Informatique & Aléatoire
Licence L3 2014-2015

Université Joseph Fourier

Durée: 2 hours, Documents non autorisés
Traiter chaque problème sur une feuille séparée

Problème 1. Problème de Monty Hall

Dans le problème de Monty Hall inspiré du jeu télévisé *Let's Make a Deal*, un présentateur est opposé à un joueur qui est placé devant trois portes fermées. Derrière l'une des portes se trouve un prix de valeur et derrière chacune des deux autres se trouve un prix sans importance. Le présentateur demande au joueur de désigner une des portes où le joueur pense que le prix de valeur se trouve. Il ouvre ensuite la porte qui n'est pas désignée par le joueur et où se trouve un des prix sans importance. Le candidat a alors le choix d'ouvrir la porte qu'il avait désignée initialement ou la troisième porte.

Posons : $A_i, i \in \{1,2,3\}$, l'événement : *le prix de valeur se trouve derrière la porte numéro i*. Supposons que le joueur désigne la porte $i = 2$ et que le présentateur choisisse d'ouvrir, parmi les portes 1 et 3, la porte 3. Posons E ce dernier événement, i.e. *E : le présentateur ouvre la porte 3 sachant que la porte désignée par le joueur est la 2*.

Question 1 : Probabilités a priori des causes

Quelles sont les probabilités a priori des causes ; $\mathbf{P}(A_i), i \in \{1,2,3\}$?

Question 2 : Probabilités conditionnelles de la conséquence

Quelles sont les probabilités conditionnelles de la conséquence E sachant ces causes ; $\mathbf{P}(E \mid A_i), i \in \{1,2,3\}$?

Question 3 : Probabilité de la conséquence

En déduire la probabilité de l'événement ; $\mathbf{P}(E)$.

Question 4 : Probabilité a posteriori des causes

Quelles sont les probabilités a posteriori de chacune des causes aboutissant à la conséquence E (i.e. le présentateur a ouvert la porte qu'il savait ne pas être celle où se trouve le prix de valeur) ; $\mathbf{P}(A_i \mid E), i \in \{1,2,3\}$?

Problème 2. Simulation de mots de passe

3 étudiants de l'UE, Ahlem, Wilfrid et Zoé ont rendu leur TP "génération" de mots de passe aléatoire. Il s'agisait d'ecrire un programme générant un mot composé de 4 signes extraits du dictionnaire $D = \cup_{j=1}^{4}(D_j)$ avec
$D_1 = \{A, \cdots, Z\}$, $D_2 = \{a, \cdots, z\}$, $D_3 = \{0, \cdots, 9\}$, $D_4 = \{"?", "!", ".", ",", ":", ""\}$.
Soient X_i le i^{eme} signe du mot de passe et Y le mot de passe, $Y = (X_1, \cdots, X_4)$.
$X_i \in D$ et $Y \in D^4$.

- A Ahlem a choisi d' utiliser le modèle suivant : les signes X_i sont indépendants et identiquement distribués (i.i.d) et tous les symboles de D sont equiprobables (la distribution est uniforme).
- B Wilfrid utilise le modèle suivant : Il n'est pas possible d'utiliser 2 symboles consécutifs qui soient 2 lettres ou 2 chiffres. Le premier signe du mot est un symbole de D_4.
- C Zoé utilise l'hypothèse suivante : elle génère un mot composé a priori pour moitié de lettres, pour moitié de chiffres ou symboles de D_4

Question 5 : Modèles

Ecrire 3 modèles qui prennent en compte les hypothèses des 3 étudiants

Question 6 : Probabilité d'avoir un caractère dans un mot

Calculez la probabilité d'avoir le signe " " (*espace*) dans un mot de passe généré par chacun de ces 3 modèles ?

Question 7 : Probabilité d'avoir un caractère au début du mot

Calculez la probabilité que le mot de passe commence par un espace ?

Question 8 : Probabilité d'avoir une séquences de caractères

Calculez pour chacun des modèles la probabilité d'obtenir les mots " ! ! ! !", " ?b7.", "aaaa" ?

Question 9 : Le modèle le plus vraisemblable

De quel modèle est il le plus vraisemblable que le mot ".A1." soit issu ?

Question 10 : Le modèle le plus vraisemblable (2)

Si on définit un critère de sécurité comme étant "le mot de passe doit contenir un symbole de chacun des 4 dictionnaires", lequel des 3 modèles générera le plus de mots acceptables ?

Problème 3. Détection de motifs suspects

On analyse une séquence de bits `S` en ligne pour détecter d'éventuels motifs suspects `P`.

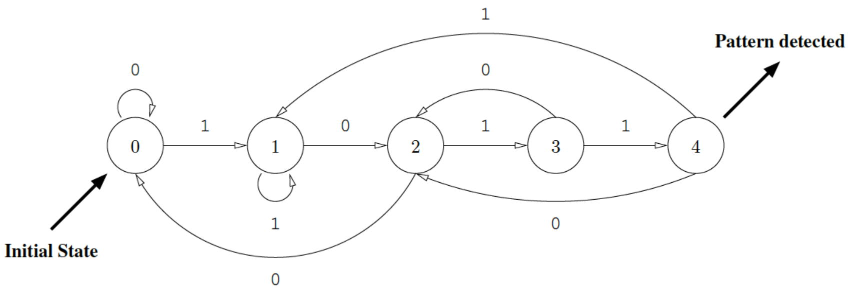
Pour ce motif, un automate des préfixes est construit et une détection à lieu à l'instant n si le motif `P` termine au n-ième bit.

Dans l'exemple suivant, avec le motif `P=[1011]` l'automate émet un signal aux dates 9 et 12.

$Time$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bit	0	0	0	0	1	0	1	0	1	1	0	1	1	0	0	1

Question 11 : Automate

Montrer rapidement que l'automate suivant détecte le motif `[1011]`. (c'est l'automate des préfixes)



Supposons maintenant que la séquence `S` de bits est aléatoire et générée de manière indépendante et uniforme sur $\{0, 1\}$.

Question 12 : Temps d'apparition du premier motif

Pour une séquence uniforme `S` calculer le temps moyen d'apparition du premier motif `[1011]`.

Question 13 : Analyse à long terme

Pour une séquence infinie de bits calculer la fréquence des détections.

Question 14 : Généralisation

Est-ce que le temps d'apparition du premier motif et la fréquence du motif dépendent du motif ? Donner un contre-exemple si ce n'est pas le cas.

Question 15 : Simulation et intervalle de confiance

On cherche maintenant à savoir lequel des motifs $P_1 = [0, 1, 0, 1, 0, 1, 0, 1, 0, 1]$ ou $P_2 = [0, 0, 0, 1, 0, 0, 0, 0, 1, 1]$ apparaît le premier en moyenne. Un calcul montre qu'en moyenne, le motif P_1 apparaît après 1364 coups. Le calcul pour le motif P_2 étant plus compliqué, nous avons programmé un simulateur qui tire des séquences indépendantes et calcul le temps d'apparition du motif P_2 pour chaque séquence. On note X_i le temps d'apparition du motif P_2 dans la $i$ème séquence.

Question 15a – le simulateur effectue n expériences et nous retourne deux valeurs x et y : $x = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $y = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(X_i - x)^2}$. Que représentent les valeurs x et y ?

Question 15b – vers quoi converge la valeur $\frac{1}{n}\sum_{i=1}^{n} X_i$ quand n grandit (justifier) ?

Question 15c – Le simulateur effectue 100 expériences et imprimes les valeurs $x = 1197.22$ et $y = 127.21$. Donner un intervalle de confiance à 95% sur la moyenne des X_i. Peut-on conclure lequel de P_1 ou P_2 apparaît le plus tôt ?

Question 15c – On refait la même expérience avec le motif $P_3 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$. Le simulateur effectue 100 expérience et imprime $x = 1860.13$ et $y = 148.96$. Que peut-on en conclure ?