

## Probabilités, Statistiques, Combinatoire : TD3

Combinatoire : codage des arbres binaires

<http://www.labri.fr/perso/duchon/Enseignements/Probas/>

Tous les arbres mentionnés dans les exercices de cette feuille sont des arbres binaires complets, même si l'énoncé ne le précise pas. Tous les mots sont sur l'alphabet  $\{a, b\}$ , même si l'énoncé ne le précise pas.

### 3.1 Définitions du cours et révisions

Les exercices de cette section ont pour but de faire pratiquer les définitions et leur application. Ils ne sont censés présenter aucune difficulté si les définitions sont connues.

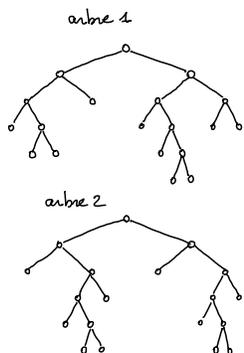
#### Exercice 3.1

On définit, pour un arbre binaire complet  $t$ , son *mot de codage*, sur l'alphabet  $\{a, b\}$ , de la manière suivante :

- l'arbre réduit à une feuille, est codé par le mot vide  $\varepsilon$  ;
- un arbre  $t$  non réduit à une feuille, dont les sous-arbres gauche et droit sont respectivement  $t_1$  et  $t_2$ , est codé par le mot  $a.w_1.b.w_2$ , où  $w_1$  est le mot de codage de  $t_1$ , et  $w_2$ , le mot de codage de  $t_2$ .

(Si vous avez vu en cours d'algorithmique un codage par des mots binaires pour les arbres binaires complets, c'est *a priori* le même, avec peut-être une différence d'une lettre à la fin ; si vous avez vu un codage pour les arbres binaires quelconques, c'est le même codage si on initialise différemment la définition récursive en disant que c'est l'arbre vide qui est codé par le mot vide)

Calculer les mots codant les arbres ci-dessous.



### Exercice 3.2

Parmi les arbres binaires complets, on distingue certains arbres particuliers :

- un arbre est un *peigne droit* si chacun de ses nœuds internes a pour fils gauche une feuille ;
  - un arbre est un *peigne gauche* si chacun de ses nœuds internes a pour fils droit une feuille ;
  - un arbre est *parfait* si toutes ses feuilles sont à la même hauteur.
1. Dessiner les peignes gauches et peignes droits à 4, 5 et 6 nœuds internes, et calculer leurs mots de codage. Proposer une formule pour le mot de codage d'un peigne droit et d'un peigne gauche, en général (c'est-à-dire une formule donnant, en fonction de  $n$ , le mot de codage de l'arbre peigne gauche à  $n$  nœuds internes, et une autre pour le mot de codage de l'arbre peigne droit à  $n$  nœuds internes).
  2. Dessiner les arbres parfaits de hauteurs 2 et 3, et calculer leurs mots de codage.

## 3.2 Compléments du cours

Les exercices de cette section sont l'occasion de refaire une partie des calculs vus en cours à l'identique, ou d'en faire d'autres selon le même principe.

### Exercice 3.3

1. Proposer une formule censée donner, pour tout arbre binaire complet, la longueur de son mot de codage, en fonction de sa taille.
2. Prouver, par récurrence (forte) sur la taille des arbres, que votre formule est valide.

### Exercice 3.4

Un mot binaire  $w$  sur l'alphabet  $\{a, b\}$  est dit *équilibré* s'il contient autant d'occurrences de  $a$  que de  $b$  :  $|w|_a = |w|_b$ .

1. En identifiant une correspondance (bijective) entre les mots équilibrés de longueur  $2n$ , et certaines parties de l'ensemble  $[[1, 2n]]$ , donner une formule pour le nombre de mots équilibrés de longueur  $2n$ .
2. Soient  $w$  et  $w'$  deux mots quelconques ; on définit alors le mot  $w'' = a.w.b.w'$ .  
Exprimer  $|w''|_a$  et  $|w''|_b$  en fonction de  $|w|_a$ ,  $|w|_b$ ,  $|w'|_a$  et  $|w'|_b$ . En déduire la propriété suivante : *Si  $w$  et  $w'$  sont équilibrés, alors  $w''$  est équilibré.*
3. Montrer, par récurrence forte sur la taille des arbres, que le codage de tout arbre binaire complet est un mot équilibré.

Un mot  $u$  est appelé *préfixe* d'un mot  $w$ , s'il existe un mot  $v$  tel que l'on ait  $w = u.v$  (le mot vide  $\varepsilon$ , et le mot  $w$  lui-même, font partie des préfixes de  $w$  ; ils correspondent respectivement aux cas  $v = w$  et  $v = \varepsilon$ ).

Un mot de longueur  $n$  a donc, pour chaque entier  $k$  compris entre 0 et  $n$ , exactement un préfixe de longueur  $k$ .

### Exercice 3.5

On convient de dire qu'un mot  $w$  sur l'alphabet  $\{a, b\}$  est *positif* si *chacun de ses préfixes contient au moins autant d'occurrences de  $a$  que de  $b$*  : pour tous mots  $u$  et  $v$  tels que  $w = u.v$ , on a  $|u|_a \geq |u|_b$ .

Soient  $w$  et  $w'$  deux mots quelconques ; on pose de nouveau  $w'' = a.w.b.w'$ .

1. Soient  $u$  et  $v$  deux mots quelconques. Décrire, en fonction des préfixes de  $u$  et de  $v$ , **tous** les préfixes du mot  $u.v$ .
2. Décrire, en fonction des préfixes de  $w$  et de  $w'$ , tous les préfixes de  $w''$ .
3. Montrer la propriété suivante : *si  $u$  et  $v$  sont deux mots positifs, alors le mot  $u.v$  est un mot positif.*
4. Montrer, par récurrence (forte) sur la taille des arbres, la propriété suivante : *le mot de codage de tout arbre binaire complet est un mot positif.*

On définit un *mot de Dyck* comme un mot qui est à la fois équilibré et positif. Les deux exercices précédents, pris ensemble, montrent que le mot de codage d'un arbre binaire est toujours un mot de Dyck.

### 3.3 Exercices et problèmes

Les exercices de cette section sont prévus pour demander plus de réflexion et d'effort.

#### Exercice 3.6

**Décodage d'un mot de Dyck en arbre binaire.** L'objectif de cet exercice est de décrire, pour tout mot de Dyck  $w$ , la construction d'un arbre binaire complet dont  $w$  soit le mot de codage, **et** de montrer que cet arbre est unique. On a donc une *bijection* entre les arbres binaires complets à  $n$  nœuds internes et les mots de Dyck de longueur  $2n$ .

1. Soit  $w$  un mot de Dyck non vide. Justifier l'existence d'un *plus court préfixe non vide de  $w$  qui soit un mot de Dyck*. (**Indication** : le mot  $w$  est un de ses propres préfixes)
2. Soit  $w'$  le plus court préfixe non vide de  $w$  qui soit un mot de Dyck. Quelles sont les première et dernière lettres de  $w'$  ? Justifier qu'il existe un mot  $w''$  tel que l'on ait  $w' = a.w''.b$ , et que  $w''$  est un mot de Dyck.
3. Expliquer rapidement pourquoi la propriété suivante est vraie (on n'attend pas une preuve complète) : *pour tout mot de Dyck non vide  $w$ , il existe une et une seule paire de mots de Dyck  $(w', w'')$  tels que l'on ait  $w = a.w'.b.w''$ .*
4. Soit  $w$  un mot de Dyck décomposé comme à la question précédente. Montrer que, si  $w'$  est le codage d'un arbre  $t_1$  et que  $w''$  est le codage d'un arbre  $t_2$ , alors  $w$  est le codage d'un arbre  $t$  (décrire  $t$ ).
5. Terminer la preuve (**indication** : on peut procéder par l'absurde, en supposant la propriété fautive et en prenant un mot de Dyck de longueur minimale qui ne soit pas le codage d'un arbre).