Лабораторная работа №2

Владимир Медведев, группа M3238

Задача №1, вариант 3

Заметим, что X - дискретная случайная величина. Тогда её можно представить только как сумму дискретных. Дадим название с.в., на которые мы раскладываем, пусть X = Y + Z, Y и Z независимы, одинаково распределены и дискретны. Какие значения могут принимать Y и Z с ненулевой вероятностью? Обозначим их множество как D. Рассмотрим случай Y = Z

$$x \in D \iff 2x \in \{-1, 0, 1\} \iff x \in \{-0.5, 0, 0.5\}$$

Выразим p_1, p_2, p_3 с помощью распределения Y и Z. Обозначим $q_1 = P(Y = -0.5), q_2 = P(Y = 0), q_3 = P(Y = 0.5)$

$$p_1 = q_1^2, p_2 = q_2^2 + 2q_1q_3, p_3 = q_3^2$$

При этом все другие исходы невозможны, то есть

$$q_1q_2 = q_2q_3 = 0$$

Тогда носитель Y и Z либо состоит из 1 числа, либо равен $\{-0.5, 0.5\}$. В терминах p_1, p_2, p_3 эти условия равносильны

$$p_2 = 1 \vee \sqrt{p_1} + \sqrt{p_3} = 1$$

Этих условий будет достаточно, так как при их выполнении можно выразить q.

$$p_2 = 1 \implies q_2 = 1, q_1 = q_3 = 0$$

$$\sqrt{p_1} + \sqrt{p_3} = 1 \implies q_1 = \sqrt{p_1}, q_3 = \sqrt{p_3}, q_2 = 0$$

Задача №2, вариант 2

Покажем, что (X, Y) - гауссовский вектор с нулевыми матожиданиями и дисперсией $\sigma^2 * E$. Для этого найдём совместную плотность $p_{X,Y}$ с помощью формулы

$$p_{X,Y}(x, y) = p_{R,\theta}(r(x, y), \theta(x, y)) \left| det \begin{pmatrix} \nabla r(x, y) \\ \nabla \theta(x, y) \end{pmatrix} \right|$$

Поскольку R и θ независимы

$$p_{R,\theta}(r, \theta) = p_R(r)p_\theta(\theta) = \frac{x}{2\pi\sigma^2} exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$r(x, y) = \sqrt{x^2 + y^2}$$

$$\theta(x, y) = \arctan\frac{Y}{X}$$

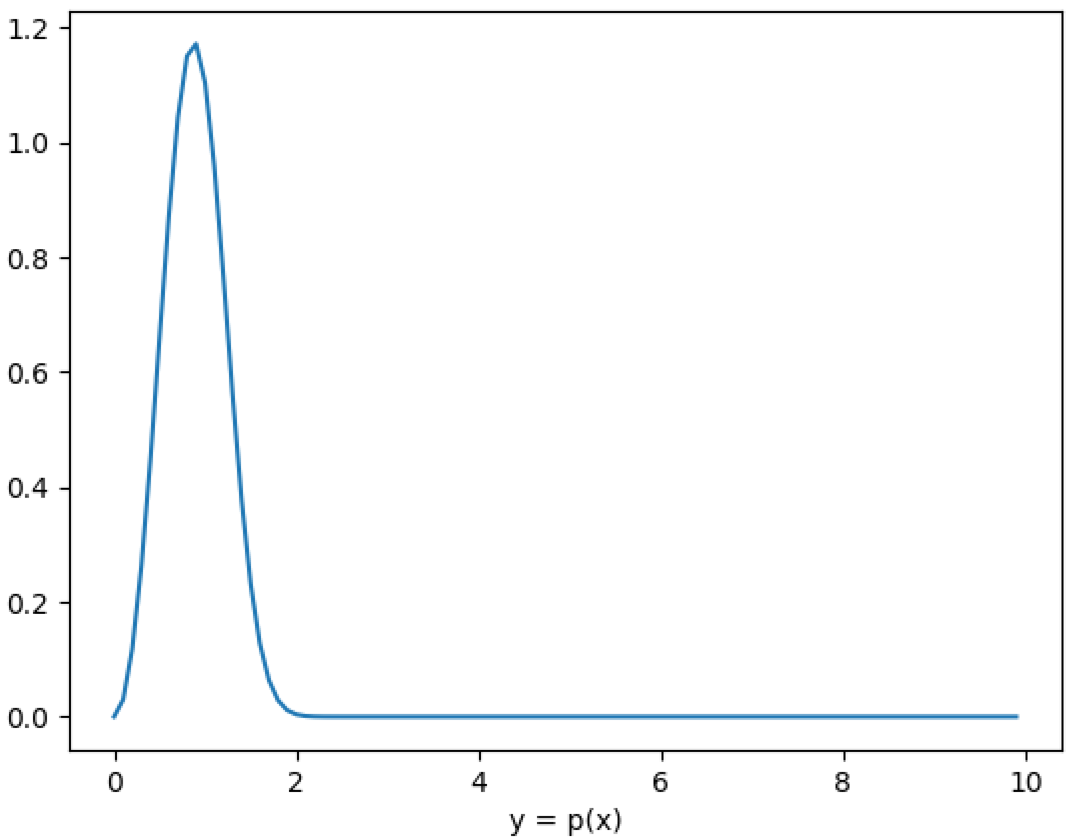
$$\left|det\begin{pmatrix}\nabla r(x,y)\\ \nabla\theta(x,y)\end{pmatrix}\right| = \left|det\begin{pmatrix}\frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}}\\ -\frac{y}{x^2}\frac{1}{1+\frac{y^2}{x^2}} & \frac{1}{x}\frac{1}{1+\frac{y^2}{x^2}}\end{pmatrix}\right| = \left|det\begin{pmatrix}\frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}}\\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2}\end{pmatrix}\right| = \left|\frac{x^2+y^2}{(x^2+y^2)^{\frac{3}{2}}}\right| = \frac{1}{\sqrt{x^2+y^2}}$$

$$p_{X,Y}(x,y) = \frac{\sqrt{x^2+y^2}}{2\pi\sigma^2}exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\frac{1}{\sqrt{x^2+y^2}} = \frac{1}{2\pi\sigma^2}exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

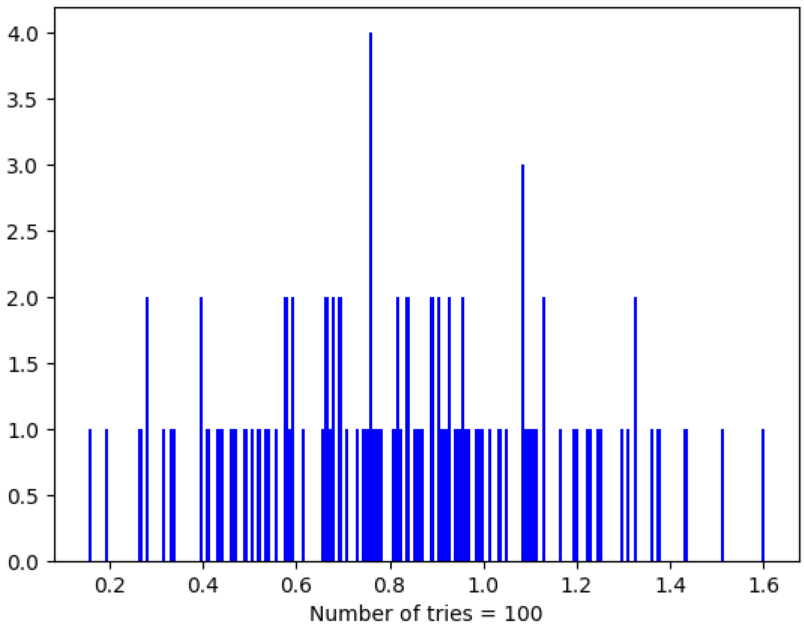
Это и есть совместная плотность распределения соответствующего гауссовского вектора. Тогда X и Y являются нормально распределёнными с матожиданием 0 и дисперсией σ^2 как компоненты гауссовского вектора и независимыми, поскольку соответсвующие ковариации в матрице равны нулю.

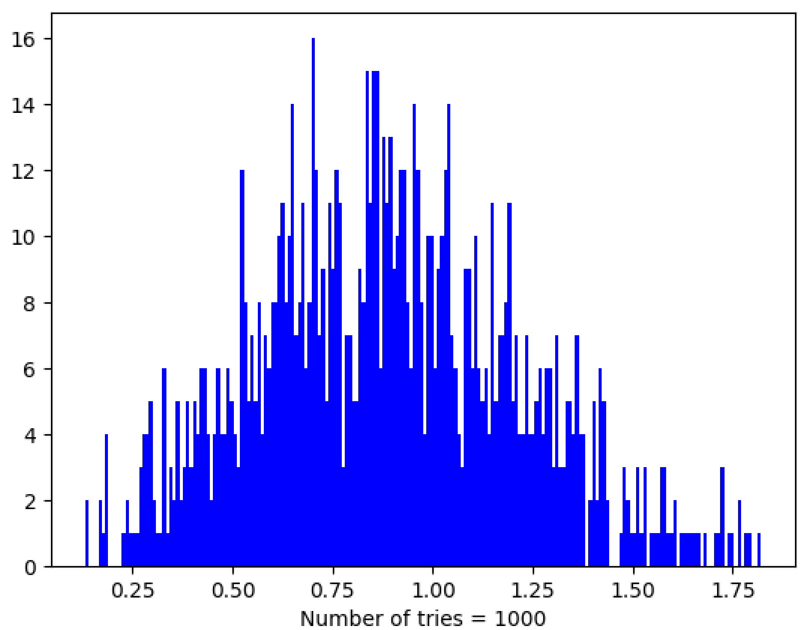
Задача №3, вариант 1

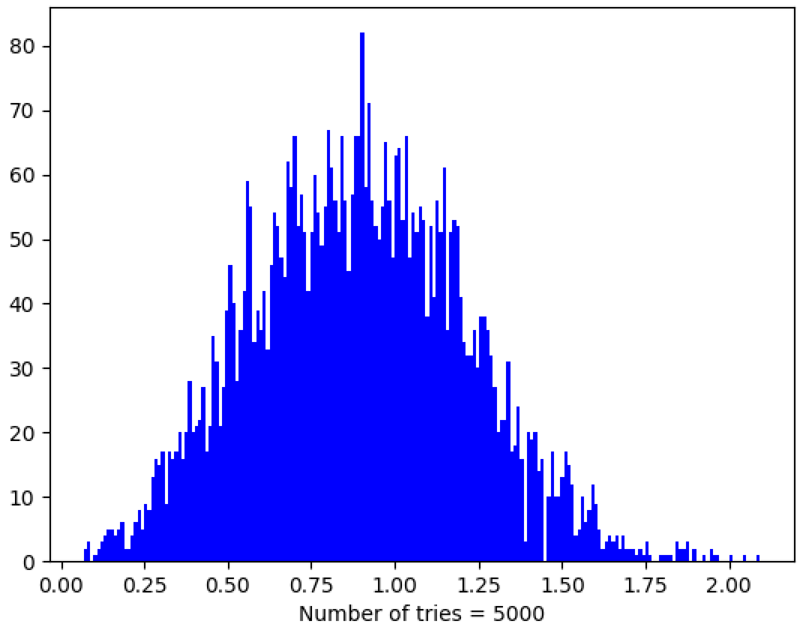
Построим график плотности, чтобы сравнивать распределения полученных значений с ним



Унаследовавшись от rv_continuous и реализовав метод _pdf получаем реализацию нужного нам распределения. Рассмотрим сэмплы разных размеров для этого распределения.


4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Number of tries = 100

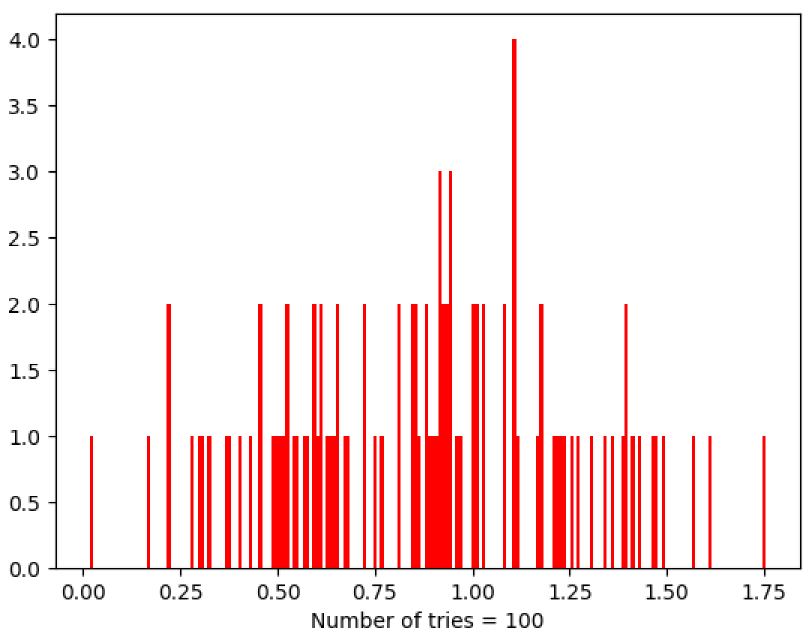

16
14
12
10
8
6
4
2
0
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Number of tries = 1000

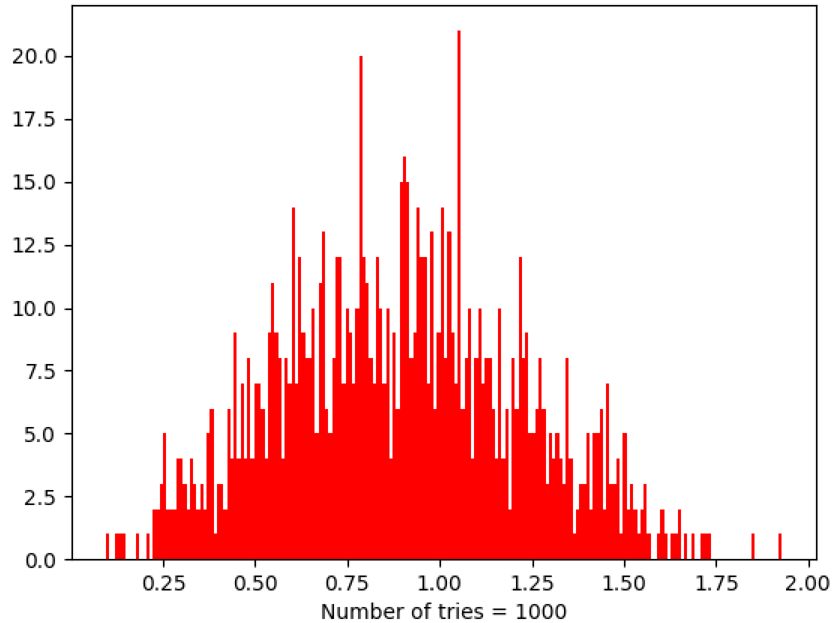


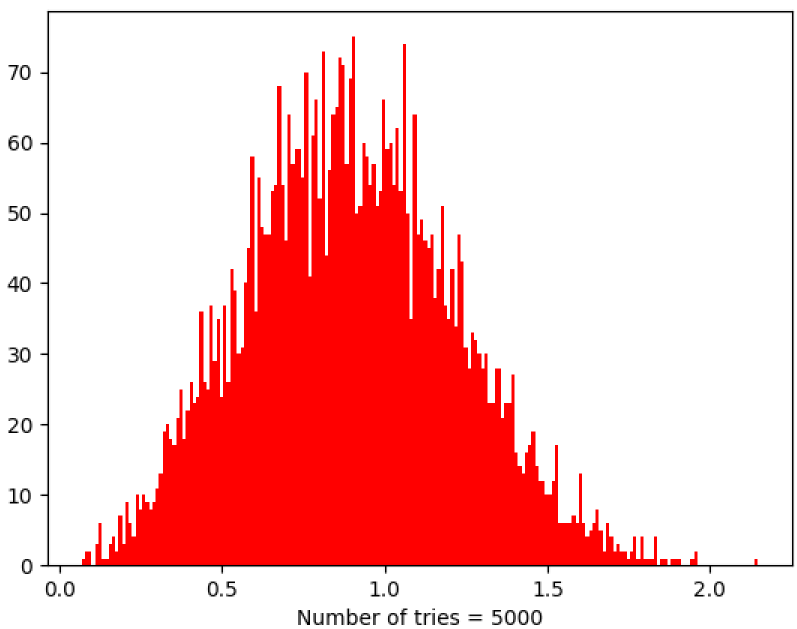
Очевидно стремление распределения к требуемому. При этом во время генерации довольно часто появляются аномально большие значения порядка 8, которые приходится отлавливать в ручную.
Для того, чтобы реализовать распределение с помощью обратной функции необходимо выразить обратную функцию в явном виде.

$$F(t) = \int_0^t 3x^2 e^{-x^3} = 1 - e^{-t^3} \implies F^{-1}(t) = -\sqrt[3]{ln(1-t)}$$

Рассмотрим сэмплы полученного распределения.



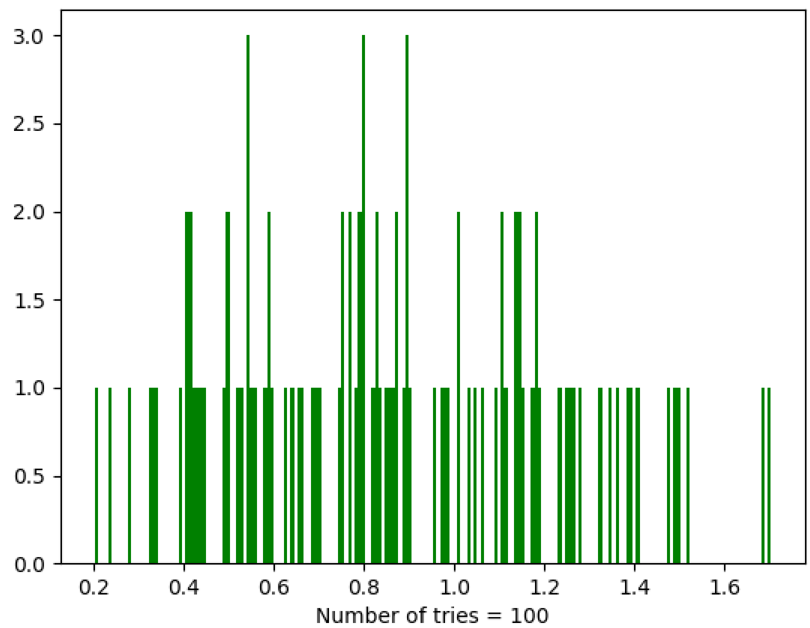


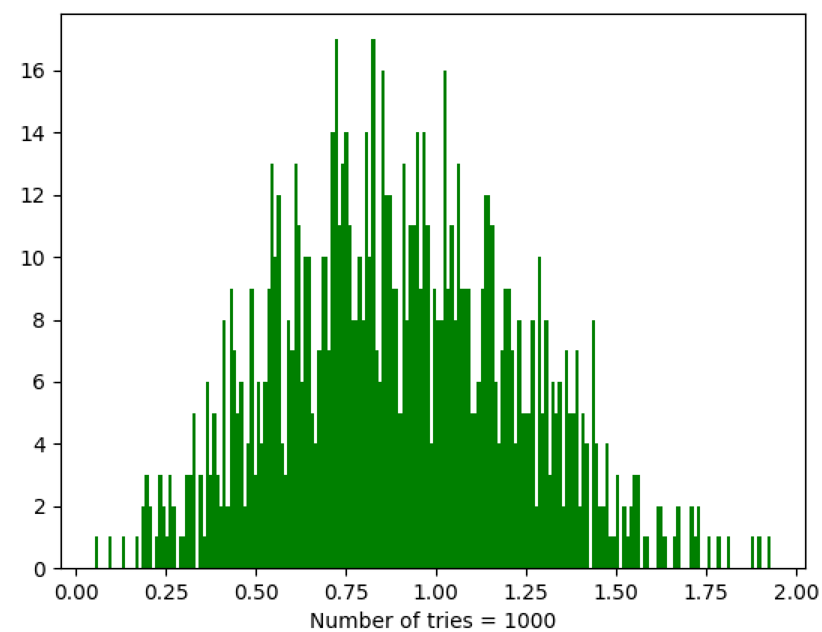


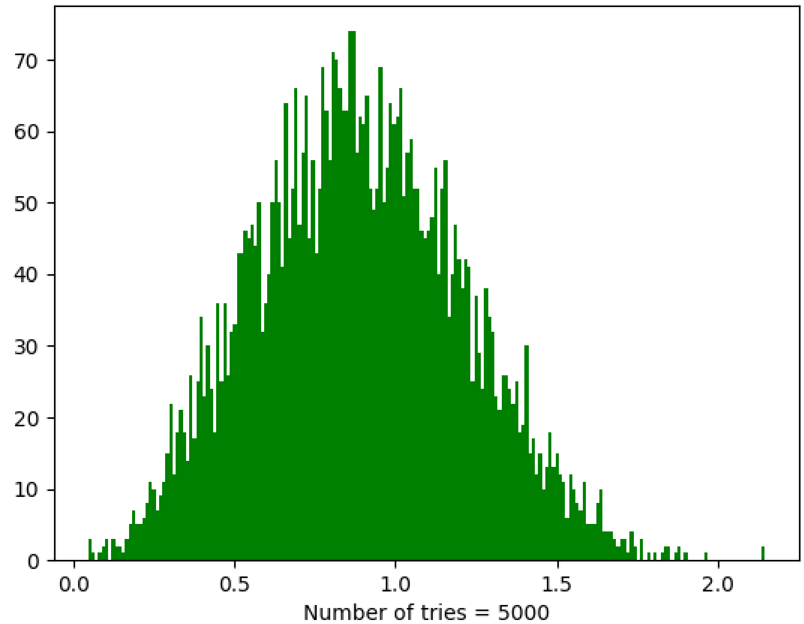
Полученные распределения не имеют аномалии в районе 8, а также распределены более равномерно, так как в предыдущем распределении при n=5000 есть блок значений выпавший $>$ 80 раз, в то время как в данном распределении самый частый блок встречался примерно 70 раз

В качестве третьего метода реализуем Rejection sampling. Суть метода такова: изобразим график плотности распределения, и выберем отрезок на оси OX, где лежит подавляющее количество значений (т.е. интеграл от плотности по отрезку почти равен единице). А так же ограничим Y сверху значением не меньше $\max p(x)$. В полученном прямоугольнике введём равномерное распределение и будем генерировать случайные точки, пока полученная точка не будет лежать ниже графика плотности. Когда получим подходящую точку, вернём её x как значение случайной величины.

Математически метод обоснован тем, что при отсечении точек над графиком мы получаем равномерное распределение на подграфике, при этом плотность x выражается как $p_X(x) = \int_0^{p(x)} p_Y(y)dy = p(x)$ в силу независимости X и Y. Приведём полученные сэмплы.







Распределение очень похоже на распределение, полученное с помощью F^{-1} и лишено недостатков распределения полученного наследованием.
Два последних метода работают намного быстрее первого, так как они используют лишь равномерное распределение, в то время как первый метод прибегает к большим вычислениям.