1. Mājasdarbs
Bezzudumu saspiešana

Vairāki uzdevumi šajā mājasdarbā iespaidojušies no MIT Open Courseware: `https://bit.ly/37Aywtf` un `https://bit.ly/2BdK8GA`

Termiņš: 2020.gada 28.septembris, līdz vakaram (23:59:59 EEST).
Iesniegšanas veids: E-studiju vide.

1.uzdevums (Gabalu garumu kodējums).
Aplūkojam *Bernulli gadījumlielumu* X, kam

$$\begin{cases} P(X=0) = \dfrac{255}{256}, \\ P(X=1) = \dfrac{1}{256}. \end{cases}$$

(Piemēram, tiek mesta asimetriska monēta, kam "cipars" (*heads* jeb bits 1) uzkrīt ar varbūtību $\frac{1}{256}$.) Virknīti ar n šī gadījumlieluma vērtībām apzīmē ar X^n. To saspiež, izmantojot *Gabalu garumu kodējumu* (*run-length encoding*). Katru vieninieku kodē atsevišķi, bet nepārtrauktiem nuļļu gabaliem izvada tikai gabalu garumus.
Algoritma apraksts:
Ievadei X^n saspiesto rezultātu $f(X^n)$ iegūst, atkārtoti izpildot šos 3 soļus, ko turpina līdzkamēr ievade pilnībā nolasīta.

1.Solis
Ja ievadē pirmais bits ir "1", tad to nolasa un izvadē raksta astoņas nulles: `00000000`.

2.Solis
Ja ievades sākumā ir $r \leq 255$ nulles, aiz kurām seko vieninieks, tad visas šīs nulles nolasa un izvadē raksta 8-bitu virknīti – skaitļa r bināro pierakstu. (Tā kā nuļļu skaits $r \neq 0$, tad šajā solī nerodas izvade `00000000`.)

3.Solis
Ja ievadē ir $r > 255$ pēc kārtas esošas nulles, tad šajā solī nolasa tikai 255 nulles (un izvadē raksta skaitli 255 bināri, kā iepriekšējā punktā: `11111111`). Turpina pildīt 3.soli (un izvadīt skaitļa 255 kodus) līdzkamēr iestājas 1. vai 2. soļa situācija.

Atrast saspiešanas attiecības robežu šim kodējumam:

$$\lim_{n\to\infty} \frac{1}{n} E(\ell(f(X^n))). \quad (1)$$

Ar E apzīmējam vidējo vērtību gadījumlielumam, bet $\ell(f(X^n))$ apzīmē saspiestās virknītes garumu.

(A) Atrast (1) vērtību ar precizitāti līdz 10^{-8}.

(B) Atrast *entropijas ātrumu* (*entropy rate*) Bernulli eksperimentu virknei: $\lim_{n\to\infty} \frac{1}{n} H(X^n)$ jeb teorētisko robežu, cik labi šo virkni var saspiest. Izteikt šo lielumu ar formulu un arī aprēķināt to ar precizitāti līdz 10^{-8}

(Šī ir parodija par 4.uzdevumu no `https://bit.ly/2Y6mUKa`.)

2.uzdevums (Gadījuma analīze).
Aplūkojam ziņojumu alfabētu ar četriem ziņojumiem: $S = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$, kuru varbūtības ir attiecīgi $\{1/3, 1/3, 2/9, 1/9\}$.

1. Izmantot Hafmana algoritmu, lai atrastu šim ziņojumu alfabētam optimālu bezprefiksu kodējumu.
2. Vai ar Hafmana algoritmu var konstruēt būtiski citādu optimālu bezprefiksu kodējumu (t.i. tādu, kas simboliem `a`, `b`, `c`, `d` piekārto citādus kodu garumus, nevis tikai kaut kur aizstāj `0` ar `1` un otrādi).
3. Atrast bezprefiksu kodējumu, kurš arī ir optimāls, bet nevar būt radies Hafmana algoritma rezultātā.

(Šis ir uzdevums 2.12. no `https://bit.ly/2Y4PKMe`, 55.lpp.)

3.uzdevums (Markova process).
Pieņemsim, ka *Markova process* (Attēls 1) ir jau sasniedzis stabilu stāvokli (jau ir notikusi pietiekami gara nejauša staigāšana pa šo varbūtisko grafu, lai simbolu proporcijas un secību vairs nenoteiktu sākumstāvoklis).

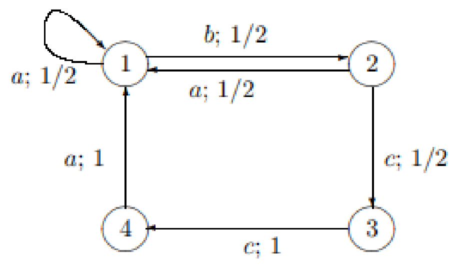

Figure 1: Markova process.

Aplūkojam stabilā Markova procesa ģenerētu simbolu virkni X^n, kur X pieņem vērtības no $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$.

(A) Atrast saspiešanas attiecības robežu, ja simbolu virkni X^n saspiež ar aritmētisko kodu.

(B) Ar datorsimulāciju izveidot virkni X^n garumā $n = 10^6$. Atrast saspiešanas attiecību, ja izmanto WinZip arhivatoru (WinZip parasti lieto *DEFLATE* algoritmu, kas ir modificēts LZ77 un vēl arī Hafmana kods.)

(C) Atrast saspiešanas attiecības robežu ($n \to \infty$), ja Markova virkni X^n saspiež ar Hafmana kodu, piešķirot prefiksu kodus atsevišķiem burtiem $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$.

(D) Atrast saspiešanas attiecības robežu ($n \to \infty$), ja Markova virkni X^n saspiež ar Hafmana kodu, piešķirot prefiksu kodus burtu pāriem

$$\{\mathtt{aa}, \mathtt{ab}, \ldots, \mathtt{cc}\}.$$

(Neiespējamie burtu pāri prefiksu kokā nav jāattēlo.)

(Šī ir parodija par 2.30.uzdevumu no `https://bit.ly/2Y4PKMe`, 60.lpp.)

5.uzdevums (I-iespēja).
Teksts satur $n = 2^k$ dažādus simbolus:

$$S = \{s_1, s_2, \ldots, s_n\}.$$

To varbūtības izkārtotas dilstošā secībā:

$$p_1 \geq \ldots \geq p_n.$$

Zināms, ka ziņojumu kopai S optimālu prefiksu kodējumu veido visas 2^k vienāda garuma k-bitu virknītes.

(A) Vai noteikti ir spēkā apgalvojums: Populārākā ziņojuma varbūtība apmierina nevienādību:

$$p_1 \leq \frac{2}{2^k + 1}?$$

(B) Vai noteikti ir spēkā apgalvojums: Populārākā ziņojuma varbūtība nepārsniedz divu retāko ziņojumu varbūtību summu: $p_1 \leq p_{n-1} + p_n$?

(C) Vai noteikti ir spēkā apgalvojums: Populārākā ziņojuma varbūtība nepārsniedz divkāršotu visretākā ziņojuma varbūtību: $p_1 \leq 2p_n$?

(Šī ir parodija par 16.3-8.uzdevumu no *Introduction to Algorithms. Third Edition.* Cormen, Leiserson, Rivest, 2009.)