

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

КАФЕДРА РАДИОЭЛЕКТРОННЫХ СРЕДСТВ

Лабораторная работа №1

**МЕТОДЫ ФОРМИРОВАНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН
С ЗАДАННЫМ РАСПРЕДЕЛЕНИЕМ**

Методические указания к выполнению лабораторной работы
по дисциплине
«Компьютерное проектирование и моделирование
радиоэлектронных устройств»,

Киров 2015

УДК 621.391

Составитель: доктор технических наук, профессор Д.Е.Прозоров

Рецензент:

Авторская редакция

1. Теоретические сведения

Введение

Чаще всего для выработки случайных чисел с равномерным распределением привлекаются генераторы, которые созданы как соответствующие программы на ЭВМ. С помощью этих программ по некоторому алгоритму получают последовательности чисел. Числа, являющиеся результатами соответствующей вычислительной процедуры, в отличие от случайных чисел, получающихся при подбрасывании монеты или вытаскивании карточек из урны, называются псевдослучайными, или квазислучайными. Генераторы, созданные как соответствующие программы для ЭВМ, называются программными генераторами.

Важным достоинством программных методов является простота практической реализации. Кроме того, возможен контроль в процессе решения задачи (т.е. возможен двойной просчет). Основным недостатком всех программных методов является то, что получаемая при их реализации последовательность оказывается периодической. Поэтому очень длинные последовательности уже не будут в полном смысле случайными.

Полученные с помощью программных методов последовательности в идеале должны состоять из равномерно распределенных, статистически независимых, воспроизводимых и неповторяющихся чисел.

Большинство алгоритмов для получения псевдослучайных чисел имеют вид $x_{i+1} = F(x_i)$, где F – совокупность операций, которые надо проделать над числом x_i , чтобы получить x_{i+1} .

Так, широко известные способы имитации равномерно распределенных случайных чисел основаны на реализации рекуррентного соотношения

$$x_{i+1} = \varphi(x_i, \dots, x_{i-r}) \bmod m \quad (1)$$

и определяют x_{i+1} как наименьший неотрицательный вычет $\varphi(x_i, \dots, x_{i-r})$ по

модулю m (т.е. остаток от деления на m).

Это разностное уравнение описывает динамику инерционной нелинейной дискретной системы, автоколебания в которой стабилизированы механизмом квантования. Причем процесс во многом определяется выбором начальных значений $x_{0-1}, \dots, x_{0-r}$. В зависимости от того, является ли функция φ линейной или нет, различают линейные и нелинейные методы.

Методы формирования случайных величин с равномерным распределением

Метод мультипликативного сравнения или метод вычетов. В данном методе задаются двумя определенным образом подобранными целыми числами: множителем a и модулем m . Последовательность случайных чисел вычисляется по следующему алгоритму:

1. Число x_i , известно с предыдущего шага. Вычисляется произведение ax_i .
2. Число ax_i делится на m . Получается целое число q и целочисленный остаток x_{i+1} , что можно представить в виде $ax_i = qm + x_{i+1}$, $0 \leq x_{i+1} \leq m-1$.
3. Так как x_{i+1} – число между 0 и m , то нужно его еще разделить на m , чтобы получить число между 0 и 1: $y_{i+1} = x_{i+1}/m$.

Таким образом, соотношение (1) записывается в виде

$$x_{i+1} = ax_i \bmod m. \quad (2)$$

При этом говорят, что числа ax_i и x_{i+1} сравнимы по модулю m , т.е. они являются равноостаточными при делении на m . Формулу (2) называют сравнением по модулю, а остаток x_{i+1} – наименьшим положительным вычетом по модулю m . Этим объясняются оба названия алгоритма – метод мультипликативного сравнения или метод вычетов.

Последовательности случайных чисел, полученные методом мультипликативного сравнения, периодически повторяются. Это связано с тем,

что числа x_i могут принимать только значения $0, 1, 2, \dots, m$. Итак, самое большее через $(m-1)$ шагов уже один раз полученное число должно появиться опять, а за ним повторяется и вся последовательность. Таким образом, длина периода при модуле m не может превышать $(m-1)$. Это означает, что на практике можно обеспечить удовлетворительную величину периода, если выбирать модуль m достаточно большим.

Статистические свойства последовательности зависят от выбора x_0 и a . Выбор m не зависит от системы счисления, а только от длины машинного слова. Так, для 32-разрядного машинного слова можно выбрать $m=2^{31}-1$ и множитель $a=16807$. Число $(2^{31}-1)$ является простым. В этом случае достигается максимально возможная длина периода $(m-1)$, т.е. $(2^{31}-1)$. Алгоритмы наиболее часто применяемых в настоящее время генераторов равномерно распределенных случайных величин в общем виде можно описать следующим образом:

$$x_{i+1} = \left(\sum_{j=0}^r a_j x_{i-j} + \mu \right) \bmod m \quad (3)$$

где коэффициенты $a_0, a_1, \dots, a_r, \mu$ и m , а также получаемые числа $x_1, x_2, \dots$ – целые числа.

В этом случае можно получить более длинные периоды в последовательности по сравнению с методом вычетов. Числа в получающейся последовательности имеют слабую корреляцию.

Мультипликативный генератор. Рассмотренный выше пример получения случайных последовательностей методом мультипликативного сравнения является частным случаем линейного генератора. Подставляя в (3) $\mu = a_1 = \dots = a_r = 0$ и положив, что $a_0 > 0$, получим:

$$x_{i+1} = a_0 x_i \bmod m \quad (4)$$

На практике генератор вида (4) обладает достаточно хорошими статистическими характеристиками.

Смешанный генератор получится, если в (3) положить $a_0 > 0$, $\mu > 0$, а $a_1 = \dots = a_r = 0$:

$$x_{i+1} = (a_0 x_i + \mu) \bmod m \quad (5)$$

Аддитивный или генератор Фибоначчи. Если в (3) задать $\mu = a_2 = \dots = a_r = 0$, $a_0 = a_1 = 1$, то генератор такого вида описывается выражением:

$$x_{i+1} = (x_i + x_{i-1}) \bmod m \quad (6)$$

Обобщенным аддитивным генератором или обобщенным генератором Фибоначчи называется генератор вида (3), где все $a_i = 1$.

Пример. Рассмотрим смешанный генератор:

$$(1255 \cdot 135 + 10555) \bmod 1000 = 179980 \bmod 1000 = 980$$

$$(1255 \cdot 980 + 10555) \bmod 1000 = 1240455 \bmod 1000 = 455$$

Таким образом, по целым числам последовательности $\{x_i\}$ можно построить последовательность $\{x_i/m\}$ рациональных чисел из единичного интервала. Например, 0,98; 0,455 и т.д.

Моделирование значений случайной величины U с равномерным распределением на отрезке $[0; 1)$ доступно в большинстве современных систем программирования. Например, в языке Matlab эту роль выполняет функция `rand()`. Строго говоря, невозможно с помощью рекуррентных соотношений получить последовательность случайных чисел, которые являются независимыми реализациями случайной величины X , равномерно распределенной на интервале $[0; 1)$, так как число используемых двоичных или десятичных разрядов в ЭВМ ограничено. На самом деле при этом получают так называемые квазиравномерно распределенные случайные числа. При достаточно большом числе разрядов квази-равномерно распределенная случайная величина хорошо аппроксимирует равномерную случайную

величину.

Для проверки, являются ли сгенерированные числа в действительности последовательностью независимых случайных величин, применяются критерии согласия между эмпирическим и теоретически распределениями, критерии стохастической независимости следующих друг за другом генерируемых чисел и критерии случайности. Эти группы критериев в основном охватывают проверку тех свойств, которыми должны обладать случайные числа.

С помощью критериев согласия проверяется гипотеза о том, что выборка $x_0, x_1, \dots, x_n$ произведена из генеральной совокупности с определенным распределением вероятностей. Очень часто для этой цели применяются критерий согласия Пирсона χ^2 (критерий частот) и критерии, учитывающие различные свойства данной генеральной совокупности. Более подробно критерии согласия будут рассмотрены ниже.

Если проверяется равномерность распределения на интервале $[0; 1)$, то случайные числа должны иметь следующие свойства:

- 1) среднее значение чисел не должно существенно отличаться от $1/2$;
- 2) среднее значение квадратов чисел не должно существенно отличаться от $1/3$;
- 3) дисперсия чисел не должна существенно отличаться от $1/12$;
- 4) коэффициенты асимметрии и эксцесса не должны существенно отличаться соответственно от 0 и -1.2

Методы формирования случайных величин с заданным законом распределения

Можно выделить несколько основных методов формирования последовательности случайных величин с произвольным законом распределения из последовательности равномерно распределенных случайных чисел на интервале $[0; 1)$.

1. Метод обратных функций или метод прямого преобразования равномерно распределенных случайных чисел.

2. Метод отсеивания чисел из первоначальной последовательности случайных чисел (метод генерации Неймана).

3. Методы, основанные на центральной предельной теореме.

Метод обратных функций. Пусть случайная величина Y непрерывна и имеет заданную плотность распределения вероятностей $w(y)$. Тогда для функции распределения можно записать

$$F(y) = \int_{-\infty}^y w(y) dy \quad (7)$$

Идея данного метода основывается на следующей теореме.

Теорема. Если случайная величина Y имеет ПРВ $w(y)$, то распределение случайной величины

$$X = \int_{-\infty}^Y w(y) dy = F(Y) \quad (8)$$

является равномерным на интервале $[0; 1)$. На основе этой теоремы получаем следующее правило.

Правило. Для того, чтобы найти возможное значение y , непрерывной случайной величины Y зная ее ПРВ $w(y)$, надо выбрать случайное число x , и решить относительно y , уравнение

$$x_i = \int_{-\infty}^{y_i} w(y) dy \quad (9)$$

Справедливо и обратное утверждение.

Если $F(y)$ – функция распределения некоторой непрерывной случайной величины Y , а X — случайная величина с равномерным распределением на интервале $[0; 1)$, то случайная величина $Y = F^{-1}(X)$ имеет функцию распределения $F(Y)$, где F^{-1} – функция, обратная по отношению к F .

Если закон распределения задан функцией распределения, то правило получения значений случайной величины X заключается в следующем:

1) реализовать случайную величину X равномерно распределенную на интервале $[0; 1)$;

2) вычислить значение случайной величины Y по формуле $Y = F^{-1}(X)$.

Пример. Случайная величина Y имеет экспоненциальный (показательный) закон распределения

$$w(y) = \lambda e^{-\lambda y}, \quad F(y) = \int_0^y w(y) dy = \int_0^y \lambda e^{-\lambda y} dy = 1 - e^{-\lambda y},$$

где $\lambda > 0$ — параметр распределения, $M(Y) = 1/\lambda$, $y \geq 0$.

Требуется найти формулу для моделирования случайной величины Y с помощью равномерно распределенной случайной величины X .

Находим обратную по отношению к F функцию. Имеем: $X = 1 - e^{-\lambda Y}$. Решая это уравнение относительно Y , получаем $Y = -(1/\lambda) \ln(1 - X)$.

Учитывая, что случайная величина $(1 - X)$ имеет также равномерный закон распределения на интервале $[0; 1)$, окончательно получаем $Y = -(1/\lambda) \ln(X)$.

В результате случайные числа с экспоненциальным распределением вычисляются по формуле

$$y_i = -(1/\lambda) \ln(x_i). \quad (10)$$

Пример. ФР случайной величины Y на интервале $[a; b)$ равна $F(y) = (y - a)/(b - a)$. Составим уравнение согласно теореме: $X = (Y - a)/(b - a)$. Откуда $Y = a + (b - a)X$.

Обращение функции распределения во многих случаях представляет собой довольно сложную задачу. Поэтому метод обратных функций не всегда применим. Однако для обратной функции F^{-1} иногда можно найти достаточно точную аппроксимацию.

Метод отсеивания (метод Неймана). Данный метод заключается в том,

что из равномерно распределенной последовательности случайных чисел отбираются числа таким образом, чтобы они подчинялись заданному закону распределения. Способ отбора применим для получения реализаций только таких случайных величин, закон распределения которых может быть задан с помощью функции плотности распределения вероятностей (ПРВ) (рис. 1).

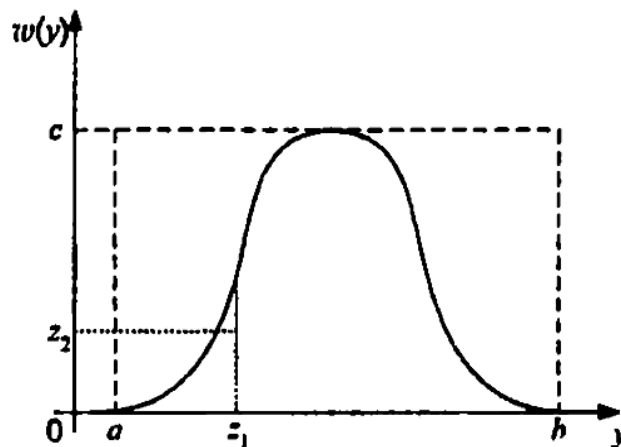


Рис.1.

Рассмотрим моделирование этим методом одномерной случайной величины Y определенной на интервале $[a; b)$, с ПРВ $w(y)$. Вне этого интервала $w(y)=0$, и, кроме того, ПРВ ограничена сверху, т.е. $w(y) \leq c$, где c – постоянная.

Алгоритм получения значений случайной величины Y иллюстрируется на рис. 1 и записывается следующим образом.

1. Получаем два независимых значения x_1 , и x_2 случайной величины X равномерно распределенной на интервале $[0; 1)$.
2. Строим точку с координатами (z_1, z_2) , где $z_1 = a + x_1(b - a)$, $z_2 = x_2 c$.
3. Если $z_2 \leq w(z_1)$, то полагаем, что случайная величина Y приняла значение z_1 ; если $z_2 > w(z_1)$, то точка $z = (x_1; x_2)$ отбрасывается и вычисления повторяем после получения новой пары случайных чисел $\{x_n\}$, возвращаясь к шагу 1.

Полученные с помощью приведенного алгоритма значения являются реализациями случайной величины Y с плотностью $w(y)$. Нужно отметить, что данный метод требует довольно большого числа машинных операций.

Метод, основанный на центральной предельной теореме. Наиболее часто встречающимся видом распределения является нормальное. В связи с этим при моделировании различных явлений возникает потребность иметь в распоряжении последовательности случайных чисел, отвечающих нормальному закону распределения.

Один из известных программных методов реализации нормально распределенной случайной величины основан на центральной предельной теореме.

Теорема (центральная предельная теорема). Распределение суммы независимых случайных величин x_i ($i=1,2,\dots,n$) приближается к нормальному при неограниченном увеличении n , если выполняются следующие условия:

- 1) все величины имеют конечные математические ожидания и дисперсии;
- 2) ни одна из величин по значению резко не отличается от всех остальных.

Согласно этой теореме, сумма независимых случайных величин $x_1, x_2, \dots, x_n$ равномерно распределенных на интервале $[0; 1)$, асимптотически нормальна с математическим ожиданием $M(Y)=n M(x_i)$ и дисперсией $D(Y)=n D(x_i)$, где $M(x_i)=1/2$; $D(x_i)=1/12$; $i=1,\dots,n$.

Как показывает практика, сумма X_i ($i=1,\dots,n$) имеет распределение, близкое к нормальному при сравнительно небольших n (около 8 — 12 слагаемых). Обычно конструируют алгоритм реализации случайной величины Y на основе аппроксимации распределения с плотностью $N(0;1)$.

Произведем нормировку рассматриваемой суммы, для чего вычтем математическое ожидание и разделим результат на среднее квадратическое отклонение:

$$\frac{\left(\sum_{i=1}^n X_i\right) - n/2}{\sqrt{n/12}} . \quad (11)$$

В силу центральной предельной теоремы при $n \rightarrow \infty$ распределение этой нормированной случайной величины стремится к нормальному с параметрами $a=0$ и $\sigma=1$. При конечном n распределение приближенно нормальное. В частности, при $n=12$ получаем достаточно хорошее и удобное для расчета приближение:

$$\sum_{i=1}^{12} X_i - 6$$

При помощи линейного преобразования $y_i = \mu + \sigma x_i$, $i=1,2,\dots,n$, при любом μ и $\sigma > 0$ можно получить последовательность случайных чисел $\{y_n\}$, отвечающих распределению $N(\mu; \sigma^2)$. Помимо увеличения n , для улучшения асимптотической нормальности случайных чисел можно воспользоваться специальными преобразованиями.

Алгоритмы моделирования часто употребляемых случайных величин

Рассмотрим алгоритмы формирования случайных величин с наиболее распространенными распределениями.

Гауссовская случайная величина. Плотность распределения гауссовской (нормальной) случайной величины (НСВ) записывается выражением

$$w(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-a)^2}{2\sigma^2}} \quad (12)$$

где σ – стандартное отклонение; a — математическое ожидание. Для удобства записи ПРВ вместо (12) часто используют сокращенную форму записи в виде $N(a; \sigma^2)$.

Функция распределения (ФР) СВ, распределенной по нормальному

закону, не выражается в элементарных функциях. Поскольку данное распределение играет огромное значение в дисциплинах так или иначе связанных с моделированием случайных величин, можно привести интегральную форму записи функции распределения нормального закона:

$$F\left(\frac{y-a}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\left(\frac{y-a}{\sigma}\right)} e^{-\frac{u^2}{2}} du \quad (13)$$

где u – действительная переменная; σ – стандартное отклонение; a — математическое ожидание. Для вычисления значений функции распределения гауссовской СВ можно воспользоваться специально составленными таблицами.

Алгоритм формирования НСВ имеет вид

$$Y = a + \sigma \sqrt{-2 \ln x_1} \sin(2\pi x_2), \quad (14)$$

где x_i и x_2 – независимые равномерно распределенная случайная величина (РРСВ) на интервале $[0; 1)$.

Гауссовская случайная величина с экспоненциальной корреляционной функцией. Алгоритм формирования коррелированной гауссовской случайно величины имеет вид

$$y_k = a_0 x_k + b_1 y_{k-1}, \quad (15)$$

где $a_0 = \sigma \sqrt{1 - \rho^2}$, $b_1 = \rho$, ρ – нормированный коэффициент корреляции СВ, σ – СКО, x_k – НСВ с параметрами $(0,1)$.

Логарифмически-нормально распределенная случайная величина. Это распределение случайной величины, логарифм которой распределен нормально. Одномерная плотность логарифмически нормальной СВ (ЛНСВ) имеет вид

$$w(y) = \frac{1}{\sqrt{2\pi} \sigma y} e^{-\frac{(\ln y - a)^2}{2\sigma^2}}, \quad (16)$$

где a — математическое ожидание; σ – параметр распределения.

ФР логарифмически-нормального распределения невозможно выразить в

элементарных функциях.

Алгоритм формирования ЛНСВ записывается в виде $Y=e^X$, где X – НСВ с параметрами $(a; \sigma^2)$.

Рэлеевская случайная величина. Плотность распределения такой случайной величины записывается выражением

$$w(y) = \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}}, \quad (17)$$

в котором σ – параметр распределения.

ФР рэлеевской СВ запишется как

$$F(y) = 1 - e^{-\frac{y^2}{2\sigma^2}}. \quad (18)$$

Алгоритм формирования такой СВ имеет вид

$$Y = \sigma \sqrt{-2 \ln X}, \quad (19)$$

где X – РПСВ на интервале $[0; 1)$.

Райсовская случайная величина. В некоторых литературных источниках распределение для таких СВ называется обобщенным законом Рэлея, законом Рэлея — Райса, распределением Райса. Плотность распределения такой СВ имеет вид

$$w(y) = \frac{y}{\sigma^2} e^{\frac{-y^2+a^2}{2\sigma^2}} I_0\left(\frac{ay}{\sigma^2}\right), \quad (20)$$

где σ – параметр распределения, характеризующий переменную компоненту райсовской СВ; a — математическое ожидание; I_0 — модифицированная функция Бесселя 1-го рода нулевого порядка.

ФР Райса невозможно выразить в элементарных функциях.

Алгоритм формирования имеет вид

$$Y = \sqrt{(X_1+a)^2 + X_2^2}, \quad (21)$$

где X_1 и X_2 – НСВ с параметрами $(0; \sigma^2)$.

Обработка результатов моделирования

При реализации моделирующих алгоритмов на ЭВМ вырабатывается информация о состояниях исследуемых систем. Эта информация является исходной для определения приближенных значений или оценок искомых величин.

Для сложных систем и большого количества реализаций, воспроизводимых при моделировании, объем информации о состояниях системы может быть настолько значительным, что ее запоминание, обработка и последующий анализ оказываются слишком трудоемким. Поэтому необходимо так организовать фиксацию и обработку результатов моделирования, чтобы оценки для искомых величин формировались постепенно по ходу моделирования, без специального запоминания всей информации о состояниях системы.

Оценки случайных величин (СВ). Если при моделировании данной системы учитываются случайные факторы, то в качестве оценок для искомых величин используются средние значения, дисперсии и другие вероятностные характеристики случайных величин, полученных по результатам многократного моделирования.

Рассмотрим, например, оценку вероятностей возможных значений случайной величины (закона ее распределения). Разобьем область возможных значений случайной величины на n интервалов. В специальных переменных (или массиве) будем накапливать количества m_k , $k=1,2,\dots,n$ попаданий СВ в эти интервалы. Тогда оценкой вероятности попадания случайной величины в интервал с номером k служит величина $p_k = m_k / N$.

Для оценки среднего значения случайной величины X будем накапливать сумму возможных значений случайной величины, которые она принимает при различных реализациях процесса. Тогда

$$M(X) = \frac{1}{N} \sum_{k=1}^N x_k. \quad (22)$$

Оценкой дисперсии случайной величины может служить выражение

$$D(X) = \frac{1}{N-1} \sum_{k=1}^N (x_k - M(X))^2 . \quad (23)$$

В случае оценки корреляционного момента для случайных величин X и Y используется выражение

$$R(X, Y) = \frac{1}{N-1} \sum_{k=1}^N (x_k - M(X))(y_k - M(Y)) , \quad (24)$$

или в более удобном виде, требующем запоминания небольшого количества величин:

$$R(X, Y) = \frac{1}{N-1} \sum_{k=1}^N x_k y_k - \frac{1}{N(N-1)} \sum_{k=1}^N x_k \sum_{k=1}^N y_k . \quad (25)$$

Непосредственное вычисление по приведенным формулам неудобно, так как среднее значение изменяется в процессе накопления значений x_k , а это, в свою очередь, приводит к усложнению расчетов либо требует запоминания всех N значений x_k . В случае анализа стационарного процесса X вычисление среднего значения $M(x)$ и дисперсии $D(X)$ производится усреднением не по времени, а по множеству N значений, измеренных по одной реализации процесса достаточной продолжительности.

В целях экономии памяти ПЭВМ, на которой производится моделирование, $M(x)$ и $D(X)$ вычисляются в ходе моделирования путем наращивания итогов при появлении очередного измерения случайной характеристики по следующим рекуррентным формулам.

Математическое ожидание случайной величины X для ее n -й выборки x_n :

$$m_n = m_{n-1} \frac{n-1}{n} + \frac{x_n}{n} . \quad (26)$$

Выборочная дисперсия соответственно запишется как

$$\sigma_n^2 = \sigma_{n-1}^2 \frac{n-2}{n-1} + \frac{1}{n} (x_n - m_{n-1})^2 . \quad (27)$$

Начальная дисперсия $\sigma_0^2 = 0$.

Точность и число реализаций. Для обеспечения статистической устойчивости результатов моделирования соответствующие оценки вычисляются как средние значения по большому числу реализаций, выбор которого производится с помощью доверительных интервалов.

Пусть некоторый параметр, имеющий математическое ожидание a и дисперсию σ^2 , оценивается по результатам моделирования СВ x_i . В качестве оценки параметра выбирается среднеарифметическое $M(X)$ (22). В силу случайных причин $M(X)$ будет в общем случае отличаться от a .

Величину ϵ , такую, что $|a - M(X)| < \epsilon$ называют точностью оценки $M(X)$, а вероятность того, что это неравенство выполняется, ее достоверностью $P(|a - M(X)| < \epsilon) = \alpha$.

Воспользуемся данным правилом для определения точности результатов, полученных методом статистического моделирования.

Пусть целью моделирования будет вычисление вероятности p появления некоторого случайного события A . В каждой из N реализаций процесса количество событий A является случайно величиной X , принимающей значение $x_1=1$ с вероятностью p , и значение $x_2=0$ с вероятностью $(1-p)$.

Математическое ожидание и дисперсия x равны:

$$M(X) = x_1 p + x_2 (1-p) = p ,$$

$$D(X) = [x_1 - M(X)]^2 p + [x_2 - M(X)]^2 (1-p) = p(1-p) .$$

В качестве оценки для искомой вероятности p принимается частота m/N наступлений события A в N реализациях:

$$\frac{m}{N} = \frac{1}{N} \sum_{i=1}^N x_i ,$$

где x_i – число наступлений события A в реализации с номером i .

В силу центральной предельной теоремы теории вероятностей частота m/N при достаточно больших N имеет распределение, близкое к нормальному, с математическим ожиданием и дисперсией:

$$M(m/N) = p,$$

$$D(m/N) = p(1-p)/N.$$

Для нормированного отклонения частоты от своего математического ожидания можно использовать выражение

$$\Phi(t) = P\left(\left|\frac{\epsilon}{\sqrt{D(m/N)}}\right| < t_\alpha\right). \quad (28)$$

Поэтому для каждого значения достоверности $\Phi(t) = \alpha$ можно выбрать из таблиц интеграла вероятностей такую величину t_α , что точность ϵ будет равна

$$\epsilon = t_\alpha \sqrt{D(m/N)}, \text{ или } \epsilon = t_\alpha \sqrt{D(X)/N}. \quad (29)$$

Отсюда число реализаций $N = t_\alpha^2 D(X) / \epsilon^2$.

Пример. Для $\alpha = 0.95 \Rightarrow t_\alpha = 1.96$. Подставляя вместо дисперсии $D(X)$ ее значение, можно определить количество реализаций N необходимых для получения оценки частоты m/N с точностью ϵ и достоверностью α :

$$N = t_\alpha^2 = \frac{p(1-p)}{\epsilon^2}. \quad (30)$$

На практике вероятность p обычно неизвестна. Поэтому для определения количества реализаций поступают следующим образом. Выбирают $N_0 = 50 \dots 100$, по результатам реализаций определяют m , а затем окончательно назначают N , принимая, что $p \approx m/N_0$. Точность ϵ обычно выбирают в пределах $0.01 \dots 0.05$.

Если оцениваем одновременно дисперсию нескольких случайных величин, то для расчета N выбираем $\max D_i(X)$. Если $D(X)$ неизвестно, то

используют статистическую оценку выборочной дисперсии. Аналогично можно определить количество реализаций, необходимых для оценки дисперсий, корреляционных моментов и других характеристик случайных величин.

Первичная статистическая обработка данных. При обработке ряда наблюдений $x_1, x_2, \dots, x_n$ исследуемой случайной величины очень важно понять механизм формирования выборочных значений x_i , т.е. подобрать и обосновать некоторую модельную функцию распределения $F_m(x)$, с помощью которой можно адекватно описать исследуемую функцию распределения $F(x)$. На определенной стадии исследования это приводит к необходимости проверки гипотезы типа

$$H: F(x) = F_m(x) . \quad (31)$$

В случае, когда число возможных значений случайной величины не велико, представление о ее распределении дает набор частот появления каждого из значений (гистограмма).

При построении гистограммы горизонтальная ось разбивается на равные интервалы, и на каждом из интервалов, как на основании, строится прямоугольник с высотой, пропорциональной частоте попадания СВ в данный интервал.

Существенным для построения гистограммы является выбор оптимального разбиения, поскольку при увеличении интервалов снижается детализация оценки плотности распределения, а при уменьшении падает точность ее значения. Для выбора оптимального количества интервалов n часто применяется правило Стёрджеса

$$n = [1 + \log_2 N] ,$$

где N — общее число наблюдений величины, $\log_2$ — логарифм по основанию 2, $[*]$ — обозначает целую часть числа x .

Также часто встречается правило, оценивающее оптимальное количество интервалов как квадратный корень из общего числа измерений: $n = [\sqrt{N}]$.

2. Лабораторная работа №1 «Методы формирования случайных величин с заданным распределением»

Цель работы: обучение методам моделирования и оценке точности случайных величин (СВ) с заданным законом распределения.

Методические указания к выполнению работы

1. В соответствии с заданием разработать программный код алгоритма моделирования случайной величины с равномерным законом распределения на интервале $[0,1)$.

2. На основании генератора, полученного в п.1 реализовать алгоритм моделирования случайной величины с заданным законом распределения.

3. Разработать программный код для эмпирической оценки математического ожидания, дисперсии и плотности распределения (гистограммы) случайной величины, полученной в п.2.

4. Вычислить точность оценки математического ожидания и дисперсии полученных в результате моделирования (п.1 и п.2) случайных величин при количестве выборок $N=10^3, 10^4, 10^5$.

Вариант	п.1	п.2
1	Мультипликативный генератор, $a=16807$, $m=2^{31}-1$	СВ с экспоненциальным законом распределения, $\lambda=0.5$
2	Смешанный генератор, $a=1103515245$, $m=32768$, $\mu=12345$	Гауссовская СВ (генератор на основе центральной предельной теоремы), $a=1$, $\sigma=1$
3	Мультипликативный генератор, $a=1664525$, $m=2^{32}$	Гауссовская СВ (генератор на основе тригонометрических преобразований), $a=-1$, $\sigma=1$
4	Смешанный генератор, $a=22695477$, $m=2^{32}$, $\mu=1$	Релеевская СВ, $\sigma=2$
5	Мультипликативный генератор,	Райсовская СВ, $a=1$, $\sigma=1$

	$a=69069$, $m=2^{32}$	
6	Смешанный генератор, $a=134775813$, $m=2^{32}$, $\mu=1$	СВ с экспоненциальным законом распределения, $\lambda=1$
7	Мультипликативный генератор, $a=1103515245$, $m=2^{64}$	Гауссовская СВ (генератор на основе центральной предельной теоремы), $a=-1$, $\sigma=1$
8	Смешанный генератор, $a=214013$, $m=2^{32}$, $\mu=2531011$	Гауссовская СВ (генератор на основе тригонометрических преобразований), $a=0$, $\sigma=\sqrt{2}$
9	Мультипликативный генератор, $a=16807$, $m=2^{31}-1$	Релеевская СВ, $\sigma=\sqrt{2}$
10	Смешанный генератор, $a=1103515245$, $m=32768$, $\mu=12345$	Райсовская СВ, $a=2$, $\sigma=\sqrt{2}$
11	Мультипликативный генератор, $a=1664525$, $m=2^{32}$	СВ с экспоненциальным законом распределения, $\lambda=2$
12	Смешанный генератор, $a=22695477$, $m=2^{32}$, $\mu=1$	Гауссовская СВ (генератор на основе центральной предельной теоремы), $a=-1$, $\sigma=2$
13	Мультипликативный генератор, $a=69069$, $m=2^{32}$	Гауссовская СВ (генератор на основе тригонометрических преобразований), $a=2$, $\sigma=\sqrt{2}$
14	Смешанный генератор, $a=134775813$, $m=2^{32}$, $\mu=1$	Релеевская СВ, $\sigma=\sqrt{3}$
15	Мультипликативный генератор, $a=1103515245$, $m=2^{64}$	Райсовская СВ, $a=\sqrt{2}$, $\sigma=2$

Оформление отчета

Отчет должен содержать:

1. Титульный лист.
2. Цель работы и задание (в соответствии с выданным вариантом).
3. График плотности СВ с равномерным законом распределения (см. п.1

задания), эмпирические оценки математического ожидания и дисперсии РРСВ, результаты вычисления точности оценок математического ожидания и дисперсии РРСВ.

4. График плотности СВ с заданным законом распределения (см. п.2 задания), эмпирические оценки математического ожидания и дисперсии СВ.

5. Программный код.

3. Контрольные вопросы

1. Метод мультипликативного сравнения (метод вычетов).
2. Смешанный генератор равномерно распределенных СВ.
3. Обобщенный генератор Фибоначчи.
4. Как оценить качество генератора равномерно распределенной СВ?
5. Метод обратных функций.
6. Метод отсеивания (метод Неймана).
7. Метод генерации гауссовских СВ на основе центральной предельной теоремы.
8. Статистические оценки статистических параметров СВ (среднее, дисперсия, корреляционная функция).
9. Что называют точностью оценки математического ожидания СВ?
10. Как зависит точность оценки математического ожидания СВ от количества опытов?
11. Как оценить плотность распределения СВ?

Пример. Программный код генератора Фибоначчи:

```
%-----  
%файл randfib.m  
%-----  
function y = randfib(N)  
    M = 2^32;  
    persistent seed;  
    if isempty(seed)  
        seed = 2^15;  
    end  
    x = zeros(1,N);  
    y = zeros(1,N);  
    x(1) = seed;  
    for i=3:N  
        x(i) = x(i-1) + x(i-2);  
        x(i) = mod(x(i),M);  
        y(i) = x(i)/M;  
        if ~isfinite(y(i)) then  
            y(i) = y(i-1);  
        end  
    end  
    seed = x(N);  
end
```

Пример. Программный код для оценки математического ожидания, дисперсии и гистограммы СВ:

```
%-----  
%файл lab1.m  
%-----  
  
%очистка памяти  
clearvars;  
  
%количество выборок  
N = 10^5;  
%генерация РРСВ  
x = randfib(N);  
  
%вычисление среднего значения  
disp('Среднее :'); disp(mean(x));  
%вычисление дисперсии  
disp('Дисперсия:'); disp(var(x));  
  
%отображение гистограммы  
M = 30; %количество столбцов  
[bar_y,bar_x] = hist(x,M);  
bar_y = bar_y / (N*(bar_x(2)-bar_x(1)));  
bar(bar_x,bar_y);
```