Devoir à la maison n°1

On considère l'algorithme suivant qui calcule le maximum des éléments d'un tableau de taille n (donné en cours) :

---

fonction max_iteratif (tableau T, entier n)
{T est un tableau d'objets, l'ensemble des objets est muni d'une relation d'ordre total $\succ$}
max $=-\infty$;
pour i = 1 to n **faire**
  si T[i]$\succ$max **alors**
    `Traiter`(T[i])
    max=T[i]
  fin si
fin pour

L'objectif est d'analyser le coût de cet algorithme, exprimé en nombre d'appels à la fonction `Traiter`, c'est à dire en nombre de fois où la variable "max" est modifiée. La taille des données en entrée est donc la taille n du tableau.

L'algorithme, générique, précise simplement que `T` est un tableau d'objets $e \in \mathcal{E}$ sur lesquels existe une relation d'ordre total $\succ$. L'algorithme fonctionne donc aussi bien, par exemple, pour un tableau d'entiers ($\mathcal{E} = \mathbb{N}$) que pour des chaînes de caractères (ordre lexicographique).

Ce coût est dépendant non seulement de la taille du tableau, mais aussi des **données** du tableau passé en paramètre. En effet :

- le **meilleur coût** possible est 1 opération `Traiter` pour un tableau de taille n si l'on tombe au 1$^{\text{er}}$ coup sur le maximum :
$$\forall i \geq 1, T[i] \leq T[1]$$
- le **pire coût** possible est n opérations `Traiter` pour un tableau de taille n si l'on doit mettre à jour le maximum à chaque étape, i.e. si le tableau est rangé par ordre strictement croissant :
$$\forall i \geq 1, T[i] < T[i+1]$$

La fonction de coût $c : \mathbb{N} \times \mathcal{E}^n \to \mathbb{N}$ est donc fonction de n et de `T`.

La méthode proposée pour analyser ce coût est expérimentale. L'idée de ce devoir est d'implémenter l'algorithme et de concevoir les expériences nécessaires pour en évaluer le coût moyen en fonction de n. Cette moyenne s'entend donc sur l'ensemble des tableaux d'objets possibles de taille n fixée. On définit donc la fonction de coût moyen :

$$\begin{array}{rcl} f : \mathbb{N} & \to & \mathbb{R} \\ n & \mapsto & \mathbb{E}_T[c(n,T)] \end{array}$$

Question 1 *Le coût moyen $f(n)$ dépend-il du* type *des objets contenus dans le tableau ?*

Question 2 *Le coût $c(n,T)$ est-il pénalisé lorsque le tableau T contient plusieurs objets identiques ?*

Question 3 *Peut-on générer la totalité des tableaux de taille n pour calculer le coût moyen ?*

Question 4 *Jusqu'à quelle valeur N_{max} de n est-il raisonnable de calculer ainsi la fonction de coût moyen f ?*

Question 5 *Faire les expériences et tracer la fonction $f(n)$ sur $[1, N_{max}]$.*

On considère maintenant une approche aléatoire.

Question 6 *Proposer un algorithme de génération uniforme d'un tableau de taille n.*

Question 7 *À n fixé, quel nombre $m(n)$ de tableaux faut-il générer pour que l'échantillon soit représentatif ?*

Question 8 *Proposer un plan d'expériences pour évaluer le coût moyen pour différentes valeurs de n (c'est-à-dire donner explicitement les différentes valeurs de n choisies ainsi que, pour chacune d'entre elles, le nombre $m(n)$ de tableaux générés).*

Question 9 *Tracer, à n fixé, le nuage de points obtenus ainsi que l'histogramme des coûts observés. Indiquer l'écart-type.*

Question 10 *Tracer la fonction $f(n)$ pour $1 \leq n \leq 10000$.*

Question 11 *Les observations sont-elles cohérentes avec les résultats de la question 5 ?*

Question 12 *Montrer expérimentalement que la complexité de l'algorithme est en $\mathcal{O}(\log(n))$.*

Les rapports de 4 pages maximum au format PDF exclusivement seront envoyés par mail :
– destinataires : Jean-Marc.Vincent@imag.fr, Florence.Perronnin@imag.fr
– sujet du mail : `[RICM4:PS] DM1 - binôme Nom1-Nom2`
– nom du fichier attaché : `RICM4-PS-2011-DM1-Nom1-Nom2.pdf`
avant le mardi 25 octobre 2011 à 23h59.