Teoria de la Informació GCED-UPC curs 2018/19
Examen parcial

5 de novembre de 2018

1. Considereu la cadena binària 11011100101110111 formada concatenant els dígits binaris dels enters entre 1 i $7 = 2^3 - 1$. Considereu les variables aleatòries del problema 3 de la primera llista: $\boldsymbol{X}_2$ correspon a agafar dos dígits consecutius de la seqüència, i Y_1 i Z_1 són el primer i segon dígit de $\boldsymbol{X}_2$, respectivament. Recordeu el conveni que, al final de la cadena, s'ha d'afegir el primer dígit per poder agafar blocs de mida 2 que comencin en qualsevol dígit de la seqüència.

1. calculeu les entropies $H(\boldsymbol{X}_2)$, $H(Y_1)$, i $H(Z_1)$;
2. calculeu les entropies condicionades $H(Z_1|Y_1 = x)$ per a $x = 0$ i $x = 1$, i $H(Z_1|Y_1)$;
3. calculeu la informació mútua $I(Y_1; Z_1)$;
4. calculeu les entropies relatives $D(p\|q)$ i $D(q\|p)$, on p i q són les distribucions de probabilitat de les variables Y_1 i Z_1, respectivament;
5. calculeu les entropies relatives $D(p\|q)$ i $D(q\|p)$, on p i q són les distribucions de probabilitat de les variables $Z_1|Y_1 = 0$ i $Z_1|Y_1 = 1$, respectivament;

Solució: La cadena té longitud 17. Comptant parells de dígits consecutius s'obtenen les taules de probabilitat conjunta i condicionada, com es va fer en el problema 1.3:

$p(y,z)$	0	1	$p(y)$
0	1/17	4/17	5/17
1	4/17	8/17	12/17
$p(z)$	5/17	12/17	1

$p(z\|y)$	0	1
0	1/5	4/5
1	1/3	2/3

Les entropies d'aquestes variables es calculen com en el problema 2.1. Recordi's que en el problema 1.1 ja es veu que les distribucions de probabilitat de Y_1 i Z_1 són sempre la mateixa, independentment de la cadena binària (o de lletres de qualsevol alfabet) que es consideri. Les entropies es donen arrodonides a quatre decimals.

1. $H(\boldsymbol{X}_2) = H(\frac{1}{17}, \frac{4}{17}, \frac{4}{17}, \frac{8}{17}) = 1.7345$;
   $H(Y_1) = H(Z_1) = H(\frac{5}{17}, \frac{12}{17}) = 0.8740$;

2. $H(Z_1|Y_1=0)=H(\frac{1}{5},\frac{4}{5})=0.7219$;
   $H(Z_1|Y_1=1)=H(\frac{1}{3},\frac{2}{3})=0.9183$;
   $H(Z_1|Y_1)=\frac{5}{17}0.7219+\frac{12}{17}0.9183=0.8605$;
3. $I(Y_1;Z_1)=H(Y_1)+H(Z_1)-H(\boldsymbol{X}_2)=0.0135$;
4. les variables tenen la mateixa distribució $p=q=(\frac{5}{17},\frac{12}{17})$ i, per tant, per una propietat vista a teoria, $D(p\|q)=D(q\|p)=0$; en efecte, si dues variables tenen la mateixa distribució, a la fórmula
$$D(p\|q)=\sum_{x\in\mathcal{X}}p(x)\log\frac{p(x)}{q(x)}$$
els quocients $p(x)/q(x)$ són tots iguals a 1 i els seus logaritmes són zero.
5. les distribucions són $p=(\frac{1}{5},\frac{4}{5})$ i $q=(\frac{1}{3},\frac{2}{3})$; a partir de la definició de la divergència es calcula
$$D(p\|q)=\sum p(x)\log\frac{p(x)}{q(x)}=\frac{1}{5}\log\frac{3}{5}+\frac{4}{5}\log\frac{6}{5}=0.0630$$
Anàlogament es calcula $D(q\|p)=\frac{1}{3}\log\frac{5}{3}+\frac{2}{3}\log\frac{5}{6}=0.0703$.

2. De les afirmacions següents digueu quines són certes i quines falses; les certes, demostreu-les, i, de les falses, doneu-ne un contraexemple.

1. $I(X;Y)=H(Y)$ si, i només si, Y és funció de X: $Y=g(X)$;
2. si X pren valors a $\mathcal{X}$ i Y pren valors a $\mathcal{Y}$ i es té una funció $g\colon\mathcal{X}\to\mathcal{Y}$ exhaustiva i $Y=g(X)$ aleshores $H(X)=H(Y)$;
3. Si X i Y prenen valors reals i $Z=X+Y$ aleshores $I(X;Y|Z)=I(X;Y)$;
4. Si X i Y prenen valors reals i $Z=X+Y$ aleshores $H(X|Z)=H(Y|Z)$;
5. Si X i Y prenen valors reals i $Z=X+Y$ aleshores $H(Z|X)=H(Z|Y)$.

Solució:

1. Cert: $I(X;Y)=H(Y)-H(Y|X)=H(Y)\Leftrightarrow H(Y|X)=0$. A problemes s'ha vist que si $Y=g(X)$ aleshores $H(Y|X)=0$ (problema 2.3) i, recíprocament, que si $H(Y|X)=0$ aleshores $Y=g(X)$ (problema 2.4). Per tant, $H(Y|X)=0\Leftrightarrow Y=g(X)$.
2. Fals: s'ha vist al problema 2.2 que quan $Y=g(X)$ aleshores $H(Y)\leq H(X)$ i que la igualtat només es dóna quan la funció g és injectiva. Hi ha funcions exhaustives que no són injectives. Per tant es pot tenir un contraexemple agafant qualsevol funció exhaustiva no injectiva.
   (a) Contraexemple: sigui $\mathcal{X}=\{1,2,3,4\}$ i $\mathcal{Y}=\{1,2\}$, la funció és $g(1)=g(2)=1$ i $g(3)=g(4)=2$, que és exhaustiva, i les distribucions de probabilitat de X i Y són uniformes. Aleshores $H(X)=2\neq 1=H(Y)$.
   (b) Altre contraexemple: sigui $\mathcal{X}=\{x_1,x_2,\dots,x_n\}$ amb $n>1$ i X una variable aleatòria no constant (pren més d'un valor amb probabilitat no nul·la). Sigui $\mathcal{Y}=\{y\}$ un conjunt amb un únic element i sigui $g\colon\mathcal{X}\to\mathcal{Y}$ l'única aplicació possible: la que envia cada x_i a y (exhaustiva no injectiva). Aleshores Y és constant i es té $0=H(Y)<H(X)\neq 0$.

3. Fals. En el problema 2.6 es va veure que hi havia ternes de variables aleatòries X, Y, Z tals la relació entre els dos valors $I(X;Y)$ i $I(X;Y|Z)$ podia ser qualsevol: el primer més petit, el segon més petit, o tots dos iguals. N'hi ha prou a trobar un exemple en què no siguin iguals amb variable $Z = X + Y$. En aquest cas es té $I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|Z)$ ja que $X = Z - Y$ és funció de Y i Z.
   (a) Contraexemple: un que es va donar a classe en el problema 2.6 era X i Y independents binaries amb distribució uniforme; $Z = X + Y$ és binomial de tres valors; $I(X;Y) = 0$ però en canvi $I(X;Y|Z) = H(X|Z) = \frac{1}{2}H(X|Z=1) = \frac{1}{2} > 0$.
   (b) Altre contraexemple, amb desigualtat en l'altre sentit: $X = Y$ variables no constants, amb $Z = X + Y = 2X$ i $X = \frac{1}{2}Z$. Es té $I(X;X) = H(X) > 0$. En canvi, $I(X;X|Z) = H(X|Z) = 0$ ja que X és funció de Z.
4. Cert: és clar que $H(X,Y,Z) = H(X,Y) = H(X,Z) = H(Y,Z)$ ja que una qualsevol de les tres variables és funció de les altres dues. Aleshores

$$H(Z) + H(X|Z) = H(X,Z) = H(Y,Z) = H(Z) + H(Y|Z)$$

   i simplificant $H(Z)$ a cada costat s'obté la igualtat de l'enunciat.
5. Fals: raonant com a l'apartat anterior es té

$$H(X) + H(Z|X) = H(X,Z) = H(Y,Z) = H(Y) + H(Z|Y)$$

   i si es complís la igualtat de l'enunciat aleshores seria $H(X) = H(Y)$, però agafant dues variables qualsevol amb entropies diferents (per exemple, amb distribucions uniformes que prenguin un nombre diferent de valors) això no es compleix.

3. En aquesta pregunta valorarem les propietats del conjunt típic $\mathcal{A}_\epsilon^{(n)}$ suposant que el calculem per a una font que genera una seqüència de variables aleatòries i.i.d. $X_1, X_2, \ldots, X_n$. Es demana:

1. Quina condició ha de complir una seqüència per tal que sigui típica?
2. Les seqüències típiques no tenen totes la mateixa probabilitat, tot i estar en un rang de valors petit. Quin valor ha de tenir ϵ per tal que la probabilitat màxima sigui com a molt un 0.1% més gran que la probabilitat mínima?
3. Demostreu que el nombre d'elements de $\mathcal{A}_\epsilon^{(n)}$ està fitat superiorment per $2^{n(H(X)+\epsilon)}$.
4. Expliqueu de quina forma faríeu servir el concepte de conjunt típic $\mathcal{A}_\epsilon^{(n)}$ per codificar una font segons el *source coding theorem* fins a una llargada mitjana igual a l'entropia.

SOLUCIÓ:

1. La definició de seqüència típica (respecte un ϵ donat) és que la seva probabilitat satisfaci:

$$\left| -\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) - H(X) \right| \leq \epsilon.$$

2. La relació entre les probabilitats màxima i mínima del conjunt és

$$\frac{Pr_{max}}{Pr_{min}} = \frac{2^{-n(H(X)-\epsilon)}}{2^{-n(H(X)+\epsilon)}} = 2^{2n\epsilon}.$$

Si volem que $Pr_{max} \leq Pr_{min}(1 + 10^{-3})$, llavors ha de ser

$$2^{2n\epsilon} \leq 1.001 \Leftrightarrow \epsilon \leq \frac{\log(1.001)}{2n} = \frac{1}{n}(7.21 \times 10^{-04})$$

3.
$$\begin{aligned} 1 = \sum_{x^n \in \mathcal{X}^n} p(x^n) &\geq \sum_{x^n \in \mathcal{A}_\epsilon^{(n)}} p(x^n) \geq \sum_{x^n \in \mathcal{A}_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} \\ &= 2^{-n(H(X)+\epsilon)} \left|\mathcal{A}_\epsilon^{(n)}\right| \end{aligned}$$

i multiplicant tots dos costats per $2^{n(H(X)+\epsilon)}$ s'obté la desigualtat demanada.

4. El total de seqüències a transmetre es la unió del conjunt típic $\mathcal{A}_\epsilon^{(n)}$ i el seu complementari $\overline{\mathcal{A}}_\epsilon^{(n)}$. La llargada de les paraules del codi per a les seqüències de $\mathcal{A}_\epsilon^{(n)}$ seria: $\lceil n(H(X)+\epsilon)\rceil$ digits binaris per a cada seqüència més un 1 per designar $\mathcal{A}_\epsilon^{(n)}$. Per a les seqüències de $\overline{\mathcal{A}}_\epsilon^{(n)}$, assignaríem $\lceil n \log |\mathcal{X}|\rceil$ digits binaris més un 0 per denotar el conjunt complementari. Això seria suficient per codificar fins l'entropia quan $n \to \infty$. Tant el codificador com el descodificador haurien de tenir una taula amb $|\mathcal{X}|^n$ entrades per a relacionar les seqüències amb les paraules codi.

4. Una font genera símbols independents amb probabilitats $\{p_1, p_2, \ldots p_m\}$ i es vol construir un codi D-ari prefix amb llargades $\{\ell_1, \ell_2, \ldots \ell_m\}$. Es demana:

1. Quina condició han de complir $\{\ell_1, \ell_2, \ldots \ell_m\}$ per tal que existeixi un codi prefix?
2. Què vol dir que el codi sigui complet?
3. Deriveu la fita inferior del *Sandwich bound* i justifiqueu com cal que siguin $\{p_1, p_2, \ldots p_m\}$ i D per tal que el codi tingui una llargada mitja igual a l'entropia en base D, $H_D(X)$.
4. Digueu com es pot codificar de manera que la diferència entre la llargada del codi i l'entropia sigui arbitràriament petita.

Solució:

1. La condició suficient és la desigualtat de Kraft: $\sum_{i=1}^m D^{-\ell_i} \leq 1$.

2. Que la desigualtat de Kraft es compleix amb igualtat i que totes les possibles paraules codi del arbre construït amb el criteri prefix es fan servir. A més, el codi serà únicament descodificable.

3.
$$\begin{aligned} L - H_D(X) &= \sum_i p_i \ell_i - \sum_i p_i \log_D \frac{1}{p_i} \\ &= -\sum_i p_i \log_D D^{-\ell_i} + \sum_i p_i \log_D p_i \end{aligned}$$

Si definim la distribució auxiliar $r_i = \frac{D^{-\ell_i}}{\sum_i D^{-\ell_i}}$ i $d = \sum_i D^{-\ell_i} \leq 1$, ens queda

$$\begin{aligned} L - H_D(X) &= \sum_i p_i \log_D \frac{p_i}{r_i} - \log_D d \\ &= D(p||r) + \log_D d^{-1} \geq 0 \end{aligned}$$

on la divergència de Kullback-Leibler i el darrer terme són positius.

4. El *Sandwich bound* ens diu que $H_D(X) \leq L < H_D(X) + 1$. Si codifiquem els símbols de la font en grups de n en n podem arribar a tant a prop com es vulgui de l'entropia:

$$\frac{1}{n} H_D(X^n) \leq \frac{1}{n} L_n < \frac{1}{n} H_D(X^n) + \frac{1}{n}$$

a costa d'incrementar el retard en la codificació i descodificació.