

Probabilités, Statistiques, Combinatoire

Philippe Duchon

Université de Bordeaux – Licence Informatique

2018-2019

Obtention d'une formule pour les coefficients binomiaux

On a défini $\binom{n}{k}$ comme étant le nombre de parties à k éléments de l'ensemble $[[1, n]]$; peut-on trouver une formule ?

Lemme : comptage de séquences d'éléments distincts

Pour tous entiers n et k , avec $k \leq n$, le nombre de séquences de longueur k d'éléments *distincts* de $[[1, n]]$, est

$$\prod_{i=0}^{k-1} (n - i) = n.(n - 1) \dots (n - k + 1) = \frac{n!}{(n - k)!}$$

Preuve : par récurrence sur k (à n fixé) : on montre que, pour peu que l'on ait $n \geq k$, il y a $(n - k + 1)$ fois plus de séquences de k éléments distincts de $[[1, n]]$, que de séquences de $k - 1$ éléments distincts de $[[1, n]]$. On initialise la récurrence pour $k = 1$: il y a n séquences de longueur 1 (ou pour $k = 0$: 1 séquence de longueur 0).

Coefficients binomiaux (2)

Pour arriver à une formule pour les coefficients binomiaux, on répond à la question suivante : si on prend la fonction $f_{n,k}$ de l'ensemble des séquences de longueur k sur $[[1, n]]$, vers l'ensemble des *parties* de cardinal k de $[[1, n]]$, qui envoie chaque séquence sur l'ensemble de ses éléments (on “oublie” simplement l'ordre), **combien chaque partie de cardinal k de $[[1, n]]$ a-t-elle d'antécédents par $f_{n,k}$?**

Coefficients binomiaux (2)

Pour arriver à une formule pour les coefficients binomiaux, on répond à la question suivante : si on prend la fonction $f_{n,k}$ de l'ensemble des séquences de longueur k sur $[[1, n]]$, vers l'ensemble des *parties* de cardinal k de $[[1, n]]$, qui envoie chaque séquence sur l'ensemble de ses éléments (on “oublie” simplement l'ordre), **combien chaque partie de cardinal k de $[[1, n]]$ a-t-elle d'antécédents par $f_{n,k}$?** **réponse : $k!$, mais pourquoi ?**

Coefficients binomiaux (2)

Pour arriver à une formule pour les coefficients binomiaux, on répond à la question suivante : si on prend la fonction $f_{n,k}$ de l'ensemble des séquences de longueur k sur $[[1, n]]$, vers l'ensemble des *parties* de cardinal k de $[[1, n]]$, qui envoie chaque séquence sur l'ensemble de ses éléments (on “oublie” simplement l'ordre), **combien chaque partie de cardinal k de $[[1, n]]$ a-t-elle d'antécédents par $f_{n,k}$?** réponse : $k!$, mais pourquoi ? Résultat, on a une formule pour $\binom{n}{k}$:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!},$$

qui a l'avantage de faire apparaître (clairement ?) le fait que $\binom{n}{k} = \binom{n}{n-k}$

Un autre exemple de comptage : mots sans aa

On considère la classe combinatoire suivante : A est l'ensemble des mots sur l'alphabet $\{a, b\}$ qui ne contiennent pas deux a consécutifs (taille : la longueur du mot).

- ▶ C'est bien une classe combinatoire. (pourquoi?)

Un autre exemple de comptage : mots sans aa

On considère la classe combinatoire suivante : A est l'ensemble des mots sur l'alphabet $\{a, b\}$ qui ne contiennent pas deux a consécutifs (taille : la longueur du mot).

- ▶ C'est bien une classe combinatoire. (pourquoi?)
- ▶ On écrit les premières tranches, pour voir à quoi ressemble le début de la suite de comptage :
 - ▶ $A_0 = \{\varepsilon\}$, $a_0 = 1$
 - ▶ $A_1 = \{a, b\}$, $a_1 = 2$
 - ▶ $A_2 = \{ab, ba, bb\}$, $a_2 = 3$
 - ▶ $A_3 = \{aba, abb, bab, bba, bbb\}$, $a_3 = 5$
 - ▶ (ça ne donne pas de formule. . .)

Un autre exemple de comptage : mots sans aa

On considère la classe combinatoire suivante : A est l'ensemble des mots sur l'alphabet $\{a, b\}$ qui ne contiennent pas deux a consécutifs (taille : la longueur du mot).

- ▶ C'est bien une classe combinatoire. (pourquoi?)
- ▶ On écrit les premières tranches, pour voir à quoi ressemble le début de la suite de comptage :
 - ▶ $A_0 = \{\varepsilon\}$, $a_0 = 1$
 - ▶ $A_1 = \{a, b\}$, $a_1 = 2$
 - ▶ $A_2 = \{ab, ba, bb\}$, $a_2 = 3$
 - ▶ $A_3 = \{aba, abb, bab, bba, bbb\}$, $a_3 = 5$
 - ▶ (ça ne donne pas de formule. . .)
- ▶ À défaut de formule, peut-on trouver une relation de récurrence satisfaite par la suite de comptage? (**indication** : se poser la double question suivante : combien A_n contient-il de mots qui commencent par la lettre a ? par la lettre b ?)

Un autre exemple de comptage : mots sans aa

On considère la classe combinatoire suivante : A est l'ensemble des mots sur l'alphabet $\{a, b\}$ qui ne contiennent pas deux a consécutifs (taille : la longueur du mot).

- ▶ C'est bien une classe combinatoire. (pourquoi?)
- ▶ On écrit les premières tranches, pour voir à quoi ressemble le début de la suite de comptage :
 - ▶ $A_0 = \{\varepsilon\}$, $a_0 = 1$
 - ▶ $A_1 = \{a, b\}$, $a_1 = 2$
 - ▶ $A_2 = \{ab, ba, bb\}$, $a_2 = 3$
 - ▶ $A_3 = \{aba, abb, bab, bba, bbb\}$, $a_3 = 5$
 - ▶ (ça ne donne pas de formule...)
- ▶ À défaut de formule, peut-on trouver une relation de récurrence satisfaite par la suite de comptage? (**indication** : se poser la double question suivante : combien A_n contient-il de mots qui commencent par la lettre a ? par la lettre b ?)
- ▶ (Bon, c'est encore la suite de Fibonacci en fait)

Encore un exemple : arbres binaires complets

L'ensemble $\mathcal{B}$ des arbres binaires complets non vides, peut être décrit **inductivement** de la manière suivante :

- ▶ un arbre t peut être composé d'un unique nœud racine ; ou
- ▶ un arbre t peut être composé d'un nœud racine r , et d'un couple de deux arbres (t_1, t_2) ; t_1 est alors le sous-arbre gauche, et t_2 le sous-arbre droit.

(Ces arbres binaires complets sont les “squelettes” des arbres binaires dits complets définis en Algorithmique des Structures arborescentes : les nœuds de nos arbres ne portent pas de valeurs, on s'intéresse seulement à leurs “formes”)

On pourrait prendre comme notion de taille diverses quantités :

- ▶ le nombre total de nœuds (forcément, strictement positif)
- ▶ le nombre de nœuds internes (peut être nul)
- ▶ le nombre de feuilles
- ▶ la hauteur

Notion usuelle de taille : le nombre de nœuds internes.

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ?

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?
 $n - 1 - k$

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?
 $n - 1 - k$
 - ▶ Si on fixe k , combien de choix possibles pour l'arbre t_1 ? Pour l'arbre t_2 ? Pour le couple (t_1, t_2) ?

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?
 $n - 1 - k$
 - ▶ Si on fixe k , combien de choix possibles pour l'arbre t_1 ? Pour l'arbre t_2 ? Pour le couple (t_1, t_2) ? respectivement, c_k , c_{n-1-k} ,
 $c_k \cdot c_{n-1-k}$

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?
 $n - 1 - k$
 - ▶ Si on fixe k , combien de choix possibles pour l'arbre t_1 ? Pour l'arbre t_2 ? Pour le couple (t_1, t_2) ? respectivement, c_k , c_{n-1-k} ,
 $c_k \cdot c_{n-1-k}$
 - ▶ On en déduit une formule pour c_n , en fonction de tout ce qui le précède dans la séquence, $(c_0, c_1, \dots, c_{n-1}) \dots$

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?
 $n - 1 - k$
 - ▶ Si on fixe k , combien de choix possibles pour l'arbre t_1 ? Pour l'arbre t_2 ? Pour le couple (t_1, t_2) ? respectivement, c_k , c_{n-1-k} ,
 $c_k \cdot c_{n-1-k}$
 - ▶ On en déduit une formule pour c_n , en fonction de tout ce qui le précède dans la séquence, $(c_0, c_1, \dots, c_{n-1}) \dots$
 - ▶ La séquence est définie par : $c_0 = 1$, et pour $n \geq 1$,

$$c_n = \sum_{k=0}^{n-1} c_k c_{n-1-k}$$

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?
 $n - 1 - k$
 - ▶ Si on fixe k , combien de choix possibles pour l'arbre t_1 ? Pour l'arbre t_2 ? Pour le couple (t_1, t_2) ? respectivement, c_k , c_{n-1-k} ,
 $c_k \cdot c_{n-1-k}$
 - ▶ On en déduit une formule pour c_n , en fonction de tout ce qui le précède dans la séquence, $(c_0, c_1, \dots, c_{n-1}) \dots$
 - ▶ La séquence est définie par : $c_0 = 1$, et pour $n \geq 1$,

$$c_n = \sum_{k=0}^{n-1} c_k c_{n-1-k}$$

- ▶ Ça ne donne pas une formule, mais ça permet de calculer efficacement les premiers termes (**suite de Catalan**) :
1, 2, 5, 14, 42, 132, 429, 1430 ...

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque.

- ▶ Si $n = 0$, il y a un seul arbre sans nœuds internes : $c_0 = 1$.
- ▶ Si $n > 0$, l'arbre a un sous-arbre t_1 ; soit k sa taille (son nombre de nœuds internes).
 - ▶ Quelles sont les valeurs possibles de k ? les entiers de 0 à $n - 1$
 - ▶ Si on fixe k , quelle est alors la taille du sous-arbre gauche ?
 $n - 1 - k$
 - ▶ Si on fixe k , combien de choix possibles pour l'arbre t_1 ? Pour l'arbre t_2 ? Pour le couple (t_1, t_2) ? respectivement, c_k , c_{n-1-k} ,
 $c_k \cdot c_{n-1-k}$
 - ▶ On en déduit une formule pour c_n , en fonction de tout ce qui le précède dans la séquence, $(c_0, c_1, \dots, c_{n-1}) \dots$
 - ▶ La séquence est définie par : $c_0 = 1$, et pour $n \geq 1$,

$$c_n = \sum_{k=0}^{n-1} c_k c_{n-1-k}$$

- ▶ Ça ne donne pas une formule, mais ça permet de calculer efficacement les premiers termes (**suite de Catalan**) :
1, 2, 5, 14, 42, 132, 429, 1430 ...

Codage des arbres binaires complets

On peut coder bijectivement les arbres binaire complets de taille n , par des mots de longueur $2n$ sur un alphabet à 2 lettres. On peut décrire une telle bijection récursivement, de la manière suivante (noter qu'on la définit ainsi pour toutes les tailles) :

- ▶ l'arbre réduit à une feuille ($n = 0$) est codé par le mot vide ϵ
- ▶ un arbre non réduit à une feuille, $t = (\bullet, t_1, t_2)$ est codé par le mot

$$w = a \cdot w_1 \cdot b \cdot w_2,$$

où w_1 est le codage de t_1 et w_2 est le codage de t_2 .

Codage des arbres binaires complets

On peut coder bijectivement les arbres binaire complets de taille n , par des mots de longueur $2n$ sur un alphabet à 2 lettres. On peut décrire une telle bijection récursivement, de la manière suivante (noter qu'on la définit ainsi pour toutes les tailles) :

- ▶ l'arbre réduit à une feuille ($n = 0$) est codé par le mot vide ϵ
- ▶ un arbre non réduit à une feuille, $t = (\bullet, t_1, t_2)$ est codé par le mot

$$w = a \cdot w_1 \cdot b \cdot w_2,$$

où w_1 est le codage de t_1 et w_2 est le codage de t_2 .

- ▶ **Peut-on décrire l'ensemble des mots qui codent des arbres ?**

Codage des arbres binaires complets

On peut coder bijectivement les arbres binaire complets de taille n , par des mots de longueur $2n$ sur un alphabet à 2 lettres. On peut décrire une telle bijection récursivement, de la manière suivante (noter qu'on la définit ainsi pour toutes les tailles) :

- ▶ l'arbre réduit à une feuille ($n = 0$) est codé par le mot vide ϵ
- ▶ un arbre non réduit à une feuille, $t = (\bullet, t_1, t_2)$ est codé par le mot

$$w = a \cdot w_1 \cdot b \cdot w_2,$$

où w_1 est le codage de t_1 et w_2 est le codage de t_2 .

- ▶ **Peut-on décrire l'ensemble des mots qui codent des arbres ?**
- ▶ Tout mot de codage a autant d'occurrences de a que de b :
 $|w|_a = |w|_b$ (clairement, si c'est vrai pour w_1 et w_2 , c'est vrai aussi pour $w = a \cdot w_1 \cdot b \cdot w_2$, on finit par une preuve par récurrence sur le nombre de nœuds de l'arbre)

Codage des arbres binaires complets

On peut coder bijectivement les arbres binaire complets de taille n , par des mots de longueur $2n$ sur un alphabet à 2 lettres. On peut décrire une telle bijection récursivement, de la manière suivante (noter qu'on la définit ainsi pour toutes les tailles) :

- ▶ l'arbre réduit à une feuille ($n = 0$) est codé par le mot vide ϵ
- ▶ un arbre non réduit à une feuille, $t = (\bullet, t_1, t_2)$ est codé par le mot

$$w = a \cdot w_1 \cdot b \cdot w_2,$$

où w_1 est le codage de t_1 et w_2 est le codage de t_2 .

- ▶ **Peut-on décrire l'ensemble des mots qui codent des arbres ?**
- ▶ Tout mot de codage a autant d'occurrences de a que de b :
 $|w|_a = |w|_b$ (clairement, si c'est vrai pour w_1 et w_2 , c'est vrai aussi pour $w = a \cdot w_1 \cdot b \cdot w_2$, on finit par une preuve par récurrence sur le nombre de nœuds de l'arbre)
- ▶ **Propriété supplémentaire** : dans tout mot w de codage, si on "coupe" w en deux mots u et v (i.e. $w = u \cdot v$), **dans le mot u , il y a au moins autant de a que de b** : $|u|_a \geq |u|_b$.

Mots de Dyck

Définition - Théorème

Un mot w sur l'alphabet $A = \{a, b\}$ est un mot de Dyck s'il satisfait aux deux conditions suivantes :

- ▶ $|w|_a = |w|_b$
- ▶ Pour tous mots u et v tels que $w = u.v$, on a $|u|_a \geq |u|_b$.

Le codage précédent établit une **bijection** entre les arbres binaires à n nœuds, et les mots de Dyck de longueur $2n$.

Preuve : pour la preuve, on va travailler sur une représentation des mots comme des *chemins*, dans lesquels la lettre a code un pas Nord-Est $(1, 1)$, et la lettre b code un pas Sud-Est ; les chemins commencent en $(0, 0)$. (“Chemins de Dyck”)

Chemins de Dyck

On dessine un mot sur l'alphabet $\{a, b\}$ comme décrivant un "chemin" partant de l'origine $(0, 0)$:

- ▶ pour chaque lettre a , on fait un pas vers le Nord-Est (de (x, y) à $(x + 1, y + 1)$)
- ▶ pour chaque lettre b , on fait un pas vers le Sud-Est (de (x, y) à $(x + 1, y - 1)$)

Dire que le mot w est un mot de Dyck, c'est équivalent à deux conditions sur le chemin :

- ▶ il fait autant de pas Nord-Est que Sud-Est : il se termine donc sur l'axe horizontal, en un point $(2n, 0)$;
- ▶ n'importe quel point du chemin (correspondant à ne lire que le début du mot) peut se trouver au-dessus de l'axe horizontal, ou sur l'axe, mais pas dessous (car parmi les k premiers pas, il y a au moins autant de Nord-Est que de Sud-Est)