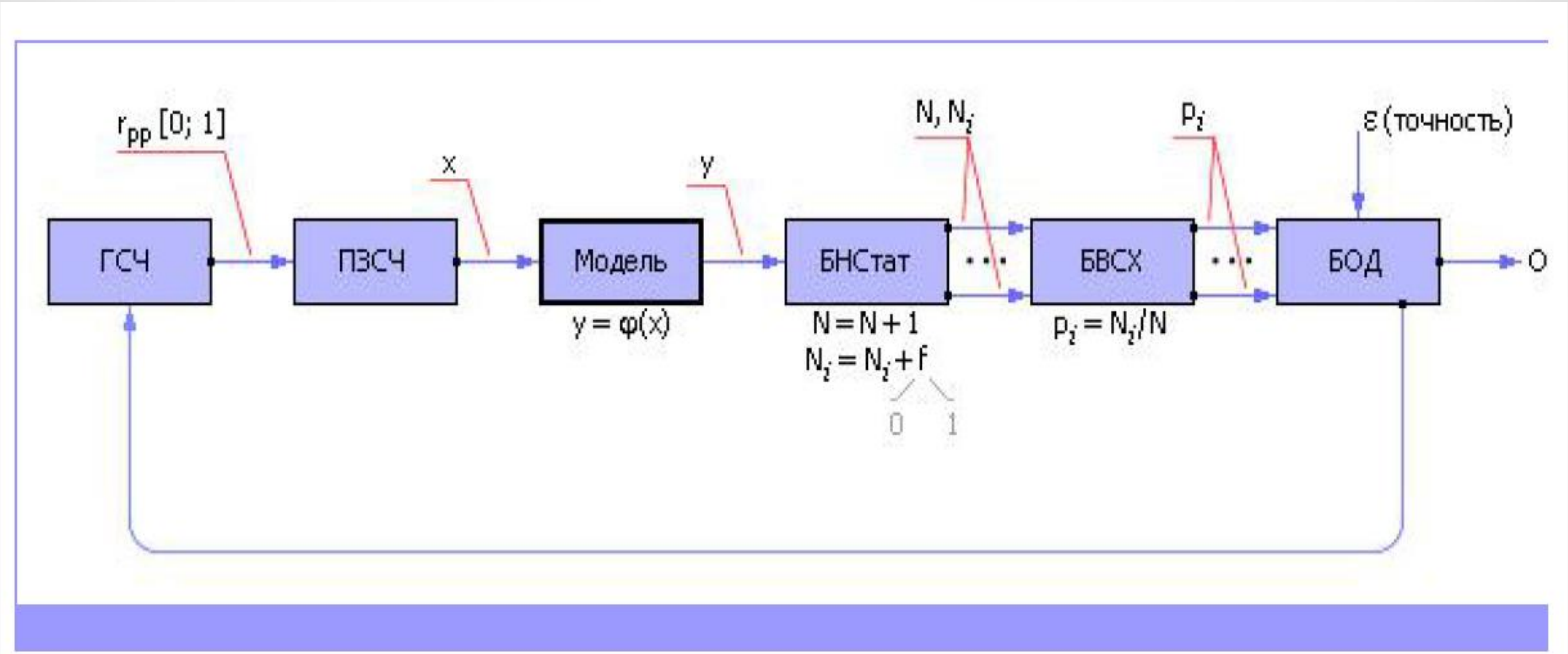
Генерирование и моделирование случайных чисел

Методы и применения

Имитационное моделирование и случайные числа

- Случайные числа находят применение при математическом моделировании стохастических систем, в поисковых алгоритмах, системах шифрования и алгоритмах управления в системах.
- Имитационное моделирование систем воспроизводит поведение систем, которое разворачивается в масштабах модельного или реального времени.
- Моделируемое поведение реализует при запуске имитационной модели конкретный сценарий, который соответствует в установленных целями исследования рамках множеству возможных вариантов реализации сценария поведения реальной системы, модель которой исследуется

Общая схема метода статистического моделирования



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

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

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

Определения наблюдаемых статистических объектов

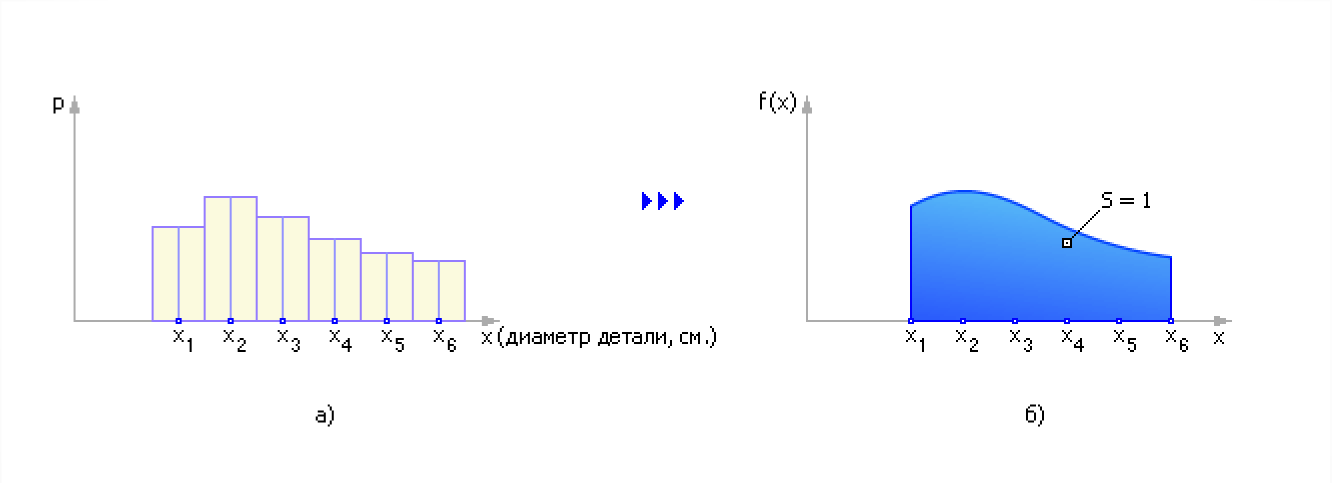
- **Генеральная совокупность** – это совокупность всех мыслимых наблюдений (или всех мысленно возможных объектов интересующего нас типа, с которых «снимаются:» наблюдения), которые могли бы быть произведены при данном реальном комплексе условий.
- **Выборка** из данной генеральной совокупности — это результаты ограниченного ряда наблюдений $\{X_1, X_2, \dots, X_n\}$ случайной величины ξ.
- **Статистический ансамбль** – это совокупность сколь угодно большого числа одинаковых физических систем многих частиц («копий» данной стохастической системы.

Оценка статистических данных по результатам полученных выборок при исследовании результатов имитационного моделирования

- 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.

Распределение частоты наблюдений

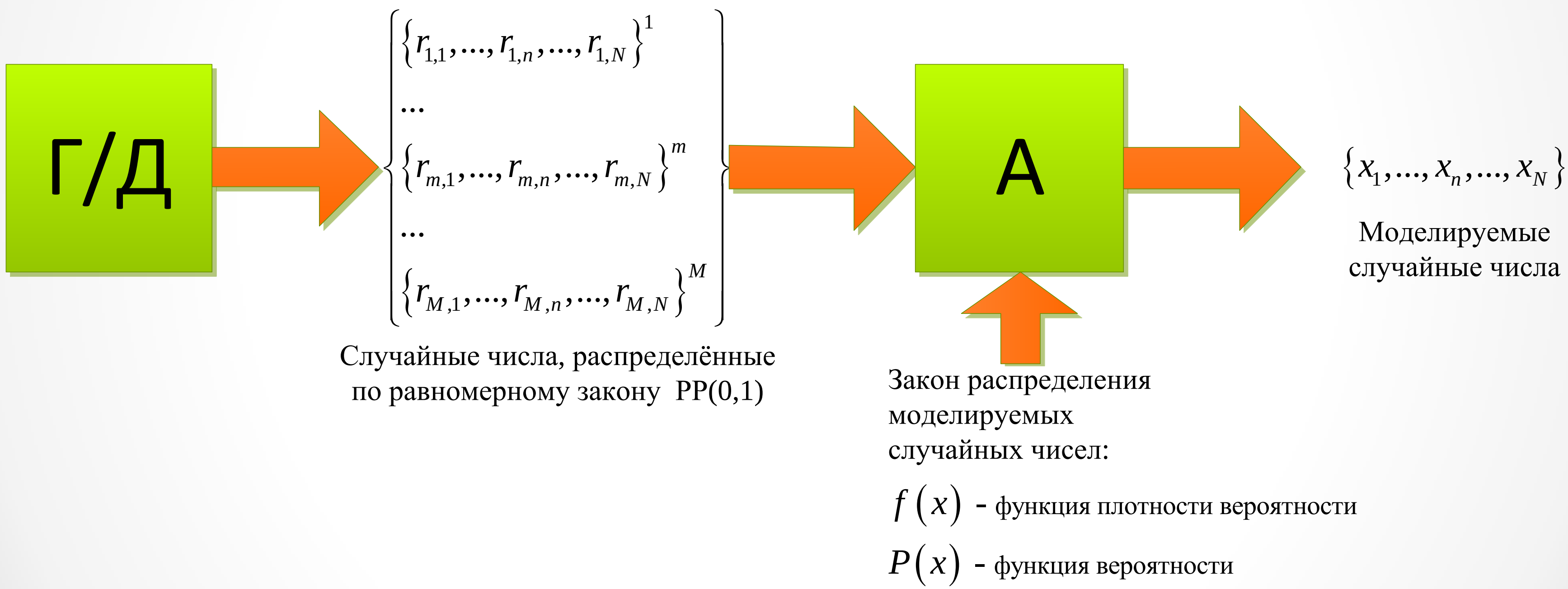
- Исследование закона распределения частоты наблюдаемых значений в интервале наблюдаемых данных производят с помощью построения гистограммы



Для построения гистограмм разбивают наблюдаемый интервал случайных значений на равные отрезки (карманы) и подсчитывают количество попаданий наблюдаемых значений в карманы: $\{\ldots, \nu_k, \ldots\}|\mathrm{k} = 1,2, \ldots \mathrm{K}$, где K – число карманов.

Общая схема алгоритма моделирования случайных чисел

На схеме представлена технология преобразования закона распределения СЧ - ПЗСЧ



- Г/Д – генератор (датчик) случайных чисел, распределённых по равномерному закону;
- А – алгоритм, воспроизводящий метод моделирования случайных чисел, требуемых при решении задачи.

Случайные процессы и статистические распределения

- В стохастических системах осуществляются случайные процессы.
- Случайный процесс характеризуется непредсказуемой последовательностью событий, которые математически можно связать с некоторой последовательностью случайных значений, взаимно однозначно соответствующих наблюдаемому потоку событий.
- Случайные значения, порождаемые потоком случайных событий в стохастических системах, подчиняются законам статистических распределений.
- Таким образом, моделирование случайных процессов связано с моделированием последовательностей случайных чисел, которые подчиняются тем или иным законам статистических распределений.

Генерирование чисел, распределённых по равномерному закону: СЧ РР( )

- Генераторы случайных чисел по способу получения чисел делятся на:
- · Физические (аппаратные);
- · табличные;
- · алгоритмические.
- При аппаратном способе случайные числа вырабатываются электронной приставкой (генератор, датчик случайных чисел).
- Случайные числа оформлены в виде таблицы, которая хранится в оперативной памяти или на внешнем носителе.
- Случайные числа генерируются на компьютере по специальным алгоритмам.

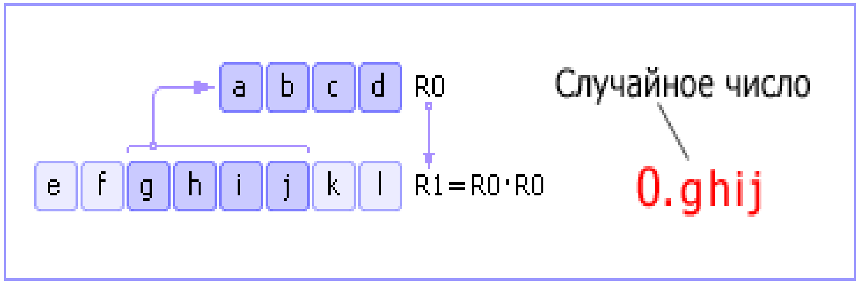
Табличный способ ГСЧ

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
...								...

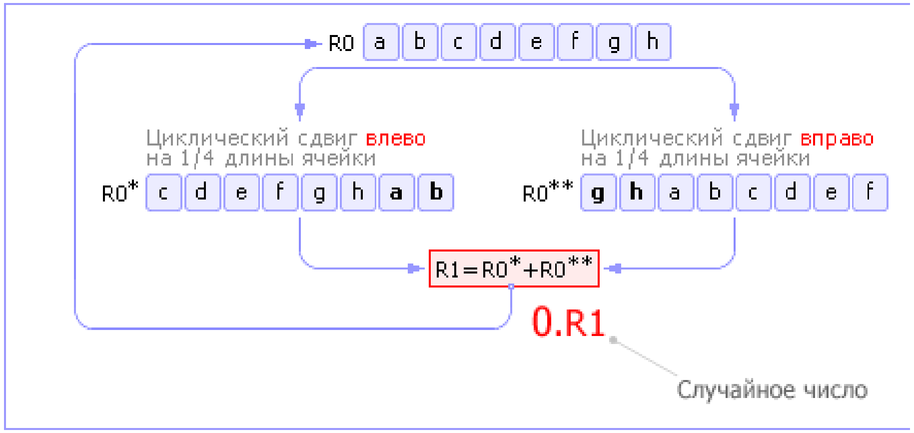
- Составить последовательность СЧ РР( ) и оценить статистические параметры выборки: выборочное среднее, выборочная дисперсия, выборочное среднее квадратическое отклонение.
- Сравнить с точными характеристиками рр СЧ на $[0,1]$

Алгоритмические способы ГСЧ

Битовые способы получения СЧ РР( )



- метод серединных квадратов



- метод перемешивания

Алгоритмические методы ГСЧ

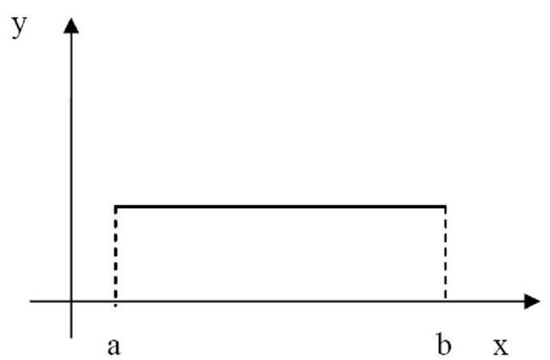
- **Линейный конгруэнтный метод**
- В линейном конгруэнтным методе случайное число вычисляется по следующей рекуррентной формуле: $X_{n+1} = (aX_n + c) \ mod \ m$, где — *m* модуль (*m*>0), — *a* множитель $(0 \leq a < m)$, c — приращение $(0 \leq c < m)$, X_0 — начальное значение, которое также иногда называют зерном (от англ. seed) $(0 \leq X_0 < m)$
- Дает при удачно подобранных коэффициентах и больших *m* достаточно непредсказуемые псевдослучайные числа
- не обладает криптографической стойкостью, так как, зная четыре подряд идущих числа, криптоаналитик может составить систему уравнений, из которой можно найти *a*, c и *m*
- Задать значения X_0, m, a, c:
- Рассчитать последовательность случайных чисел – выборку
- Проверить статистические характеристики: диапазон СЧ $\in [0, m-1]$

Алгоритмические методы ГСЧ

- Генератор псевдослучайных чисел на основе **алгоритма BBS**
- 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 $x_0 = x^2 mod \mathrm{M}$. x_0 называется стартовым числом генератора.
- На каждом n-м шаге работы генератора вычисляется $x_{n+1} = x_n^2$ mod M. Результатом n-го шага является один (обычно младший) бит числа x_{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

Статистические характеристики РР СЧ

$$f(x) = \begin{cases} \dfrac{1}{b-a} & при \quad a < x \le b \\ 0 & вне\ отрезка \end{cases}$$



- Плотность равномерного непрерывного распределения случайных чисел в интервале $x \in [a, b]$
- График функции плотности равномерного непрерывного распределения случайных чисел в интервале $x \in [a, b]$

Равномерное распределение:

$$\mathrm{PP}(a, b): F(x) = \int_a^{x<b} f(x)dx = \frac{x-a}{b-a}\ ;\ x \in [a, b]$$

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

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

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

Критерий качества ГСЧ рр

- **ГСЧ должен выдавать близкие к следующим значения статистических параметров, характерных для равномерного случайного закона для** $\text{РР СЧ} \in [\mathbf{0}, \mathbf{1}]$**:**

$$m_\gamma = \frac{\sum_{i=1}^{n} r_i}{n} \approx 0.5$$ — математическое ожидание;

$$D_\gamma = \frac{\sum_{i=1}^{n} (r_i - m_\gamma)^2}{n} \approx 0.0833$$ — дисперсия;

$$\sigma_\gamma = \sqrt{D_\gamma} \approx 0.2887$$ — среднеквадратичное отклонение.

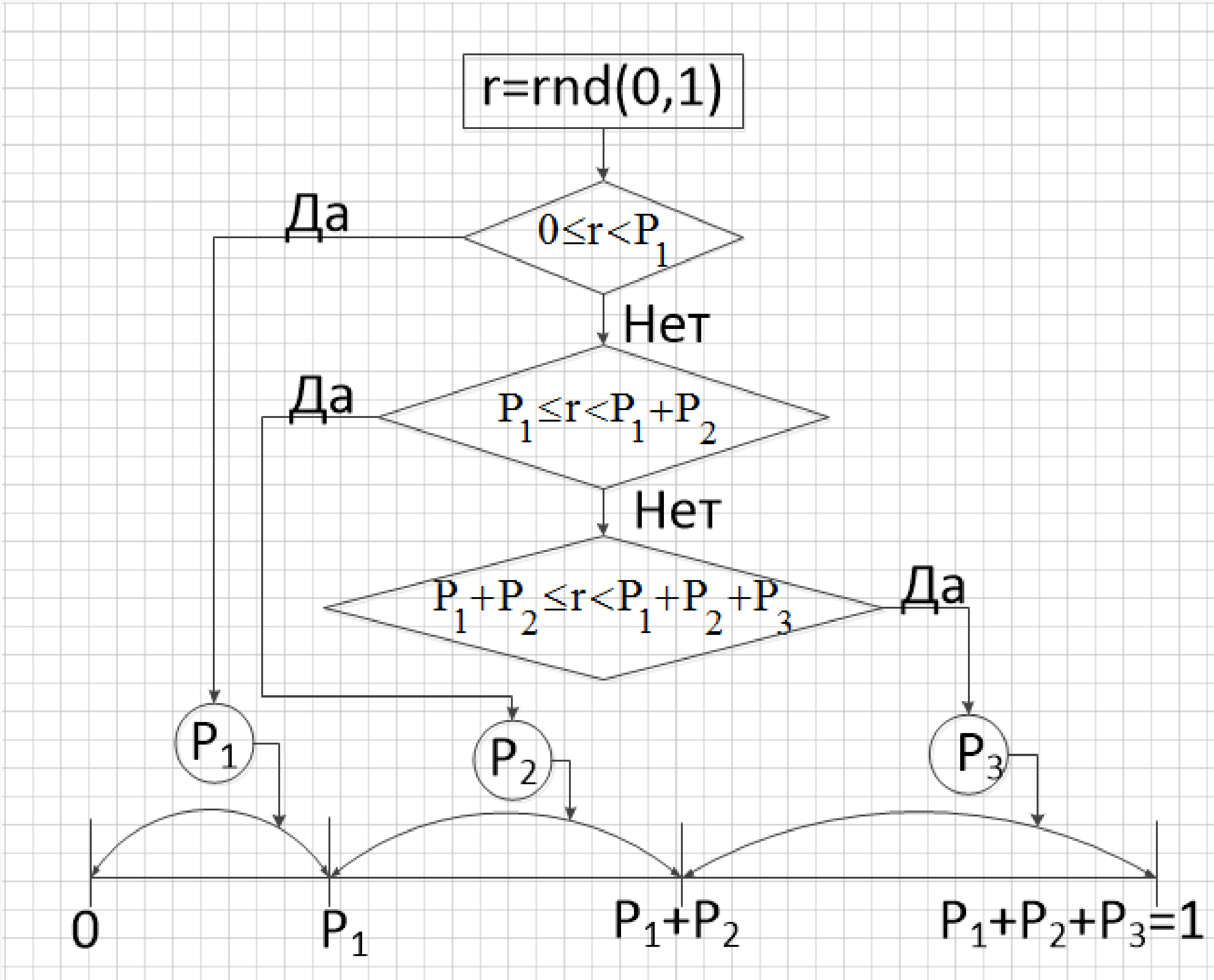
- **В общем случае качество получаемой с помощью ГСЧ выборки (последовательности) случайных чисел на соответствие требуемому закону распределения проверяют с использованием статистических критериев. Проверка осуществляется на основе доверительной оценки справедливости предложенных статистических гипотез**

Моделирование СЧ с произвольным законом распределения
Общие методы

- С помощью случайных равномерно распределенных чисел можно моделировать случайные события, составляющие конечные или счётные множества исходов и наступающие с заданной вероятностью, – **дискретные распределения**
- 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 ξ_i
- 2) проверкой неравенства $\xi_i \leq p$ (1)
- устанавливается принадлежность этого числа отрезку $[0, p]$.
- Если число ξ_i удовлетворяет неравенству (1), говорят, что событие А наступило, в противном случае — не наступило.

Моделирование дискретных случайных чисел

- ГСА: Метод «розыгрыша жребия» (ГСА-графическая схема алгоритма)



Пример реализации метода «жребия»

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:	1	2	3
	0,5	0,2	0,3
А: Метод «жребия»	$\{x\} = \{1; 1; 3; 3; 1; 3; 2; 3; 1; 1\}$		

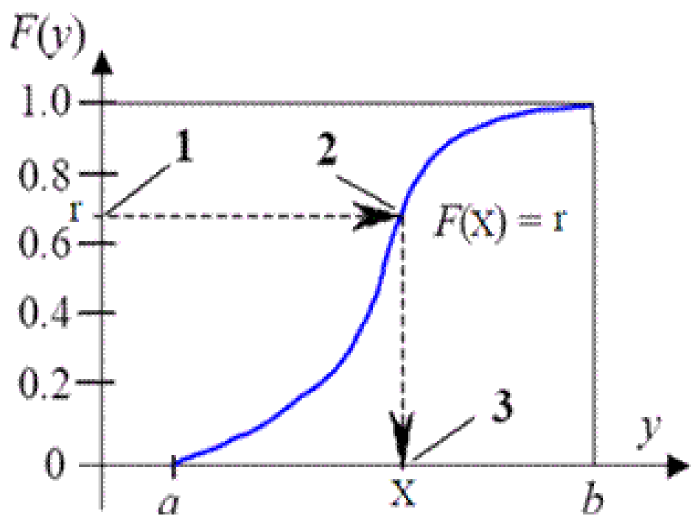
- Пояснения:
- Разыгрывается 10 исходов случайной реализации с заданным в таблице законом.
- Для этого получают (генерируют) порождающую последовательность равномерно распределённой величины РР(0,1): $\{r\}$.
- Затем, пользуясь методом «жребия», получаем выходную последовательность случайных исходов моделируемого процесса.
- Количество исходов связано с числом чисел в $\{r\}$.

Моделирование СЧ с произвольным законом распределения
Общие методы

- Метод моделирования **непрерывно распределённых случайных величин**, связанных с бесконечным множеством случайных исходов
- Требуется получить случайные числа y_i являющиеся возможными значениями случайной величины η с законом распределения, заданным функцией плотности $f(y)$ или функцией вероятности $F(y)$:
$f(y) = \frac{dF(y)}{dy}$; $F(y) = \int_{-\infty}^{y} f(y)dy$
- Случайная величина $\mathrm{F}(\eta) = \xi$, являющаяся значением интеграла от плотности распределения $f(y)$:
$\mathrm{F}(\eta) = \int_{-\infty}^{\eta} f(y)dy = \xi$,
- и определяющая величину вероятности нахождения значения случайной величины $\mathrm{y} \leq \eta$, распределена равномерно в интервале (0,1)

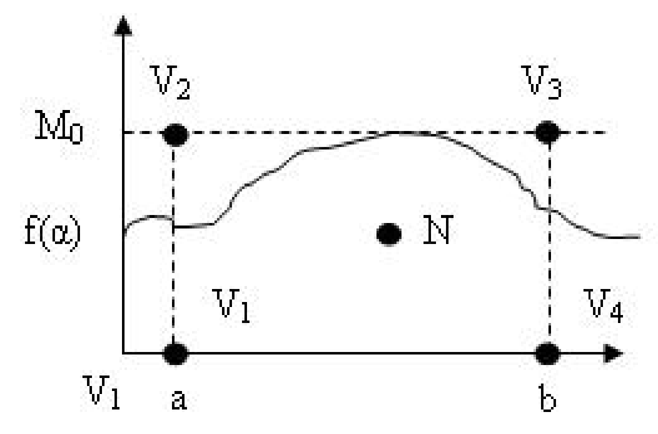
Моделирование СЧ с произвольным законом распределения
Общие методы

- Соотношением $\int_{-\infty}^{\eta} f(y)dy = F(\eta) = \xi$ можно воспользоваться получения случайных чисел с заданным законом распределения.
- **Решение интегрального уравнения:** $\int_{-\infty}^{\eta} f(y)dy = F(\eta) = \xi$ относительно η при заданных значениях ξ, которые получают с помощью ГСЧ рр: $\xi \in [0,1]$, – называемое **методом обратных функций** проиллюстрировано на графике: $r \to \xi; X \to y = \eta$.



Моделирование СЧ с произвольным законом распределения Общие методы

- Можно разыграть **случайное** число, **непрерывно** распределённое по заданному закону ***методом усечения Неймана***
- 1. **Ограничим** интервал распределения: сделаем его конечным. Пусть x – случайная величина распределенная на интервале (a, b) и моделируемая плотность сверху ограничена значением моды M_0.
- 2. **Генерируем** 2 значения: r_1, r_2 – равномерно-распределенной в интервале [0; 1] случайной величины r.
- 3. На плоскости $f(y)$ и y **отложим случайную точку** $(\cdot)N$ с координатами (α, β): $N(\alpha, \beta)$: $\alpha = a + r_1(b - a)$ и $\beta = r_2 M_0$.
- Точка находится внутри прямоугольника.



4. Если $(\cdot)N$ лежит под кривой, т.е. β<f(α), то **разыгранное полученное значение** x считается равным α: $X \to \alpha$. Если $(\cdot)N$ лежит над кривой, то пара r_1 и r_2 отбрасывается и генерируется новая пара значений r_3 и r_4, и т.д.

Статистические оценки законов распределения

- **Оценка качества** получаемых случайных чисел **на соответствие заданным законам распределения** производится путем **определения выборочных оценок** и последующего их **сравнения** с **теоретическими параметрами распределения**, см. слайд 5.
- Выборочные:
- **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**, см. слайд 6. Это исследование можно проводить с помощью построения гистограмм, или проверкой статистических гипотез, например, по критерию Пирсона.

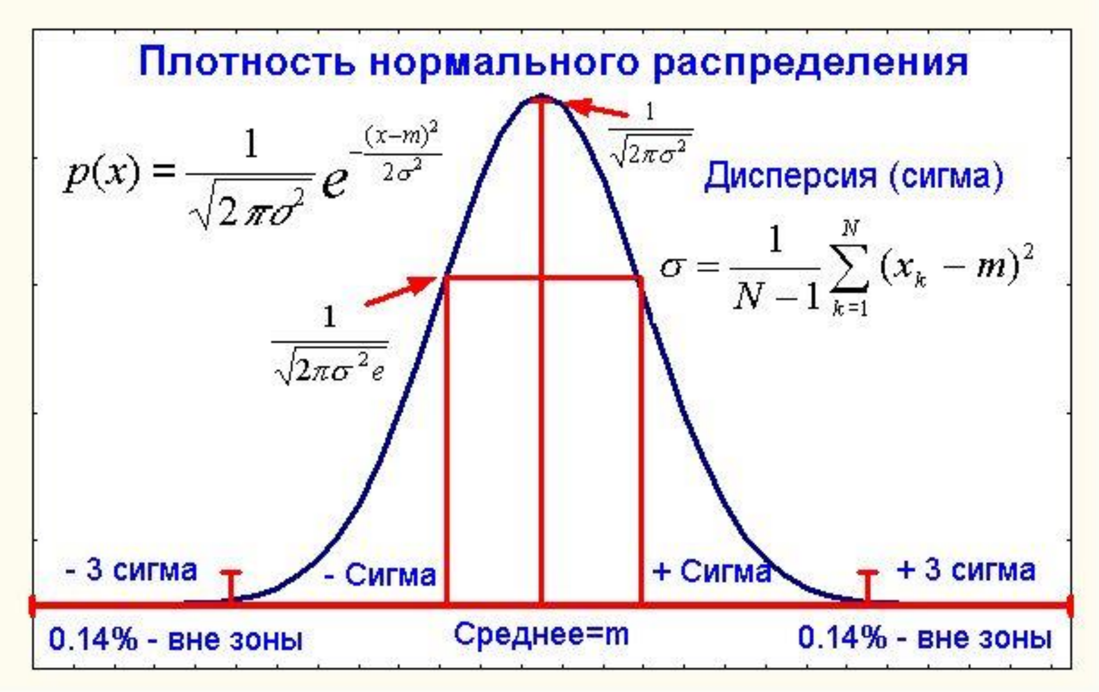
Моделирование СЧ с произвольным законом распределения
Частные методы

- **Метод генерации нормально распределенных чисел, использующий центральную предельную теорему**
- Допустим, что нам надо в целях имитации получить ряд случайных чисел *x*, распределенных по нормальному закону с заданными математическим ожиданием m_x и среднеквадратичным отклонением σ_x.
- Сложим *n* случайных чисел, используя стандартный ГСЧ рр [0; 1] : $V = \sum_{i=1}^{n} r_i$
- Согласно ЦПТ числа *V* образуют ряд значений, распределенный по нормальному закону. Эти числа тем лучше описывают нормальный закон, чем больше параметр *n*. На практике *n* берут равными 6 или 12. Заметим, что закон распределения чисел *V* имеет математическое ожидание $m_V = n/2$, $\sigma_V = \sqrt{\frac{n}{12}}$. Поэтому он является смещенным относительно заданного произвольного.

Моделирование СЧ с произвольным законом распределения
Частные методы

- С помощью формулы $z = (V - m_V)/\sigma_V$ нормализуем этот ряд. Получим нормализованный закон нормального распределения чисел Z. То есть $m_z = 0$, $\sigma_z = 1$.
- Формулой (сдвиг на m_x и масштабирование на σ_x) преобразуем ряд Z в ряд x: $x = z \cdot \sigma_x + m_x$.
- Промоделировать СЧ нр, используя ГСЧ рр компьютера
- Сравнить выборочные статистические характеристики с характеристиками распределения.
- **Метод генерации нормально распределенных чисел - Метод Бокса-Мюллера**
- Совсем простым методом получения нормализованных нормальных чисел является метод Мюллера, использующий формулы: $Z = \sqrt{-2Ln(r_1)}cos(2\pi r_2)$, где r_1 и r_2 — случайные числа из $ГСЧ_{рр}$ [0; 1].
- Можно также воспользоваться аналогичной формулой: $Z = \sqrt{-2Ln(r_1)}sin(2\pi r_2)$, где r_1 и r_2 — случайные числа из $ГСЧ_{рр}$ [0; 1].

Статистические характеристики НР СЧ



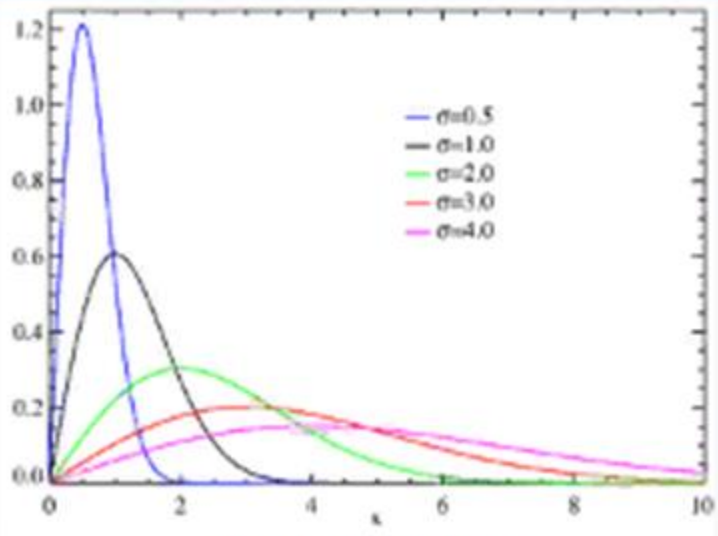
- График функции плотности непрерывного нормального распределения случайных чисел на числовой оси:
- $x \in (-\infty, \infty)$

Математическое ожидание: $m_x = m$

Дисперсия: ${\sigma_x}^2 = \sigma^2$

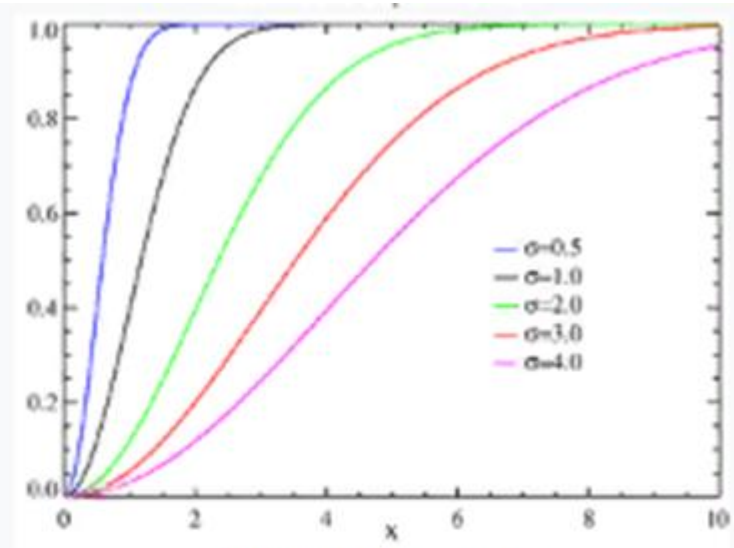
Среднее квадратическое отклонение: $\sigma_x = \sigma$

Распределение Рэлея:
Функция плотности и функция вероятности



Плотность вероятности

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	$(2 - \pi/2)\,\sigma^2$



Функция распределения

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	$1 + \sigma t\, e^{\sigma^2 t^2/2}\sqrt{\frac{\pi}{2}}\left(\operatorname{erf}\left(\frac{\sigma t}{\sqrt{2}}\right) + 1\right)$
Характеристическая функция	$1 - \sigma t e^{-\sigma^2 t^2/2}\sqrt{\frac{\pi}{2}}\left(\operatorname{erfi}\left(\frac{\sigma t}{\sqrt{2}}\right) - i\right)$

Задания

- Результаты оформить в формате документа:
- Описать методические основы
- Представить условие задачи
- Описать шаги решения
- Представить результаты
- Сделать выводы
- **Задание 1:**
  - Смоделировать выборку (последовательность не менее 20-ти) равномерно распределённых случайных чисел – СЧ РР(0,1) любым способом, см. слайды 9-15.
  - Проверить качество полученного распределения, руководствуясь сведениями со слайдов 15,5-6,22.
- **Задание 2:**
  - Промоделировать методом ЦПТ и Бокса-Мюллера случайные числа (последовательность не менее 20-ти) , распределённые по нормальному закону, используя компьютерный ГСЧ рр(0,1) и заданные значения m_x и σ_x (значения выбираем произвольно).
  - Сравнить выборочные статистические характеристики с характеристиками распределения, см. слайды 15,5-6,22.

Задания

- **Задание 3:**
- Смоделировать методом усечения - Неймана случайные числа (последовательность не менее 20-ти) по закону распределения, заданного в ограниченной области $y \in [a, b]$ плотностью распределения:
- $f(y) = \frac{3\sqrt{A}}{4C\sqrt{C}}(C - A(y-m)^2)$, где $\mathrm{A} > 0, \mathrm{m} > 0, C > 0$
- Задать числа A, C, m ; вычислить $a = m - \sqrt{\frac{C}{A}}, b = m + \sqrt{\frac{C}{A}}$ и использовать СЧ рр, полученные ранее
- Сравнить выборочные статистические характеристики с характеристиками распределения, см. слайды 15,5-6,22.
- Математическое ожидание: $m_y = m$
- Дисперсия: $\sigma_y{}^2 = \bar{y}^2 - m_y{}^2$, где средний квадрат случайной величины:
- $\bar{y}^2 = \frac{20Cm^2 + 4\frac{C^2}{A}}{20C}$

Задания

- **<u>Задание 4:</u>**
- Смоделировать методом обратных функций случайные числа (последовательность не менее 20-ти) , распределённые по закону Рэлея.
- Функция плотности вероятности:
- $f(x) = \frac{x}{\sigma^2} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \forall x \in [0,\infty)$. Параметр σ выбираем произвольно (например, берем с графика)
- Функция вероятности:
- $F(X < x) = \int_0^x f(x)dx = 1 - e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \in [0,1] \forall x \in [0,\infty)$
- Сравнить выборочные статистические характеристики с характеристиками распределения, см. слайды 15,5-6,22.
- Математическое ожидание: $m_x = \sqrt{\frac{\pi}{2}}\sigma$
- Дисперсия: ${\sigma_x}^2 = \left(2 - \frac{\pi}{2}\right)\sigma^2$

Задания

- **Задание 5:**
- Смоделировать методом «розыгрыша жребия» случайные числа (последовательность не менее 20-ти) , распределённые по дискретному геометрическому закону.
- Член последовательности распределения вероятностей по геометрическому закону описывается по формуле:
- $P_k = q^{k-1}p \; \forall k = 1,2,\ldots,n,\ldots \in [1,\infty)$, 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 ${\sigma_K}^2 = \frac{q}{p^2}$
- Вычислить выборочные характеристики и сравнить с теоретическими