МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«МИРЭА – Российский технологический университет»

РТУ МИРЭА

Методические указания к выполнению практических работ

Моделирование информационно-аналитических систем
Практическая работа 17

(наименование дисциплины (модуля) в соответствии с учебным планом)

Уровень	специалитет *(бакалавриат, магистратура, специалитет)*
Форма обучения	очная *(очная, очно-заочная, заочная)*
Направление(-я) подготовки	10.05.04 Информационно-аналитические системы безопасности, специализации: специализация N 1 "Автоматизация информационно-аналитической деятельности"; специализация N 3 "Технологии информационно-аналитического мониторинга». *(код(-ы) и наименование(-я))*
Институт	Кибербезопасности и цифровых технологий *(полное и краткое наименование)*
Кафедра	Информационно-аналитические системы кибербезопасности (КБ-2) *(полное и краткое наименование кафедры, реализующей дисциплину (модуль))*
Лектор	к.т.н., доцент Лебедев Владимир Владимирович *(сокращенно – ученая степень, ученое звание; полностью – ФИО)*

Используются в данной редакции с учебного года 2022/23
(учебный год цифрами)

Проверено и согласовано «____» ________20___г. ____________ ______________________
(подпись директора Института/Филиала с расшифровкой)

Москва 20__ г.

Практическое занятие №17.

Основы имитационного моделирования случайных процессов

Тема: Генерирование и моделирование случайных чисел

Цель: Изучить основные методы и алгоритмы моделирования случайных чисел.

Назначение: Случайные процессы в математических моделях представляются последовательностями случайных чисел, которые подчиняются заданным законам распределения. Метод имитационного моделирования – это технология моделирования случайных процессов в системах. Чтобы разрабатывать такие модели, или использовать готовые инструменты моделирования, надо понимать основы методов, применяемых в алгоритмах.

Принципы имитационного моделирования случайных процессов

Случайные числа находят применение при математическом моделировании стохастических систем, в поисковых алгоритмах, системах шифрования и алгоритмах управления в системах.

Имитационное моделирование случайных процессов непосредственно связано с моделированием случайных чисел. Случайный процесс в имитационной модели представляет собой последовательность случайных чисел.

Моделирование случайного процесса в имитационной модели стохастической системы часто связано с генерацией некоторого количества случайных процессов, статистические параметры которых известны априорно.

Эти процессы выступают в имитационных моделях как первичные, которые и определяют случайность поведения стохастической системы. Вырабатываемые моделью вторичные случайные процессы представляют интерес как предмет статистического исследования.

Имитационное моделирование систем воспроизводит поведение систем, которое разворачивается в масштабах модельного или реального времени.

Моделируемое поведение исследуемой системы при запуске имитационной модели реализует конкретный сценарий, который соответствует в рамках, установленных целями исследования, множеству возможных сценариев.

Сам статистический эксперимент носит характер статистического испытания, т.к. требует многократного прогона модели с получением отдельного отклика, проявление которого случайно. В результате накапливается множество результатов повторных испытаний, называемое выборкой. Такой метод статистического эксперимента носит название метод Монте-Карло.

Метод Монте-Карло основан на применении генератора псевдослучайных чисел, распределённых по равномерному закону. Обычно применяют генераторы физического, табличного и алгоритмического типов.

Полученные при генерации случайные числа, распределённые по равномерному закону, используют для моделирования случайных чисел, распределённых по другим законам и имеющих другие статистические характеристики.

Различают методы моделирования случайных чисел, распределённых по дискретным или непрерывным законам распределения.

Для чисел или событий, распределённых по дискретным законам распределения используют метод «разыгрывания жребия».

Для моделирования чисел, имеющих непрерывный закон распределения, используют различные методы: метод обратных функций, метод Неймана, метод интервальной гистограммы или аппроксимации, и многие др.

Общая схема имитационной модели по методу статистического моделирования представлена на рис. 1.

На схеме:

ГСЧ – генераторы случайных чисел модели (в методе Монте-Карло это равномерно распределённые числа rрр$\in[0,1]$).

ПЗСЧ – преобразование закона распределения с.ч. (моделирование с.ч.).

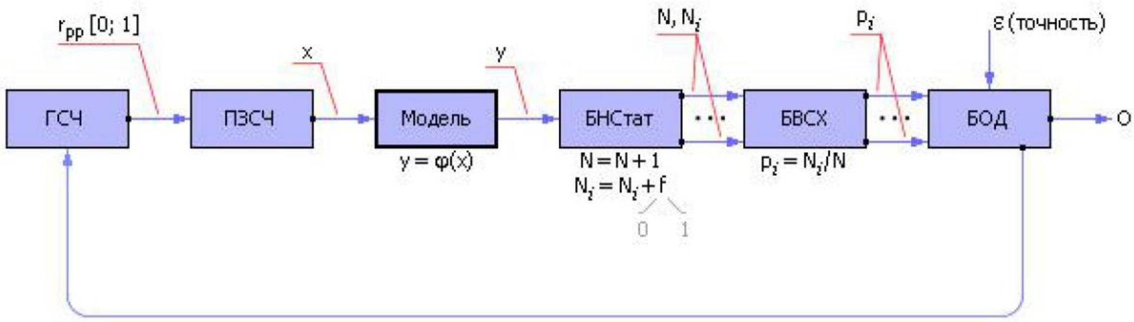


Рисунок 1 Общая схема имитационного статистического моделирования

Модель – модель стохастической системы (имитационная модель случайного процесса).

БНСтат – блок накопления статистических данных.

БВСХ – блок вычисления статистических характеристик.

БОД – блок оценки данных, полученных при использовании модели.

Наблюдаемые в эксперименте значения в зависимости от их полноты характеризуют:

Генеральная совокупность – это совокупность всех мыслимых наблюдений (или всех мысленно возможных объектов интересующего нас типа, с которых «снимаются:» наблюдения), которые могли бы быть произведены при данном реальном комплексе условий.

Выборка из данной генеральной совокупности — это результаты ограниченного ряда наблюдений $\{X_1, X_2, \ldots, X_n\}$ случайной величины ξ.

Статистический ансамбль – это совокупность сколь угодно большого числа одинаковых физических систем многих частиц («копий» данной стохастической системы.

Обработка данных эксперимента по полученной выборке значений сводится к определению выборочных оценок:

- Оценка выборочного (эмпирического) среднего:
- $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$, где N – объём выборки

- Оценка выборочной дисперсии:
- $S^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}$ – это смещённая оценка выборочной дисперсии
- Полученные на разных выборках наблюдений $V_j = \{\dots, X_i, \dots\} | i = 1,2,\dots,N \; j = 1,2,\dots,M$,где $M > 1$ – количество наблюдаемых выборок, множества оценок $\{\dots, \bar{X}_j, \dots\}$ и $\{\dots, \overline{S^2}_j, \dots\}$ представляют собой случайные величины.

Кроме вычисления указанных оценок осуществляют исследование закона распределения частоты наблюдаемых значений в наблюдаемом диапазоне данных. Исследование производят с помощью построения гистограммы, см. рис. 2. Подтверждение соответствия наблюдаемого закона заданному или предполагаемому надо производить на основании проверки статистических гипотез с заданным уровнем доверительной вероятности, например по критерию Пирсона.

Рисунок 2. Гистограмма

Гистогра́мма — способ представления табличных данных в графическом виде — в виде столбчатой диаграммы. Количественные соотношения некоторого показателя представляют в виде прямоугольников, площади которых пропорциональны значениям показателя. Чаще всего для удобства восприятия ширину прямоугольников берут одинаковую, при этом их высота определяет соотношения отображаемого параметра.

В описательной статистике гистограмма распределения — наглядное представление функции плотности вероятности некоторой случайной величины, построенное по выборке. Иногда её называют частотным распределением, так как гистограмма показывает частоту появления измеренных значений параметров объекта, наблюдаемых в случайном процессе. Данное понятие и название для него введены Карлом Пирсоном в 1895 году.

Для построения гистограмм разбивают наблюдаемый интервал случайных значений на равные отрезки (интервалы группирования или карманы) и подсчитывают число наблюдаемых значений, попадающих в карманы: $\{\dots, \mathrm{v_k}, \dots\}|\mathrm{k} = 1,2, \dots \mathrm{K}$, где K – число карманов, см. рис. 3.

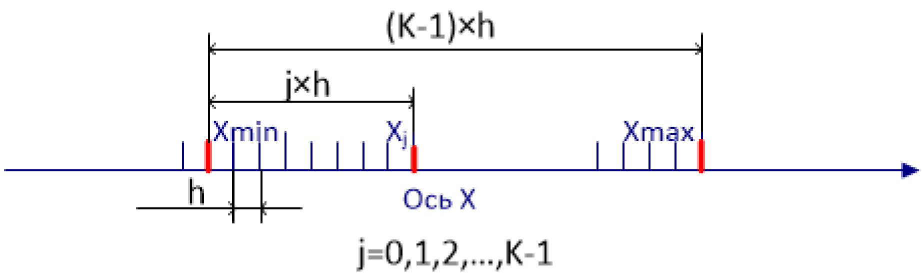


Рисунок 3. Шкала интервалов группирования (карманов) гистограммы

Прежде чем воспользоваться процедурой построения гистограммы необходимо найти границы интервалов группирования (выборки объёмом $\boldsymbol{n}$). Число интервалов группирования k в Excel вычисляется по формуле:

$$\boldsymbol{k} = \lfloor \sqrt{\boldsymbol{n}} + \mathbf{1} \rfloor$$

Для вычисления количества интервалов группирования часто рекомендуют формулу Стерджеса:

$$\boldsymbol{k} = \lfloor \mathbf{1} + \mathbf{log_2} \boldsymbol{n} \rfloor$$

Известна также формула Манна и Вальда:

$$\boldsymbol{k} = \left\lfloor \mathbf{4} \left(\frac{\mathbf{3}}{\mathbf{4}} (\boldsymbol{n} - \mathbf{1})^2 \right)^{\mathbf{1}/\mathbf{5}} \right\rfloor$$

Вычислим интервалы группирования.

Величина интервала группирования вычисляется по формуле:

$$\boldsymbol{h = \frac{x_{max} - x_{min}}{k - 1}}$$

Узлы сетки разбиения диапазона значений выборки (шкалы карманов) определяют по формуле:

$$\boldsymbol{x_j = x_{min} + j \cdot h; \quad j = 0, 1, 2, \dots, k - 1}$$

$$\boldsymbol{x_{max} = x_{min} + (k - 1) \cdot h}$$

Алгоритм вычисления частотной характеристики сканирует выборку значений, проверяет попадание значения в какой-нибудь из интервалов группирования, и подсчитывает число попаданий значений в интервалы:

$$j = 1,2, \dots, k - 1 \colon N_j = 0$$

$$i = 1,2, \dots, n; \;\; j = 1,2, \dots, k - 1 \colon \text{если } x_{j-1} < x_i < x_j \rightarrow N_j = N_j + 1$$

Полученные частоты наблюдения выборочных значений в интервалах группирования $N_j; j = 1,2, \dots, k$ представляют выборочную статистическую частотную характеристику распределения случайной величины. По ним строится столбиковая диаграмма – гистограмма распределения.

При моделировании первичных случайных процессов с заданным законом распределения рассмотренные оценки применяют для проверки адекватности результатов моделирования. При получении на модели вторичных случайных процессов этими методами изучают статистические законы выходных процессов.

Объём выборки влияет на точность выборочных оценок. При увеличении объёма точность приближения оценок к теоретическим значениям растёт. Поэтому для получения заданной точности результатов моделирования устанавливают необходимый объём получаемых данных.

Генерация случайных чисел

Методом генерации случайных чисел получают числа, которые распределены равномерно и непрерывно на интервале значений [0;1]. Эти числа называют базовыми, их часто обозначают символически в виде $\mathrm{r_{pp}}(0; 1)$. Все существующие алгоритмы моделирования случайных чисел опираются на использование базовых случайных чисел, что тесно связано со свойствами статистического распределения этих чисел.

Статистические характеристики закона равномерного непрерывного распределения

Функция плотности непрерывного равномерного распределения в диапазоне $x \in [a, b]$ имеет вид:

$$f(x) = \begin{cases} \dfrac{1}{b-a} \ \forall a \leq x \leq b \\ 0 \text{ вне отрезка} \end{cases}$$

Ниже на рис. 4 представлен график функции плотности.

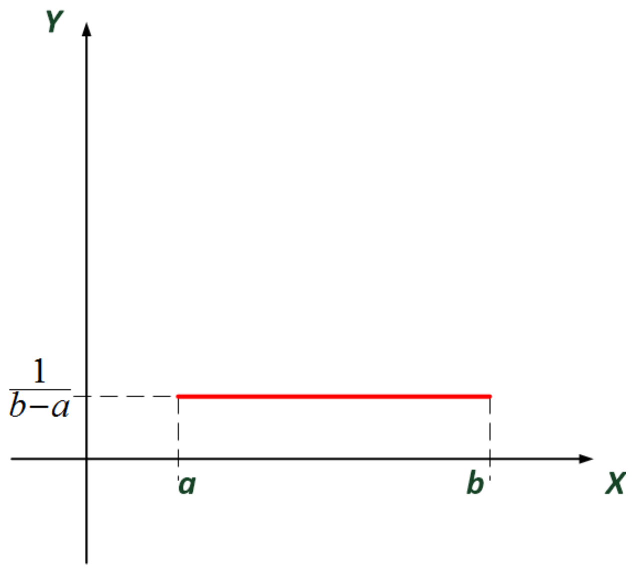


Рисунок 4 Функция плотности равномерного распределения

Равномерное распределение имеет функцию вероятности:

$$\text{PP}(a, b): F(x) = \begin{cases} \displaystyle\int_a^{x<b} f(x)dx = \frac{x-a}{b-a}\ ;\ x \in [a, b] \\ 0; x < a \\ 1; x > b \end{cases}$$

График линейной функции вероятности равномерного распределения мы видим на рис. 5.

Математическое ожидание: $m_x = \frac{a+b}{2}$

Дисперсия: $\sigma_x{}^2 = \frac{(b-a)^2}{12}$

Среднее квадратическое отклонение: $\sigma_x = \frac{b-a}{\sqrt{12}}$

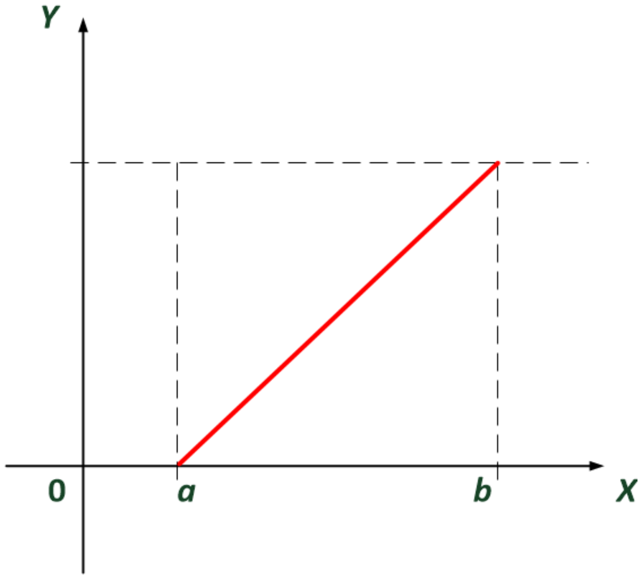


Рисунок 5 Функция вероятности равномерного распределения

ГСЧ должен выдавать близкие к следующим значения выборочных оценок статистических параметров, характерных для равномерного случайного закона для РР СЧ $\in [0,1]$.

Оценка выборочного (эмпирического) среднего:

$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} \approx m_x = \frac{0+1}{2} = 0{,}5$, где N – объём выборки

Оценка выборочной дисперсии:

$S^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1} \approx \sigma_x{}^2 = \frac{(1-0)^2}{12} = 0{,}0833$ – это смещённая оценка

Оценка выборочного среднего квадратического отклонения:

$$S \approx \sigma_x = 0{,}2887$$

В общем случае качество получаемой с помощью ГСЧ выборки (последовательности) случайных чисел на соответствие требуемому закону распределения проверяют с использованием статистических критериев.

Проверка осуществляется на основе доверительной оценки справедливости предложенных статистических гипотез.

Качество генерации псевдослучайных последовательностей проверяется различными тестами на равномерность распределения получаемых чисел, т.е. на соответствие получаемых при генерации выборок закону равномерного непрерывного распределения на отрезке [0,1]. Например, визуальный тест на равномерность должен показывать равномерность покрытия случайными точками (x,y) квадрата 1×1, если $\mathrm{x} = \mathrm{r}_{\mathrm{pp}}(0;1)$ и $y = \mathrm{r}_{\mathrm{pp}}(0;1)$, см. рис. 6.

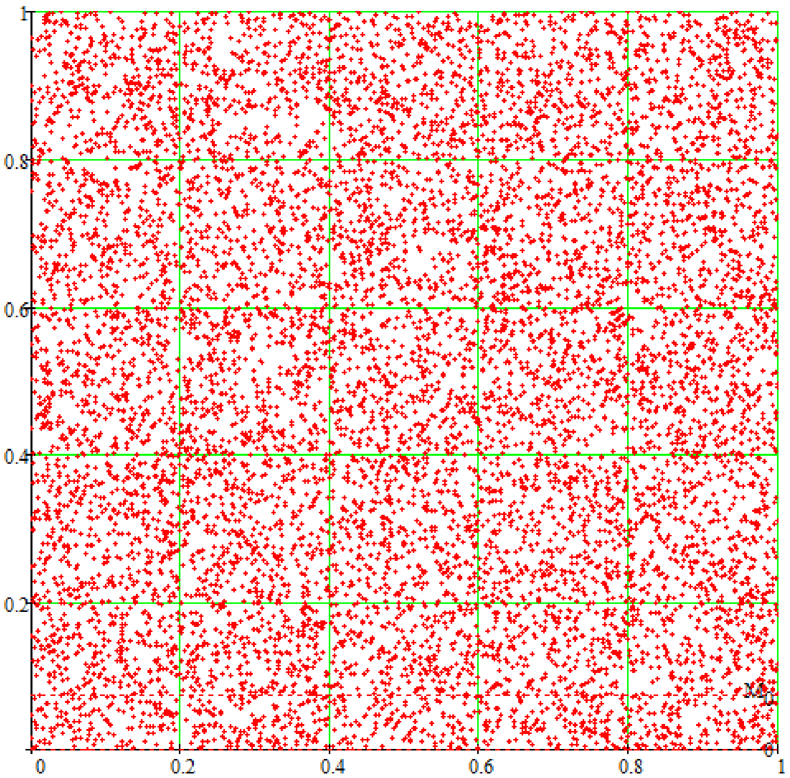


Рисунок 6 Визуальный тест равномерности ГСПЧ генератора MathCad (10000 точек)

Результат, представленный на рис. 6, действительно показывает, что равномерность распределения случайно сгенерированных точек по площади прямоугольника свидетельствует о равномерности генератора. Такое свойство должно быть присуще всем генераторам базовых случайных чисел, применяемым при имитационном моделировании, потому что это гарантирует качество статистических характеристик модели.

Методы генерации базовых случайных чисел

Генераторы случайных чисел по способу получения чисел делятся на:

1) физические (аппаратные);
2) табличные;
3) алгоритмические.

При аппаратном способе случайные числа вырабатываются электронной приставкой (генератор, датчик случайных чисел).

При табличном способе генерации случайные числа оформлены в виде таблицы, которая хранится в оперативной памяти или на внешнем носителе, см. рис. 7. Получение случайных чисел осуществляется выбором цифр при обходе таблицы заданным способом.

Случайные цифры								Равномерно распределенные от 0 до 1 случайные числа
9	2	9	2	0	4	2	6	0.929
9	5	7	3	4	9	0	3	0.204
5	9	1	6	6	5	7	6	0.269
...								...

Рисунок 7 Табличный способ генерации базовых СЧ

Алгоритмические методы генерации псевдослучайных чисел

При применении алгоритмических способов случайные числа генерируются на компьютере по специальным алгоритмам.

Рассмотрим далее некоторые примеры алгоритмических способов генерации псевдослучайных чисел.

Битовые способы

На рис. 8 представлены некоторые алгоритмические битовые способы получения СЧ РР(0,1). Алгоритмические способы генерируют псевдослучайные последовательности, т.к. используют рекуррентные алгоритмы.

Линейный конгруэнтный метод

В линейном конгруэнтным методе (ЛКМ) случайное число вычисляется по следующей рекуррентной формуле: $X_{\mathrm{n}+1} = (aX_{\mathrm{n}} + c)\mathrm{mod}m$, где — m модуль (m>0), — а множитель ($0 \leq a < m$), с — приращение ($0 \leq c < m$), X_0 — начальное значение, которое также иногда называют зерном (от англ. seed) ($0 \leq X_0 < m$). Операция $a\mathrm{mod}m$ означает вычисление остатка от деления целого числа a на целое число m.

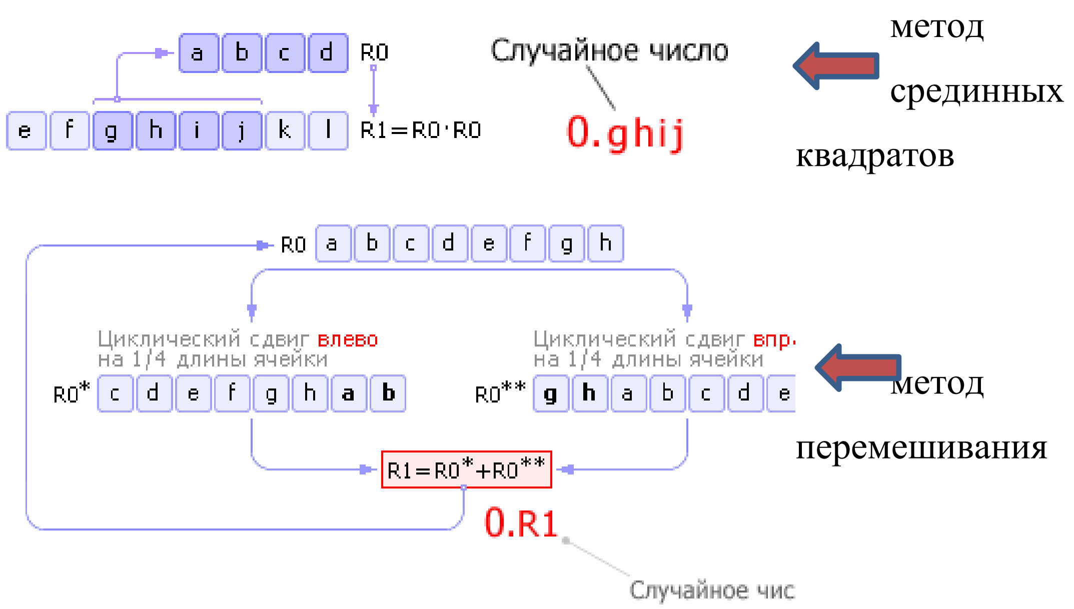


Рис. 8 Алгоритмические битовые способы генерации базовых СЧ

По данной схеме вычисляется последовательность значений остатков целочисленного деления, т.е. множество кольца простых вычетов. Схема определяет псевдослучайную последовательность чисел.

Метод производит при удачно подобранных коэффициентах и больших m большие последовательности псевдослучайных чисел, но не обладает криптографической стойкостью, так как, зная четыре подряд идущих числа, криптоаналитик может составить систему уравнений, из которой можно найти a, c и m

Например, при следующих параметрах алгоритма:

***m*=**	103
***a*=**	53
***c*=**	73
X_0=	23

Вырабатывается последовательность ПСЧ:

№	**X**	**№**	**X**	**№**	**X**	**№**	**X**	**№**	**X**
1	56	22	13	43	100	64	15	85	72
2	54	23	41	44	17	65	44	86	78
3	51	24	83	45	47	66	36	87	87
4	98	25	43	46	92	67	24	88	49
5	14	26	86	47	5	68	6	89	95
-------	---------	---------	---------	-------	---------	----------	----------	----------	----------
17	11	38	25	59	35	80	101	101	1
18	38	39	59	60	74	81	70	102	23
19	27	40	7	61	81	82	75	103	56
20	62	41	32	62	40	83	31	104	54
21	63	42	18	63	30	84	68	105	51

Последовательность имеет период, включающий 103 разных числа.

На диаграмме, рис. 9, показана полученная по ЛКМ последовательность ПСЧ в пределах периода 103.

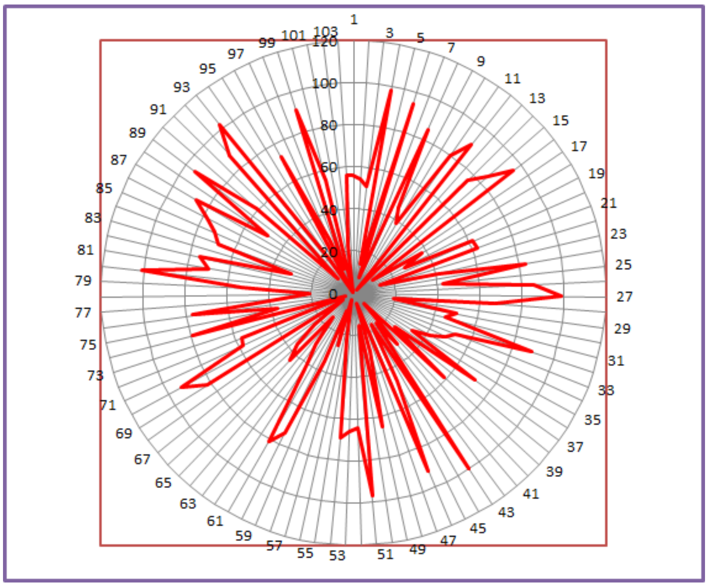


Рис. 9 Диаграмма последовательности ПСЧ по ЛКМ

Гистограмма значений псевдослучайной последовательности от1 –го до 102 представлена ниже на рис. 10.

Выбран ряд неповторяющихся значений, сгенерированных по линейному конгруэнтному алгоритму в пределах периода 103.

Гистограмма показывает, что частотное распределение очень близко к равномерному распределению.

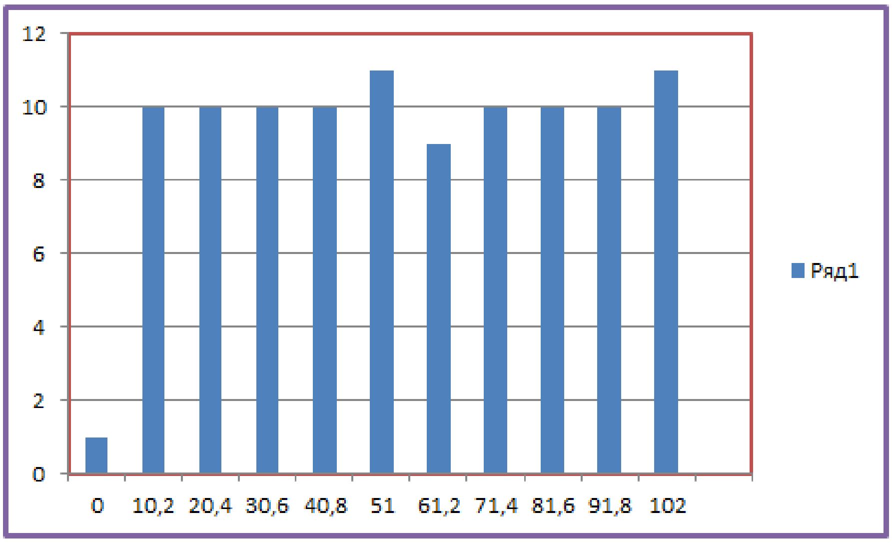


Рис. 10 Гистограмма чисел ряда линейного конгруэнтного генератора

Генератор псевдослучайных чисел на основе алгоритма BBS

Название происходит от англ. Algorithm Blum — Blum — Shub, авт. – Ленор Блюм, Мануэль Блюм и Майкл Шуб, – по русск. алг. Блюма-Блюма-Шуба.

В этом алгоритме вычисляется множество кольца квадратичных вычетов. Выбираются два больших простых числа p и q. Числа p и q должны быть оба сравнимы с 3 по модулю 4, то есть при делении p и q на 4 должен получаться одинаковый остаток 3. Далее вычисляется число M = p · q, называемое целым числом Блюма. Затем выбирается другое случайное целое число x, взаимно простое (то есть не имеющее общих делителей, кроме единицы) с M.

Вычисляем $x_0 = x^2 \cdot \mathrm{mod} M$. x_0 называется стартовым числом генератора.

На каждом $(n+1)$-м шаге работы генератора вычисляется $\mathrm{x}_{\mathrm{n}+1}=\mathrm{x}_{\mathrm{n}}^{2}\cdot \mathrm{mod} M$. Результатом $(n+1)$-го шага является один (обычно младший) бит числа $\mathrm{x}_{\mathrm{n}+1}$.

Иногда в качестве результата принимают бит чётности, то есть количество единиц в двоичном представлении элемента. Если количество единиц в записи числа четное – бит четности принимается равным 0, нечетное – бит четности принимается равным 1

Например, при параметрах:

p	215
q	295
M=pq	63425
x=	317
x_0=	37064

Вырабатывается последовательность ПСЧ:

№	**X**	**№**	**X**	**№**	**X**	**№**	**X**	**ё**	**X**
1	18021	21	38096	41	16571	61	41671	81	25196
2	20441	22	14366	42	31216	62	22591	82	17591
3	54006	23	60431	43	40381	63	35731	83	56131
4	49411	24	21111	44	31836	64	22536	84	52286
5	28396	25	50271	45	62821	65	27321	85	18021
------	------	------	------	------	------	------	------	------	------
16	45111	36	60886	56	661	76	36511	96	40461
17	11196	37	40596	57	56371	77	49896	97	29846
18	22616	38	16	58	33716	78	52716	98	43016
19	24256	39	256	59	2381	79	10281	99	15306
20	23236	40	2111	60	24336	80	32911	100	45111

Последовательность имеет период неповторимости, включающий 84 значения последовательности псевдослучайных чисел.

На диаграмме, рис. 11, показана полученная по BBS последовательность ПСЧ в пределах периода 84.

Гистограмма значений псевдослучайной последовательности генератора BBS от 1 –го до 84 -го представлена ниже на рис. 12.

Выбран ряд неповторяющихся значений последовательности.

Качество частотной равномерности здесь видимо хуже, чем у линейного конгруэнтного генератора.

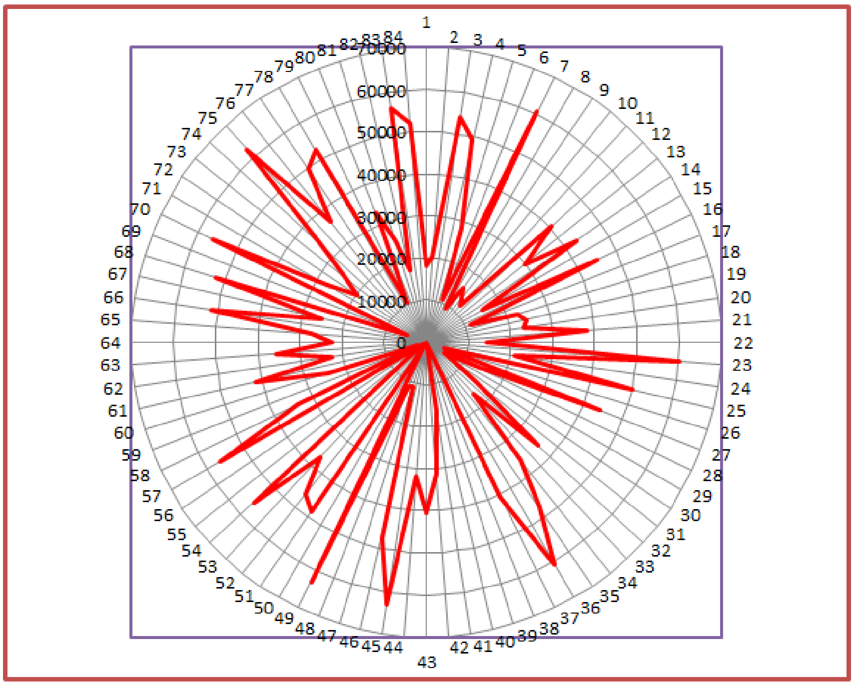


Рис. 11 Диаграмма последовательности ПСЧ по BBS

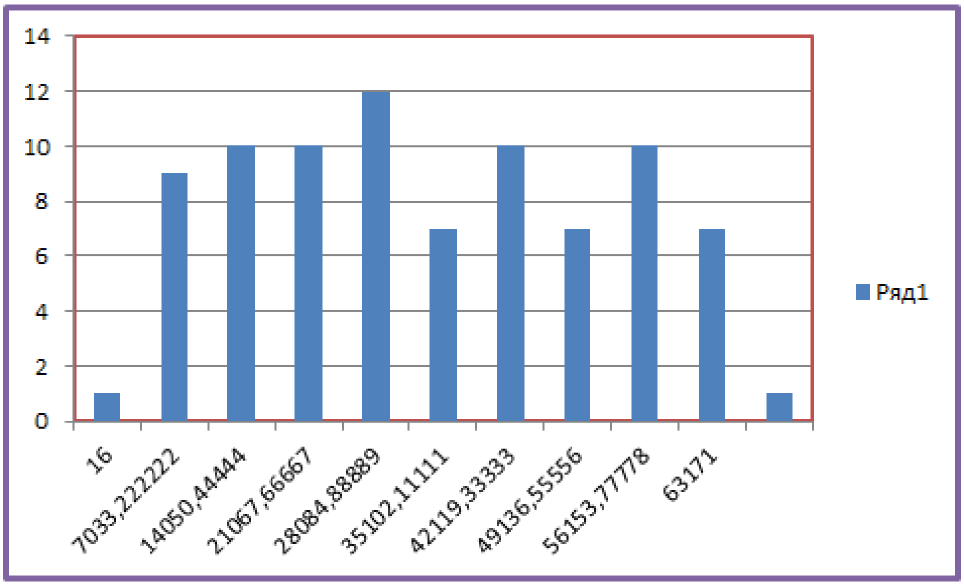


Рис. 12 Гистограмма чисел ряда генератора BBS

Моделирование случайных чисел

Общая схема алгоритма метода моделирования случайных чисел представлена на рис. 13. Она раскрывает алгоритм работы блока ПЗСЧ (преобразование закона распределения случайных чисел), входящего в общую схему алгоритма имитационного статистического моделирования, см. выше рис. 1.

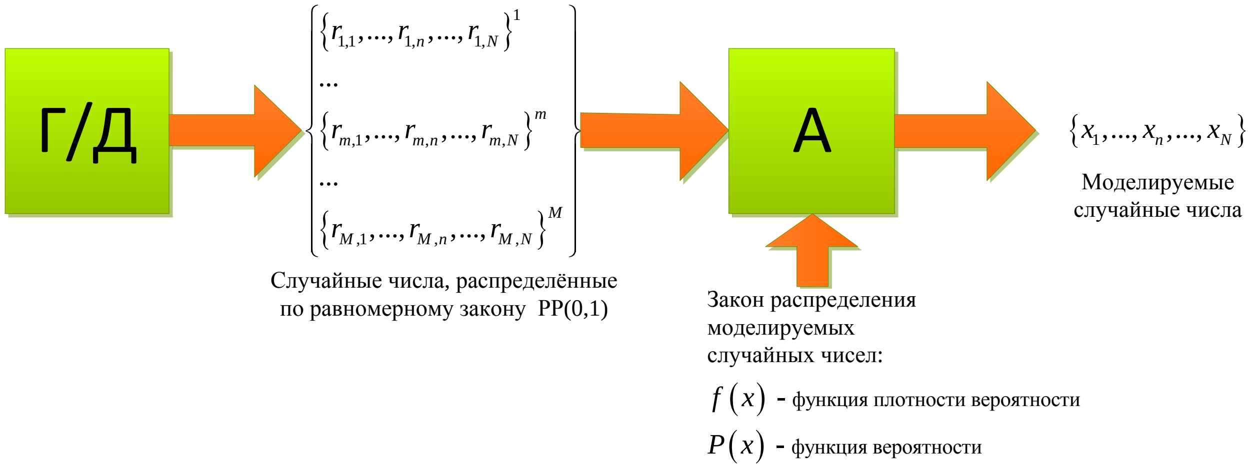


Рисунок 13. Общая схема моделирования случайных чисел

На схеме:

Г/Д – генератор (датчик) случайных чисел, распределённых по равномерному закону(ГПСЧ);

А – алгоритм, воспроизводящий метод моделирования случайных чисел, требуемых при решении задачи.

При моделировании производится преобразование закона распределения.

Любой закон распределения характеризуется множеством значений и частотным распределением их проявления в случайном процессе по множеству.

Различают моделирование с.ч., распределённых по дискретному и по непрерывному закону.

Дискретные распределения характеризуют случайные процессы, множества признаков которых (объектов, событий, состояний, исходов, значений) есть конечные или счётные множества. Частоту явления в процессе отдельных элементов множеств характеризуется заданной вероятностью.

Случайные числа, распределённые по дискретному закону моделируют методом «розыгрыша жребия».

Эту процедуру называют еще «***реализацией жребия***».

Пусть событие А наступает с вероятностью p, тогда процедура моделирования этого события с помощью равномерно распределенных в интервале (0,1) случайных чисел выглядит следующим образом:

1) выбирается очередное сгенерированное равномерно распределённое случайное число r_{ppi}

2) проверкой неравенства $r_{ppi} \le p$ (1)

устанавливается принадлежность этого числа отрезку $[0, p]$.

Если число r_{ppi} удовлетворяет неравенству (1), говорят, что событие А наступило, в противном случае — не наступило.

Общая схема применения алгоритма выбора «жребия» состоит в реализации следующих операций:

1. Разбивают отрезок [0;1] на неперекрывающиеся отрезки, длина которых равна вероятностям элементарных исходов дискретного распределения:

$[0, p_1)$; $\left[p_1, (p_1 + p_2)\right)$; …; $\left[(p_1 + p_2 + \cdots + p_{k-1}), (p_1 + p_2 + \cdots + p_{k-1} + p_k)\right)$; …; $\left[(p_1 + p_2 + \cdots + p_{N-1}), (p_1 + p_2 + \cdots + p_{N-1} + p_N)\right)$, причём $p_1 + p_2 + \cdots + p_{N-1} + p_N = 1$ как сумма всех элементарных исходов данного выбора, где p_1; p_2; …; p_{N-1}; p_N – вероятности дискретного распределения.

Длина каждого отрезка, как легко видеть, равна вероятности выбора:

$[0, p_1) \to p_1$; $\left[p_1, (p_1 + p_2)\right) \to p_2$;…; $\left[(p_1 + p_2 + \cdots + p_{k-1}), (p_1 + p_2 + \cdots + p_{k-1} + p_k)\right) \to p_k$; $\left[(p_1 + p_2 + \cdots + p_{N-1}), (p_1 + p_2 + \cdots + p_{N-1} + p_N)\right) \to p_N$.

2. Генерируют, используя алгоритм приложения, случайное число r, распределённое непрерывно по равномерному закону на отрезке $r_{pp} \in [0; 1]$, т.е. базовое псевдослучайное число, и затем определяют путём перебора, какому отрезку принадлежит это число по схеме:

Цикл по $i \in \{1,2, \dots, N\}$ с шагом 1

$\forall r \in [0, p_1) \Rightarrow$ выбор $i \to 1$;

$$\forall r \in \left[p_1, (p_1 + p_2)\right) \Rightarrow \text{выбор } i \to 2;$$

$$\forall r \in \left[(p_1 + p_2), (p_1 + p_2 + p_3)\right) \Rightarrow \text{выбор } i \to 3;$$

…

$$\forall r \in \left[(p_1 + p_2 + \cdots + p_{k-1}), (p_1 + p_2 + \cdots + p_k)\right) \Rightarrow \text{выбор } i \to k \in \{1,2,\dots,N\};$$

…

$$\forall r \in \left[(p_1 + p_2 + \cdots + p_{N-1}), (p_1 + p_2 + \cdots + p_N)\right) \Rightarrow \text{выбор } i \to N;$$

Выход из цикла перебора происходит по достижении актуального выбора.

Пример результатов численной реализации метода «жребия» представлен в таблице 1.

Таблица 1

Г/Д: [РР(0,1)]		$\{r\} = \{0{,}26; 0{,}28; 0{,}90; 0{,}74; 0{,}44; 0{,}77; 0{,}65; 0{,}80; 0{,}14; 0{,}48\}$		
Закон распределения:	Исход	1	2	3
	Вероятность	0,5	0,2	0,3
Разбиение отрезка [0;1] под выбор и условие выбора по жребию		$0 \le r < 0{,}5$	$0{,}5 \le r < 0{,}5 + 0{,}2 = 0{,}7$	$0{,}7 \le r \le 0{,}7 + 0{,}3 = 1$
А: Метод «жребия»		$\{x\} = \{1; 1; 3; 3; 1; 3; 2; 3; 1; 1\}$		

Разыгрывается 10 исходов случайной реализации процесса в соответствии с заданным в таблице законом дискретного распределения.

Для этого получают (генерируют) с помощью ГПСЧ порождающую последовательность равномерно распределённой величины РР(0,1): $\{r\}$.

Затем, пользуясь методом «жребия», получаем выходную последовательность случайных исходов моделируемого процесса.

Количество исходов связано с числом чисел в $\{r\}$.

Графическая схема алгоритма и разметки отрезка [0,1] для реализации метода жребия для моделирования дискретного распределения, имеющего три исхода, представлены на рис. 14.

Таким образом, разметка отрезка [0;1] вероятностями дискретного распределения задаёт правило «игры» ГПСЧ на разметке при работе алгоритма.

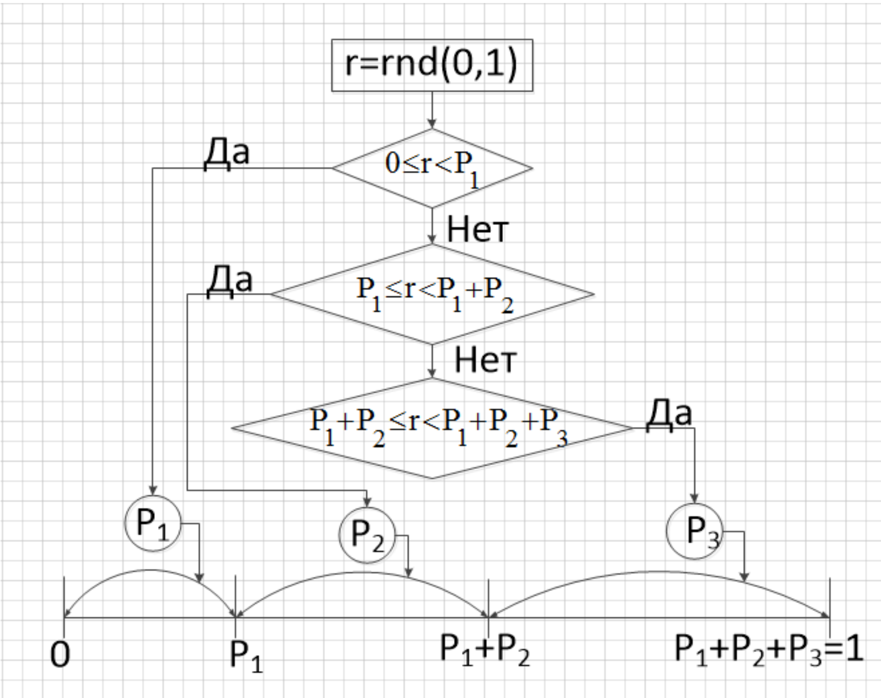


Рисунок 14   Схема алгоритма моделирования методом «жребия»

Перечень методов ***моделирования непрерывных распределений*** более обширен:

I. Универсальные методы:

1) Метод обратных функций
2) Метод ступенчатой аппроксимации
3) Метод усечения Неймана

и др.

II. Специальные методы применяют для моделирования отдельных видов распределений.

Например, для моделирования выборок случайных значений по нормальному закону распределения используют метод ЦПТ и метод Бокса-Мюллера.

Методы моделирования непрерывно распределённых случайных величин, связанных с бесконечным множеством случайных исходов.

Метод обратных функций

Требуется получить случайные числа y_i, являющиеся возможными значениями случайной величины η с законом распределения, заданным функцией плотности $f(y)$ или функцией вероятности $F(y)$:$f(y)$=$dF(y)/dy$; $F(y) = \int_{-\infty}^{y} f(y)dy$. Случайная величина F($\eta$)=$\xi$, являющаяся значением интеграла от плотности распределения $f(y)$: $F(\eta) = \int_{-\infty}^{\eta} f(y)dy = \xi$, и определяющая величину вероятности нахождения значения случайной величины y≤η, распределена равномерно в интервале (0,1). Схему реализации метода см. на рис. 15.

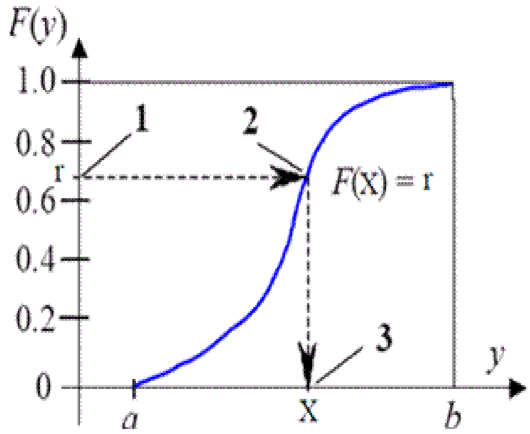


Рисунок 15 Метод обратных функций

Для моделирования этим методом надо получить ***формулу обратной функции***, решив уравнение. **Решение уравнения:** $\int_{-\infty}^{\eta} f(y)dy = F(\eta) = \xi$ аналитически даёт формулу обратной функции $y = \eta = F^{-1}(\xi)$, а численное решение $\eta(\xi)$ задаёт численный алгоритм, где F^{-1} – условное обозначение обратной функции: не путать со степенью (-1), т.е. $F^{-1} \neq \frac{1}{F}$. Корень решения есть искомое, моделируемое по заданному закону число. Основанный на этом решении алгоритм моделирования случайных чисел называется **методом обратных функций**.

Алгоритм метода: генерируем базовое случайное число ГПСЧрр: $r \in [0,1]$; подставляем это число в решение уравнения $r \rightarrow \xi; Y \rightarrow y = \eta = F^{-1}(\xi)$.

Искомыми случайными числами являются $y = \eta = F^{-1}(\xi) \in Y$. Практическое применение метод находит, если имеется аналитическая формула обратной функции.

Пример применения метода к функциям вероятности ряда законов непрерывного распределения приведён на рисунке 16.

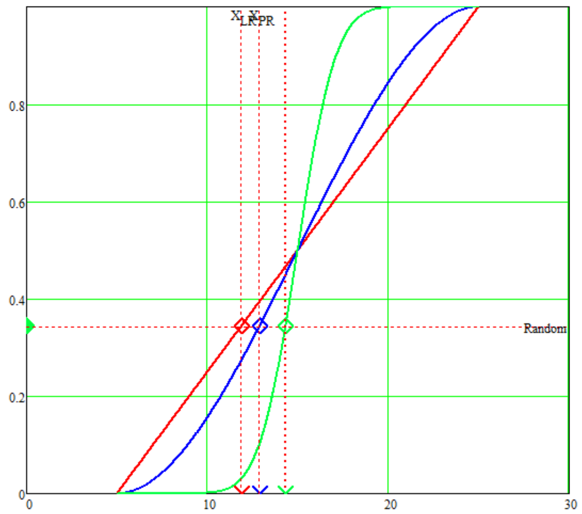


Рисунок 16 Красная линия – график функции вероятности равномерного закона; синяя – график функции параболического закона; зелёная – график функции нормального распределения

Функции вероятности законов непрерывных распределений, представленных на рис. 16 имеют следующие формулы:

$$F(x) = \frac{x-5}{20}$$ – функция равномерного закона распределения на отрезке [5,25];

$$F(x) = \int_5^x \frac{3}{40}\left[1 - \frac{(x-15)^2}{100}\right]dx = -\frac{(x-5)\left(4x^2 - 160x + 700\right)}{16000}$$ – функция параболического закона распределения на отрезке [5,25];

$$F(x)=\frac{1}{\sqrt{2\pi}\left(\frac{5}{3}\right)}\int_{-\infty}^{x} e^{-\frac{9(x-15)^2}{50}}dx$$

– функция вероятности нормального закона, который имеет математическое ожидание – $m=15$ и СКО – $\sigma=\frac{5}{3}$.

На графике случайному значению функций $y=\mathrm{r_{pp}}(0;1)$ соответствуют разные случайные значения аргументов, которые при воспроизведении операции в цикле производят значения генерируемых случайных последовательностей, представляющих выборки значений, распределённых по моделируемым законам.

Метод усечения Неймана

Можно разыграть случайное число, непрерывно распределённое по заданному закону ***методом усечения Неймана***.

Алгоритм метода:

1. Выбираем ***функцию плотности вероятности*** $f(y)$ для описания заданного в моделировании закона распределения.

2. Ограничим интервал распределения (эта процедура называется усечением): сделаем его конечным, если он бесконечен. Отсечение крайних диапазонов производят, если есть возможность пренебречь вероятностями наблюдения с.ч. в них.

Пусть x – случайная величина, распределенная на усечённом интервале $(a,\ b)$ или $[a;b]$, а плотность вероятности сверху ограничена значением в моде M_0.

3. Заключаем функцию на отрезке усечения в прямоугольник $a-V_2-V_3-b-a$, см. рис. 17.

4. Генерируем два числа с помощью ГПСЧ: r_1, r_2 – базовые случайные числа, представляющие значения равномерно-распределенной в интервале [0; 1] случайной величины r.

5. На плоскости $f(y)$ и y отложим случайную точку $(\cdot)N$ с координатами $(\alpha,\ \beta)$: $N(\alpha,\ \beta)$. Случайные значения координат α и β точки моделируем по равномерному закону $\alpha = \text{РР}(a, b)$ и $\beta = \text{РР}(0, M_0)$ методом обратных функций. Формулы обратных функций: координата на оси аргументов (y – моделируемых чисел, горизонтальная) – $\alpha = a + r_1 \cdot (b - a)$ и координата на оси значений функции плотности вероятности (f – вертикальная) – $\beta = r_2 \cdot M0$.

Точка находится внутри прямоугольника $a - V_2 - V_3 - b - a$, см. рис. 17.

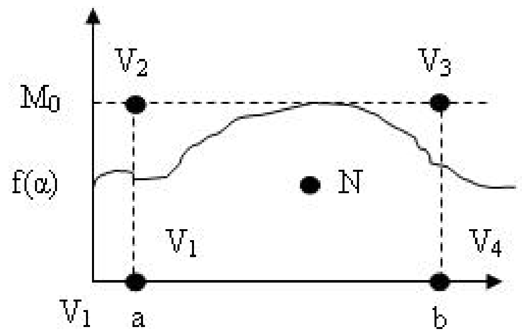


Рисунок 17 Метод усечения Неймана

6. Если точка $(\cdot)N(\alpha,\ \beta)$ лежит под кривой, т.е. $\beta<f(\alpha)$, то разыгранное случайное значение x считается равным α: $X \to \alpha$.

Если точка $(\cdot)N(\alpha,\ \beta)$ лежит над кривой, при $\beta>f(\alpha)$, то пара r_1 и r_2 отбрасывается и генерируется новая пара значений r_3 и r_4, и т.д.

Пример применения генерации случайных точек по методу Неймана применительно к функции плотности параболического закона непрерывного распределения на отрезке [a=5,b=20] мы видим на рис. 18.

Формула функции плотности параболического закона распределения, график которой виден на рис. 18 (красная линия):

$$f\left(x,a,b,m=\frac{a+b}{2}\right)=\frac{3}{2(b-a)}\left[1-4\left(\frac{x-m}{b-a}\right)^2\right]$$

На рисунке 18 кроме графика функции изображена выборка множества случайных точек, сгенерированных по равномерному закону внутри прямоугольника, охватывающего график функции плотности параболического закона. При моделировании производится отбор точек, которые попадают под кривую (подмножество синих точек).

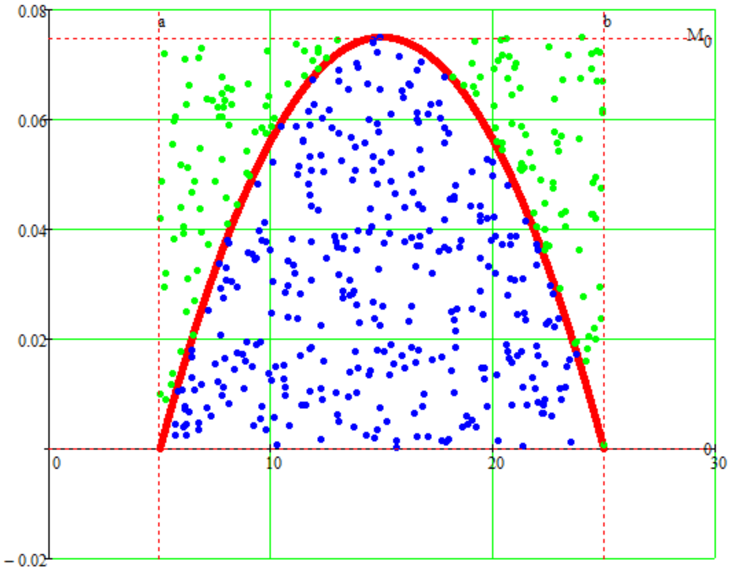


Рисунок 18 Сгенерировано 500 точек:
синие – точки, попавшие в область прямоугольника под кривую;
зелёные – точки, попавшие в область прямоугольника над кривой

Вероятность появления таких точек соответствует данному закону распределения, поэтому их X-координаты из отрезка [5;25] формируют моделируемую выборку случайных значений процесса. Алгоритм имеет значительную долю промахов (появление зелёных точек), что повышает его трудоёмкость, или, как ещё говорят, вычислительную сложность, при реализации имитационного моделирования.

Метод ***ступенчатой аппроксимации***.

Схема метода показана на рис. 19.

Метод основан на комбинированном применении двух уже знакомых нам алгоритмов:

1. Метод жребия используют для выбора интервала. Разметка метода производится по ступенчатой гистограмме.

2. В выбранном интервале случайное число определяют методом обратных функций по равномерному распределению: $\eta = x_i + r(x_{i+1} - x_i)$: $r =$ ГПСЧрр(0,1).

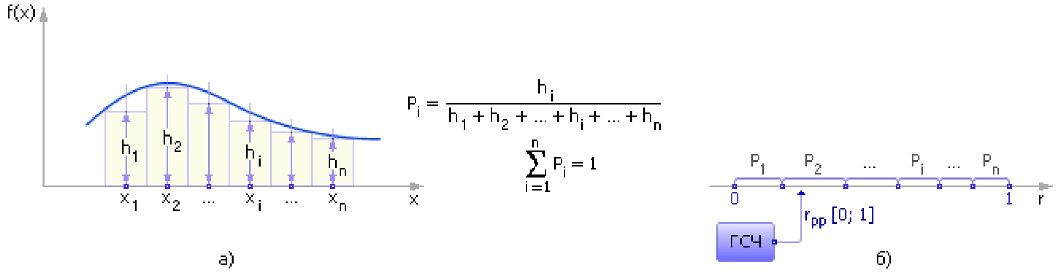


Рисунок 19 Метод ступенчатой аппроксимации по гистограмме

В ряде случаев случайные числа, распределенные по некоторым видам непрерывных распределений, моделируют, используя не общие, а специальные алгоритмы.

Для примера рассмотрим два способа эффективного моделирования с.ч., распределённых непрерывно по нормальному закону.

Моделирование случайных чисел, распределённых по нормальному закону

Методы моделирования случайных чисел, распределённых по нормальному закону:

1) Метод генерации нормально распределенных чисел, использующий ***центральную предельную теорему***.
2) Метод ***Бокса-Мюллера***.

Плотность нормального распределения (функция Гаусса $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$) представлена на рис. 20.

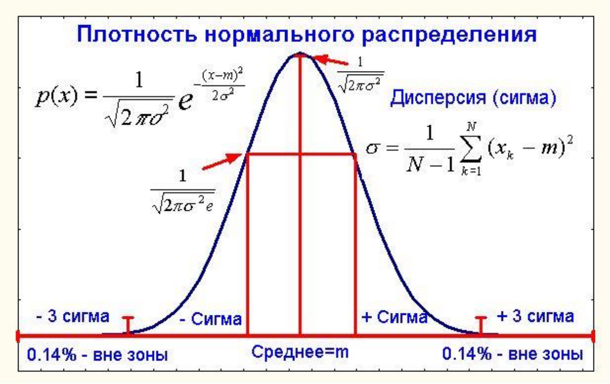


Рисунок 20 Функция плотности вероятности нормального распределения

Метод моделирования нормально распределенных чисел, использующий ***центральную предельную теорему*** работает по алгоритму:

Сложим n случайных чисел, используя стандартный ГПСЧрр [0; 1]: $v = \sum_{i=1}^{n} r_i$.

Согласно ЦПТ при сложении происходит преобразование закона распределения: числа v образуют случайные значения, распределенные по нормальному закону: $v \in V = N(m_V, \sigma_V)$.

Заметим, что закон распределения чисел V имеет математическое ожидание $m_V = n/2$, $\sigma_V = \sqrt{\frac{n}{12}}$.

Эти числа тем лучше описывают нормальный закон, чем больше параметр n. На практике n берут равными 6 или 12. Поэтому для моделирования заданного распределения с параметрами m_x и σ_x делают две операции:

1. С помощью формулы $z=(v-m_V)/\sigma_V$ нормализуем этот ряд. Получим нормализованный закон нормального распределения чисел Z. То есть $m_z=0$, $\sigma_z=1$.

2. Формулой (сдвиг на m_x и масштабирование на σ_x) преобразуем ряд Z в ряд X: $x=z\cdot\sigma_x+m_x$.

Метод Бокса-Мюллера использует формулы: $Z = \sqrt{-2Ln(r_1)}\, cos(2\pi r_2)$, где r_1 и r_2 — случайные числа из ГСЧрр [0; 1].

$Z = \sqrt{-2Ln(r_1\ )}\, sin(2\pi r_2)$, где r_1 и r_2 — случайные числа из ГПСЧрр [0; 1].

Числа Z – распределены непрерывно по нормализованному нормальному закону распределения.

Формулой (сдвиг на m_x и масштабирование на σ_x) преобразуем ряд Z в ряд X: $x = z \cdot \sigma_x + m_x$.

Оценка качества получаемых случайных чисел **на соответствие заданным законам распределения** производится путем **определения выборочных оценок** и последующего их **сравнения** с **теоретическими параметрами распределения**.

Выборочные оценки:

Среднее выборочное: $\bar{x} = \frac{\sum_i x_i}{n}$

Выборочная дисперсия:$S_n^2 = \frac{\sum_i (x_i - \bar{x})^2}{n} = \frac{\sum_i {x_i}^2}{n} - \bar{x}^2$

Смещенная выборочная дисперсия (уточненная оценка):

$$S_{n-1}^2 = \frac{n}{n-1} S_n^2$$

Выборочное среднее квадратическое отклонение: S_{n-1}

Кроме вышерассмотренных оценок необходимо проводить **исследование распределения частот наблюдаемых значений по интервалу наблюдаемых значений**. Это исследование можно проводить с помощью построения гистограмм, или проверкой статистических гипотез, например, по критерию Пирсона.

Задания

Пять заданий представлены ниже.

Требуется результаты оформить в формате документа:

1). Описать методические основы

2). Представить условие задачи

3). Описать шаги решения

4). Представить результаты

5). Сделать выводы

Задание 1:

Смоделировать выборку (последовательность не менее 20-ти) равномерно распределённых случайных чисел – СЧ РР(0,1) и чисел, распределённых на отрезке $[a,b]$, методом обратных функций.

Функция вероятности закона равномерного распределения:

$$F(x) = \frac{x-a}{b-a} \ ; \quad x \in [a,b]$$

Формула обратной функции равномерного распределения РР(a,b):

$$x = a + r \cdot (b-a)$$

Математическое ожидание: $m_{\text{x}} = \frac{a+b}{2}$

Дисперсия: ${\sigma_{\text{x}}}^2 = \frac{(b-a)^2}{12}$

Проверить качество полученного распределения, сравнив выборочные статистические характеристики с теоретическими параметрами.

Задание 2:

Промоделировать методом ЦПТ и Бокса-Мюллера случайные числа (последовательность не менее 20-ти), распределённые непрерывно по нормальному закону, используя компьютерный ГПСЧрр(0,1) и заданные значения m_{x} и σ_{x} (значения выбираем произвольно).

Сравнить выборочные статистические характеристики с теоретическими характеристиками распределения.

Задание 3:

Смоделировать методом усечения – Неймана случайные числа (последовательность не менее 20-ти) по непрерывному закону распределения, заданного в ограниченной области $y \in [a,b]$ плотностью распределения:

$f(y) = \frac{6}{4\cdot(b-a)} \cdot \left(1 - 4 \cdot \left(\frac{y-m}{b-a}\right)^2\right)$

Задать числа a и b, вычислить $m = \frac{a+b}{2}$.

Математическое ожидание: $m_y = m$

Дисперсия: ${\sigma_y}^2 = \overline{y^2} - {m_y}^2$, где средний квадрат случайной величины:

$$\overline{y^2} = \frac{20m^2 + (b-a)^2}{20}$$

Сравнить выборочные статистические характеристики с теоретическими параметрами закона распределения.

<u>Задание 4:</u>

Смоделировать методом обратных функций случайные числа (последовательность не менее 20-ти), распределённые непрерывно по закону Рэлея.

Распределение Рэлея: Функция плотности и функция вероятности, см. рис. 21 ниже.

Функция плотности вероятности распределения Рэлея:

$f(x) = \frac{x}{\sigma^2} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \forall x \in [0,\infty)$. Параметр σ выбираем произвольно (например, берем с графика)

Функция вероятности распределения Рэлея:

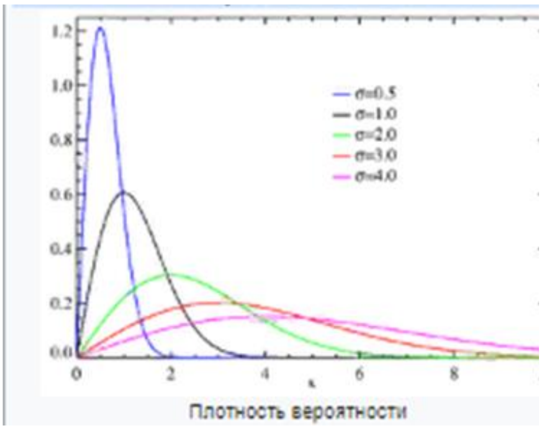
$$F(X < x) = \int_0^x f(x)dx = 1 - e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \in [0,1] \forall x \in [0,\infty)$$

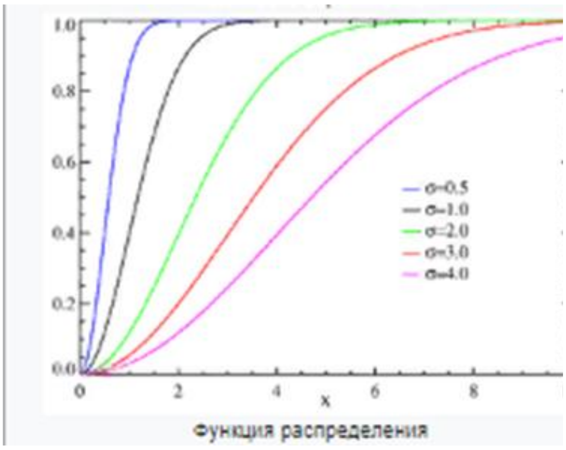
Математическое ожидание: $m_x = \sqrt{\frac{\pi}{2}}\sigma$

Дисперсия: ${\sigma_x}^2 = \left(2 - \frac{\pi}{2}\right)\sigma^2$

Формула обратной функции закона распределения Рэлея:

$$x = \sigma\sqrt{2\ln\left(\frac{1}{1-r}\right)}$$





Параметры	$\sigma > 0$
Носитель	$x \in [0; \infty)$
Плотность вероятности	$\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$
Функция распределения	$1 - \exp\left(\frac{-x^2}{2\sigma^2}\right)$
Математическое ожидание	$\sqrt{\frac{\pi}{2}}\sigma$
Медиана	$\sigma\sqrt{\ln(4)}$
Мода	σ
Дисперсия	$(2 - \pi/2)\sigma^2$

Коэффициент асимметрии	$\frac{2\sqrt{\pi}(\pi - 3)}{(4-\pi)^{3/2}}$
Коэффициент эксцесса	$-\frac{6\pi^2 - 24\pi + 16}{(4-\pi)^2}$
Дифференциальная энтропия	$1 + \ln\left(\frac{\sigma}{\sqrt{2}}\right) + \frac{\gamma}{2}$
Производящая функция моментов	$1 + \sigma t\, e^{\sigma^2 t^2/2}\sqrt{\frac{\pi}{2}}\left(\mathrm{erf}\left(\frac{\sigma t}{\sqrt{2}}\right) + \right.$
Характеристическая функция	$1 - \sigma t\, e^{-\sigma^2 t^2/2}\sqrt{\frac{\pi}{2}}\left(\mathrm{erfi}\left(\frac{\sigma t}{\sqrt{2}}\right) - \right.$

Рисунок 21 Распределение Рэлея

Сравнить выборочные статистические характеристики с теоретическими характеристиками распределения.

<u>Задание 5:</u>

Смоделировать методом «розыгрыша жребия» случайные числа (последовательность не менее 20-ти), распределённые по дискретному геометрическому закону.

Пояснения к выполнению задания:

Член последовательности распределения вероятностей по геометрическому закону описывается по формуле:

$P_k = q^{k-1}p \; \forall k = 1,2,\dots,n,\dots \in [1,\infty)$, где

$q, p \in [0; 1]$ – параметры распределения, причём $q + p$=1, а

$$\sum_{k=1}^{\infty} P_k = 1$$

Математическое ожидание $m_K = \frac{1}{p}$

Дисперсия ${\sigma_K}^2 = \frac{q}{p^2}$

При моделировании выбора жребием случайного k-го исхода, соответствующего распределению согласно заданным вероятностям P_k, производим разметку отрезка [0,1], откладывая слева направо от точки 0 по порядку $k = 1,2, \dots, n, \dots \in [1,\infty)$ отрезки длиной P_k. Так как количество отрезков в этой задаче бесконечно, то разметку мы производим динамически в численном алгоритме.

Вероятности P_k при увеличении k сходятся к 0: $\lim_{k\to\infty} P_k = 0$. Поэтому отрезки длиной P_k, откладываемые по порядку слева направо также по длине неуклонно уменьшаются до 0. Частичная сумма: $\sum_{j=1}^{k} P_j = P_1 + P_2 + \cdots + P_k = 1 - q^k$ – используется при разметке слева направо интервала $(0;1)$ в методе жребия. Сумма $\sum_{j=1}^{k} P_j$ – есть левая граница k-го отрезка, причём $\lim_{k\to\infty} \sum_{j=1}^{k} P_j = \lim\limits_{k\to\infty}(P_1 + P_2 + \cdots + P_k) = \lim\limits_{k\to\infty}(1 - q^k) = 1$, т.е. левая граница последнего отрезка в пределе при $k = \infty$ равна 1. Поэтому все отрезки полностью без перекрытий покрывают при разметке весь отрезок [0,1]. Границы k-го отрезка – $[1 - q^{k-1}; 1 - q^k]$. При моделировании методом жребия проверяем появление базового случайного числа $r = PP(0;1)$ в каком-нибудь k-ом отрезке разбиения по условию: $1 - q^{k-1} < r < 1 - q^k$. При выполнении этого условия случайным исходом будет k.

Вычислить выборочные характеристики и сравнить с теоретическими.