# Convergence de l'échantillonneur de Gibbs pour les distributions uniformes

**Jeffrey S. Rosenthal**

**Department of Statistics**

**University of Toronto**

http://markov.utstat.toronto.edu/jeff/

jeff@math.toronto.edu

Basé sur trois articles, tous avec **G.O. Roberts** :

*Convergence of slice sampler Markov chains.* (JRSSB 61:643–660, 1999.)

*On convergence rates of Gibbs samplers for uniform distributions.* (Ann. Appl. Prob. 8:1291–1302, 1998.)

*The polar slice sampler.* Soumis.

**Chaîne de Markov Monte Carlo :**

On s'intéresse à une distribution (compliquée) $\pi(\cdot)$ sur un espace (de grande dimension) $\mathcal{X}$ (par ex. $\mathcal{X} \subseteq \mathbf{R}^d$).

On crée une chaîne de Markov $P(x, \cdot)$ sur $\mathcal{X}$, qui a la proprieté que $\pi P = \pi$, c.à.d. $\pi$ est une distribution stationnaire pour la chaîne.

**Pour l'éxecution :**

Prendre $X^{(0)}$ d'une distribution initiale $\mu_0$, puis prendre $X^{(k)}$ de $P(X^{(k-1)}, \cdot)$, pour $k = 1, 2, 3, \ldots$.

Si $k$ est assez grand, nous espérons que $X^{(k)}$ est à peu près un échantillon de $\pi(\cdot)$, c.à.d. $\mathcal{L}(X^{(k)}) \approx \pi(\cdot)$.

**Questions :**

Comment choisir $P(x, \cdot)$?

Quelle est la vitesse de convergence? Est-ce qu'on peut trouver des bornes sur

$$\|\mu_k - \pi\|_{\text{var}} \equiv \sup_{A \subseteq \mathcal{X}} \left| \mathbf{P}(X^{(k)} \in A) - \pi(A) \right| ?$$

**L'Échantillonneur de Gibbs pour les distributions uniformes :**

Soit $R \subseteq \mathbf{R}^d$, et supposons qu'on s'intéresse a $\mathbf{Unif}(R)$, c.à.d. la distribution uniforme dans la région $R$.

<u>L'échantillonneur de Gibbs</u> (à balayage aléatoire) marche ainsi : Étant donné $X^{(k)}$, choisir d'abord $I_k$ uniforme dans $\{1, 2, \ldots, d\}$, puis remplacer $X^{(k)}$ par

$$X^{(k+1)} \sim \mathbf{Unif}\{x \in R;\ x_j = X_j \text{ pour } j \neq I_k\}\,,$$

c.à.d. choisir $X^{(k+1)}$ uniformément du segment de droite dans $R$ qui touche $X^{(k)}$, et qui est parallèlle à l'axe de la $I_k$-ième coordonnée.

Distribution stationnaire de cet échantillonneur : $\mathbf{Unif}(R)$.

(C'est clair, parce que c'est réversible relatif à $\mathbf{Unif}(R)$ : $\mathbf{Unif}(R)(dx)\, P(x, dy) = \mathbf{Unif}(R)(dy)\, P(y, dx)\ \forall\, x, y$.)

<u>Question spécifique :</u> Quelle est la vitesse de convergence à $\mathbf{Unif}(R)$ de cet échantillonneur?

**Motivation et cas spécial :**

L'Échantillonneur par tranches ($\ll$ Slice Sampler $\gg$)

[*Développé par Edwards/Sokal, Besag/Green, Higdon, Damien/Walker, Mira/Tierney, Neal, etc. Lié aussi à l'algorithme de Swendsen/Wang.*]

Supposons que $\nu$ est une distribution sur $\mathbf{R}^d$, avec une densité proportionnelle à

$$f : \mathbf{R}^d \to \mathbf{R}^{\geq 0}\,.$$

Observons que si $(X, Y)$ est uniforme dans

$$R \;=\; \{(x,y);\; x \in \mathbf{R}^d,\; 0 \leq y \leq f(x)\}$$

(la région sous le graphe de $f$), alors $X \sim \nu$.

(Ici, $Y$ s'appelle une $\ll$ variable auxiliaire $\gg$.)

Alors, ≪ l'échantillonnage par tranches ≫ est une forme d'échantillonnage de Gibbs pour $(X, Y)$ pour la distribution $\mathbf{Unif}(R)$.

C'est-à-dire, étant donné $(X^{(k)}, Y^{(k)})$, avec probabilité $\frac{1}{2}$ :

(a) remplacer $Y^{(k)}$ avec

$$Y^{(k+1)} \sim \mathbf{Unif}[0,\ f(X^{(k)})]\,,$$

ou bien (b) remplacer $X^{(k)}$ avec

$$X^{(k+1)} \sim \mathbf{Unif}\{x \in \mathbf{R}^d;\ f(x) \geq Y^{(k)}\}\,,$$

sans changer l'autre variable.

Nous notons qu'en grande dimension, l'étape (b) peut être difficile; mais pour l'instant nous ne nous en préoccupons pas.

**Convergence géometrique :**

Rappellons qu'une chaîne de Markov, avec une distribution stationnaire $\pi(\cdot)$, est géométriquement ergodique s'il y a $\rho < 1$ avec

$$\|P^k(x,\cdot) - \pi(\cdot)\| \ \leq \ M(x)\, \rho^k\,, \qquad k = 1, 2, \ldots .$$

(La convergence est uniforme si $M(x) \equiv M$.)

**Proposition 1.** Soit $R \subseteq \mathbf{R}^2$ le ≪ triangle tranchant ≫

$$R \ = \ \Big\{(x,y) \in \mathbf{R}^2;\ 0 < x < 1,$$

$$x \tan\theta < y < (x + x^\alpha)\tan\theta\Big\}\,,$$

où $0 < \theta < \pi/2$ et $1 < \alpha < \infty$. L'échantillonneur de Gibbs pour $\mathbf{Unif}(R)$ n'est pas géométriquement ergodique, c.à.d. qu'il converge lentement!

[D'autres exemples, avec vitesse de convergence aussi lente que desirée, sont presentés par C. Bélisle, RCS 1998.]

Par contraste, nous avons

**Proposition 2.** Soit $R \subseteq \mathbf{R}^2$ le vrai triangle

$$R = \Big\{(x,y) \in \mathbf{R}^2;\ 0 < x < 1, \\ x \tan\theta < y < x \tan\phi\Big\},$$

où $0 < \theta < \phi < \pi/2$. L'échantillonneur de Gibbs pour $\mathbf{Unif}(R)$ est géométriquement ergodique. (Mais ce n'est pas uniformément ergodique.)

Ces deux résultats suggèrent que l'aspect lisse de la frontière de la région $R$ contrôle le taux de convergence.

Si la frontière de $R$ est bien lisse, nous avons :

**Théorème 3.** Soit $R$ une région ouverte, connexe, et bornée dans $\mathbf{R}^d$, avec frontière qui est une variété (de $d-1$ dimensions) qui est deux fois différentiable. Alors l'échantillonneur de Gibbs pour $\mathbf{Unif}(R)$ est uniformément géométriquement ergodique.

Ensuite, soit

$$R_\epsilon \;=\; \{\mathbf{x} \in \mathbf{R}^d\,;\, B(\mathbf{x},\epsilon) \subseteq R\}$$

$\ll$ l'intérieur $\epsilon$ de $R \gg$, où $B(\mathbf{x},\epsilon)$ est une boule $L^\infty$ (ou $L^2$ ou $L^1$). Alors nous avons

**Théorème 4.** Soit $R$ une région ouverte, connexe, et bornée dans $\mathbf{R}^d$. Supposons qu'il y a $m \in \mathbf{N}$, $\epsilon > 0$ et $\delta > 0$ pour que

$$P^m(\mathbf{x}, R_\epsilon) \;\geq\; \delta\,, \qquad \mathbf{x} \in R$$

(où $P$ est l'échantilloneur de Gibbs pour $\mathbf{Unif}(R)$). Alors $P$ est uniformément géométriquement ergodique.

Ce théorème dit que l'échantillonneur de Gibbs est géométriquement ergodique s'il n'est jamais $\ll$ attrapé $\gg$ près de la frontière de $R$.

**Retournons à l'échantillonneur par tranches :**

Il est bien connu (Mira/Tierney) que, si $f$ est bornée, et le support de $f$ est aussi borné, alors l'échantillonneur par tranches pour $f$ est uniformément géométrique.

C'est bon! Mais, la condition que le support de $f$ soit borné est très forte. Comment l'enlever?

Soit

$$Q(y) \;=\; \lambda_d\{x \in \mathbf{R}^d;\; f(x) \geq y\}\,,$$

où $\lambda_d$ est la mesure de Lebesgue sur $\mathbf{R}^d$.

[En effet, la fonction $Q$ contient toute l'information sur la convergence de l'échantillonneur par tranches pour $f$ : Si $f$ et $\widehat{f}$ sont deux densités différentes (peut-être même en dimensions différentes), mais leurs échantillonneurs par tranches $\{X_n\}$ et $\{\widehat{X}_n\}$ correspondent à la même fonction $Q$, et $f(X_0) = \widehat{f}(\widehat{X_0})$, alors

$$\|\mathcal{L}(X_n) - \pi(\cdot)\| \;=\; \|\mathcal{L}(\widehat{X}_n) - \widehat{\pi}(\cdot)\|\,, \qquad n \in \mathbf{N}\,;$$

voir G.O. Roberts et J.S. Rosenthal (1998), *Markov chains and de-initialising processes*, à paraître dans *Scandinavian Journal of Statistics*.]

**Théorème 5.** Supposons que $f$ est bornée, et qu'il y a $\alpha > 1$ avec $Q'(y)y^{1+\frac{1}{\alpha}}$ non croissant pour $y$ suffisamment petit. (En dimension $d = 1$, ça veut dire à peu près que les ailes (queues?) de $f$ sont aussi légères que $x^{-\alpha}$.) Alors, l'échantillonneur par tranches pour $f$ est géométriquement ergodique.

Ce théorème nécessite une condition pour les ailes de $f$ qui est beaucoup moins forte que ≪ support borné ≫.

(Mais, le théorème perd l'uniformité de la convergence.)

Aussi, la condition que $f$ elle-même soit bornée n'est pas essentielle; il suffit que $f^{-1}$ aussi satisfasse une condition comme $Q'(y)y^{1+\frac{1}{\alpha}}$ non croissant.

**Des Bornes quantitatives :**

**Théorème 6.** Supposons que $f$ est bornée, et que $Q'(y)y$ est non croissant pour toutes valeurs de $y$. (C'est la limite $\alpha \to \infty$ de l'autre condition. En dimension $d = 1$, ça veut dire à peu près que les ailes de $f$ sont aussi légères que $e^{-x}$.) Si $[f(X^{(0)})/\sup_y f(y)] \geq 0.0025$, alors

$$\|\mu_k - \pi\|_{\text{var}} \;\leq\; 0.0324\,(0.9897)^k(k{-}7.79)\,, \quad k \geq 23\,.$$

Par exemple, $\|\mu_k - \pi\|_{\text{var}} < 0.01$ si $k \geq 530$. Alors, on peut dire que 530 itérations ≪ suffisent ≫ pour la convergence de l'échantillonnage par tranches, si $Q'(y)y$ est non croissant.

Bien utile! Et, applicable à beaucoup de fonctions $f$, y compris certaines fonctions en grande dimension.

[Ce résultat utilise des bornes générales d'autres auteurs, y compris Meyn/Tweedie (Ann. Appl. Prob., 1994); Rosenthal (JASA, 1995); Lund/Tweedie (Math. Op. Res., 1996); et Roberts/Tweedie (Stoch. Proc. Appl., 1999; J. Appl. Prob., 2000). La preuve utilise des conditions de minorisation et des conditions de tendence (≪ drift ≫).]

**LÉchantillonneur par tranches polaire :**

(≪ Polar Slice Sampler ≫)

Pour utiliser le Théorème 6, il faut vérifier que $Q'(y)y$ est non croissant. Que veut dire cette condition?

Si $d = 1$, il suffit que $\log f$ soit concave, une condition qui n'est pas rare. Mais si $d > 1$, ça ne suffit pas. C'est dommage!

Alors, nous *changeons* l'algorithme ainsi :

Étant donné $(X^{(k)}, Y^{(k)})$, avec probabilité $\frac{1}{2}$

(a) remplacer $Y^{(k)}$ par

$$Y^{(k+1)} \sim \mathbf{Unif}[0,\ f(X^{(k)})\ |X^{(k)}|^{d-1}],$$

ou bien (b) remplacer $X^{(k)}$ par

$$X^{(k+1)} \sim |x|^{-(d-1)} \mathbf{1}_{\{f(x)\,|x|^{d-1} \geq Y^{(k)}\}} \lambda_d(dx)\,.$$

sans changer l'autre variable.

(Ça correspond à une autre ≪ factorisation ≫ de $f$.)

L'échantillonneur par tranches polaire change $Q(y)$ à

$$Q_{\text{polaire}}(y) \;\equiv\; \int_{\{f(x)\,|x|^{d-1} \geq y\}} |x|^{-(d-1)}\,\lambda_d(dx)\,.$$

On a

**Théorème 7.** Si $\log f$ est concave (en n'importe quelle dimension), alors $Q'_{\text{polaire}}(y)y$ est non croissant. Alors, l'échantillonneur par tranches polaire pour $f$ converge (avec $\|\mu_k - \pi\|_{\text{var}} \leq 0.01$) en $k \leq 530$ itérations.

[Il y a aussi une condition technique, liée à la $\ll$ symmétrie sphéricale $\gg$ de $f$ ...]

**Résumé :**

- Nous avons étudié l'échantillonneur de Gibbs pour les distributions uniformes dans les régions $R$.
- Cet échantillonneur est parfois géométriquement ergodique, mais parfois pas; ça dépend de l'aspect lisse de la frontière de $R$. Si la frontière de $R$ est deux fois différentiable, alors l'échantillonneur est uniformément géométriquement ergodique.
- L'échantillonneur par tranches est un cas spécial de l'échantillonneur de Gibbs dans la région $R$ sous le graphe de la densité $f$ d'une distribution.
- Si les ailes de $f$ sont polynômiales (à peu près), alors l'échantillonneur par tranches est géométriquement ergodique.
- Si les ailes de $f$ sont exponentielles (à peu près) en dimension 1, alors l'échantillonneur par tranches converge en 530 itérations.
- Si les ailes de $f$ sont exponentielles (à peu près) en *n'importe quelle* dimension, alors l'échantillonneur par tranches *polaire* converge en 530 itérations.

**Résumé très bref :**

- Pour les régions $R \ll$ raisonnables $\gg$, et les fonctions $f \ll$ raisonnables $\gg$, l'échantillonneur de Gibbs et l'échantillonneur par tranches ont de très bonnes propriétés de convergence.

---

http://markov.utstat.toronto.edu/jeff/

jeff@math.toronto.edu