

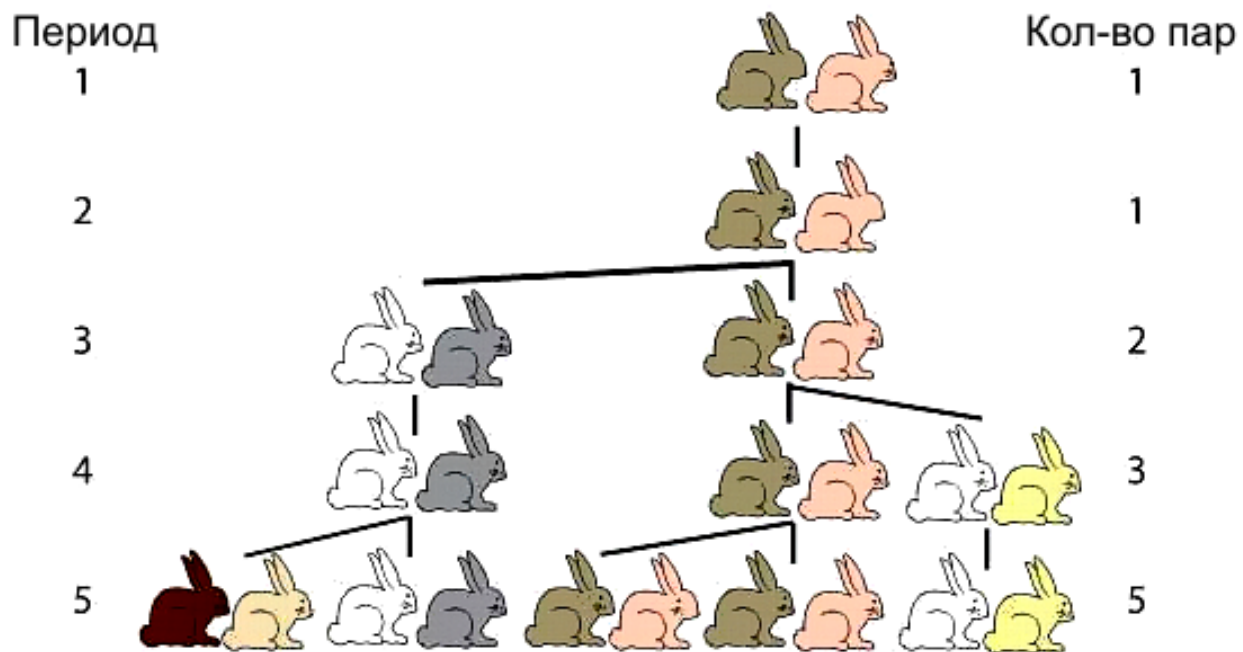
Лабораторная работа № 1

Определение 1. Рекуррентное соотношение (рекуррентная формула) – это соотношение вида $a_n = f(n, a_{n-1}, a_{n-2}, \dots, a_{n-p})$ которое позволяет вычислять все члены последовательности $1, 2, 3, \dots$, если заданы ее первые p членов.

Пример. Последовательность Фибоначчи задается линейным рекуррентным соотношением:

$$F_0 = 0, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}, \quad n \geq 2, \quad n \in \mathbb{N}.$$

На Западе эта последовательность была исследована Леонардо Пизанским, известным как Фибоначчи, в его труде «Liber Abaci» (1202). Он рассматривает развитие идеализированной (биологически нереальной) популяции кроликов, предполагая что: изначально есть новорожденная пара кроликов (самец и самка), со второго месяца после своего рождения кролики начинают спариваться и каждый месяц производить новую пару кроликов, кролики никогда не умирают. В конце n -го месяца число кроликов будет равно числу кроликов в предыдущем месяце плюс числу новорожденных пар, которых будет столько же, сколько пар было два месяца назад. Таким образом: $F_n = F_{n-2} + F_{n-1}$.



Кролики Фибоначчи

Задание 1.

Входные данные: Два натуральных числа $n \leq 40$ и $k \leq 5$. Число k – число рожающихся пар крольчат у пары кроликов, n – число месяцев, в течении которых размножаются кролики.

Выходные данные: Общее число пар кроликов которое образуется после n месяцев если изначально присутствует только одна пара и в каждом поколении каждая пара репродуктивного возраста производит потомство, состоящее из k пар кроликов (вместо одной пары из последовательности Фибоначчи).

Пример входных данных: $n = 5$, $k = 3$

Пример выходных данных: 19

Определение 2. Строка – это набор упорядоченных символов, взятых из некоторого алфавита, символы могут повторяться, длина строки – число символов, из которых она состоит.

Пример. Алфавит генетического кода – четыре азотистых основания – 'А' (аденин), 'С' (цитозин), 'G' (гуанин), 'Т' (тимин), строка – последовательность нуклеотидов с основаниями 'А', 'С', 'G' и 'Т' в цепи ДНК, например, "ATGCTTCAGAAAGGTCTTACG".

Задание 2.

Входные данные: s – последовательность ДНК.

Выходные данные: четыре целых числа, разделённые пробелами, представляющие собой количество символов 'А', 'С', 'G' и 'Т', содержащихся в s .

Пример входных данных: AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAGAGTGTCTGATAGCAGC

Пример выходных данных: 20 12 17 21

Определение 3. Расстояние Хэмминга – это число позиций, в которых соответствующие символы двух слов одинаковой длины различны.

Задание 3. Даны две строки s и t одинаковой длины. Найти расстояние Хэмминга $d_H(s, t)$ между ними.

Входные данные: две последовательности ДНК s и t одинаковой длины

Выходные данные: расстояние Хэмминга $d_H(s, t)$

Пример входных данных:

GAGCCTACTAACGGGAT
CATCGTAATGACGGCCT

Пример выходных данных: 7

Определение 4. Последовательность РНК – это строка, состоящая из букв алфавита 'А', 'С', 'G' и 'U', где 'U' – урацил.

Задание 4. Дана последовательность ДНК t . Заменяя в t тимин 'Т' на урацил 'U'.

Входные данные: последовательность ДНК.

Выходные данные: последовательность, полученная из t заменой 'Т' на 'U'.

Пример входных данных: GATGGAACCTTGACTACGTAAATT

Пример выходных данных: GAUGGAACUUGACUACGUAAAUU

Белки практически всех живых организмов построены из аминокислот всего 20 видов, которые обозначаются 20 буквами английского алфавита: G (глицин), L (лейцин), Y (тирозин), S (серин), E (глутаминовая кислота), Q (глутамин), D (аспарагиновая кислота), N (аспарагин), F (фенилаланин), A (аланин), K (лизин), R (аргинин), H (гистидин), C (цистеин), V (валин), P (пролин), W (триптофан), I (изолейцин), M (метионин), T (треонин).

Белок представляется в виде последовательности этих букв.

Таблица 1.

Кодон	АК	Кодон	АК	Кодон	АК	Кодон	АК
UUU	F	CUU	L	AUU	I	GUU	V
UUC	F	CUC	L	AUC	I	GUC	V
UUA	L	CUA	L	AUA	I	GUA	V
UUG	L	CUG	L	AUG	M	GUG	V
UCU	S	CCU	P	ACU	T	GCU	A
UCC	S	CCC	P	ACC	T	GCC	A
UCA	S	CCA	P	ACA	T	GCA	A
UCG	S	CCG	P	ACG	T	GCG	A
UAU	Y	CAU	H	AAU	N	GAU	D
UAC	Y	CAC	H	AAC	N	GAC	D
UAA	Stop	CAA	Q	AAA	K	GAA	E
UAG	Stop	CAG	Q	AAG	K	GAG	E
UGU	C	CGU	R	AGU	S	GGU	G
UGC	C	CGC	R	AGC	S	GGC	G
UGA	Stop	CGA	R	AGA	R	GGA	G
UGG	W	CGG	R	AGG	R	GGG	G

Генетический код — свойственный всем живым организмам способ кодирования аминокислотной последовательности белков при помощи последовательности нуклеотидов. Матричная рибонуклеиновая кислота (мРНК) — РНК, содержащая информацию о первичной структуре (аминокислотной последовательности) белков. Реализация генетической информации в живых клетках (то есть синтез белка, кодируемого геном) осуществляется при помощи двух матричных процессов: транскрипции (то есть синтеза мРНК на матрице ДНК) и трансляции генетического кода в аминокислотную последовательность (синтез полипептидной цепи на мРНК). Для кодирования 20 аминокислот, а также сигнала «стоп», означающего конец белковой последовательности, достаточно трёх последовательных нуклеотидов. Набор из трёх нуклеотидов называется триплетом. Единица генетического кода, тройка нуклеотидных остатков (триплет) в ДНК или РНК, обычно кодирующих включение одной аминокислоты называется кодоном. В таблице 1 приведены все 64 кодона и указаны соответствующие аминокислоты (АК).

Задание 5.

Входные данные: последовательность мРНК t

Выходные данные: соответствующая t белковая последовательность s

Пример входных данных:

AUGGCCAUGGCGCCCAGAACUGAGAUCAAUAGUACCCGUAUUAACGGGUGA

Пример выходных данных: MAMAPRTEINSTRING