; внизу: разность распределений (параметры функций: $x_{\min} = -10$, $x_{\max} = 10$, $\mu = 0$, $1 - \sigma = 2; 2 - \sigma = 7; 3 - \sigma = 12; 4 - \sigma = 17$)

Из рис. 2–4 видно, что наименьшие отличия между функциями оказываются при расположении точечного источника в точке $x = 0$, соответственно, $\mu = 0$. При этом увеличение значения σ приводит к увеличению данных отличий, максимальные значения которых могут достигать 1,8% и 4,8% для плотности распределения и функции распределения, соответственно. Смещение положения источника от центра области рассеяния приводит к увеличению отличий между анализируемыми зависимостями: максимальные значения разностей между

плотностью распределения и функцией распределения для случая $\sigma = 2, \mu = -9$ составляют 10% и 9.8%, соответственно; для случая $\sigma = 2, \mu = -3$ – 3% и 7.5%, соответственно. Таким образом, в связи с тем, что при анализе случайной последовательности, у которой априори неизвестны положение источника и параметр σ , использование усеченного нормального закона может приводить к возникновению значимых погрешностей при оценке плотностей и функций распределения.

6. Заключение

Проведенное сравнение двух возможных подходов к построению математических моделей плотностей распределений и функций распределений ограниченных случайных последовательностей позволяет сделать вывод о том, что использование модели броуновского движения в ограниченной области для величин, область значений которых ограничена с физической точки зрения, является более обоснованным, чем ограниченное нормальное распределение.

В тоже время размерность пространства параметров модели броуновского движения в ограниченной области оказывается на единицу больше, чем аналогичная величина ограниченного нормального распределения. Данное обстоятельство увеличивает вычислительную сложность нахождения параметров распределения реальных последовательностей и определяет целесообразность исследования возможности использования в рассматриваемой задаче современных алгоритмов поиска экстремумов функций многих переменных, например, генетических алгоритмов [8].

Следует также отметить, что представляет самостоятельный интерес математическое моделирование движения броуновской частицы в ограниченной области, а также анализ возможности использования для

описания функций распределения и плотностей распределения аппроксимации Парзена-Розенблатта [9–10].

Список литературы

1. Матвиевский В.Р. Надежность технических систем [Текст]: учебное пособие/ В.Р. Матвиевский// –М: Московский государственный институт электроники и математики, 2002. 113 с.
2. Бородачев Н.А. Точность производства в машиностроении и приборостроении [Текст]/ Н.А. Бородачев, Р.М. Абрашитов, И.М. Веселова// М.: Машиностроение, 1973. –567 с.
3. Эйнштейн А., Смолуховский М. Броуновское движение [Текст]: сб. статей. – Ленинград: ОНТИ - Главная редакция общетехнической литературы, 1936. –606 с.
4. Поршнев С.В. Теория и алгоритмы аппроксимации эмпирических зависимостей и распределений [Текст]/ С.В. Поршнев, Е.В. Овечкина, В.Е. Каплан// –Екатеринбург: УрО РАН, 2006. –166 с.
5. Эйнштейн А. О движении взвешенных в покоящейся жидкости частиц, требуемом молекулярно-кинетической теории теплоты [Текст]// –Ленинград: ОНТИ - Главная редакция общетехнической литературы, 1936. –С. 13–25.
6. Гнеденко Б.В. Курс теории вероятностей. –М.: Физматгиз, 1988. –407 с.
7. Тобочник Я., Гулд Х. Компьютерное моделирование в физике. Ч. 2. –М.: Мир, 1990. –399 С.
8. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы [Текст]/Л.А. Гладков, В.В. Курейчик, В.М. Курейчик//–М.: Физматлит, 2004. –407 с.
9. Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33, 1065–1076.
10. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 832–837.