

Вариант 1

1. Проверить закон Ципфа на материале абстрактов PubMed с 2001 по 2017. Оценить величину поправки в закон Ципфа.
2. Имеют ли сети №2 и №3 одинаковую медиану? Описать использованные статистические тесты

Вариант 2

1. Просемплировать распределение χ^2 для произвольных параметров распределения.
2. Определить принадлежат ли сети №2 и №3 к одному распределению. Описать использованные статистические тесты

Вариант 3

1. Построить скрытую марковскую модель для сета №4. Построить график зависимости вероятности от итерации алгоритма.
2. Сгенерировать псевдоДНК последовательность с помощью марковской цепи. Использовать цепи длиной 2, 4, 6, 8 нуклеотидов. Частоты оценить из референсного генома человека.

Вариант 4

1. Просемплировать трехмерное распределение Максвелла при произвольных значениях параметров
2. Определить форму распределения для сета №1.

Вариант 5

1. Просемплировать обратное биномиальное распределение при произвольных значениях параметров
2. Сгенерировать псевдоДНК последовательность с помощью марковской цепи для экзонов. За один шаг марковской цепи должны генерироваться 3 символа.

Вариант 6

1. Дана скрытая марковская модель с параметрами

$$A = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}, \quad B = \begin{pmatrix} 0.55 & 0.45 \\ 0.3 & 0.7 \end{pmatrix}, \quad \pi = (0.5 \quad 0.5)$$

Найти оптимальный путь

2. Построить гистограммы для сетов №2 и №3. Определить оптимальные параметры распределения, считая что сеты распределены согласно смешанному гауссову закону $f(x) = \sum w_i N_i(m, \sigma)$, где $N_i(m, \sigma)$ - нормальное распределение, а $\sum w_i = 1$

Вариант 7

1. Используя первые 30 значений из сета №4 и модель $A = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}$, $B = \begin{pmatrix} 0.55 & 0.45 \\ 0.3 & 0.7 \end{pmatrix}$, $\pi = (0.5 \quad 0.5)$ вычислить апостериорные вероятности последовательности состояний.
2. Просемплировать биномиальное распределение при произвольных значениях параметров