DCC/ICEx/UFMG
INFORMATION THEORY
2ND SEMESTER OF 2017

PROF. MÁRIO S. ALVIM
SECOND EXAM
2017/10/24

Student: Solução do Professor	Registration number:

Instructions:

i. This exam contains 6 page(s). If your exam is missing any page, please ask the instructor for another copy of the exam.

ii. You can use the back of the exam's sheets as a draft, but the final answer for each question must be provided within the designated boxes. **Only answers provided within these boxes will be considered for grading.**

iii. Your answers will be evaluated in their clarity and conciseness.

---

1. Preencha cada quadrado abaixo com V ou F de acordo com se a afirmativa correspondente é *verdadeira* ou *falsa*, respectivamente.

5pts

Não é necessário justificar suas repostas, mas **cada resposta errada anulará uma correta.**

(a) V O código $\{00, 010, 10, 111\}$ é instantaneamente decodificável.

(b) F O código $\{01, 011, 100, 110\}$ é instantaneamente decodificável.

(c) V Seja uma fonte que pode emitir 4 símbolos $A_X = \{a_1, a_2, a_3, a_4\}$. É possível construir uma codificação instantaneamente decodificável para esta fonte se fixarmos o seguinte número de bits para a codificação de cada símbolo: $\ell_1 = 3$, $\ell_2 = 1$, $\ell_3 = 2$, $\ell_4 = 3$.

(d) F Seja uma fonte que pode emitir 4 símbolos $A_X = \{a_1, a_2, a_3, a_4\}$. É possível construir uma codificação instantaneamente decodificável para esta fonte se fixarmos o seguinte número de bits para a codificação de cada símbolo: $\ell_1 = 3$, $\ell_2 = 1$, $\ell_3 = 2$, $\ell_4 = 2$.

(e) F Códigos aritméticos projetados para uma fonte tendem a produzir boa compactação mesmo se aplicados em fontes distintas.

(f) F Códigos de Lempel-Ziv produzem excelente compactação mesmo que a fonte não possua muita redundância.

(g) — Seja arquivo X de entropia de $H(X)$ bits por símbolo, e considere que um código binário de compressão ótimo é aplicado a esta fonte, produzindo um arquivo binário compactado Y. Então Y tem entropia de aproximadamente $H(X)$ bits por símbolo.

(h) — Dadas duas distribuições de probabilidade p e q, a divergência de Kullback-Liebler $D_{KL}(p \parallel q)$ mede a quantidade extra de bits por símbolo necessária para codificar uma fonte distribuída de acordo com q quando produzimos um código assumindo que a distribuição da fonte é p.

(i) V Dado um enxemble conjunto (X, Y) é impossível que $H(X \mid Y) > H(X)$.

(j) F Dado um enxemble conjunto (X, Y) é impossível que $H(X \mid Y = y) > H(X)$.

Obs: As questões (g) e (h) foram anuladas por estarem ambíguas. Todos ganham os pontos destas questões.

2. **(Códigos de Huffman)** O ensemble X representa a frequência de letras em um texto, em que o alfabeto é $\mathcal{A}_X = \{\mathtt{a}, \mathtt{b}\}$, e a distribuição é $\mathcal{P}_X = \{1/3, 2/3\}$. O ensemble X^3 representa a escolha independente e aleatória de três letras (não necessariamente consecutivas) do ensemble X.

1pt

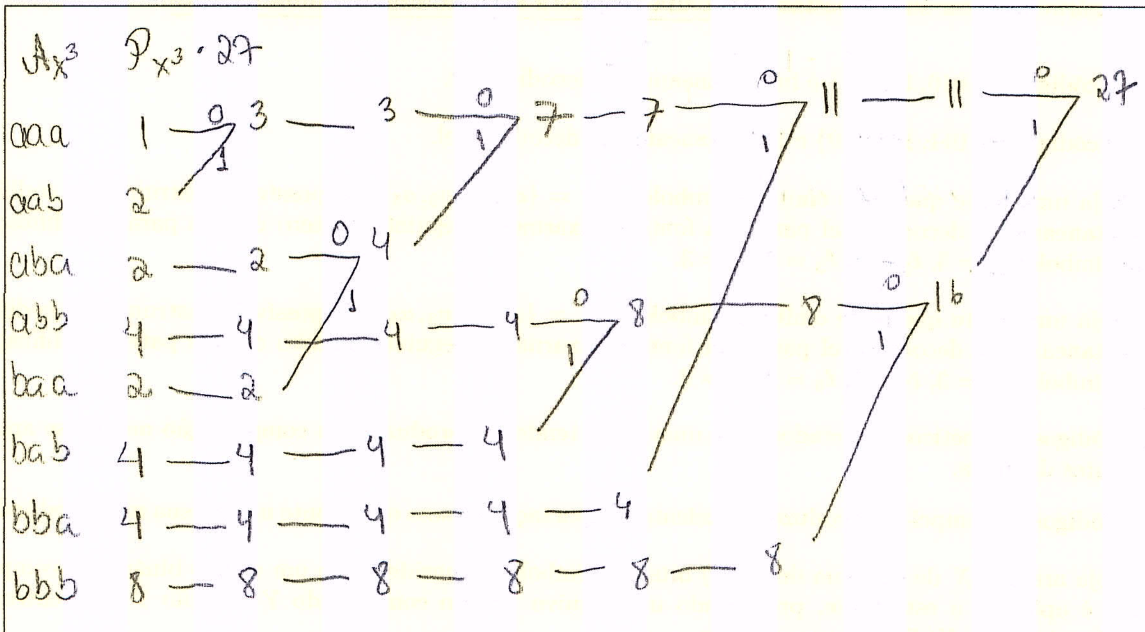
(a) Determine o conjunto de resultados $\mathcal{A}_{X^3}$ e as probabilidades $\mathcal{P}_{X^3}$ correspondentes para o ensemble X^3.

$\mathcal{A}_{X^3} = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$

$\mathcal{P}_{X^3} = \{1/27, 2/27, 2/27, 4/27, 2/27, 4/27, 4/27, 8/27\}$

3pts

(b) Crie um código de Huffman C para X^3 e compute seu tamanho esperado $L(C, X)$.



O código C é:

aaa $\mapsto$ 0000 aab $\mapsto$ 0001 aba $\mapsto$ 0010 abb $\mapsto$ 100

baa $\mapsto$ 0011 bab $\mapsto$ 101 bba $\mapsto$ 01 bbb $\mapsto$ 11

$L(C,X) = \sum_i p_i l_i = (1 \cdot 4 + 2 \cdot 4 + 2 \cdot 4 + 4 \cdot 3 + 2 \cdot 4 + 4 \cdot 3 + 4 \cdot 2 + 8 \cdot 2)/27$

$= 76/27 \approx 2.8148$ bits por símbolo.

3pts

(c) Usando o resultado $L(C, X)$ do item anterior, e usando *Shannon's Source Coding Theorem for Symbol Codes*, estime o valor da entropia $H(X^3)$, provendo valores $m, n \in \mathbb{R}$ tais que $m \leq H(X^3) \leq n$. (Faça isso sem calcular a entropia de $H(X^3)$ diretamente.)

Pelo Teorema de Shannon para códigos de símbolo sabemos que num código ótimo C' para X: $H(X) \leq L(C, X) < H(X) + 1$.

Isso significa que:

- $H(X) \leq L(C', X)$
- $L(C', X) < H(X) + 1 \Rightarrow H(X) > L(C', X) - 1$.

Como o código C é ótimo para X^3, temos que

- $H(X^3) \leq L(C, X^3) = 2.8148$
- $H(X^3) > L(C, X^3) - 1 = 1.8148$

Logo $\boxed{1.8148 < H(X^3) \leq 2.8148}$

3. **(Variáveis aleatórias conjuntas)** Usando as definições de entropia, entropia condicional e informação mútua e suas propriedades estudadas no curso ("*information can't hurt*", "*data-processing inequality*", "*chain rules*", etc.) demonstre quais inequalidades ($=$, $\leq$, $\geq$) se aplicam entre as quantidades abaixo.

Considere os ensembles

$$X, \qquad Y = \cos X, \qquad \text{e} \qquad Z = \sin Y.$$

2pts. (a) $H(X, Y \mid Z)$ vs. $H(X \mid Z)$.

$H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$ (regra da cadeia)

$= H(X|Z) + 0$ (como Y é função de X, temos $H(Y|X,Z) = 0$)

$= H(X|Z)$

Assim $\boxed{H(X,Y|Z) = H(X|Z)}$

Obs: Para ver que $H(Y|X,Z) = 0$ note que:

$H(Y|X) = 0$ (Y é função de X)

$H(Y|X) \geq H(Y|X,Z)$ ("information can't hurt")

Logo $H(Y|X,Z) \leq 0 \Rightarrow H(Y|X,Z) = 0$ (entropia é não-negativa).

2pts. (b) $I(X; Y)$ vs. $I(X; Z)$.

Como Y é uma função de X e Z é uma função de Y, temos que $X \to Y \to Z$ são uma cadeia de Markov. (Isto é, $p(z|y,x) = p(z|y)$).

Logo, pela data-processing inequality, temos que

$\boxed{I(X;Y) \geq I(X;Z)}$

4. **(Comunicação em canais ruidosos.)** Considere o canal de comunicação ruidoso C com entrada X, saída Y, e matriz de probabilidades condicionais abaixo.

(4 pts)

C	y_1	y_2
x_1	1/3	2/3
x_2	2/3	1/3
x_3	1/3	2/3

(a) Considere que o canal C seja utilizado com a distribuição de entrada $\mathcal{P}_X = \{1/2, 1/3, 1/6\}$. Calcule $H(X)$, $H(X \mid Y)$ e $I(X;Y)$.

p(x)

x_1	1/2
x_2	1/3
x_3	1/6

p(y|x)

	y_1	y_2
x_1	1/3	2/3
x_2	2/3	1/3
x_3	1/3	2/3

$\Longrightarrow$ p(x,y)

	y_1	y_2
x_1	1/6	2/6
x_2	2/9	1/9
x_3	1/18	3/18

=

	y_1	y_2
x_1	3/18	6/18
x_2	4/18	2/18
x_3	1/18	3/18

	y_1	y_2
p(y)	8/18	10/18

	y_1	y_2
p(y)	4/9	5/9

p(x|y)

	y_1	y_2
x_1	3/8	6/10
x_2	4/8	2/10
x_3	1/8	2/10

= p(x|y)

	y_1	y_2
x_1	3/8	3/5
x_2	1/2	1/5
x_3	1/8	1/5

$H(X) = -\left(1/2 \log(1/2) + 1/3 \log(1/3) + 1/6 \log(1/6)\right) = 1.4591$

$H(X \mid Y=y_1) = -\left(3/8 \log(3/8) + 1/2 \log(1/2) + 1/8 \log(1/8)\right) = 1.4056$

$H(X \mid Y=y_2) = -\left(3/5 \log(3/5) + 1/5 \log(1/5) + 1/5 \log(1/5)\right) = 1.3709$

$H(X \mid Y) = p(y_1)\, H(X \mid Y=y_1) + p(y_2)\, H(X \mid Y=y_2) = 4/9 \cdot 1.4056 + 5/9 \cdot 1.3709$

$= 0.6247 + 0.7616 = 1.3863$

$I(X;Y) = H(X) - H(X \mid Y) = 1.4591 - 1.3863 = 0.0728$

$H(X) = 1.4591$	$H(X \mid Y) = 1.3863$	$I(X;Y) = 0.0728$

(2 pts.)

(b) **(Questão extra +10%: "Desperdiçando entradas".)** Pode-se mostrar que a capacidade do canal C é de 0.08 bits por uso do canal e que esta capacidade é atingida sob a distribuição de entrada $\mathcal{P}_X^* = \{1/2, 1/2, 0\}$. Argumente porque no canal C a capacidade pode ser atingida em uma distribuição de probabilidade que ignora a entrada x_3.
Prove formalmente seu resultado. (Dica: mostre que qualquer que seja a distribuição de entrada p_1, p_2, p_3 para x_1, x_2, x_3, respectivamente, existe uma outra distribuição p'_1, p'_2, p'_3 que atinge a mesma informação mútua entre X e Y, mas em que $p'_3 = 0$.)

No canal C as linhas correspondentes a x_1 e a x_3 são idênticas, logo nenhum output é capaz de distinguir entre x_1 e x_3 e eles se comportam como um input só.
Logo, qualquer uso de x_3 pode ser substituido por um uso de x_1 sem impactar a informação mútua do canal.

Formalmente, considere a distribuicão de entrada p_1, p_2, p_3.
Para o canal C:

$$p(y_1) = p_1 \cdot p(y_1|x_1) + p_2 \cdot p(y_1|x_2) + p_3\, p(y_1|x_3) =$$
$$= (p_1 + p_3)\, p(y_1|x_1) + p_2\, p(y_1|x_2) \quad \text{pois } p(y_1|x_1) = p(y_1|x_3)$$

$$p(y_2) = (p_1 + p_3)\, p(y_2|x_1) + p_2\, p(y_2|x_2)$$

Logo a mesma distribuição em Y pode ser atingida com o p_1, p_2, p_3 original ou com $p'_1 = p_1 + p_3$, $p'_2 = p_2$, $p'_3 = 0$, e $H(Y)$ não se altera.
Da mesma forma,

$$H(Y|X) = p_1\, H(Y|X=x_1) + p_2\, H(Y|X=x_2) + p_3\, H(Y|X=x_3)$$
$$= (p_1 + p_3)\, H(Y|X=x_1) + p_2\, H(Y|X=x_2)$$
pois $H(Y|X=x_1) = H(Y|X=x_3)$ pois $p(y|x_1) = p(y|x_3)$.

Logo $H(Y|X)$ se mantém com $p'_1 = p_1 + p_3$, $p'_2 = p_2$, $p'_3 = 0$.
Assim $I(X;Y)$ é o mesmo com p_1, p_2, p_3 ou $p'_1 = p_1 + p_3$, $p'_2 = p_2$, $p'_3 = 0$ e a capacidade também é a mesma.