Note sur la démonstration de la profondeur moyenne d'un ABR

Je reviens sur la preuve de la profondeur moyenne d'un ABR dans le sens vu en cours, *i.e.*, d'un arbre moyen obtenu en insérant un à un les éléments d'une permutation de n entiers de 1 à n. L'hypothèse que l'on fait sur les données est que les $n!$ permutations des n clés introduites sont équiprobables. En particulier la racine de l'ABR (qui est toujours la première clé insérée) peut être de façon équiprobable le 1, le 2, ou le nombre n (on suppose ici que les clefs sont les entiers de 1 à n).

On remarque que si la racine est l'entier i, alors le sous-arbre gauche contient les clefs/entiers de 1 à $i-1$ et contient donc $i-1$ sommets. De façon similaire, le sous-arbre droit contient les clefs de $i+1$ à n et contient donc $n-i$ sommets.

On va ici regarder la somme de la profondeur $P(\mathcal{A})$ de tous les sommets d'un l'arbre binaire de recherche $\mathcal{A}$ (et non pas la moyenne E_n mais en fait c'est la même chose et la preuve est à l'identique car prendre deux arbres et les coller à une racine ne fait que rajouter exactement 1 à chaque noeud, donc 1 à la moyenne). On va considérer que la profondeur de la racine est 0.

Pour tout arbre binaire de recherche $\mathcal{A}$ de n sommets, si $\mathcal{A} = \langle \text{left}, r, \text{right} \rangle$ alors $P(\mathcal{A}) = P(\text{left}) + P(\text{right}) + (n-1)$.

$$\begin{aligned}
P(\mathcal{A}) &= \sum_{u \in \text{left}} d(x,r) + \sum_{u \in \text{right}} d(x,r) + d(r,r) \\
&= \sum_{u \in \text{left}} (d(x, \text{root}(\text{left})) + 1) + \sum_{u \in \text{right}} (d(x, \text{root}((\text{right})) + 1) + 0 \\
&= P(\text{left}) + |\text{left}| + P(\text{right}) + |\text{right}| \\
&= P(\text{left}) + P(\text{right}) + (n-1)
\end{aligned}$$

Si maintenant on s'intéresse à la moyenne $\tilde{P}_n$ de la la somme de la profondeur $P(\mathcal{A})$ de tous les sommets d'un l'arbre binaire de recherche $\mathcal{A}$ pour tous les arbre $\mathcal{A}$ ayant n sommets alors l'équation de récurrence s'applique toujours et on peut employer l'hypothèse d'équipartition de la racine pour un arbre $\mathcal{A} = \langle \text{left}, r, \text{right} \rangle$ puisque r peut prendre de façon équiprobable toutes les valeurs de 1 à n. La moyenne $\tilde{P}_n(\mathcal{A})$ est donc égale à :

$$\begin{aligned} \tilde{P}_n &= \frac{1}{n} \sum_{i=1}^{n} \left(\tilde{P}_{i-1} + \tilde{P}_{n-i} + n - 1 \right) \\ &= n - 1 + \frac{2}{n} \sum_{i=0}^{n-1} \tilde{P}_i \end{aligned}$$

En calculant $n\tilde{P}_n - (n-1)\tilde{P}_{n-1}$ comme en cours on trouve :

$$\begin{aligned} \frac{\tilde{P}_n}{n+1} &= 2 \sum_{i=1}^{n} \frac{1}{i} + \frac{2}{n+1} - 4 \\ &\approx 2 \log(n) \end{aligned}$$

donc $\tilde{P}_n = O(n \log n)$ et donc la profondeur moyenne d'un sommet est en $O(\log n)$