Université Laval
Faculté des sciences et de génie
Département d'informatique et de génie logiciel
IFT-4003/7023

Danny Dubé
Hiver 2019
Version : 24 février

Travail pratique #1
Introduction – codage arithmétique

Questions

Donnez le développement pour vos réponses, en particulier quand le calcul est long.

1. (*15 points.*) **Modèles de Markov.** Analysez les phénomènes aléatoires suivants et décrivez-les en tant que modèles de Markov. À cette fin, il faut les modéliser sous la forme d'automates ayant des états et des transitions entre les états (comme la modélisation à deux états de l'image en noir et blanc du petit garçon devant le tableau). Vos automates doivent avoir le plus petit nombre d'états possible. Aussi, veuillez indiquer l'ordre de chaque modèle de Markov en justifiant votre réponse.

   (a) On joue avec un assortiment de quatre dés : Δ_1, Δ_2, Δ_3 et Δ_4. Les dés Δ_1, Δ_2, Δ_3 et Δ_4 ont 2, 3, 4 et encore 4 faces, respectivement. Les quatre dés sont équitables. Un dé qui a k faces possède des faces marquées avec les nombres 1 à k. Si, à un tour donné, on obtient la face marquée i, alors on doit utiliser le dé Δ_i au tour suivant.

   (b) On définit la suite aléatoire $\{N_i\}_{i=0}^{\infty}$ de nombres naturels. On initialise la suite à $N_0 = 1$. Au $i$ème tour, on lance un dé équitable à N_i faces, où les faces sont numérotées de 1 à N_i. Si le résultat est impair, alors N_{i+1} est égal à $N_i + 1$, sinon, N_{i+1} est égal à $N_i - 1$.

(c) On définit la suite aléatoire $\{C_i\}_{i=1}^{\infty}$ de chiffres dans $\Sigma = \{1, 2, 3\}$. Le tirage fixant la valeur prise par chaque variable aléatoire de la suite est fait de manière passablement compliquée. Supposons qu'après avoir procédé au tirage des n premières variables aléatoires $C_1, \ldots, C_n$, on souhaite faire le tirage de C_{n+1}. Soient $c_1, \ldots, c_n$ les chiffres obtenus lors des n premiers tirages. On calcule une clé $w_n \in \Sigma^*$ à partir de ces chiffres et enfin on tire C_{n+1} à partir de la clé. La clé est calculée ainsi :

$$w_n \quad = \quad f(2121,\, c_1 \ldots c_n)$$

et le chiffre C_{n+1} résulte du choix aléatoire d'une position, où une position est l'une de $1, \ldots, |w_n|$ avec probabilités uniformes, et de l'extraction du chiffre situé en cette position dans w_n. Le processus peut sembler exagérément complexe mais il suffit de maintenir à jour la clé à chaque étape (ou une connaissance suffisante de celle-ci). Il n'existe qu'un nombre restreint de clés possibles.

Voici la définition de la fonction f :

$$\begin{aligned} f(w,\, \epsilon) &= w \\ f(w,\, c_1\, c_2 \ldots c_k) &= f(\text{suffix}(w,\, c_1 + 1) \cdot d_{c_1},\, c_2 \ldots c_k) \end{aligned}$$

où $\text{suffix}(w, l)$ est une fonction qui extrait le suffixe de longueur l de la chaîne w et où $\{d_i\}_{i=1}^{3}$ est un dictionnaire de morceaux de clé défini ainsi :

$$d_1 = 21 \;, \qquad d_2 = 312 \;, \qquad d_3 = 1 \;.$$

2. (*10 points.*) **Entropie.**

(a) Vous devez concevoir une suite infinie $\{\Delta_n\}_{n=2}^{\infty}$ de dés en respectant les contraintes suivantes. Premièrement, le dé Δ_n possède n faces. Deuxièmement, chaque dé possède une face préférée. Les autres faces sont moins sujettes à se faire sélectionner (excepté chez Δ_2) et elles sont équiprobables entre elles. Autrement dit, la face préférée du dé Δ_n a la probabilité p_n de se faire sélectionner tandis que chacune des autres faces a la probabilité $\frac{1-p_n}{n-1}$ de se faire sélectionner. Troisièmement, vous devez fixer la probabilité p_n de chaque dé Δ_n de telle façon à ce que l'entropie de Δ_n soit 1. Donnez une formule (ou, à tout le moins, un système d'équations) qui détermine chacune des p_n. Donnez explicitement la valeur de p_2, p_3, p_4 et p_5. Utilisez une approximation numérique pour ce faire, si nécessaire.

(b) Une image en trois tons de gris est modélisée par un modèle de Markov à trois états, un peu comme dans l'exemple de l'image du petit garçon. Supposons que, lorsqu'on prédit le ton de gris d'un pixel, on se fie au ton de gris de son voisin de gauche. La matrice suivante montre quelle est la probabilité de chaque ton de gris chez le pixel prédit en fonction du ton de gris du voisin de gauche. Calculez l'entropie d'ordre 1 de ce modèle.

voisin\prédit	noir	gris	blanc
noir	80%	15%	5%
gris	15%	70%	15%
blanc	5%	15%	80%

Rappel : il faut d'abord calculer les probabilités stationnaires des trois états, puis calculer l'entropie conditionnelle de chaque état et enfin calculer l'entropie du modèle complet.

3. (*20 points.*) **Codes uniquement décodables.** Pour chacun des codes suivants, indiquez s'il est uniquement décodable. Si oui, donnez une justification (par exemple, une trace de l'algorithme vu en classe). Si non, donnez aussi une justification (par exemple, une chaîne de bits qui est décodable de deux façons différentes).

Code 1	Code 2	Code 3	Code 4	Code 5
a : 000	a : 0000	a : 000	a : 00	a : 000
b : 10	b : 0001	b : 001	b : 01000	b : 00001
c : 100	c : 001	c : 010	c : 010010	c : 000111
d : 1000	d : 00110	d : 0101000	d : 010011	d : 001
e : 110	e : 0100	e : 0101001	e : 0101	e : 001011
f : 1100	f : 0101	f : 011	f : 011	f : 001111
g : 11000	g : 010110	g : 0111000	g : 100	g : 0101
h : 1110	h : 011	h : 0111001	h : 1010	h : 01111
i : 11100	i : 01110	i : 10001000	i : 1011	i : 10
j : 111000	j : 1010	j : 10001001	j : 110	j : 110
k : 1111	k : 11	k : 10011000	k : 111	k : 11001
		l : 10011001		l : 110111
		m : 101		m : 11101
		n : 1011000		n : 111111
		o : 1011001		
		p : 110		
		q : 1110		

4. (*10 points.*) **Codes préfixes.** Soit l'alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{f}, \mathtt{g}, \mathtt{h}, \mathtt{i}\}$. Dans chacun des cas si-dessous, indiquez s'il existe un code préfixe dont la longueur du mot de code de chaque symbole est celle indiquée dans le tableau. Si oui, donnez un tel code préfixe. Si non, expliquez pourquoi.

   (a)

a	b	c	d	e	f	g	h	i
3	4	5	3	5	4	2	3	2

   (b)

a	b	c	d	e	f	g	h	i
6	5	3	2	6	3	4	3	2

5. (*10 points.*) **Codes préfixes.** Considérez l'alphabet Σ suivant ainsi que les fréquences d'apparition associées.

a	b	c	d	e	f	g	h	i	j	k	l
6	5	2	16	13	1	11	8	12	15	7	4

   (a) Construisez un code préfixe pour Σ en utilisant l'algorithme de Shannon-Fano. Calculez la longueur moyenne des mots de code.

   (b) Construisez un code préfixe pour Σ en utilisant l'algorithme de Huffman. Calculez la longueur moyenne des mots de code.

6. (*10 points.*) **Codes de Golomb.** Considérons l'alphabet des nombres naturels. La probabilité du nombre i, $i \geq 0$, est :

$$p_i \quad = \quad \begin{cases} \frac{3}{5} \cdot \left(\frac{1}{2}\right)^{\lfloor i/2 \rfloor + 1}, & \text{si } i \text{ est pair,} \\ \frac{2}{5} \cdot \left(\frac{1}{2}\right)^{\lfloor i/2 \rfloor + 1}, & \text{si } i \text{ est impair.} \end{cases}$$

Considérez le code de Golomb de paramètre $m = 2$. Montrez la forme des mots de code pour les divers naturels. Aussi, donnez explicitement les mots de code pour les naturels 0 à 7. Enfin, calculez la longueur moyenne du mot de code d'un naturel tiré aléatoirement en fonction des probabilités.

Voici deux identités qui peuvent vous être utiles.

$$\sum_{i=0}^{\infty} r^i \quad = \quad \frac{1}{1-r} \qquad\qquad \sum_{i=0}^{\infty} i \cdot r^i \quad = \quad \frac{r}{(1-r)^2}$$

7. (*10 points.*) **Codes de Tunstall.** Construisez le code de Tunstall dont le dictionnaire a la capacité de contenir 12 mots. L'alphabet source et les probabilités des symboles sont les suivants.

a	b	c
$\frac{3}{10}$	$\frac{2}{10}$	$\frac{5}{10}$

8. (*15 points.*) **Codage arithmétique.** Soient l'alphabet et les probabilités des symboles suivants.

a	b	c
$\frac{3}{10}$	$\frac{2}{10}$	$\frac{5}{10}$

Vous devez décrire la gestion de l'intervalle de travail au cours de l'encodage de la chaîne `cacb`, en commençant par l'intervalle complet. Il y a cinq moments distincts : le moment précédant l'encodage du premier symbole et chacun des moments qui suivent immédiatement l'encodage d'un symbole. Vous devez faire les redimensionnements de l'intervalle de travail (à l'aide de E_1, E_2 et E_3) au fur et à mesure. Faites deux fois l'exercice : une fois en manipulant des fractions exactes et une autre fois en nombres entiers, où l'intervalle complet comporte 16 tuiles.

Remise des travaux

Vous devez remettre le travail dans la boîte de dépôt de **monPortail** au plus tard le **6 mars**. Veuillez ignorer la date de remise inscrite dans le plan de cours. Les autres modalités de remise sont inscrites dans le plan de cours.