

Samoopravné kódy

Jméno:

Domácí úkol 11

Termín odevzdání: 11. května 2020 do půlnoci

Problém 1. Na genetický kód v RNA se můžeme dívat jako na blokový kód délky 3, který kóduje jednu z 20 aminokyselin nebo příkaz „stop“ (to budou naše zdrojová slova). Abeceda RNA kódu sestává ze čtyř znaků (tzv. bází) A, G, C, U.

Předpokládejme, že máme informační zdroj, který nám posílá RNA sekvenci délky $3n$, ve které jsou znaky na různých pozicích nezávislé náhodné a každá ze čtyř má bázi stejnou pravděpodobnost výskytu. My tuto posloupnost RNA bází po trojicích překládáme do posloupnosti aminokyselin (nebo příkazů stop, tj. celkem je 21 možností). Jaká je entropie výsledné posloupnosti aminokyselin v závislosti na n ?

Tabulku pro to, jaké proteiny odpovídají jakým trojicím znaků genetického kódu najdete například na

<https://www.khanacademy.org/science/biology/gene-expression-central-dogma/central-dogma-transcription/a/the-genetic-code-discovery-and-properties>

K řešení této úlohy můžete použít počítač; pokud to uděláte, popište ve svém řešení pro člověka srozumitelně a matematicky exaktně, jak entropii počítáte.

Při řešení úloh je možné se poradit s dalšími lidmi (nejlépe s Vašimi spolužáky a spolužačkami), ale svá řešení *píšte samostatně* a před termínem odevzdání úloh sepsaná řešení nikomu *neukazujte*.