87/100

Ayoub Echchahed
(111 274 558)

Théorie de l'information
GEL-7062

Devoir 2
Résolution de Problèmes

Travail présenté à
Mr. Jean-Yves Chouinard

Faculté des sciences et de génie
Université Laval
Hiver 2022

- **<u>Problème 1 : Codage de source sans perte</u>** 15/20

Problème 4.1 : Une source d'information discrète X génère des symboles avec la distribution d'entrée suivante : $p(x_1) = 0.37$, $p(x_2) = 0.28$, $p(x_3) = 0.18$, $p(x_4) = 0.12$, $p(x_5) = 0.05$.

a) Donnez l'entropie de la source $H(X)$ en *shannons* (ou bits).

b) Construisez un code de Huffman binaire pour cette source, écrivez les mots de code et leur longueur respective. Quelle est la longueur moyenne des mots de code ?

c) Construisez un *code de Shannon-Fano* binaire, où $l_k = \lceil -\log_b p(x_k) \rceil$, pour cette source et donnez encore une fois les mots de code, leur longueur respective et la longueur moyenne des mots de code.

d) Considérons maintenant des codes de Huffman and Shannon-Fano M-aires pour la même source d'information.

   i) Déterminez l'entropie de la source $H(X)$ en unités M-aires.

   ii) Quel est le plus petit entier M tel que la longueur moyenne des mots de code du code Shannon-Fano M-aire sera la même que pour le code Huffman$M$-aire ? Expliquez et justifiez clairement votre réponse.

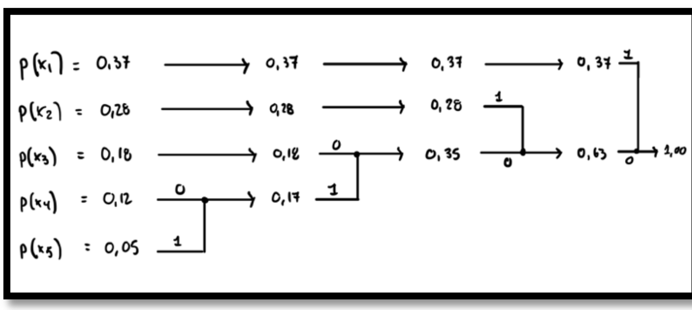
a) $H(X) = -\sum p(x_k) \log_2 p(x_k)$

$= -[(0.37)\log_2(0.37) + (0.28)\log_2(0.28) + (0.18)\log_2(0.18) + (0.12)\log_2(0.12) + (0.05)\log_2(0.05)]$

$= -[(-0.5307) + (-0.5142) + (-0.4453) + (-0.3671) + (-0.2161)]$

= **2.07 Sh** 5/5

b)



Symbole x_k	**Probabilité $p(x_k)$**	**Mot-code c_k**	**Longueur l_k**
x1	0.37	1	1
x2	0.28	01	2
x3	0.18	000	3
x4	0.12	0010	4
x5	0.05	0011	4

- Longueur moyenne :

$L = \sum p(x_k)\, l_k$

$= (0.37)(1) + (0.28)(2) + (0.18)(3) + (0.12)(4) + (0.05)(4)$

= **2.15 bits / symbole de source** 5/5

c) - Using the binary expansion of the cumulative probabilities

Symbole x_k	**Probabilité $p(x_k)$**	**Probabilités cumulées**	**Probabilités cumulées (binaire)**	**Longueur $l_k = \lceil -\log_2 p(x_k) \rceil$**	**Mot-code c_k**
x1	0.37	0.00	0.00000	2	00
x2	0.28	0.37	0.010111	2	01
x3	0.18	0.65	0.101001	3	101
x4	0.12	0.83	0.110101	4	1101
x5	0.05	0.95	0.111100	5	11110

- Longueur moyenne :

$L = \Sigma\, p(x_k)\, l_k$
$= (0.37)(2) + (0.28)(2) + (0.18)(3) + (0.12)(4) + (0.05)(5)$
$=$ **2.57 bits / symbole de source**

5/5

d)

i) Entropie de la source

(Calculer entropie pour base 3, base 4, base 5...)

$H(X) = -\Sigma\, p(x_k) \log_3 p(x_k)$
$= -[(0.37)\log_3(0.37) + (0.28)\log_3(0.28) + (0.18)\log_3(0.18) + (0.12)\log_3(0.12) + (0.05)\log_3(0.05)$
$=$ **1.3082**

$H(X) = -\Sigma\, p(x_k) \log_4 p(x_k)$
$= -[(0.37)\log_4(0.37) + (0.28)\log_4(0.28) + (0.18)\log_4(0.18) + (0.12)\log_4(0.12) + (0.05)\log_4(0.05)]$
$=$ **1.0367**

$H(X) = -\Sigma\, p(x_k) \log_5 p(x_k)$
$= -[(0.37)\log_5(0.37) + (0.28)\log_5(0.28) + (0.18)\log_5(0.18) + (0.12)\log_5(0.12) + (0.05)\log_5(0.05)]$
$=$ **0.893**

ii) Plus petit entier M où L (Huffman) = L (Shannon-Fano)

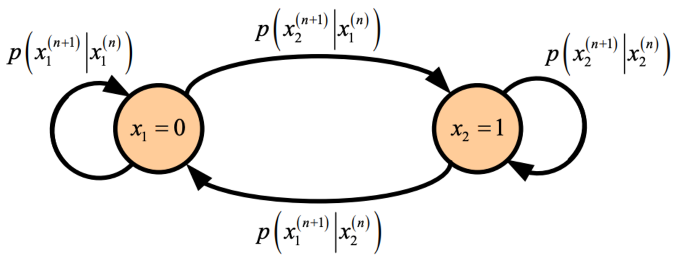
0/5 M=20

- **Problème 2: Source d'information markovienne**

16/20

Problème 4.5 : Soit X une source d'information binaire avec mémoire, stationnaire et markovienne. Sa matrice de probabilité de transition est donnée par :

$$\mathbf{P} = \begin{bmatrix} p\left(x_1^{(n+1)} \middle| x_1^{(n)}\right) & p\left(x_2^{(n+1)} \middle| x_1^{(n)}\right) \\ p\left(x_1^{(n+1)} \middle| x_2^{(n)}\right) & p\left(x_2^{(n+1)} \middle| x_2^{(n)}\right) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.9 & 0.1 \end{bmatrix}$$



a) Déterminez les probabilités des symboles x_1 et x_2.

b) Calculez les entropies conditionnelles $H(X|x_1)$ et $H(X|x_2)$.

c) Donnez l'entropie $H_{\text{mémoire}}(X)$ de ce canal avec mémoire.

d) Maintenant supposez que le canal est sans mémoire mais avec les mêmes probabilités $p(x_1)$ et $p(x_2)$ trouvées en (a). Quelle est l'entropie $H(X)$ de ce canal sans mémoire ?

e) Comparez les valeurs d'entropie avec et sans mémoire, $H_{\text{mémoire}}(X)$ et $H(X)$ et expliquez les résultats obtenus.

$$\mathbf{p} = \left\{ p(x_1) = \frac{\beta}{\alpha+\beta}, p(x_2) = \frac{\alpha}{\alpha+\beta} \right\}$$

L'entropie $H(X_n)$ au temps discret n :

$$\begin{aligned} H(X_n) &= -\left[p(x_1)\log_b p(x_1) + p(x_2)\log_b p(x_2)\right] \\ H(X_n) &= -\left[\left(\frac{\beta}{\alpha+\beta}\right)\log_b\left(\frac{\beta}{\alpha+\beta}\right) + \left(\frac{\alpha}{\alpha+\beta}\right)\log_b\left(\frac{\alpha}{\alpha+\beta}\right)\right] \end{aligned}$$

Le taux d'entropie $H_R(X)$ est donc :

$$\begin{aligned} H_R(X) &= H(X_2|X_1) = -\sum_{k=1}^{2}\sum_{j=1}^{2} p(x_k)\, p(x_j|x_k) \log_b p(x_j|x_k) \\ H_R(X) &= -\sum_{k=1}^{2} p(x_k) \sum_{j=1}^{2} p(x_j|x_k) \log_b p(x_j|x_k) \end{aligned}$$

$$H_R(X) = -\left\{ \left(\frac{\beta}{\alpha+\beta}\right)\left[(1-\alpha)\log_b(1-\alpha) + \alpha\log_b\alpha\right] + \left(\frac{\alpha}{\alpha+\beta}\right)\left[\beta\log_b\beta + (1-\beta)\log_b(1-\beta)\right] \right\}$$

$$\mathbf{P} = \begin{bmatrix} p\left(x_1^{(n+1)} \middle| x_1^{(n)}\right) & p\left(x_2^{(n+1)} \middle| x_1^{(n)}\right) \\ p\left(x_1^{(n+1)} \middle| x_2^{(n)}\right) & p\left(x_2^{(n+1)} \middle| x_2^{(n)}\right) \end{bmatrix} = \begin{bmatrix} (1-\alpha) & \alpha \\ \beta & (1-\beta) \end{bmatrix}$$

a) 4/4

p(x1) = 0.9 / (0.3 + 0.9)
= 0.75

p(x2) = 0.3 / (0.3 + 0.9)
= 0.25

b) 4/4

$$H(\tilde{Y}|x = x_1) = -[p(y = y_1|x = x_1) \log p(y = y_1|x = x_1) + p(y = y_2|x = x_1) \log p(y = y_2|x = x_1)]$$

$H(X \mid x1) = - \Sigma\Sigma\, p(x_k, y_j) \log_b p(x_k \mid y_j)$
$= - [0.7 \log_2 (0.7) + 0.3 \log_2 (0.3)]$
$=$ 0.8813 Sh

p(x1) = 1

$H(X \mid x2) = - \Sigma\Sigma\, p(x_k, y_j) \log_b p(x_k \mid y_j)$
$= - [0.9 \log_2 (0.9) + 0.1 \log_2 (0.1)]$
$=$ 0.469 Sh

p(x2) = 1

c) $H_R(X) = -\{(\beta / (\alpha + \beta) [(1 - \alpha) \log_b (1 - \alpha) + \alpha \log_b \alpha] + ((\alpha / (\alpha + \beta)) [\beta \log_b \beta + (1 - \beta) \log_b (1 - \beta)]\}$
$= -\{(0.9 / (0.3 + 0.9) [(0.7) \log_b (0.7) + 0.3 \log_b 0.3] + ((0.3 / (0.3 + 0.9)) [0.9 \log_b 0.9 + (0.1) \log_b (0.1)]\}$
$= -\{(0.75) [-0.88129] + (0.25) [-0.468996]\}$
$=$ 0.5437 Sh

0/4

d) *p(x1) = 0.75*
p(x2) = 0.25

$H(X) = -[p(x1) \log_2 p(x1) + (p(x2)) \log_2 (p(x2))]$
$= -[0.75 \log_2 0.75 + (0.25 \log_2 0.25]$
$=$ 0.8113 Sh 4/4

e) 4/4

- 0.5437 Sh (avec mémoire)
- 0.8113 Sh (sans mémoire)

L'entropie avec mémoire est **moindre** que l'entropie sans mémoire puisque celle-ci se voit réduite en raison de l'information additionnelle disponible du passé, ce qui réduit donc l'incertitude du canal.

- **Problème 3 : Convexité de l'information mutuelle** 5/5

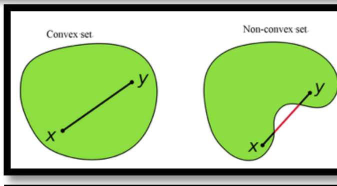
Problème 5.1 : L'information mutuelle $I(X;Y)$ est une fonction convexe $\cap$ (i.e. *fonction concave*) définie sur l'ensemble convexe des distributions des symboles d'entrée $\mathcal{S}_{\mathbf{p}} = \{\mathbf{p}\}$. Cependant, $I(X;Y)$ est une fonction convexe $\cup$ (*fonction convexe*) sur l'ensemble convexe des matrices de probabilités de transition d'un canal $\mathcal{S}_{\mathbf{P}} = \{\mathbf{P}\}$.

a) Démontrez que l'ensemble convexe des matrices de probabilités de transition $\mathcal{S}_{\mathbf{P}}$ forme un ensemble convexe.

b) Démontrez que sur cet ensemble convexe $\mathcal{S}_{\mathbf{P}}$, l'information mutuelle est une fonction convexe $\cup$.

$$\mathbf{P} = \begin{bmatrix} p\left(x_1^{(n+1)} \middle| x_1^{(n)}\right) & p\left(x_2^{(n+1)} \middle| x_1^{(n)}\right) \\ p\left(x_1^{(n+1)} \middle| x_2^{(n)}\right) & p\left(x_2^{(n+1)} \middle| x_2^{(n)}\right) \end{bmatrix} = \begin{bmatrix} (1-\alpha) & \alpha \\ \beta & (1-\beta) \end{bmatrix}$$

a) Ensemble de toutes les distributions possibles → Matrice de probabilités de transition P = S_p → Paramètre α & β



Définition (Fonction convexe) : Un ensemble de points S est convexe si, pour toute paire de points x1 ∈ S et x2 ∈ S, tout point x reliant ces deux points x1 et x2 est lui aussi compris dans l'ensemble de points S.

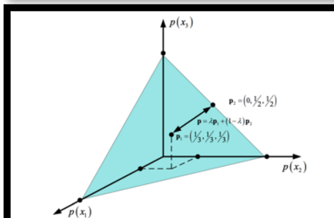


FIGURE 5.2 – Exemple d'ensemble convexe.

- Deux conditions:

1) $0 \leqslant p(x_k) \leqslant 1$, pour k = 1, 2, 3, 4
2) $\Sigma\, p(x_k) = 1$

1) Soient x1 et x2, deux points dans un espace à 2 dimensions.

Tout point x sur la droite reliant les points x1 et x2 peut s'exprimer par :

$x = \lambda x1 + (1 - \lambda)\, x2$, où $\lambda \in [0,1]$

$\lambda = 1$ → $x = x1$
$\lambda = 0$ → $x = x2$
$0 \leqslant \lambda \leqslant 1$, le point x est situé sur la droite reliant x1 et x2.

- Sachant que dû aux lois des probabilités :
$0 \leqslant (\alpha) \leqslant 1$
$0 \leqslant (\beta) \leqslant 1$,
alors : $p(x_k) = \lambda p1\,(x_k) + (1 - \lambda)\, p2\,(x_k) \geqslant 0$, pour k = 1, 2.

2) Since the total of transition probability from a state *i* to all other states must be 1

- $\Sigma\, \beta + (1-\beta) = 1$

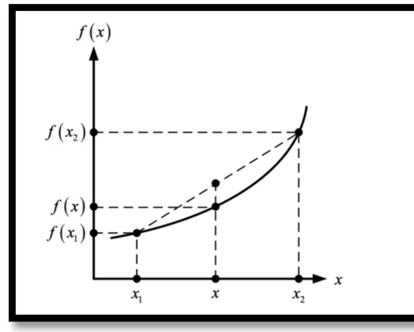
- $\Sigma\, \alpha + (1-\alpha) = 1$

Donc tout point $p = \lambda p1 + (1 - \lambda)\, p2$,
Défini avec deux distributions:

P1 = $[\beta_1, \alpha_1]$ ou {p1 (y | x)}
P2 = $[\beta_2, \alpha_2]$ ou p2 (y | x)}

est lui-même une distribution valide, et ce pour tout choix de paires de distributions p1 et p2. **Donc l'ensemble Sp de toutes les distributions à dimensions (ici N = 2) forment un ensemble convexe.**

b) Démontrez que sur cet ensemble convexe S_P, l'information mutuelle est une fonction convexe ∪.



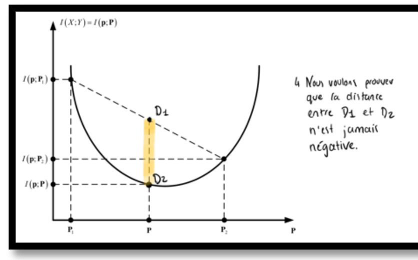
- Définition (Fonction convexe) : Une fonction réelle f (x), définie sur un ensemble convexe S est dite convexe (convexe vers le haut ou convexe ∪) si, pour tout point x sur la droite reliant la paire de points x1 et x2, i.e. $x = \lambda x1 + (1 - \lambda)\, x2$, $(\lambda \in [0, 1])$:

$$f(x) \leqslant \lambda f(x1) + (1 - \lambda)\, f(x2)$$

$$I_M(X,Y) = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)\, p(y_j)}$$

$I(X; Y) = f(p, P) = I(p; P)$

Fixons la distribution de symbole p et faisons varier les probabilités de transition P



Donc assumons 2 distributions différentes de la matrice de probabilités de transition P :

- P1 avec = {p1 (y | x)}
- P2 avec = {p2 (y | x)}

La distribution P, entre les distributions P1 et P2 dans l'ensemble convexe Sp est donnée par

$P = \lambda P1 + (1 - \lambda)\, P2$

$P(yj \mid xk) = \lambda\, P1(yj \mid xk) + (1 - \lambda)\, P2(yj \mid xk)$ pour un $\lambda \in [0, 1]$

Soit I, I_1, I_2 l'information mutuelle moyenne de P, P1, et P2 respectivement. Il est donc nécessaire de prouver que: $\lambda I_1 + (1 - \lambda) I_2 \geq I$

Donc si l'information I (X; Y) = I (p; P) est une fonction convexe de la distribution de l'ensemble des paramètres α & β, alors :

$I(X; Y) \leqslant \lambda I(p(x), p1(y|x)) + (1 - \lambda)\, I((p(x), p2(y|x))$

Si c'est le cas, alors :

→ $\lambda I(p(x), p1(y|x) + (1 - \lambda)\, I(p(x), p2(y|x) - I(X; Y) \geqslant 0$

→ $\lambda I(p(x), p1(y|x) + (1 - \lambda)\, I(p(x), p2(y|x) - I(X; Y) =$

$$\lambda \Sigma\Sigma\, p(x_k)\, p1(y_j|x_k) \log_b (p1(y_j|x_k) / p(y_j)) + (1 - \lambda)\, \Sigma\Sigma\, p(x_k)\, p2(y_j|x_k) \log_b (p2(y_j|x_k) / p(y_j)) - \Sigma\Sigma\, p(x_k)\, p(y_j|x_k) \log_b (p(y_j|x_k) / p(y_j))$$

…

(Même démarche que p.90/p.91 des notes)

De ce fait, **l'information mutuelle moyenne I (X; Y) est une fonction convexe (∪) sur l'ensemble convexe S_P de toutes les matrices de probabilité de transition possibles {P}**

- **Autre méthode pour prouver convexité de la fonction I (X; Y) ?**

J'ai remarqué à plusieurs reprises l'utilisation d'une seconde méthode beaucoup plus courte que celle que vous avez utilisez afin de prouver la convexité de l'information sur l'ensemble convexe S_P de toutes les matrices de probabilité de transition possibles. Elle consiste à introduire une variable aléatoire Z où I (X ; Z) = 0... Voilà un exemple:

1 Convexity/Concavity of Mutual Information

In the previous lecture, we saw that mutual information is concave in p. To be more precise, let (X,Y) have a joint probability distribution $p(x,y) = p(x)p(y|x)$. Write $\alpha = \alpha(x) = p(x)$ and $\pi = \pi(x,y) = p(y|x)$. Then the pair (α, π) specifies the distribution $p(x,y)$.

Lemma 1 (Mutual information is concave in p)**.**
Let I_1 be $I(X;Y)$ where $(X,Y) \sim (\alpha_1, \pi)$,
let I_2 be $I(X;Y)$ where $(X,Y) \sim (\alpha_2, \pi)$,
let I be $I(X;Y)$ where $(X,Y) \sim (\lambda\alpha_1 + (1-\lambda)\alpha_2, \pi)$, for some $0 \leq \lambda \leq 1$.
then $I \geq \lambda I_1 + (1-\lambda) I_2$.

Now, we prove that mutual information is convex in $p(y|x)$. More formally, we have the following. Let (X,Y) have a joint probability distribution $p(x,y) = p(x)p(y|x)$. Write $\alpha = \alpha(x) = p(x)$ and $\pi = \pi(x,y) = p(y|x)$. Then the pair (α, π) specifies the distribution $p(x,y)$.

Lemma 2 (Mutual information is convex in π)**.** *Let I_1 be $I(X;Y)$ where $(X,Y) \sim (\alpha, \pi_1)$,*
let I_2 be $I(X;Y)$ where $(X,Y) \sim (\alpha, \pi_2)$,
let I be $I(X;Y)$ where $(X,Y) \sim (\alpha, \lambda\pi_1 + (1-\lambda)\pi_2)$, for some $0 \leq \lambda \leq 1$.
then $I \leq \lambda I_1 + (1-\lambda) I_2$.

Proof Let us draw X first according to α. Let S be a B_λ random variable such that S is 1 with probability λ and 0 with probability $1-\lambda$. If $S = 1$ we select Y using π_1, and otherwise we select Y using π_2. Note that $I(X;Y) = I$.

$$I(SY;X) = I(Y;X) + I(S;X|Y) \geq I(Y;X) = I$$

Also, we have

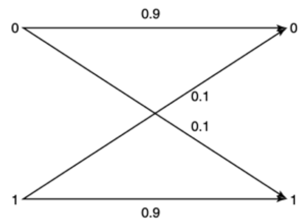
$$\begin{aligned} & I(SY;X) \\ = \; & I(S;X) + I(Y;X|S) \\ = \; & 0 + I(Y;X|S) \\ = \; & \lambda I(Y;X|S=1) + (1-\lambda) I(Y;X|S=0) \\ = \; & \lambda I_1 + (1-\lambda) I_2 \end{aligned}$$

Thus, we have $I \leq \lambda I_1 + (1-\lambda) I_2$.

- **Problème 4 : Séquences conjointement typique**

20/20

Jointly typical sequences. As we did in Problem 3.13 for the typical set for a single random variable, we will calculate the jointly typical set for a pair of random variables connected by a binary symmetric channel, and the probability of error for jointly typical decoding for such a channel.



We consider a binary symmetric channel with crossover probability 0.1. The input distribution that achieves capacity is the uniform distribution [i.e., $p(x) = (\frac{1}{2}, \frac{1}{2})$], which yields the joint distribution $p(x, y)$ for this channel is given by

$X \backslash Y$	0	1
0	0.45	0.05
1	0.05	0.45

The marginal distribution of Y is also $(\frac{1}{2}, \frac{1}{2})$.

(a) Calculate $H(X)$, $H(Y)$, $H(X, Y)$, and $I(X; Y)$ for the joint distribution above.

(b) Let $X_1, X_2, \ldots, X_n$ be drawn i.i.d. according the Bernoulli($\frac{1}{2}$) distribution. Of the 2^n possible input sequences of length n, which of them are typical [i.e., member of $A_\epsilon^{(n)}(X)$ for $\epsilon = 0.2$]? Which are the typical sequences in $A_\epsilon^{(n)}(Y)$?

(c) The jointly typical set $A_\epsilon^{(n)}(X, Y)$ is defined as the set of sequences that satisfy equations (7.35-7.37). The first two equations correspond to the conditions that x^n and y^n are in $A_\epsilon^{(n)}(X)$ and $A_\epsilon^{(n)}(Y)$, respectively. Consider the last condition, which can be rewritten to state that $-\frac{1}{n}\log p(x^n, y^n) \in (H(X, Y) - \epsilon, H(X, Y) + \epsilon)$. Let k be the number of places in which the sequence x^n differs from y^n (k is a function of the two sequences). Then we can write

$$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i) \qquad (7.156)$$

$$= (0.45)^{n-k}(0.05)^k \qquad (7.157)$$

$$= \left(\frac{1}{2}\right)^n (1-p)^{n-k} p^k. \qquad (7.158)$$

An alternative way at looking at this probability is to look at the binary symmetric channel as in additive channel $Y = X \oplus Z$, where Z is a binary random variable that is equal to 1 with probability p, and is independent of X. In this case,

$$p(x^n, y^n) = p(x^n)p(y^n|x^n) \qquad (7.159)$$

$$= p(x^n)p(z^n|x^n) \qquad (7.160)$$

$$= p(x^n)p(z^n) \qquad (7.161)$$

$$= \left(\frac{1}{2}\right)^n (1-p)^{n-k} p^k. \qquad (7.162)$$

Show that the condition that (x^n, y^n) being jointly typical is equivalent to the condition that x^n is typical and $z^n = y^n - x^n$ is typical.

(d) We now calculate the size of $A_\epsilon^{(n)}(Z)$ for $n = 25$ and $\epsilon = 0.2$. As in Problem 3.13, here is a table of the probabilities and numbers of sequences with k ones:

k	$\binom{n}{k}$	$\binom{n}{k}p^k(1-p)^{n-k}$	$-\frac{1}{n}\log p(x^n)$
0	1	0.071790	0.152003
1	25	0.199416	0.278800
2	300	0.265888	0.405597
3	2300	0.226497	0.532394
4	12650	0.138415	0.659191
5	53130	0.064594	0.785988
6	177100	0.023924	0.912785
7	480700	0.007215	1.039582
8	1081575	0.001804	1.166379
9	2042975	0.000379	1.293176
10	3268760	0.000067	1.419973
11	4457400	0.000010	1.546770
12	5200300	0.000001	1.673567

[Sequences with more than 12 ones are omitted since their total probability is negligible (and they are not in the typical set).] What is the size of the set $A_\epsilon^{(n)}(Z)$?

a)

5/5

H (X) = **1 bit**
H (Y) = **1 bit**

H (X, Y) = H (X) + H (Y | X)
= 1 − [0.9 log $_2$ 0.9 − 0.1 log $_2$ 0.1]
= **1.469 bits**

I (X; Y) = H (Y) – H (Y|X)
= **0.531 bits**

b) Of the 2^n possible input sequences of length n, which of them are typical [member of $A^{(n)}_{\epsilon}(X)$] for $\epsilon = 0.2$? Which are the typical sequences in $A^{(n)}_{\epsilon}(Y)$?

— l'ensemble $\mathcal{T}_X(\delta)$ des séquences typiques de longueur N :

$$\mathcal{T}_X(\delta) \triangleq \left\{ \mathbf{x} : \left| -\frac{1}{N} \log_b p(\mathbf{x}) - H(X) \right| \leqslant \delta \right\}$$

- If uniform distribution:

$p = (1/2)^n$

$[(-1/N) \log_b p(x)] = 1$

$H(X) = 1$

$1 - 1 = 0$

$0 \le \delta$

5/5

→ Every sequence x^n is typical $\in A^{(n)}_{\epsilon}(X)$

→ Every sequence y^n is typical $\in A^{(n)}_{\epsilon}(Y)$

c) Show that the condition that (x^n, y^n) being jointly typical is equivalent to the condition that x^n is typical and $z^n = y^n - x^n$ is typical.

7.6 JOINTLY TYPICAL SEQUENCES

Roughly speaking, we decode a channel output Y^n as the ith index if the codeword $X^n(i)$ is "jointly typical" with the received signal Y^n. We now define the important idea of joint typicality and find the probability of joint typicality when $X^n(i)$ is the true cause of Y^n and when it is not.

Definition The set $A_\epsilon^{(n)}$ of *jointly typical* sequences $\{(x^n, y^n)\}$ with respect to the distribution $p(x, y)$ is the set of n-sequences with empirical entropies ϵ-close to the true entropies:

$$A_\epsilon^{(n)} = \Big\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n :$$

1 $$\left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon, \qquad (7.35)$$

196 CHANNEL CAPACITY

2 $$\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| < \epsilon, \qquad (7.36)$$

3 $$\left| -\frac{1}{n} \log p(x^n, y^n) - H(X, Y) \right| < \epsilon \Big\}, \qquad (7.37)$$

- Conditions (x^n, y^n) are jointly typical (image):

1 | 2 | 3

Every sequence xn and yn satisfies 1 & 2

Hence looking at condition 3

and replacing $p(x^n, y^n)$

We get:

5/5

$$| -(1/n) \log ((1/2)^n p^k (1-p)^{n-k}) - H(X, Y) | < \varepsilon$$

Which is similar to condition for typicality of $z^n = y^n \oplus x^n$

$$\begin{aligned} p(x^n, y^n) &= \prod_{i=1}^{n} p(x_i, y_i) && (7.73) \\ &= (0.45)^{n-k}(0.05)^k && (7.74) \\ &= \left(\frac{1}{2}\right)^n (1-p)^{n-k} p^k && (7.75) \end{aligned}$$

An alternative way at looking at this probability is to look at the binary symmetric channel as in additive channel $Y = X \oplus Z$, where Z is a binary random variable that is equal to 1 with probability p, and is independent of X. In this case,

$$\begin{aligned} p(x^n, y^n) &= p(x^n)p(y^n|x^n) && (7.76) \\ &= p(x^n)p(z^n|x^n) && (7.77) \\ &= p(x^n)p(z^n) && (7.78) \\ &= \left(\frac{1}{2}\right)^n (1-p)^{n-k} p^k && (7.79) \end{aligned}$$

d) What is the size of the set $A^{(n)}_{\epsilon}(Z)$?

$n = 25$

$\varepsilon = 0.2$

5/5

$H(Z) = 0.469$

Typical set for Z: $-(1/n) \log_b p(z^n) \in (H(Z) - \varepsilon, H(Z) + \varepsilon)$

$\in (0.269; 0.669)$

-- Typical Z sequences → $k = 1,2,3,4$

Total probability of set $A^{(n)}_{\epsilon}(Z) = 0.902 - 0.071$

$= 0.830$ %

Size = $15276 - 1$

= 15 275 sequences

- **Problème 5 : Capacité d'un canal asymétrique** 15/15

Problème 6.5 : Pour ce problème, écrivez un programme pour calculer la capacité C de canaux asymétriques. Utilisez l'algorithme de Blahut-Arimoto vue en classe. Donnez le listing de votre programme. Note : vous pouvez vous référer entre autres à la section 5.4 du livre "Principles and Practice of Information Theory" de Richard E. Blahut ou à la section correspondante de livre "Elements of Information Theory" de Thomas M. Cover et Joy A. Thomas. Calculez la capacité des canaux de communication suivants :

a) Déterminez la capacité C_1 du canal caractérisé par la matrice de probabilité de transition $\mathbf{P}_1$ et donnez la distribution des symboles de source $\mathbf{p}_1^*$ correspondante.

$$\mathbf{P}_1 = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.7 \end{pmatrix}$$

b) Déterminez la capacité C_2 du canal caractérisé par la matrice de probabilité de transition $\mathbf{P}_2$ et donnez la distribution des symboles de source $\mathbf{p}_2^*$ correspondante.

$$\mathbf{P}_2 = \begin{pmatrix} 0.4 & 0.2 & 0.4 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.2 & 0.4 & 0.0 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.4 & 0.0 & 0.0 & 0.4 & 0.2 \\ 0.2 & 0.4 & 0.0 & 0.0 & 0.4 \end{pmatrix}$$

c) Déterminez la capacité C_3 du canal caractérisé par la matrice de probabilité de transition $\mathbf{P}_3$ et donnez la distribution des symboles de source $\mathbf{p}_3^*$ correspondante.

$$\mathbf{P}_3 = \begin{pmatrix} 0.11 & 0.25 & 0.04 & 0.10 & 0.20 & 0.00 & 0.12 & 0.18 \\ 0.27 & 0.00 & 0.14 & 0.05 & 0.13 & 0.26 & 0.10 & 0.05 \\ 0.11 & 0.09 & 0.18 & 0.32 & 0.16 & 0.08 & 0.04 & 0.02 \\ 0.00 & 0.10 & 0.15 & 0.25 & 0.25 & 0.15 & 0.10 & 0.00 \\ 0.11 & 0.27 & 0.09 & 0.03 & 0.03 & 0.09 & 0.27 & 0.11 \\ 0.25 & 0.00 & 0.00 & 0.25 & 0.25 & 0.00 & 0.00 & 0.25 \\ 0.27 & 0.00 & 0.14 & 0.05 & 0.13 & 0.26 & 0.10 & 0.05 \\ 0.11 & 0.25 & 0.04 & 0.10 & 0.20 & 0.00 & 0.12 & 0.18 \end{pmatrix}$$

a)

C1 = 0.2091
Distribution p1* = [0.5255, 0.4744] 5/5

b)

C2 = 0.5293
Distribution p2* = [0.2564, 0.2436, 3.42435657e-06, 0.2438, 0.2563] 5/5

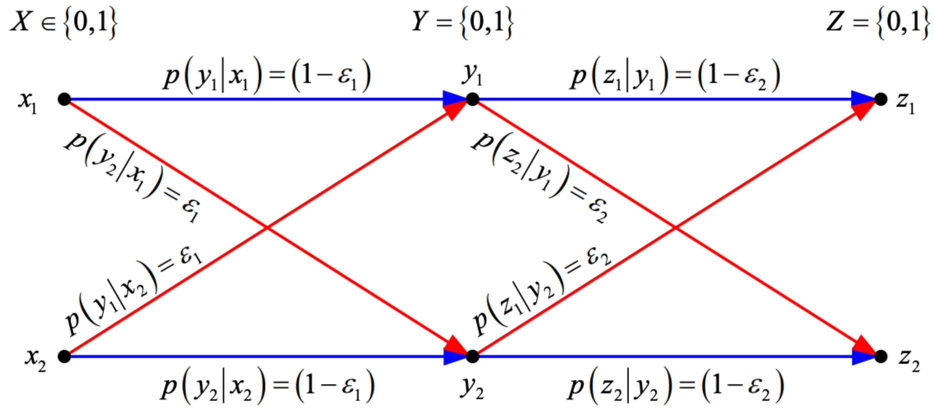
c) 5/5

C3 = 0.4106
Distribution p3* = [4.268e-09, 0.04525, 7.234e-08, 0.1926, 0.2792, 0.4529, 0.030038, 3.458e-09]

Problème 6 : Capacité de canaux symétriques en cascade

16/20

La figure ci-dessous montre deux canaux symétriques binaires en cascade.



a) Écrivez l'expression de l'information mutuelle du canal $I(X;Y)$ en fonction de ϵ_1.

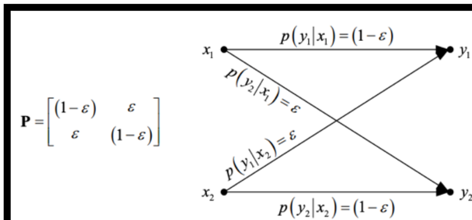
b) Écrivez l'expression de l'information mutuelle du canal $I(X;Z)$ en fonction de ϵ_1 de ϵ_2.

c) Donnez capacité du canal C_{XY} pour $\epsilon_1 = 0.1$.

d) Donnez capacité du canal C_{XZ} pour $\epsilon_1 = 0.1$ et $\epsilon_2 = 0.2$.

e) À l'aide d'un logiciel, tracez les courbes suivantes :
- C_{XY} pour $0 \leq \epsilon_1 \leq 1$.
- C_{XZ} pour $0 \leq \epsilon_1 \leq 1$ et $0 \leq \epsilon_2 \leq 1$ (graphique 3D). Expliquez les résultats obtenus.

a) Mutual information can be seen as X uncertainty less error sequence uncertainty



canal symétrique binaire

Because it is a symmetric channel → input distribution p∗ is an equiprobable source distribution: p(x1) = p(x2) = 1/2 X

H (X) = 1 Sh

H (E) = - Σ p (yj |xk) log b p (yj |xk)

= - [(1 − ε1) log $_b$ (1 − ε1) + ε1 log $_b$ ε1]

 # E as row of the transition matrix

I (X; Y) = H (X) - H (E)

= 1 + [(1 − ε) log $_b$ (1 − ε) + ε log $_b$ ε]

0/4

b) The channel has the transition matrix:

$$P = \begin{bmatrix} 1-\varepsilon_1 & \varepsilon_1 \\ \varepsilon_1 & 1-\varepsilon_1 \end{bmatrix} \begin{bmatrix} 1-\varepsilon_2 & \varepsilon_2 \\ \varepsilon_2 & 1-\varepsilon_2 \end{bmatrix}$$

$$= \begin{bmatrix} (1-\varepsilon_1)(1-\varepsilon_2) + \varepsilon_1 \varepsilon_2 & (1-\varepsilon_1)\varepsilon_2 + \varepsilon_1(1-\varepsilon_2) \\ \varepsilon_1 \cdot (1-\varepsilon_2) + (1-\varepsilon_1)(\varepsilon_2) & \varepsilon_1 \cdot \varepsilon_2 + (1-\varepsilon_1)(1-\varepsilon_2) \end{bmatrix}$$

Again, because it is a symmetric channel → input distribution p∗ is an equiprobable source distribution: p(x1) = p(x2) = 1/2 X

H (X) = 1 Sh

H (E) = - Σ p (yj |xk) log b p (yj |xk)

= - [(1 − ε) log $_b$ (1 − ε) + ε1 log $_b$ ε] *# E as row of the transition matrix*

The two cascaded BSC channels can be viewed as a single BSC channel with an overall loss parameter ε = ε1(1 - ε2) + ε2 (1 - ε1) 4/4

I (X; Z) = H (X) - H (E)

= 1 + [(1 − ε) log $_b$ (1 − ε) + ε log $_b$ ε]

= 1 + [(1 − [ε1(1 - ε2) + ε2 (1 - ε1)]) log $_b$ (1 − [ε1(1 - ε2) + ε2 (1 - ε1)]) + [ε1(1 - ε2) + ε2 (1 - ε1)] log $_b$ [ε1(1 - ε2) + ε2 (1 - ε1)]]

c) ε1 = 0.1

C = Max I (X; Y)

= 1 + [(1 − 0.1) log $_2$ (1 − 0.1) + 0.1 log $_2$ 0.1]

= 1 + [-0.1368 - 0.33219]

= 0.5310 Sh

4/4

d) ε1 = 0.1 & ε2 = 0.2

The two cascaded BSC channels can be viewed as a single BSC channel with an overall loss parameter ε:

ε = ε1(1 - ε2) + ε2 (1 - ε1)

ε = 0.1 (1 - 0.2) + 0.2 (1 - 0.1)

ε = 0.26

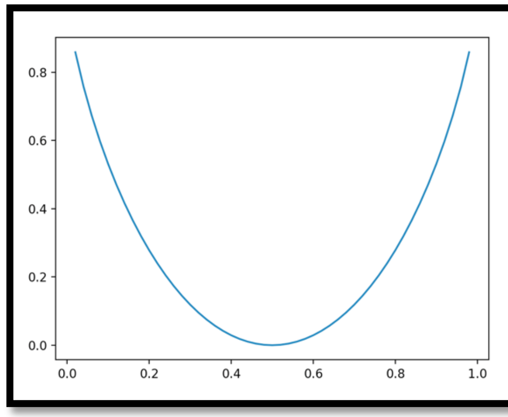
C = Max I (X; Y)

= 1 + [(1 − ε) log $_b$ (1 − ε) + ε log $_b$ ε]

$= 1 + [(1 - 0.26) \log_b (1 - 0.26) + 0.26 \log_b 0.26]$
$= 1 + [(0.74) \log_2 (0.74) + 0.26 \log_2 0.26]$
$= 1 + [- 0.32146 - 0.5053]$
$=$ 0.1732 Sh

4/4

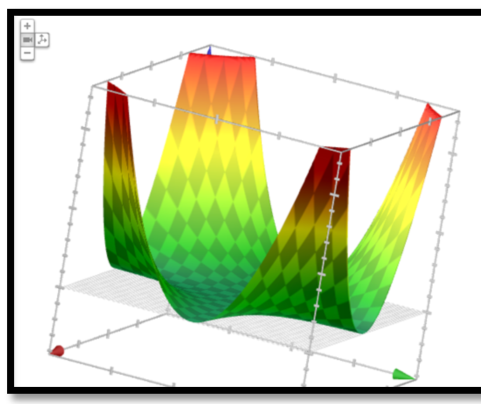
e) 4/4



- **C_{XY} pour 0 <= ε1 <= 1**

0 <= ε <= 1
$C = 1 + [(1 - \varepsilon) \log_b (1 - \varepsilon) + \varepsilon \log_b \varepsilon]$

- **C_{XZ} pour 0 <= ε1 <= 1 & 0 <= ε2 <= 1**



0 <= ε1 <= 1
0 <= ε2 <= 1

$C_{XZ} = 1 + [(1 - [\varepsilon 1(1 - \varepsilon 2) + \varepsilon 2 (1 - \varepsilon 1)]) \log_b (1 - [\varepsilon 1(1 - \varepsilon 2) + \varepsilon 2 (1 - \varepsilon 1)]) + [\varepsilon 1(1 - \varepsilon 2) + \varepsilon 2 (1 - \varepsilon 1)] \log_b [\varepsilon 1(1 - \varepsilon 2) + \varepsilon 2 (1 - \varepsilon 1)]]$

- **Explanation**

The more the error is close to 0.5, the less information can be transmitted.
(symmetry of p & 1 - p)
If we gain certainty about a high or a low error rate, we remove uncertainty from the channel.
Example: if error rate approach 1, the information can be almost intact when we reverse it.