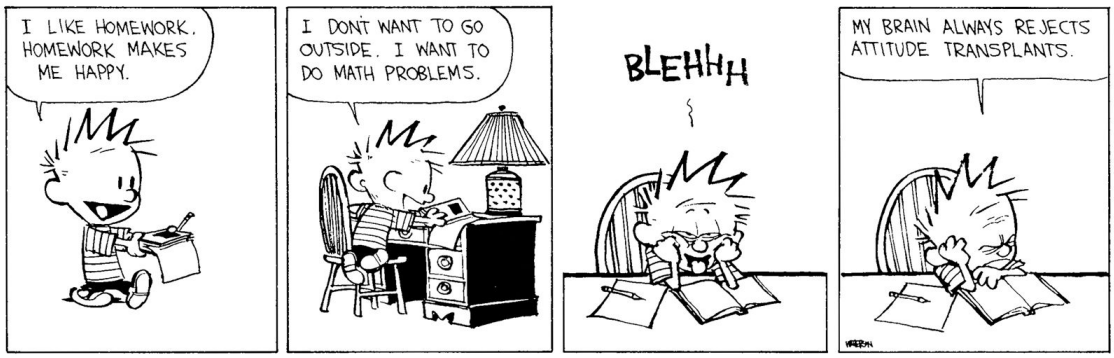
L3 Mention Informatique / Mathématiques pour l'informatique *One-shot*

Pierre Karpman
Université Grenoble-Alpes

2023-09-07



1 Notations *Oo*

Les notations O, Ω, Θ, o, ω, $\tilde{O}$ sont, à des degrés divers, couramment utilisées pour exprimer le coût asymptotique d'un algorithme. On ne définit ici ces notations que pour des fonctions réelles et où le régime asymptotique caractérisé par la notation se trouve en l'infini (positif) ; les cas de régimes asymptotiques différents (par exemple en zéro) se traitent de façon similaire. Dans le même esprit de simplification, les définitions formelles sont uniquement données pour le cas de fonctions en une variable.

§ Exemple informel & notations

Soit f, g deux fonctions, la notation $f = O(g)$ (parfois aussi écrite $f \in O(g)$), qui se lit « f est un grand Oh de g » exprime informellement le fait qu'à une constante près, f est asymptotiquement « inférieure ou égale » à g. On utilise souvent la notation alternative qui remplace « g » par son expression. Par exemple, pour $g : n \mapsto n^2$, on peut écrire $f = O(n^2)$ (cela sous-entend que f est une fonction en au moins une variable n). De même, on écrit $f = O(1)$ pour $f = O(g)$ où g est une fonction en les même variables que f et à valeur constante 1. Dans le cas où f et g sont des fonctions de plusieurs variables, on note simplement (par exemple) $f = O(n^2 + m \log(t))$, ce qui exprime quelque chose sur le comportement asymptotique de f quand les trois variables n, m et t tendent vers l'infini (ou possiblement d'autres valeurs en général, *cf. supra*).

1.1 Notation O

On peut informellement interpréter $f = O(g)$ comme voulant dire que « $f \leq g$ », où cette inégalité doit être comprise comme asymptotique et *à une constante près*.

Formellement, on a :

Définition 1.1 Soit f, g deux fonctions (que l'on rappelle de $\mathbb{R}$ dans lui-même) en une variable n :

$$f = O(g) \Rightarrow \exists C, n_0 \in \mathbb{R}^+ : \forall n \geq n_0, f(n) \leq Cg(n)$$

Dans le cas où f et g admettent toutes deux une limite en ∞, on peut de façon équivalente utiliser la définition :

$$f = O(g) \Rightarrow \exists c \in \mathbb{R}^+ : \lim_{n\to\infty} g(n)/f(n) \geq c$$

R Il est important de ne pas interpréter cette notation comme une *égalité* (asymptotique, à une constante près).

Exemple 1.2

1. $f : n \mapsto 3n$. On a immédiatement $f = O(n)$ puisque par exemple (suivant les notations ci-dessus) $C = 3$ et $n_0 = 0$ garantissent pour tout $n \geq n_0$ que $f(n) = 3n \leq Cn = 3n$. On a tout aussi immédiatement $f = O(n^2)$ puisqu'on a par exemple que pour tout $n \geq 12$, $f(n) = 3n \leq 257n^2$.
2. $f : n \mapsto 7n^2 + 3n$. Montrons que $f = O(n^2)$. En prenant $C = 8$, l'expression $Cn^2 - f(n)$ vaut $n^2 - 3n$, et la fonction associée est croissante pour $n \geq 3/2$, et positive pour $n \geq 3$. Il s'ensuit que $C = 8$ et $n_0 = 3$ garantissent pour tout $n \geq n_0$ que $f(n) \leq Cn^2$.

   R On aurait pu montrer de la même façon que $f = O(8n^2)$, et il aurait alors été suffisant de prendre $C = 1$. Cependant, un des intérêts de la notation O est justement de permettre une écriture concise qui ne nécessite pas d'expliciter les constantes C et n_0. Il est donc d'usage d'utiliser une expression la plus « simple » possible à l'intérieur du grand Oh. Ceci se justifie d'autant plus que $f = O(8n^2)$ n'indique en aucune façon que la constante C de la notation peut être prise à 1, puisqu'on aurait par exemple tout aussi bien pu écrire $f = O(2n^2)$. La notation O n'est donc pas adaptée pour exprimer les constantes apparaissant dans l'expression de f.
3. $f : n \mapsto n-1$. On a évidemment $f = O(n-1)$, mais tout autant $f = O(n-2)$ ou $f = O(n)$; cette dernière expression est préférable car plus concise.
4. $f : n \mapsto n^2$. Montrons que $f \neq O(n)$, c'est à dire qu'il n'existe pas de C, n_0 tel que pour tout $n \geq n_0$ on a $n^2 \leq Cn$. Pour cela, on montre que *pour tout* couple C, n_0 il existe un $n_1 \geq n_0$ tel que $(n_1)^2 > Cn_1$. Ceci se fait aisément en raisonnant par cas : si $n_0 < 1$, il suffit de prendre $n_1 = C + 1$ et de constater que pour tout $C \in \mathbb{R}^+$, $(C+1)^2 > C + 1$; sinon, il suffit de prendre $n_1 = n_0(C+1)$, qui vérifie dans ce cas $n_0^2(C+1)^2 > n_0(C+1)$.

   R Comme on le verra plus tard, on a ici en fait montré (à peu de choses près) que la fonction $n \mapsto n$ est un $o(n^2)$.

R En combinant cet exemple avec le premier, on remarque que $n \mapsto n$ est un $O(n^2)$, mais que $n \mapsto n^2$ n'est pas un $O(n)$. La notation O n'est donc pas symétrique, ce qui est attendu pour une notation traduisant une «inégalité».

5. Montrons que pour toute constante entière $k > 1$, $f : n \mapsto n^k$, $g : n \mapsto e^n$, on a $f = O(g)$[a]. On commence par constater que pour $n_0 = k^4 \ln(k)$, on a $g(n_0) = e^{k^4 \ln(k)} = k^{k^4}$ et $f(n_0) = (k^4 \ln(k))^k = k^{4k} \ln(k)^k$. Puisque f est croissante sur $\mathbb{R}^+$ et que pour tout $a > 1$, $x \geq 1$ on a $\log_a(x) < x$, on déduit $f(n_0) \leq k^{5k}$. De plus, comme la fonction $n \mapsto k^n$ est croissante et que $5k < k^4$ pour $k \geq 2$, on obtient $f(n_0) \leq g(n_0)$.

   Il suffirait maintenant pour conclure de montrer que $n \mapsto g(n)/f(n)$ est une fonction croissante pour $n \geq n_0$, ce qui peut par exemple se faire en étudiant sa dérivée. On montre à la place de façon directe que $g(n) > f(n)$ pour tout $n > n_0$. On pose pour cela $\alpha := n/n_0 > 1$, et on considère $g(n) = k^{\alpha k^4}$ et $f(n) = \alpha^k f(n_0) \leq \alpha^k k^{5k}$. On va montrer que la différence des logarithmes en base k de ces deux expressions est supérieure à 1, ce qui assurera que $g(n)/f(n) > 1$. Ces logarithmes sont respectivement αk^4 et $k \log_k(\alpha) + 5k$; en utilisant le fait que $\log_k(\alpha) < \alpha$, leur différence est minorée par $\alpha k^4 - k\alpha - 5k = \alpha(k^4 - k) - 5k$. Notant $\delta = \alpha - 1 > 0$, ceci se réécrit $\delta(k^4 - k) + k^4 - 6k$, qui pour $k \geq 2$ est supérieur à $\delta + 1$. On a donc $g(n)/f(n) \geq k^{\delta+1} > 1$, et le rapport (exponentiellement croissant) entre $g(n)$ et $f(n)$ peut donc être rendu arbitrairement grand.

   R Comme on le verra plus tard, La dernière remarque ci-dessus fait qu'on a en fait montré que $f = o(e^n)$.

a. Le cas particulier $k = 1$ est laissé en exercice.

Exercice 1.1. Montrez que pour tout entier $d \geq 1$, toute fonction polynomiale (en une variable n) de degré d est un $O(n^k)$ pour tout $k \geq d$. Déduisez-en que toute fonction polynomiale est un $O(e^n)$, c'est à dire que toute fonction polynomiale est asymptotiquement inférieure à la fonction exponentielle naturelle. Sous quelle condition ce dernier résultat reste-t'il valide si la base de l'exponentielle est changée ?

1.2 Notation Ω

La notation $f = \Omega(g)$ est équivalente à $g = O(f)$. On peut donc informellement l'interpréter comme voulant dire que «$f \geq g$», où cette inégalité doit être comprise comme asymptotique et *à une constante près*.

Formellement on a :

Définition 1.3 Soit f, g deux fonctions en une variable n :

$$f = \Omega(g) \Rightarrow \exists C, n_0 \in \mathbb{R}^+ : \forall n \geq n_0, Cf(n) \geq g(n)$$

Dans le cas où f et g admettent toutes deux une limite en ∞, on peut de façon équivalente utiliser la définition :

$$f = \Omega(g) \Rightarrow \exists c \in \mathbb{R}^+ : \lim_{n\to\infty} f(n)/g(n) \geq c$$

Exemple 1.4 On se contentera de quelques exemples, donnés sans démonstration :

1. $f : n \mapsto n^2$. On a $f = \Omega(n)$, $f = \Omega(n^2)$, mais $f \neq \Omega(n^2 \log(n))$ et $f \neq \Omega(n^3)$.
2. $f : n \mapsto 2^n$. On a que pour tout entier positif k, $f = \Omega(n^k)$.

1.3 Notation Θ

Les notations O et Ω traduisent toutes deux des inégalités (asymptotiques, à constante près). La notation Θ combine ces deux notations afin de traduire une égalité (asymptotique, à constante près), en définissant $f = \Theta(g)$ comme $f = O(g) \wedge f = \Omega(g)$ (ou de façon strictement équivalente, $f = O(g) \wedge g = O(f)$).

Contrairement aux notations précédentes, la notation Θ est symétrique (par construction), et $f = \Theta(g)$ ssi. $g = \Theta(f)$.

Exemple 1.5 On se contentera de quelques exemples, donnés sans démonstration :

1. Soit f une fonction polynomiale de degré d, k un entier positif, alors $f = \Theta(n^k)$ ssi. $k = d$.
2. $f : n \mapsto 2^n$. Il n'existe aucun entier positif k tel que $f = \Theta(n^k)$.

1.4 Notation *o*

On a vu précédemment que par exemple, la fonction $f : n \mapsto n$ est à la fois un $O(n)$ et un $O(n^2)$, mais seulement un $\Theta(n)$ et non un $\Theta(n^2)$. La notation o permet d'exprimer concisément le fait que $f = O(n^2) \wedge f \neq \Theta(n^2)$. Informellement, $f = o(g)$ peut s'interpréter comme voulant dire que « $f \ll g$ » (f est « très inférieure », ou « dominée », ou « négligeable » devant g). Comme précédemment, cette inégalité est asymptotique ; contrairement à précédemment, celle-ci n'est plus à une constante près, car son caractère « fort » permet de dominer toutes les constantes.

Formellement, on a :

Définition 1.6 Soit f, g deux fonctions en une variable n :

$$f = o(g) \Rightarrow \forall C \in \mathbb{R}^+, \exists n_C : \forall n \geq n_C, Cf(n) < g(n)$$

Autrement dit, quelque soit la constante par laquelle on multiplie f, celle-ci sera toujours inférieure à g à partir d'un certain point.
Dans le cas où f et g admettent toutes deux une limite en ∞, on peut de façon équivalente utiliser la définition :

$$f = o(g) \Rightarrow \lim_{n \to \infty} f(n)/g(n) = 0$$

Exemple 1.7

1. $f : n \mapsto n^2$. Montrons $f = o(n^2 \log n)$. Il suffit de constater que pour toute constante positive C, $n_C := 2^C + 1$ est tel que $n \geq n_C \Rightarrow \log(n) > C$.
2. La notation o permet de reformuler plus clairement notre comparaison entre fonctions

polynomiales et exponentielles : toute fonction polynomiale est un $o(e^n)$.

1.5 Notation ω

La notation $f = \omega(g)$ est équivalente à $g = o(f)$. On peut donc informellement l'interpréter comme voulant dire que « $f \gg g$ », où cette inégalité doit être comprise comme asymptotique.

Formellement, on a :

Définition 1.8 Soit f, g deux fonctions en une variable n :

$$f = \omega(g) \Rightarrow \forall C \in \mathbb{R}^+, \exists n_C : \forall n \geq n_C, f(n) > Cg(n)$$

Exemple 1.9 On se contentera de quelques exemples, donnés sans démonstration :

1. $f : n \mapsto 542n^2$. On a $f = \Omega(n)$, $f = \omega(n)$ et $f = \Omega(n^2)$, mais $f \neq \omega(n^2)$.
2. $f : n \mapsto \log(n)$. On a $f = \omega(1)$.

1.6 Notation $\tilde{O}$

La notation $f = \tilde{O}(g)$ (qui se lit « f est un soft-Oh de g ») est similaire à la notation O, si ce n'est qu'elle « ignore » les facteurs polylogarithmiques en g : $f = \tilde{O}(g)$ si $f = O(g \log^k g)$ pour une certaine constante positive k. Elle est principalement utilisée pour raccourcir les notations dans des contextes où les facteurs (poly)logarithmiques sont courants.

On peut de la même façon définir une notation $\tilde{\Omega}$, et de là une notation $\tilde{\Theta}$.

Exemple 1.10 On se contentera de quelques exemples, donnés sans démonstration :

1. $f : n \mapsto n^2 \log^{18}(n)$. On a $f = \tilde{O}(n^2)$, bien que $f \neq O(n^2)$.
2. $f : n \mapsto n^4 2^n$. On a $f = \tilde{O}(2^n)$, bien que $f \neq O(2^n)$.

2 Somme des termes d'une suite

Le coût d'un algorithme se calcule souvent en partie ou en totalité comme la somme des termes d'une suite. On rappelle ici les expressions pour les suites arithmétiques, géométriques (exponentielles) et harmonique, et donne quelques illustrations.

§ Changements de variable

On utilise par la suite deux règles de changement de variable dans l'écriture d'une somme, qu'on résume ici [1].

1. Soit $\sum_{i=a}^{b} i$, on peut procéder au changement de variable $i \mapsto (i + c)$ pour réécrire cette somme comme $\sum_{i=a-c}^{b-c} i + c$.

1. Un changement de variable doit s'appliquer pour toute occurrence de la variable d'index dans le terme de la somme, mais on les illustre ici avec des termes ne comportant qu'une seule occurrence, pour plus de clarté.

Exemple. $\sum_{i=1}^{n} i-1 = \sum_{i=0}^{n-1} i$.

2. Soit $\sum_{i=a}^{b} i$, on peut procéder au changement de variable $i \mapsto (a+b-i)$ pour réécrire cette somme comme $\sum_{i=a}^{b} a+b-i$.

   Exemples. $\sum_{i=0}^{n} i = \sum_{i=0}^{n} n-i$; $\sum_{i=3}^{7} i = \sum_{i=3}^{7} 10-i$.

2.1 Suite arithmétique

Définition 2.1 Une *suite arithmétique* (réelle) (u_i) est une fonction $\mathbb{N} \to \mathbb{R}$ dont les termes sont liés par la relation de récurrence $u_{i+1} = u_i + t$, pour une certaine constante $t \in \mathbb{R}$, la *raison (arithmétique)* de la suite. De façon équivalente, c'est une suite dont les termes sont définis comme $u_i = u_0 + ti$.

Proposition 2.2 On note S_n la somme $\sum_{i=0}^{n-1} u_i$ des n premiers termes d'une suite arithmétique (u_i). On a :

$$S_n = \frac{n(2u_0 + t(n-1))}{2} = \frac{n(u_0 + u_{n-1})}{2}$$

Démonstration. Par définition, $S_n = \sum_{i=0}^{n-1} u_0 + ti$. On peut aussi écrire cette somme «à l'envers» en utilisant le changement de variable $i \to (n-1-i)$, ce qui donne $S_n = \sum_{i=0}^{n-1} u_0 + t(n-1-i)$; en sommant ces deux expressions on obtient $2S_n = \sum_{i=0}^{n-1} u_0 + ti + u_0 + t(n-1-i) = \sum_{i=0}^{n-1} 2u_0 + t(n-1)$. Puisque cette dernière expression a n termes constants, on obtient $2S_n = n(2u_0 + t(n-1))$, d'où $S_n = n(2u_0 + t(n-1))/2$. ∎

R Cette formule peut se mémoriser comme : «(nombre de termes) fois (premier terme plus dernier terme) divisé par deux».

R En considérant u_0 et t comme des constantes, on a que $S_n = \Theta(n^2)$.

R On a le cas particulier notable de la suite des entiers consécutifs, définie par $u_0 = 1$ et $t = 1$. Dans ce cas, la somme des n premiers termes (parfois appelée le $n^{\text{ième}}$ nombre *triangulaire*, noté T_n) vaut $n(n+1)/2$.

R La somme des n termes consécutifs à partir du $j^{\text{ième}}$ s'obtient simplement en remplaçant u_0 par u_j dans l'expression de S_n ci-dessus. Ceci se voit facilement par un calcul direct, ou en remarquant que cette somme est égale à celle des n premiers termes de la suite arithmétique (v_i) de premier terme u_j et de raison t.

R La somme des n premiers termes de la sous-suite *décimée* (v_i), définie par $v_i = u_{ki}$ pour une certaine constante k, s'obtient en remplaçant t par kt dans l'expression de S_n ci-dessus.

§ **Exemple d'application**

On considère l'algorithme de tri par sélection suivant :

Algorithme 2.3

```
def SelSort(T, n):
    for i in range(n - 1):
        mi = i
        for j in range(i + 1, n):
            if T[j] < T[mi]:
                mi = j
        T[i], T[mi] = T[mi], T[i]
```

On constate qu'à sa $i^{\text{ème}}$ itération, le corps de la boucle intérieure (indexée par j) sera exécuté $n-i-1$ fois. Puisque la boucle extérieure est exécutée pour $i \in \{0,\dots,n-2\}$, le nombre total d'exécution du corps de la boucle intérieure est de $I_n := \sum_{i=0}^{n-2} n-i-1 = \sum_{i=0}^{n-2} n-(i+1)$. Par changement de variable $i \to (i-1)$, $I_n = \sum_{i=1}^{n-1} n-i$, ce qui peut se réécrire « à l'envers » (ou plutôt « à l'endroit ») *via* le changement de variable $i \to (n-i)$ comme $\sum_{i=1}^{n-1} i = T_{n-1}$. Le coût de `SelSort` est alors majoré par $aT_{n-1} + b(n-2)$, pour certaines constantes a, b, et (puisqu'il est minoré par $T_{n-1} + (n-2)$) il est donc en $\Theta(n^2)$.

2.2 Suite géométrique

Définition 2.4 Une *suite géométrique* (réelle) (u_i) est une fonction $\mathbb{N} \to \mathbb{R}$ dont les termes sont liés par la relation de récurrence $u_{i+1} = u_i q$, pour une certaine constante $q \in \mathbb{R} \setminus \{0,1\}$, la *raison (géométrique)* de la suite[a]. De façon équivalente, c'est une suite dont les termes sont définis comme $u_i = u_0 q^i$.

a. Les cas $q \in \{0,1\}$ sont bien définis mais peu intéressants, et les exclure simplifie la formule de la somme des termes de la suite.

Proposition 2.5 On note S_n la somme $\sum_{i=0}^{n-1} u_i$ des n premiers termes d'une suite géométrique (u_i). On a :

$$S_n = u_0 \frac{q^n - 1}{q-1} = u_0 \frac{1-q^n}{1-q}$$

Démonstration. Pour q entier, on a par analogie avec l'écriture des nombres en base q que $q^n = 1 + (q-1)\sum_{i=0}^{n-1} q^i$, ce qui donne le résultat voulu. Il suffit pour conclure de montrer que cette expression reste valide pour un $q \in \mathbb{R} \setminus \{0,1\}$ quelconque, ce qu'on fait par récurrence.

Pour le cas dégénéré $n = 0$, on a bien $1 + (q-1)\cdot 0 = 1 = q^0$ (en prenant la convention qu'une somme vide s'évalue à zéro), et pour $n = 1$ on vérifie également que $1 + (q-1)\sum_{i=0}^{0} q^i = q$. On suppose

maintenant que l'expression est vraie jusqu'à un certain rang $n-1$, et on calcule :

$$\begin{aligned} 1+(q-1)\sum_{i=0}^{n} q^i &= 1+(q-1)\cdot q^n+(q-1)\sum_{i=0}^{n-1} q^i \\ &= 1+q^{n+1}-q^n+q^n-1 \\ &= q^{n+1} \end{aligned}$$

■

§ **Exemple d'application**

On considère un algorithme capable de résoudre un certain problème à condition qu'une borne B soit connue pour une certaine quantité, et qui échoue si cette borne s'avère être fausse. On représente cela par une fonction `BSolve(P, B)` qui retourne une solution à P, ou `False` si B n'est pas une bonne borne pour cette instance du problème. On suppose également que le coût d'invocation de `BSolve` est uniquement fonction de B, et on le note c_B.

Pour résoudre une instance arbitraire du problème sans connaître la borne appropriée, une approche possible est de commencer par invoquer `BSolve` avec $B = 2$, et de doubler la valeur de B tant que `BSolve` renvoie `False`. Ceci donne l'algorithme suivant :

Algorithme 2.6

```
def SloStartSolve(P):
    S = False
    B = 2
    while not S:
        S = BSolve(P, B)
        B = 2 * B
    return S
```

Pour analyser le coût total de cet algorithme `SloStartSolve`, on note B_P la borne minimale permettant de résoudre l'instance P, et on définit $d := \lceil \log(B_P) \rceil$. Le nombre d'itération de la boucle de `SloStartSolve` est donc de d, et le coût total C_P de l'exécution de l'algorithme sur l'entrée P sera alors $\sum_{i=1}^{d} c_{2^i}$.

Supposons maintenant que $c_B = B$; ceci donne $C_P = \sum_{i=1}^{d} 2^i = 2^{d+1} - 2$. Puisque $B_P > 2^{d-1}$, ne pas connaître la borne et utiliser `SloStartSolve` n'a fait perdre dans ce cas qu'un facteur d'au plus 4 par rapport à une unique exécution de `BSolve` avec la meilleure borne possible, et cette majoration ne dépend notablement pas de B_P.

2.3 Suite harmonique

Définition 2.7 Une *suite harmonique* (u_i) est une fonction $\mathbb{N} \to \mathbb{R}$ dont les termes sont les inverses d'une suite arithmétique (dont aucun terme n'est nul). De façon équivalente, c'est une suite dont les termes sont définis comme $u_i = \frac{1}{u_0+ti}$ pour une certaine raison t.

Il n'existe pas de formule close connue pour la somme des termes d'une suite harmonique. On se contentera donc de donner le résultat d'approximation suivant pour la suite harmonique la plus courante,

qui prend $u_0 = t = 1$. La somme des n premiers termes de cette suite, souvent appelé $n^{\text{ième}}$ *nombre harmonique*, et noté H_n, est approchée par :

$$H_n := \sum_{i=1}^{n} \frac{1}{i} = \Theta(\log(n))$$

C'est une conséquence du comportement limite (que l'on admettra) $\lim_{n\to\infty} H_n = \ln(n) + \gamma$, avec $\gamma \approx 0.57722$.

§ **Exemple d'application**

On considère l'algorithme `Sieve` suivant, qui prend en entrée une liste initialisée tout à `True` et calcule la liste des nombres premiers inférieurs à n :

Algorithme 2.8

```
def Sieve(P, n):
    for i in range(2, n - 1):
        if (P[i] == True):
            for j in range(2 * i, n, i):
                P[j] = False
    return P
```

Une analyse grossière du coût de `Sieve` montre facilement que (dans un modèle de coût algébrique) le temps d'exécution est un $O(n^2)$, puisque l'algorithme comprend deux boucles imbriquées de borne supérieure n (ou $n-1$). Cependant, on peut nettement améliorer cette analyse pour montrer que c'est un $\Theta(n\log(\log(n)))$.

On montre dans un premier temps que c'est un $O(n\log(n))$ en constatant qu'à la $i^{\text{ème}}$ itération de la boucle extérieure, le corps de boucle exécute au plus n/i instructions (à une constante près). Le coût de `Sieve(P,n)` est donc majoré (pour une certaine constante C) par $\sum_{i=1}^{n} Cn/i = CnH_n = \Theta(n\log(n))$. Cette borne peut encore être améliorée en constatant que la boucle intérieure n'est exécutée que quand i est premier. Le coût exact de `Sieve(P,n)` est donc (à une constante et à un terme $\Theta(n)$ près) $\sum_{i\in\mathscr{P}_n} n/i$, où $\mathscr{P}_n$ désigne l'ensemble des nombres premiers inférieurs à n. Il est alors possible de montrer que $HP_n := \sum_{i\in\mathscr{P}_n} 1/i = \Theta(\log(\log(n)))$ (ce que l'on admettra), ce qui donne le résultat voulu.

3 Probabilités

Notation. Dans cette section, on utilisera parfois la notation $[\phi]$ avec ϕ un prédicat quelconque pour désigner une quantité valant 1 si ϕ est vrai, et 0 sinon.

3.1 Définitions

§ **Univers et distributions**

On appelle *univers*, et note généralement Ω, l'ensemble des *événements* sur lesquels porte une *distribution de probabilité*. Dans tout ce qui suit, on considérera uniquement des univers **finis**.

Définition 3.1 Une *distribution de probabilité* pour un univers (fini) Ω est une fonction $\mathfrak{D} : \Omega \to [0,1]$ qui satisfait la condition $\sum_{\omega\in\Omega} \mathfrak{D}(\omega) = 1$.

Notation. Pour tout sous-ensemble d'événements $\mathscr{A} \subseteq \Omega$ et une distribution $\mathfrak{D}$ sur Ω, la notation $\mathfrak{D}(\mathscr{A})$ désigne la somme $\sum_{\omega\in\mathscr{A}} \mathfrak{D}(\omega)$ des probabilités des événements de $\mathscr{A}$. Cette notation s'étend naturellement aux expressions sur les ensembles, par exemple $\mathfrak{D}(\mathscr{A} \cap \mathscr{B})$.
Une conséquence particulière de cette notation est que $\forall \mathscr{A}, \mathscr{B} \subseteq \Omega$, $\mathscr{A} \cap \mathscr{B} = \varnothing$, on a $\mathfrak{D}(\mathscr{A} \cup \mathscr{B}) = \mathfrak{D}(\mathscr{A}) + \mathfrak{D}(\mathscr{B})$.

Exemple 3.2

1. Soit $\Omega = \{0,1\}$, la distribution $\mathfrak{D} : 1 \mapsto p \ ; \ 0 \mapsto 1-p$ pour $p \in [0,1]$ est généralement appelée *distribution de Bernoulli de paramètre* p, et notée Ber_p.
2. Soit Ω un univers quelconque de N événements, la *distribution uniforme* sur Ω est la distribution $\mathfrak{U}$ telle que $\forall \omega \in \Omega, \mathfrak{U}(\omega) = 1/N$.

R Lors de l'analyse probabiliste d'un problème, il est important de bien définir l'univers qui capture correctement les phénomènes à modéliser, et notamment de ne pas en oublier. Par exemple, l'univers capturant le tirage d' `1d6` (un tirage d'un dé à six faces) n'est pas le même que celui capturant le tirage de `6d6` (six tirages indépendants d'un dé à six faces) : bien que le processus de tirage soit le même dans les deux cas, il faut considérer plus d'événements dans le second (6^5 fois plus !).

R L'univers à lui-seul ne suffit pas à modéliser un problème : il est essentiel d'également préciser quelle distribution est considérée sur cet univers. Par exemple, « tirer un nombre aléatoire entre 1 et 6 » donne trop peu d'informations sur le processus considéré pour qu'il soit possible de conclure grand chose.

§ Variables aléatoires

La plupart du temps, les résultats en probabilités sont exprimés par l'intermédiaire de *variables aléatoires*, qui sont des fonctions de Ω vers un ensemble quelconque.

Définition 3.3 Une *variable aléatoire* est une fonction $X : \Omega \to \mathscr{S}$, pour un certain ensemble $\mathscr{S}$.

Notation. On utilise généralement une lettre latine capitale pour désigner une variable aléatoire. De plus, quand l'univers de définition est clair, on note généralement simplement X à la place de $X(\omega)$ (pour un certain ω).

Notation. Pour X, Y des variables aléatoires, les expressions $X + Y$, XY etc. ont le même sens usuel que par exemple $f + g$ pour f, g deux fonctions quelconques. C'est à dire que $X + Y$ est la fonction $\omega \mapsto X(\omega) + Y(\omega)$. Toutefois, pour que cette notation ait du sens, il faut que les opérations correspondantes sur X, Y, etc. soient bien définies sur leurs espaces de valeur.

Vocabulaire. Une variable aléatoire est *échantillonnée* (ou *tirée*) suivant une distribution pour son univers de définition. La *réalisation* d'une variable aléatoire est la valeur qu'elle prend à l'issue de ce tirage.

Un peu plus formellement, ceci donne :

Définition 3.4 Soit $\mathfrak{D}$ une distribution sur Ω et $x \in \mathscr{S}$, on note $\Pr_{\mathfrak{D}}[X = x]$ (ou simplement $\Pr[X = x]$ quand $\mathfrak{D}$ est claire dans le contexte) la probabilité $\sum_{\omega \in \Omega} \mathfrak{D}(\omega)[X(\omega) = x]$ que X se *réalise* en x quand elle est *échantillonnée* suivant $\mathfrak{D}$, ce qu'on note $X \sim \mathfrak{D}$.

Notation. La notation $\Pr[X = x]$ s'étend naturellement à n'importe quel prédicat portant sur une ou plusieurs variables aléatoires. On notera par exemple $\Pr[(X = 0) \vee (X = 1)]$ pour la probabilité qu'X se réalise en 0 ou en 1, et $\Pr[(X = 0) \wedge (Y + Z = 0)]$ pour la probabilité qu'X et la somme $Y + Z$ se réalisent toutes deux en 0.

R Il est toujours possible de construire une nouvelle variable aléatoire pour représenter une expression comme celles données en exemple ci-dessus. En conséquence, on se contente par la suite d'énoncer les résultats en fonction de variables aléatoires « élémentaires », ce qui se fait sans perte de généralité.

Vocabulaire. Soit X et Y deux variables aléatoires à valeur dans un ensemble fini $\mathscr{S}$, on dit que X et Y *suivent la même distribution* ssi. $\forall s \in \mathscr{S}$, $\Pr[X = s] = \Pr[Y = s]$

Un intérêt des variables aléatoires est qu'une unique variable (en tant que fonction) peut être échantillonnée suivant plusieurs distributions pour son univers de définition. Par exemple, pour $X : \omega \mapsto \omega$, $\mathfrak{D} = \mathrm{Ber}_{1/2}$, $\mathfrak{D}' = \mathrm{Ber}_{1/4}$, on a par définition que $\Pr_{\mathfrak{D}}[X = 1] = 1/2$, et $\Pr_{\mathfrak{D}'}[X = 1] = 1/4$.

Un autre intérêt des variables aléatoires est qu'elles permettent facilement d'exprimer le fait qu'une même distribution peut être échantillonnée plusieurs fois. Pour cette raison, il est important de distinguer l'égalité de fonction entre deux variables aléatoires, et l'égalité des variables aléatoires elles-mêmes. De fait la fonction « représentée » par une variable aléatoire est souvent (comme dans l'exemple ci-dessus) la fonction identité, et sauf mention du contraire c'est ce qui doit être compris lors de sa définition. Pour autant, même si $X \sim \mathfrak{D}$ et $Y \sim \mathfrak{D}$ sont toutes deux des variables aléatoires représentant la fonction identité, elles sont distinctes en tant que variables aléatoires, et notamment leurs réalisations ne sont pas nécessairement identiques.

Exemple 3.5 On tire deux fois `1d20` (un dé à 20 faces), et note x et y les résultats respectifs des deux tirages. Pour tout entier $a \in [\![1, 20]\!]$, les probabilités que x et y valent a sont toutes deux égales à $1/20$. Pour autant, on n'a pas nécessairement $x = y$.
Ceci peut se formaliser en représentant les deux tirages par deux variables aléatoires *indépendantes* X et Y (représentant chacune la même fonction identité de $[\![1, 20]\!]$ dans lui-même) échantillonnées suivant la distribution uniforme sur $[\![1, 20]\!]$.
Dans cet exemple, l'univers capturant *les deux* tirages est $[\![1, 20]\!] \times [\![1, 20]\!]$, et non pas $[\![1, 20]\!]$, qui pourrait correspondre à l'univers pour un seul tirage. S'il est donc licite de raisonner sur un seul des deux tirages en considérant uniquement l'univers $[\![1, 20]\!]$, raisonner conjointement sur les deux tirages (par exemple pour calculer $\Pr[X = Y]$) ne peut se faire que dans l'univers « complet » $[\![1, 20]\!] \times [\![1, 20]\!]$.

§ Indépendance

L'exemple 3.5 ci-dessus utilise la notion d'indépendance de variables aléatoires, qu'on définit de la façon suivante :

Définition 3.6 Deux variables aléatoires X et Y définies sur un même univers Ω et à valeur respectivement dans $\mathscr{X}$, $\mathscr{Y}$ sont *indépendantes* ssi. :

$$\forall x \in \mathscr{X}, y \in \mathscr{Y},\ \Pr[X = x \wedge Y = y] = \Pr[X = x] \times \Pr[Y = y]$$

Cette définition s'étend à n variables aléatoires de façon évidente : $X_1, \ldots, X_n$ sont *mutuellement indépendantes* ssi. $\forall x_1 \in \mathscr{X}_1, \ldots, x_n \in \mathscr{X}_n$, $\Pr[\wedge_{i=1}^{n}(X_i = x_i)] = \prod_{i=1}^{n} \Pr[X_i = x_i]$.

R Il n'est pas nécessaire de supposer ci-dessus que les variables aléatoires sont échantillonnées suivant la même distribution.

Proposition 3.7 Soit $\mathscr{S}$ un ensemble fini, $\mathfrak{U}$ (resp. $\mathfrak{U}^n$) la distribution uniforme sur $\mathscr{S}$ (resp. $\mathscr{S}^n$). Soit $\vec{X} \sim \mathfrak{U}^n$ et $\vec{Y} := (Y_1, \ldots, Y_n)$, où les Y_i sont mutuellement indépendantes et échantillonnées suivant $\mathfrak{U}$. Alors $\vec{X}$ et $\vec{Y}$ suivent la même distribution.

Démonstration. Par application directe de la définition 3.6. ■

R On verra ci-dessous une définition équivalente de l'indépendance de deux variables aléatoires en utilisant le langage des probabilités conditionnelles.

L'indépendance de deux (ou plus) variables aléatoires peut être comme dans l'exemple 3.5 prise comme prémisse : si l'on définit X et Y comme deux variables aléatoires indépendantes, on pose comme vrai le fait qu'elles satisfont la définition 3.6. Ce peut aussi être une conséquence plus indirecte d'un ensemble de prémisses et de la définition des variables aléatoires ; dans ce cas, l'indépendance doit être prouvée. Par exemple, on peut montrer que soit $X, Y \sim \mathrm{Ber}_{1/2}$ indépendantes, $Z := X \oplus Y$ (où '$\oplus$' dénote le OU EXCLUSIF), alors X et Z sont indépendantes, mais X, Y, Z ne sont pas mutuellement indépendantes.

§ **Probabilité conditionnelle**

La définition 3.6 donne la valeur de $\Pr[X = x \wedge Y = y]$ quand X et Y sont indépendantes, mais ne dit rien dans le cas contraire. Pour cela, on introduit la notion de *probabilité conditionnelle* :

Définition 3.8 Soit X, Y deux variables aléatoires définies sur un même univers Ω et à valeur respectivement dans $\mathscr{X}$, $\mathscr{Y}$; $x \in \mathscr{X}$, $y \in \mathscr{Y}$. On définit la *probabilité conditionnelle qu'X se réalise en x sachant qu'Y se réalise en y*, notée $\Pr[X = x \mid Y = y]$ à travers l'identité :

$$\Pr[X = x \wedge Y = y] = \Pr[X = x \mid Y = y] \times \Pr[Y = y]$$

On a alors immédiatement :

Proposition 3.9 Deux variables aléatoires X et Y définies sur un même univers Ω et à valeur respectivement dans $\mathscr{X}$, $\mathscr{Y}$ sont indépendantes ssi. :

$$\forall x \in \mathscr{X}, y \in \mathscr{Y},\ \Pr[X = x \mid Y = y] = \Pr[X = x]$$

Démonstration. C'est une conséquence immédiate des définitions 3.6 et 3.8. ■

Dans le cas où $\Pr[Y = y] \neq 0$, il peut être utile de réécrire l'identité de la définition 3.8 pour mettre directement en avant la définition de $\Pr[X = x \mid Y = y]$ comme :

$$\Pr[X = x \mid Y = y] = \frac{\Pr[X = x \wedge Y = y]}{\Pr[Y = y]}$$

Ceci suggère notamment l'interprétation d'une probabilité conditionnelle comme une probabilité restreinte à un certain « sous-univers », où pour calculer $\Pr_{\mathfrak{D}}[X = x \mid Y = y]$ on considère une distribution renormalisée définie sur le sous-univers (qu'on supposera non vide) constitué des événements tels que $Y = y$. Un peu plus formellement, on peut introduire l'univers $\Omega_y := \{\omega \in \Omega \mid Y(\omega) = y\}$ et la distribution $\mathfrak{D}_y : \Omega_y \to [0,1]$, $\omega \mapsto \mathfrak{D}(\omega)/\mathfrak{D}(\Omega_y)$ et simplement exprimer $\Pr_{\mathfrak{D}}[X = x \mid Y = y]$ comme $\Pr_{\mathfrak{D}_y}[X = x]$.

3.2 Quelques propriétés

§ Probabilités totales

Lors de l'analyse probabiliste d'un problème, il est courant de partitionner l'univers et de raisonner en termes de probabilités conditionnelles relativement à l'appartenance de certaines variables aléatoires à un sous-univers. Par exemple, pour étudier le comportement d'un algorithme probabiliste qui commence par tirer un bit B suivant une certaine distribution, on peut analyser séparément les deux cas $B = 0$ et $B = 1$ pour ensuite les recombiner en pondérant par $\Pr[B = 0]$ et $\Pr[B = 1]$. Ceci s'exprime formellement par le :

Proposition 3.10 Soit X, Y deux variables aléatoires définies sur un univers Ω, et à valeur respectivement dans les ensembles (finis) $\mathscr{X}$ et $\mathscr{Y}$, on a :

$$\forall x \in \mathscr{X},\ \Pr[X = x] = \sum_{y \in \mathscr{Y}} \Pr[X = x \mid Y = y] \Pr[Y = y]$$

Démonstration. Par définition des probabilités conditionnelles, $\sum_{y\in\mathscr{Y}} \Pr[X = x \mid Y = y]\Pr[Y = y] = \sum_{y\in\mathscr{Y}} \Pr[X = x \wedge Y = y]$. On note $\Omega_x := \{\omega \in \Omega \mid X(\omega) = x\}$: c'est l'union disjointe de ses partitions $\Omega_{x,y} := \{\omega \in \Omega \mid X(\omega) = x \wedge Y(\omega) = y\}$. On a alors bien pour toute distribution $\mathfrak{D}$ sur Ω que :

$$\begin{aligned}\Pr_{\mathfrak{D}}[X = x] &= \mathfrak{D}(\Omega_x) = \mathfrak{D}(\cup_{y\in\mathscr{Y}} \Omega_{xy}) \\ &= \sum_{y\in\mathscr{Y}} \mathfrak{D}(\Omega_{xy}) = \sum_{y\in\mathscr{Y}} \Pr_{\mathfrak{D}}[X = x \wedge Y = y]\end{aligned}$$

■

Exemple 3.11 Comme application de la proposition 3.10, on peut calculer pour deux variables aléatoires X et Y à valeur dans un groupe commutatif fini $\mathbb{G}$ la distribution suivie par leur somme $X+Y$. En utilisant $\Pr[X+Y = \alpha \mid Y = \beta] = \Pr[X = \alpha-\beta \mid Y = \beta]$, on a par application immédiate de la proposition 3.10 que :

$$\forall \alpha \in \mathbb{G},\ \Pr[X + Y = \alpha] = \sum_{\beta\in\mathbb{G}} \Pr[X = \alpha - \beta \mid Y = \beta]\Pr[Y = \beta]$$

Dans le cas particulier où X et Y sont indépendantes, on a $\Pr[X = \alpha - \beta \mid Y = \beta] = \Pr[X = \alpha - \beta]$, et l'expression ci-dessus se simplifie en :

$$\forall \alpha \in \mathbb{G},\ \Pr[X + Y = \alpha] = \sum_{\beta \in \mathbb{G}} \Pr[X = \alpha - \beta] \Pr[Y = \beta]$$

La distribution de $X + Y$ n'est alors rien d'autre que la convolution discrète des distributions de X et Y.

§ **Formule de Bayes**

La formule de Bayes permet d'exprimer une probabilité conditionnelle $\Pr[X = x \mid Y = y]$ en fonction de $\Pr[Y = y \mid X = x]$:

Proposition 3.12 Soit X, Y deux variables aléatoires définies sur un univers Ω, et à valeur respectivement dans les ensembles (finis) $\mathscr{X}$ et $\mathscr{Y}$, on a :

$$\forall x \in \mathscr{X}, y \in \mathscr{Y} \text{ t.q. } \Pr[Y = y] \neq 0,\ \Pr[X = x \mid Y = y] = \frac{\Pr[Y = y \mid X = x] \Pr[X = x]}{\Pr[Y = y]}$$

Démonstration. C'est une conséquence immédiate du fait que d'après la définition 3.8, on a :

$$\Pr[X = x \wedge Y = y] = \Pr[X = x \mid Y = y] \Pr[Y = y] = \Pr[Y = y \mid X = x] \Pr[X = x]$$ ■

Bien qu'élémentaire, la « réécriture » permise par cette proposition 3.12 est souvent utile.

§ **Borne de l'union**

Étant données deux expressions comme $X = x$ et $Y = y$, il est en général faux que $\Pr[X = x \vee Y = y] = \Pr[X = x] + \Pr[Y = y]$. Cependant, on montre aisément que l'*inégalité* $\Pr[X = x \vee Y = y] \leq \Pr[X = x] + \Pr[Y = y]$ est toujours valide. En généralisant à n expressions, on obtient la *borne de l'union* :

Proposition 3.13 Soit $X_1, \ldots, X_n$ n variables aléatoires définies sur un univers Ω et à valeur respectivement dans les ensembles $\mathscr{X}_1, \ldots, \mathscr{X}_n$, on a :

$$\forall x_1 \in \mathscr{X}_1, \ldots, x_n \in \mathscr{X}_n,\ \Pr[\vee_{i=1}^{n} (X_i = x_i)] \leq \sum_{i=1}^{n} \Pr[X_i = x_i]$$

Démonstration. On montre le résultat par induction. Pour $n = 1$, celui-ci est trivialement vrai. Pour $n = 2$, on pose $\Omega_{x_1}^{X_1} = \{\omega \in \Omega \mid X_1(\omega) = x_1\}$, $\Omega_{x_2}^{X_2} = \{\omega \in \Omega \mid X_2(\omega) = x_2\}$, $\Omega_{x_1,x_2}^{X_1,X_2} = \{\omega \in \Omega \mid X_1(\omega) = x_1 \vee X_2(\omega) = x_2\} = \Omega_{x_1}^{X_1} \cup \Omega_{x_2}^{X_2}$. On a alors pour n'importe quelle distribution $\mathfrak{D}$ sur Ω que :

$$\mathfrak{D}(\Omega_{x_1,x_2}^{X_1,X_2}) = \mathfrak{D}(\Omega_{x_1}^{X_1}) + \mathfrak{D}(\Omega_{x_2}^{X_2}) - \mathfrak{D}(\Omega_{x_1}^{X_1} \cap \Omega_{x_2}^{X_2})$$

d'où (puisque $\mathfrak{D}(\Omega_{x_1}^{X_1} \cap \Omega_{x_2}^{X_2}) \geq 0$) :

$$\Pr_{\mathfrak{D}}[X_1 = x_1 \vee X_2 = x_2] = \mathfrak{D}(\Omega_{x_1,x_2}^{X_1,X_2}) \leq \mathfrak{D}(\Omega_{x_1}^{X_1}) + \mathfrak{D}(\Omega_{x_2}^{X_2}) = \Pr_{\mathfrak{D}}[X_1 = x_1] + \Pr_{\mathfrak{D}}[X_1 = x_1]$$

Supposons maintenant que le résultat tienne pour un certain entier n, soit (omettant les quantificateurs pour alléger l'écriture) : $\Pr[\vee_{i=1}^{n}(X_i = x_i)] \leq \sum_{i=1}^{n} \Pr[X_i = x_i]$; on veut montrer $\Pr[\vee_{i=1}^{n+1}(X_i = x_i)] \leq \sum_{i=1}^{n+1} \Pr[X_i = x_i]$. On introduit pour cela une « nouvelle » variable aléatoire Y telle que $Y = (x_1, \ldots, x_n)$ ssi. $\bigvee_{i=1}^{n}(X_i = x_i)$ s'évalue à vraie. L'inégalité étant vraie pour $n = 2$, on a alors :

$$\Pr[\vee_{i=1}^{n+1}(X_i = x_i)] = \Pr[Y = (x_1, \ldots, x_n) \vee X_{n+1} = x_{n+1}] \leq \Pr[Y = (x_1, \ldots, x_n)] + \Pr[X_{n+1} = x_{n+1}]$$

ce qui en utilisant l'hypothèse d'induction est bien inférieur à $\sum_{i=1}^{n+1} \Pr[X_i = x_i]$. ■

R L'égalité est atteinte quand (en utilisant les notations de la preuve) les intersections $\Omega_{x_i}^{X_i} \cap \Omega_{x_j}^{X_j}$ sont toutes vides.

R Un intérêt de la borne de l'union est notamment qu'elle ne requiert pas que les variables aléatoires soient indépendantes.

Exemple 3.14 Considérons une table de hachage utilisant une fonction de hachage h et comprenant N compartiments où l'on insère Q éléments distincts $x_1, \ldots, x_Q$. On souhaite majorer la probabilité p_c qu'au moins un compartiment contienne plus d'un élément ou, autrement dit, qu'il y ait une *collision* entre les hachés des Q éléments.
On modélise pour cela les hachés $h(x_i)$ par autant de variables aléatoires H_i (échantillonnées par l'intermédiaire du tirage de h, *cf. infra*), et on introduit $Q(Q+1)/2$ variables aléatoires $C_{i,j}$ chacune définie (pour tout i, $j > i$) comme le prédicat $H_i = H_j$. On peut noter que les $C_{i,j}$ ne sont *pas* indépendantes.
On remarque alors que $p_c = \Pr[\vee_{1 \leq i < j \leq Q} C_{i,j}]$, qu'on peut (par la borne de l'union) majorer par $\sum_{1 \leq i < j \leq Q} \Pr[C_{i,j}]$.
Si l'on suppose maintenant que la fonction h a été tirée uniformément dans une famille de fonctions *universelles*, on a par définition d'une telle famille que pour tout i, $j > i$, $\Pr[C_{i,j}] \leq 1/N$.
Ceci donne finalement la majoration :

$$p_C \leq \frac{Q(Q+1)}{2N}$$

R Cette majoration s'exprime parfois asymptotiquement comme : $p_C \leq O(Q^2/N)$.

R Ce résultat est souvent appelé *borne des anniversaires*, en référence au « paradoxe » du même nom.

Exercice 3.1. Dans le jeu vidéo *Baldur's Gate 3* (et le système *Dungeon & Dragons* sur lequel il est basé), un *jet de compétence* se fait en tirant 1d20, et un *succès critique* (respectivement *échec critique*) a lieu quand la valeur de 20 (respectivement 1) est obtenue. Un personnage ayant un *avantage* pour un jet tire deux fois 1d20, et garde le meilleur résultat des deux.

1. Calculez la probabilité d'obtenir un échec critique, avec et sans avantage.
2. En utilisant la borne de l'union, calculez un majorant de la probabilité p_{crit} d'obtenir un succès critique en cas d'avantage.
3. Calculez exactement p_{crit}. (*Indice : il peut être intéressant de considérer pour chaque tirage les probabilités de ne* ***pas*** *obtenir 20.*)

§ **Échantillonnage par rejet**

L'*échantillonnage par rejet* est une méthode générale permettant de tirer un élément d'un certain ensemble suivant une certaine distribution par l'intermédiaire d'un échantillonnage d'un surensemble suivant la même distribution. Un cas concret se posant souvent en programmation est le besoin de tirer un nombre entier uniformément dans un intervalle $[\![a, b]\!]$ quelconque, alors que les primitives de génération d'aléa disponibles ne permettent que de tirer un nombre dans un intervalle fixé (par exemple $[\![0, 2^{64} - 1]\!]$).

Le principe de l'échantillonnage par rejet est le suivant : soit $\mathscr{S}' \subset \mathscr{S}$ l'ensemble dans lequel on souhaite tirer un élément, $\mathfrak{D}$ une certaine distribution sur $\mathscr{S}$ suivant laquelle on sait échantillonner, $\phi_{\mathscr{S}'} : \mathscr{S} \to \{\top, \bot\}$ un prédicat permettant de tester si un élément de $\mathscr{S}$ est inclus dans $\mathscr{S}'$, on tire simplement $X \sim \mathfrak{D}$, applique $\phi_{S'}$ à sa réalisation, et renvoie celle-ci si le prédicat s'évalue à vrai ; sinon, on recommence.

> **Proposition 3.15** L'algorithme ci-dessus échantillonne $\mathscr{S}'$ suivant la distribution $\mathfrak{D}_{\mathscr{S}'}$ d'univers $\mathscr{S}'$, définie comme $\omega \mapsto \mathfrak{D}(\omega)/\mathfrak{D}(\mathscr{S}')$.

Démonstration. Soit n le nombre total de tirages (suivant la distribution $\mathfrak{D}$) faits par l'algorithme d'échantillonnage par rejet avant qu'il ne retourne une valeur, on note $X_1, \ldots, X_n$ les variables aléatoires correspondant à ces tirages successifs. Par définition de l'algorithme, seule la réalisation de X_n est retournée, et on a $X_n \in \mathscr{S}'$. On peut donc calculer pour tout élément $x \in \mathscr{S}'$ la probabilité qu'il soit renvoyé par l'algorithme comme $\Pr_{\mathfrak{D}}[X_n = x \mid X_n \in \mathscr{S}',\ X_1, \ldots, X_{n-1} \notin \mathscr{S}']$, ce qui par indépendance mutuelle des X_i et par définition de la probabilité conditionnelle est bien égal à $\mathfrak{D}(x)/\mathfrak{D}(\mathscr{S}')$. ■

> **Exemple 3.16** On a à notre disposition un dé à six faces, et on souhaite tirer un nombre uniformément entre 1 et 35. En appliquant le principe de l'échantillonnage par rejet, il suffit de tirer deux fois le dé, d'interpréter le résultat en base six comme un nombre entre 1 et 36, et de considérer deux cas : retourner la valeur obtenue si elle est différent de 36 ; sinon recommencer.

Un inconvénient possible de l'algorithme défini ci-dessus est qu'il se peut que $\Pr_{\mathfrak{D}}[\phi_{\mathscr{S}'}(X)]$ soit (très) faible, ce qui induit alors un coût d'exécution important. On utilise donc parfois une étape intermédiaire consistant à d'abord appliquer une fonction de projection $\rho : \mathscr{S} \to \mathscr{S}''$ telle que $\mathscr{S}' \subseteq \mathscr{S}''$ et que la distribution $\mathfrak{D}_{\mathscr{S}'}$ est préservée par l'application préalable de ρ.

> **Exemple 3.17** On a à notre disposition un générateur aléatoire permettant de tirer uniformément des nombres dans $[\![0, 2^{64}]\!]$, et on souhaite tirer un nombre uniformément entre 1 et 35.
> On remarque tout d'abord que tirer un nombre X, le réduire modulo 35, puis lui ajouter 1, ne permet pas d'échantillonner $[\![1, 35]\!]$ uniformément : en effet, un calcul nous montre qu'on aura de cette façon une probabilité 527049830677415761/18446744073709551615 de générer les 16 premiers nombres, mais seulement 527049830677415760/18446744073709551615 de générer les 19 suivants.
> On peut cependant remarquer que soit $Y := \rho(X)$ avec ρ la fonction réduisant sont entrée modulo 64, on a bien que Y est distribuée uniformément sur $[\![0, 63]\!]$. On peut donc appliquer la méthode d'échantillonnage par rejet sur Y, ce qui permet à chaque itération d'avoir une probabilité de rejet de seulement 29/64 au lieu de 18446744073709551581/18446744073709551616 si l'on

avait directement appliqué la méthode sur X

R Bien que directement renvoyer X réduite modulo 35 (et incrémentée de 1) ne produise pas une distribution uniforme sur $[\![1,35]\!]$, on peut néanmoins constater que celle-ci en semble «très proche». Pour quantifier rigoureusement cette proximité, on pourrait par exemple calculer la *distance de variation totale* entre la distribution uniforme et celle obtenue, mais ce sujet dépasse le cadre de ce cours.

Une variante de la modification ci-dessus consiste à échanger les étapes de rejet et de projection : on définit à nouveau un ensemble intermédiaire $\mathscr{S}'' \supseteq \mathscr{S}'$ et une fonction de projection $\rho : \mathscr{S}'' \to \mathscr{S}'$ telle que (par un abus de notation) $\rho(\mathfrak{D}_{\mathscr{S}''}) = \mathfrak{D}_{\mathscr{S}'}$, et on échantillonne $\mathfrak{D}_{\mathscr{S}''}$ par rejet.

Exemple 3.18 On reprend l'exemple 3.17, avec maintenant $\mathscr{S}'' = [\![0, 2^{64} - 17]\!]$, et ρ la fonction qui réduit son entrée modulo 35 et y ajoute 1. L'échantillonnage uniforme de $\mathscr{S}''$ peut se faire par rejet, avec une probabilité de rejet ridicule de 2^{-60}. De plus, comme $2^{64} - 16$ est un multiple de 35, ce processus échantillonne bien $[\![1,35]\!]$ uniformément.

3.3 Espérance

§ Définition

L'*espérance* d'une variable aléatoire X à valeur dans $\mathbb{R}$ (ou plus généralement n'importe quelle structure dotée d'une multiplication externe par $\mathbb{R}$) se définit comme la somme des réalisations possible de X pondérée par les probabilités de ces réalisations. Formellement :

Définition 3.19 Soit une variable aléatoire $X : \Omega \to \mathbb{R}$, $\mathfrak{D}$ une distribution sur Ω, l'*espérance* de $X \sim \mathfrak{D}$, notée $E[X]$, est définie par :

$$E[X] := \sum_{\omega \in \Omega} X(\omega)\mathfrak{D}(\omega)$$

Exemple 3.20

1. On considère la distribution uniforme sur $[\![1,6]\!]$ et $X : \omega \mapsto \omega$; on a :
$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2}$$
2. Soit le même univers et la même distribution que ci-dessus, mais avec $X : \omega \mapsto \omega \div 2$ (où l'opérateur « $\div$ » représente le reste de la division euclidienne de son opérande de gauche par son opérande de droite), on a :
$$E[X] = 1 \times \frac{1}{6} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 0 \times \frac{1}{6} = \frac{3}{2}$$

Outre son interprétation «naturelle» en tant que «valeur moyenne attendue», l'espérance d'une variable aléatoire est une quantité qui intervient dans l'énoncé de nombreux résultats en probabilités

(qui consistent justement souvent à quantifier à quel point l'espérance « résume » la distribution suivant laquelle la variable aléatoire est tirée). Ces résultats font également souvent intervenir la *variance*, qui est une autre fonction d'une variable aléatoire. On n'abordera cependant pas ces sujets plus en détail dans le cadre de ce cours.

Exercice 3.2. Japin est un *Assassin* de niveau 3. Pour son arme secondaire, il hésite entre équiper une *dague + 1* et une *épée courte*. Lors d'une attaque, les dégâts infligés par la dague + 1 sont calculés comme `1d4 + 1` (contre `1d4` pour une dague conventionnelle), et ceux de l'épée courte comme `1d6`.

1. Soit X une variable aléatoire tirée uniformément dans un intervalle $[\![a, b]\!]$, montrez que $E[X] = (a+b)/2$.
2. Calculez l'espérance des dégâts infligés par la dague conventionnelle, la dague + 1 et l'épée courte. Quel est l'effet du bonus +1 sur l'espérance ?
3. En sachant que la dague + 1 a également un bonus aux *jets d'attaque* (ce qui lui permet d'infliger plus souvent des dégâts que l'épée courte), quelle arme conseillerez vous à Japin d'équiper ?

On a vu dans l'exercice ci-dessus que l'espérance d'une variable aléatoire uniforme se calcule facilement comme une fonction des seuls paramètres de la distribution. Ceci est vrai en général pour un grand nombre de distributions et les variables aléatoires qui leur sont naturellement associées. On donne ci-dessous comme autre exemple celui d'une distribution *géométrique*, qui intervient souvent dans l'analyse d'algorithmes probabilistes.

Définition 3.21 La distribution *géométrique* $\mathfrak{G}_p$ de paramètre $p \in]0,1[$ [a] est définie sur l'univers $\mathbb{N} \setminus \{0\}$ par $\mathfrak{G}_p : k \mapsto (1-p)^{k-1}p$.

a. On exclut ici les cas dégénérés $p \in \{0,1\}$ qui ont peu d'intérêt pour notre propos.

L'interprétation naturelle d'une distribution géométrique est que $\mathfrak{G}_p(k)$ représente la probabilité que si l'on tire un nombre arbitrairement grand de variables aléatoires indépendantes suivant une distribution de Bernoulli de paramètre p, alors le $k^{\text{ième}}$ tirage est le premier où la variable de Bernoulli se réalise en 1.

En préparation au calcul de l'espérance d'une telle distribution, on va d'abord montrer le résultat suivant, intéressant en lui-même :

Proposition 3.22 Soit X une variable aléatoire à valeur dans $\mathbb{N}$, $E[X] = \sum_{k=1}^{\infty} \Pr[X \geq k]$.

Démonstration. Par définition de l'espérance, et par hypothèse sur le co-domaine de X, on a :

$$E[X] = \sum_{k=0}^{\infty} k \Pr[X = k] = \sum_{k=1}^{\infty} k \Pr[X = k]$$

En utilisant l'égalité $\Pr[X = k] = \Pr[X \geq k] - \Pr[X \geq k+1]$, et en notant respectivement (u_n) et (v_n) les suites des sommes partielles $\sum_{k=1}^{n} k \Pr[X = k]$ et $\sum_{k=1}^{n} \Pr[X \geq k]$, on a que :

$$v_n = n \Pr[X \geq n+1] + u_n$$

On considère alors deux cas :

1. E[X] est infinie. Dans ce cas $\lim_{n\to\infty} u_n = \infty$, et puisque $n\Pr[X \geq n+1] \geq 0$, l'égalité ci-dessus nous donne $\lim_{n\to\infty} v_n = \infty$.
2. E[X] est finie. Dans ce cas la série $\sum_{k=1}^{\infty} k\Pr[X=k]$ converge absolument vers $E[X]$. On va montrer que $\sum_{k=1}^{\infty} \Pr[X \geq k]$ converge vers la même valeur en étudiant la limite de la différence des suites (u_n) et (v_n). Pour étudier cette différence, on utilise l'inégalité :

$$n\Pr[X \geq n+1] = \sum_{k=n+1}^{\infty} n\Pr[X=k] \leq \sum_{k=n+1}^{\infty} k\Pr[X=k]$$

Or la convergence de la série $\sum_{k=1}^{\infty} k\Pr[X=k]$ implique que $\lim_{n\to\infty} \sum_{k=n+1}^{\infty} k\Pr[X=k] = 0$; en particulier, il existe un rang n à partir duquel cette quantité (et donc la différence entre (u_n) et (v_n)) est inférieure à 1. On a donc que la suite (v_n) est monotone et bornée par $E[X]+1$, ce qui garantit sa convergence ; sa différence avec (u_n) tendant vers zéro, elle converge vers la même valeur, soit $E[X]$. ■

On peut maintenant montrer :

Proposition 3.23 Soit $X \sim \mathfrak{G}_p$, $E[X] = \frac{1}{p}$.

Démonstration. On applique directement la proposition 3.22 pour calculer l'espérance comme :

$$E[X] = \sum_{k=1}^{\infty} \Pr[X \geq k] = \sum_{k=1}^{\infty} (1-p)^{k-1} = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$$

où l'avant-dernière égalité est donnée par la valeur de la série géométrique de raison $(1-p)$. ■

Exemple 3.24 La proposition 3.23 peut immédiatement s'utiliser dans l'analyse de l'algorithme d'échantillonnage par rejet de la section précédente : à chaque itération de la boucle d'échantillonnage, on a une certaine probabilité p de retourner une valeur suivant la distribution cible, et cette probabilité est indépendante des précédentes itérations. La distribution du nombre d'itération est donc géométrique de paramètre p, et l'espérance du nombre d'itération est infini (ou non défini) si $p = 0$, 1 si $p = 1$, et $1/p$ sinon.

§ Linéarité de l'espérance

On conclut cette section par montrer que l'espérance est *linéaire*, ce qui est une propriété souvent utile lors du calcul de l'espérance d'une variable aléatoire qui s'exprime comme la somme d'autres variables aléatoires *pas nécessairement indépendantes*, mais suivant la même distribution.

Proposition 3.25 Soit X, Y deux variables aléatoires : $\Omega \to \mathbb{R}$ échantillonnées suivant une distribution $\mathfrak{D}$, on a $E[X+Y] = E[X] + E[Y]$.

Démonstration.

$$E[X+Y] = \sum_{\omega\in\Omega} (X(\omega)+Y(\omega))\mathfrak{D}(\omega) = \sum_{\omega\in\Omega} X(\omega)\mathfrak{D}(\omega) + \sum_{\omega\in\Omega} Y(\omega)\mathfrak{D}(\omega) = E[X] + E[Y]$$ ■

Exemple 3.26 On considère le même scénario qu'à l'exemple 3.14, et cherche à calculer l'espérance du nombre de collisions, c'est à dire du nombre de paires $(H_i, H_{j\neq i})$ telles que $H_i = H_j$. Soit $C'_{i,j}$ les variables aléatoires $[C_{i,j}]$ valant 1 si $C_{i,j}$ est vraie et 0 sinon, on introduit la variable aléatoire $K := \sum_{1\leq i<j\leq Q} C'_{i,j}$ qui compte le nombre de collisions. Par application directe de la proposition 3.25, on a alors :

$$E[K] = \sum_{1\leq i<j\leq Q} E[C'_{i,j}] = \frac{Q(Q-1)}{2N}$$

R On peut remarquer que la valeur de $E[K]$ est égale au majorant calculé dans l'exemple 3.14 de la probabilité p_c d'obtenir *une* collision. En effet, les deux calculs menant à ces résultats sont exactement les mêmes, et notamment ignorent tous deux les dépendances entre les $C_{i,j}$.

