**Задание 1.**

Генератор псевдослучайных чисел предполагает получение последовательностей действительных чисел, равномерно распределённых на отрезке от 0 до 1. Обычно такой генератор обозначается U(0;1), U – от слова Uniform (равномерный). Имея в распоряжении генератор случайных чисел, можно получать случайные величины с любым другим заданным законом распределения, для этого необходимо аналитически или численно решить уравнение вида F(x)=U(0;1), где F(x) – функция распределения случайной величины.

Схема работы генераторов случайных чисел обычно предполагает получение последовательностей не действительных, а целых неотрицательных чисел. Переход к действительным величинам на отрезке от 0 до 1 достигается делением на максимальное целое число из рабочего диапазона.

Одной из наиболее простых (и в то же время распространённых) разновидностей генераторов псевдослучайных чисел являются так называемые линейные конгруэнтные (однотипно работающие на каждом шаге) генераторы. Схема их работы определяется следующей формулой:

Каждый очередной элемент *Xi* последовательности вычисляется как линейная функция (определяемая коэффициентами *a* и *c*) на основе предыдущего значения *Xi*-1, к которой применяется операция взятия остатка по модулю *m*. Набор из трёх указанных констант задаёт независимый генератор (поток) случайных чисел. В частности, в GPSS при обращении к библиотечным функциям генерации случайных величин, первым их аргументом указывается номер потока – т.е. фактически выбирается набор таких констант. В большинстве систем программирования общего назначения в стандартных библиотеках вообще реализован единственный поток случайных чисел. Для того, чтобы получать разные последовательности при многократном запуске программ, изменяют всего лишь начальное значение X0 (например, с использованием функции srand из состава стандартной библиотеки языка C).

Очевидно, выбор модуля *m* определяет наибольшее возможное число разных значений в последовательности *Xi*. Однако может оказаться, что повторение значений начнётся гораздо раньше (другими словами начнётся новый период последовательности), поэтому интересует анализ работы генератора с конкретным набором констант с точки зрения длины периода.

Постройте свой генератор с параметрами *a* = R1, *c* = G1, *X0* = B1, *m* = 100 (здесь и далее числовые значения берутся из таблиц исходных данных к первому домашнему заданию). Составьте таблицу элементов последовательности до первого повторения, определите период генератора.

Доказано, что полный период (равный модулю *m*) может быть достигнут при наложении условий на другие параметры генератора:

* числа *c* и *m* взаимно простые (не имеют общих делителей);
* *b* = (*a* – 1) кратно *p* для каждого простого *p*, являющегося делителем *m*;
* *b* кратно 4, если *m* кратно 4.

Очевидно, для *m* = 100, параметр *а* может принимать одно из значений: 21, 41, 61 или 81.

Постройте свой генератор с рационально выбранными параметрами *a* и *c* (согласно таблицам ниже), *X0* = B1, *m* = 100. Составьте таблицу элементов последовательности до первого повторения, убедитесь в достижении максимального периода генератора.

|  |  |  |  |
| --- | --- | --- | --- |
| Группа | Параметр *а* | Группа | Параметр *а* |
| РК6-31Б | 21 | РК6-34Б | 81 |
| РК6-32Б | 41 | РК6-35Б | 21 |
| РК6-33Б | 61 | РК6-36Б | 41 |

|  |  |  |  |
| --- | --- | --- | --- |
| Вариант | Параметр *c* | Вариант | Параметр *c* |
|  | 1 |  | 51 |
|  | 3 |  | 53 |
|  | 7 |  | 57 |
|  | 9 |  | 59 |
|  | 11 |  | 61 |
|  | 13 |  | 63 |
|  | 17 |  | 67 |
|  | 19 |  | 69 |
|  | 21 |  | 71 |
|  | 23 |  | 73 |
|  | 27 |  | 77 |
|  | 29 |  | 79 |
|  | 31 |  | 81 |
|  | 33 |  | 83 |
|  | 37 |  | 87 |
|  | 39 |  | 89 |
|  | 41 |  | 91 |
|  | 43 |  | 93 |
|  | 47 |  | 97 |
|  | 49 |  | 99 |

Период является не единственной характеристикой качества работы генератора. Если рассмотреть последовательность 0, 1, 2, …, *m*-1, то формально она будет иметь максимально возможный период *m*, но очевидно случайной её считать нельзя. Поэтому требуется сохранение свойств распределения (в данном случае – равномерного от 0 до *m*-1) для отдельных частей последовательности в рамках периода.

Для этого возьмите первые *n* = 50 значений из ранее полученной таблицы. Разбейте отрезок [0;99] на *r* = 10 равных частей [0;9], [10;19], …, [90;99]. Определите число элементов усечённой последовательности *ni*, попавших в соответствующий диапазон и постройте гистограмму.

Очевидно, если бы мы имели дело с идеальным генератором, то в каждый из отрезков попало бы по 5 значений. Для реального генератора высота отдельных столбиков будет отличаться. Нужно численно оценить, насколько критичными являются эти отклонения. Для этого воспользуемся методом проверки статистических гипотез на основе критерия Пирсона (также известный как «хи-квадрат», ). Выдвинем гипотезу, что полученная усечённая последовательность представляется собой статистическую выборку, подчиняющуюся равномерному закону распределения. Нужно численно оценить совокупную величину отклонения элементов выборки от теоретически ожидаемых результатов.

Для этого рассчитаем значение коэффициента по *n* = 50 точкам:

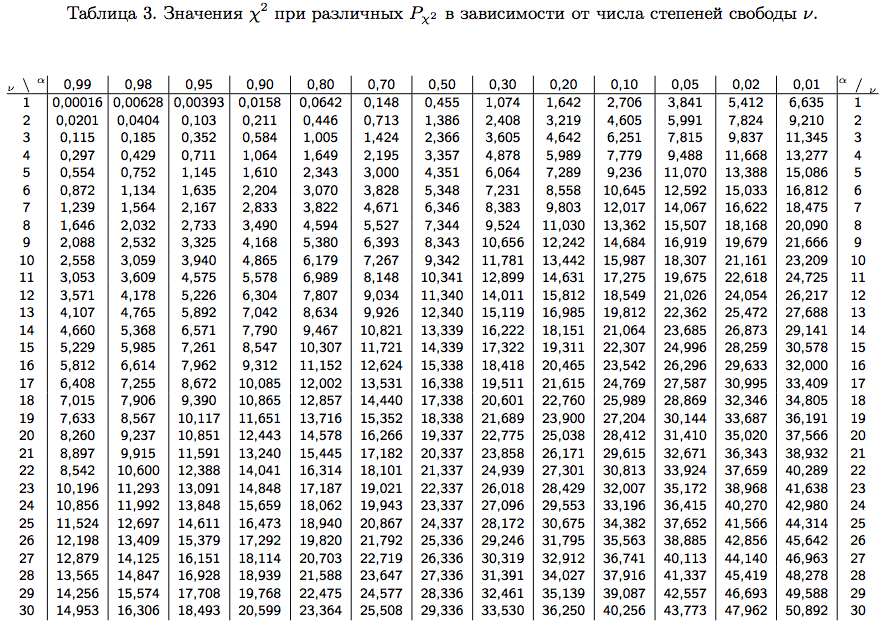
где *pi* – вероятность попадания случайной величины в соответствующий диапазон (численно соответствует площади под графиком плотности распределения для рассматриваемого диапазона).

Для равномерного распределения , и поэтому в рассматриваемой задаче .

Значение рассчитанного коэффициента сравнивается с табличным, при этом учитываются 2 параметра. Так называемое число степеней свободы (обычно обозначается ) в данном случае выбирается как *r*-1. Уровень значимости (обычно обозначается ) определяет степень соответствия (или несоответствия) выдвинутой гипотезе и выражается в долях 1. Если уровень значимости задан, то проверяется, не превышает ли рассчитанное значение коэффициента табличного значения . Если это условие выполняется, то гипотеза принимается, если нет – опровергается.

В нашем случае требуется определить такое значение уровня значимости, с которым можно принять гипотезу о том, что статистическая выборка соответствует равномерному распределению. Полученный уровень значимости можно будет рассматривать как характеристику качества работы генератора случайных чисел, с помощью которого была получена статистическая выборка.

Таблицы критических значений распределения в часто ограничены представлением уровней значимости, близкими к 0 или к 1. Поэтому в рамках решаемой задачи рекомендуется пользоваться расширенным вариантом этой таблицы, в котором представлены и промежуточные значения (приводится ниже).



Оценим также изменение статистических свойств выборки по мере увеличения её размера. Поскольку эти свойства определяются для выборки, то их называют выборочными характеристиками, что позволяет подчеркнуть их отличия от некоторых аналогичных характеристик теоретических распределений.

Выборочное среднее определяется как среднее арифметическое элементов выборки и по смыслу соответствует математическому ожиданию *M* для теоретического распределения.

Характеристики рассеивания обычно произносятся как дисперсия, но записываются в терминах среднеквадратичных отклонений. Выборочная дисперсия определяется как . Нижний индекс *n* в записи подчёркивает, что речь идёт именно о выборочной характеристике. Однако выборочная дисперсия, вычисленная по приведённой формуле, является так называемой смещённой оценкой (ведь считается сумма квадратов отклонений от выборочного среднего, а не от математического ожидания теоретического распределения), поэтому вводится ещё одна величина – исправленная оценка выборочной дисперсии: . Очевидно, по мере увеличения размера выборки, смещённая и исправленная оценки сближаются.

Требуется рассчитать выборочные характеристики (выборочное среднее, смещённую и исправленную оценки выборочной дисперсии) для *n* = 5, 10, 25 и 50 и сравнить их с соответствующими характеристиками теоретического равномерного распределения (математическим ожиданием и дисперсией). Результаты свести в таблицу, с указанием величины отклонений от теоретических значений.

**Задание 2.**

Рассматривается имитационная модель системы массового обслуживания на GPSS. Смоделируем поведение покупателя в магазине, в котором работают 2 кассы, причём к каждой из них выстраивается отдельная очередь, а квалификация сотрудников немного отличается, поэтому время обслуживания распределено с разными параметрами. Все случайные интервалы времени для простоты будем считать равномерно распределёнными (но независимыми, привязанными к разным потокам случайных чисел). Каждая касса будет представлена одноканальным устройством, обращение к которым будем осуществлять по номерам. Очереди также будут идентифицироваться номерами, без введения символьных имён.

Моделирование будем проводить в течение 1 часа, в качестве единицы времени будем выбирать секунду. Время между приходом покупателей распределено на отрезке [0; R1+G1+B1]. Время обслуживания на первой кассе распределено на отрезке [R1; R1+G1+B1]. Время обслуживания на второй кассе распределено на отрезке [G1; R1+G1+B1].

При принятии решения покупатель сперва проверяет, есть ли свободная касса, и, если есть, направляется к ней. Если же обе кассы заняты, то выбирает кассу, очередь к которой в данный момент короче (очередь понимается с бытовой точки зрения, хотя модель можно было бы упростить, если иначе выбрать расположение блоков DEPART). Если же свободны обе кассы, или очередь к ним одинакова, то выбирается первая касса.

Перед описанием модели используем конструкцию EQU (сокращение от слова «эквивалентность») для удобства изменения привязки к потокам случайных чисел. По смыслу она аналогична директиве define препроцессора языка C.

rnd EQU 1

GENERATE (uniform(rnd,0,20))

GATE U 1,metka1

GATE U 2,metka2

TEST LE Q1,Q2,metka2

metka1 QUEUE 1

SEIZE 1

DEPART 1

ADVANCE (uniform(rnd+1,5,20))

RELEASE 1

TERMINATE

metka2 QUEUE 2

SEIZE 2

DEPART 2

ADVANCE (uniform(rnd+2,8,20))

RELEASE 2

TERMINATE

GENERATE 3600

TERMINATE 1

START 1

Требуется провести 100 экспериментов, меняя значение rnd. Результаты моделирования оформляются в виде таблицы, в которой предусматриваются следующие столбцы:

* коэффициент загрузки первого кассира;
* коэффициент загрузки второго кассира;
* средняя длина первой очереди;
* средняя длина второй очереди.

Рассчитайте выборочные средние и исправленные выборочные оценки дисперсии для каждой собранной характеристики при *n* = 10, 25, 50, 100.

На основе полученных выборок для *n* = 100 построить гистограммы. Ширину интервалов выбирать не более половины исправленной оценки среднеквадратичного отклонения соответствующей величины. При попадании в крайние интервалы менее 5 значений объединять их с соседними.

Для каждой пары собранных характеристик рассчитайте выборочные ковариации и коэффициенты корреляции (для значений *n* = 10, 25, 50, 100). Аналогично выборочной дисперсии, используйте исправленные оценки указанных характеристик (т.е. необходимо делить не на *n*, а на *n*-1). Исправленная выборочная ковариация пары случайных величин записывается в виде:

Исправленный выборочный коэффициент корреляции представляет собой отношение исправленной выборочной ковариации к произведению исправленных выборочных среднеквадратичных отклонений:

Для тех же значений *n* = 10, 25, 60 требуется рассчитать доверительные интервалы для математических ожиданий каждой из собранных характеристик с уровнями значимости = 0,1 и 0,01 (для двусторонней симметричной области). Уровень значимости означает вероятность того, что исследуемая величина (в данном случае – математическое ожидание исследуемой характеристики) выйдет за пределы доверительного интервала (в некоторых случаях в статистических таблицах используется не уровень значимости, а уровень надёжности, который может также обозначаться , но чаще используется , при этом они связаны между собой как ; также при работе с таблицами необходимо учитывать, что может быть задана для односторонней области, т.е. уже предварительно разделена пополам). Если предположить нормальный характер распределения исследуемой характеристики, то расчёт строится по аналогии с решением задач на приложения интегральной теоремы Муавра-Лапласа:

где

Однако это возможно лишь при больших *n*. Для большинства практических опытов это не реализуется, поэтому вместо теоретического среднеквадратичного отклонения необходимо использовать исправленную оценку выборочного среднеквадратичного отклонения и вводить в расчёт коэффициент Стьюдента, зависящий от *n*. Он обозначается как , где *k* называется числом степеней свободы распределения Стьюдента, а сам коэффициент определяется по статистическим таблицам, при этом *k* принимается равным *n*-1.

Таблица критических точек распределения Стьюдента, приведена ниже.

