

(public 2009)

Résumé : Ce texte présente un modèle de répartition de composés chimiques avec forte répulsion et un algorithme qui permet de simuler une répartition typique de ces composés. La méthode est basée sur la simulation d'une chaîne de Markov auxiliaire et sur une méthode de couplage.

Mots clefs : Simulation de variables aléatoires, chaîne de Markov, probabilité invariante.

- *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. Vous êtes laissé(e) libre d'organiser votre discussion comme vous l'entendez. Des suggestions de développement, largement indépendantes les unes des autres, vous sont proposées en fin de texte. Vous n'êtes pas tenu(e) de les suivre. Il vous est conseillé de mettre en lumière vos connaissances à partir du fil conducteur constitué par le texte. Le jury appréciera que la discussion soit accompagnée d'exemples traités sur ordinateur.*

Nous étudions la répartition de deux composés chimiques A et B avec répulsion forte. Nous supposons que ces deux composés se répartissent sur une surface plane, et pour simplifier nous supposons qu'ils se répartissent sur un quadrillage, c'est-à-dire que leurs positions sont représentées par l'ensemble

$$\Lambda = \{1, \dots, N\}^2$$

où N est un entier positif (les points de Λ seront appelés sites). On note $d(x, y)$ la distance entre deux sites x et y donnée par

$$d(x, y) = |a - a'| + |b - b'|,$$

si $x = (a, b)$ et $y = (a', b')$. On dit que deux sites x et y sont voisins si $d(x, y) = 1$. Nous modéliserons la répartition des composés par un élément σ de $\{-1, 0, 1\}^\Lambda$ (appelé configuration). Si le site x ne porte pas de composé alors $\sigma(x) = 0$, si le site x porte le composé A (resp. le composé B) alors $\sigma(x) = -1$ (resp. $\sigma(x) = +1$). Nous modéliserons la répulsion en supposant que seules les configurations où A et B ne sont jamais portés par des sites voisins sont autorisées : plus précisément, on notera \mathcal{A} l'ensemble des configurations autorisées, c'est-à-dire que \mathcal{A} est l'ensemble des configurations $\sigma \in \{-1, 0, +1\}^\Lambda$ telles qu'il n'existe aucun site x et y dans Λ avec $d(x, y) = 1$ et $\sigma(x)\sigma(y) = -1$.

Pour décrire la répartition des composés les chimistes ont proposé la mesure de probabilité μ sur \mathcal{A} suivante. (Le modèle décrit la répartition des composés A et B dans un système non fermé ; le nombre de chaque composé n'est donc pas fixé). Cette loi dépend d'un paramètre $q > 0$ fixé, et est donnée par

$$\mu(\{\sigma\}) = \frac{q^{N(\sigma)}}{Z_q}, \quad \forall \sigma \in \mathcal{A},$$

où $N(\sigma) = \text{card}(\{x \in \Lambda, |\sigma(x)| = 1\})$ est le nombre de sites occupés, et Z_q est le facteur de normalisation

$$Z_q = \sum_{\sigma \in \mathcal{A}} q^{N(\sigma)}.$$

Il est clair que $q > 1$ favorise les configurations denses et $q < 1$ les configurations plus clairsemées.

Le but de ce texte est de présenter plusieurs méthodes de simulation de tirage sous la loi μ . La méthode la plus élémentaire est celle du rejet. On commence par choisir une configuration σ de la façon suivante : indépendamment pour chaque x dans Λ on choisit $\sigma(x) = 0$ avec probabilité $\frac{1}{1+2q}$, $\sigma(x) = -1$ avec probabilité $\frac{q}{1+2q}$ et $\sigma(x) = +1$ avec probabilité $\frac{q}{1+2q}$. On rejette alors la configuration si elle n'est pas dans \mathcal{A} . Cette méthode présente l'avantage de la simplicité mais elle est très inefficace numériquement car la probabilité de tomber sur une configuration autorisée est très faible quand N est grand. Nous allons présenter deux méthodes plus efficaces : une méthode de Métropolis qui permet de faire un tirage aléatoire sous une loi proche de μ , et une méthode qui permet de faire un tirage exactement sous la loi μ .

1. Méthode de Metropolis.

Pour tout site $x \in \Lambda$ on définit l'application $\tau_0^x : \mathcal{A} \mapsto \mathcal{A}$ qui correspond à enlever le composé éventuellement présent sur le site x , c'est-à-dire que pour une configuration autorisée σ on définit $\tau_0^x(\sigma)$ par

$$\tau_0^x(\sigma)(y) = \begin{cases} \sigma(y), & \text{si } y \neq x, \\ 0, & \text{si } y = x. \end{cases}$$

Pour toute configuration autorisée σ il est clair que $\tau_0^x(\sigma)$ est autorisée.

Pour tout site x on définit l'application $\tau_{-1}^x : \mathcal{A} \mapsto \mathcal{A}$ qui correspond à remplacer le composé éventuellement présent en x par un composé A si la configuration ainsi obtenue est autorisée et à ne rien faire sinon. Plus formellement, pour tout σ dans \mathcal{A} on définit $\tau_{-1}^x(\sigma)$ de la façon suivante : on commence par définir $\tilde{\sigma} \in \{-1, 0, 1\}^\Lambda$ par

$$\tilde{\sigma}(y) = \begin{cases} \sigma(y), & \text{si } y \neq x, \\ -1, & \text{si } y = x, \end{cases}$$

On pose alors

$$\tau_{-1}^x(\sigma) = \begin{cases} \tilde{\sigma}, & \text{si } \tilde{\sigma} \text{ est une configuration autorisée} \\ \sigma, & \text{sinon.} \end{cases}$$

On définit l'application τ_{+1}^x de la même façon pour le composé B .

Pour un site $x \in \Lambda$ et un réel $u \in [-1, 1]$ on définit $F_u^x : \mathcal{A} \mapsto \mathcal{A}$ par

$$F_u^x(\sigma) = \begin{cases} \tau_{-1}^x(\sigma), & \text{si } u < \frac{-1}{1+2q}, \\ \tau_{+1}^x(\sigma), & \text{si } u > \frac{1}{1+2q}, \\ \tau_0^x(\sigma), & \text{si } \frac{-1}{1+2q} \leq u \leq \frac{1}{1+2q}. \end{cases}$$

On se donne sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, deux suites indépendantes de variables aléatoires indépendantes identiquement distribuées $(X_n)_{n \geq 1}$ et $(U_n)_{n \geq 1}$ telles que les variables X_n suivent la loi uniforme sur Λ et les U_n suivent la loi uniforme sur $[-1, 1]$. Pour une configuration initiale σ_0 , on construit la suite de variables aléatoires $(\Sigma_n)_{n \geq 0}$ à valeurs dans \mathcal{A} , par

$$\Sigma_0 = \sigma_0, \quad \Sigma_n = F_{U_n}^{X_n}(\Sigma_{n-1}).$$

Théorème 1. (i) La suite $(\Sigma_n)_{n \geq 0}$ forme une chaîne de Markov à valeurs dans \mathcal{A} , récurrente et apériodique.

(ii) La mesure μ est l'unique probabilité invariante de cette chaîne de Markov.

Démonstration. (i) Les variables X_n et U_n étant indépendantes de toutes les variables X_k et U_k pour $k < n$, il est clair que Σ_n est une chaîne de Markov. Pour prouver la récurrence on remarque d'abord que pour toute configuration autorisée σ on peut trouver un chemin de probabilité positive entre σ et la configuration 0 (en effet, il suffit d'enlever un à un les composés présents dans σ); de même on peut trouver un chemin de probabilité positive entre la configuration 0 et σ (on ajoute un à un les composés présents dans σ). La chaîne est donc récurrente. L'apériodicité vient de ce qu'à partir de toute configuration σ , la probabilité de rester sur place est positive.

(ii) Notons $p_{\sigma, \sigma'}$ les probabilités de transition de la chaîne de Markov $(\Sigma_n)_{n \geq 0}$. On va montrer que la mesure μ est réversible pour la chaîne de Markov, c'est-à-dire que

$$\mu(\sigma) p_{\sigma, \sigma'} = \mu(\sigma') p_{\sigma', \sigma},$$

pour toutes configurations autorisées σ et σ' . Ceci implique que μ est la mesure invariante de $(\Sigma_n)_{n \geq 0}$. On note que $p_{\sigma, \sigma'} = p_{\sigma', \sigma} = 0$ si σ et σ' diffèrent sur plus d'un site. Supposons maintenant que σ et σ' diffèrent seulement sur le site x : si $\sigma(x)\sigma'(x) = -1$ alors $\mu(\{\sigma\}) = \mu(\{\sigma'\})$ et

$$p_{\sigma, \sigma'} = p_{\sigma', \sigma} = \frac{1}{N^2} \frac{q}{1+2q}.$$

Si par exemple $\sigma(x) = 0$ et $\sigma'(x) = +1$, alors $\mu(\{\sigma'\}) = q\mu(\{\sigma\})$ et

$$p_{\sigma, \sigma'} = \frac{1}{N^2} \frac{q}{1+2q}, \quad p_{\sigma', \sigma} = \frac{1}{N^2} \frac{1}{1+2q}.$$

Les autres cas sont identiques. □

On sait donc que pour toute configuration autorisée σ , $\lim_{n \rightarrow \infty} \mathbb{P}(\Sigma_n = \sigma) = \mu(\{\sigma\})$. On peut donc simuler une loi proche de μ en partant d'une configuration autorisée connue (par exemple la configuration ne contenant aucun composé : $\sigma_0 \equiv 0$) et pour un temps n "grand", choisir X_n comme variable aléatoire simulant un tirage sous μ : la difficulté de cette méthode est de choisir

n ; il faut pour cela avoir une estimée correcte de la vitesse de convergence vers la probabilité invariante. Nous n'allons pas étudier la vitesse de convergence pour cette chaîne mais plutôt présenter, dans les parties 2 et 3, une méthode permettant de contourner ce problème.

2. La méthode de couplage

Soient $(X_n)_{n \geq 1}$ et $(U_n)_{n \geq 1}$ deux suites indépendantes de variables i. i. d. définies comme dans la section 1. On définit $Y_n : \mathcal{A} \mapsto \mathcal{A}$ par récurrence par

$$Y_0 = \text{Id}, \quad Y_n = Y_{n-1} \circ F_{U_n}^{X_n},$$

de telle sorte que

$$(1) \quad Y_n(\sigma) = F_{U_1}^{X_1} \circ \dots \circ F_{U_n}^{X_n}(\sigma).$$

Remarque 1. On prendra garde à l'inversion de l'ordre de composition par rapport à la première section.

Les Y_n sont des variables aléatoires à valeurs dans les fonctions de \mathcal{A} dans \mathcal{A} . On définit T comme la variable aléatoire donnée par

$$T = \inf\{n \geq 0, Y_n(\sigma) = Y_n(\sigma'), \forall \sigma, \sigma' \in \mathcal{A}^2\}$$

(avec la convention $\inf \emptyset = \infty$). La variable aléatoire T est un temps aléatoire appelé temps de coalescence (car c'est le premier temps où toutes les valeurs de Y_n coalescent). Il est clair que sur l'événement $T < \infty$ on a

$$Y_{T+k}(\sigma) = Y_{T+k}(\sigma') = Y_T(\sigma),$$

pour tout $k \geq 0$ et toutes configurations autorisées σ et σ' ; c'est-à-dire que après le temps T , Y_n reste égal à la fonction constante prenant comme valeur la valeur commune des $(Y_T(\sigma))$.

Les transformations F_u^x possèdent une propriété remarquable de monotonie que nous allons maintenant décrire : on place sur l'ensemble des configurations l'ordre suivant

$$(\sigma \leq \sigma') \Leftrightarrow (\sigma(x) \leq \sigma'(x), \forall x \in \Lambda).$$

Proposition 1. Pour tout u dans $[-1, 1]$ et tout x dans Λ , la fonction F_u^x est croissante pour l'ordre défini précédemment sur \mathcal{A} , c'est-à-dire que

$$\sigma \leq \sigma' \Rightarrow F_u^x(\sigma) \leq F_u^x(\sigma').$$

Démonstration. Cela revient à prouver que pour tout site x , les fonctions τ_0^x , τ_{+1}^x , τ_{-1}^x sont croissantes. Considérons deux configurations ordonnées $\sigma \leq \sigma'$. Comme les applications τ_0^x , τ_{+1}^x , τ_{-1}^x ne modifient pas la configuration en dehors du site x il suffit de vérifier que l'ordre est conservé sur le site x . Concernant τ_0^x la propriété est évidente puisque

$$\tau_0^x(\sigma)(x) = \tau_0^x(\sigma')(x) = 0.$$

Concernant τ_{-1}^x la propriété découle des remarques suivantes : soit $\tilde{\sigma}$ définie à partir de σ comme dans la première section. Si $\tilde{\sigma}$ est autorisée alors $\tau_{-1}^x(\sigma)(x) = -1 \leq \tau_{-1}^x(\sigma')(x)$. Si $\tilde{\sigma}$ n'est pas autorisée alors il faut remarquer que la configuration $\tilde{\sigma}'$ construite à partir de σ' ne sera

pas non plus autorisée. Ainsi, dans ce cas $\tau_{-1}^x(\sigma) = \sigma$ et $\tau_{-1}^x(\sigma') = \sigma'$. Le même raisonnement s'applique à τ_{+1}^x . \square

On note alors $-\mathbf{1}$ (resp. $\mathbf{1}$) la configuration dont tous les sites sont occupés par des composés A (resp. par des composés B). Les configurations $-\mathbf{1}$ et $\mathbf{1}$ sont respectivement le minimum et le maximum de \mathcal{A} . On déduit de la croissance des fonctions F_u^x que pour tout $\sigma \in \mathcal{A}$

$$Y_n(-\mathbf{1}) \leq Y_n(\sigma) \leq Y_n(\mathbf{1}).$$

Lemme 1. (i) Le temps T est aussi le premier temps où les valeurs de $Y_n(-\mathbf{1})$ et $Y_n(\mathbf{1})$ coïncident :

$$T = \inf\{n \geq 0, Y_n(-\mathbf{1}) = Y_n(\mathbf{1})\}.$$

(ii) Le temps T est fini, \mathbb{P} presque sûrement.

Démonstration. (Esquisse) : (i) est une conséquence immédiate de la dernière inégalité.

(ii) On regarde les intervalles de temps $[kN^2 + 1, (k+1)N^2]$. Il est clair que si sur cet intervalle de temps les variables X_n parcourent tous les N^2 sites de Λ et si U_n reste dans $[-\frac{1}{1+2q}, \frac{1}{1+2q}]$ alors

$$F_{U_{kN^2+1}}^{X_{kN^2+1}} \circ \dots \circ F_{U_{(k+1)N^2}}^{X_{(k+1)N^2}}$$

est la fonction constante égale à la configuration ne contenant aucun composé. Les configurations $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ coïncident donc pour $n \geq (k+1)N^2$. Ainsi, on en déduit que la probabilité que T soit plus grand que kN^2 est plus petite que $(1 - (N^2)! (\frac{1}{N^2} \frac{1}{1+2q})^{N^2})^{k-1}$. Bien que très grossière, cette estimation permet de conclure le lemme en faisant tendre k vers l'infini. \square

On a alors le résultat fondamental suivant

Théorème 2. La valeur de Y_T est distribuée suivant μ , c'est-à-dire que pour tout σ dans \mathcal{A}

$$\mathbb{P}(Y_T \text{ est la fonction constante égale à } \sigma) = \mu(\{\sigma\}).$$

Démonstration. Considérons une variable aléatoire Γ_0 à valeur dans \mathcal{A} et de loi μ . Le théorème 1 nous dit que si X est une variable aléatoire uniforme sur Λ et U une variable aléatoire uniforme sur $[-1, 1]$, alors $F_U^X(\Gamma_0)$ est encore de loi μ . On en déduit que $\Gamma_n = Y_n(\Gamma_0)$ est de loi μ . Mais, pour $k \geq 0$ on sait que Γ_{T+k} est égale à la valeur de la fonction constante Y_T , on en déduit que

$$\mathbb{P}(Y_T \text{ est la fonction constante égale à } \sigma) = \lim_{n \rightarrow \infty} \mathbb{P}(\Gamma_n = \sigma) = \mu(\{\sigma\}).$$

\square

3. Mise en place de l'algorithme

On va déduire des résultats de la section 2 un algorithme assez efficace qui permet de faire un tirage exact sous la probabilité μ . Commençons par quelques remarques : grâce au lemme 1, on voit qu'il suffit de calculer la valeur de Y_n sur les configurations $\mathbf{1}$ et $-\mathbf{1}$ pour trouver le temps de coalescence T . Le problème est que pour calculer la valeur de $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ par récurrence il faudrait avoir calculé et gardé en mémoire la valeur de Y_{n-1} sur toutes les configurations σ . On utilisera donc plutôt la formule complète (1) pour calculer $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$. On se heurte alors

à un second problème : calculer pour chaque n la valeur de $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ par la formule (1) nécessite beaucoup de calculs. Une solution efficace est de ne pas calculer $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ pour tous les entiers n , mais seulement pour des n très espacés : par exemple nous pouvons calculer $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ seulement pour les n du type

$$n = 2^k, \quad k \in \mathbb{N}.$$

En effet, on sait que après le temps de coalescence T les suites $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ restent constantes et égales à la valeur commune de $Y_T(\mathbf{1})$ et $Y_T(-\mathbf{1})$. Nous n'avons donc pas besoin de connaître exactement le temps T , mais simplement de trouver un temps pour lequel les valeurs de $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ coïncident. On en déduit donc l'algorithme suivant.

0] Faire le tirage de X_1 et U_1 . $k \leftarrow 1$.

1] Faire le tirage de $X_{2^{k-1}+1}, \dots, X_{2^k}$ et $U_{2^{k-1}+1}, \dots, U_{2^k}$.

2] Calculer $Y_{2^k}(\mathbf{1})$ et $Y_{2^k}(-\mathbf{1})$ par la formule (1).

3] Si $Y_{2^k}(\mathbf{1}) = Y_{2^k}(-\mathbf{1})$ alors Résultat = $Y_{2^k}(\mathbf{1})$. FIN

Sinon $k \leftarrow k + 1$ et aller en 1].

Remarque : On prendra garde à ne pas retirer $X_1, \dots, X_{2^{k-1}}$ et $U_1, \dots, U_{2^{k-1}}$ à chaque nouvelle boucle.

Suggestions pour le développement

- *Soulignons qu'il s'agit d'un menu à la carte et que vous pouvez choisir d'étudier certains points, pas tous, pas nécessairement dans l'ordre, et de façon plus ou moins fouillée. Vous pouvez aussi vous poser d'autres questions que celles indiquées plus bas. Il est très vivement souhaité que vos investigations comportent une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats.*
 - Justifier la méthode du rejet décrite dans l'introduction.
 - Détailler la preuve du théorème 1, section 1 (on pourra en particulier expliquer pourquoi la réversibilité de μ implique qu'elle est mesure invariante).
 - Détailler les preuves de la proposition 1, du lemme 1 et du théorème 1, section 2.
 - En utilisant les arguments de la preuve du lemme 1, montrer que T est d'espérance finie.
 - En notant d_n le nombre de sites sur lesquels les configurations $Y_n(\mathbf{1})$ et $Y_n(-\mathbf{1})$ diffèrent, montrer que d_n est décroissant.
 - Simuler la chaîne de Markov Σ_n .
 - Programmer l'algorithme décrit dans la partie 3 (on commencera par des N très petits, et on ne dépassera pas $N = 20$).
 - Quelle est la complexité de l'algorithme décrit dans la partie 3, en fonction du temps de coalescence T .