Лабораторная работа по теории вероятности

1. X – случайная величина, принимающая значения только значения вида k^2, где $k \in \mathbb{N}$. При этом $\forall k \in \mathbb{N}\, P(X = k^2) > 0$. Показать, что X нельзя разложить в сумму двух независимых невырожденных случайных величин.

   Пусть X представим в виде $Y + Z$, где Y и Z – случайные величины со свойствами из условия. Из независимости следует, что если Y принимает значение y, а Z принимает значение z, то $Y + Z$ принимает значение $y + z$ с ненулевой вероятностью. Из матана известно, что конечная сумма по семейству обладает счетным числом ненулевых слагаемых, поэтому можем пронумеровать множество значений случайных величин Y и Z. Заметим, что так как $\forall i, j.\, y_i + z_j \in \mathbb{N}$, то все дробные части y_i и z_j должны совпадать. Тогда все y_i и все z_j можно сдвинуть на значения дробной части и получить целые числа. Так как $\forall i, j.\, y_i + z_j \geq 0$, то отрицательных значений конечное количество, причем они есть либо только среди y_i, либо среди только z_j. Пусть отрицательные значения только в z_j, тогда ко всем z_j можно прибавить $|\min z_j|$, а у всех y_i вычесть $|\min z_j|$. После сдвигов случайные величины Y и Z принимают только неотрицательные целые значения. Так как $Y + Z$ должна принимать все значения k^2, то Y и Z принимают 0. Либо Y, либо Z должны принимать 1. Пусть Y принимает 1, тогда Z не может принимать других значений, так как $\nexists n \in \mathbb{N} \setminus \{0\} : n = k_1^2$ и $n + 1 = k_2^2$. Получили противоречии с невырожденностью.

2. Даны $U, V \sim U[0, 1]$. Показать, что $X = \sqrt{-2 \ln V} \cos(2\pi U)$, $Y = \sqrt{-2 \ln V} \sin(2\pi U)$ образуют стандартный гауссов вектор.

   Воспользуемся формулой преобразования плотности:

   Для этого нужно предварительно выразить старые координаты через новые. Заметим, что

$$\begin{cases} X^2 + Y^2 = -2 \ln V \\ \dfrac{Y}{X} = \tan(2\pi U) \end{cases}$$

   Тогда преобразование получает вид:

$$\begin{cases} V = e^{-\frac{X^2+Y^2}{2}} \\ U = \dfrac{1}{2\pi} \arctan\big(\dfrac{Y}{X}\big) \end{cases}$$

   Тогда множитель при переходе в другие координаты равен:

$$\left\| \begin{matrix} -xe^{-\dfrac{x^2+y^2}{2}} & -ye^{-\dfrac{x^2+y^2}{2}} \\ -\dfrac{1}{2\pi}\dfrac{y}{x^2+y^2} & \dfrac{1}{2\pi}\dfrac{x}{x^2+y^2} \end{matrix} \right\| = \frac{1}{2\pi} e^{-\dfrac{x^2+y^2}{2}}$$

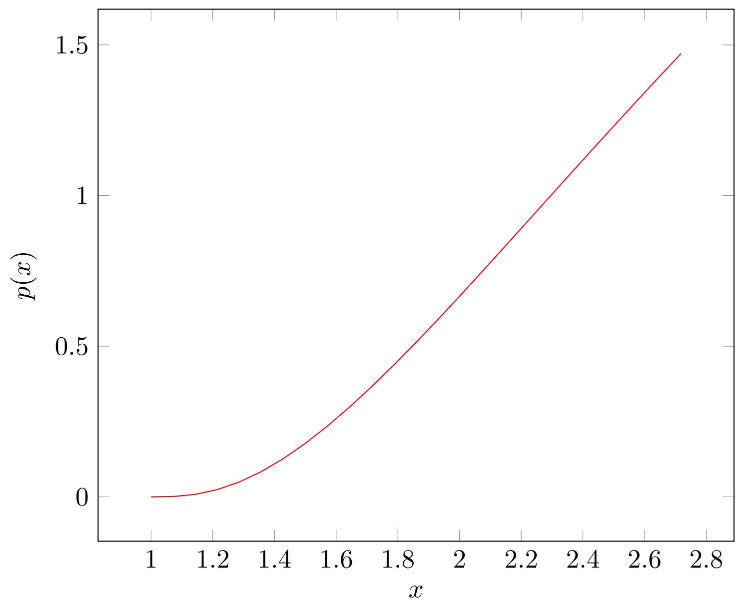
$p_{U,V}(u, v) = 1$, поэтому $p_{X,Y}(x, y) = \dfrac{1}{2\pi} e^{-\dfrac{x^2+y^2}{2}}$, что совпадает с формулой плотности для стандартного гауссового вектора. $(EX = 0, EY = 0, VarX = 1, VarY = 1, \rho(X, Y) = 0)$.

3. Реализация генератора случайных чисел для распределения с плотностью $p(x) = 4\frac{\ln^3(x)}{x}$ на $[1, e]$.

   Рассматриваются три реализации:

   1. Стандартная – задана только плотность.
   2. Через квантили.
   3. Rejection sample метод.

   Принцип работы rejection sample метода: рассмотрим прямоугольник, в котором полностью укладывается функция плотности. Будем равновероятностно генерировать точку (x, y) пока она не попадет в подграфик $p(x)$, при этом x – случайно сгенерированнное число, соответствующее распределению. Рассмотрим вероятность того, что x лежит в некотором отрезке $[a, b]$. Очевидно, что вероятность этого – это площадь под графиком $p(x)$, то есть $\int\limits_a^b p(x)dx$. Поэтому сгенерированные числа действительно соответствуют распределению. Вероятность попасть под график $p(x)$ равна $\frac{1}{S(\text{прямоугольника})} = \frac{e}{4(e-1)}$, а тогда матожидание числа попыток $= \frac{4(e-1)}{e} \approx 2.5$



   Замеры в репозитории показывают ожидаемые результаты линейного времени работы при генерации n случайных чисел. Rejection sampling работает медленнее метода через квантили, потому что методу через квантили требуется только одно случайное число, а rejection sampling не детерминирован. При этом стандарнтый метод показывает самый худший результат, так как скорее всего реализован через квантили, но обратная функция считается приближенно численным интегрированием.

   Стоит отметить, что графики распределения генераций чисел всеми 3 методами практически совпадают и очень близки к идеальному.

4. Оценка коорректности теорем о больших числах для распределения Пуассона при $\lambda \in (0, 10)$:

   Оценка через ЦПТ:

$$P\left(|\overline{X}_n - \mu| \leq a\sqrt{\frac{VarX_1}{n}}\right) \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{a} e^{-\frac{1}{2}x^2} dx$$

   В данном случае $VarX_1 = \lambda$, тогда для оценки нужно $\Phi(a) \geq 1 - \delta \Leftrightarrow a \geq \Phi^{-1}(1-\delta)$ и $a\sqrt{\frac{\lambda}{n}} \leq \varepsilon \Leftrightarrow \dfrac{\Phi^{-1}(1-\delta)\lambda}{\varepsilon^2} \leq n$
   При заданных $\varepsilon = 0.01$, $\delta = 0.05$ для выполнения при всех допустим λ неравенство на n должно выполняться при $\max E$, тогда $n \approx 270000$.

   Оценка через неравенство Чебышева:

$$P\left(|\overline{X}_n - \mu| \leq \varepsilon\right) \geq 1 - \frac{Var\overline{X}_n}{\varepsilon^2} = 1 - \frac{\lambda}{n\varepsilon^2}$$

   Тогда для получение необходимой оценки должно выполняться условие $\dfrac{\lambda}{n\varepsilon^2} \leq \delta$. Так как оно должно выполняться при всех λ, то оценка получиться при подстановке $\lambda = \max E$, тогда $n \approx 2000000$.

   При эмуляции на 100 выборках получается 0.94 вероятность выполнения условия для оценки через ЦПТ и 1.0 вероятность для оценки Чебышева. Стоит отметить, что неравенство Чебышева дает гораздо более грубую оценку (n отличается в почти в 10 раз). При этом неравенство Чебышева делает "не хуже так как условие дает знак больше. Оценка через ЦПТ дает апроксимацию, поэтому при разных запусках может получаться вероятность 0.9, что меньше ожидаемого в 0.95.