cours SOD333

Filtrage Bayésien et Approximation Particulaire

Méthodes de Monte Carlo

François Le Gland
INRIA Rennes et IRMAR

people.rennes.inria.fr/Francois.Le_Gland/ensta/

1er octobre 2021

Introduction
- motivation
- méthodes de Monte Carlo (rappel)

Distributions de Gibbs–Boltzmann

Mélanges finis

Approximation particulaire

Motivation

pour approximer le filtre bayésien

$$\mu_n(dx) = \mathbb{P}[X_n \in dx \mid Y_{0:n}]$$

on utilise ici la relation de récurrence preuve à la séance précédente

$$\mu_{k-1} \longrightarrow \eta_k = \mu_{k-1}\, Q_k \longrightarrow \mu_k = g_k \cdot \eta_k$$

avec la condition initiale $\mu_0 = g_0 \cdot \eta_0$

ici, la notation

$$\mu_{k-1}\, Q_k(dx') = \int_E \mu_{k-1}(dx)\, Q_k(x, dx')$$

désigne l'action du noyau markovien $Q_k(x, dx')$ sur la distribution de probabilité $\mu_{k-1}(dx)$, et la notation

$$g_k \cdot \eta_k = \frac{g_k\, \eta_k}{\langle \eta_k, g_k \rangle}$$

désigne le produit projectif de la distribution de probabilité a priori $\eta_k(dx')$ et de la fonction de vraisemblance $g_k(x')$

idée : rechercher une approximation sous la forme de distributions de probabilité empiriques (éventuellement pondérées)

$$\eta_k \approx \eta_k^N = \sum_{i=1}^N v_k^i \, \delta_{\xi_k^i} \quad \text{et} \quad \mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i}$$

associées à une population de N particules caractérisée par

- les *positions* $(\xi_k^1, \cdots, \xi_k^N)$ dans E
- et les *poids* positifs normalisés $(v_k^1, \cdots, v_k^N)$ et $(w_k^1, \cdots, w_k^N)$

initialisation : par échantillonnage pondéré

$$\mu_0 = g_0 \cdot \eta_0 \approx g_0 \cdot S^N(\eta_0) = \sum_{i=1}^{N} \frac{g_0(\xi_0^i)\ \delta_{\xi_0^i}}{\sum_{j=1}^{N} g_0(\xi_0^j)} = \sum_{i=1}^{N} w_0^i\ \delta_{\xi_0^i}$$

où les variables aléatoires $(\xi_0^1, \cdots, \xi_0^N)$ sont i.i.d. de distribution commune η_0

étape de *correction* : clairement, à partir de la définition

$$\mu_k^N = g_k \cdot \eta_k^N = \sum_{i=1}^{N} \frac{v_k^i\, g_k(\xi_k^i)\ \delta_{\xi_k^i}}{\sum_{j=1}^{N} v_k^j\, g_k(\xi_k^j)} = \sum_{i=1}^{N} w_k^i\ \delta_{\xi_k^i}$$

est automatiquement de la forme recherchée

étape de *prédiction* : à partir de la définition

$$\begin{aligned}\langle \mu^N_{k-1}\, Q_k, \phi\rangle &= \int \mu^N_{k-1}(dx) \int Q_k(x, dx')\, \phi(x') \\ &= \sum_{i=1}^N w^i_{k-1} \int Q_k(\xi^i_{k-1}, dx')\, \phi(x') \\ &= \int [\sum_{i=1}^N w^i_{k-1}\; Q_k(\xi^i_{k-1}, dx')\,]\, \phi(x')\end{aligned}$$

pour toute fonction ϕ, de sorte que

$$\mu^N_{k-1}\, Q_k(dx') = \sum_{i=1}^N w^i_{k-1}\; m^i_k(dx')$$

s'exprime comme un mélange fini, avec

$$m^i_k(dx') = Q_k(\xi^i_{k-1}, dx') \qquad \text{pour tout } i = 1 \cdots N$$

qu'il s'agit d'approximer / échantillonner, selon une méthode appropriée

question : comment approcher

- une distribution de Gibbs–Boltzmann de la forme

$$\mu = g \cdot \eta = \frac{g\,\eta}{\langle \eta, g \rangle} \quad \text{telle que} \quad \langle \mu, \phi \rangle = \frac{\langle \eta, g\,\phi \rangle}{\langle \eta, g \rangle}$$

- ou un mélange fini de la forme

$$\eta = \sum_{i=1}^{M} w_i\, m_i \quad \text{tel que} \quad \langle \eta, \phi \rangle = \sum_{i=1}^{M} w_i\, \langle m_i, \phi \rangle$$

à l'aide d'un échantillon ?

Méthodes de Monte Carlo (rappel)

s'il est difficile de *calculer* une intégrale (ou une espérance)

$$\langle \mu, \phi \rangle = \int_E \phi(x)\, \mu(dx) = \mathbb{E}[\phi(X)] \qquad \text{avec} \qquad X \sim \mu(dx)$$

mais qu'il est facile de *simuler* une v.a. selon la distribution μ, alors on peut former la distribution empirique

$$S^N(\mu) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i}$$

où $(\xi_1, \cdots, \xi_N)$ est un N–échantillon distribué selon μ, d'où l'approximation sans biais

$$\langle \mu, \phi \rangle \approx \langle S^N(\mu), \phi \rangle = \frac{1}{N} \sum_{i=1}^{N} \phi(\xi_i)$$

expression comme somme de variables aléatoires i.i.d. centrées

$$\langle S^N(\mu) - \mu, \phi \rangle = \frac{1}{N} \sum_{i=1}^{N} (\phi(\xi_i) - \langle \mu, \phi \rangle)$$

d'où l'expression de la variance (non–asymptotique)

$$\mathbb{E} | \langle S^N(\mu) - \mu, \phi \rangle |^2 = \frac{1}{N} \operatorname{var}(\phi, \mu)$$

facile, compte tenu de l'identité

$$\frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}[\, (\phi(\xi_i) - \langle \mu, \phi \rangle)\, (\phi(\xi_j) - \langle \mu, \phi \rangle)\,] = \frac{1}{N^2} \sum_{i=1}^{N} \underbrace{\mathbb{E} | \phi(\xi_i) - \langle \mu, \phi \rangle |^2}_{\operatorname{var}(\phi, \mu)}$$

plus généralement (inégalités de Marcinkiewicz–Zygmund) : pour $p \geq 2$

$$\{ \mathbb{E} | \langle S^N(\mu) - \mu, \phi \rangle |^p \}^{1/p} \leq \frac{c_p}{\sqrt{N}} \langle \mu, |\phi - \langle \mu, \phi \rangle|^p \rangle^{1/p}$$

théorèmes limites (pour une grande taille d'échantillon N)

loi (forte) des grands nombres

$$\langle S^N(\mu), \phi \rangle \longrightarrow \langle \mu, \phi \rangle$$

en probabilité (presque sûrement) quand $N \uparrow \infty$, à vitesse $1/\sqrt{N}$

théorème central limite

$$\sqrt{N}\, \langle S^N(\mu) - \mu, \phi \rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\phi(\xi_i) - \langle \mu, \phi \rangle) \Longrightarrow \mathcal{N}(0, \mathrm{var}(\phi, \mu))$$

en distribution quand $N \uparrow \infty$

Introduction

Distributions de Gibbs–Boltzmann
- distributions de Gibbs–Boltzmann
- échantillonnage pondéré
- estimations et théorèmes limites

Mélanges finis

Approximation particulaire

Distributions de Gibbs–Boltzmann

cas particulier important : distribution de Gibbs–Boltzmann

$$\mu = g \cdot \eta = \frac{g\,\eta}{\langle \eta, g \rangle} \qquad \text{c–à–d} \qquad \langle \mu, \phi \rangle = \frac{\langle \eta, g\,\phi \rangle}{\langle \eta, g \rangle} \qquad (\star)$$

avec (décomposition non unique)

- une distribution de probabilité η
- une fonction positive g

on introduit aussi la mesure positive (non–normalisée) définie par

$$\langle \gamma, \phi \rangle = \langle \eta, g\,\phi \rangle = \mathbb{E}[g(\Xi)\,\phi(\Xi)] \qquad \text{d'où} \qquad \langle \mu, \phi \rangle = \frac{\langle \eta, g\,\phi \rangle}{\langle \eta, g \rangle} = \frac{\langle \gamma, \phi \rangle}{\langle \gamma, 1 \rangle}$$

où la v.a. Ξ a pour loi η
motivation : formule de Bayes

"loi a posteriori" $\propto$ *"fonction de vraisemblance"* $\times$ *"loi a priori"*

s'il est difficile de simuler une v.a. selon μ, mais qu'il est facile de

- *simuler* une v.a. selon η
- *évaluer* pour tout x, la fonction positive $g(x)$

alors il est possible de

- générer exactement une v.a. de loi $\mu = g \cdot \eta$ [acceptation / rejet]
- approcher μ par une distribution empirique pondérée associée à un échantillon de loi η [échantillonnage pondéré]

même si la constante de normalisation $\langle \eta, g \rangle$ est inconnue

Échantillonnage pondéré

idée : approximer numérateur et dénominateur dans ($\star$) à l'aide d'un unique échantillon distribué selon η : on introduit les approximations suivantes

$$\langle \gamma, \phi \rangle = \langle \eta, g\,\phi \rangle \approx \langle S^N(\eta), g\,\phi \rangle = \frac{1}{N} \sum_{i=1}^{N} g(\xi^i)\,\phi(\xi^i)$$

d'où

$$\langle \mu, \phi \rangle = \langle g \cdot \eta, \phi \rangle \approx \langle g \cdot S^N(\eta), \phi \rangle = \frac{\sum_{i=1}^{N} g(\xi^i)\,\phi(\xi^i)}{\sum_{i=1}^{N} g(\xi^i)}$$

pour toute fonction ϕ, où les variables aléatoires $(\xi^1, \cdots, \xi^N)$ sont i.i.d. de distribution de probabilité commune η

ceci définit implicitement les approximations Monte Carlo pondérées

$$\gamma \approx \gamma^N = g\, S^N(\eta) = \frac{1}{N} \sum_{i=1}^{N} g(\xi^i)\, \delta_{\xi^i}$$

et

$$\mu \approx \mu^N = g \cdot S^N(\eta) = \sum_{i=1}^{N} \frac{g(\xi^i)\, \delta_{\xi^i}}{\sum_{j=1}^{N} g(\xi^j)} = \sum_{i=1}^{N} w^i\, \delta_{\xi^i}$$

où les poids $(w^1, \cdots, w^N)$ sont définis pour tout $i = 1 \cdots N$ par

$$w^i = \frac{g(\xi^i)}{\sum_{j=1}^{N} g(\xi^j)} \qquad \text{parfois noté} \qquad w^i \propto g(\xi^i)$$

à une constante multiplicative près

solution : changement de distribution de probabilité
soit ν une distribution de probabilité dominant η, c'est–à–dire que si $\nu(A) = 0$ alors nécessairement $\eta(A) = 0$

$$\langle \eta, g\,\phi \rangle = \langle \nu, \frac{d\eta}{d\nu}\, g\,\phi \rangle \qquad \text{et en particulier} \qquad \langle \eta, g \rangle = \langle \nu, \frac{d\eta}{d\nu}\, g \rangle$$

de sorte que

$$\langle \mu, \phi \rangle = \frac{\langle \eta, g\,\phi \rangle}{\langle \eta, g \rangle} = \frac{\langle \nu, \frac{d\eta}{d\nu}\, g\,\phi \rangle}{\langle \nu, \frac{d\eta}{d\nu}\, g \rangle} = \frac{\langle \nu, h\,\phi \rangle}{\langle \nu, h \rangle} \qquad \text{avec} \qquad h = \frac{d\eta}{d\nu}\, g$$

d'où l'expression alternative en terme de distribution de Gibbs–Boltzmann

$$\mu = h \cdot \nu = \frac{h\,\nu}{\langle \nu, h \rangle}$$

et il s'agit de proposer une paire (ν, h) présentant une incohérence moindre que celle de la paire (η, g)

Estimations et théorèmes limites

la v.a. $\langle \gamma^N, \phi \rangle$ est un estimateur non–biaisé de $\langle \gamma, \phi \rangle$

Théorème

$$\mathbb{E}| \frac{\langle \gamma^N - \gamma, \phi \rangle}{\langle \gamma, 1 \rangle} |^2 = \frac{1}{N} \frac{\mathrm{var}(g\,\phi, \eta)}{\langle \eta, g \rangle^2}$$

et pour tout $p \geq 2$

$$\{ \mathbb{E}| \frac{\langle \gamma^N - \gamma, \phi \rangle}{\langle \gamma, 1 \rangle} |^p \}^{1/p} \leq \frac{c_p}{\sqrt{N}} \left(\frac{\langle \eta, |g\,\phi - \langle \eta, g\,\phi \rangle|^p \rangle}{\langle \eta, g \rangle^p} \right)^{1/p}$$

Preuve il suffit de remarquer que

$$\gamma = g\,\eta \qquad \text{et} \qquad \gamma^N = g\,S^N(\eta)$$

de sorte que

$$\langle \gamma^N - \gamma, \phi \rangle = \langle S^N(\eta) - \eta, g\,\phi \rangle$$

et d'appliquer le résultat du cas général

en revanche, comme *rapport* de deux estimateurs non–biaisés, la v.a. $\langle \mu^N, \phi \rangle$ est un estimateur biaisé de $\langle \mu, \phi \rangle$

Théorème

$$\mathbb{E}[\, \langle \mu^N, \phi \rangle \,] = \langle \mu, \phi \rangle + O(1/N)$$

$$\{\, \mathbb{E}|\, \langle \mu^N - \mu, \phi \rangle \,|^2 \,\}^{1/2} = \frac{1}{\sqrt{N}} \, (\frac{\langle \eta, g^2 \, |\phi - \langle \mu, \phi \rangle|^2 \rangle}{\langle \eta, g \rangle^2})^{1/2} + O(1/N)$$

et pour tout $p \geq 2$

$$\{\, \mathbb{E}|\, \langle \mu^N - \mu, \phi \rangle \,|^p \,\}^{1/p} = O(1/\sqrt{N})$$

Remarque

$$\frac{\langle \eta, g^2 \, |\phi - \langle \mu, \phi \rangle|^2 \rangle}{\langle \eta, g \rangle^2} \leq \frac{\sup_{x \in E} g(x)}{\langle \eta, g \rangle} \, \mathrm{var}(\phi, \mu)$$

Preuve on vérifie que

$$\Delta_N = \langle \mu^N - \mu, \phi \rangle = \langle \mu^N, \phi - \langle \mu, \phi \rangle \rangle$$

et on a la majoration grossière

$$|\Delta_N| = |\, \langle \mu^N, \phi - \langle \mu, \phi \rangle \rangle \,| \leq \| \, \phi - \langle \mu, \phi \rangle \, \| \leq \mathrm{osc}(\phi)$$

on vérifie aussi que

$$T_N^0 = \langle \gamma^N - \gamma, \phi - \langle \mu, \phi \rangle \rangle = \langle \gamma^N, \phi - \langle \mu, \phi \rangle \rangle$$

compte tenu que

$$\langle \gamma, \phi - \langle \mu, \phi \rangle \rangle = \langle \gamma, 1 \rangle \, \langle \mu, \phi - \langle \mu, \phi \rangle \rangle = 0$$

et on pose

$$D = \langle \gamma, 1 \rangle \qquad \text{et} \qquad D_N = \langle \gamma^N, 1 \rangle$$

de sorte que

$$\Delta_N = \langle \mu^N, \phi - \langle \mu, \phi \rangle \rangle = \frac{\langle \gamma^N, \phi - \langle \mu, \phi \rangle \rangle}{\langle \gamma^N, 1 \rangle} = \frac{T_N^0}{D_N}$$

on rappelle que

$$\frac{T_N^0}{D} = \frac{\langle \gamma^N - \gamma, \phi - \langle \mu, \phi \rangle \rangle}{\langle \gamma, 1 \rangle} \qquad \text{et} \qquad \frac{D_N - D}{D} = \frac{\langle \gamma^N - \gamma, 1 \rangle}{\langle \gamma, 1 \rangle}$$

et on remarque que, pour $\phi' = \phi - \langle \mu, \phi \rangle$

$$g\,\phi' - \langle \eta, g\,\phi' \rangle = g\,(\phi - \langle \mu, \phi \rangle) - \langle \eta, g\,\phi \rangle + \langle \mu, \phi \rangle\,\langle \eta, g \rangle = g\,(\phi - \langle \mu, \phi \rangle)$$

d'après le Théorème précédent appliqué à $\phi - \langle \mu, \phi \rangle$ et à $\phi \equiv 1$, on a

$$\| \frac{T_0^N}{D} \|_2 = \{\, \mathbb{E} | \frac{T_0^N}{D} |^2 \,\}^{1/2} = \frac{1}{\sqrt{N}} \, (\frac{\langle \eta, g^2\, |\phi - \langle \mu, \phi \rangle|^2 \rangle}{\langle \eta, g \rangle^2})^{1/2}$$

$$\| \frac{T_0^N}{D} \|_p \leq \frac{c_p}{\sqrt{N}} \, (\frac{\langle \eta, g^p\, |\phi - \langle \mu, \phi \rangle|^p \rangle}{\langle \eta, g \rangle^p})^{1/p}$$

$$\| \frac{D^N - D}{D} \|_p \leq \frac{c_p}{\sqrt{N}} \, (\frac{\langle \eta, |g - \langle \eta, g \rangle|^p \rangle}{\langle \eta, g \rangle^p})^{1/p}$$

on remarque que

$$\Delta_N = \frac{T_N^0}{D_N} = \frac{T_N^0}{D} - \frac{T_N^0}{D_N}\,\frac{D_N - D}{D} = \frac{T_N^0}{D} - \Delta_N\,\frac{D_N - D}{D}$$

et en itérant

$$\begin{aligned}\Delta_N &= \frac{T_N^0}{D} - \left(\frac{T_N^0}{D} - \Delta_N\,\frac{D_N - D}{D}\right)\frac{D_N - D}{D} \\ &= \frac{T_N^0}{D} - \frac{T_N^0}{D}\,\frac{D_N - D}{D} + \Delta_N\left(\frac{D_N - D}{D}\right)^2\end{aligned}$$

- pour le biais, on remarque que T_N^0 est de moyenne nulle de sorte que

$$\mathbb{E}[\Delta_N] = -\mathbb{E}[\frac{T_N^0}{D}\frac{D_N - D}{D}] + \mathbb{E}[\Delta_N (\frac{D_N - D}{D})^2]$$

et en utilisant l'inégalité triangulaire puis l'inégalité de Hölder, on a

$$\begin{aligned} |\mathbb{E}[\Delta_N]| &\leq \mathbb{E}|\frac{T_N^0}{D}\frac{D_N - D}{D}| + \mathbb{E}|\Delta_N (\frac{D_N - D}{D})^2| \\ &\leq \|\frac{T_N^0}{D}\|_2 \|\frac{D_N - D}{D}\|_2 + \mathrm{osc}(\phi) \|\frac{D_N - D}{D}\|_2^2 \end{aligned}$$

où les deux termes dans la majoration sont d'ordre $1/N$

▶ pour le moment d'ordre 2 (variance), en utilisant l'inégalité triangulaire puis l'inégalité de Hölder, on a

$$
\begin{aligned}
|\,\|\Delta_N\|_2 - \|\frac{T_N^0}{D}\|_2\,| \;&\leq\; \|\Delta_N - \frac{T_N^0}{D}\|_2 \\
&\leq\; \|\frac{T_N^0}{D}\;\frac{D_N - D}{D}\|_2 + \|\Delta_N\,(\frac{D_N - D}{D})^2\|_2 \\
&\leq\; \|\frac{T_N^0}{D}\|_4\;\|\frac{D_N - D}{D}\|_4 + \mathrm{osc}(\phi)\;\|\frac{D_N - D}{D}\|_4^2
\end{aligned}
$$

où les deux termes dans la majoration sont d'ordre $1/N$

- pour le moment d'ordre p, une majoration grossière suffit et en utilisant l'inégalité triangulaire, on a

$$\begin{aligned}\|\Delta_N\|_p &\leq \|\frac{T_N^0}{D}\|_p + \|\Delta_N \frac{D_N - D}{D}\|_p \\ &\leq \|\frac{T_N^0}{D}\|_p + \mathrm{osc}(\phi)\, \|\frac{D_N - D}{D}\|_p\end{aligned}$$

où les deux termes dans la majoration sont d'ordre $1/\sqrt{N}$ $\square$

Théorème

$$\sqrt{N}\,[\frac{\langle \gamma^N, 1\rangle}{\langle \gamma, 1\rangle} - 1] \Longrightarrow \mathcal{N}(0, V) \qquad \text{et} \qquad \sqrt{N}\,\langle \mu^N - \mu, \phi\rangle \Longrightarrow \mathcal{N}(0, v(\phi))$$

en distribution quand $N \uparrow \infty$, pour toute fonction ϕ, avec l'expression suivante pour la variance asymptotique

$$V = \frac{\langle \eta, g^2\rangle}{\langle \eta, g\rangle^2} - 1 \qquad \text{et} \qquad v(\phi) = \frac{\langle \eta, g^2\,|\phi - \langle \mu, \phi\rangle|^2\rangle}{\langle \eta, g\rangle^2}$$

Preuve on remarque que

$$\gamma^N = g\,S^N(\eta) \qquad \text{et} \qquad \mu^N = \frac{\gamma^N}{\langle \gamma^N, 1\rangle}$$

de sorte que pour toute fonction ϕ

$$\sqrt{N}\,\frac{\langle \gamma^N - \gamma, \phi\rangle}{\langle \gamma, 1\rangle} = \sqrt{N}\,\frac{\langle S^N(\eta) - \eta, g\,\phi\rangle}{\langle \eta, g\rangle} \Longrightarrow \mathcal{N}(0, \frac{\mathrm{var}(g\,\phi, \eta)}{\langle \eta, g\rangle^2})$$

en distribution quand $N \uparrow \infty$, pour toute fonction ϕ

en particulier pour $\phi \equiv 1$

$$\sqrt{N}\,[\frac{\langle \gamma^N, 1\rangle}{\langle \gamma, 1\rangle} - 1] \Longrightarrow \mathcal{N}(0, \frac{\mathrm{var}(g,\eta)}{\langle \eta, g\rangle^2})$$

en distribution quand $N \uparrow \infty$, et il résulte de la décomposition suivante

$$\langle \mu^N - \mu, \phi\rangle = \frac{\langle \gamma^N - \gamma, \phi - \langle \mu, \phi\rangle\rangle}{\langle \gamma^N, 1\rangle}$$

que

$$\sqrt{N}\,\langle \mu^N - \mu, \phi\rangle = \frac{\langle \gamma, 1\rangle}{\langle \gamma^N, 1\rangle}\,\sqrt{N}\,\frac{\langle \gamma^N - \gamma, \phi - \langle \mu, \phi\rangle\,\rangle}{\langle \gamma, 1\rangle}$$

d'après la loi des grands nombres

$$\langle \gamma^N, 1\rangle = \frac{1}{N}\,\sum_{i=1}^{N} g(\xi^i) \longrightarrow \langle \eta, g\rangle = \langle \gamma, 1\rangle$$

en probabilité quand $N \uparrow \infty$, et d'après le lemme de Slutsky

$$\sqrt{N}\,\langle \mu^N - \mu, \phi\rangle \Longrightarrow \mathcal{N}(0, \frac{\mathrm{var}(g\,(\phi - \langle \mu, \phi\rangle), \eta)}{\langle \eta, g\rangle^2})$$

en distribution quand $N \uparrow \infty$

finalement, on remarque que

$$\langle \eta, g\,(\phi - \langle \mu, \phi \rangle)\,\rangle = \langle \eta, g\,\phi \rangle - \langle \eta, g \rangle\,\langle \mu, \phi \rangle = 0$$

de sorte que

$$\mathrm{var}(g\,(\phi - \langle \mu, \phi \rangle), \eta) = \langle \eta, g^2\,|\phi - \langle \mu, \phi \rangle|^2 \rangle \qquad \square$$

Remarque la variance asymptotique V s'interprète en terme de la distance du χ^2 entre les distributions de probabilité μ et η, définie par

$$\chi^2(\mu, \eta) = \int_E (\frac{d\mu}{d\eta}(x) - 1)^2\,\eta(dx) = \int_E (\frac{d\mu}{d\eta}(x))^2\,\eta(dx) - 1$$

compte tenu que

$$\mu = g \cdot \eta = \frac{g\,\eta}{\langle \eta, g \rangle} \qquad \text{de sorte que} \qquad \frac{d\mu}{d\eta}(x) = \frac{g(x)}{\langle \eta, g \rangle}$$

Introduction

Distributions de Gibbs–Boltzmann

Mélanges finis
- mélanges finis
- stratégies d'échantillonnage
- lois discrètes
- comparaisons

Approximation particulaire

Mélanges finis

étant donné un mélange fini

$$\eta = \sum_{i=1}^{M} w_i \; m_i \qquad \text{avec} \qquad \sum_{i=1}^{M} w_i = 1$$

de M distributions de probabilité $(m_1, \cdots, m_M)$ avec les poids positifs $(w_1, \cdots, w_M)$, et s'il est facile

- de simuler pour tout $i = 1 \cdots M$ une variable aléatoire distribuée selon m_i

alors il est facile de simuler une variable aléatoire distribuée selon η : il suffit

- de simuler une variable aléatoire I à valeurs dans l'ensemble fini $\{1, \cdots, M\}$ et distribuée selon les poids $(w_1, \cdots, w_M)$, c'est–à–dire

$$\mathbb{P}[I = i] = w_i \qquad \text{pour tout } i = 1 \cdots M$$

- et de générer une variable aléatoire distribuée selon m_I

la probabilité de sélectionner une composante du mélange est d'autant plus grande que le poids de cette composante est grand

objectif : simuler un N–échantillon distribué selon le mélange fini η, ou bien approximer la distribution de probabilité η par un mélange fini de N masses de Dirac, appelées particules, où

> *le nombre N de particules n'est pas nécessairement égal au nombre M de composantes du mélange*

il restera à savoir

- simuler une variable aléatoire à valeurs dans l'ensemble fini $\{1, \cdots, M\}$ et distribuée selon des poids $(w_1, \cdots, w_M)$ donnés
- voire simuler globalement un N–échantillon à valeurs dans l'ensemble fini $\{1, \cdots, M\}$ et distribué selon des poids $(w_1, \cdots, w_M)$ donnés, plus efficacement qu'en répétant N fois la simulation d'une seule variable aléatoire

Échantillonnage multinomial

on simule un N–échantillon $(\xi_1, \cdots, \xi_N)$ distribué selon η, et on pose

$$S^N(\eta) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i} \qquad \texttt{(multi)}$$

utilisation des poids pour sélectionner (avec remise) les composantes du mélange de plus forts poids, avec l'effet attendu que

- les composantes de plus forts poids seront sélectionnées plusieurs fois
- à l'inverse, les composantes de moins forts poids pourront même être éliminées et ne plus être représentées du tout dans l'approximation

si N_i désigne le nombre de fois que la i–ème composante du mélange est sélectionée, ou de manière équivalente son nombre N_i de représentants dans l'approximation, pour tout $i = 1 \cdots M$, alors

le vecteur aléatoire $(N_1, \cdots, N_M)$ suit une loi multinomiale

Théorème la variable aléatoire $\langle S^N(\eta), \phi \rangle$ est un estimateur non–biaisé de $\langle \eta, \phi \rangle$, et les moments de l'erreur d'estimation vérifient

$$\mathbb{E}| \langle S^N(\eta) - \eta, \phi \rangle |^2 = \frac{1}{N} \operatorname{var}(\phi, \eta)$$

ou de manière équivalente

$$\mathbb{E}| \langle S^N(\eta) - \eta, \phi \rangle |^2 = \frac{1}{N} \sum_{i=1}^{M} w_i \operatorname{var}(\phi, m_i) + \frac{1}{N} W_M$$

$$W_M = \sum_{i=1}^{M} w_i \, |\langle m_i, \phi \rangle|^2 - | \sum_{i=1}^{M} w_i \, \langle m_i, \phi \rangle|^2$$

pour toute fonction ϕ

interprétation de W_M comme variance des moyennes *intra–composantes* affectées du poids de chaque composante

preuve de l'équivalence :

$$\begin{aligned}
\mathrm{var}(\phi,\eta) &= \langle \eta, |\phi|^2 \rangle - |\langle \eta, \phi \rangle|^2 \\
&= \sum_{i=1}^{M} w_i \, \langle m_i, |\phi|^2 \rangle - |\sum_{i=1}^{M} w_i \, \langle m_i, \phi \rangle|^2 \\
&= \sum_{i=1}^{M} w_i \, [\, \langle m_i, |\phi|^2 \rangle - |\langle m_i, \phi \rangle|^2 \,] \\
&\qquad + [\, \sum_{i=1}^{M} w_i \, |\langle m_i, \phi \rangle|^2 - |\sum_{i=1}^{M} w_i \, \langle m_i, \phi \rangle|^2 \,] \\
&= \sum_{i=1}^{M} w_i \, \mathrm{var}(\phi, m_i) + [\, \sum_{i=1}^{M} w_i \, |\langle m_i, \phi \rangle|^2 - |\sum_{i=1}^{M} w_i \, \langle m_i, \phi \rangle|^2 \,]
\end{aligned}$$

Remarque intuitivement, si tous les poids sont égaux à (ou proches de) $1/M$, c'est–à–dire si la répartition des poids de mélange est proche de l'équidistribution, alors il peut être contre–productif de sélectionner les composantes du mélange

Stratification par composante

on décide de conserver les poids et de simuler un représentant exactement pour chaque composante du mélange (ce qui impose que N est nécessairement égal au nombre M de composantes du mélange initial), et on pose

$$\eta_M = \sum_{i=1}^{M} w_i \, \delta_{\xi_i} \qquad \text{(strata)}$$

où indépendamment pour tout $i = 1 \cdots M$ la variable aléatoire ξ_i est distribuée selon m_i

Théorème la variable aléatoire $\langle \eta_M, \phi \rangle$ est un estimateur non–biaisé de $\langle \eta, \phi \rangle$, et la variance de l'erreur d'estimation vérifie

$$\mathbb{E}|\langle \eta_M - \eta, \phi \rangle|^2 = \sum_{i=1}^{M} w_i^2 \, \mathrm{var}(\phi, m_i)$$

pour toute fonction ϕ

Preuve par indépendance

$$\mathbb{E}|\langle \eta_M - \eta, \phi \rangle|^2 = \mathbb{E}|\sum_{i=1}^{M} w_i \, [\phi(\xi^i) - \langle m_i, \phi \rangle]|^2 = \sum_{i=1}^{M} w_i^2 \underbrace{\mathbb{E}|\phi(\xi^i) - \langle m_i, \phi \rangle|^2}_{\mathrm{var}(\phi, m_i)}$$

pour toute fonction ϕ $\square$

Remarque intuitivement, cette approche est pertinente dans le cas où la répartition des poids de mélange est proche de l'équidistribution, mais en revanche peu appropriée dans le cas extrême où presque tous les poids sont nuls sauf quelques uns

soit à comparer les variances des estimateurs (`multi`) et (`strata`) avec $N = M$

$$V_{\texttt{multi}} \geq \frac{1}{M} \sum_{i=1}^{M} w_i \, \mathrm{var}(\phi, m_i) \qquad \text{et} \qquad V_{\texttt{strata}} = \sum_{i=1}^{M} w_i^2 \, \mathrm{var}(\phi, m_i)$$

- à l'équidistribution, c'est–à–dire si tous les poids sont égaux entre eux (et égaux à $1/M$), alors

$$V_{\texttt{multi}} \geq \frac{1}{M^2} \sum_{i=1}^{M} \mathrm{var}(\phi, m_i) = V_{\texttt{strata}}$$

ce qui confirme l'intuition que redistribuer est contre–productif dans ce cas

- à l'inverse, si la distribution des poids est complètement dégénérée, c'est–à–dire si tous les poids sont nuls sauf le poids $w_a = 1$ pour une certaine composante du mélange, alors

$$V_{\texttt{multi}} = \frac{1}{M} \mathrm{var}(\phi, m_a) \leq \mathrm{var}(\phi, m_a) = V_{\texttt{strata}}$$

ce qui confirme l'intuition que redistribuer est pertinent dans ce cas

Échantillonnage adaptatif

étant donné le mélange fini

$$\eta = \sum_{i=1}^{M} w_i \; m_i$$

il n'est véritablement intéressant de sélectionner les différentes composantes que si les poids $(w_1, \cdots, w_M)$ sont très déséquilibrés plusieurs critères été proposés pour mesurer l'écart à l'équidistribution, et pour décider de redistribuer ou non les particules

- taille effective de l'échantillon
- entropie de l'échantillon

la distance du χ^2 entre deux vecteurs de probabilité $p = (p_1, \cdots, p_M)$ et $q = (q_1, \cdots, q_M)$ est définie par

$$\chi^2(p,q) = \sum_{i=1}^{M} q_i \, (\frac{p_i}{q_i} - 1)^2 = \sum_{i=1}^{M} \frac{p_i^2}{q_i} - 1$$

et en particulier pour $p = (w_1, \cdots, w_M)$ et $q = (1/M, \cdots, 1/M)$, il vient

$$0 \leq M \sum_{i=1}^{M} w_i^2 - 1 = \frac{M}{M_{\text{eff}}} - 1$$

où M_{eff} est la *taille effective* de l'échantillon, définie par

$$1 \leq M_{\text{eff}} = 1 \; / \; [\sum_{i=1}^{M} w_i^2 \,] \leq M$$

et où l'égalité est atteinte à l'*équidistribution*, ce qui suggère de redistribuer si

$$\frac{M}{M_{\text{eff}}} - 1 \geq \chi^2_{\text{red}} \geq 0 \qquad \text{c–à–d si} \qquad M_{\text{eff}} \leq c\, M$$

avec $c = 1/(1 + \chi^2_{\text{red}}) \leq 1$ où le seuil χ^2_{red} reste à déterminer

ré–interprétation des résultats précédents
pour l'estimateur (`strata`) avec stratification par composante

$$\mathbb{E}|\,\langle \eta_M - \eta, \phi\rangle\,|^2 = \sum_{i=1}^{M} w_i^2 \,\mathrm{var}(\phi, m_i) = \frac{1}{M_{\mathrm{eff}}} \sum_{i=1}^{M} w_i^{\square} \,\mathrm{var}(\phi, m_i)$$

pour toute fonction ϕ, où les nouveaux poids $(w_1^{\square}, \cdots, w_M^{\square})$ sont définis pour tout $i = 1 \cdots M$ par

$$w_i^{\square} = \frac{w_i^2}{\sum_{j=1}^{M} w_j^2} \qquad \text{c–à–d} \qquad w_i^{\square} \propto w_i^2$$

à une constante de normalisation près

on introduit l'approximation adaptative

$$\eta_M = \begin{cases} \displaystyle\sum_{i=1}^{M} w_i \, \delta_{\xi_i} & \text{si } M_{\text{eff}} > c\, M \\ \quad \text{avec } \xi_i \sim m_i \text{ pour tout } i = 1 \cdots M & \\ \displaystyle\frac{1}{M} \sum_{i=1}^{M} \delta_{\xi_i} & \text{si } M_{\text{eff}} \leq c\, M \\ \quad \text{avec } \xi_i \sim \eta \text{ pour tout } i = 1 \cdots M & \end{cases}$$

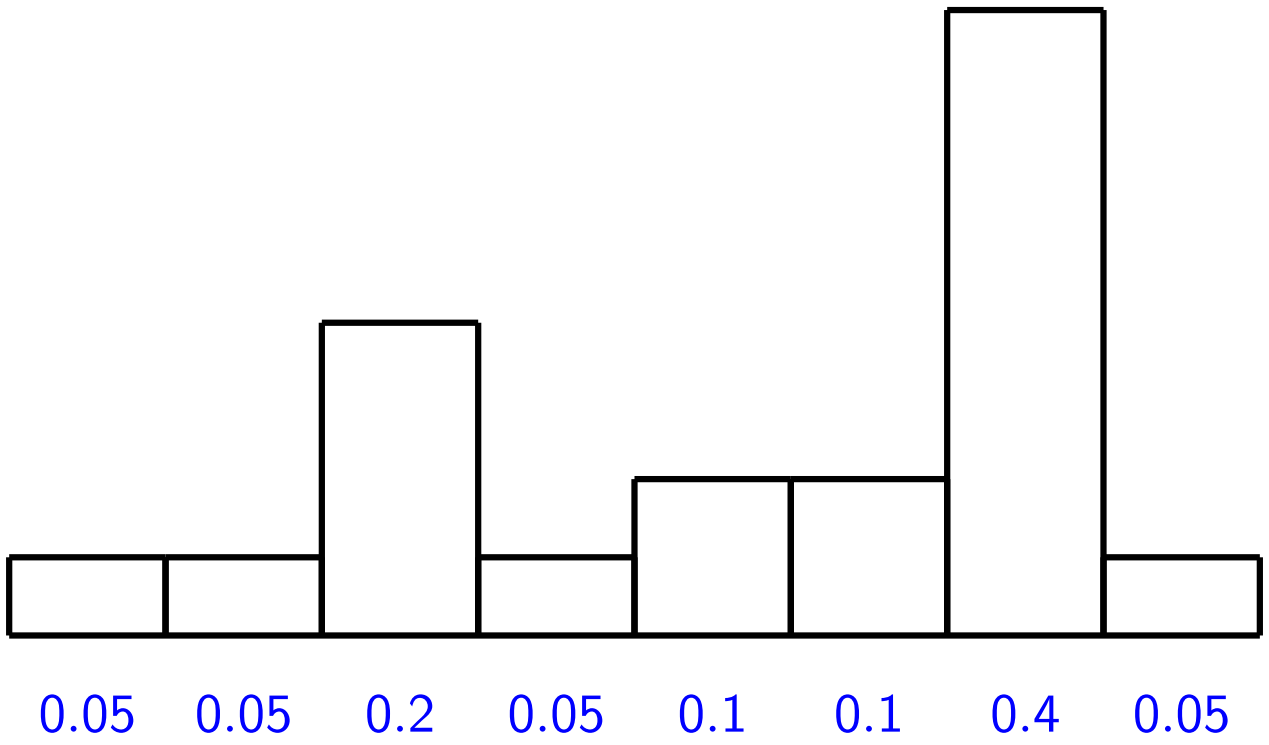
avec l'expression suivante pour la variance de l'erreur d'estimation

$$\mathbb{E}|\,\langle \eta_M - \eta, \phi \rangle\,|^2 = \begin{cases} \displaystyle\frac{1}{M_{\text{eff}}} \sum_{i=1}^{M} w_i^{\square} \, \mathrm{var}(\phi, m_i) & \text{si } M_{\text{eff}} > c\, M \\ \displaystyle\frac{1}{M} \, \mathrm{var}(\phi, \eta) & \text{si } M_{\text{eff}} \leq c\, M \end{cases}$$

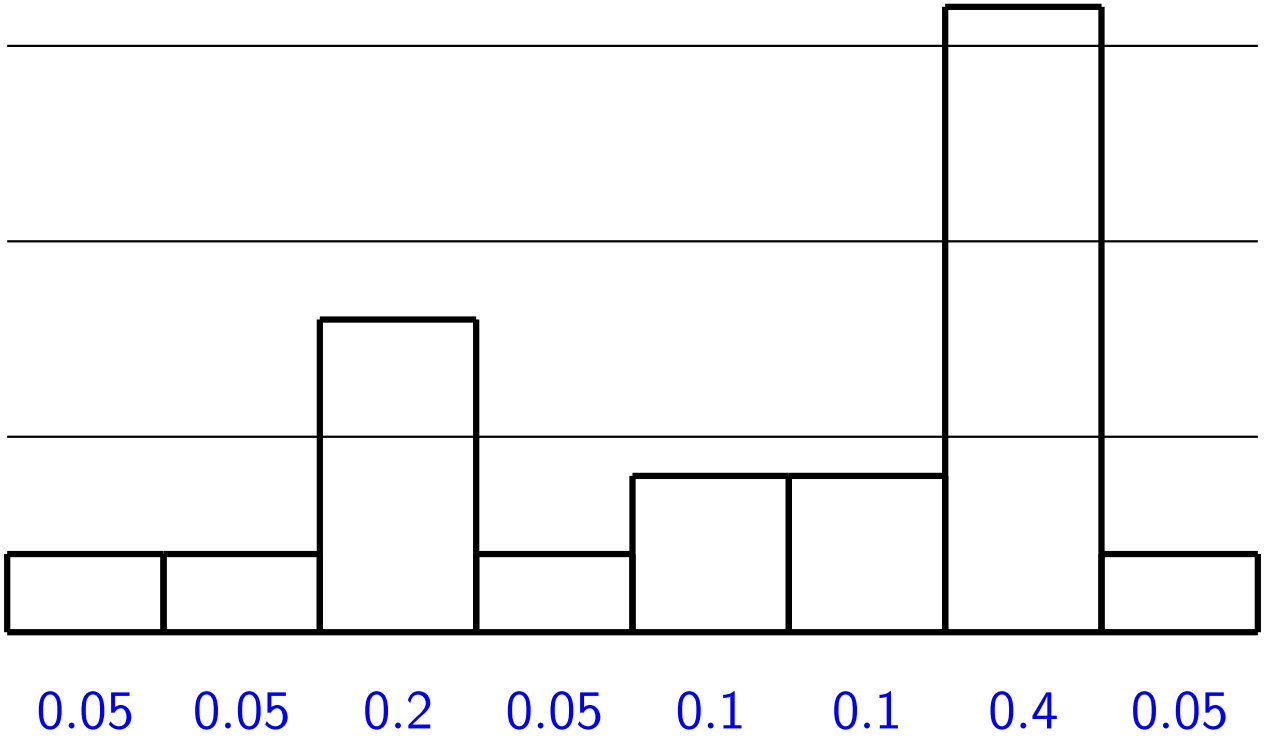
Remarque échantillonner selon les poids respectifs n'est donc approprié que dans les cas où la répartition des poids de mélange est éloigné de l'équidistribution, mais introduit de toute manière un aléa supplémentaire

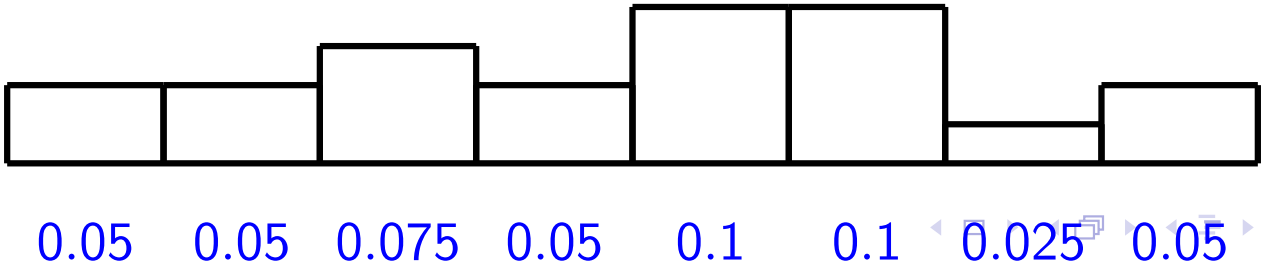
pour limiter cette source d'aléa, on peut par exemple affecter de manière déterministe à chaque composante $i = 1 \cdots M$ un nombre de représentants égal au nombre N_i de fois que le poids $1/N$ est contenu dans le poids w_i de la composante (le poids $1/N$ est celui qui sera affecté à chaque particule dans l'approximation finale)

à l'issue de cette première passe, $(N - N_0)$ représentants sont déjà affectés et il reste ensuite à compléter la population de particules de manière à assurer un effectif de la taille N désirée, par exemple en simulant un N_0–échantillon selon la distribution résiduelle des poids non encore affectés



ici $N = M = 8$, d'où $1/N = 0.125$, et à l'issue de la première passe :
la composante 3 a reçu 1 représentant, la composante 7 a reçu 3 représentants,
et il reste $N_0 = 4$ représentants à affecter


0.05
0.05
0.2
0.05
0.1
0.1
0.4
0.05


0.05
0.05
0.075
0.05
0.1
0.1
0.025
0.05

Échantillonnage résiduel multinomial

pour toute composante $i = 1 \cdots M$ on définit le nombre $N_i = \lfloor N\, w_i \rfloor$ de ses représentants affectés à l'issue de la première passe, comme résultat de la division euclidienne

$$N\, w_i = N_i + q_i \qquad \text{avec} \qquad 0 \leq q_i < 1$$

et compte tenu des identités

$$N = \sum_{i=1}^{M} N\, w_i = \sum_{i=1}^{M} (N_i + q_i) = \sum_{i=1}^{M} N_i + N_0 \qquad \text{avec} \qquad N_0 = \sum_{i=1}^{M} q_i$$

et

$$\eta = \sum_{i=1}^{M} w_i\, m_i = \frac{1}{N} \sum_{i=1}^{M} (N_i + q_i)\, m_i = \sum_{i=1}^{M} \frac{N_i}{N}\, m_i + \frac{N_0}{N}\, m_0$$

avec

$$m_0 = \sum_{i=1}^{M} \frac{q_i}{N_0}\, m_i$$

on déduit qu'il reste N_0 représentants à affecter de manière à approcher la distribution de probabilité résiduelle convenablement renormalisée m_0

l'approximation proposée consiste à simuler

- pour tout $i = 1 \cdots M$, un N_i–échantillon $(\xi^{i,1}, \cdots, \xi^{i,N_i})$ distribué selon m_i
- un N_0–échantillon $(\xi^{0,1}, \cdots, \xi^{0,N_0})$ distribué selon le mélange fini m_0

toutes les variables aléatoires étant simulées de manière indépendantes, et à poser

$$\eta_N = \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{N_i} \delta_{\xi^{i,j}} + \frac{1}{N} \sum_{j=1}^{N_0} \delta_{\xi^{0,j}} \qquad \texttt{(residu)}$$

c'est–à–dire

$$\langle \eta_N, \phi \rangle = \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{N_i} \phi(\xi^{i,j}) + \frac{1}{N} \sum_{j=1}^{N_0} \phi(\xi^{0,j})$$

et par différence

$$\langle \eta_N - \eta, \phi \rangle = \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{N_i} [\phi(\xi^{i,j}) - \langle m_i, \phi \rangle] + \frac{1}{N} \sum_{j=1}^{N_0} [\phi(\xi^{0,j}) - \langle m_0, \phi \rangle]$$

pour toute fonction ϕ

Théorème la variable aléatoire $\langle \eta_N, \phi \rangle$ est un estimateur non–biaisé de $\langle \eta, \phi \rangle$, et la variance de l'erreur d'estimation vérifie

$$\mathbb{E}|\,\langle \eta_N - \eta, \phi \rangle\,|^2 = \frac{1}{N}\,[\sum_{i=1}^{M} \frac{N_i}{N}\,\mathrm{var}(\phi, m_i) + \frac{N_0}{N}\,\mathrm{var}(\phi, m_0)\,]$$

ou de manière équivalente

$$\mathbb{E}|\,\langle \eta_N - \eta, \phi \rangle\,|^2 = \frac{1}{N}\,\sum_{i=1}^{M} w_i\,\mathrm{var}(\phi, m_i) + \frac{N_0}{N^2}\,W_M$$

$$W_M = \sum_{i=1}^{M} \frac{q_i}{N_0}\,|\langle m_i, \phi \rangle\,|^2 - |\,\sum_{i=1}^{M} \frac{q_i}{N_0}\,\langle m_i, \phi \rangle\,|^2$$

pour toute fonction ϕ

interprétation de W_M comme variance des moyennes *intra–composantes* affectées du poids *résiduel* de chaque composante

preuve de l'équivalence :

$$\frac{N_i}{N}\,\mathrm{var}(\phi, m_i) = w_i\,\mathrm{var}(\phi, m_i) - \frac{q_i}{N}\,[\,\langle m_i, |\phi|^2\rangle - |\langle m_i, \phi\rangle|^2\,]$$

$$\mathrm{var}(\phi, m_0) = \sum_{i=1}^{M} \frac{q_i}{N_0}\,\langle m_i, |\phi|^2\rangle - |\sum_{i=1}^{M} \frac{q_i}{N_0}\,\langle m_i, \phi\rangle\,|^2$$

de sorte que

$$\sum_{i=1}^{M} \frac{N_i}{N}\,\mathrm{var}(\phi, m_i) + \frac{N_0}{N}\,\mathrm{var}(\phi, m_0) =$$

$$= \sum_{i=1}^{M} w_i\,\mathrm{var}(\phi, m_i) - \sum_{i=1}^{M} \frac{q_i}{N}\,[\,\langle m_i, |\phi|^2\rangle - |\langle m_i, \phi\rangle|^2\,]$$

$$+ \frac{N_0}{N}\,[\sum_{i=1}^{M} \frac{q_i}{N_0}\,\langle m_i, |\phi|^2\rangle - |\sum_{i=1}^{M} \frac{q_i}{N_0}\,\langle m_i, \phi\rangle\,|^2\,]$$

$$= \sum_{i=1}^{M} w_i\,\mathrm{var}(\phi, m_i) + \frac{N_0}{N}\,[\sum_{i=1}^{M} \frac{q_i}{N_0}\,|\langle m_i, \phi\rangle|^2 - |\sum_{i=1}^{M} \frac{q_i}{N_0}\,\langle m_i, \phi\rangle|^2\,]$$

Preuve du Théorème par indépendance

$$\mathbb{E}|\,\langle \eta_N - \eta, \phi\rangle\,|^2 =$$

$$= \mathbb{E}|\,\frac{1}{N}\,\sum_{i=1}^{M}\sum_{j=1}^{N_i}[\phi(\xi^{i,j}) - \langle m_i, \phi\rangle] + \frac{1}{N}\,\sum_{j=1}^{N_0}[\phi(\xi^{0,j}) - \langle m_0, \phi\rangle]\,|^2$$

$$= \frac{1}{N^2}\,\sum_{i=1}^{M}\sum_{j=1}^{N_i}\mathbb{E}|\phi(\xi^{i,j}) - \langle m_i, \phi\rangle|^2 + \frac{1}{N^2}\,\sum_{j=1}^{N_0}\mathbb{E}|\phi(\xi^{0,j}) - \langle m_0, \phi\rangle|^2$$

$$= \frac{1}{N}\,[\sum_{i=1}^{M}\frac{N_i}{N}\,\mathrm{var}(\phi, m_i) + \frac{N_0}{N}\,\mathrm{var}(\phi, m_0)\,]$$

pour toute fonction ϕ $\square$

Remarque on rappelle que

$$\eta = \sum_{i=1}^{M} \frac{N_i}{N}\, m_i + \frac{N_0}{N}\, m_0$$

de sorte que

$$\sum_{i=1}^{M} \frac{N_i}{N}\, \mathrm{var}(\phi, m_i) + \frac{N_0}{N}\, \mathrm{var}(\phi, m_0)$$

$$= \sum_{i=1}^{M} \frac{N_i}{N}\, \langle m_i, |\phi|^2 \rangle - \sum_{i=1}^{M} \frac{N_i}{N}\, |\langle m_i, \phi \rangle|^2$$

$$+ \frac{N_0}{N}\, \langle m_0, |\phi|^2 \rangle - \frac{N_0}{N}\, |\langle m_0, \phi \rangle\,|^2$$

$$= \langle \eta, |\phi|^2 \rangle - [\sum_{i=1}^{M} \frac{N_i}{N}\, |\langle m_i, \phi \rangle|^2 + \frac{N_0}{N}\, |\langle m_0, \phi \rangle\,|^2\,]$$

tandis que

$$\mathrm{var}(\phi,\eta) = \langle \eta, |\phi|^2 \rangle - \left| \sum_{i=1}^{M} \frac{N_i}{N} \langle m_i, \phi \rangle + \frac{N_0}{N} \langle m_0, \phi \rangle \right|^2$$

d'après l'inégalité de Cauchy–Schwartz, on a

$$\left| \sum_{i=1}^{M} \frac{N_i}{N} \langle m_i, \phi \rangle + \frac{N_0}{N} \langle m_0, \phi \rangle \right|^2$$

$$\leq \sum_{i=1}^{M} \frac{N_i}{N} |\langle m_i, \phi \rangle|^2 + \frac{N_0}{N} |\langle m_0, \phi \rangle|^2$$

on en déduit que la variance de l'erreur d'estimation pour l'algorithme d'échantillonnage résiduel multinomial est inférieure à la variance de l'erreur d'estimation pour l'algorithme d'échantillonnage multinomial

Échantillonnage selon une loi discrète

objectif : simuler une variable aléatoire I, ou un N–échantillon $(I_1, \cdots, I_N)$, à valeurs dans l'ensemble fini $\{1, \cdots, M\}$ et distribué selon les poids $(w_1, \cdots, w_M)$

la méthode la plus directe est la méthode d'inversion :
on découpe l'intervalle $[0, 1]$ en M segments adjacents de longueurs respectives $(w_1, \cdots, w_M)$,

- on simule une variable aléatoire U uniforme sur $[0, 1]$
- si U appartient au j–ème segment, alors on pose $I = j$

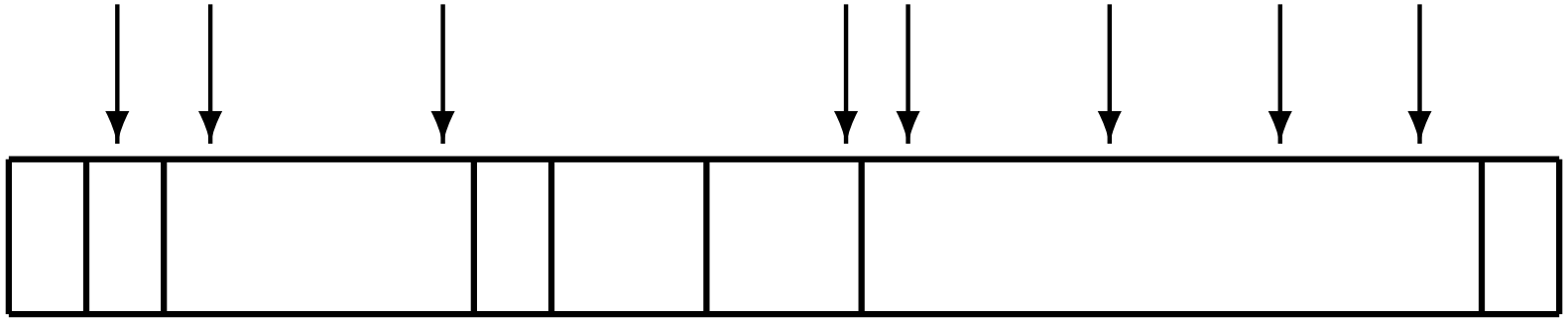
une recherche binaire en $O(\log_2 M)$ opérations permet d'obtenir ce résultat, et il suffit donc de $N\, O(\log_2 M)$ opérations pour générer un N–échantillon à valeurs dans l'ensemble fini $\{1, \cdots, M\}$ et distribué selon les poids $(w_1, \cdots, w_M)$

amélioration : au lieu de répéter N fois l'opération de

- générer une variable aléatoire uniforme sur $[0, 1]$
- puis effectuer une recherche binaire

on simule un N–échantillon $(U_1, \cdots, U_N)$ de variables aléatoires uniformes sur $[0, 1]$, on ordonne cet échantillon, ce qui nécessite $O(N \log_2 N)$ opérations, et on applique la méthode d'inversion à l'échantillon ré–ordonné $U_{(1)} \leq \cdots \leq U_{(N)}$

si $U_{(i)}$ appartient au j–ème segment, alors on pose $I_i = j$, de sorte que pour simuler I_{i+1} il suffit de tester l'appartenance de $U_{(i+1)}$ aux segments situés à partir du j–ème segment



amélioration supplémentaire : pour éviter l'étape préalable de ré–ordonner les variables aléatoires uniformes, on simule directement une N–statistique d'ordre uniforme, c'est–à–dire un vecteur aléatoire $(V_1, \cdots, V_N)$ distribué comme le vecteur aléatoire obtenu en ré–ordonnant un N–échantillon $(U_1, \cdots, U_N)$ de variables aléatoires uniformes sur $[0, 1]$, ce qui peut être effectué en $O(N)$ opérations

Proposition soit $(U_1, \cdots, U_N)$ un N–échantillon de variables aléatoires uniformes sur $[0, 1]$, et on définit

$$V_i = U_N^{1/N} \cdots U_i^{1/i} \qquad \text{pour tout } i = N \cdots 1$$

ou bien par récurrence : $V_N = U_N^{1/N}$ et

$$V_i = V_{i+1}\, U_i^{1/i} \qquad \text{pour tout } i = N - 1 \cdots 1$$

alors le vecteur aléatoire $(V_1, \cdots, V_N)$ est distribué comme le vecteur aléatoire $(U_{(1)}, \cdots, U_{(N)})$ obtenu en ré–ordonnant $(U_1, \cdots, U_N)$

fonction Python `resampling_multi.py`

Échantillonnage systématique ou universel

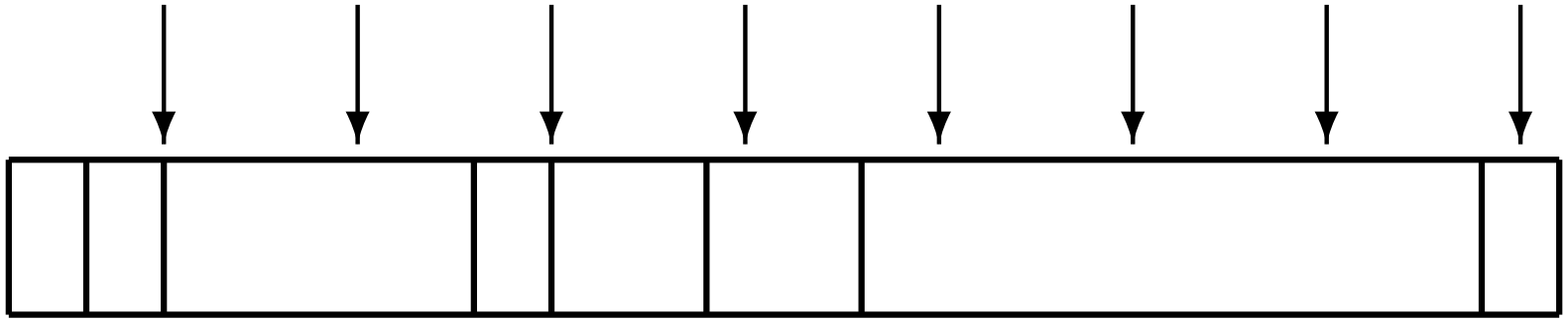
toutes ces méthodes tentent de mettre en œuvre l'idée que le nombre O_j de représentants attribués la j–ème composante devrait être 'proportionnel' au poids w_j

il existe une manière très radicale d'atteindre cet objectif : on simule une unique variable aléatoire U uniforme sur $[0, 1/N]$, et on pose

$$U_i = (i-1)/N + U \qquad \text{pour tout } i = 1 \cdots N$$

clairement, la suite $U_1 \leq \cdots \leq U_N$ obtenue est ordonnée

si U_i appartient au j–ème segment, alors on pose $I_i = j$



Introduction

Distributions de Gibbs–Boltzmann

Mélanges finis

Approximation particulaire

Motivation (rappel)

objectif : approximer la suite $\{\mu_k\}$ (par exemple le filtre bayésien) définie par la relation de récurrence

$$\mu_{k-1} \xrightarrow{\text{mutation}} \eta_k = \mu_{k-1}\, Q_k \xrightarrow{\text{pondération}} \mu_k = g_k \cdot \eta_k$$

avec la condition initiale $\mu_0 = g_0 \cdot \eta_0$
idée : rechercher une approximation sous la forme de distributions de probabilité empiriques (éventuellement pondérées)

$$\eta_k \approx \eta_k^N = \sum_{i=1}^{N} v_k^i\, \delta_{\xi_k^i} \quad \text{et} \quad \mu_k \approx \mu_k^N = \sum_{i=1}^{N} w_k^i\, \delta_{\xi_k^i}$$

associées à une population de N particules caractérisée par

- les *positions* $(\xi_k^1, \cdots, \xi_k^N)$ dans E
- et les *poids* positifs normalisés $(v_k^1, \cdots, v_k^N)$ et $(w_k^1, \cdots, w_k^N)$

initialisation : par échantillonnage pondéré

$$\mu_0 = g_0 \cdot \eta_0 \approx g_0 \cdot S^N(\eta_0) = \sum_{i=1}^{N} \frac{g_0(\xi_0^i)\ \delta_{\xi_0^i}}{\sum_{j=1}^{N} g_0(\xi_0^j)} = \sum_{i=1}^{N} w_0^i\ \delta_{\xi_0^i}$$

où les variables aléatoires $(\xi_0^1, \cdots, \xi_0^N)$ sont i.i.d. de distribution commune η_0

étape de *correction* : clairement, à partir de la définition

$$\mu_k^N = g_k \cdot \eta_k^N = \sum_{i=1}^{N} \frac{v_k^i\, g_k(\xi_k^i)\ \delta_{\xi_k^i}}{\sum_{j=1}^{N} v_k^j\, g_k(\xi_k^j)} = \sum_{i=1}^{N} w_k^i\ \delta_{\xi_k^i}$$

est automatiquement de la forme recherchée (noter la mise–à–jour multiplicative des poids)

étape de *prédiction* : à partir de la définition

$$\begin{aligned}\langle \mu^N_{k-1}\, Q_k, \phi\rangle &= \int \mu^N_{k-1}(dx) \int Q_k(x, dx')\, \phi(x') \\ &= \sum_{i=1}^N w^i_{k-1} \int Q_k(\xi^i_{k-1}, dx')\, \phi(x') \\ &= \int [\sum_{i=1}^N w^i_{k-1}\; Q_k(\xi^i_{k-1}, dx')\,]\, \phi(x')\end{aligned}$$

pour toute fonction ϕ, de sorte que

$$\mu^N_{k-1}\, Q_k(dx') = \sum_{i=1}^N w^i_{k-1}\; m^i_k(dx')$$

s'exprime comme un mélange fini, avec

$$m^i_k(dx') = Q_k(\xi^i_{k-1}, dx') \qquad \text{pour tout } i = 1 \cdots N$$

qu'il s'agit d'approximer / échantillonner, selon une méthode appropriée

première méthode : ré–échantillonnage multinomial

$$\eta_k^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_k^i}$$

où $(\xi_k^1, \cdots, \xi_k^N)$ est un N–échantillon distribué selon le mélange fini, est de la forme recherchée avec $v_k^i = 1/N$ pour tout $i = 1 \cdots N$

deuxième méthode : stratification par composante

$$\eta_k^N = \sum_{i=1}^{N} w_{k-1}^i \, \delta_{\xi_k^i}$$

où indépendamment pour tout $i = 1 \cdots N$ la v.a. ξ_k^i est distribué selon la i–ème composante du mélange fini, est de la forme recherchée avec $v_k^i = w_{k-1}^i$ pour tout $i = 1 \cdots N$

troisième méthode : ré–échantillonnage adaptatif
on calcule la *taille effective* de l'échantillon

$$N_{k-1}^{\mathrm{eff}} = 1 \,/\, [\, \sum_{i=1}^{N} (w_{k-1}^{i})^2 \,]$$

si $N_{k-1}^{\mathrm{eff}} > c\, N$, alors

$$\eta_k^N = \sum_{i=1}^{N} w_{k-1}^{i}\, \delta_{\xi_k^i}$$

où indépendamment pour tout $i = 1 \cdots N$ la v.a. ξ_k^i est distribuée selon la i–ème composante du mélange fini

si $N_{k-1}^{\mathrm{eff}} \leq c\, N$, alors

$$\eta_k^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_k^i}$$

où $(\xi_k^1, \cdots, \xi_k^N)$ est un N–échantillon distribué selon le mélange fini
(noter la ré–initialisation des poids)

Filtres SIR, SIS, adaptatif

filtre SIR (*sampling with importance resampling*)

- pour $k = 0$, indépendamment pour tout $i = 1 \cdots N$
  on simule une v.a. ξ_0^i distribuée selon $\eta_0(dx)$, et on définit

$$w_0^i = g_0(\xi_0^i) \;/\; \left[\sum_{j=1}^{N} g_0(\xi_0^j) \right]$$

- pour tout $k = 1 \cdots n$, indépendamment pour tout $i = 1 \cdots N$
  - on sélectionne un individu $\widehat{\xi}_{k-1}^i$ au sein de la population $(\xi_{k-1}^1, \cdots, \xi_{k-1}^N)$ et selon les poids $(w_{k-1}^1, \cdots, w_{k-1}^N)$
  - on simule une v.a. ξ_k^i distribuée selon $Q_k(\widehat{\xi}_{k-1}^i, dx')$

  et on définit

$$w_k^i = g_k(\xi_k^i) \;/\; \left[\sum_{j=1}^{N} g_k(\xi_k^j) \right]$$

filtre SIS (*sequential importance sampling*)

- pour $k = 0$, indépendamment pour tout $i = 1 \cdots N$
  on simule une v.a. ξ_0^i distribuée selon $\eta_0(dx)$, et on définit

$$w_0^i = g_0(\xi_0^i) \;/\; \Big[\sum_{j=1}^{N} g_0(\xi_0^j) \Big]$$

- pour tout $k = 1 \cdots n$, indépendamment pour tout $i = 1 \cdots N$
  - on simule une v.a. ξ_k^i distribuée selon $Q_k(\xi_{k-1}^i, dx')$

  et on définit

$$w_k^i = w_{k-1}^i \, g_k(\xi_k^i) \;/\; \Big[\sum_{j=1}^{N} w_{k-1}^j \, g_k(\xi_k^j) \Big]$$

filtre particulaire adaptatif

- pour $k = 0$, indépendamment pour tout $i = 1 \cdots N$
  on simule une v.a. ξ_0^i distribuée selon $\eta_0(dx)$, et on définit

$$w_0^i = g_0(\xi_0^i) \;/\; \Big[\sum_{j=1}^{N} g_0(\xi_0^j) \Big]$$

- pour tout $k = 1 \cdots n$, on calcule la taille effective N_{k-1}^{eff}
  si $N_{k-1}^{\text{eff}} > c\, N$, alors indépendamment pour tout $i = 1 \cdots N$
  - on simule une v.a. ξ_k^i distribuée selon $Q_k(\xi_{k-1}^i, dx')$

  et on définit

$$w_k^i = w_{k-1}^i \, g_k(\xi_k^i) \;/\; \Big[\sum_{j=1}^{N} w_{k-1}^j \, g_k(\xi_k^j) \Big]$$

filtre particulaire adaptatif

- pour $k = 0$, indépendamment pour tout $i = 1 \cdots N$
  on simule une v.a. ξ_0^i distribuée selon $\eta_0(dx)$, et on définit

$$w_0^i = g_0(\xi_0^i) \;/\; \left[\sum_{j=1}^{N} g_0(\xi_0^j) \right]$$

- pour tout $k = 1 \cdots n$, on calcule la taille effective N_{k-1}^{eff}
  si $N_{k-1}^{\text{eff}} \leq c\, N$, alors indépendamment pour tout $i = 1 \cdots N$
  - on sélectionne un individu $\widehat{\xi}_{k-1}^i$ au sein de la population $(\xi_{k-1}^1, \cdots, \xi_{k-1}^N)$ et selon les poids $(w_{k-1}^1, \cdots, w_{k-1}^N)$
  - on simule une v.a. ξ_k^i distribuée selon $Q_k(\widehat{\xi}_{k-1}^i, dx')$

  et on définit

$$w_k^i = w_{k-1}^i \, g_k(\xi_k^i) \;/\; \left[\sum_{j=1}^{N} w_{k-1}^j \, g_k(\xi_k^j) \right]$$

filtre particulaire adaptatif (formulation équivalente)

- pour $k = 0$, indépendamment pour tout $i = 1 \cdots N$
  on simule une v.a. ξ_0^i distribuée selon $\eta_0(dx)$, et on définit

$$w_0^i = g_0(\xi_0^i) \;/\; \Big[\sum_{j=1}^{N} g_0(\xi_0^j) \Big]$$

- pour tout $k = 1 \cdots n$, on calcule la taille effective N_{k-1}^{eff}
  si $N_{k-1}^{\mathrm{eff}} > c\, N$, alors indépendamment pour tout $i = 1 \cdots N$
  - on simule une v.a. ξ_k^i distribuée selon $Q_k(\xi_{k-1}^i, dx')$

  si $N_{k-1}^{\mathrm{eff}} \leq c\, N$, alors indépendamment pour tout $i = 1 \cdots N$
  - on sélectionne un individu $\widehat{\xi}_{k-1}^i$ au sein de la population $(\xi_{k-1}^1, \cdots, \xi_{k-1}^N)$ et selon les poids $(w_{k-1}^1, \cdots, w_{k-1}^N)$
  - on simule une v.a. ξ_k^i distribuée selon $Q_k(\widehat{\xi}_{k-1}^i, dx')$
  - on ré–initialise le poids $w_{k-1}^i = 1/N$

  et (quel que soit le cas) on définit

$$w_k^i = w_{k-1}^i \, g_k(\xi_k^i) \;/\; \Big[\sum_{j=1}^{N} w_{k-1}^j \, g_k(\xi_k^j) \Big]$$

en résumé dans le filtre SIR, les particules $(\xi_{k-1}^1, \cdots, \xi_{k-1}^N)$

- *sont sélectionnées* selon leurs poids respectifs $(w_{k-1}^1, \cdots, w_{k-1}^N)$ [étape de *sélection*]
- *évoluent* selon les probabilités de transition $Q_k(x, dx')$ [étape de *mutation*]
- et *sont pondérées* en évaluant la fonction de vraisemblance $g_k(x')$ [étape de *pondération*]

pros : au lieu de s'accumuler le long de chaque trajectoire, les poids sont ici utilisés pour redistribuer les particules
les particules de plus fort poids sont multipliées et les particules de plus faible poids sont éliminées
en ne conservant à chaque pas de temps que les particules les plus pertinentes, on espère concentrer la puissance de calcul disponible dans les régions d'intérêt

cons : introduction d'un aléa supplémentaire, à l'étape de ré–échantillonnage

solutions proposées

- autres procédures de redistribution / allocation d'un nombre (quasiment) déterministe de descendants à chaque particule
- ré–échantillonnage adaptatif, seulement quand les poids $(w_k^1, \cdots, w_k^N)$ sont trop déséquilibrés

cons : réduction du nombre de positions différentes

- *dégénérescence des positions* : en pratique, on compte sur l'étape de mutation pour recréer de la diversité

solution alternative

- régularisation : après l'étape de ré–échantillonnage, on déplace d'une petite quantité aléatoire chaque particule sélectionnée