

Imię i nazwisko:

Zadanie 1 (3 pkt)

	A	B
A	$2 * n$	$-2 * n$
B	$-2 * n$	$2 * n$

	A	B
A	2	-2
B	-2	2

	A	B	A	B	A
B	0	0	0	0	0
B	0	0	2	1	2
B	0	0	2	1	3
B	0	0	2	1	3
A	0	2	1	4	3

B	A	B	A
B	B	B	A

Narysuj graf 4 rzędu.



nie

AAAAAB, ABABABBBBBB

Zadanie 3 (3 pkt)

Znajdź motyw o długości $l = 3$, w 4 sekwencjach przedstawionych poniżej.

$$\begin{aligned}s_0 &= A & A & T & T, \\s_1 &= T & A & A & T, \\s_2 &= T & A & T & A, \\s_3 &= A & A & A & T.\end{aligned}$$

Motyw:

A	A	T
---	---	---

Pozycje występowania motywu, indeksujemy od 0:

$$P(s_0) = \mathbf{0}, P(s_1) = \mathbf{1}, P(s_2) = \mathbf{0}, P(s_3) = \mathbf{1},$$

Zadanie 4 (3 pkt)

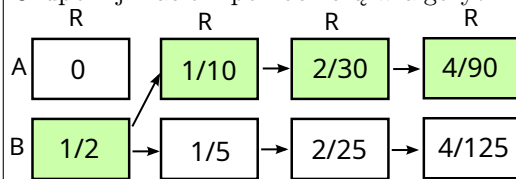
Posługujemy się monetami A i B, obserwując sekwencje rzutów $RRRR$ (orły i reszki). Podaj najbardziej prawdopodobną sekwencję użytych monet, zakładając, że to doświadczenie może być opisane ukrytym modelem Markowa pokazanym niżej.

$$\begin{aligned}Q &= \{A, B\}, \\V &= \{O, R\}, \\P_A &= 0, P_B = 1.\end{aligned}$$

	A	B
A	$\frac{2}{3}$	$\frac{1}{3}$
B	$\frac{1}{5}$	$\frac{4}{5}$

	O	R
A	0	1
B	$\frac{1}{2}$	$\frac{1}{2}$

Uzupełnij macierz pomocniczą w algorytmie Viterbiego



Podaj najbardziej prawdopodobną sekwencję użytych monet:

B	A	A	A
---	---	---	---

Pytania (2 pkt)

Jakie zagadnienie / zagadnienia warto byłoby dodać do MBI

Ile godzin poświęciłeś na MBI (sumarycznie wykłady, ćwiczenia, praca własna):

W jakim obszarze, Twoim zdaniem, przedstawione na wykładzie metody mogą przynieść największą korzyść.

Notatki lub uwagi do R. Nowaka:

Imię i nazwisko:

Zadanie 5 (3 pkt)

Poniżej podano fragment sekwencji referencyjnej oraz pliku SAM (kolumny POS i SEQ, odpowiadające pozycji oraz sekwencji) dla czterech kolejnych odczytów.

>chr1:101-124
accacatggtgttccgctttatg

101 accacatggtgttccgctt
102 ccacatggtgttccgcttt
103 cacatggtgttccgctttt
104 acatggtgttccgctttta

- Określ sekwencję aminokwasów, które są kodowane przez tą sekwencję zakładając, że na pozycji 1 zaczyna się nowy kodon.

TTCCSAFM
TTWCSAFM

- Dla pozycji 1-24 określ głębokość pokrycia odczytami. Na których pozycjach ta głębokość jest największa a na których najmniejsza?

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
o 1	2	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	1	0	0

- Czy dane z pliku SAM pozwalają wnioskować o wystąpieniu wariantu(ów)? tak

Jeżeli tak:

- a) Określ typ wariantu(ów): SNV
- b) Jaki jest ich efekt na białko? missense
- c) Jaka jest prawdopodobna zygotyczność wariantu (homozygotyczny czy heterozygotyczny)? heterozygotyczny

typ wariantu: SNV
efekt na białko: missense
prawdopodobna zygotyczność wariantu: heterozygotyczny

Tablica kodonów:

		Second base				
		U	C	A	G	
First base	U	UUU } Phenyl- alanine F UUC UUA } Leucine L UUG	UCU } Serine S UCC UCA UCG	UAU } Tyrosine Y UAC UAA } Stop codon UAG } Stop codon	UGU } Cysteine C UGC UGA } Stop codon UGG } Tryptophan W	U C A G
	C	CUU } Leucine L CUC CUA CUG	CCU } Proline P CCC CCA CCG	CAU } Histidine H CAC CAA } Glutamine Q CAG	CGU } Arginine R CGC CGA CGG	U C A G
	A	AUU } Isoleucine I AUC AUA } Methionine M start codon	ACU } Threonine T ACC ACA ACG	AAU } Asparagine N AAC AAA } Lysine K AAG	AGU } Serine S AGC AGA } Arginine R AGG	U C A G
	G	GUU } Valine V GUC GUA GUG	GCU } Alanine A GCC GCA GCG	GAU } Aspartic acid D GAC GAA } Glutamic acid E GAG	GGU } Glycine G GGC GGA GGG	U C A G

Notatki lub uwagi do Tomasza Gambina:

MBI, 25.06.2025, czas: 60 min

Imię i nazwisko:

[illegible]

Zadanie 6 (3 pkt)

Zaproponuj ile przestrzeni dyskowej należy przygotować na nieskompresowany plik z odczytami DNA w formacie FASTQ.

- sekwencjonowanie Whole Genome Sequencing;
- długość genomu referencyjnego: 2 Gbp;
- pokrycie odczytami: 30×;
- długość odczytu 100 bp.

W szacowaniu pomiń znaki nowych linii, przyjmij że każdy identyfikator sekwencji to 50 B.

Rozwiązanie krok po kroku:

$(2\text{Gbp} / 100 \text{ bp}) * 30 = 600\,000\,000$ – tyle odczytów

odczyt zajmuje 4 linie, zakładając 100 bajtów na sekwencję, 100 bajtów na jakość, oraz 50 bajtów

na opisy, więc 300B

daje to $600M * 300B \approx 180GB$

Notatki lub uwagi do W. Kuśmirka: