Teoria de la Informació GCED-UPC curs 2019/20
Problemes; full número 4

13 de novembre de 2019

4.1. Es té un codi prefix de sis paraules amb longituds $1, 1, 2, 2, 3, 3$. Quantes lletres ha de tenir l'alfabet?

4.2. Es considera una variable aleatòria X que pren 7 valors amb distribució de probabilitats $\boldsymbol{p} = (0.49, 0.26, 0.12, 0.04, 0.04, 0.03, 0.02)$.

1. Trobeu un codi de Huffman binari per a aquesta variable.
2. Calculeu la longitud mitjana d'aquest codi de font.
3. Trobeu un codi de Huffman ternari per a aquesta variable i calculeu la seva longitud mitjana.

4.3. Trobeu un codi de Huffman binari per a una font de cinc lletres amb distribució de probabilitats $\boldsymbol{p} = (\frac{1}{3}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{2}{15})$. Comproveu que aquest codi també és optimal per a la distribució uniforme sobre les cinc lletres.

4.4. Digueu quins dels codis següents poden ser codis de Huffman per a alguna font:

1. $\{0, 10, 11\}$;
2. $\{00, 01, 10, 110\}$;
3. $\{01, 10\}$.

4.5. Es considera una cadena de Markov (d'ordre 1) estacionària $X_1, X_2, \ldots$ de tres estats s_1, s_2, s_3 amb matriu de transició

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Es consideren tres codis prefixos binaris C_1, C_2, C_3 que permetin codificar els estats de la cadena segons l'esquema següent: el primer estat es codifica amb qualsevol dels tres codis C_i; un cop codificat fins a l'estat n-èsim, i suposant que $X_n = s_i$, l'estat següent $X_{n+1} = s_j$ es codifica usant el codi C_i.

1. Calculeu els codis C_i que minimitzin la longitud d'aquesta codificació.
2. Calculeu la longitud mitjana per estat d'aquesta codificació.
3. Compareu el resultat anterior amb l'entropia del procés.

4.6. El teorema de codificació de font assegura que tot codi optimal per a una variable X té longitud mitjana L fitada de la forma $H(X) \leqslant L < H(X) + 1$. Doneu exemples de variables i codis optimals en els dos extrems: uns amb $L = H(X)$ i altres amb $L = H(X) + 1 - \epsilon$ per a un nombre $\epsilon > 0$ donat tan petit com es vulgui.

4.7. Una variable aleatòria X es codifica amb un codi D-ari prefix i s'obté la longitud mitjana més petita possible:

$$L = H_D(X) = H(X) \log_D 2.$$

Demostreu que

1. la distribució de probabilitats ha de ser de la forma $p(x_i) = D^{-n_i}$ per a enters $n_i \in \mathbb{Z}$;
2. el codi és complet: la desigualtat de Kraft-McMillan és una igualtat;
3. el nombre de paraules del codi de longitud màxima és múltiple de D.

4.8. Sigui X una variable aleatòria amb distribució de probabilitats $(\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12})$.

1. Construïu un codi de Huffman per a la variable X.
2. Construïu un codi de Shannon per a la variable X i compareu-lo amb l'anterior.
3. Calculeu la longitud mitjana d'un codi de bloc per a aquesta variable i compareu-lo amb els anteriors.
4. Traieu conclusions de tot plegat.

4.9. Trobeu codis de Huffman binaris i ternaris per a una variable aleatòria amb distribució de probabilitats $\frac{1}{21}(1, 2, 3, 4, 5, 6)$, calculeu les longituds mitjanes corresponents i compareu amb l'entropia.

4.10. Construïu un codi de Huffman binari per a la distribució de probabilitats $\frac{1}{10}(3, 3, 2, 1, 1)$, calculeu la seva longitud mitjana i compareu-la amb l'entropia.

Trobeu una distribució de probabilitats $(p_1, p_2, p_3, p_4, p_5)$ tal que el codi construït abans tingui longitud mitjana igual a l'entropia d'aquesta distribució.

4.11. Doneu un exemple de codi no prefix amb descodificació única per al qual hi ha missatges arbitràriament llargs tals que per a descodificar la primera paraula cal llegir tot el missatge.

4.12. Construïu codis de Huffman i de Shannon binaris per a la distribució de probabilitats $(0.6, 0.3, 0.1)$ i compareu les longituds mitjanes amb l'entropia.

Trobeu els enters $D \geqslant 2$ tals que els codis de Huffman i de Shannon D-aris són tots dos optimals per a aquesta distribució de probabilitats.

4.13. Considereu una variable que pren 19 valors amb distribució de probabilitats

$$\boldsymbol{p} = (0.4, 0.26, 0.02, 0.02, 0.02, \dots, 0.02).$$

Trobeu codis de Shannon, Shannon-Fano i Huffman binaris per a aquesta variable, calculeu les seves longituds mitjanes i compareu amb l'entropia.

Compareu les mides de les paraules en tots tres codis.

4.14. Implementeu en Python funcions que a partir d'una distribució de probabilitat calculin codis de Shannon, de Shannon-Fano i de Huffman. Podeu fer-les servir per resoldre els problemes de la llista, tot i que aquests es poden fer a mà, o per calcular codis per a distribucions en què treballar a mà no és possible. Apliqueu-los per exemple a distribucions de probabilitats de variables $\boldsymbol{X}_n$ del problema 1.3.

Alternativament, podeu buscar a Internet implementacions ja fetes i estudiar el codi procurant entendre'l.

4.15. Sigui X una variable aleatòria binària amb distribució de Bernoulli amb $p(1) = 0.9$. Quina és la seva entropia? Calculeu codis de Huffman per a codificar blocs de dígits emesos per aquesta font de longituds 1, 2, 3 i 4 i calculeu la seva longitud mitjana per dígit.

4.16. Sigui $\mathcal{C}$ un binari format per n paraules de longituds $\ell_1, \dots, \ell_n$. Demostreu que

$$\ell_1 + \ell_2 + \cdots + \ell_n \geqslant n \log n.$$

Indicació: Useu aquest codi per codificar una font d'informació apropiada.