1 Домашняя работа №1

1.1 Задание 1

С помощью $R[0,1]$ реализовать датчик распределения $Bi(5, \frac{1}{2})$. По смоделированным выборкам длины 100 и 100 000 построить на одном рисунке графики ряда частот и теоретический график. По выборкам вычислить выборочные среднее и дисперсию, сравнить их с теоретическими значениями.

1.1.1 Решение

Для того, чтобы получить распределение $Bi(5, \frac{1}{2})$ из генератора $R[0,1]$, возьмем 5 генераторов и полим выборку размерности 5, то есть $X = (X_1, X_2, X_3, X_4, X_5)$, где X_i - случайная величина, причем $X_i \sim R[0,1]$.

Чтобы выполнялось второе условие, то есть "вероятность успеха $\frac{1}{2}$ выберем на отрезке $[0,1]$ точку $\frac{1}{2}$, тогда отрезок будет разделен на две равные части. Теперь будем предпологать, что все значения на полуинтервале $[0, \frac{1}{2})$ соответствуют нулю, а $[\frac{1}{2}, 1]$ единице. Вернем сумму полученных генераторов, что будет той самой случайной величиной $Y \sim Bi(5, \frac{1}{2})$

Рассмотрим графики полученного распределения.

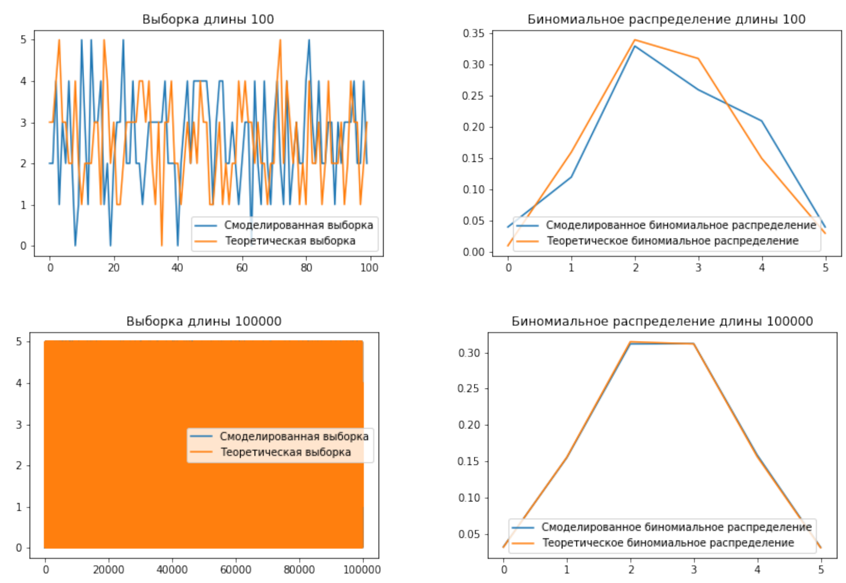


Рис. 1: Распределение Бернулли

Вычислим выборочное среднее:

$$X = (X_1, X_2, ..., X_n); n = \{100, 100000\}; E_X = \sum_{i=0}^{n} \frac{1}{n} * X_i$$

И дисперсию:

$$X = (X_1, ..., X_n); n = \{100, 100000\}; D_X = E_(X^2) - (E_X)^2$$

Для выборки длины 100:

- Выборочное среднее: 2.6, теоретическое: 2.52
- Дисперсия смоделированной выборки: 1.38, теоретическое: 1.1095999999999995

Для выборки длины 100000:

- Выборочное среднее: 2.50395, теоретическое: 2.498
- Дисперсия смоделированной выборки: 1.2560643974999994, теоретическое: 1.2472359999999991

1.2 Задание 2

С помощью $R[0, 1]$ реализовать датчик стандартного распределения Коши. По смоделированным выборкам длины 100 и 100 000 построить на одном рисунке теоретических и эмпирических функций распределения. На одном рисунке построить гистограммы смоделированных выборок (число и положение разрядов выбрать самостоятельно, но осмысленно!) и теоретическую плотность распределения. Вычислить выборочные медианы и сравнить их с теоретическим значением.

1.2.1 Решение

Для моделлирования случайных величин некоторого распределения через генератор равномерного распределение рассмотрим функцию, обратную функции распределения Коши:

$$F^{-1}(X) = \mathrm{tg}\left(\pi * (X - \frac{1}{2})\right), X \sim R[0, 1]$$

По решению задачи №7 полученная случайная величина $Y = F^{-1}(X)$ распределена по закону Коши, то есть $Y \sim C(0, 1)$. Для вычисления эмперической функции распределения используем формулу:

$$F_n(x) = \frac{1}{n} * \sum_{k=1}^{n} I(x - X_k)$$

$$I(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ 1 & \text{при } x < 0. \end{cases}$$

Для построения теоретического распределения воспользуемся распределением $C(0,1)$:

$$F(x) = \frac{1}{\pi} * \operatorname{arctg}(x) + \frac{1}{2}$$

Рассмотрим полученные графики для выборки длины 100 и 100000:

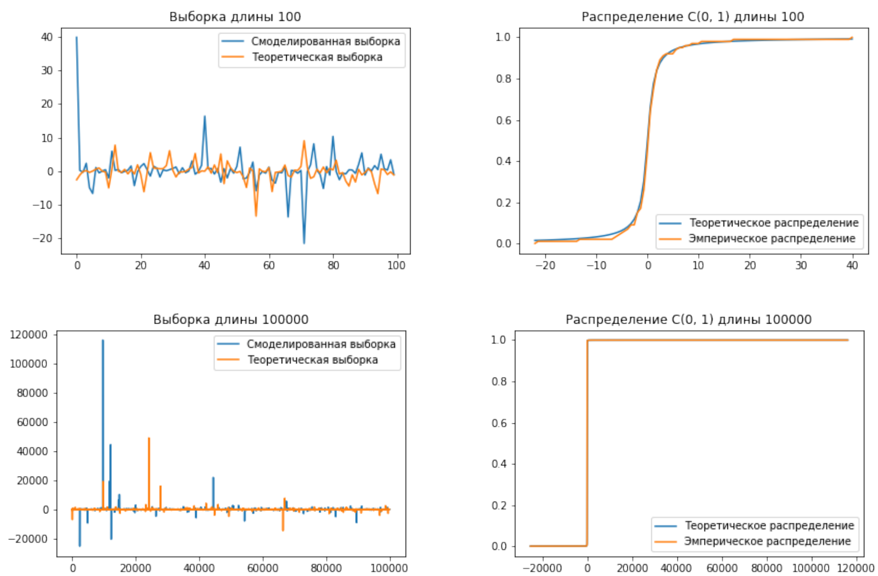


Рис. 2: Распределение Коши

Построим гистограмму и плотность распределения Коши. Для плотности справедлива формула:

$$f(x) = \frac{1}{\pi * (1 + x^2)}$$

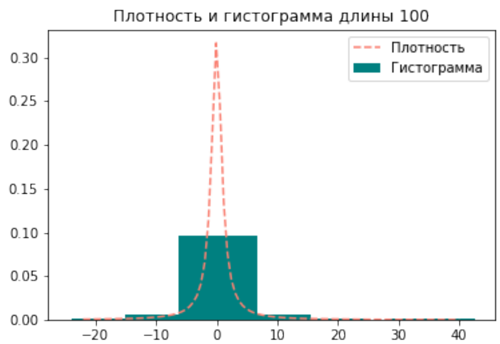
Чтобы построить гистограмму, определим число разрядов и выберем отрезок так, чтобы выполнялось условие нормировки. Данный вопрос рассмотрен в задаче №3, поэтому выберем отрезок так, что левая границы соответствует минимуму из полученной выборки, а правая - максимуму:

$$X = (X_1, ..., X_n), [a, b] = \left[\min_{i \in \overline{1,n}} X_i, \max_{i \in \overline{1,n}} X_i\right]$$

Для определения числа разрядов гистограммы воспользуемся правилом Стёрджеса. Число разрядов n определяется через размер выборки N:

$$n(N) = 1 + \lfloor \log_2 N \rfloor$$

Используя полученные значения, построим плотность и гистограмму распределения Коши:



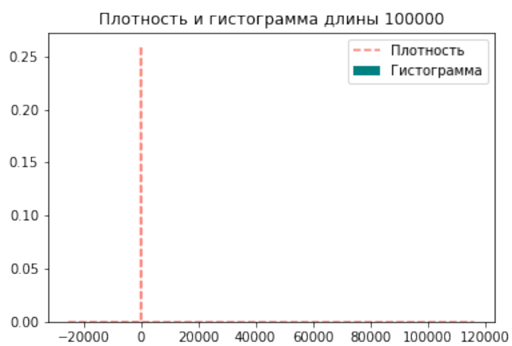


Рис. 3: Плотность и гистограмма распределения Коши

Вычислим выборочную медиану и сравним значение с теоретическим аналогом. Теоретическое значение получается при подстановке в обратную функцию распределения число $\frac{1}{2}$:

$$M = F^{-1}(\frac{1}{2})$$

- Для выборки длины 100 выборочная медиана: 0.21787135837578478; Теоретическое значение: 0
- Для выборки длины 100000 выборочная медиана: -0.0005758960784728971; Теоретическое значение: 0

1.3 Задание 3

В условиях задачи 8 выбрать произвольную кусочно-постоянную плотность с 5 разрядами. помощью $R[0,1]$ реализовать соответствующий генератор. По смоделированным выборкам длины 1000 и 100 000 построить на одном рисунке теоретических и эмпирических функций распределения. По выборкам вычислить выборочные среднее и дисперсию, сравнить их с теоретическими значениями. На одном рисунке построить гистограммы смоделированных выборок и теоретическую плотность распределения. Разряды гистограмм выбирать двумя способами: 1) совпадающими с теоретическими разрядами распределения, 2) разбивающими теоретические разряды пополам. По выборкам вычислить выборочные среднее и дисперсию, сравнить их с теоретическими значениями.

1.3.1 Решение

Чтобы смоделировать кусочно-постоянную плотность распределения, рассмотрим отрезок $[0,1]$. На данном отрезке возьмем 5 точек, не равных нулю, и ноль. Например:

$$[0,1] \to \left[0, \frac{5}{16}\right) \cup \left[\frac{5}{16}, \frac{7}{16}\right) \cup \left[\frac{7}{16}, \frac{10}{16}\right) \cup \left[\frac{10}{16}, \frac{11}{16}\right) \cup \left[\frac{11}{16}, 1\right)$$

Теперь мы получили разбиение отрезка на 5 частей. Выберем на каждом плотность так, чтобы выполнялось условие нормировки. Положим, что:

$$\mathrm{f(x)} = \begin{cases} 0 & \text{при } x < 0, \\ 1 & \text{при } 0 \leq x < \frac{5}{16}, \\ \frac{1}{2} & \text{при } \frac{5}{16} \leq x < \frac{7}{16}, \\ \frac{2}{3} & \text{при } \frac{7}{16} \leq x < \frac{10}{16}, \\ 2 & \text{при } \frac{10}{16} \leq x < \frac{11}{16}, \\ \frac{6}{5} & \text{при } \frac{11}{16} \leq x < 1, \\ 0 & \text{при } 1 \leq x. \end{cases}$$

Проверим условие нормировки:

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{5}{16} \times 1 + \frac{2}{16} \times \frac{1}{2} + \frac{3}{16} \times \frac{2}{3} + \frac{1}{16} \times 2 + \frac{5}{16} \times \frac{6}{5} = 1$$

Поэтому выбранное разбиение верное и у полученной случайной величины кусочно-постоянная плотность. Чтобы смоделировать генератор полученной случайной величины используем два генератора

- Первый генератор выдаст точку на отрезке $[0,1]$. Используем наше разбиение отрезка на 5 частей и определим, какому множеству принадлежит полученная точка

- Второй генератор покажет на какую позицию разбиения попадет точка. Уже определено, какому отрезку принадлежит данная точка, известна длина отрезка и значение левой границы. Найденные значения подставим в формулу случайной величины:

$$X() = \sum_{i=1}^{5} |X_i - X_{i-1}| * I_{\Delta[X_{i-1}, X_i]}, X_0 = 0$$

Например, если после работы первого генератора точка находится на множестве $\left[0, \frac{5}{16}\right)$, то позиция точки принадлежит этому же интервалу, тогда второй генератор дает число от 0 до 1, которое умножается на длину замыкания данного множества, то есть на $\frac{5}{16}$, затем прибавляется начало множества - точка 0.

По определению распределения получим, что

$$F_X(x) = \int_{-\infty}^{x} f(x)dx$$

Так как плотность кусочно-постоянная, то функция распределения будет линейной функцией с множителями перед x, соответствующими значению плотности на данном множестве, сложенная с константой, равной $\int_{-\infty}^{a_l} f(x)dx$, где a_l - левая граница данного множества.

Так как в точке $\frac{7}{16}$ плотность меняется с $\frac{1}{2}$ на $\frac{2}{3}$, для большей наглядности добавленая вспомогательная функция $y(x) = \frac{2}{3} * x + \frac{1}{12}$, показывающая, что перелом в точке $\frac{7}{16}$, хоть и незначительный, есть.

Для эмперической функции распределения используем уже упомянутую в предыдущих заданиях этого документа формулу. Получим:

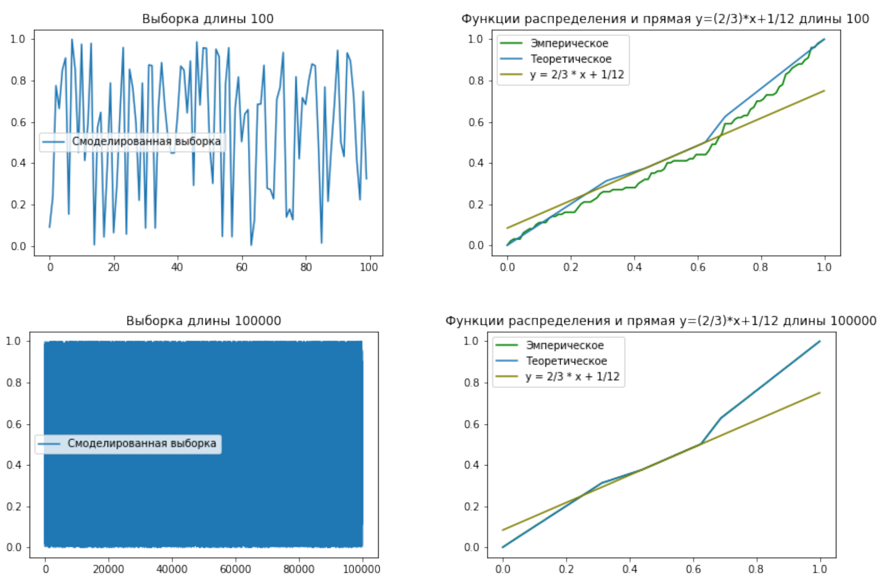


Рис. 4: Распределение с кусочно-постоянной плотностью

Сравним полученные выборочные среднее и дисперсию, сравним их с еоретическими значениями:

- Для выборки длины 100 выборочное среднее составило: 0.4252; теоретическое - 0.4632670454545454

- Для выборки длины 100000 выборочное среднее составило: 0.4633432522; теоретическое - 0.462890996097461

- Для выборки длины 100 выборочная дисперсия составила: 0.07501495999999996; теоретическая - 0.07843229629324083

- Для выборки длины 100000 выборочная дисперсия составила: 0.0766235435949392; теоретическая - 0.07666826120296805

При построении гистограммы использовалось два способа, как описано в условии задачи. Плотность данной случайной величины, как следует из требования задачи, кусочно-постоянная функция с значениями указанными ранее на множествах описанных выше. Число разрядов фиксированно и в 1-ом случае равно 5(по условию задачи), во втором случае равно 10.

Для построения гистограммы использовалась описанная ранее в заданиях этого документа формула. Для выбора отрезка, на котором строится гистограмма, использовалась ранее описанная идея с максимумом и минимумом выборки. Получим:

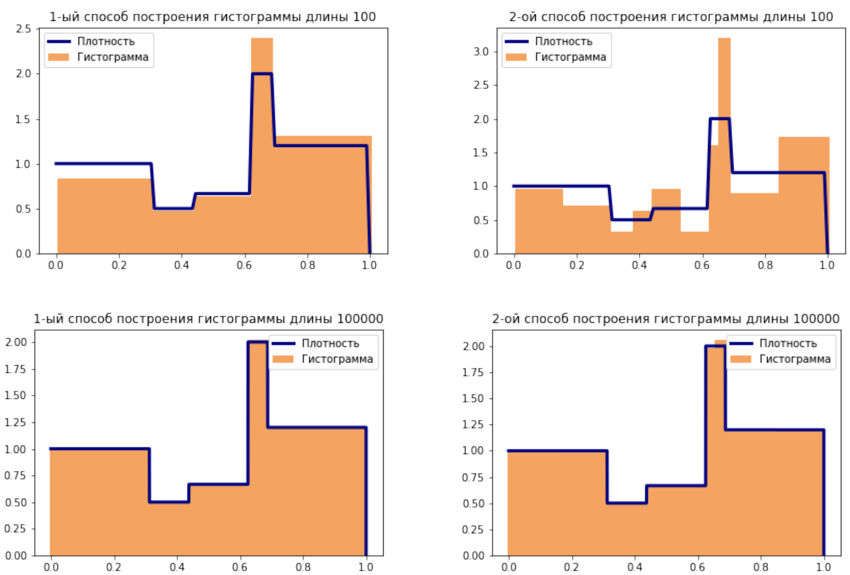


Рис. 5: Плотность и гистограмма распределения с кусочно-постоянной плотностью

1.4 Задание 4