Лабораторная работа № 1

Анализ и генерация случайных чисел. Основы имитационного моделирования

Цель работы: Изучение основных характеристик случайных величин на базе теории вероятностей и математической статистики; изучение и программирование способов получения псевдослучайных чисел.

Теоретическая часть

Основные понятия и характеристики стохастического процесса

1. <u>Вероятность</u> - мера возможности осуществления результата. Формально это функция P(x), которая ставит в соответствие результатам некоторые вещественные числа и удовлетворяет аксиомам:

1. $0 \le P(x) \le 1$, для любого результата x,
2. P(s) = 1, где s – пространство выборки,
3. если x_1, x_2, x_3 – взаимоисключающие результаты, то вероятность их появления равна сумме вероятностей появления каждого из x_i, то есть $P(x_1 \cup x_2 \cup x_3 \ldots) = P(x_1)+P(x_2)+P(x_3)+\ldots$

2. <u>Случайной величиной</u> называется функция, которая ставит в соответствие каждому результату из пространства выборки некоторое вещественное число.

Случайные величины, имеющие конечное или счетное множество значений, называются <u>дискретными</u>: а если они имеют континуум значений, то являются <u>непрерывными</u> случайными величинами.

3. <u>Вероятностное</u> распределение представляет собой некоторое правило задания вероятности для каждого из всех возможных значений случайной переменной.

Вероятностное распределение характеризуется <u>функцией</u> <u>вероятности</u> p(x) и <u>функцией распределения</u> F(x), Функция вероятности устанавливает конкретную вероятность того, что случайная переменная X принимает значение x_i; функция, распределения определяет вероятность того, что случайная величина X примет значение меньше заданного x.

Для дискретной случайной величины эти характеристики определяются следующим образом:

- $p(x_i) = P(X = x_i)$ со следующими ограничениями:
    - $0 \le P(x_i) \le 1 \forall i$
    - $\sum P(x_i) = 1$

- F(x) = P(X < x)   со следующими свойствами:
    - $0 \le F(x) \le 1, \forall i$
    - $F(-\infty) = 0, F(+\infty) = 1$

Функция распределения и функция вероятностей связаны следующим образом: $F(x) = \sum_{x_i < x} P(x_i)$

Для непрерывных случайных величин функция вероятности заменяется на непрерывную функций плотности вероятности f(x), определяемую как: $P(a \le x \le b) = \int_a^b f(x)dx$;

Функция распределения:

$$\boldsymbol{F(x) = \int_{-\infty}^{x} f(y)dy = P(X < x)}$$

4. Математическое ожидание - взвешенная по вероятности средняя величина всех возможных значений X, определяющая меру центральности распределения.

- $M(X) = \sum_i [x_i * p(x_i)]$ для дискретной X
- $M(X) = \int_{-\infty}^{+\infty} x f(x)dx$ если X непрерывна.

5. Математическое ожидание X^n называется n-м моментом случайной переменной и определяется следующим образом:

- $M(X^n) = \sum_i [x_i{}^n * p(x_i)]$ для дискретной X
- $M(X^n) = \int_{-\infty}^{+\infty} x^n f(x)dx$ если X непрерывна.

Вариацией n-го момента называется n-й момент среднего: $M([X - M(X)]^n)$.

6. Дисперсией случайной переменной называется второй момент среднего, являющийся мерой разброса вероятностного распределения:

$D(X) = \sum_i \left[\left(x_i - M(x)\right)^2 p(x_i)\right]$ для дискретной X

$D(X) = \int_{-\infty}^{+\infty} \left(x - M(x)\right)^2 f(x)dx$ для непрерывной X

$D(X) = M(X^2) - (M(X))^2$ для любой.

7. Среднеквадратичным отклонением случайной величины называется квадратный корень из дисперсии этой величины : $\sigma = \sqrt{D(X)}$.

Функционирование элементов системы, подверженных случайным воздействиям, задается *генераторами случайных чисел,* реализуемых программными методами, вырабатывающими псевдослучайные последовательности. *Псевдослучайными последовательностями* называют вполне детерминированные числа, обладающие статистическими свойствами случайных чисел, определяемых путем их проверки специальными тестами, а также периодичностью, т. е. повторяемостью через определенные промежутки времени. При моделировании используются интервалы последовательностей псевдослучайных чисел, в которых нет одного числа, встречающегося более одного раза.

Равномерно распределенные случайные числа в процессе генерации случайных величин с другими вероятностными распределениями, в том числе нормальным, пуассоновским и биномиальным. В настоящее время известно множество методов машинной имитации равномерно распределенной случайной величины r (где $0 \le r \le 1$). Соответствующие программы-генераторы вычисляют так называемые псевдослучайные числа, которые, хотя и определяются вполне детерминированными отношениями, обладают статистическими свойствами случайных чисел, равномерно распределенных на интервале (0,1). Псевдослучайную последовательность, выдержавшую ряд специальных статистических тестов (частотных, автокорреляционных и т. д.), можно использовать как «истинно» случайную, хотя в действительности она таковой не является.

Прежде чем перейти к описанию конкретных генераторов псевдослучайных чисел, сформулируем набор требований, которым обязан удовлетворять *идеальный* генератор. Полученные с его помощью последовательности должны состоять из:

(1) равномерно распределенных,

(2) статистически независимых,

(3) воспроизводимых и

(4) неповторяющихся чисел.

Кроме того, генератор должен (5) работать быстро и (6) занимать минимальный объем машинной памяти. Конгруэнтные процедуры, описанные здесь, удовлетворяют *всей* совокупности перечисленных требований лучше, чем любые другие методы генерирования случайных чисел.

Мы рассмотрим 3 метода построения псевдослучайных последовательностей на ЭВМ — (1) мультипликативный, (2) смешанный и (3) комбинационный методы. После описания механизмов и некоторых свойств этих процедур мы кратко остановимся на проблеме автокорреляции псевдослучайных чисел и различных тестах, с помощью которых оценивается их «случайность».

Методы генерирования псевдослучайных чисел. Конгруэнтные методы

Для любой конгруэнтной процедуры можно получить формулу, позволяющую найти i-й член определяемой ею последовательности псевдослучайных чисел $\{n_0, n_1, n_2, \ldots, n_i, \ldots\}$, не зная никаких членов, кроме нулевого. Тем не менее, в практических приложениях эту последовательность можно рассматривать как реализацию случайного процесса, если только она выдержит ряд специальных статистических тестов, состав которых определяется способом ее применения. Иногда, например, достаточно, чтобы псевдослучайные числа были равномерно распределены и статистически независимы. (Можно показать, что конгруэнтные методы достаточно хорошо удовлетворяют этим условиям.) Все конгруэнтные генераторы автоматически удовлетворяют также требованиям (3) и (6), перечисленным выше, поскольку генерируемые с их помощью последовательности полностью воспроизводимы и занимают минимальный объем машинной памяти. Только степень выполнения (4) и (5) целиком определяется свойствами конкретной схемы. Поэтому рассматриваемые в дальнейшем процедуры анализируются именно с этих позиций.

В основе конгруэнтных процедур генерирования псевдослучайных чисел лежит математическое понятие *сравнения.*

Говорят, что два целых числа a и b *сравнимы по модулю m,* если их разность кратна числу m. Отношение сравнения записывают a≡b(mod m) и читают: «а сравнимо с b по модулю m». Это значит, что разность $a - b$ делится на m без остатка, т. е. числа a и b дают одинаковые, остатки при делении на m. Например, 1897≡7 (mod 5) и 4339≡39 (mod 10^2).

Все конгруэнтные методы опираются на рекуррентную формулу:

$$n_{i+1} \equiv \lambda n_i + \mu \ (mod\ m), \qquad (1)$$

где n_i, m, λ и μ — неотрицательные целые числа. Запишем (1) при i=0, 1, 2...:

$$n_1 \equiv \lambda n_0 + \mu \ (mod\ m),$$

$$n_2 \equiv \lambda n_1 + \mu = \lambda^2 n_0 + (\lambda+1)\mu \ (mod\ m),$$

$$n_3 \equiv \lambda^3 n_0 + (\lambda^2+\lambda+1)\mu = \lambda^3 n_0 + \frac{\mu(\lambda^3-1)}{(\lambda-1)} (mod\ m), \quad (2)$$

…

$$n_i \equiv \lambda^i n_0 + \frac{\mu(\lambda^i-1)}{(\lambda-1)} (mod\ m).$$

Если дано *начальное значение* n_0, *множитель* λ и *аддитивная константа* μ, то (2) определяет последовательность целых чисел $\{n_1, n_2, \ldots, n_i, \ldots\}$, составленную из остатков от деления на m членов последовательности $\left\{\lambda^i n_0 + \frac{\mu(\lambda^i-1)}{(\lambda-1)}\right\}$. Таким образом, для любого i≥1 справедливо неравенство $n_i < m$. По целым числам последовательности $\{n_i\}$ можно построить последовательность $\{r_i\}= = \{n_i/m\}$ рациональных чисел из единичного интервала.

Естественно возникает вопрос о существовании наименьшего положительного числа h, при котором $n_h \equiv n_0$(mod m) (число h называют периодом последовательности $\{n_i\}$). Если такое число b, существует, то как надо выбрать параметры n_0 , т, λ и μ , чтобы сделать период максимальным? Величина h существенна, поскольку $n_h \equiv n_0$(mod m) влечет за собой $n_{h+1}=n_1$,

$n_{h+2}=n_2$,.... Таким образом, последовательность $\{n_i\}$ будет периодической, т. е. значения ее членов будут повторяться через h номеров.

Можно показать, что период h всегда существует, причем его максимальное значение является монотонно возрастающей функцией модуля m. Это значит, что конгруэнтные методы не позволяют строить последовательности неповторяющихся псевдослучайных чисел. Тем не менее, в практических ситуациях можно обеспечить удовлетворительную величину периода h, выбирая достаточно большой модуль или применяя другие способы.

Мультипликативный конгруэнтный алгоритм задает последовательность неотрицательных целых чисел $\{n_i\}$, не превосходящих *m,* по формуле:

$$n_{i+1} \equiv \lambda n_i (mod\ m) \qquad (3)$$

представляющей собой частный случай формулы (1) при $\mu=0$. Оказалось, что этот метод обладает достаточно хорошими статистическими характеристиками. Выбирая соответствующим образом параметры λ и n_0, с его помощью можно получить последовательности равномерно распределенных некоррелированных псевдослучайных чисел.

Более того, при выполнении определенных условий на эти параметры генерируемая мультипликативным алгоритмом последовательность обладает максимальным при данном модуле периодом. Поскольку метод полностью детерминирован, последовательности полностью воспроизводимы. Мультипликативный алгоритм требует минимального объема машинной памяти. С вычислительной точки зрения он сводится к последовательному подсчету произведения двух целых чисел — операции, которая очень быстро выполняется современными ЭВМ.

Для численной реализации наиболее удобна версия алгоритма, в которой модуль m равен p^e, где p — число цифр в системе счисления, используемой в машине, а e — длина машинного слова, отводимого под запись числа. Для двоичной машины $p=2$, для десятичной $p=10$. Величина e для ЭВМ с фиксированной длиной слова задана, а для ЭВМ с переменной длиной слова ее выбор остается на усмотрение программиста. В дальнейшем вместо e мы будем употреблять символы b и d, относящиеся соответственно к двоичной и десятичной ЭВМ.

Выбор $m=p^e$ удобен по двум причинам.

1. Во-первых, вычисление остатка от деления на т сводится к выделению е младших разрядов делимого;
2. во-вторых, преобразование целого числа в рациональную дробь из интервала (0,1) осуществляется подстановкой слева от него двоичной или десятичной запятой.

Таким образом, специальный выбор числа m позволяет исключить из процесса счета две операции деления.

Так как в большинстве ЭВМ используются двоичная и десятичная системы счисления, мы рассмотрим мультипликативный метод именно в этих двух случаях. При этом нас будут интересовать только последовательности, порождаемые отношением сравнения вида

$$n_{i+1} \equiv \lambda n_i (mod\ p^e)$$

Двоичные машины

В этом случае m==2^b, где b — число двоичных цифр (битов) в машинном слове. Максимальный период последовательности, генерируемой мультипликативной процедурой, равен 2^{b-2}. Алгоритм построения последовательности с максимальным периодом таков:

1. Выбрать в качестве параметра n_0 произвольное нечетное число.
2. Вычислить коэффициент λ по формуле λ=8t±3, где t — любое целое положительное число.
3. Вычислить произведение λn_0. Полученное число содержит не более *2b* значащих разрядов. Взять b младших в качестве первого члена последовательности, n_1 остальные отбросить.
4. Вычислить дробь из интервала (0,1) по формуле $r_1=n_1/2^b$.
5. Вычислить очередное псевдослучайное число n_1 +1 как b правых разрядов произведения λn_1 и вернуться к п. 4.

В качестве иллюстрации описанной схемы рассмотрим пример, в котором b=4. Мультипликативная процедура должна определить 4 разных числа ($h=2^{4-2}=4$).

1. Положим n_0=7. В двоичной форме записи n_0=0111.
2. При t=1 коэффициент λ можно взять равным 11 или 5. Мы выберем λ =5, или в двоичной форме, λ = 0101.
3. λn_0= (0101)(0111) = 00100011. Отсюда n_1 = 0011 и r_1 = 3/16 = 0,1875.
4. λn_1= (0101)(0011) = 00001111. Отсюда n_2 = 1111 и r_2 = 15/16 = 0,9375.
5. λn_2= (0101)(1111) = 01001011. Отсюда n_3 = 1011 и r_3 = 11/16 = 0,6875.
6. λn_3= (0101)(1011) = 00110111. Отсюда n_4 = 0111 =n_0 и r_4 = 7/16 = 0,4375.

Для 32х разрядного машинного слова m=2^{31} = 2147483648. При этом в 32хразрядном машинном слове, максимальное целое число, размещающееся в машинном слове, равно 2^{31} -1, следовательно m = 2147483647, b - простое число относительно m: b=2531011, λ=8t±3, где t целое число, λ = 214013.

Метод Неймана (метод квадратов)

В квадрат возводится, текущее случайное число и из результатов средних разрядов выделяется следующее случайное число, т.е. используется итерация:

- z_0 равно некоторому n- разрядному целому числу (например, 9876)
- z_1 определяется следующим образом : $z_0 * z_0 = 97\underline{5353}76$, $z_1 = 5353$ $z_2 = 6546$ и т.д.

$r_{i+1} = z_{i+1} / 10^n$

3. Метод произведений

Два следующих друг за другом случайных числа умножают и из произведения средних разрядов выделяют следующее случайное число.

Проверка генераторов равномерно распределенных псевдослучайных чисел

Выделяют три вида проверки: на периодичность, на случайность, на равномерность. При проверке на случайность используется совокупность тестов проверки: 1) частот, 2) пар, 3) комбинаций, 4) серий, 5) корреляции.

Тест проверки частот предполагает разбиение диапазона распределения на q интервалов и подсчет количества попаданий случайных чисел в выделенные интервалы. Возможно использование критериев согласия. Вероятность попаданий в заданный интервал теоретического распределения определяется по формуле

$$P_i = \int_{x_{i-1}}^{x_i} \frac{1}{b-a} dx = \frac{x_i - x_{i-1}}{b-a}$$

где x_i - верхний предел i-го интервала.

Тест проверки пар заключается в подсчете количества «1» для каждого разряда случайного числа. В этом случае используется критерий согласия χ^2 с одной степенью свободы. Теоретическая вероятность появления «1» для равномерно распределенных случайных чисел $p_i = 1/2$. Поразрядный анализ позволяет отбросить неслучайные разряды, в качестве которых часто оказываются младшие разряды с преобладанием в них «1».

Тест проверки комбинаций сводится к подсчету количества «1» в случайных числах. Можно использовать также критерии согласия. Теоретическая вероятность появления комбинации с i-м количеством «1» будет: $P_i = \frac{k!}{i!(k-j)!2^k}$, где k — количество разрядов случайного числа.

Тест проверки серий заключается в подсчете количества различных длин последовательностей одинаковых значений случайных чисел. Возможно использование критериев согласия. Теоретическая вероятность $P_i = \frac{R_i}{N_s}$, где R_i — количество серий i-и длины в N случайных числах; N_s— общее количество серий в N случайных числах при гипотетическом распределении.

Характерная особенность приведенных формул теста серий — независимый учет серий различной длины. Например, одна и та же серия, состоящая из четырех единиц, учитывается как одна серия из четырех, две — из трех и три — из двух единиц.

Тест проверки корреляции заключается в определении коэффициента корреляции. При этом выполняют следующие действия:

- запускают два генератора случайных чисел на отрезке апериодичности с некоторой разницей между собой;
- подсчитывают коэффициент корреляции между этими последовательностями.

Проверка на равномерность. При проверке на равномерность можно использовать тест проверки частот, так как гистограмма частот хорошо отражает равномерность распределения случайных чисел по всему диапазону изменения. Для равномерного распределения случайных чисел $m_1 = \frac{(a_q+a_0)}{2}$ и $\sigma = \frac{(a_q-a_0)}{2\sqrt{3}}$.

Задаемся доверительной вероятностью w того, что оценка математического ожидания M^* не выйдет за пределы доверительного интервала:

$P\{|M - M^*| < e\} = w$ или $P\{M - e < M^* < M + e\}= w$,

где M=(a+b)/2, - теоретические величины математического ожидания и среднеквадратичного отклонения, e - бесконечно малая величина. Величина w должна быть достаточно большой и составлять 0.9, 0.95, 0.99.

ЗАДАНИЕ 1

Для стандартного генератора случайных чисел выбранного Вами языка программирования получить три последовательности N случайных чисел (N={100, 1000, 10000}), для которых определить следующие характеристики: математическое ожидание M, дисперсию D и среднеквадратичное отклонение .

Выполнить проверку частотности и равномерности генератора.

Построить графики Функций P(X) для оценки частотности генератора. Для получаемой выборки N чисел P(X) – вероятность попадания генерируемой случайной величины в соответствующий интервал ее области определения.

Сравнить результаты с теоретическими.

Для оценки равномерности генератора случайных чисел выполнить расчет математического ожидания M_i для i последовательностей из 1000 случайных чисел (i=1,2,..., 10) и для i последовательностей случайных чисел переменной длины (длина i-ой последовательности задается как i*1000; i = 1, 2, ..., 10). Построить графики зависимости разности ($M-M_i$) от номера последовательности i, где M - теоретическое математическое ожидание равномерного распределения случайных чисел, M_i - расчетное математическое ожидание для i-й последовательности случайных чисел, полученных от генератора. По данным результатам определить $P\{|M-M_i|<\sigma\}$ - вероятность того, что отклонения расчетного математического ожидания от теоретического не превышают величину теоретического среднеквадратичного отклонения.

ЗАДАНИЕ 2

Запрограммировать заданный вариант генератора случайных чисел и выполнить для него задание 1.

Порядок выполнения работы:

1. Изучить теоретическую часть, ответить на контрольные вопросы.
2. Получить у преподавателя свой вариант задания
3. Выполнить задания 1 и 2.
4. Оформить отчет.

Содержание отчета:

1. Описание и текст процедур заданий 1 и 2. Полученные результаты и графики.

2. Сравнение характеристик стандартного и запрограммированного генераторов случайных чисел.

3. Выводы.

Контрольные вопросы

1. Пусть стохастический процесс заключается в 3-х бросаниях монеты. Случайная величина X, характеризующая его, пусть обозначает число орлов, выпавших в результате 3-х бросаний. Построить для X функции P(x) и F(x).

2. Построить F(x) по заданной f(x).

3. Может ли быть случайная величина одновременно и дискретной и непрерывной?

4. Каков "физический" смысл понятий математическое ожидание и дисперсия?

5. Какие способы генерации случайных чисел Вы знаете?

6. Каким требованиям должны удовлетворять генераторы псевдослучайных чисел?

Варианты заданий

1. Метод серединных квадратов
2. Метод серединных произведений
3. Метод перемешивания
4. Линейный конгруэнтный метод
5. Инверсный конгруэнтный метод

Лабораторная работа № 2

Получение случайной величины, распределенной по заданному закону

Цель работы. Изучение различных законов распределения случайных величин и способов их получения при моделировании случайных процессов.

Теоретическая часть

ГЕНЕРИРОВАНИЕ ДАННЫХ

В программе имитации на ЭВМ часто применяются численные методы (т.е. методы, которые можно запрограммировать на вычислительной машине) генерирования данных. Информацию, используемую в имитационном эксперименте, можно либо ввести в ЭВМ с внешних источников, таких, как перфокарты и магнитные ленты, либо генерировать при помощи специальных программ. Если среди экзогенных переменных модели есть случайные величины с известным вероятностным распределением, то надо построить численный процесс случайного выбора из совокупности с заданным распределением. Результатом повторения этого процесса на цифровой вычислительной машине должно быть такое вероятностное распределение выборочных значений, которое соответствует вероятностному распределению изучаемой, переменной.

При рассмотрении дискретных или непрерывных случайных процессов вводят функцию F*(x),* называемую *кумулятивной функцией распределения* величины *X.* Эта функция задает вероятность того, что случайная величина *X* принимает значение, не превосходящее число *x.* Если случайная величина дискретна, т.е. *X* принимает конечное число значений, то функция F*(x)* является ступенчатой. Если функция F*(x)* непрерывна, то ее можно продифференцировать и положить f(x)=dF(x)/dx. Функция f*(x)* называется функцией плотности вероятностей. Кумулятивную функцию распределения можно определить как $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$, где F*(x)* изменяется на отрезке [0,1], а *f(t)* представляет собой значение функции плотности вероятностей случайной величины *X* при *X=t.*

При генерировании случайных величин, имеющих различные функции распределения, используются равномерно распределенные случайные величины. Равномерно распределенные случайные величины будем обозначать через r*,* $0 \leq r \leq 1$*, F(r)=r.*

В лабораторной работе №1 дан обзор методов генерирования случайных величин, равномерно распределенных на интервале (0,1). Числа, получаемые таким образом, называются псевдослучайными, так как, хотя они и генерируются на ЭВМ при помощи чисто детерминированной рекурсивной формулы, их статистические свойства совпадают со статистическими свойствами чисел, генерированных идеальным случайным механизмом, выбирающим числа из интервала (0,1) независимо и с одинаковой вероятностью. Пока эти псевдослучайные числа удовлетворяют некоторому набору статистических критериев (частотный, сериальной корреляции, интервалов, пар и др.), отражающих свойства идеального случайного механизма, их можно считать «истинно» случайными числами, хотя на самом деле это не так.

Генераторы псевдослучайных чисел в виде подпрограмм есть во всех вычислительных машинах и в большинстве языков программирования.

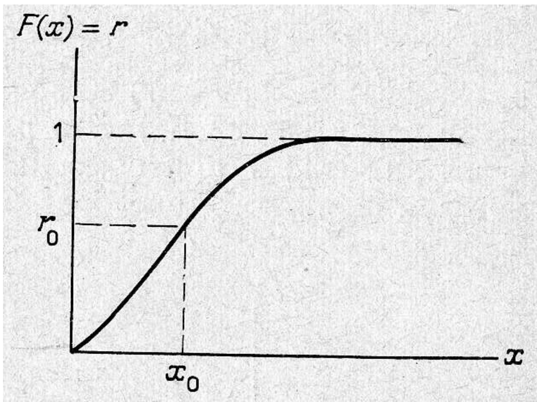

Рисунок 1. Кумулятивная функция распределения

Если требуется генерировать случайные числа x_i из некоторой статистической совокупности с функцией плотности вероятностей f*(x),* то сначала строят кумулятивную функцию распределения F*(x)* (рис. 1). Так как F*(x)* изменяется на отрезке [0,1], то, чтобы получить случайные числа с этим распределением, можно генерировать равномерно распределенные случайные числа r и полагать F*(x)=r.* Ясно, что величина *x* однозначно определяется из этого соотношения. Следовательно, для конкретного значения r*,* скажем r0*,* можно найти величину *x,* в данном случае x_0*,* связанную с r_0 обратной функцией к F (если она известна):

$$x_0 = F^{-1}(r_0)$$

где $F^{-1}(r)$ — обратное отображение величины r*,* заданной на единичном интервале, в область изменения *x.* Математически этот метод можно выразить следующим образом: если мы генерируем равномерно распределенные случайные числа и ставим их в соответствие данной функции F(x), то есть

$$r = F(x) = \int_{-\infty}^{x} f(t)dt, \text{ то}$$

$$P(X \le x) = F(x) = P[r \le F(x)] = P[F^{-1}(r) \le x],$$

и следовательно $F^{-1}(r_0)$ есть случайная величина с функцией плотности вероятностей f(x). Это равносильно выражению величины x через значение r при помощи (2.3). Такая процедура называется методом обратного преобразования.

Самым простым непрерывным распределением является, по-видимому, распределение с функцией плотности вероятностей, постоянной на интервале *(a, b)* и равной нулю вне его. Эта функция плотности вероятностей определяет так называемое равномерное, или прямоугольное, распределение. Равномерное распределение часто применяется в имитационных методах, во-первых, потому, что оно просто, а во-вторых, потому, что его можно использовать для генерирования случайных величин с другими вероятностными распределениями.

Функция плотности вероятностей равномерного распределения имеет вид

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, \text{в противном случае} & \end{cases}$$

Здесь *X* — случайная величина, определенная на интервале *(a, b),* График равномерного распределения изображен на рис, 2.2,

Кумулятивная функция распределения F(x) равномерно распределенной случайной величины *X* равна

$$F(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}$$

Для имитации равномерного распределения на интервале *(a, b)* сначала в соответствии с формулой (2.2) надо получить обратное преобразование для (2.6):

$x=a+(b—a)r$, 0<r<1. (2.7)

Далее генерируются случайные числа, равномерно распределенные в интервале (0,1). Каждое случайное число r однозначно определяет реализацию равномерно распределенной случайной величины X.

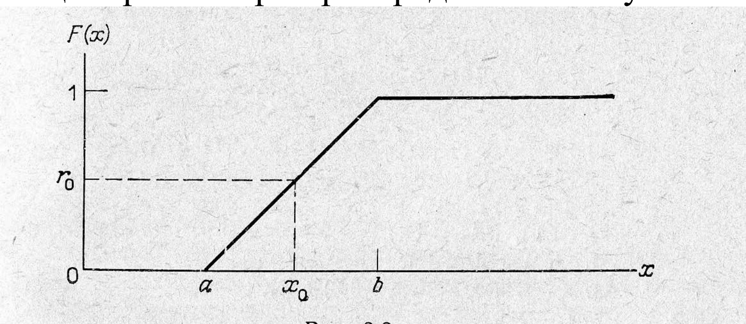

Рис. 2.3.

На рис. 2.3 видно, что каждому значению r соответствует единственное значение x. Так, конкретное значение кумулятивной функции распределения, равное r_0, определяет значение x, равное x_0. Очевидно, что процедуру можно повторять нужное число раз, и каждый раз она будет давать новое значение x.

Основные виды распределений случайных величин, используемые в моделировании

1. Равномерное распределение (прямоугольное распределение).

Функция плотности вероятности этого распределения задает вероятность того, что некоторое значение попадает в заданный интервал [a,b], и эта вероятность пропорциональна длине этого интервала.

Это распределение применяют часто в условиях полного отсутствия информации о случайной величине кроме ее предельных значений. Равномерное распределение характеризуется

a. функцией плотности вероятности: f(x)=1/(b-a), a≤x≤b;
b. математическим ожиданием: M=(a+b)/2;
c. дисперсией: D=(b - a)^2/12.

2. Треугольное распределение

Для этого распределения определяют 3 величины: минимум a, максимум b, и моду m ($m = \frac{a+b}{2}$). График функции плотности вероятности состоит из двух отрезков прямых, один из которых возрастает при изменении x от минимального значения до моды, а другой убывает при изменении x от значения моды до максимума. Это распределение используется тогда, когда известно наиболее вероятное значение на некотором интервале и предполагается кусочно-линейный характер функции плотности вероятности.

Треугольное распределение характеризуется:

a. $f(x) = \begin{cases} \frac{2*(x-a)}{(m-a)*(b-a)} & a < x < m \\ \frac{2*(b-x)}{(b-m)*(b-a)} & m < x < b \end{cases}$

b. M = (a + b + m)/3;

c. $D = \frac{a(a-m)+b(b-a)+m(m-b)}{18} \qquad a \leq m \leq b$

3. Экспоненциальное распределение

Если вероятность того, что один и только один результат наступит на интервале Δt пропорциональна Δt, и если наступление результата не зависит от наступления других результатов (т.е. процесс характеризуется отсутствием последействия), то величины интервалов между результатами распределены экспоненциально.

Это распределение характеризуется функцией плотности распределения: $f(x) = \frac{1}{M} * e^{-\frac{x}{M}}$, где x>0, M – математическое ожидание; D=M^2.

4. Распределение Пуассона

Это распределение является дискретным и связано обычно с числом результатов за определенный период времени. Если интервалы между появлением результатов распределены экспоненциально, то число, появившихся результатов в данный отрезок времени будет распределено в соответствии с распределением Пуассона.

Характеристики распределения:
$P(X = x) = \frac{e^{-M} * M^{x}}{x!}$, x=0,1,2,...;
D = M.

5. Нормальное (гауссово) распределение

Широко применяется в моделировании. Это определено значением центральной предельной теоремы, которая утверждает, что при весьма нестрогих условиях распределение средней величины или суммы N-независимых наблюдений из любого распределения стремится к нормальному по мере увеличения N. (**Центральные предельные теоремы** — класс теорем в теории вероятностей, утверждающих, что сумма достаточно большого количества слабо зависимых случайных величин, имеющих примерно одинаковые масштабы (ни одно из слагаемых не доминирует, не вносит в сумму определяющего вклада), имеет распределение, близкое к нормальному. Так как многие случайные величины в приложениях формируются под влиянием нескольких слабо зависимых случайных факторов, их распределение считают нормальным. При этом должно соблюдаться условие, что ни один из факторов не является доминирующим. Центральные предельные теоремы в этих случаях обосновывают применение нормального распределения. [Википедия])

Распределение характеризуется $f(x) = \left\lfloor \frac{1}{\sqrt{2\pi D}} \right\rfloor * exp\left\lfloor -\frac{(\mathrm{x}-M)^{2}}{2D} \right\rfloor$, x∈(–∞, +∞).

M=m;

D=σ^2.

6. Логарифмическое нормальное распределение

Это распределение такой случайной величины, натуральный логарифм которой нормально распределен. Оно пригодно для моделирования мультипликативных процессов так же, как нормальное - для аддитивных.

Распределение характеризуется функцией плотности вероятности: $f(x) = \left\lfloor \frac{1}{x\sqrt{2\pi D}} \right\rfloor * exp\left\lfloor -\frac{(\ln(x)-M)^{2}}{2D} \right\rfloor$

7. Биномиальное распределение .

Это распределение вероятностей случайной величины X с целочисленными значениями x=0,1,2,...,n.

Пусть проводится эксперимент, в результате которого нас интересует, произошло событие A или не произошло. Случай, в котором событие A произошло, назовем успехом, вероятность этого события P(A) = p. Если же событие A не произошло, то его вероятность P($\bar{A}$) = 1 – p = q. Предположим теперь, что серия независимых испытаний такого типа проводится n раз. Нас интересует вероятность события, состоящего в том, что успех произошел ровно m раз, или вероятность того, что дискретная случайная величина X, равная числу успехов, примет значение x. Решение этой задачи имеет вид:

Функция вероятности характеризуется формулой: $P(X = x) = C_n^x p^x (1-p)^{n-x}$, где $C_n^x = \frac{n!}{x!(n-x)!}$; 0≤p≤1 ; n≥1.

Функция распределения: $F(x) = \begin{cases} \sum_{k=1}^{l} C_n^k p^k (1-p)^{n-k}, l \le x < l+1 \\ 1, x \ge n \\ 0, x < 0 \end{cases}$

Математическое ожидание: M(x) = n*p

Дисперсия: D(x) = n*p*(1-p).

8. Распределение Вейбулла

Это распределение характеризуется функцией распределения: $F(x) = 1 - \exp\left(\frac{1}{b} * x^a\right)$, где b - параметр масштабирования, a -параметр кривизны.

Распределение Вейбула часто используется в теории надежности для описания времени безотказной работы приборов.

9. Распределение Коши

Это распределение характеризуется плотностью вероятности: $f(x) = \frac{1}{\pi} * \frac{l}{l^2 + (x-m)^2}$, где l и m (мода и медиана) - параметры данной функции, –∞<m<+∞ и l>0.

10. Распределение Эрланга

Это распределение характеризуется плотностью вероятности

$$f(x) = \frac{(kx/M)^{k-1} * e^{-kx/M}}{\left(\frac{M}{k}\right) * (k-1)}$$

где k > 0 - коэффициент Эрланга, m - математическое ожидание.

Методы получения случайных величин, распределенных по заданному закону

1. Метод обратной функции

В основе метода лежит факт того, что случайная величина r=F(x) равномерно распределена на интервале [0,1]. Для генерации случайной величины из распределения x генерируется случайное число r и решается уравнение r = F(x) относительно значения $x=F^{-1}(r)$.

Например, функция экспоненциального распределения имеет вид $F(x) = 1 - e^{-\lambda x}$, где 1/λ – математическое ожидание. Приравнивая F(x) = r и решая уравнения относительно x, получаем $x = -\frac{1}{\lambda} * \ln(r)$.

Этот метод применим и для дискретных распределений (см. рис. 1).

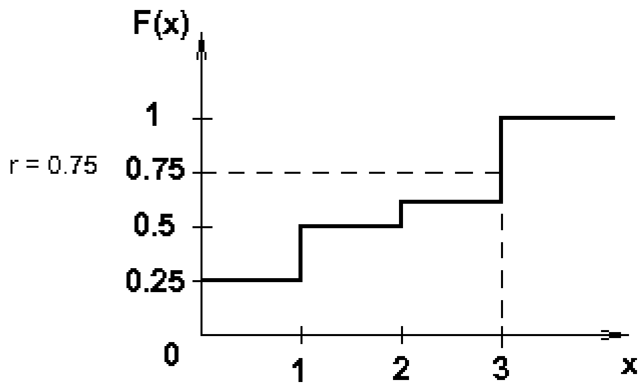

рис 1

Достоинства метода: точность метода; не требуется составления и хранения в памяти таблиц.

Основная трудность этого метода - поиск обратного преобразования $F^{-1}(r)$. Для ряда непрерывных распределений представление обратной функции в явном виде отсутствует. Для всех основных распределений, не имеющих явного представления обратной функции, разработаны специальные методы генерации.

2. Табличный метод

В качестве аргумента используется равномерно распределенное случайное число r, в качестве функции – последовательность чисел x_i, задающих закон распределения. Для этого формируется таблица $<F(x_i), x_i>$, i =1, 2, ..., N.

Значение случайного числа X с заданным законом распределения находят методом линейной интерполяции по формуле:

$X = x_{i-1} + \frac{(r_j - F(x_i))}{F(x_i) - F(x_{i-1})} * (x_i - x_{i-1})$, где $F(x_{i-1}) < r_j \le F(x_i)$; i=1, 2, . . , N.

Поиск нужного интервала производится методом последовательного сравнения j-го случайного числа с границами интервалов $F(x_i)$, i =1,2,...,N до выполнения условия $F(x_{i-1}) < r_j \le F(x_i)$.

Достоинства табличного метода: имеется возможность генерировать случайные последовательности с любым заданным законом; любую заданную точность можно получить при увеличении количества интервалов; требуется только одно случайное равномерное распределенное число и выполнение несложных операций, занимающих мало времени.

3. Использование функциональных особенностей распределений

Этот метод используется в тех случаях, когда аналитически не удается вычислить интеграл от функции плотности вероятности.

Так, для генерации случайных чисел X, имеющих специальное эрланговское распределение, можно воспользоваться k равномерно распределенными случайными числами r_i:

$$X = -\frac{1}{k\mu}\sum_{i=1}^{k}\ln(r_i) = -\frac{1}{k\mu}\ln\left(\prod_{i=1}^{k} r_i\right)$$

Для генерации нормально распределенных случайных чисел используется центральная предельная теорема, на основании которой суммируются N равномерно распределенных случайных чисел для получения нормально распределенного случайного числа X. Обычно, принимают N = 12..20.

$X = \frac{\sum_{i=1}^{k} r_j - \frac{N}{2}}{\sqrt{N}/2\sqrt{3}} * \sigma + M$, где σ – требуемое среднеквадратичное отклонение, а M – требуемое математическое ожидание генерируемых случайных чисел.

Оценка качества случайных последовательностей

Для оценки качества случайных последовательностей с заданным законом распределения используется тест проверки частот и метод доверительного интервала для математического ожидания.

ЗАДАНИЕ

Запрограммировать генерацию случайных величин по заданному закону распределения, определяемую вариантом задания. Для выборки из 100 случайных величин определить их характеристики: математическое ожидание М, дисперсию D и среднеквадратичное отклонение σ; построить графики функций плотности вероятностей f(X) / p(X) / и F(X). Выполнить оценку качества полученной случайной последовательности.

Порядок выполнения работы:

1. Изучить теоретическую часть.

2. Ответить на контрольные вопросы.

3. Получить у преподавателя номер варианта задания.

4. Выполнить задание и оформить отчет.

Содержание отчета.

1. Описание исходных данных и хода выполнения задания.

2. Результаты моделирования (характеристики, графики).

3. Выводы.

Контрольные вопросы

1. Что такое распределение случайной величины и какими функциями оно характеризуется?

2. Как вычисляются функции f(x) по F(x) и наоборот F(x) по f(x)?

3. Качественно изобразите графики функций f(x) и F(x) для основных видов распределений случайных величин.

4. Сформулируйте основные положения метода обратной функции.

5. Сформулируйте основные положения табличного метода.

6. Постройте графики p(x) и F(x) для дискретного распределения, заданного следующей функцией вероятности: p(0)=0,5 p(1)=0,3 p(2)=0,2 и, используя метод, обратной функции преобразуйте случайные числа 0,025; 0,91; 0,37; 0,26; 0,31 в выборку заданного распределения.

Варианты задания для лаб. раб. 2

/ дискретный закон распределения случайной величины /

Номер варианта	Вид закона распределения	Характеристики		
		Мат. Ожидание	Max	p
1	З-н Пуассона	26	78	
2	З-н Бернулли		78	0,3
3	З-н Пуассона	38	114	
4	З-н Бернулли		114	0,9
5	З-н Пуассона	22	66	
6	З-н Бернулли		66	0,04
7	З-н Пуассона	30	90	
8	З-н Бернулли		90	0,71
9	З-н Пуассона	31	93	
10	З-н Бернулли		93	0,92
11	З-н Пуассона	33	99	
12	З-н Бернулли		99	0,63
13	З-н Пуассона	23	69	
14	З-н Бернулли		69	0,19
15	З-н Пуассона	53	159	
16	З-н Бернулли		159	0,52
17	З-н Пуассона	40	120	
18	З-н Бернулли		120	0,81
19	З-н Пуассона	28	84	
20	З-н Бернулли		84	0,89
21	З-н Пуассона	24	72	
22	З-н Бернулли		72	0,66
23	З-н Пуассона	39	117	

24	З-н Бернулли		117	0,66
25	З-н Пуассона	52	156	