

Génération aléatoire de polytopes en nombres entiers en utilisant les chaînes de Markov

Julien David, Lionel Pournin, Rakotonarivo Rado

24 janvier 2018

Résumé. Ce document va décrire une approche pour faire de la génération aléatoire de polytope entiers en utilisant les chaînes de Markov, de la construction du générateur à l'optimisation de celui-ci. Les objets à engendrer sont les polytopes entiers contenus dans l'hypercube $[0, k]^d$, notés (d, k) -polytopes. La construction du générateur aléatoire se fait de la manière suivante : modéliser une chaîne de Markov dont les états seront les (d, k) -polytopes, lancer des marches sur la chaîne jusqu'à ce qu'on atteigne une distribution stationnaire. On va montrer que cette distribution stationnaire est unique et uniforme. De ce fait, on aura construit un générateur aléatoire uniforme de (d, k) -polytopes.

1 Introduction

Une chaîne de Markov est un processus qui évolue en temps sur un ensemble d'états Ω et est caractérisée par sa matrice de transition P , qui décrit les transitions entre les états de Ω .

L'idée principale de la génération aléatoire en utilisant les chaînes de Markov est la suivante : modéliser une chaîne de Markov dont les états seront les objets à engendrer, lancer des marches sur la chaîne jusqu'à ce qu'on atteigne une distribution stationnaire. L'uniformité du générateur sera conditionnée par la distribution stationnaire de la chaîne de Markov ainsi modélisée. Cette approche va être utilisée pour notre cas dans la génération de polytopes en nombres entiers contenus dans l'hypercube $[0, k]^d$, notés (d, k) -polytopes.

Bien que l'idée de laquelle on part soit simple, il y a cependant plusieurs difficultés dans le cas des (d, k) -polytopes : nous n'avons pas d'information exacte sur la taille de notre espace des états, la matrice de transition n'est pas

entièrement déterminée, et enfin, assurer une distribution uniforme nécessite que certaines conditions sur cette dernière soient vérifiées.

Toutefois, se baser sur une chaîne de Markov semble être une piste sérieuse vu les nombreux résultats et différentes techniques qui sont déjà à notre disposition [1]. Il y a également le fait qu'on peut faire des essais expérimentaux : modéliser notre chaîne et lancer une batterie de tests dessus.

Le travail sur cette approche sera principalement de :

- Trouver des opérations locales sur nos (d, k) -polytopes pour pouvoir passer d'un état à un autre.
- Utiliser les techniques et les résultats connus sur les chaînes de Markov pour estimer le temps de mélange.
- Réfléchir sur des idées de preuves pour l'uniformité du générateur.

Ce document sera structuré de la manière suivante. Dans un premier temps, un bref rappel des principaux résultats sur les chaînes de Markov - tous essentiellement proviennent de notre lecture de [1]. Et dans un deuxième temps notre modèle concret de générateur en utilisant les chaînes de Markov.

2 Chaînes de Markov

Le nombre de polytopes entiers dans $[0, k]^d$ est fini, par conséquent on va s'intéresser aux chaînes de Markov à espace d'états fini. Dans toute la suite du document, on supposera que la chaîne de Markov étudiée est une chaîne à espace d'états fini.

2.1 Définitions

Une chaîne de Markov à espace d'états fini $(X_t)_{t \geq 0}$ est un processus qui évolue en temps sur les éléments d'un espace Ω fini. Le déplacement d'un état $x \in \Omega$ à un autre état de Ω suit une distribution fixe $P(x, \cdot)$. C'est une séquence de variables aléatoires (X_0, X_1, \dots) à valeur dans Ω .

La valeur de (X_t) étant l'état de la chaîne à l'instant t , la **propriété de Markov** sur (X_t) assure que l'état à l'instant $t + 1$ ne dépend que de l'état à l'instant t .

La matrice de transition P de (X_t) qui décrit les transitions entre les états de Ω est une matrice stochastique dont la x -ème ligne désigne la distribution $P(x, \cdot)$. On a :

$$\sum_{y \in \Omega} P(x, y) = 1 \quad \forall x \in \Omega \quad (1)$$

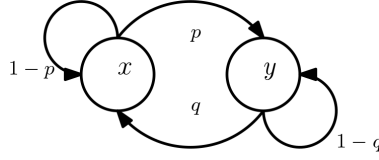


FIGURE 1 – Une chaîne de Markov telle que : $\Omega = \{x, y\}$, $P(x, x) = 1 - p$, $P(x, y) = p$, $P(y, x) = q$ et $P(y, y) = 1 - q$

2.2 Propriétés

Soient $(X_t)_{t \geq 0}$ une chaîne de Markov, Ω son espace des états et P sa matrice de transition.

2.2.1 Irréductibilité

La matrice P de X_t est **irréductible** si tout état x de Ω est accessible depuis n'importe quel état de Ω . En d'autres mots, cela signifie que le graphe de X_t , dont les sommets sont les états de Ω et les arêtes les relations de transitions entre les états de Ω , est fortement connexe. Formellement, on a :

$$\exists r_0 \quad : \quad \forall r > r_0, \forall x, y \in \Omega \quad P^r(x, y) > 0 \quad (2)$$

2.2.2 Périodicité

On définit par $\mathcal{T}(x) := \{t \geq 1 : P^t(x, x) > 0\}$, l'ensemble des temps où la chaîne retourne en x . La **période** de x est défini par $\text{pgcd}(\mathcal{T}(x))$.

Propriété 2.2.1. *Si P est irréductible alors tous les états de Ω ont la même période.*

X_t est dite **apériodique** si tous les états de Ω ont pour période 1, sinon elle est dite **périodique**.

2.2.3 Symétrie

On dit que X_t est **symétrique** si pour tout x et y de Ω , la probabilité de passer de x en y et la probabilité de passer de y en x sont égales ; i.e. :

$$P(x, y) = P(y, x) \quad \forall x, y \in \Omega \quad (3)$$

2.2.4 Hitting time et premier temps de retour

Pour tout état $x \in \Omega$ on définit le **hitting time** par :

$$\tau_x := \min\{t \geq 0 : X_t = x\}, \quad (4)$$

le premier moment où la chaîne visite x . en particulier, le **premier temps de retour** en x est défini par :

$$\tau_x^+ := \min\{t \geq 1 : X_t = x\} \quad \text{si } X_0 = x \quad (5)$$

2.3 Récurrence d'un état

Un état x est dit **récurrent positif** si $E_x(\tau_x^+) < \infty$, où $E_x(\tau_x^+)$ est l'espérance du premier temps de retour en x en partant de x . Cette quantité représente l'intervalle de temps typique entre deux passages consécutifs en x .

2.4 Distribution stationnaire

On note par μ_t le vecteur ligne qui donne la distribution de probabilité sur Ω à l'instant t et par $\mu_t(x)$ la probabilité d'être en x à l'instant t , i.e. $\mu_t(x) = \mathbf{P}\{X_t = x\} \quad \forall x \in \Omega$. Obtenir la distribution à l'instant $t+1$ revient à appliquer une fois à μ_t la matrice de transition P par la droite. On a :

$$\mu_t = \mu_{t-1}P \quad \forall t \geq 1 \quad (6)$$

En partant d'une distribution initiale μ_0 , on a :

$$\mu_t = \mu_0 P^t \quad \forall t \geq 0 \quad (7)$$

Une distribution π est dite **stationnaire** pour (X_t) si elle satisfait

$$\pi = \pi P \quad (8)$$

De manière équivalente :

$$\pi(y) = \sum_{x \in \Omega} \pi(x)P(x, y) \quad \forall y \in \Omega \quad (9)$$

La distribution stationnaire peut être perçue comme étant la représentation du temps passé en chaque état x de Ω .

2.5 Temps de mélange

Le **temps de mélange** est le temps minimum pour atteindre la distribution stationnaire si elle existe. Dans notre cas, comme on veut atteindre une distribution uniforme, le temps de mélange nous donne une idée sur l'efficacité de notre générateur.

Plusieurs résultats de [1] nous donnent des bornes (inférieures et supérieures) sur le temps de mélange. Ces résultats seront rappelés et utilisés plus tard dans nos preuves.

2.6 Existence et unicité de π

L'existence de la distribution stationnaire pour une chaîne irréductible est fortement liée à τ_x^+ , en particulier à la récurrence de x (2.3). Pour une chaîne (X_t) irréductible et apériodique on a les propriétés suivantes :

Propriété 2.6.1. *Si P est irréductible et apériodique, les deux assertions suivantes sont équivalentes :*

i Il existe π , une distribution sur Ω tel que $\pi = \pi P$ et $\pi(x) > 0$ pour tout x de Ω .

ii $\pi(x) = \frac{1}{E_x(\tau_x^+)}$

Propriété 2.6.2. *Une chaîne irréductible à espace d'états fini est récurrente positive, i.e. tous ces états sont récurrents positifs. Elle possède une unique distribution stationnaire π .*

Propriété 2.6.3. *Si (X_t) est symétrique alors la distribution uniforme sur Ω est stationnaire.*

3 Modèle des (d, k) -polytopes

Rappelons que notre objectif principal est de construire un générateur aléatoire uniforme de (d, k) -polytopes. Et cela en utilisant les chaînes de Markov. Considérons notre hypercube $[0, k]^d$. L'idée générale de la construction du générateur aléatoire est de lancer des marches sur une chaîne de Markov (X_t) dont l'espace des états Ω est l'ensemble des (d, k) -polytopes pour k et d fixés et la matrice de transition P , n'étant pas entièrement déterminée, est définie par des opérations locales sur nos (d, k) -polytopes.

Nous allons prouver que (X_t) ainsi définie aura l'uniforme comme distribution stationnaire et ainsi obtenir un générateur uniforme de (d, k) -polytopes.

3.1 Notations et contexte du générateur

Il est utile de préciser les quelques notations que l'on va utiliser dans toute la suite du document.

L'espace des états Ω de (X_t) est l'ensemble des (d, k) -polytopes. Un état x de Ω est un ensemble de l points qui constituent exactement une enveloppe convexe à l sommets, i.e. $x = \{v_1, v_2, \dots, v_l\}$ avec $v_i \in [0, k]^d$ pour tout $i \in [1, l]$. On observe alors que $\text{Conv}(x) = x$. Dans certains contextes on préférera faire référence à l'enveloppe convexe $\text{Conv}(\cdot)$ quand des opérations locales seront appliquées à nos états.

On notera par $|x|$ le nombre de sommets de x .

Soit $x = \{v_1, v_2, \dots, v_l\} \in \Omega$. On considère :

1. $x - \{v_j\}$ l'enveloppe convexe auquel on a retiré exactement le j -ème sommet de x . On remarque que $x - \{v_j\}$ n'est de dimension pleine que si $|x| > d + 1$.
2. $x \cup \{v\}$, où $v \in [0, k]^d$, l'enveloppe convexe auquel on a ajouté exactement un point à x sans avoir retiré aucun des l sommets de x . On observe que $|x \cup \{v\}| = |x| + 1$.

3.2 Règles de transition

Les règles de transition sur notre espace Ω est définie par des opérations locales sur nos (d, k) -polytopes, des opérations qui consistent à ajouter ou supprimer un sommet pour passer d'un état à un autre.

Soient x et y deux états de Ω . On tire uniformément un point v dans $[0, k]^d$. On se donne les règles suivantes :

- Si v est un point intérieur à x ; v appartient à l'enveloppe convexe de x mais n'est pas dans x , on boucle sur x
- Si v appartient à x :
 - Si x a au moins $d + 2$ sommets alors on supprime v de x . On a une transition de x vers $y = x - \{v\}$
 - Sinon on boucle sur x
- Si v est tiré à l'extérieur de l'enveloppe convexe de x , on calcule l'enveloppe convexe de $x \cup \{v\}$:
 - Si l'ensemble des sommets de l'enveloppe convexe de $x \cup \{v\}$ est exactement $x \cup \{v\}$, alors on a une transition de x vers $y = x \cup \{v\}$
 - Sinon on boucle sur x

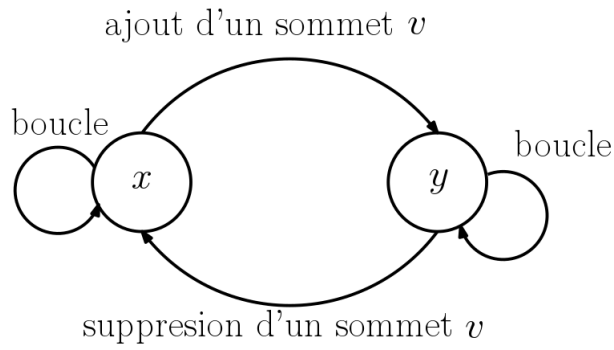


FIGURE 2 – Relation de transition entre deux états x et y de Ω : $y = x \cup \{v\}$ et $x = y - \{v\}$, où $v \in [0, k]^d$.

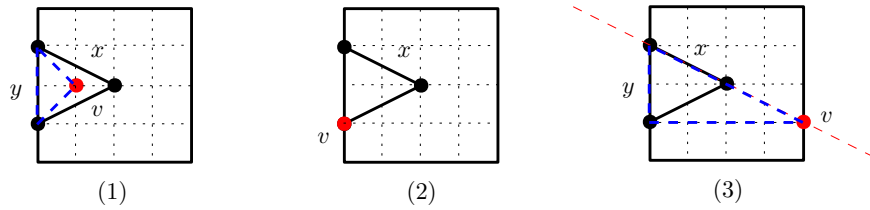


FIGURE 3 – Pour $[0, 4]^2$. Exemples de cas où on a une boucle sur x : (1) v est un point intérieur à x , $P(x, y) = 0$. (2) $v \in x$ et $|x| < d + 2$. (3) v est tiré à l'extérieur de $\text{Conv}(x)$ et $|\text{Conv}(x \cup \{v\})| \neq |x| + 1$, $P(x, y) = 0$.

3.3 Marches sur (X_t)

Considérons maintenant notre chaîne (X_t) ainsi définie. Engendrer aléatoirement un (d, k) -polytope consiste à effectuer une marche aléatoire sur Ω avec les règles de transition que l'on a précédemment définies jusqu'à ce qu'on atteigne la distribution stationnaire.

Et justement ce temps d'arrêt de notre marche, qui est exactement le temps de mélange de X_t , sera déterminé par des conditions qu'il nous faut encore établir de manière à optimiser l'efficacité de notre générateur. Plusieurs techniques sont à notre disposition mais elles seront étudiées plus tard.

Engendrer aléatoirement un (d, k) -polytope, en se basant sur ce modèle de chaîne de Markov est donné par l'algorithme suivant :

Algorithme 1 : $\Gamma(d, k)$: Engendrer aléatoirement un (d, k) -polytope

Entrées : la dimension d , le côté k de l'hypercube

Sortie : (d, k) -polytope tiré aléatoirement dans $[0, k]^d$

```

1 état initial arbitraire  $x_0$ 
2 tant que on n'a pas atteint la distribution stationnaire faire
3   | tirer aléatoirement un point  $v$  dans  $[0, k]^d$ 
4   | si  $v \in x_0$  et  $|x_0| > d + 1$  alors
5   |   |  $x_0 = x_0 - \{v\}$ 
6   | sinon si  $|Conv(x_0 \cup \{v\})| == |x_0| + 1$  alors
7   |   |  $x_0 = x_0 \cup \{v\}$ 
8 renvoyer  $x_0$ 
```

Théorème 1. *La chaîne de Markov $(X_t)_{t \geq 0}$ est irréductible, apériodique et a l'uniforme comme distribution stationnaire.*

Avant de passer à la preuve du théorème 1, on va introduire la notion de **différence symétrique** entre deux polytopes de Ω . Elle nous sera utile dans la preuve de l'irréductibilité de (X_t) .

Soient deux polytopes $x = \{u_1, \dots, u_l\}$ et $y = \{v_1, \dots, v_k\} \in \Omega$.

Définition 3.3.1. *On définit par $x \triangle y$ la **différence symétrique** entre x et y telle que*

$$x \triangle y = \{u \in x : u \notin y \text{ et } v \in y : v \notin x\} \quad (10)$$

On peut voir la différence symétrique de manière ensembliste comme étant $x \triangle y = x \cup y \setminus x \cap y$ cependant quelques précisions sont à mentionner :

- $x \cup y$ ne constitue pas forcément une enveloppe convexe.
- $|x \cup y| = |x| + |y|$ si $x \cap y = \emptyset$.
- $x \triangle y$ est maximal quand x et y n'ont aucun sommet en commun.

Proposition 3.3.1. *Le cardinal de $x \triangle y$ constitue une borne inférieure de la distance entre x et y dans le graphe de X_t , on notera cette distance $\delta(x, y)$ et on a :*

$$\delta(x, y) \geq |x \triangle y| \quad (11)$$

Démonstration. Considérons x et $y \in \Omega$. Comme $x \triangle y$ constitue l'ensemble des sommets sur lesquels x diffère de y et réciproquement, passer de x en y avec un nombre minimal d'étapes consiste à choisir un chemin qui fera en sorte de réduire $x \triangle y$ d'un sommet à chaque étape. Par conséquent, il faut au moins $|x \triangle y|$ étapes pour passer de x en y . ■

Remarque 3.3.1. *Pour passer d'un état x à un état y de Ω , l'idéal serait de directement ajouter des sommets de y et de supprimer ceux de x , mais certaines configurations ne le permettent pas. Il faut alors trouver des états transitoires entre x et y .*

Lemme 3.3.1. *Pour tout simplexe $x \in \Omega$ et pour tout $y \in \Omega$, il existe un simplexe $z \in \Omega$ tel que $\delta(x, z) = 2$ et $|x \triangle y| = |z \triangle y|$.*

Démonstration. Soit $X = \{v_1, \dots, v_{d+1}\}$, un simplexe de Ω . Considérons pour chaque facette \mathcal{F}_i de X , le parallélépipède \mathcal{P}_i définit de la manière suivante,

$$\mathcal{P}_i = \text{Conv}(X \cup \{u_i\}) \quad (12)$$

où u_i est le symétrique du sommet $v_i \in X \setminus \mathcal{F}_i$ par rapport à \mathcal{F}_i . Le point u_i ici n'est pas forcément entier. Si il ne l'est pas, \mathcal{P}_i n'est pas un état de la chaîne X_t .

Considérons ensuite l'ensemble $\mathcal{K} = [0, k]^d \cap \mathcal{P}_i$ pour tout i . ■

Proposition 3.3.2. *Pour tout x et $y \in \Omega$, $\exists z \in \Omega$, tel que $|x \triangle y| > |z \triangle y|$, pour lequel on a $\delta(x, z) \leq 3$.*

Démonstration. Considérons X et $Y \in \Omega$, tel que $P(X, Y) = 0$. Passer de X en Y consiste en à trouver un nombre fini d'opérations d'ajouts et de suppressions de sommets ; chaque opération correspond à une transition vers un état Z qui doit être à priori plus proche de Y . On observe alors les cas suivants :

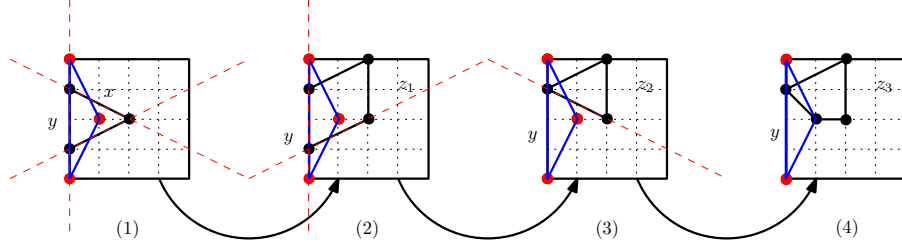


FIGURE 4 – Ici pour x et $y \in [0, 4]^2$, avec $|x \triangle y| = 6$. On trouve un $z_3 \in \Omega$, tel que $|x \triangle y| > |z_3 \triangle y| = 5$, pour lequel on a $\delta(x, z_3) = 3$.

1. X n'est pas un simplexe.
 - (a) $X \subset Y$: On ajoute $v \in Y \setminus X$ et $Z = X \cup \{v\}$, alors $\delta(X, Z) = 1$
 - (b) $X \not\subset Y$: On supprime $v \in X \setminus Y$ et $Z = X - \{v\}$, alors $\delta(X, Z) = 1$
2. X est un simplexe.
 - (a) Si on peut ajouter $v \in Y \setminus X$ alors on le fait, alors $Z = X \cup \{v\}$ et $\delta(X, Z) = 1$
 - (b) Sinon :
 - i. Ajouter un point extérieur à $X \triangle Y$
 - ii. Supprimer un élément de $X \setminus Y$
 - iii. Ajouter un élément de $Y \setminus X$
 Dans ce cas on trouve un Z tel que $\delta(X, Z) = 3$

L'idée de la preuve consiste à montrer que si on ne trouve pas un état Z à distance 1 de X alors on trouvera toujours un point extérieur à $X \triangle Y$ que l'on pourra ajouter. Dans ce cas, passer de X à Z nécessitera au plus 3 étapes.

Les cas où l'on ne puisse pas ajouter un point de $Y \setminus X$ sont les cas où les éléments de $Y \setminus X$ sont des points intérieurs à $\text{Conv}(X)$ et/ou des points sur les droites qui supportent les arêtes de X . Pour exemple voir le point (1) de la figure 4. Comme on considère le cas de figure où X est un simplexe, ces droites sont au nombre de $\frac{d(d+1)}{2}$.

Ajouter un point u extérieur à $X \triangle Y$ puis supprimer un point de $X \setminus Y$ permet de lever cette contrainte sur l'ajout dans $Y \setminus X$. Montrons alors qu'il est toujours possible de trouver ce point u . Il suffit de prendre un point u dans l'ensemble \mathcal{K} du lemme 3.3.1.

■

Une fois ces propositions établies, passons à la preuve du théorème 1.

Démonstration. La preuve du théorème 1 se fera en trois étapes :

i *Irréductibilité*

Rapellons que X_t est irréductible si à partir d'un état on peut atteindre n'importe quel état de Ω en un temps fini, i.e. on peut trouver un r_0 fini tel que pour tout $x_1, x_2 \in \Omega$, $P^r(x_1, x_2) > 0$ dès que $r \geq r_0$. Soient $x_1, x_2 \in \Omega$. Considérons ensuite deux simplexes y_1 et y_2 de Ω .

À compléter ...

ii *Symétrie*

On rappelle que X_t est symétrique si et seulement si pour tout état x et y de Ω , $P(x, y) = P(y, x)$.

Soient $x, y \in \Omega$, deux états distincts. La relation de transition entre x et y est donnée par nos règles de transition : elle consiste à ajouter ou retirer un unique sommet - voir Figure 2.

Observons que si il n'existe pas de transition en une étape de x en y , $P(x, y) = 0$, par conséquent si $P(x, y) = 0$ alors $P(y, x) = 0$.

Prouvons ensuite que si $P(x, y) > 0$ et $P(y, x) > 0$ alors $P(x, y) = P(y, x)$. $P(x, y) > 0$ signifie que l'on a une transition de x en y et deux cas peuvent se présenter : soit $y = x - \{v\}$, soit $y = x \cup \{v\}$. Étudions le premier cas où $y = x - \{v\}$. Considérons v , un point tiré uniformément sur l'ensemble des points entiers de $[0, k]^d$, la probabilité de tirer v et de l'ajouter à l'enveloppe convexe de x i.e. avoir une transition de x en y est $P(x, y) = \frac{1}{(k+1)^d}$. De même, pour avoir une transition de y en x , on tire v dans $[0, k]^d$ tel que $x = y \cup \{v\}$ avec une probabilité $P(y, x) = \frac{1}{(k+1)^d}$. On a alors

$$P(x, y) = P(y, x) = \frac{1}{(k+1)^d} \quad (13)$$

On prouve l'autre cas par un raisonnement analogue.

iii *Apériodicité*

Montrer que X_t est apériodique revient à montrer que tous les états de Ω ont tous pour période 1. On rappelle que la période d'un état x est défini comme étant le pgcd($\mathcal{T}(x)$) où $\mathcal{T}(x) = \{t \geq 1 : P^t(x, x) > 0\}$, c'est l'ensemble des temps t pour lesquels on a $\mathbf{P}\{X_t = x | X_0 = x\} > 0$.

Si X_t est irréductible la propriété 2.2.1 nous dit que tout les états ont la même période. Par conséquent pour tout $x, y \in \Omega$, $\text{pgcd}(\mathcal{T}(x)) =$

$\text{pgcd}(\mathcal{T}(y))$. Pour prouver que cette période vaut 1, il nous suffit de trouver un x tel que $\text{pgcd}(\mathcal{T}(x)) = 1$.

Soit x un simplexe de Ω avec $|x| = d + 1$. Par définition de nos règles de transition, après avoir tiré le point v dans $[0, k]^d$, on a la possibilité de boucler sur x pour les cas suivants : soit v est un point intérieur à x , soit $v \in x$, soit v est tiré à l'extérieur de x mais $|x \cup \{v\}| \neq |x| + 1$. En termes de probabilités on a : $P(x, x) = \mathbf{P}\{v \text{ un point intérieur à } x\} + \mathbf{P}\{v \in x\} + \mathbf{P}\{|x \cup \{v\}| \neq |x| + 1\}$.

On observe que $\mathbf{P}\{v \in x\} = \frac{1}{(k+1)^d}$ mais puisqu'on a choisi x tel que $|x| = d + 1$, on boucle sur x avec probabilité :

$$P(x, x) \geq \frac{1}{(k+1)^d} > 0 \quad (14)$$

(14) nous dit qu'avec une probabilité non nulle on peut revenir en x en partant de x en une étape. Dès lors, $\mathcal{T}(x)$ est de la forme $\mathcal{T}(x) = \{1, \dots\}$. Donc $\text{pgcd}(\mathcal{T}(x)) = 1$.

■

Corollaire 3.3.1. *Le générateur $\Gamma(d, k)$ décrit par l'algorithme 1 est un générateur aléatoire uniforme de (d, k) -polytopes.*

Démonstration. La construction de $\Gamma(d, k)$ se base sur le modèle d'une marche sur X_t . La preuve est immédiate et découle du théorème 1. ■

Références

- [1] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.