1. МОДЕЛИРОВАНИЕ НЕПРЕРЫВНЫХ И ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Цели работы. Цель лабораторной работы – получить знание о генераторах случайных величин и их практической реализации. Работа также закрепляет практические навыки применения теории вероятности и основ статистического анализа.

Основные сведения

Непредсказуемость – одна из ключевых концепций систем реального мира и имеет место практически в каждой относительно сложной биомедицинской модели. Случайные величины используются для описания влияния различных факторов на систему, флюктуации параметров системы или случайной компоненты входных или выходных сигналов. Знание о применении случайных величин необходимо для моделирования любой относительно сложной биомедицинской системы; это один из основных составных компонентов, которые используется в практике построения моделей.

Примечание. Для выполнения лабораторной работы необходимы базовые знания в области теории вероятности. Необходимо понимание следующих понятий: случайная величина, функция распределения, плотность распределения, математическое ожидание и дисперсия. Эти определения могут быть изучены с помощью следующих пособий: “Basic Probability Theory for Biomedical Engineers” и “Intermediate Probability Theory for Biomedical Engineers”, обе авторства John D. Enderle.

Во многих языках программирования случайные числа или псевдослучайные числа генерируются при помощи специальных функций. Например:

Java: Math.random().

C++: rand().

MATLAB: rand(n) возвращает n-на-n матрицу случайных чисел.

MATLAB: randn(n) возвращает n-на-n матрицу нормально распределенных случайных чисел.

MATLAB: randi(imax) возвращает псевдослучайное целое числе между 1 и imax.

Как правило, эти функции возвращают равномерно распределенную переменную на интервале от 0 до 1. В данной лабораторной работе вы научитесь создавать случайные величины, имеющие различные непрерывные и дискретные законы распределения на основе стандартных генераторов равномерно распределенных чисел.

Генерирование непрерывно распределенных случайных величин

Для генерации различных случайных величин используется метод обратных функций. Положим, мы имеем случайную переменную X, определенную на интервале $[a;b]$, чья плотность распределения $f(X) > 0$. Функция распределения этой переменной в таком случае имеет вид:

$$F(X) = \int_a^X f(u)du\,, x \in [a,b] \qquad (1.1)$$

Тогда, имея реализацию z для переменной Z, которая равномерно распределена на интервале $[0;1]$, мы можем получить выборочное значение случайной величины X решая уравнение:

$$F(x) = z \qquad (1.2)$$

В случае, если уравнение (1.2) может быть аналитически решена, мы можем использовать формулу $x = G(z)$ для получения значений переменной X. В противном случае, решение может быть получено аппроксимацией и другими методами.

Пример 1. Случайная переменная X равномерно распределена на интервале $[a;b]$ так, что ее функция распределения задана следующим образом:

$$F(X) = \begin{cases} 0, x < a \\ \dfrac{x-a}{b-a}, a \le x \le b \\ 1, x > b \end{cases}$$

Чтобы получить алгоритм генерирования этой переменной, нам необходимо решить уравнение:

$$\frac{x-a}{b-a} = z$$

Решение этого уравнения:

$x = a + (b-a)z.$

Пример 2. Случайная величина X задана на интервале $X > a$ функцией плотности распределения:

$$f(X) = \lambda \cdot e^{-\lambda(X-a)}$$

В данном случае уравнение (1.2) имеет вид:

$$1 - e^{-\lambda(X-a)} = z$$

Решая уравнение для X, получаем:

$x = a - \frac{ln(1-z)}{\lambda}$

Так как z - это равномерно распределенная случайная величина, z аналогично $(1-z)$, и мы можем использовать формулу:

$$x = a - \frac{ln(z)}{\lambda}$$

Генерирование дискретных случайных величин

Положим, мы имеем дискретную случайную величину X, которая принимает конечное или счетное число значений $\{x_i\}, i = 1,2,\dots,n$ с вероятностями $p_i = P\{X = x_i\}, \sum_{i=1}^{n} p_i = 1$ (в общем случае, n может быть равно ∞).

Поделим интервал $[0; 1]$ на подынтервалы a_i с длинами p_i. Равномерно распределенная случайная величина Y на интервале $[0; 1]$ может принимать любое из значений с вероятностью, равной длине интервала, так, что с вероятностью p_i случайная величина Y принимает значение x_i. Тогда для получения реализации x случайной переменной X, нам нужно взять значение z случайной величины Z и определить интервал, в который z попадает. Если значение z попадает в интервал a_i, величина x равна x_i.

Алгоритм для генерации дискретной случайной величины представлен на рисунке 1.1.

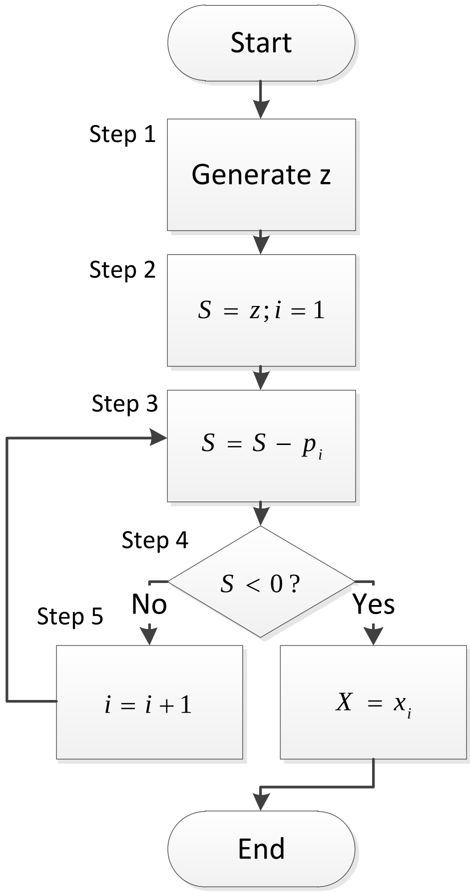

Рисунок 1.1. Алгоритм генерации дискретной случайной переменной

Шаг 1. Генерируется значение случайной переменной z.

Шаг 2. Переменная S берет значение z и инициируется целочисленная переменная i (начинается цикл).

Шаг 3. Переменная S декрементируется по $p_i = P\{X = x_i\}$.

Шаг 4. Если переменная S при итерации $i \leq n$ становится меньше, чем 0, это означает, что переменная z попадает в интервал a_i, а значит принимает значение x_i.

Шаг 5. В противном случае, переменная i инкрементируется на 1 и шаги 3, 4 повторяются.

Если $n = \infty$ обычно применяются формулы для оценки p_i. В некоторых случаях имеет смысл оценки нескольких первых значений $p_{1,}p_2, \dots, p_k$ для ускорения вычислений. Тогда программа только будет осуществлять оценки при помощи описанной формулы для случаев $i > k$. Если значение $\sum_{i=1}^{k} p_i$ близко к 1, эти вычисления редко будут иметь место, так как вероятность событий $i > k$ равна $1 - \sum_{i=1}^{k} p_i$, и потому очень мала.

Оценка среднего и дисперсии

Имея ряд сгенерированных значений случайной величины, мы можем рассчитать точечные оценки среднего и дисперсии случайной величины, а также доверительные интервалы этих оценок. Положим, мы имеем N независимых значений случайной величины X.

Точечная оценка математического ожидания рассчитывается с помощью следующей формулы:

$$M_N^* = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1.3)$$

Точечная оценка дисперсии может быть рассчитана следующим образом:

$$D_N^* = \frac{1}{N-1}\sum_{i=1}^{N} x_i^2 - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} x_i\right)^2 \qquad (1.4)$$

Примечание. В MATLAB имеются специальные функции для вычисления данных оценок. Используйте справку MATLAB, чтобы найти их.

В случае, если мы знаем дисперсию реального значения и число N достаточно большое ($N > 30$), чтобы посчитать закон распределения выборки равным распределению Стьюдента, мы можем использовать следующее выражение для нахождения интервальной оценки среднего:

$$[M_N^* - t_{1-\frac{\alpha}{2}}^{N-1}\frac{S}{\sqrt{N}}; M_N^* + t_{1-\frac{\alpha}{2}}^{N-1}\frac{S}{\sqrt{N}}] \qquad (1.5)$$

Где S – выборочная дисперсия; $t_{\frac{1-\alpha}{2}}^{N-1}$ – коэффициент Стьюдента для $(N - 1)$ степеней свободы и уровня значимости α.

Для нахождения интервальной оценки дисперсии для описанного выше случая, может быть использовано следующее:

$$[\frac{S^2(N-1)}{\chi^2_{\frac{\alpha}{2},N-1}}; \frac{S^2(N-1)}{\chi^2_{1-\frac{\alpha}{2},N-1}}] \qquad (1.6)$$

Где $\chi^2_{\frac{1-\alpha}{2},N-1}$ – коэффициент хи-квадрат для $(N - 1)$ степеней свободы и уровня значимости α.

Для оценки выборочной дисперсии можно использовать следующую формулу:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - M)^2 \qquad (1.7)$$

Примечание. В MATLAB есть специальные функции для нахождения значений распределения Стьюдента и Хи-квадрат для заданных уровней значимости и степеней свободы (вы также использовать табличные значения этих величин или эти формулы). Используйте справку MATLAB, чтобы найти эти функции.

Задание к лабораторной работе

1. Создайте .m файл в MATLAB и постройте алгоритм, который будет рассчитывать выборочные значения случайной переменной, заданной законом распределения соответственно вашему варианту (варианты задания представлены в таблице 1.1 в конце описания работы). Обратите внимания, что в таблице указаны формулы для плотности распределения. Для применения метода обратных функций, необходимо сначала получить функции распределения, что можно сделать путем взятия интеграла от плотности распределения (что можно сделать на бумаге, либо при помощи специальной функции MATLAB).
2. Сгенерируйте три выборки случайной величины следующих размеров: N=50, N=200, N=1000.
3. Рассчитайте точечные оценки среднего, дисперсии и среднеквадратического отклонения (СКО).
4. Рассчитайте интервальные оценки среднего и дисперсия для уровней значимости α: 0.1, 0.05 и 0.01 (всего 18 чисел для среднего и для дисперсии).
5. Сведите все результаты, полученные в пунктах 3 и 4 в таблицы, сохраните их для отчета.
6. Постройте гистограммы, описывающие закон распределения случайной величины по каждой из трех выборок. Для построения гистограммы, необходимо разделить интервал, на котором распределена величина, на заданное количество подынтервалов и рассчитать количество попаданий сгенерированных значений величины в полученные подынтервалы. Высота гистограммы на каждом из подынтервалов определяет количество попаданий в этот подынтервал. Рекомендуется брать подынтервалы одинакового размера. Для расчета подходящего количества интервалов на гистограмме, используется следующая формула:

$$k = [1 + 3.2 \cdot ln(N)] \qquad (1.8)$$

Где k – подходящее число интервалов, N – величина выборки.

7. Постройте поверх каждой построенной гистограммы график теоретической плотности распределения вероятности. Пример гистограммы вместе с теоретической плотностью представлен на рисунке 1.2.

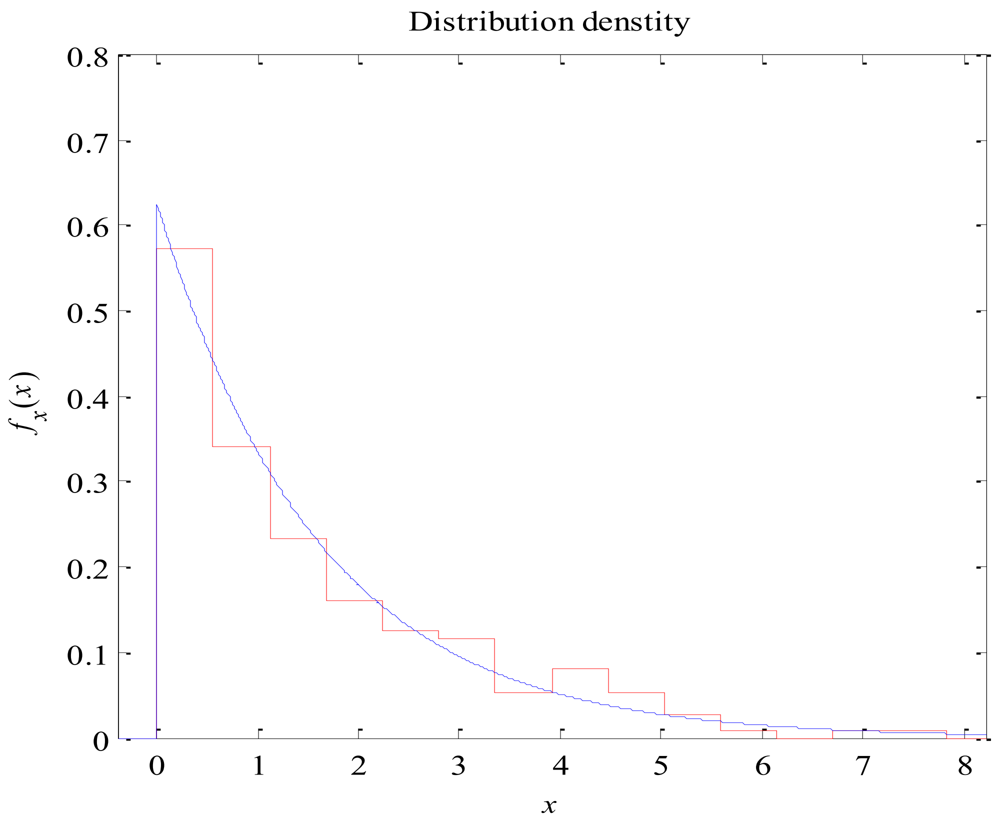

Рисунок 1.1. Гистограмма и теоретически полученная плотность вероятности для непрерывной случайной величины

8. Создайте отдельный график, на котором постройте вместе теоретические функцию распределения вероятности и плотность распределения вероятности.

9. Сделайте выводы по полученным результатам (0,5 страницы).

10. Повторите шаги 1-9 для дискретно распределенной случайной величины, с учетом того, что некоторые шаги потребуют других вычислений (варианты законов распределения представлены в конце описания работы, в таблице 1.2). При построении гистограммы для дискретной переменной поместите теоретически заданный и эмпирически полученный законы распределения на одном графике. На рисунке 1.3. показан пример такого графика.

Примечание. Если дискретная случайная величина распределена на интервале, равном бесконечности, гистограмму следует строить для первых 20-30 значений.

11. Сделайте выводы по лабораторной работе (0,5 страницы) и подготовьте отчет.

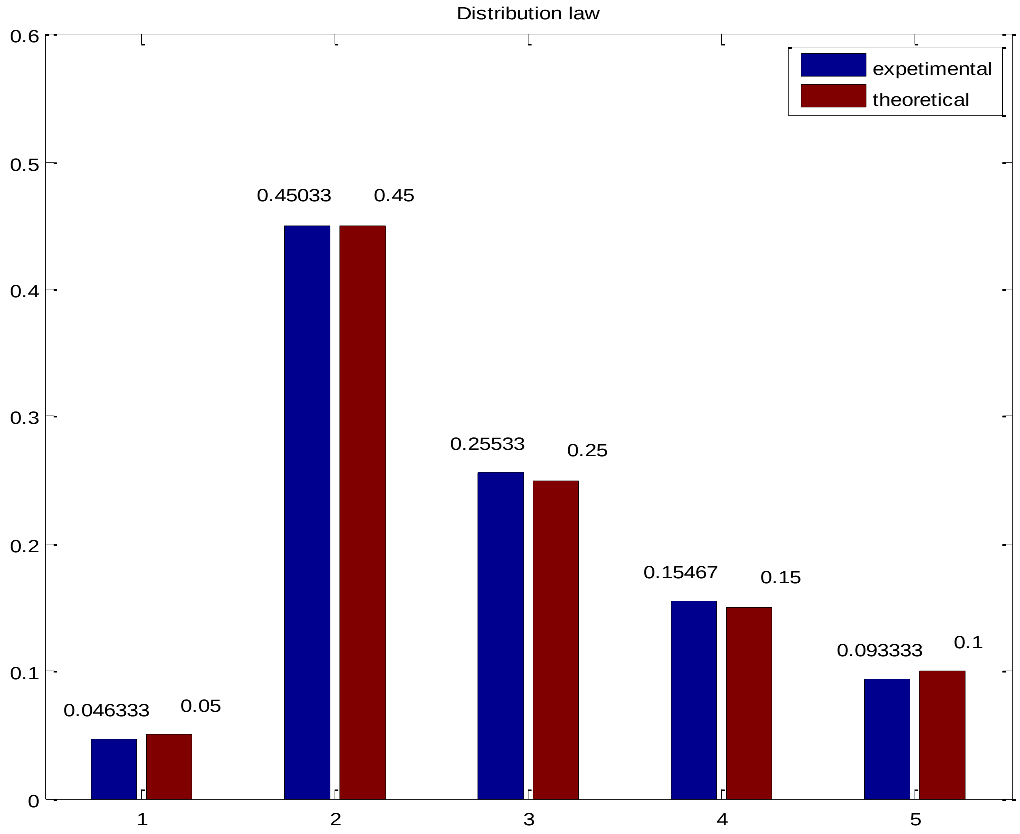

Рисунок 1.3. Теоретический и экспериментальный законы распределения для дискретной случайной величины

Дополнительные задания:

1. Рассчитайте по формулам теоретические математическое ожидание и дисперсию для вашего закона распределения. Сравните эти значения с экспериментально полученными значениями.

2. Сгенерируйте нормально распределенную случайную величину. Рассчитайте коэффициенты асимметрии и эксцесса для выборки, полученной на основе заданного закона распределения. Информацию об этих величинах можно посмотреть по ссылкам:

http://www.mathworks.com/moler/random.pdf

http://www.mathworks.com/help/stats/kurtosis.html

http://www.mathworks.com/help/stats/skewness.html

3. Оцените, имеет ли полученное экспериментальное распределение случайной величины статистически значимое различие с теоретическим для уровней значимости 0.9, 0.95, 0.99. Существует несколько критериев для выполнения этой оценки. Один из наиболее распространенных критериев –

тест Холмогорова-Смирного. О его имплементации в MATLAB можно почитать по следующей ссылке:

http://www.mathworks.com/help/stats/kstest.html

Другой подход – критерий хи-квадрат Пирсона:

http://www.mathworks.com/help/stats/chi2gof.html

Используйте один из этих тестов для оценки выборок различного размера N.

Отчет в конце лабораторной работы должен включать:

1. Цель работы и базовые сведения о генераторах случайных чисел.
2. Таблицы и графики, полученные в ходе выполнения работы.
3. Работающие MATLAB программы для каждого варианта (m-файлы).
4. Выводы по работе, которые должны включать аналитические заключения по полученным результатам.

Таблица 1.1

Непрерывные случайные величины

Вариант	Плотность распределения	Интервал распределния
1	0.25	$[0;4]$
2	$\frac{x}{4.5}$	$[0;3]$
3	$0.5e^{\frac{x}{2}}$	$[-\infty;0]$
4	$\frac{1}{x\cdot ln\,10}$	$[0;10]$
5	$1.5\sqrt{x}$	$[0;1]$
6	$\frac{3x^2}{8}$	$[0;2]$
7	$\frac{1}{2\sqrt{x}}$	[1;4]
8	$\frac{2}{x^3}$	$[1;\infty]$
9	$Sin(x)$	$[0;\frac{\pi}{2}]$
10	$2Cos(2x)$	$[0;\frac{\pi}{4}]$

Таблица 1.2

Дискретные случайные величины

Вариант	Закон распределения
1	$P\{X = k\} = 0.2;\ k = 1,2,3,4,5$
2	$P\{X = k_i\} = p_i;\ p = \{0.1; 0.1; 0.8\};\ k = \{1; 5; 7\}$
3	$P\{X = k_i\} = p_i; p = \{0.05; 0.1; 0.05; 0.2; 0.6\};\ k = \{1; 2; 3; 4; 10\}$
4	$P\{X = k_i\} = p_i; p = \{0.05,0.1,0.2,0.3,0.2,0.1,0.05\}\ k = 4,5,\ldots,10$
5	$P\{X = k\} = 1/15;\ k = 1,2,\ldots,15$
6	$P\{X = k\} = 0.1^k(1 - 0.1);\ k = 0,1,\ldots$
7	$P\{X = k\} = \frac{3^k}{4^{k+1}};\ k = 0,1,\ldots$
8	$P\{X = k\} = 2^{-k-1};\ k = 0,1,\ldots$
9	$P\{X = k\} = C_4^k 0.2^k \cdot 0.8^{4-k};\ k = 0,1,2,3,4$
10	$P\{X = k\} = \frac{0.5^k}{k!} e^{-0.5};\ k = 0,1,\ldots$