Методы моделирования
непрерывных случайных величин с заданными законами распределений

Основные понятия и определения

Случайная величина - это величина, которая в результате опыта со случайным исходом принимает то или иное значение, причем неизвестно заранее, какое именно

Все возможные значения случайной величины образуют **множество возможных значений случайной величины.**

Закон распределения случайной величины - это любое правило (таблица, функция), устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Событие – это всякий факт, который в результате опыта может произойти или не произойти.

Наиболее распространенными событиями являются следующие:

- случайная величина X принимает значение x;
- случайная величина X попадает в интервал (a, b);
- случайная величина X меньше значения x.

Несколько событий в данном опыте образуют **полную группу событий**, если в результате опыта обязательно появляется одно из них.

События называются несовместными, если они не могут появиться одновременно в данном опыте.

Различают дискретные и непрерывные случайные величины

Возможные значения дискретной величины можно перечислить.
Возможные значения непрерывной случайной величины непрерывно заполняют некоторый промежуток (конечный или бесконечный).

При моделировании случайных величин

Базис - совокупности случайных величин с заданными характеристиками для получения случайной величины на любом наперед заданном множестве возможных значений и при любом наперед заданном законе распределения.

1. В качестве такого базиса часто выбирают
равномерно распределенную случайную величину в интервале (0, 1).

2. Невозможно воспроизвести опыт с истинно случайным исходом. Поэтому вместо случайных величин используются псевдослучайные величины.

3. В различных программных продуктах Randomize(), RND(), в Python - RANDOM(). Вызов таких функций в цикле вычислительного алгоритма позволяет получить **последовательность псевдослучайных чисел.**

Пусть случайная величина - $\boldsymbol{X} \in (a, b)$

Вероятность события $\boldsymbol{F(x)} = P(X < x)$,

x – некоторое число,

$\boldsymbol{F(x)}$ **-** функция распределения случайной величины $\boldsymbol{X}$

Свойства функции распределения:

1. Функция распределения *F(x)* не убывает, т.е. при $x_2 > x_1$ $F(x_2) \geq F(x_1)$
2. $F(-\infty) = 0$, $(F(a) = 0)$.
3. $F(\infty) = 1$, $(F(b) = 1)$

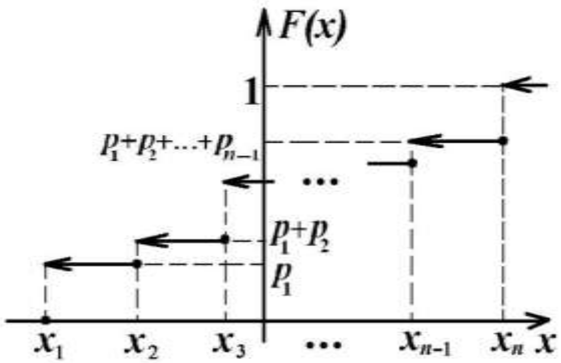


Рис.1

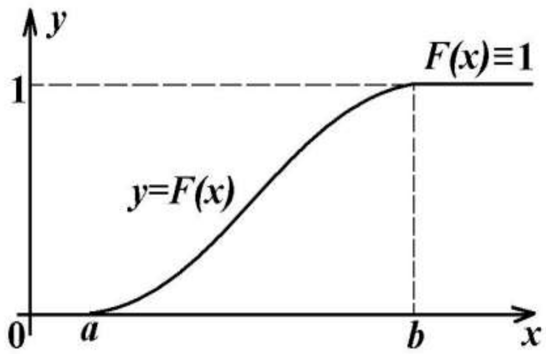


Рис.2

Плотность вероятности непрерывной случайной величины –

$w(x) = dF(x)/dx.$

Свойства плотности вероятности:

1. Плотность распределения есть неотрицательная функция $w(x) \geq 0$.
2. $\int_{-\infty}^{\infty} w(x)dx = 1$ или $(\int_{a}^{b} w(x)dx = 1)$
3. Если X – случайная величина, заданная на интервале *(a, b)*, то, если $x \in (a, b)$ иначе $w(x) = 0$. Откуда следует, что.
4. По заданной плотности распределения $w(x)$ величины X можно найти ее функцию распределения *F(x)* :

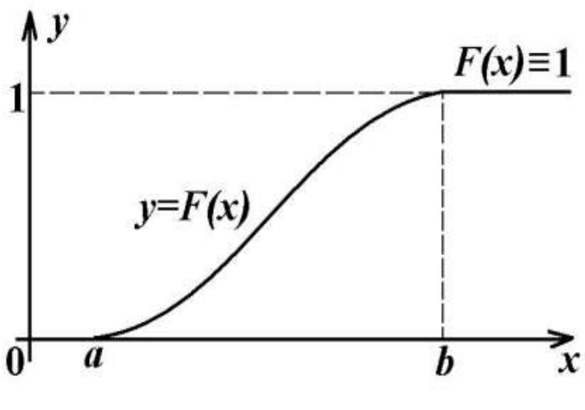
- $F(x) = 0$, при $x < a$,
- $F(x) = 1$ при $x > b$, и
- $F(x) = \int_{a}^{x} w(x)dx.$

Моделирование непрерывных случайных величин $\{x_1, x_2, \dots, x_i, \dots, x_N\}$ с заданной функцией распределения $F(x)$ методом обратной функции

$F(x),\ x \in (a,b),\ F(x) \in (0,1)$ - монотонна

Однозначная обратная функция:

$G(y),\ y \in (0,1),\ G(y) \in (a,b)$ - монотонна



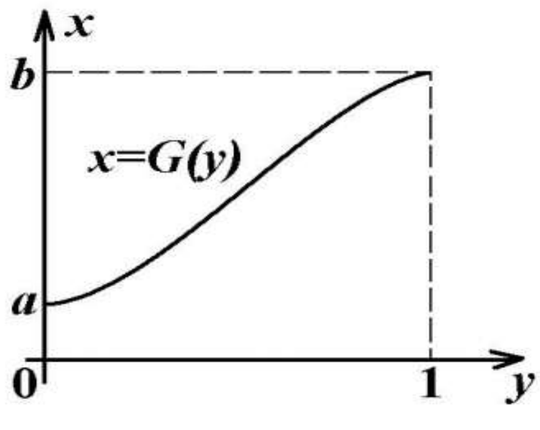


Рис.3

Пусть R – равномерно распределенная на интервале случайная величина.

Функция распределения такой величины на некотором интервале $R \in (a, b)$

$F(x) = F(R < x) = {}^{(x-a)}\!\big/_{(b-a)}$,

Плотность вероятности - $w(x) = {}^{1}\!\big/_{(b-a)}$

Среднее значение случайной величины - $m = (b + a)/2$,

Дисперсия – $D^2 = (b - a)^2/12$.

Выберем в качестве величины интервала $(\boldsymbol{a}, \boldsymbol{b})$ значениия (0, 1),

$\boldsymbol{F(x) = x, m = 1/2, D^2 = 1/12}$

Получаем выборку N независимых значений $\{r_1, r_2, \dots, r_i, \dots, r_N\}$ равномерно распределенной на интервале (0, 1) случайной величины R.

Далее для каждого равномерного случайного числа r_i находим соответствующее значение x_i через преобразование $x_i = G(r_i)$,

$G(r_i)$ – функция, **обратная** к необходимой функции распределения, т.е. $G(r_i) = F^{-1}(r_i)$.

Пример:

Сформировать релеевскую случайную величину x с заданным параметром σ и плотностью вероятности

$w(x) = \frac{x}{\sigma^2} \exp\left(-x^2 / 2\sigma^2\right), 0 \leq x < \infty.$

Релеевская функция распределения - $F(x) = 1 - \exp\left(-x^2 / 2\sigma^2\right)$.

Математическое ожидание $m = \sqrt{\pi\sigma^2/2}$.

Дисперсия - $(2 - \pi/2)\ \sigma^2$

Уравнение для нахождения x:

$$R = 1 - \exp\left(-x^2 / 2\sigma^2\right)$$

Обратная функция:

$$x_i = \sigma\sqrt{-2\ln(r_i)}$$

Формирование непрерывной случайной величины методом Неймана

Используется - для моделирования случайных величин, возможные значения которых не выходят за пределы некоторого ограниченного интервала $X \in (a, b)$.

Это - случайные величины X с усеченными законами распределения

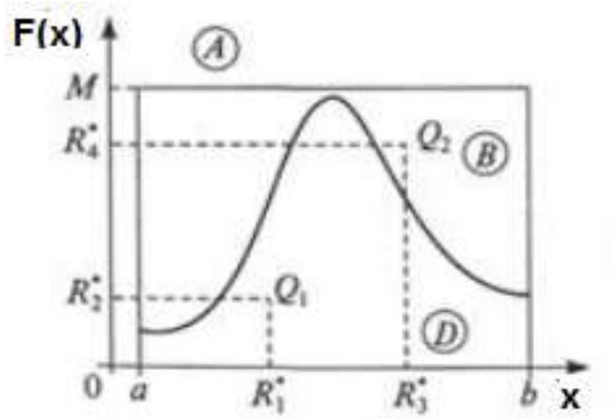


Рис.4

1. Выполняется усечение исходного распределения случайной величины:

$g(x) = c \cdot w(x), a < x < b, g(x) < M$. Таким образом кривая $g(x)$ вписана в прямоугольник

2. Формируем два равномерно распределенных числа:

$X = a + (b - a)r_i, Y = Mr_j$, - координаты случайной точки Q_i в этом прямоугольнике.

3. Если $Y \leq g(X)$, т.е. точка лежит под кривой, то $x_i \in X$. Иначе точка отбрасывается.

Пример:

сформировать случайную величину x с плотностью вероятности $\mathrm{w(x)} = \pi \cdot \cos(2\pi\mathrm{x})$, на интервале $\mathrm{x} \in [-0.25, 0.25]$.

Решение:

$\mathrm{g(x)} = \cos(2\pi\mathrm{x})$,

$\mathrm{X} = 0.25(2\mathrm{r_i} - 1)$,

$\mathrm{Y} = \mathrm{r_j}$.

Если $\mathrm{Y} < \cos(2\pi\mathrm{x})$, то $x_i \in X$

Формирование стандартных гауссовских случайных величин на основе центральной предельной теоремы

Функция распределения стандартного нормального закона: $m = 0$, $D^2 = 1$

$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)\,dt$ - не выражается через элементарные функции в конечном виде, а значит, нет возможности явно записать обратную к ней.

Плотность вероятности:

$w(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$

Центральная предельная теорема (ЦПТ) –

Закон распределения суммы независимых случайных величин приближается к нормальному закону распределения при неограниченном увеличении объема, если выполняются следующие условия:

1) все независимые случайные величины имеют конечные математические ожидания и дисперсии:

2) ни одна из этих величин по значению резко не отличается от остальных.

Тогда для формирования стандартных гауссовских случайных величин $x \to N(0,1)$ на основании этой теоремы:

1. Сложим n равномерно распределенных на интервале (0, 1) случайных величин $v = \sum_{i=1}^{n} r_i$
2. 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 - 6 или 12**. Закон распределения чисел v имеет математическое ожидание $m_v = n/2$, дисперсию $D_v^2 = n/12$ Поэтому он является смещенным относительно заданного произвольного.
3. С помощью формулы замены $x = {(v - m)}/{D}$ нормализуем этот ряд. Получим нормализованный закон нормального распределения чисел:

$x = \sqrt{12/n \sum_{i=1}^{n}(r_i - 0.5)}$ с m =0, D = 1.

Коэффициент использования датчика равен $1/n$.

Формирование гауссовского случайного числа с произвольными параметрами среднего и дисперсии $N(m, D^2)$

Преобразование вида $X = D\xi + m$ от стандартной гауссовской величины.

Его функция распределения имеет вид $F(x) = \frac{1}{\sqrt{2\pi D^2}} \int_{-\infty}^{x} \exp(-(t - m)^2/2\, D^2)\, dt$

ЗАДАНИЕ

1. Сформировать выборку непрерывных равно распределенных на интервале $(a.b)$ случайных величин методом обратных функций.

2. Сформировать выборку величин, распределенных по гауссовскому закону с параметрами $N(m, D^2)$ на основе ЦПТ

3. Методом Неймана сформировать выборку величин, распределенных по релеевскому закону заданным σ.

4. Объем выборки $N \geq 10^3$, параметры моделирования задаются при входе в программу.

Построение гистограмм

Гистограмма - это эмпирическая плотность вероятности. Она представляет собой ступенчатую фигуру в виде прямоугольников. Длина каждого прямоугольника имеет равный частотный интервал. При ее моделировании:

1. Берется сформированная по определенному вероятностному закону выборка $\bar{x} = (x_1, x_2, \ldots, x_N)$ объемом N и проводится ее упорядочивание.
2. Определяется число интервалов группировки, например, как $k = (4 \ldots 5)\lg(N)$. Обычно выбирается число $9 < k < 21$. Если распределение предполагается симметричным, то число интервалов вбирается нечетным.
3. Определяется длина и границы интервалов группировки. Для всех $x_i : a < x_i < b$ задается $a = (1 \mp 0.02)x_{min}$, $b = (1 \mp 0.02)x_{max}$. $d = 1.02(x_{max} - x_{min})$, $\Delta x = d/k$. Поправочный множитель 0.02 может быть и другим, исходя из конкретных требований задачи.
4. Границы i-ого интервала Δ_i : $\Delta_i = (a + (i-1)\Delta x, a + i\Delta x)$, $i = 1, \ldots, k$. Для односторонних распределений очевидно $a=0$.
5. Подсчитывается количество k_i элементов выборки $\bar{x} = (x_1, x_2, \ldots, x_N)$, попавших в интервал группировки Δ_i. При этом $k_i > 5 \ldots 10$.
6. Определяются частоты $\delta_i = k_i/N$ либо относительные частоты $\omega_i = k_i/N\Delta x = \delta_i/\Delta x$ и строится диаграмма столбцов высотой δ_i или ω_i , для всех интервалов группировки $i = 1 \ldots k$.

Диаграмма меньше всего отличается от теоретической плотности вероятности в центре каждого из интервалов группировки $(a + (i - 0.5)\Delta x)$

<u>Построение полигона накопленных частот – это эмпирическая функция распределения</u>

Накопленные относительные частоты на i-ом интервале вычисляются как накопленная частота на всех предыдущих интервалах, поделенная на объем выборки. Поведение эмпирической функции распределения описывается ступенчатой функцией.

1. Берется сформированная по определенному вероятностному закону выборка $\bar{x} = (x_1, x_2, \ldots, x_N)$ объемом N и, также как и при построении гистограммы определяется число интервалов.
2. Подсчитывается количество элементов k_q выборки, попавших в интервал $(a, a + q\Delta x)$. Очевидно, что $k_q = \sum_{i=1}^{q} k_i$, $q = 1 \ldots k$.
3. На каждом шаге определяются выборочные вероятности: $F_q = k_q/N = \sum_{i=1}^{q} \delta_i$
4. Далее строится ступенчатая диаграмма, высота которой равна F_q на каждом интервале $\Delta_q = (a + (q-1)\Delta x, a + q\Delta x)$.

Полигон меньше всего отличается от теоретической функции распределения в конце интервала группировки

ЗАДАНИЕ

Для сформированных выборок построить

- гистограмму
- полигон накопленных частот.

На графики нанести кривые теоретических распределений, рассчитанные для выбранных при формировании значений параметров.

Расчет выборочных моментов

Набор значений $\bar{x} = (x_1, x_2, \ldots, x_N)$ измеряемой величины x, полученный при N-кратном повторении эксперимента называется реализацией выборки.

Первый выборочный момент - приближенное измерение математического ожидания m случайной величины x, для выборки $\bar{x}$ объемом N:

$m_x = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Второй выборочный центральный момент - приближенное значение дисперсии D^2 величины x. Если нам заранее известно математическое ожидание m случайной величины x, то выборочная дисперсия может быть посчитана по значениям $\bar{x}$ как:

$D_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2$.

При неизвестном заранее среднем значении вначале по выборке следует рассчитать m_x:

$m_x = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Затем, подставив в формулу вычисленное выборочное среднее, можно рассчитать выборочную дисперсию как:

$D_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m_x)^2$.

ЗАДАНИЕ

Для сформированных выборок с заданными законами распределений рассчитать выборочные средние и дисперсии случайных величин x. Сравнить расчетные данные с теоретическими значениями