

Метод 1

Нуклеотидам А, G, C і Т ставляться у відповідність вектори: А (1, 0.8), G (1, 0.6), C (1, 0.4), Т (1, 0.2). Послідовність отримуємо, сумуючи вектори, що відповідають нуклеотидам з послідовності.

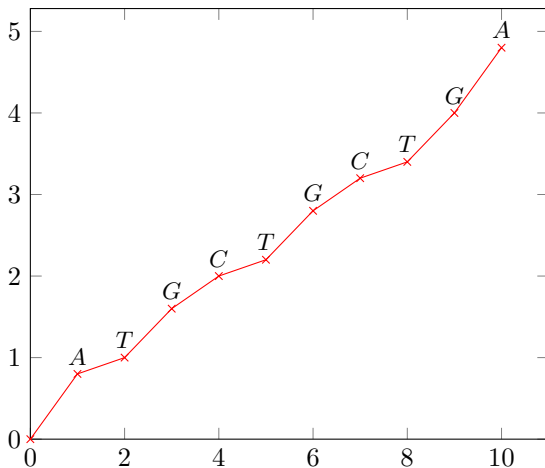


Рис.: Графічне представлення послідовності ATGCTGCTGA

Метод 2

Спочатку використаємо попередній метод щоб отримати числову послідовність (x_i, y_i) . Далі використаємо наступну формулу для обчислення результуючої послідовності:

$$\frac{x_i - \vec{y_i}}{\frac{1}{2}n(n+1) - y_n},$$

де $\vec{y_i}$ це y -компонента вектора, що відповідає i -тому нуклеотиду при використанні методу 1, n це розмір ДНК послідовності.

Метод 3

Нуклеотидам А, G, C, T ставимо у відповідність вектори $(-1, 0)$, $(1, 0)$, $(0, 1)$, $(0, -1)$. Починаємо з точки $(0, 0)$ і рухаємось по відповіднім векторам. Точки через які ми проходимо утворюють послідовність, причому точка стільки разів зустрічається у послідовності, скільки разів ми в неї потрапили.

Метод 4

Розташовуємо нуклеотиди у вершинах квадрата зі стороною 1: $A=(0,0)$, $G=(1,1)$, $C=(0,1)$, $T=(1,0)$. Координати послідовності рахуються ітеративно, рухаючись на половину відстані між попередньою позицією і точкою квадрата, якій відповідає наступний нуклеотид у напрямку цієї точки. Ітеративну процедуру можна задати наступним чином:

$$p_i = p_{i-1} - 0.5(p_{i-1} - g_i)$$

$$i = 1, \dots, n; p_0 = (0.5, 0.5),$$

де g_i - координати, що відповідають i -тому нуклеотиду, n - довжина послідовності ДНК.

Метод 4

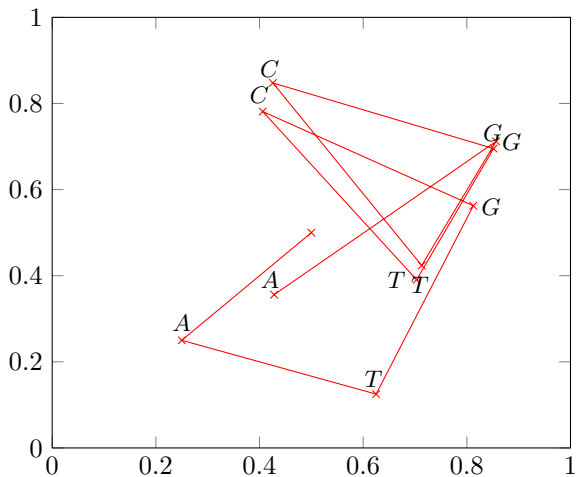


Рис.: Графічне представлення послідовності ATGCTGCTGA

Метод 5,6

Використовуємо попередній метод, щоб отримати послідовність p_i , отримуємо результуючу, як суму всіх попередніх:

$$z_i = \sum_{j=1}^i p_i$$

Отримуємо за допомогою методу 4 послідовність p_i і, щоб отримати результуючу послідовність, кожній точці ставимо у відповідність число:

$$z_i = x_i + y_i,$$

де $p_i = (x_i, y_i)$.

Метод 7

Нуклеотидам А,С ставимо у відповідність -1 , а нуклеотидам Т,Г ставимо у відповідність 1 . Починаючи з точки 0 рухаємось ітеративно:

$$p_i = p_{i-1} - \frac{(g_i - p_{i-1})}{2} \text{sign}(g_i)$$

де g_i число яке відвідає i -тому нуклеотиду. Тобто ми, подібно до методу 4, рухаємось на пів відстань до числа яке відповідає i -тому нуклеотиду.

p -СТАТИСТИКА

G і G' -генеральні сукупності. $x = (x_1, \dots, x_n) \in G$ і $x' = (x'_1, \dots, x'_m) \in G'$, $x_{(1)} < \dots < x_{(n)}$, $x'_{(1)} < \dots < x'_{(n)}$ - порядкові статистики. Припустимо, що $F_G(u) = F_{G'}(u)$.

$A_{ij}^{(k)} = \{x'_k \in (x_{(i)}, x_{(j)})\}$. Якщо $F_G(u) = F_{G'}(u)$:

$$P\left(A_{ij}^{(k)}\right) = P\left(x'_k \in (x_{(i)}, x_{(j)})\right) = p_{ij}^{(n)} = \frac{j-i}{n+1} = \frac{q}{n+1}, q = j-i$$

$$p_{ij}^{(1)} = \frac{h_{ij}^{(n)} m + g^2/2 - g\sqrt{h_{ij}^{(n)}(1-h)m + g^2/4}}{m + g^2}, \quad (1)$$

$$p_{ij}^{(2)} = \frac{h_{ij}^{(n)} m + g^2/2 + g\sqrt{h_{ij}^{(n)}(1-h)m + g^2/4}}{m + g^2}, \quad (2)$$

де $h_{ij}^{(n)}$ — частота події $A_{ij}^{(n)}$ в m випробуваннях. N кількість інтервалів $I_{ij}^{(n,m)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right)$ ($N = n(n-1)/2$) і L - кількість інтервалів $I_{ij}^{(n,m)}$, які містять ймовірності $p_{ij}^{(n)}$.

$h^{(n,m)} = \rho(x, x') = \frac{L}{N}$ будемо називати p -статистикою.