СОДЕРЖАНИЕ

МОДЕЛИРОВАНИЕ И АНАЛИЗ ПАРАМЕТРОВ СТОХАСТИЧЕСКИХ МОДЕЛЕЙ 2

Моделирование параметров стохастических систем 2
Теоретические основы оценки стохастических параметров 3
Оценка вероятности случайного события *4*
Оценка математического ожидания случайной величины *4*
Центральная предельная теорема *5*
Статистический метод определения числа π 6

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН 10

Генераторы случайных чисел 13
Физические генераторы случайных чисел *13*
Табличные генераторы случайных чисел *15*
Алгоритмические генераторы случайных чисел *16*
Алгоритм середины квадрата Неймана 18
Линейный конгруэнтный метод 19
Моделирование дискретной случайной величины 20
Моделирование непрерывной равномерно распределённой случайной величины 21
Моделирование непрерывной случайной величины с заданным законом распределения 22
Метод обратной функции *22*
Моделирования гауссовой случайной величины *24*
Метод ступенчатой аппроксимации *24*
Метод усечения *26*

МОДЕЛИРОВАНИЕ И АНАЛИЗ ПАРАМЕТРОВ СТОХАСТИЧЕСКИХ МОДЕЛЕЙ

Моделирование параметров стохастических систем

Стохастические модели характеризуются наличием в них параметров, принимающих случайные значения. Применение аналитических методов для моделирования сложных стохастических систем обычно сталкивается с серьёзными трудностями адекватного формального выражения вероятностных характеристик стохастических параметров модели M_S. Как правило, для использования аналитических методов моделирования и анализа сложных стохастических систем приходится при моделировании отказываться от учёта значительного количества "несущественных деталей", представленных в семантической модели. Например, получение уравнений Колмогорова для анализа вероятностей состояний стохастической модели M_S, оказывается возможным, если допустить, что моделируемая система представляется марковским стохастическим автоматом, а события перехода автомата из одного состояния в другое описываются простейшими потоки. При этом для получения с помощью таких моделей результатов анализа функционирования систем близких к реальным такое допущение не является убедительным. В связи с этим, количественные результаты анализа характеристик стохастических систем, полученные формальными методами с использованием упрощённых моделей, допускают лишь качественную интерпретацию полученных на модели значений стохастических параметров, но не позволяют использовать их как репрезентативную[1] оценку характеристик реальной системы. Поэтому для получения реальных характеристик предмета моделирования S с необходимой репрезентативностью стохастическую модель M_S приходится «усложнять», стремясь полнее учесть и отразить в ней реальные свойства предмета моделирования. В результате сложность модели обычно возрастает настолько, что использование классических аналитических или численных методов анализа характеристик таких моделей, базирующихся на анализе марковских цепей, становится невозможным. Выходом из этого является

[1] **Репрезентати́вный** - дающий объективное представление о чём-либо.

построение адекватных моделей в виде компьютерных моделей, максимально отражающих их структурные особенности и особенности функционирования. Модели сложных стохастических автоматов реализуются в виде программ, компьютерное выполнение которых имитирует их поведение во времени в виде фазовой траектории перехода модели из одного состояния в другое. С такими моделями осуществляются компьютерные эксперименты, в процессе выполнения которых осуществляется сбор необходимого статистического материала для анализа параметров стохастических моделей *статистическими методами*.

Для создания адекватных моделей поведения стохастических систем важным является возможность формировать (генерировать) *валидные*[2] значения стохастических параметров модели с репрезентативными законами распределения для экзогенных параметров. При этом наборы получаемых случайных значений для экзогенных параметров должны быть достаточно многочисленными, чтобы иметь возможность статистически оценивать значения эндогенных параметров с необходимой степенью точности. В связи с этим наличие эффективных средств генерации случайных чисел с требуемыми законами распределения является важным аспектом эффективного применения статистических методов анализа параметров имитационных стохастических моделей.

Теоретические основы оценки стохастических параметров

Оценка параметров стохастических систем осуществляется с использованием статистических методов получения вероятностных характеристик параметров стохастической системы на основе статистического материала, получаемого в результате экспериментов с моделью M_S, как с абстрактным *стохастическим автоматом*, играющим роль абстрактного оператора, связывающего эндогенные параметры M_S с экзогенными. Эксперименты с M_S выражаются в многократном циклическом выполнении во времени модели M_S и накоплении статистического материала для эндогенных параметров. В итоге

[2] **Валидность** (от англ. valid) - действительный, пригодный, имеющий силу.

накопленный экспериментальный материал обрабатывается статистическими методами анализа.

Теоретической основой статистических методов анализа стохастических параметров являются фундаментальные результаты теории вероятностей и математической статистики, известные как *предельные теоремы теории вероятностей* или *законы больших чисел*.

Оценка вероятности случайного события

Статистический метод оценки вероятности случайного события базируется на *теореме Бернулли*.

Теорема Бернулли. Пусть производится N независимых испытаний, в каждом из которых с вероятностью $p(A)$ может наступить событие A, тогда $\forall \varepsilon > 0$ справедливо равенство:

$$\lim_{N\to\infty} P\{\left|\frac{m}{N} - p(A)\right| < \varepsilon\} = 1,$$

где $\frac{m}{N}$ - относительная частота события *A*. Иными словами, относительная частота события *A* по вероятности стремится к вероятности события *A*.

Оценка математического ожидания случайной величины

Статистический метод оценки математического ожидания случайной величины базируется на *теореме Чебышёва*.

Теорема Чебышёва. Пусть x_1, x_2, x_3, ... – попарно независимые одинаково распределённые случайные величины с математическим ожиданием a и конечной дисперсией $\sigma^2 < \infty$. Тогда $\forall \varepsilon > 0$ справедливо равенство:

$$\lim_{N\to\infty} P\{\left|\frac{1}{N}\sum_{i=1}^{N} x_i - a\right| < \varepsilon\} = 1.$$

Иными словами, среднее арифметическое большого числа случайных слагаемых "стабилизируется" с ростом этого числа. Как бы сильно каждая случайная величина не отклонялась от своего среднего значения, при суммировании эти отклонения "взаимно компенсируются" так, что среднее арифметическое приближается к постоянной величине.

Центральная предельная теорема

Статистический метод оценки закона распределения случайной величины, значение которой формируется как сумма попарно независимых одинаково распределённых случайных величин, базируется на *центральной предельной теореме*.

Если x_1, x_2, x_3, ... – неограниченный набор попарно независимых одинаково распределённых случайных величин, имеющих математическое ожидание a и дисперсию σ^2 , то закон распределения случайной величины

$$X_N = \frac{1}{N}\sum_{i=1}^{N} x_i \, , N \to \infty;$$

неограниченно приближается к *нормальному закону* распределения с математическим ожиданием a и дисперсией $\frac{\sigma^2}{N}$. Соответственно нормализованная случайная величина

$$X_N = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i - a}{\frac{\sigma}{\sqrt{N}}}$$

стремится к *стандартному нормальному закону распределения* - **N**(0,1), с математическим ожиданием 0 и дисперсией 1.

Предельные теоремы выражают объективную закономерность, определяющую формирование значений случайных событий. Если количество N реализаций наблюдаемого случайного значения достаточно велико, то его вероятностные характеристики приобретают статистическую устойчивость, отражающие закон его формирования. Поэтому полученные на их основе результаты моделирования могут быть приняты в качестве оценок искомых характеристик с некоторой вероятностной точностью. Иными словами, случайные явления (события, величины) подчиняются определённым закономерностям, выявление которых позволяет не только прогнозировать их поведение, но и количественно оценивать некоторые средние их характеристики, проявляющие определённую устойчивость. Выражением этих закономерностей и устойчивости средних показателей и являются законы больших чисел или предельные теоремы

теории вероятностей. Принципиальное практическое значение предельных теорем состоит в том, что они гарантируют при весьма большом числе испытаний (реализаций) N приемлемую валидность статистических оценок. На практике, при компьютерной реализации стохастической модели, часто валидные оценки искомых характеристик могут быть получены при сравнительно "небольших количествах испытаний" N. Тем не менее, необходимо учитывать тот факт, что увеличение вероятностной точности результатов стохастического моделирования связанно со значительным увеличением необходимого объёма вычислений. Примерно это выглядит так, что для увеличения вероятностной точности в 10 раз необходимо увеличить число экспериментов с моделью в 100 раз. В связи с этим эффективность практического использования имитационных стохастических моделей принципиально связана с возможностью компьютерной реализации моделирующего алгоритма и статистическими методами анализа стохастических параметров.

При исследовании характеристик стохастических моделей M_S статистическими методами анализа можно выделить следующие основные этапы:

- многократное генерирование случайных значений экзогенных стохастических параметров модели M_S с заданными функциями распределения;
- преобразование моделью M_S полученных случайных значений экзогенных параметров модели в эндогенные и накопление статистического материала для оценки необходимых характеристик;
- обработка накопленного статистического материала методами математической статистики для получения по каждой из анализируемых характеристик устойчивых значений.

Статистический метод определения числа π

В качестве примера применения статистического метода анализа детерминированных характеристик стохастической модели может служить способ

определения числа π, который был предложен Бюффоном[3] ещё в 1777 году. Он придумал натурную модель реализации случайного события, вероятность которого теоретически выражалась через число π, и, экспериментируя с моделью, статистически оценил эту вероятность, основываясь на теореме Бернулли. В результате ему удалось статистическим методом приближённо с вероятностной точностью оценить значение числа π.

Натурная модель реализации случайного события выражалась в многократном бросании иглы длиной l на плоскость, расчерченную параллельными прямыми линиями, расположенными друг от друга на расстоянии, равном длине иглы (рисунок 1).

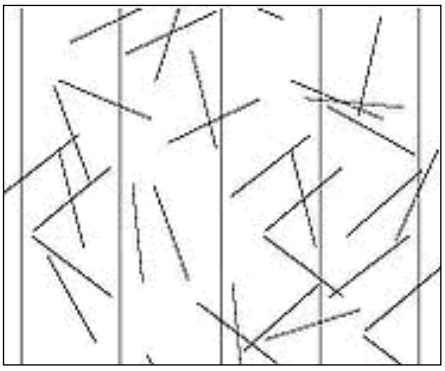


Рисунок 1

В качестве случайного события A рассматривается событие, когда брошенная на плоскость игла случайно пересекает любую из параллельных линий на плоскости.

По теореме Бернулли вероятность $P\{A\}$ события A статистически оценивается по формуле:

$$P\{A\} \approx \frac{m}{N},$$

где m - число реализаций события A, N - общее число бросаний иглы (число испытаний).

Чтобы теоретически выразить вероятность события пересечения иглой любой из параллельных линий, найдём множество всех возможных комбинаций пересечения иглой одной из параллельных линий (рисунок 2).

[3] **Жорж-Луи Леклéрк, граф де Бюффóн (**1707 — 1788) — французский натуралист, биолог, математик, естествоиспытатель и писатель XVIII века.

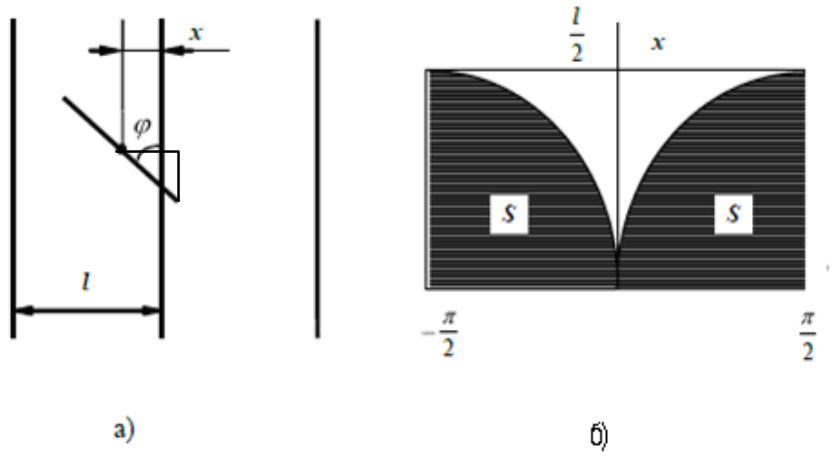


Рисунок 2

Положение иглы относительно пересечённой линии выразим парой (x,φ), которую будем рассматривать как точку на координатной плоскости с координатами (x,φ), где:

$x \in \left[0,\frac{l}{2}\right]$ - расстояние от середины иглы до точки пересечения линии;

$\phi \in \left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ - острый угол между иглой и пересечённой линией (рисунок 2,а). Острый угол слева от пересекаемой линии - отрицательный, справа - положительный. Множество таких точек будет равно декартовому произведению - $(x,\varphi)\in\left[0,\frac{l}{2}\right]\times\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. Очевидно, что расстояние x от середины иглы до пересечения иглой линии будет меньше или равно катету прямоугольного треугольника, гипотенузу которого составляет половина иглы. Поэтому событие A пересечения иглой некоторой линии возникает, если точка (x,φ), соответствующая положению иглы, будет удовлетворять условию, выраженному неравенством:

$$x \le \frac{l}{2}|\sin\phi|.$$

Множеству точек (x,φ), удовлетворяющих этому условию, будет соответствовать заштрихованная часть прямоугольной области $\left[0,\frac{l}{2}\right]\times\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ (рисунок 2, б). Положим, что в процессе бросания иглы реализации случайных величин x и φ являются независимыми и подчиняются *равномерному закону распределения*. Тогда получаемые при бросании иглы значения x и φ будут равномерно заполнять соответственно отрезки $\left[0,\frac{l}{2}\right]$ и $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$, и, следовательно, точки (x,φ) будут равномерно заполнять всю прямоугольную область. При равномерном распределении точек (x,φ) в прямоугольной области вероятность реализации события A будет равна отношению площадей "заштрихованной" области S к площади прямоугольника. Найдём площадь заштрихованной области S:

$$S = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{l}{2}|\sin\phi|\,d\phi = 2\int_0^{\frac{\pi}{2}} \frac{l}{2}\sin\phi\,d\phi = -l\cos\phi\Big|_0^{\frac{\pi}{2}} = -l(0-1) = l.$$

Так как площадь прямоугольной области равна $\frac{l}{2}\pi$, то вероятность события A вычисляется по формуле:

$$P\{A\} = \frac{l}{\frac{l}{2}\pi} = \frac{2}{\pi}.$$

С учётом ранее найденной статистической оценки вероятности $P\{A\}$ получим:

$$\frac{2}{\pi} \approx \frac{m}{N}.$$

В итоге число π приближённо вычисляется по формуле $\pi \approx \frac{2N}{m}$. Точность вычисления π зависит от числа испытаний N. При достаточно большом числе испытаний N можно получить приемлемый результат оценки числа π с установившимися дробными разрядами.

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

Применение статистических методов для анализа имитационных стохастических моделей связано с необходимостью генерации последовательностей случайных чисел, являющихся реализациями случайных величин с заданными законами распределений вероятностей. При этом результаты статистического анализа стохастических моделей существенно зависят от качества используемых генераторов случайных чисел, то есть от того насколько точно они соответствуют заданному закону распределения. Поэтому наличие простых и экономичных способов формирования нужных последовательностей случайных чисел требуемого качества во многом определяет возможность эффективного практического использования статистических методов анализа.

Далее будет показано, что методы моделирования непрерывных или дискретных случайных чисел с требуемым законом распределения базируются на преобразовании непрерывной случайной величины, равномерно распределённой на отрезке [0,1]. Поэтому проблема в итоге сводится к использованию хорошего генератора действительно случайных и действительно равномерно распределённых чисел. Если базовый генератор выдаёт числа, связанные по вероятности или смещённые в какую-то часть отрезка [0,1] (одни числа выпадают чаще других), то результаты статистического анализа будут не состоятельными.

Так как непрерывное множество чисел является математической абстракцией, то используемый на практике базовый генератор будет лишь с определённой степенью точности генерировать последовательность стохастических значений *дискретной* случайной величины равномерно распределённой на отрезке [0,1], характеристики которой должны соответствовать теоретическим характеристикам *непрерывной* случайной величины r с равномерным законом распределения на отрезке [0,1]. Для *непрерывной равномерно распределённой* на отрезке [0,1] случайной величины r функция плотности распределения её значений x имеет вид:

$$f_r(x) = \begin{cases} 1, & 0 \le x \le 1; \\ 0, & x < 0, x > 1. \end{cases}$$

Ниже показан график плотности распределения значений случайной величины *r* (рисунок 3):

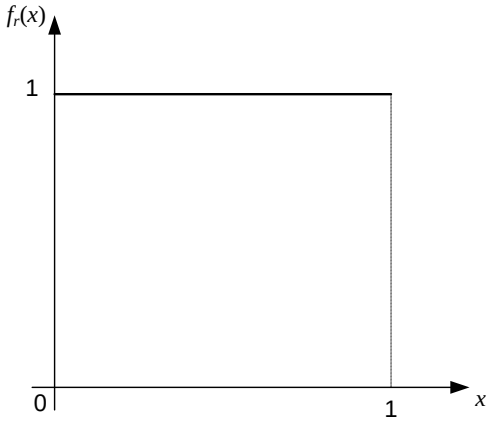


Рисунок 3

Непрерывная случайная величина *r* имеет следующие теоретические характеристики - математическое ожидание $M[r]$ и дисперсия $D[r]$:

$$M[r] = \int_0^1 x f_r(x)dx = \int_0^1 x dx = \frac{1}{2},$$

$$D[r] = \int_0^1 (x - M[r])^2 f_r(x)dx = \int_0^1 (x - M[r])^2 dx = \frac{1}{12}.$$

Очевидно, что на практике невозможно моделировать непрерывную случайную величину *r*, принимающую любое из возможных значений на отрезке [0,1]. Прежде всего, это связано ограниченной разрядностью практического представления значений (например, ограниченная разрядность ячеек памяти компьютера, а также практическая невозможность в численном виде представить числа с бесконечной дробной частью). Поэтому для численного моделирования теоретически непрерывная случайная величина *r* заменяется аналогом в виде дискретной *n*-разрядной *двоичной* случайной величины $\dot{r}$, принимающей на отрезке [0,1] лишь только 2^n (n – разрядность представления двоичных чисел) одинаково отстоящих друг от друга и равномерно "засеивающих" отрезок [0,1] значений $x_k = \frac{k}{2^n - 1}$ (k=0,1,2,…,2^n – 1). Вероятность выдачи генератором любого из значений x_k должна быть одна и та же и равна $p_k = \frac{1}{2^n}$. Дискретную случайную величину $\dot{r}$ называют *квазинепрерывной*, а её распределение вероятностей -

законом квазинепрерывного равномерного распределения. Ниже показан график квазиравномерного распределения на отрезке [0,1] (рисунок 4):

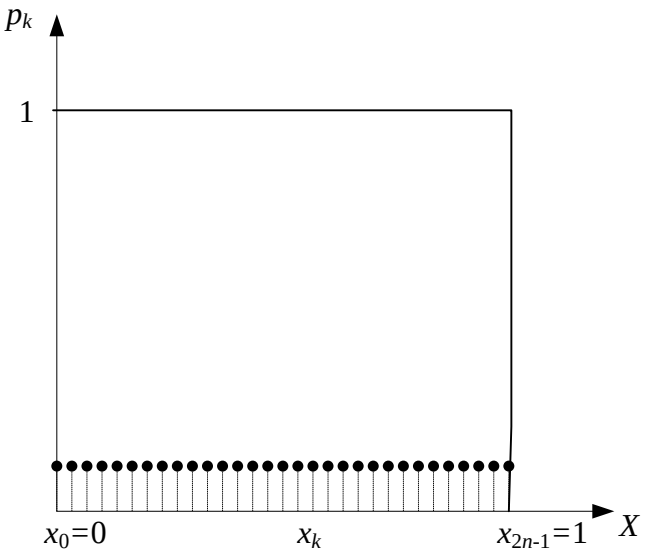


Рисунок 4

Математическое ожидание и дисперсия случайной величины $\dot{r}$ с *квазинепрерывным равномерным* распределением соответственно имеют вид:

$$M[\dot{r}] = \sum_{k=0}^{2^n-1} \frac{k}{(2^n-1)} \frac{1}{2^n} = \frac{1}{2}$$

$$D[\dot{r}] = \sum_{k=0}^{2^n-1} \left[\frac{k}{(2^n-1)} - \frac{1}{2}\right]^2 \frac{1}{2^n} = \frac{1}{12} \cdot \frac{2^n+1}{2^n-1} = \frac{1}{12} \cdot \left(1 + \frac{2}{2^n-1}\right) \underset{n\to\infty}{\longrightarrow} \frac{1}{12}$$

Заметим, что математическое ожидание квазинепрерывной равномерно распределённой случайной величины $\dot{r}$ совпадает с математическим ожиданием соответствующей непрерывной случайной величины r, равномерно распределённой на отрезке [0,1], а дисперсия при $n\to\infty$ стремится к $\frac{1}{12}$. Следовательно, при достаточно большом значении n характеристики квазинепрерывного аналога будут практически мало отличаться от теоретических характеристик непрерывной равномерно распределённой случайной величины.

Таким образом, на практике при моделировании непрерывной равномерно распределённой случайной величины она будет заменяться её квазинепрерывным дискретным аналогом с заданной точностью получения случайных значений.

Генераторы случайных чисел

Для получения при статистическом моделировании значений случайных величин с различными законами распределения используются соответствующие *генераторы случайных чисел* (ГСЧ). Как будет показано в дальнейшем, генератор значений случайной величины с любым заданным законом распределения можно получить на основе генератора значений случайной величины с *равномерным* законом распределения на отрезке [0,1]. Такой генератор используется в качестве базового ГСЧ.

Базовые ГСЧ по способу получения значений делятся на:

- физические;
- табличные;
- алгоритмические.

Физические генераторы случайных чисел

Основой построения физического базового ГСЧ служат физические генераторы равновероятных случайных событий.

Рассмотрим пример генерации значений равномерно распределённой случайной величины, используя в качестве физического ГСЧ результат подбрасывания монеты с двумя исходами: «орёл» — 1, «решка» — 0.

Первый способ

Подбросим монету 4 раза. Если монета упала решкой, то пишем **0**, если орлом, то - **1**. Допустим, что в результате испытаний получили случайную последовательность из 4-х двоичных цифр: **1001**.

Полученную последовательность двоичных цифр будем использовать следующим образом. Возьмём отрезок от 0 до 1. Считывая цифры в последовательности слева направо, будем разбивать текущий отрезок пополам и выбирать каждый раз одну из половин (если выпал **0** - левую, если выпала **1** - правую). Действуя таким же образом по отношению к выделенной половине, используя очередную двоичную цифру, вновь получаемые половины будут

становиться всё меньше и меньше. Половина, выделенная в результате анализа последней 4-ой цифры, определит на отрезке [0, 1] интервал, куда попадает искомое случайное число. Любое число этого интервала с соответствующей точностью можно принять в качестве сгенерированного значения случайной величины. Увеличивая количество бросаний монеты, можно получать значения случайной величины со сколь угодно заданной точностью.

Итак, **1**: отрезок [0, 1] делится пополам — (0, 0.5] и (0.5, 1], — выбирается правая половина, интервал сужается: (0.5, 1]. Следующее число **0**: интервал (0.5, 1] делится пополам — (0.5; 0.75) и (0.75; 1], — выбирается левая половина (0.5; 0.75). Следующее число, **0**: интервал (0.5; 0.75) делится пополам — (0.5, 0.625) и (0.625, 0.75), — выбирается левая половина (0.5, 0.625). Следующее число, **1**: интервал (0.5; 0.625) делится пополам — (0.5; 0.5625) и (0.5625; 0.625), — выбирается правая половина (0.5625; 0.6250). В итоге, получаем интервал (0.5625; 0.6250). Любое число из интервала (0.5625; 0.6250), например, 0.6, является искомым значением.

Если подходить более строго и искать число с заданной точностью (количество знаков после запятой), то исходная последовательность случайных нулей и единиц должна быть значительно больше и такой, чтобы деление интервалов продолжалось до тех пор, пока левая и правая границы найденного интервала в пределах заданной точности не *совпадут* между собой, например, до третьего знака после запятой. То есть с этой точностью все числа сгенерированного интервала будут неразличимы.

Второй способ

Подбросим монету количество раз кратное трём, например - 9-ть раз. Разобьём полученную последовательность двоичных цифр, например – 100110100, на триады: 100, 110, 100. После перевода этих двоичных чисел в восьмеричные цифры получаем: 4, 6, 4. Подставив спереди «0.», получим: 0.464_8. Таким методом могут получаться только числа от 0.000_8 до 0.777_8 (так как максимум, что можно «выжать» из трёх двоичных разрядов — это $111_2 = 7_8$). Выполним перевод *восьмеричного* числа в *десятичное* представление:

$$0.464_8 = 4 \cdot 8^{-1} + 6 \cdot 8^{-2} + 4 \cdot 8^{-3} = 0.6015625_{10} = 0.602_{10}.$$

Итак, искомое число равно: 0.602_{10}.

Генератором равновероятных случайных событий может служить и аппаратный генератор "случайного шума" (ГШ) электрического напряжения, обработка которого позволяет получить случайную последовательность нулей и единиц (рисунок 5).

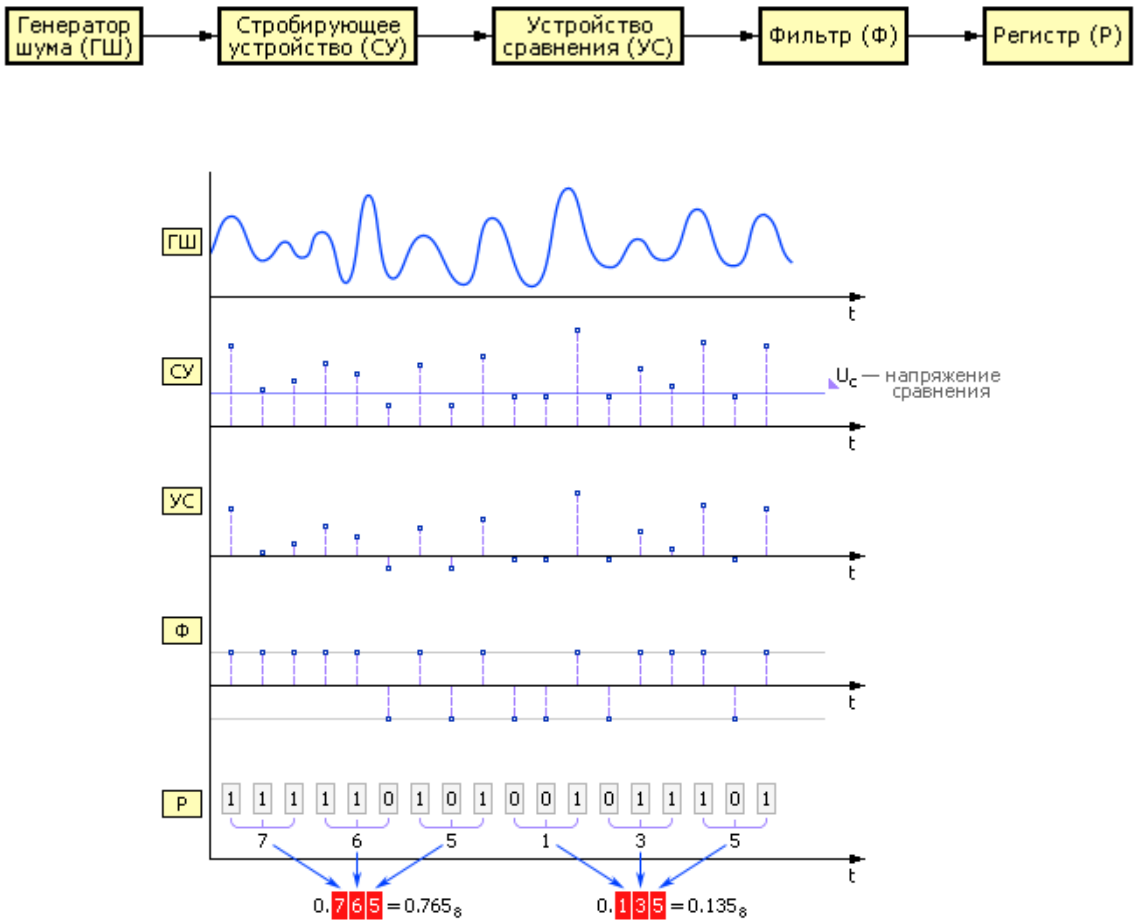


Рисунок 5

Табличные генераторы случайных чисел

Табличные ГСЧ в качестве источника случайных чисел используют специальным образом составленные таблицы, содержащие проверенные некоррелированные, то есть никак не зависящие друг от друга, цифры. В табл. 1 приведён небольшой фрагмент такой таблицы. Обходя таблицу слева направо и сверху вниз, можно получать равномерно распределённые от 0 до 1 случайные числа с нужным числом знаков после запятой (в нашем примере мы используем для каждого числа по три знака). Так как цифры в таблице не зависят друг от друга, то таблицу можно обходить разными способами, например, сверху вниз, или справа налево, или, скажем, можно выбирать цифры, находящиеся на четных позициях. Ниже в правом столбце таблицы приведены реализации случайных

чисел в виде трёхзначных десятичных дробей, полученных последовательной выборкой троек подряд идущих цифр, представленных в левой части таблицы

Таблица 1.

Случайные цифры								Равномерно распределённые 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
								0.573
…								…

Достоинство данного метода в том, что он даёт действительно случайные числа с заданной разрядностью, так как таблица содержит проверенные некоррелированные цифры. Недостаток метода в том, что для хранения большого количества цифр требуется много памяти. Большие трудности с вязаны с порождением и проверкой такого рода таблиц. Повторы при использовании таблицы уже не гарантируют случайности числовой последовательности, а значит и качество генерируемой последовательности.

Алгоритмические генераторы случайных чисел

Алгоритмические ГСЧ базируются на использовании алгебраического рекуррентного соотношения вида:

$$r_{i+1} = f(r_i), i = 0,1,2, \ldots,$$

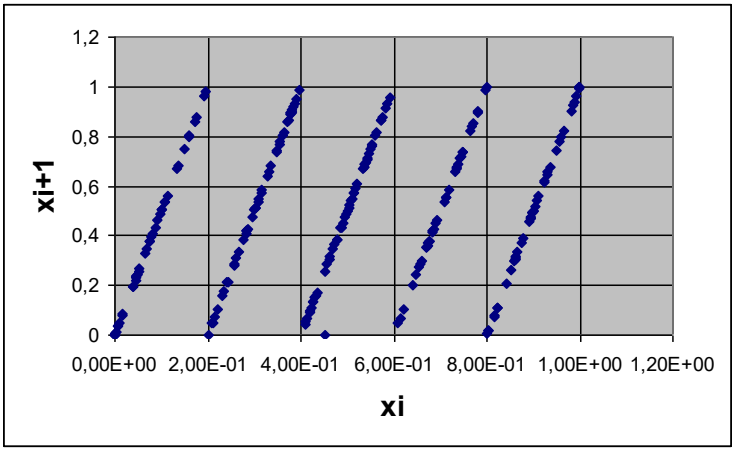
где каждое последующее вычисленное число r_{i+1} зависит от предыдущего вычисленного - r_i. Из теории чисел известно, что последовательности чисел, формируемые такими рекуррентными соотношениями, образуют *петли*, то есть обязательно существует цикл, повторяющейся бесконечное число раз одной и той же последовательности чисел, который называют *периодом* рекуррентной последовательности. Поэтому получаемые с помощью алгоритмических ГСЧ последовательности случайных чисел являются квазислучайными числами псевдонепрерывной случайной величины, степень валидности которых требуемой

точности решения задачи статистического анализа стохастической модели, требует отдельного рассмотрения. В то же время, важным достоинством алгоритмических ГСЧ является их простота компьютерной реализации и быстродействие.

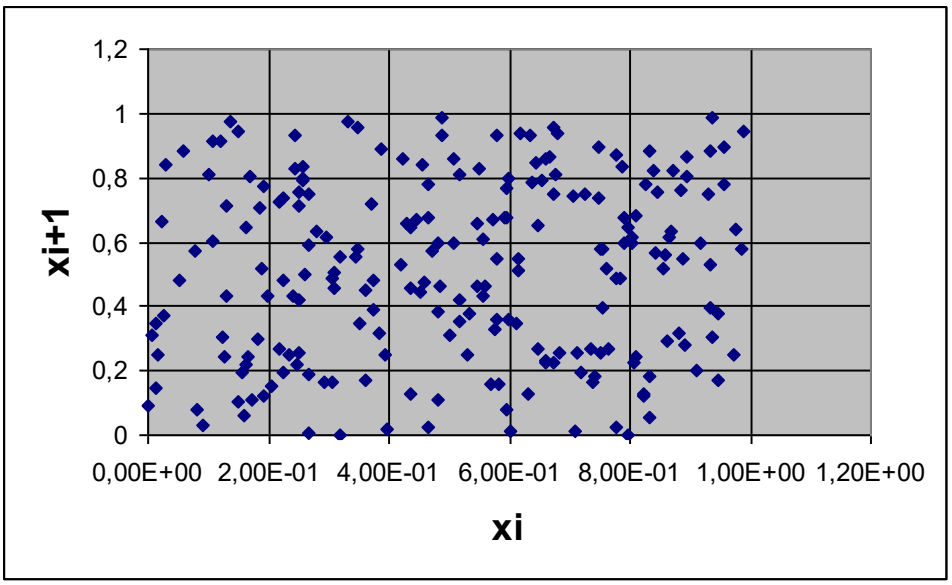
Идею получения псевдослучайных чисел, квазиравномерно распределённых на отрезке [0,1], используя рекуррентное соотношение $r_{i+1} = f(r_i)$, проиллюстрируем на примере рекуррентной функции, основанной на операции получения дробной части десятичного числа:

$$r_{i+1} \equiv x_{i+1} = Д(A \cdot x_i),$$

$Д(u)$ – дробная часть числа $u = A \cdot x_i$, где A – некоторое целое положительное число. Если, например, в качестве начального значения взять x_0=0,0000123456789, а в качестве A=5, то график рекуррентной зависимости $x_{i+1} = Д(A \cdot x_i)$ будет иметь вид:



Глядя на "пилообразный" график этой рекуррентной функции, можно сделать вывод, что порождаемая последовательность чисел $r_i \equiv x_i$, не подходит для имитации последовательности случайных чисел последовательность не является псевдослучайных хорошей, просматриваются ярко выраженные области сильной корреляции между числами. Однако, если изменить исходные данные: x_0=0,135711234567, A=134775813, то график существенно меняется:



Очевидно, что теперь генерируемая последовательность чисел больше похожа на реализацию случайной величины с равномерным законом распределения на отрезке [0,1]. Из этого можно сделать вывод, что качество генерации последовательности псевдослучайных чисел существенно зависит от выбора исходных значений параметров используемого рекуррентного преобразования. Очевидно, что поиск эффективного рекуррентного преобразования и настройка его параметров становятся отдельной теоретической проблемой.

Алгоритм середины квадрата Неймана

Простой алгоритм для получения равномерно распределённых псевдослучайных чисел был предложен венгеро-американским учёным Джоном Фон Нейманом в 1949 году. Он называется методом середины квадрата. Пусть задано произвольно выбранное 4-значное число r_0=0.9876. Возведя его в квадрат, получим 8-значное число r_{02}=0.97535376. Выбрав четыре "серединных" цифры дробной части этого числа, получим следующее псевдослучайное число: r_1=0.5353. Повторив операцию возведения в квадрат уже с r_1, получим: r_{12}=0,28654609. Опять выберем серединные цифры дробной части числа и получим очередное псевдослучайное число: r_2=0.6546. Таким образом, получен очень простой рекуррентный алгоритм генерирования псевдослучайных точек, который называется методом Неймана или методом серединных квадратов. Для запуска алгоритма достаточно задать некоторое произвольное начальное число r_0. Разным

стартовым числам, вообще говоря, будут соответствовать разного качества последовательности чисел с разными периодами рекуррентной последовательности.

Линейный конгруэнтный метод

Решение проблемы качества алгоритмических ГСЧ выразилось в нахождении рекуррентного преобразования, известного как линейный конгруэнтный метод генерации псевдослучайных квазинепрерывных чисел, распределённых на отрезке [0,1]. Линейный конгруэнтный метод является одним из простейших и наиболее эффективных методов, используемых в настоящее время процедурами генерации псевдослучайных чисел. В этом методе в качестве случайного значения r используется результат бинарной операции $\mathrm{mod}(x, y)$, возвращающей остаток от деления x на y:

$$r=\mathrm{mod}(x, y).$$

Каждое последующее случайное число рассчитывается на основе предыдущего случайного числа (полагая, что $x_i= k\cdot r_i + b$, а $y=M$ – заданное целое число - *модуль*) по следующей формуле:

$$r_{i+1} = \mathrm{mod}(k\cdot r_i + b, M).$$

M — заданный параметр - *модуль* ($M > 0$);
k — множитель ($0 \le k < M$);
b — заданный параметр - *приращение* ($0 \le b < M$);
r_0 — заданное начальное значение ($0 \le r_0 < M$).

Последовательность случайных чисел, полученных с помощью данной формулы, называется *линейной конгруэнтной последовательностью*. Принято получение линейной конгруэнтной последовательности при $b=0$ называть *мультипликативным конгруэнтным методом*, а при $b\neq0$ — *смешанным конгруэнтным методом*. Мультипликативный конгруэнтный метод генерации

псевдослучайных чисел был предложен американским учёным Д. Г. Лемером (D. H. Lehmer) в 1949 году.

Для качественного генератора требуется подобрать подходящие параметры: M, k, b; и начальное значение r_0 формируемой последовательности случайных чисел r. Было установлено, что число M должно быть довольно большим, так как теоретически показано, что *период* генерируемой последовательности *не может иметь больше* M элементов. С учётом того, что в общем случае деление, использующееся в этом методе, для компьютера является довольно медленной операцией, то логично выбирать M равным степени двойки - 2^N, поскольку в этом случае нахождение остатка от деления для компьютера упрощается. Кроме того, широко распространён выбор в качестве M наибольшего *простого* числа, меньшего, чем 2^N (*число Мерсенна*). Теоретически доказано, что в этом случае младшие разряды получаемого случайного числа r_{i+1} ведут себя так же случайно, как и старшие, что положительно сказывается на всей последовательности случайных чисел в целом.

Одним из требований к линейным конгруэнтным последовательностям является как можно большая длина периода. Теоретически доказано, что длина периода существенно зависит от выбранных значений M, k, b и r_0. После долгих поисков были найдены эффективные параметры настройки линейного конгруэнтного метода с практически большим периодом генерации псевдослучайных чисел.

Линейный конгруэнтный метод используется в современных операционных системах компьютеров для реализации генераторов случайных чисел, равномерно распределённых на отрезке [0,1].

Моделирование дискретной случайной величины

Реализацию дискретной случайной величины можно рассматривать как реализацию одного из полной группы несовместных случайных событий. Отсюда следует, что суммарная вероятность полной группы несовместных событий равна 1.

Обозначим через a_1, a_2, …, a_n несовместные случайные события, а через P_1, P_2, …, P_n — вероятности появления этих событий, $P_1 + P_2 + … + P_n = 1$.

Отложим на единичном отрезке [0; 1] последовательные отрезки длиной P_1, P_2, …, P_n. Понятно, что в сумме отрезки составят точно единичный интервал.

Воспользуемся для имитации выпадения одного из событий генератором равномерно распределённых случайных чисел, значение которых находится в диапазоне от 0 до 1. Точка, соответствующая выпавшему числу из ГСЧ на этом интервале, попадёт в один из отрезков P_i, что ассоциируем с реализацией события a_i. Вероятность попасть в тот или иной отрезок равна длине этого отрезка. В большие отрезки генерируемые случайные числа будут попадать чаще (вероятность появления этих событий больше!), в меньшие отрезки — реже (рисунок).

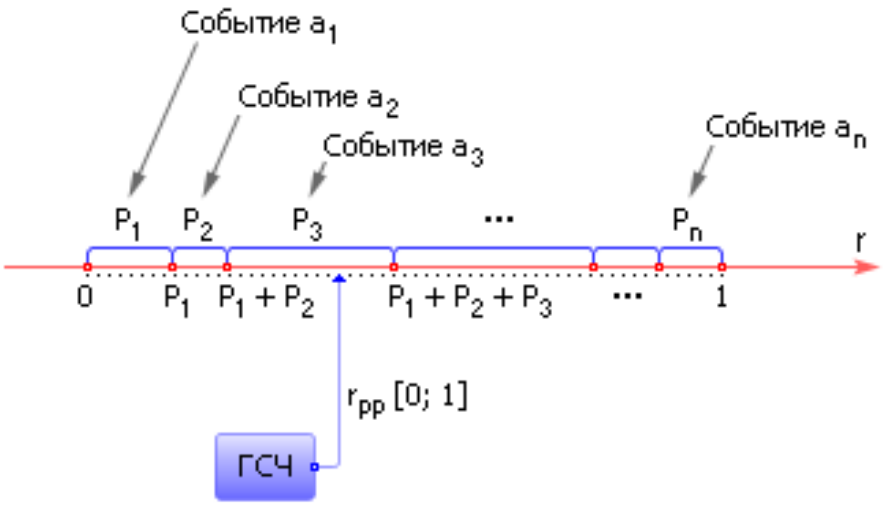


Рисунок 6

Моделирование непрерывной равномерно распределённой случайной величины

Если пользователю потребуется, чтобы случайное число x было равномерно распределено на отрезке $[a; b]$, нужно воспользоваться формулой $x = a + (b - a)r$, где r — равномерно распределённое случайное число на отрезке [0; 1]. Законность данного преобразования демонстрируется ниже на рисунке (рисунок).

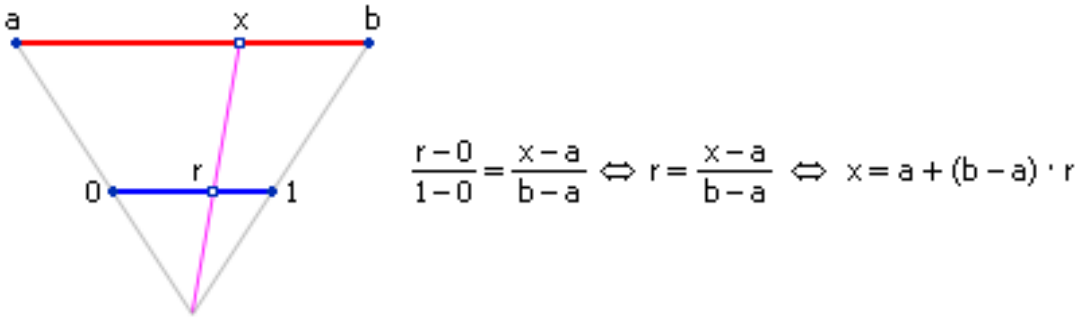


Рисунок 7

Теперь $x = a + (b - a)r$ — случайное число, распределённое в диапазоне от a до b.

Моделирование непрерывной случайной величины с заданным законом распределения

Метод обратной функции

Требуется построить генератор случайных чисел непрерывной случайной величины η, генерирующий последовательность случайных чисел x с интегральным законом распределения $y = F_\eta(x)$, $y \in [0,1]$. Построим этот генератор как преобразователь значений генератора случайной величины $r_{\text{pp}}[0,1]$, равномерно распределённой на $[0,1]$, в случайное значение x случайной величины η с интегральным законом распределения $F_\eta(x)$.

Положим, что для заданного интегрального закона распределения $F_\eta(x)$ существует обратная функция: $x = F_\eta^{-1}(y)$. Доказано, что такая обратная функция и является искомым преобразованием значений y генератора $r_{\text{pp}}[0,1]$ в значения x случайной величины η с интегральным законом распределения $F_\eta(x)$

Ниже суть метода иллюстрируется графически (рисунок 8).

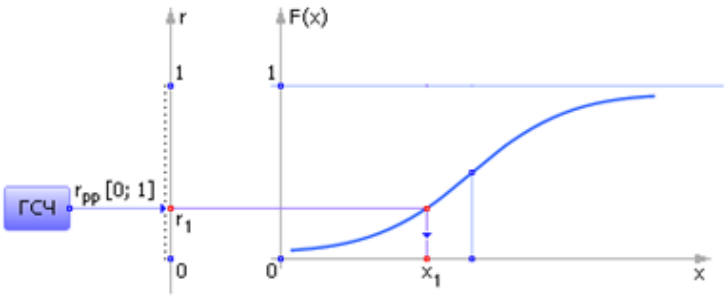


Рисунок 8

На рисунке показан интегральный закон распределения F(x)≡$F_\eta(x)$ некой случайной величины η. Для получения последовательности случайных значений x_1, x_2, x_3,… с интегральный закон распределения F(x) воспользуемся ГСЧ с равномерным законом распределения - r_{pp}[0;1] для получения последовательности случайных значений y_1, y_2, y_3,… Для получения значений x_i случайной величины с интегральным законом распределения F(x) (рисунок 8) необходимо подставлять сгенерированные r_{pp}[0;1] значения y_i в функцию, обратную этому интегральному закону распределения - $x_i = F_\eta^{-1}(y_i)$. В результате будет порождаться последовательность значений x_1, x_2, x_3,… случайной величины η с интегральным законом распределения $F_\eta(x)$.

Пример

Пусть требуется построить ГСЧ с экспоненциальным законом распределения $F_\eta(x) = 1 - e^{-\lambda x}$. Тогда $F_\eta^{-1}(r) = -\frac{1}{\lambda} ln(1 - r)$, откуда имеем $\eta = -\frac{1}{\lambda} ln(1 - r)$, где r - равномерно распределённая случайная величина на отрезке [0,1]. Так как в статистическом смысле значения (1–r) и значения r одинаково распределены, то в качестве обратной функции воспользуемся функцией $\eta = -\frac{1}{\lambda} ln(r)$.

Метод обратной функции является теоретически строгим методом моделирования непрерывных случайных величин с заданным интегральным законом распределения. Он позволяет получать случайные числа даже в тех случаях, когда величина определена на бесконечных интервалах числовой оси (как, например, для экспоненциальной случайной величины, когда $\xi \in (0,\infty)$. Однако, не всегда можно получить аналитическое выражение обратной функции. Известно, например, что для гауссовой случайной величины вообще не существует аналитического выражения интегрального закона распределения и, следовательно, аналитического представления обратной функции. В таких случаях приходится применять численные методы задания обратной функции, что повышает

сложность получения адекватной последовательности случайных значений.

Моделирования гауссовой случайной величины

Для моделирования гауссовой случайной величины можно воспользоваться центральной предельной теоремой. Пусть ξ_i - независимые, равномерно распределённые на отрезке [0,1] случайные величины с математическим ожиданием $M[\xi]=1/2$ и дисперсией $D[\xi]=1/12$, тогда случайная величина $\eta = \sum_{i=1}^{n} \xi_i$ будет иметь при $n \to \infty$ нормальный закон распределения с $M[\eta]=1/2$ и $D[\eta]=D[\xi]/n$. После нормализации случайной величины $\eta = \sum_{i=1}^{n} \xi_i$ получим:

$$\eta = \sqrt{\frac{12}{n}} \sum_{i=1}^{n} (\xi_i - \frac{1}{2})$$

при этом случайная величина η имеет нулевое математическое ожидание и единичную дисперсию. Если взять n=12, то:

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

Как показали статистические исследования даже при n=12 данный метод хорошо согласуется с гипотезой о нормальности распределения случайной величины η.

Метод ступенчатой аппроксимации

Метод ступенчатой аппроксимации для моделирования *ограниченной* непрерывной случайной величины ξ основан на использовании заданной для неё функции плотности распределения $f_\xi(x)$, которая аппроксимируется дискретным приближением в виде ступенчатой функции. Для этого *ограниченная* область определения функции $f_\xi(x)$ разбивается на n непересекающиеся равных интервалов размером Δ, на которых строится ступенчатая функция. Высота прямоугольной ступени i-го интервала - h_i, подбирается так, чтобы ступень находилась под функцией плотности распределения $f_\xi(x)$ и имела бы с ней точку касания (рисунок 9, а).

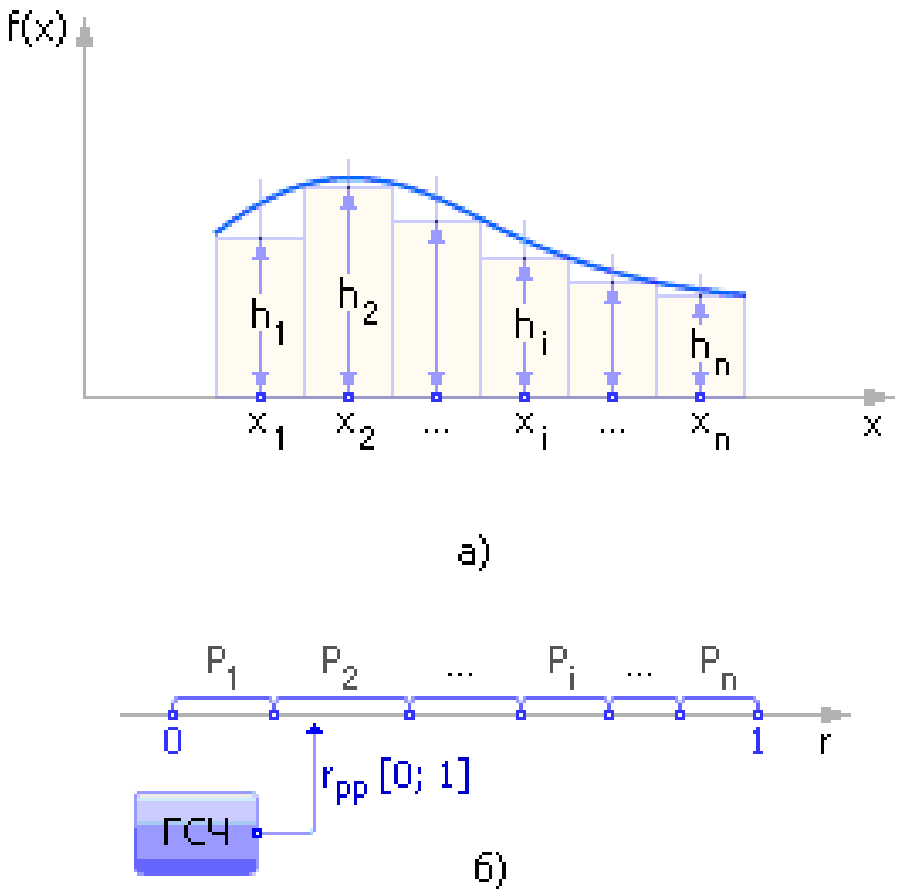


Рисунок 9

Далее осуществляется переход от непрерывной случайной величины ξ к дискретной случайной величине $\ddot{\xi} \in \{x_1, x_2, \dots, x_n\}$ с вероятностями $P(x_i)$ появления дискретного значения x_i, равного отношению площади прямоугольника с основанием Δ и высотой h_i, к сумме площадей всех прямоугольников под аппроксимирующей ступенчатой функцией. В итоге вероятность $P(x_i)$ получения непрерывной случайной величиной ξ значения x_i, принадлежащего i-ому интервалу будет равна:

$$P(x_i) = \frac{\Delta \cdot h_i}{\sum_{k=1}^{n} \Delta \cdot h_k} = \frac{h_i}{\sum_{k=1}^{n} h_k}, i = 1,2, \dots, n.$$

Для устранения погрешностей дискретизации необходимо выполнить операцию нормировки полученных вероятностей, чтобы выполнялось условие:

$$\sum_{i=1}^{n} P(x_i) = 1$$

На рисунок 9, а графически показан переход от функции плотности распределения непрерывной случайной величины ξ к распределению вероятностей аппроксимирующей дискретной случайной величины $\ddot{\xi} \in \{x_1, x_2, \dots, x_n\}$.

Далее генерация значений *x* непрерывной случайной величины ξ происходит как генерация значений дискретной случайной величины $\ddot{\xi}$. Для этого осуществляется разбиение отрезка [0; 1] на последовательно примыкающие друг к другу интервалы, по величине соответствующие $P(x_i)$. Далее, для получения значения x непрерывной случайной величины ξ, с использованием эталонного ГСЧ - $r_{\text{рр}}$[0; 1], генерируется случайное значение равномерно распределённое на отрезке [0; 1] - (рисунок 9, б). Попадание полученного значения в некий i-ый интервал внутри отрезка [0; 1] определяет получение случайной величиной ξ значения x_i.

Очевидно, что погрешность метода ступенчатой аппроксимации для генерации значений *ограниченной* непрерывной случайной величины ξ с заданной функцией распределения плотности вероятности $f_{\xi}(x)$ зависит от количества сформированных интервалов n при переходе от непрерывного распределения к дискретному. Недостатком метода является и то, что он применим только для генерации значений случайных величин с ограниченной областью значений.

Метод усечения

Рассмотренный ранее метод ступенчатой аппроксимации предполагает, что функция плотности вероятности $f_{\xi}(x)$ непрерывной случайной величины ξ определена на ограниченной области значений *x*. Если область значений случайной величины ξ не ограничена, то применение метода ступенчатой аппроксимации становится проблематичным. Кроме того, необходимость ассоциировать попадание в некий интервал с реализацией серединного значения этого интервала так же является недостатком метода, так как принципиально не позволяет получать

другие значения. Это существенный недостаток данного метода. Указанные недостатки учтены в методе усечения, который применим для генерации случайных значений с заданной функцией плотности распределения на любом множестве. Не вдаваясь в теоретическое обоснование метода, рассмотрим только алгоритм его реализации.

Если область значений не ограничена, то предварительно график функции плотности распределения $f_\xi(x)$ вписывают в прямоугольник (рисунок 10) со сторонами $[x_0,x_n]$ и $[y_0,y_n]$, усекая маловероятные значения случайной величины ξ.

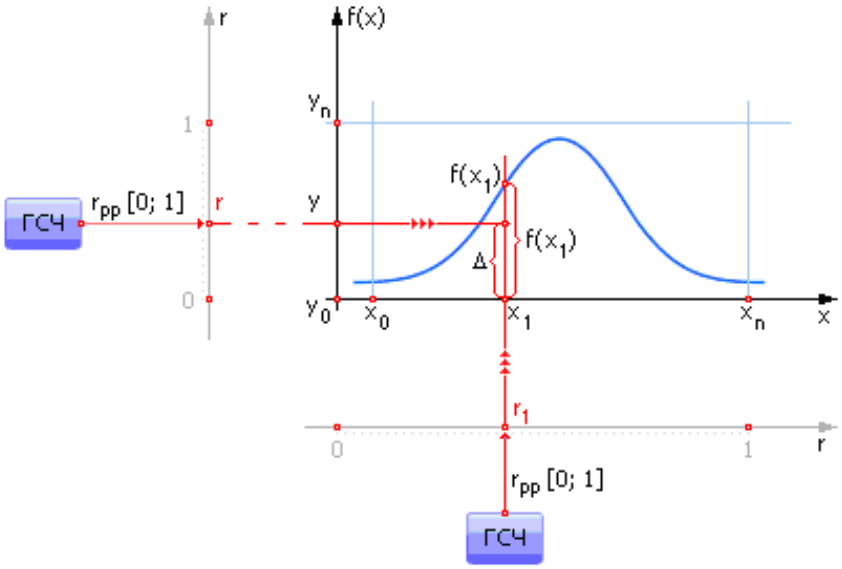


Рисунок 10

Далее на ось $f(x) \equiv f_\xi(x)$ подают число y_i случайной величины, равномерно распределённой на отрезке $[y_0,y_n]$, полученное с использованием соответствующего ГСЧ. Затем на ось x подают число x_i случайной величины, равномерно распределённой на отрезке $[x_0,x_n]$, полученное из соответствующего ГСЧ. Если точка (x_i,y_i) лежит ниже кривой плотности вероятности $f_\xi(x)$, то x_i берётся в качестве значения случайной величины ξ, иначе - x_i отбрасывается и опыт повторяется.

Достоинство данного метода в том, что нет необходимости в ступенчатой аппроксимации плотности вероятности $f_\xi(x)$ и теоретически все значения непрерывной случайной величины с вычислительной точностью могут быть

реализованы. Однако существенным недостатком метода является то, что те точки, которые оказались выше кривой $f_\xi(x)$, отбраковываются, тем самым увеличивая трудоёмкость генерации случайных значений.