MAPSI — cours 9 : Échantillonnage et MCMC

Christophe Gonzales

LIP6 – Université Paris 6, France

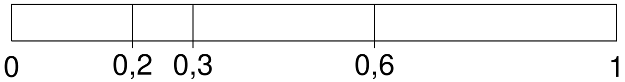
Échantillonnage avec une distribution discrète

- *Problème :* échantillonner selon :

$$\text{distribution } \overset{\infty}{\pi}(X) = \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \\ \hline \multicolumn{1}{|c|}{0{,}2} & \multicolumn{1}{c|}{0{,}1} & \multicolumn{1}{c|}{0{,}3} & \multicolumn{1}{c|}{0{,}4} \\ \hline \end{array}$$

- *Solution :*
  1. Calculer la cumulative :

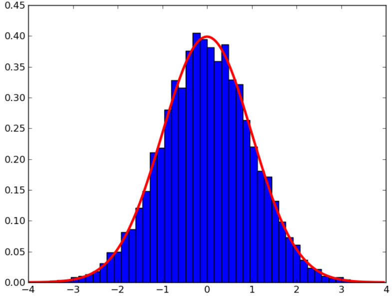
$$F(X_i) = \sum_{Y \leq X_i} \overset{\infty}{\pi}(Y) = \begin{array}{|c|c|c|c|} \hline 0{,}2 & 0{,}3 & 0{,}6 & 1 \\ \hline \end{array}$$



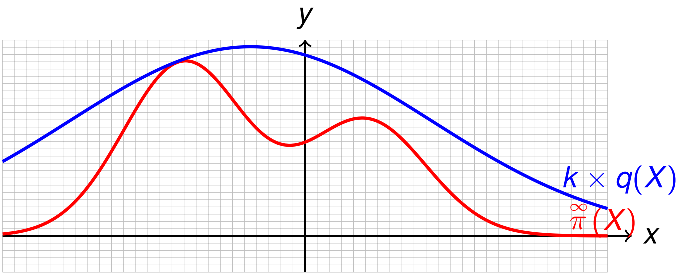
  2. Tirer un nombre z selon une distribution uniforme sur [0,1[
  3. Soit i tel que $F(X_{i-1}) \leq z < F(X_i)$ (en posant $X_0 = 0$)
  4. Renvoyer X_i

Échantillonnage d'une loi normale

- Faire la cumulative de la fonction de densité (cf. table)



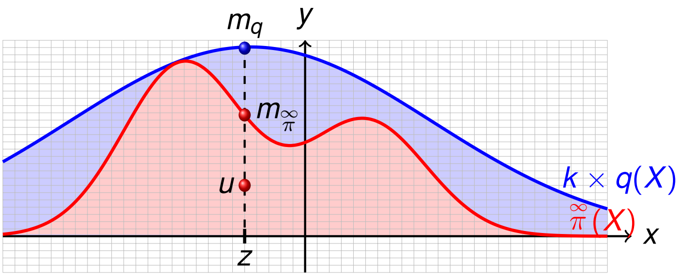
Il existe des algos dédiés performants (Ziggurat, Box-Muller)



Hypothèses :

- $\overset{\infty}{\pi}(\cdot)$ difficile à échantillonner
- Mais pour tout $x \in X$, $\overset{\infty}{\pi}(x)$ facile à calculer
- $q(\cdot)$ facile à échantillonner
- il existe $k \in \mathbb{R}$ tel que $\overset{\infty}{\pi}(x) \leq k \times q(x)$ pour tout $x \in X$

Distributions complexes : Rejection Sampling



Alogrithme « rejection sampling » :

1. Tirer un nombre z selon $q(\cdot)$
2. Calculer $m_q = k \times q(z)$
3. Tirer un nombre u selon une loi uniforme sur $[0, m_q[$
3. Accepter z si $u \leq \overset{\infty}{\pi}(z) = m_{\overset{\infty}{\pi}}$

Avantage : fonction de partition inconnue

- $\overset{\infty}{\pi}(x) = \dfrac{1}{Z_p} p(x)$
- Seul $p(x)$ connu

- Nouvelle règle : $k \times q(x) \geq p(x)$ pour tout x
- Rejection sampling $\Longrightarrow$ échantillon $\langle z_1, \ldots, z_n \rangle \sim \hat{\overset{\infty}{\pi}}(\cdot)$
- $\hat{\overset{\infty}{\pi}}(z) \propto q(z) \times \dfrac{p(z)}{k \times q(z)}$
- $\hat{\overset{\infty}{\pi}}(z) \propto \dfrac{p(z)}{k} \propto p(z) \propto \overset{\infty}{\pi}(z)$

$\Longrightarrow$ on peut échantillonner sans connaître la fonction de partition

Distributions complexes : Rejection Sampling

Calcul du taux d'acceptation :

$$
\begin{aligned}
P(\text{acceptation}) &= \int q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz \\
&= \int q(z) \times \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} dz \\
&= \frac{1}{k} \int \overset{\infty}{\pi}(z) dz = \frac{1}{k}
\end{aligned}
$$

Exemple précédent : $k = 1,96 \Longrightarrow$ seulement 1 z sur 2 accepté !

k augmente exponentiellement avec la dimension de $\overset{\infty}{\pi}(\cdot)$!

Principe de l'échantillonnage par MCMC

MCMC : Markov Chain Monte Carlo

- But : échantillonner selon une loi $\overset{\infty}{\pi}(\cdot)$
- Principe : construire une suite (X_i) de variables aléatoires tirées selon des lois $\overset{\infty}{\pi}_i(\cdot)$ tendant vers $\overset{\infty}{\pi}(\cdot)$

  et sélectionner un échantillon $\langle x_i, \dots, x_{m+i} \rangle$
  ou sous-échantillonner : $\langle x_{\sigma(i)}, \dots, x_{\sigma(m+i)} \rangle \Longrightarrow \approx$ i.i.d.
- Solution : construire une chaîne de Markov de loi stationnaire $\overset{\infty}{\pi}(\cdot)$

Loi stationnaire

- soit $P(X_{t+1}|X_t)$ la probabilité de transition (chaîne homogène)

Loi stationnaire $\overset{\infty}{\pi}(\cdot)$

$$\overset{\infty}{\pi}(x) = \int_y P(x|y)\,\overset{\infty}{\pi}(y)dy$$

ici, on connaît $\overset{\infty}{\pi}(\cdot)$ et on cherche $P(\cdot|\cdot)$

Problème : sous quelles conditions $P(\cdot|\cdot)$ existe-t-elle ?

Ergodicité ?

Une condition possible

Réversibilité

$$\overset{\infty}{\pi}(x)P(y|x) = \overset{\infty}{\pi}(y)P(x|y),\ \forall x, y$$

propriété également connue sous le nom de « detailed balance »

conséquence :

$$\begin{aligned}
\int_y P(x|y)\,\overset{\infty}{\pi}(y)dy &= \int_y P(y|x)\,\overset{\infty}{\pi}(x)dy \\
&= \overset{\infty}{\pi}(x)\int_y P(y|x)dy \\
&= \overset{\infty}{\pi}(x)
\end{aligned}$$

$\Longrightarrow \overset{\infty}{\pi}(\cdot)$ loi stationnaire !

Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

Intérprétation de $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$

Le processus markovien va évoluer plus souvent de x vers y que de y vers $x \Longrightarrow$ non réversible.

Correction : diminuer $P(y|x)$ ou augmenter $P(x|y)$

$\Longrightarrow$ créer deux nombres $\alpha(x,y)$ et $\alpha(y,x)$ tels que :

$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y,x)$$

on veut que $P(y|x)\alpha(x,y)$ soit une probabilité de transition !

Remarque : $y = x \Longrightarrow \overset{\infty}{\pi}(x)P(x|x)\alpha(x,x) = \overset{\infty}{\pi}(x)P(x|x)\alpha(x,x)$
pour tout $\alpha(x,x)$

Si $P(x|x)\alpha(x,x) = 1 - \displaystyle\int_{y \neq x} P(y|x)\alpha(x,y)dy$, on a bien une proba !

Garantir la réversibilité

$$P(x|x)\alpha(x,x) = 1 - \int_{y\neq x} P(y|x)\alpha(x,y)dy$$

Pour assurer que $P(x|x)\alpha(x,x) \geq 0$, on impose $\alpha(x,y) \leq 1$

$$\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$$
$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y,x)$$

$\Longrightarrow$ pour augmenter $P(x|y)$, on fixe $\alpha(y,x) = 1$

$$\Longrightarrow \overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)$$

$$\Longrightarrow \alpha(x,y) = \frac{\overset{\infty}{\pi}(y)P(x|y)}{\overset{\infty}{\pi}(x)P(y|x)}$$

Garantir la réversibilité

Résumé :

- Si $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$:

  $$\text{Fixer } \alpha(x,y) = \frac{\overset{\infty}{\pi}(y)P(x|y)}{\overset{\infty}{\pi}(x)P(y|x)} \text{ et } \alpha(y,x) = 1$$

- Par symmétrie, si $\overset{\infty}{\pi}(x)P(y|x) < \overset{\infty}{\pi}(y)P(x|y)$:

  $$\text{Fixer } \alpha(x,y) = 1 \text{ et } \alpha(y,x) = \frac{\overset{\infty}{\pi}(y)P(x|y)}{\overset{\infty}{\pi}(x)P(y|x)}$$

Garantir la réversibilité

Intérprétation de α : probabilité de mouvement

$\alpha(x, y) =$ la probabilité de réaliser la transition de x vers y

$\Longrightarrow$ à l'étape t, on a 2 choix :

- transiter de x vers un y avec la probabilité $P(y|x)\alpha(x, y)$
- ne pas réaliser de transition

Résumé

si $\alpha(x, y) = \min\left\{1, \dfrac{\overset{\infty}{\pi}(y)P(x|y)}{\overset{\infty}{\pi}(x)P(y|x)}\right\}$ alors :

$\overset{\infty}{\pi}(x)P(y|x)\alpha(x, y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y, x)$

$\Longrightarrow$ réversibilité $\Longrightarrow \overset{\infty}{\pi}(\cdot)$ distribution stationnaire

Algorithme de Metropolis-Hastings

Metropolis-Hastings

Algorithme pour générer x_{t+1} à partir de x_t :

1. tirer z selon la distribution $P(\cdot|x_t)$
2. calculer $\alpha(x_t, z) = \min\left\{1, \dfrac{\overset{\infty}{\pi}(z)P(x_t|z)}{\overset{\infty}{\pi}(x_t)P(z|x_t)}\right\}$
3. tirer un nombre u selon une loi uniforme sur $[0, 1[$
4. renvoyer $x_{t+1} = \begin{cases} z & \text{si } u \leq \alpha(x_t, z) \\ x_t & \text{sinon} \end{cases}$

Références :

- N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller et E. Teller (1953) "Equations of State Calculations by Fast Computing Machines". Journal of Chemical Physics, 21 (6), pp. 1087–1092
- W.K. Hastings (1970) "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". Biometrika, 57 (1), pp. 97–109

Choix de $P(\cdot|x_t)$

- $P(\cdot|x_t)$ doit être simple à échantillonner
- 1ère possibilité [Metropolis *et al.* (1953), Müller (1993)]

  $P(z|x_t) = q(z - x_t)$ avec $q(\cdot)$ densité multivariée

  autrement dit $z = x_t + y$ avec $y \sim q(\cdot)$

  $q(\cdot)$ indépendante de x_t !

$\Longrightarrow$ random walk chain

- choix possible de $q(\cdot)$: loi normale
- si q est symmétrique : $q(y) = q(-y)$ et

$$\alpha(x_t, z) = \min\left\{1, \frac{\overset{\infty}{\pi}(z)P(x_t|z)}{\overset{\infty}{\pi}(x_t)P(z|x_t)}\right\} = \min\left\{1, \frac{\overset{\infty}{\pi}(z)}{\overset{\infty}{\pi}(x_t)}\right\}$$

Choix de $P(\cdot|x_t)$

- 2ème possibilité [Hastings (1970)]

  $P(z|x) = q(z)$ avec $q(\cdot)$ densité multivariée

$\Longrightarrow$ independent chain

$\Longrightarrow$ généralisation de rejection sampling

- 3ème possibilité : l'algorithme Langevin [Roberts et Rosenthal (1998)]

  $$z = x_t + \frac{\sigma^2}{2} \nabla \log(\overset{\infty}{\pi}(x_t)) + \sigma y \text{ avec } y \sim q(\cdot)$$

  σ : facteur d'échelle

⚠ Il existe plein d'autres possibilités...

Choix de l'étalement/variance de $P(\cdot|x_t)$

⚠ important pour la vitesse de convergence

Influence de l'étalement

- Taux d'acceptation
- Région couverte par la chaîne de Markov

Choix de l'étalement/variance de $P(\cdot|x_t)$

Roberts, Gelman, Gilks (1994)

- cadre : random walk
- $\overset{\infty}{\pi}$ et $P(\cdot|\cdot)$: lois normales mono-dimensionnelles

  affiner l'étalement de $P(\cdot|x_t)$ pour obtenir un taux d'acceptation $\approx 0,45$
- $\overset{\infty}{\pi}$ et $P(\cdot|\cdot)$: lois normales n-dimensionnelles

  affiner l'étalement de $P(\cdot|x_t)$ pour obtenir un taux d'acceptation $\approx 0,23$ lorsque n tend vers $+\infty$

Müller (1993)

Random walk $\Longrightarrow$ taux d'acceptation $\approx$ 0,5.

Initialisation et burn in

Initialisation :

Partir de n'importe quelle valeur x_0

au départ l'échantillon ne suit pas $\overset{\infty}{\pi}(\cdot)$

$\Longrightarrow$ burn in nécessaire :

Ne conserver dans l'échantillon que les x_t pour $t > t_0$

En général, t_0 est de l'ordre de quelques milliers

Metropolis-Hastings par bloc

- supposons que $x_t = (x_t^1, x_t^2)$

Précédemment :

- stationnarité : $\overset{\infty}{\pi}(x_{t+1}) = \int_{x_t} P(x_{t+1}|x_t) \overset{\infty}{\pi}(x_t) dx_t$

Maintenant :

- $P(x_{t+1}|x_t) = P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2)$
- stationnarité :

$$\overset{\infty}{\pi}(x_{t+1}^1, x_{t+1}^2) = \int_{x_t^1} \int_{x_t^2} P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2) \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

- Or $P(x_{t+1}|x_t) = P(x_{t+1}^2 | x_{t+1}^1, x_t^1, x_t^2) \times P(x_{t+1}^1 | x_t^1, x_t^2)$

$$= P(x_{t+1}^2 | x_{t+1}^1, x_t^2) \times P(x_{t+1}^1 | x_t^1, x_t^2) \quad \text{(prop. Markov)}$$

Metropolis-Hastings par bloc : stationnarité

$$\overset{\infty}{\pi}(x_{t+1}^2, x_{t+1}^2) = \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

Rappel : stationnarité pour 1 variable

$$\overset{\infty}{\pi}(x_{t+1}) = \int_{x_t} P(x_{t+1}|x_t)\ \overset{\infty}{\pi}(x_t) dx_t$$

Stationnarité par bloc

Généralisation en rajoutant toutes les variables sauf celle en x_{t+1}^i à droite des signes de conditionnement :

- $\overset{\infty}{\pi}(x_{t+1}^1|y^2) = \int_{x_t^1} P(x_{t+1}^1|x_t^1, y^2)\ \overset{\infty}{\pi}(x_t^1|y^2) dx_t^1$ pour tout y^2
- $\overset{\infty}{\pi}(x_{t+1}^2|y^1) = \int_{x_t^2} P(x_{t+1}^2|x_t^2, y^1)\ \overset{\infty}{\pi}(x_t^2|y^1) dx_t^2$ pour tout y^1

conséquences de la stationnarité par bloc :

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\, \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\, \overset{\infty}{\pi}(x_t^1|x_t^2)\, \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2}\int_{x_t^1} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\, \overset{\infty}{\pi}(x_t^1|x_t^2)\, \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \left[\int_{x_t^1} P(x_{t+1}^1|x_t^1, x_t^2)\, \overset{\infty}{\pi}(x_t^1|x_t^2) dx_t^1\right] \overset{\infty}{\pi}(x_t^2) dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\, \overset{\infty}{\pi}(x_{t+1}^1|x_t^2)\, \overset{\infty}{\pi}(x_t^2) dx_t^2 \qquad \text{(stationnarité par bloc)}$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\, \overset{\infty}{\pi}(x_t^2|x_{t+1}^1)\, \overset{\infty}{\pi}(x_{t+1}^1) dx_t^2 \qquad \text{(formule de Bayes)}$$

$$= \overset{\infty}{\pi}(x_{t+1}^2|x_{t+1}^1)\, \overset{\infty}{\pi}(x_{t+1}^1) = \overset{\infty}{\pi}(x_{t+1}^1, x_{t+1}^2) \qquad \text{(stationnarité par bloc)}$$

Metropolis-Hastings par bloc

Conclusion du transparent précedent

Stationnarité par bloc $\Longrightarrow$ Stationnarité de la loi jointe

Metropolis-Hastings par bloc

Algorithme pour générer $x_{t+1} = (x_{t+1}^1, \ldots, x_{t+1}^n)$ à partir de x_t :

1. choisir une permutation $\sigma : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$
2. pour tout $i \in \{1, \ldots, n\}$ faire :
   - a. Posons $y = (x_{t+1}^{\sigma(1)}, \ldots, x_{t+1}^{\sigma(i-1)}, x_t^{\sigma(i+1)}, \ldots, x_t^{\sigma(n)})$
   - b. tirer $z^{\sigma(i)}$ selon la distribution $P(\cdot | x_t^{\sigma(i)}, y)$
   - c. calculer $\alpha(x_t^{\sigma(i)}, z^{\sigma(i)} | y) = \min \left\{ 1, \dfrac{\overset{\infty}{\pi}(z^{\sigma(i)} | y) P(x_t^{\sigma(i)} | z^{\sigma(i)}, y)}{\overset{\infty}{\pi}(x_t^{\sigma(i)} | y) P(z^{\sigma(i)} | x_t^{\sigma(i)}, y)} \right\}$
   - d. tirer un nombre u selon une loi uniforme sur $[0, 1[$
   - e. $x_{t+1}^{\sigma(i)} = \begin{cases} z^{\sigma(i)} & \text{si } u \leq \alpha(x_t^{\sigma(i)}, z^{\sigma(i)} | y) \\ x_t^{\sigma(i)} & \text{sinon} \end{cases}$

Échantillonneur de Gibbs

- Metropolis-Hastings par bloc
- Choix de la proba de transition : $P(z^{\sigma(i)}|x_t^{\sigma(i)}, y) = \overset{\infty}{\pi}(z^{\sigma(i)}|y)$

Conséquence :

$$\alpha(x_t^{\sigma(i)}, z^{\sigma(i)}|y) = \min\left\{1, \frac{\overset{\infty}{\pi}(z^{\sigma(i)}|y)P(x_t^{\sigma(i)}|z^{\sigma(i)}, y)}{\overset{\infty}{\pi}(x_t^{\sigma(i)}|y)P(z^{\sigma(i)}|x_t^{\sigma(i)}, y)}\right\} = 1$$

$\Longrightarrow z^{\sigma(i)}$ est toujours accepté

Échantillonneur de Gibbs

Algorithme

Algorithme pour générer $x_{t+1} = (x_{t+1}^1, \ldots, x_{t+1}^n)$ à partir de x_t :

1. choisir une permutation $\sigma : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$
2. pour tout $i \in \{1, \ldots, n\}$ faire :
   - a. posons $y = (x_{t+1}^{\sigma(1)}, \ldots, x_{t+1}^{\sigma(i-1)}, x_t^{\sigma(i+1)}, \ldots, x_t^{\sigma(n)})$
   - b. tirer $x_{t+1}^{\sigma(i)}$ selon la distribution $\overset{\infty}{\pi}(\cdot|y)$