TD 11 : Arbres binaires de recherche

Exercice 1. Rappel *: un arbre binaire de recherche (ABR) est un arbre binaire dans lequel chaque nœud n possède un identifiant id et tel que*

- *chaque nœud du sous-arbre gauche de n a un identifiant inférieur à id*
- *chaque nœud du sous-arbre droit de n a un identifiant supérieur à id*

1. *Dans un arbre binaire de recherche à n nœuds et de profondeur p, quelle est la complexité de :*
   - *la recherche d'un identifiant ?*
   - *l'insertion d'un identifiant ?*
   - *la suppression d'un identifiant ?*
2. *Quelle est la profondeur d'un arbre binaire de recherche à n nœuds dans le pire des cas ?*

Exercice 2. *Pour apprécier l'intérêt des arbres, il faut procéder à une analyse probabiliste de la profondeur des nœuds, vue comme une variable aléatoire. L'hypothèse probabiliste que nous faisons est :*

« L'arbre est obtenu par ajout successif de n nœuds, où le nœud à ajouter est à chaque fois choisi de manière équiprobable parmi les nœuds restants. »

Dans la suite nous notons les identifiants des n nœuds par $id_1 < \ldots < id_n$. Soit $1 \leq i < j \leq n$.

- *$X_{i,j}$ désigne la v.a. qui vaut 1 si id_i et id_j ont été comparés et 0 sinon.*
- *$p(i)$ désigne la v.a. correspondant à la profondeur de id_i (on suppose que la profondeur de la racine est 1, la profondeur de ses fils est 2 et ainsi de suite).*
- *$p_m = \frac{1}{n}\sum_{i=1}^{n} p(i)$ désigne la v.a. correspondant à la profondeur moyenne d'un identifiant dans un arbre aléatoire.*
- *$p_{max} = max(\{p(i) \mid 1 \leq i \leq n\})$ désigne la v.a. correspondant à la profondeur d'un arbre aléatoire.*

1. *Soit $1 \leq i < j \leq n$. Montrer que les identifiants id_i et id_j seront comparés si et seulement si l'un des deux est le premier à être inséré parmi l'ensemble des identifiants compris entre id_i et id_j (i.e. parmi $ID(i,j) = \{id_i, id_{i+1}, \ldots, id_{j-1}, id_j\}$).*
2. *Montrer que $p_m = 1 + \frac{1}{n}\sum_{i<j} X_{i,j}$.*
3. *Calculer $\mathbf{E}(X_{i,j})$.*
4. *Montrer que $\mathbf{E}(p_m) = -3 + 2(1+\frac{1}{n})\sum_{k=1}^{n}\frac{1}{k}$.*
5. *Montrer que $2(1+\frac{1}{n})\ln(n+1) - 3 \leq \mathbf{E}(p_m) \leq 2(1+\frac{1}{n})(1+\ln(n)) - 3$.*
6. *Quelle est la complexité d'une recherche fructueuse (en supposant que les identifiants sont recherchés avec la même probabilité) ?*
7. *Quelle est la complexité d'une recherche infructueuse (identique au cas de l'ajout) ? (Indication : supposons que l'on recherche un identifiant placé entre id_i et id_{i+1}) ?*

8. *Dans ce qui suit, on note* p_n *la v.a.* p_{max} *d'un arbre obtenu par ajout aléatoire de* n *identifiants et* $f_n = 2^{p_n}$ *la profondeur exponentielle de l'arbre. Notons* $f_{n,i}$ *la v.a. représentant la profondeur exponentielle sachant que le premier identifiant inséré est* id_i*. Montrer que*

$$f_{n,i} = 2^{1+max(p_{i-1},p_{n-i})}$$

et que

$$f_{n,i} \leq 2(f_{i-1} + f_{n-i}) \ .$$

9. *Montrer que* $\mathbf{E}(f_n) \leq \frac{4}{n} \sum_{i=0}^{n-1} \mathbf{E}(f_i)$.
10. *Conclure que* $\forall n, \mathbf{E}(f_n) \leq n^3 + 1$
11. *Conclure que* $\forall n > 1, \mathbf{E}(p_n) \leq 3log(n) + 1$