Лабораторная 2. Теорвер

Юрий Баринов

June 2023

Задача 1. Вариант 1

Предположим, что существуют такие Y и Z что $X = Y + Z$. Тогда для x_i найдутся такие y_i и z_i, что $x_i = y_i + z_i$. Тогда $y_i = x_i - z_i$.
Так как Y и Z являются имеют невырожденное распределение найдется такое i и j что $j = i + 1$ и $z_i \neq z_j$. Тогда

$$\begin{aligned} y_i &= x_i - a & z_i &= a \\ y_j &= x_j - b & z_j &= b \end{aligned}$$

Тогда предположим что $x_i - a + b = x_m$ а значит

$$-a + b = x_m - x_j = m^2 - i^2 = (i+\alpha)^2 - i^2 = 2\alpha i + \alpha^2$$

Попробуем подставить иначе $x_j - b + a$ и снова предположим что оно равно еще одному квадрату x_k. Тогда

$$\begin{gathered} -b + a = x_k - x_j = k^2 - (i+1)^2 = 2\alpha i + \alpha^2 \\ (i+\beta)^2 - i^2 - 2i + 1 = -2\alpha i - \alpha^2 \\ 2\beta i + \beta^2 - 2i + 1 = -2\alpha i - \alpha^2 \\ 2i(\alpha+\beta) + (\beta^2 + \alpha^2) = 2i + 1 \end{gathered}$$

Решим эту систему

$$\begin{cases} \alpha + \beta = 1 \\ \beta^2 + \alpha^2 = 1 \end{cases}$$

Получаем

$$\left[\begin{aligned} &\begin{cases} \alpha = 0 \\ \beta = 1 \end{cases} \\ &\begin{cases} \alpha = -1 \\ \beta = 2 \end{cases} \end{aligned}\right.$$

Если рассмотреть первый случай то выйдет что

$$\begin{gathered} x_i - a + b = x_i \\ -a + b = 0 \\ a = b \end{gathered}$$

Но мы изначально взяли различные a и b. Так что не рассматриваем этот случай. Рассмотрим же другой

$$\begin{cases} x_i - a + b = x_{i-1} \\ x_{i+1} - b + a = x_{i+2} \end{cases}$$

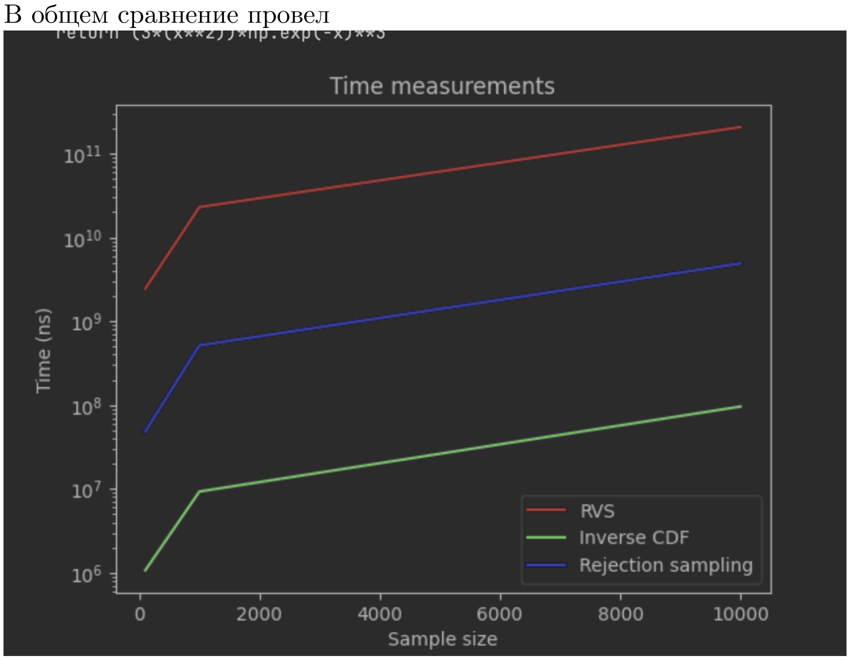
$$\begin{cases} -a + b = -2i + 1 \\ 2i + 1 - b + a = 4i + 4 \end{cases}$$

$$\begin{cases} -a + b = -2i + 1 \\ -b + a = 2i + 3 \end{cases}$$

Сложим два уравнения и получим $0 = 4$. Что является противоречием. Значит наше изначальное допущение не верно, а значит, что такие x_m и x_k не найдутся, а значит X и Z не существуют.

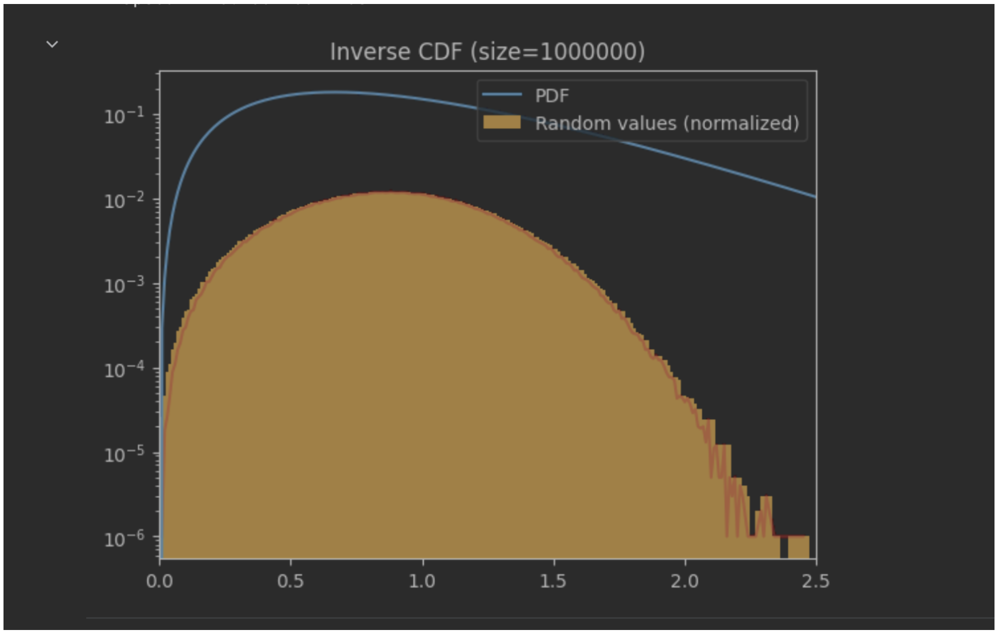
Задача 3. Вариант 1

В общем сравнение провел

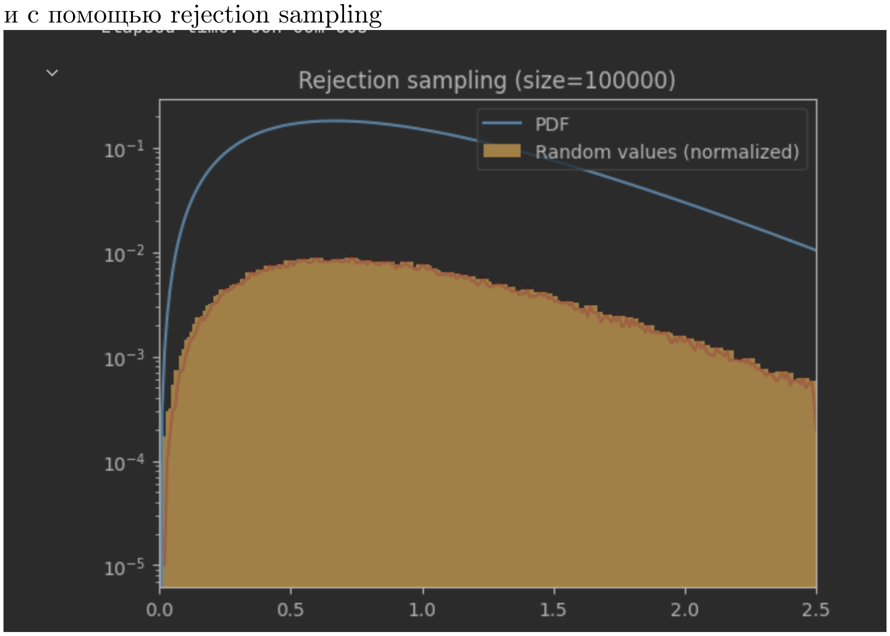


Шкала логарифмическая и там получается что генерация случайных чисел используя метод *rvs* из SciPy без предоставления ему *cdf* работает просто ужасающе медленно и дает довольно плохое распределение. Красивые графики можно глянуть тут. Там же можно посмотреть на графики того как выглядят распределения с ростом n

Из интересного можно посмотреть на графики распределения с помощью инвертированного *pdf*

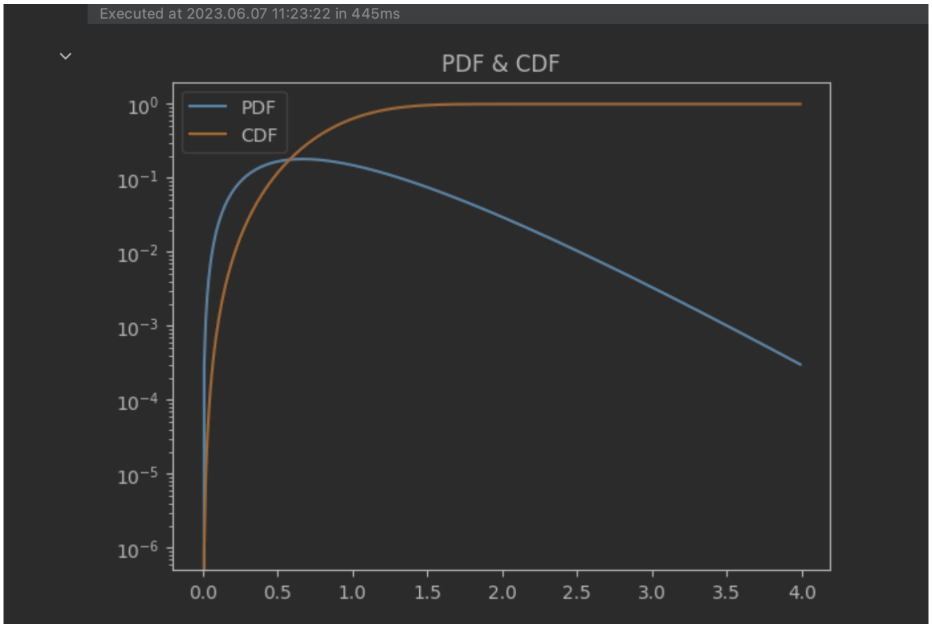


и с помощью rejection sampling



За счет того что у моей функции я не мог найти обратной я использовал поиски по *cdf* а она довольно быстро становилась равной единице потому кажется что сэмплилось хуже чем rejection sampling, у которого таких проблем нет вообще, но при этом rejection sampling работает дольше хоть и не так значительно как SciPy евский rvs.

Вот кстати как выглядят CDF и PDF



Ну использовал я как было выяснено rejection sampling.

Почему rejection sampling работает

Теорема. Если верхняя граница M для плотности распределения $p(x)$ существует, то алгоритм rejection sampling может быть использован для сэмплирования из $p(x)$, если $q(x)$ удовлетворяет следующим условиям:

1. Существует константа $c \geq 0$, такая что $p(x) \leq cq(x)$ для всех x.
2. Распределение $q(x)$ доступно для сэмплирования.

Доказательство:

Предположим, что $p(x)$ и $q(x)$ удовлетворяют условиям теоремы. Тогда, для всех x можно написать:

$P(u \leq p(x)) = \int_0^M P(u \leq p(x)) dq(x)$

$= \int_0^M P(u \leq p(x) \mid q(x)) q(x) dx$

$= \int_0^M \dfrac{p(x)}{cq(x)} q(x) dx$ (используя условие 1)

$= \dfrac{1}{c} \int_0^M p(x) dx$

$= \dfrac{1}{c} \int p(x) dx$

$= \frac{1}{c}$ (так как $p(x)$ - плотность вероятности, она должна интегрироваться до 1)

Это означает, что вероятность отброса (перехода в шаг 4) $u > p(x)$ не превосходит $\frac{1}{c}$. Из этого следует, что вероятность принятия сэмпла (перехода к шагу 3) равна $\frac{c}{c} = 1$. Таким образом, алгоритм rejection sampling может быть использован для сэмплирования из $p(x)$, если $q(x)$ удовлетворяет условиям теоремы.