

Examen de Probabilité.

Lundi 9 décembre 1996

durée 2heures

documents autorisés

Le plus grand compte sera tenu de la qualité de la rédaction et de la précision du raisonnement. En particulier toute référence explicite et justifiée à des formules ou des propriétés vues en cours sera du meilleur effet!

Gestion de table.

On se propose d'étudier la complexité d'une gestion particulière de table.

Le problème est le suivant: un processus lit un texte (un compilateur lit un programme), ce faisant il élabore une table qui devra permettre de stocker certains mots (les identificateurs), permettre de savoir s'il sont déjà apparu dans le texte, etc...

On sait qu'on ne rencontrera pas plus de n mots différents, on prévoit donc pour les stocker un tableau T de taille n .

On dispose d'une fonction de dispersion h (h comme hachage) qui à chaque mot associe une permutation σ de l'ensemble des indices du tableau T $\{1, \dots, n\}$.

Stockage d'un nouvel élément dans le tableau.

A un instant donné le tableau T contient p éléments stockés dans p de ses cases, et donc, $n - p$ cases du tableau sont marquées "vide".

Question 1

Si on regarde une case au hasard, quelle est la probabilité pour qu'elle soit vide?

La stratégie de stockage d'un nouveau mot m dans T contenant strictement moins de n éléments est la suivante:

```
i:=0;
repeat
  i:=i+1;
until T[h(m)(i)] = ''vide'';
T[h(m)(i)] := m;
```

où $h(m)$ est la permutation de $\{1, \dots, n\}$ associée par h à m , et donc $h(m)(i)$ est l'image de i par cette permutation. On rappelle que $i \neq j \rightarrow h(m)(i) \neq h(m)(j)$.

On notera $\alpha = p/n$ le taux de remplissage de T . On rappelle que $\alpha < 1$.

Question 2

Après avoir explicité des hypothèses raisonnables, montrer que le nombre moyen de tours faits dans la boucle de recherche d'une place vide est inférieur ou égal à $1/(1 - \alpha)$.

On pourra utiliser, après l'avoir justifié, le fait que $\frac{p-i}{n-i} \leq \frac{p}{n}$.

Question 3

En déduire que le placement d'un nouvel élément dans T peut être considéré comme fait en temps moyen constant si on limite le taux de remplissage de T .

Remplissage de la table.

On fait l'hypothèse que la fonction h est uniforme, c'est à dire que toutes les permutations σ de l'ensemble $\{1, \dots, n\}$ ont la même probabilité d'être obtenues comme image d'un mot m par h . On suppose que les mots que l'on cherche à placer dans la table T ont été tiré aléatoirement dans l'ensemble des mots possibles. La table T étant vide, on cherche à y inscrire p mots ($p < n$).

Question 4

Montrer que le nombre total de tours dans la boucle de recherche d'une place vide au cours de l'insertion des p mots est en moyenne inférieur à

$$n \cdot \left(\sum_{k=n-p+1}^n 1/k \right)$$

On rappelle que

$$\sum_{k=a}^b 1/k \leq \int_{a-1}^b (1/t) dt.$$

Question 5

Montrer que le nombre moyen par élément de tours de boucle de recherche d'une place vide au cours de l'insertion de p éléments dans une table initialement vide est majoré par

$$\frac{1}{\alpha} \ln\left(\frac{1}{1-\alpha}\right)$$

Où $\ln$ désigne le logarithme népérien.

Application.

Question 6

Dans un tableau de 10000 cases on veut placer par ce mode d'adressage 9000 mots. Combien fait-on en moyenne de tours de boucle par élément à placer.

On rappelle que $\ln(10) = 2,3$.

Prolongement algorithmique.

En cours d'algorithmique on montrera l'intérêt de ce mode d'adressage, et le fait que la recherche d'un élément se fait de la même manière que son insertion.

En particulier le temps moyen de recherche d'un élément présent est en

$$\frac{1}{\alpha} \ln\left(\frac{1}{1-\alpha}\right)$$

si α est le taux de remplissage de la table. Il est donc indépendant de la taille de la table!