УДК 621.396

Исследование методов аппроксимации плотностей распределения

Иванов А.Д., Жиганов С.Н., Ракитин А.В.

В статье проводится анализ точности формирования случайной величины с заданным законом распределения с использованием метода обратной функции на основе анализа гистограммы распределения и фазовых портретов.

Ключевые слова: гистограмма распределения, фазовый портрет, метод обратной функции.

Введение

При разработке и анализе сложных систем обработки радиотехнической информации требуется формировать случайные величины с заданным законом распределения, проводить статистический анализ качества обработки. При этом от характеристик сформированного случайного процесса зависит работоспособность полученных алгоритмов.

В данной работе проведено исследование качества формирования случайной величины с распределением Лапласа на основе метода анализа гистограммы распределения и фазовых портретов.

Статистическое описание случайных величин

Значение случайной величины невозможно предсказать при единичном наблюдении. Свойство случайных процессов принято описывать, рассматривая не просто те величины, которые наблюдаются в какой-нибудь момент времени, а изучая совокупности этих величин, относящихся к различным фиксированным моментам времени. Строго говоря, любая физическая величина случайна, а статистическая трактовка оправдана лишь точностью выполнения расчетов. Из этого следует, что большинство параметров, характеризующих работу любой системы, случайны. Тогда, используя аппарат теории вероятностей, а в более сложных случаях и аппарат теории случайных функций, можно предсказать и количественно описать случайные закономерности [1-3].

Числовые характеристики случайной величины определяются из анализа единичных реализаций, происходящих в период наблюдения за этой величиной. Поэтому методы обработки случайных величин называются статистическими. Для непрерывных случайных величин наряду с законом распределения вероятностей рассматривают плотность вероятностей, которую обозначают $f(x)$.

Любая форма закона распределения случайной величины, например, самая универсальная - функция распределения, полностью определяет случайную величину с вероятностной точки зрения. Существуют более простые, хотя и менее исчерпывающие, характеристики случайной величины, в определенной мере характеризующие её существенные черты, например, среднее значение или степень разбросанности значений относительно среднего. Такие характеристики, назначение которых - выразить числом наиболее существенные особенности распределения, называются числовыми характеристиками случайной величины.

В теории вероятностей основными числовыми характеристиками случайной величины являются её математическое ожидание и дисперсия.

Методы формирования случайных величин на ЭВМ

Необходимость в решении данной задачи возникает при исследовании методом имитационного моделирования алгоритмов для анализа законов распределения случайных величин, процессов, событий. Исходным материалом для формирования на ЭВМ случайных величин с различными законами распределения служат равномерно распределенные в интервале от 0 до 1 случайные числа, которые вырабатываются на ЭВМ программным или же физическим датчиком случайных чисел.

Существуют различные методы для генерирования с определенным законом распределения [4, 5]:

- метод обратной функции (нелинейного преобразования);
- приближенный метод;
- метод исключения и др.

Для решения задачи моделирования с заданным законом распределения, случайный процесс с равномерным законом распределения подвергается нелинейному преобразованию.

Для большинства случаев функцию распределения нельзя найти аналитически и тогда применяют приближенный метод моделирования, который основан на использовании последовательности случайных величин с равномерным законом распределения, кусочно-линейной интерполяции функции распределения и решении задачи обратной интерполяции [4].

Анализ характеристик случайных величин

Провести приближенный анализ качества формируемой последовательности случайных чисел можно при помощи построения гистограммы распределения и фазовых портретов, которые являются уникальными для различных случайных величин. Рассмотрим оба эти подхода.

Гистограммы. Гистограмма представляет собой фигуру, состоящую из прямоугольников, у которой по оси абсцисс откладывают частичные интервалы возможных значений случайной величины, а по оси ординат отношение относительной частоты попадания случайной величины в соответствующий частичный интервал к длине этого интервала. В результате получается фигура, площадь которой равна единице, как и у фигуры, полученной из кривой плотности распределения и осью абсцисс. Поэтому гистограмма является приближением к функции плотности распределения, и чем она ближе к плотности распределения вероятностей (ПРВ), тем точнее сформирована случайная величина. По полученной гистограмме можно подобрать вид теоретической функции распределения для решения задачи аппроксимации ПРВ.

Фазовые портреты. Исследования показали [6], что более наглядным способом проверки качества генерирования является способ, основанный на анализе фазовых портретов законов распределения. Под фазовым портретом будем понимать графическую зависимость, построенную в координатах: $f(x)$ и $f'(x)$.

В работе было проведено моделирование процесса формирования случайных величин на примере случайной величины с законом Лапласа. Был использован метод обратной функции. Исходными для моделирования являлись случайные числа, равномерно распределенные в интервале от 0 до 1. Нелинейное преобразование для получения случайных чисел с распределением Лапласа [4] имеет вид

$$x = \begin{cases} \frac{1}{\lambda} \cdot \ln(2y) + \mu, 0 < y < \frac{1}{2}, \\ -\frac{1}{\lambda} \ln\big(2(1-y)\big) + \mu, \quad \frac{1}{2} < y < 1. \end{cases} \quad (1)$$

Средствами программы MathCAD сформировали $N = 100$ отсчетов случайных величин, равномерно распределенный в интервале от 0 до 1, из них по соотношению (1) получили 100 отсчетов случайных величин с распределением Лапласа. С использованием встроенной в программу MathCAD функции histogram построили гистограмму при количестве частичных интервалов 20. На рис. 1 приведена гистограмма и график плотности распределения Лапласа при $\mu = 5$ и $\lambda = 3$. Из рис. 1 видно, что графики похожи друг на друга. На рис. 2 сплошной кривой приведен фазовый портрет случайной величины, полученной по соотношению (1), а пунктирной кривой представлен полученный фазовый портрет по экспериментальным данным при $N = 100$.

Точность формирования случайных величин можно оценить средней ошибкой полученной гистограммы по соотношению

$$\delta g = \sqrt{\frac{1}{M} \sum_{k=0}^{M-1} (g_k - f(x_k))^2}, \quad (2)$$

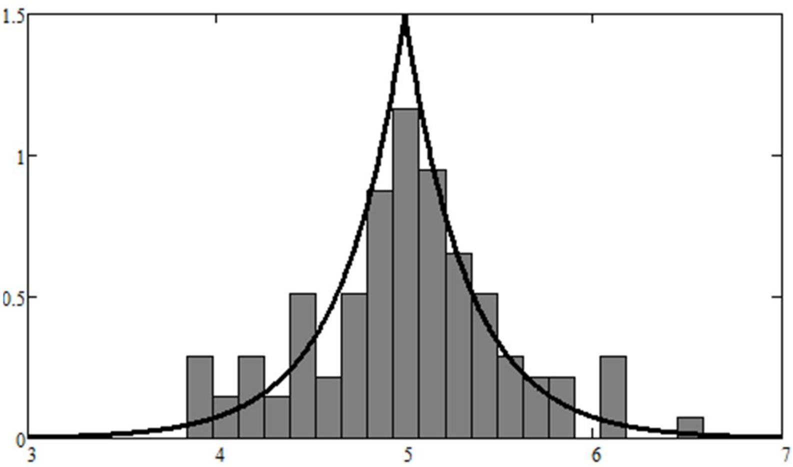

Рис 1 – Гистограмма и график плотности распределения Лапласа при N=100

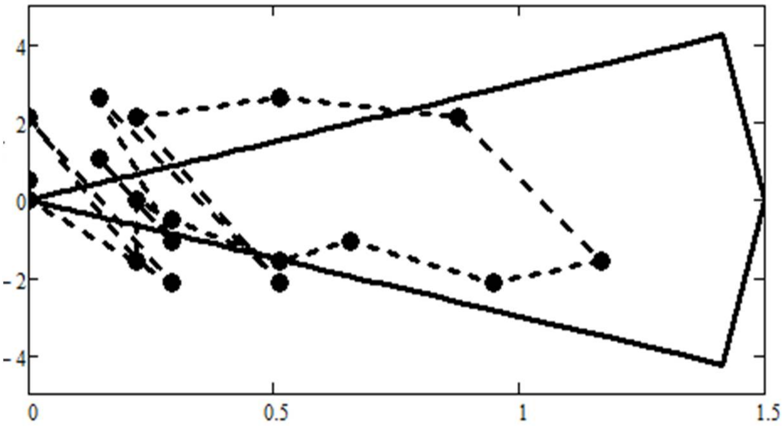

Рис. 2 – Теоретический и экспериментальный фазовые портреты распределения Лапласа при N=100

где M – количество частичных интервалов при построении гистограммы; g_k – значение гистограммы; $f(x_k)$ – значение плотности распределения вероятности в середине частичного интервала, а также ошибкой фазового портрета, вычисляемой по формуле

$$\delta f = \sqrt{\frac{1}{M}\left[\sum_{k=0}^{M-1}\left(\frac{dg_k}{dx} - \frac{df(x_k)}{dx}\right)^2 + \sum_{k=0}^{M-1}(g_k - f(x_k))^2\right]}. \quad (3)$$

В таблице 1 приведены значения ошибок при различных объемах выборок N.

Таблица 1. Ошибки формирования случайных величин с распределением Лапласа

N	100	500	1000	5000	10000	50000	10^5
δg	0,113	0,125	0,137	0,156	0,108	0,154	0,156
δf	1,321	1,31	1,392	0,838	1,211	0,921	0,868

Как видно из таблицы 1 объем выборки существенно не влияет на ошибки формирования случайной величины, меняется только внешний вид гистограммы и фазового портрета. Для примера на рис. 3 и 4 приведены гистограмма и фазовый портрет соответственно при N = 1000. Из сравнения рис. 1 и 3, 2 и 4 видно, что изменение объема выборки существенно не изменяет результирующую гистограмму и фазовый портрет.

Заключение

объем выборки существенно не влияет на ошибки формирования случайной величины, меняется только внешний вид гистограммы и фазового портрета

Литература

1. Теория вероятностей и математическая статистика/ Н.Ш. Кремер и др.- М.: ЮНИТИ-ДАНА, 2000. - 420 с.

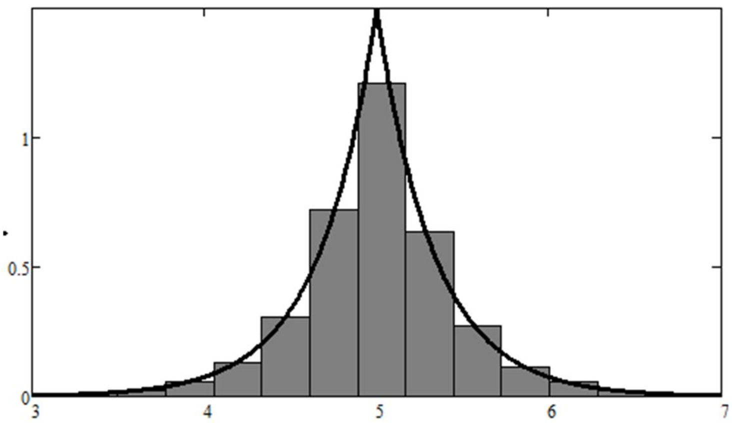

Рис.2 – Гистограмма и график плотности распределения Лапласа при N=10000

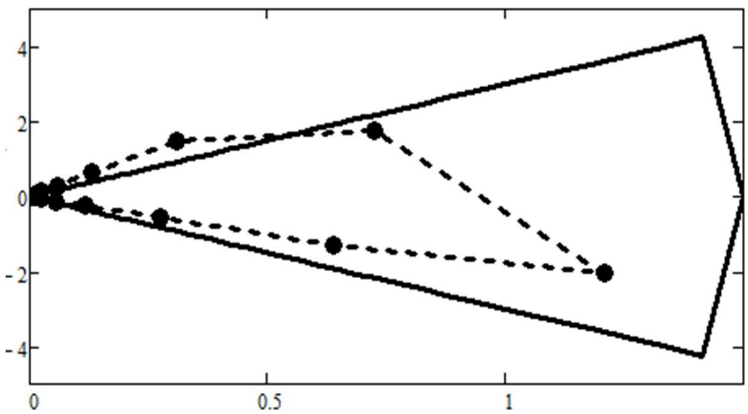

Рис. 4 – Теоретический и экспериментальный фазовые портреты распределения Лапласа при N=10000

2. Гмурман В.Е. Теория вероятностей и математическая статистика. - М.: Высшая школа, 1998.- 368 с.

3. Гмурман В.Е. Руководство по решению задач по теории вероятностей и математической статистике. - М.: Высшая школа, 1998. - 400 с.

4. Быков В.В. Цифровое моделирование в статистической радиотехнике.-М.: Сов. радио, 1971. - 326 с.

5. Полляк Ю.Г. Вероятностное моделирование на электронно-вычислительных машинах. - М.: Сов. радио, 1971. - 400 с.

6. Прохоров С.А. Моделирование и анализ случайных процессов. Лабораторный практикум // Самарский государственный аэрокосмический университет, 2001. 191 с

Поступила 15 октября 2018 г.

The article analyzes the accuracy of the formation of a random variable with a given distribution law using the inverse function method based on the analysis of the histogram of distribution and phase portraits.

Key words: a histogram of the distribution of the phase portrait, of inverse functions.

Иванов Алексей Геннадьевич – магистрант Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

Жиганов Сергей Николаевич – кандидат технических наук, доцент кафедры радиотехники Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

Ракитин Алексей Валерьевич – кандидат технических наук, доцент кафедры радиотехники Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

Адрес: 602264, г. Муром, ул. Орловская, д. 23.