

# Bayesian modeling of earthquakes

---

Ibrahim SEYDI

## Restricted case

### Spatial data without aftershocks or foreshocks

Let observed seismic events be :

$$x_1, \dots, x_n \in \mathbb{R}^2.$$

We assume that each observation  $x_i = (x_i^{(1)}, x_i^{(2)})$  is a spatial coordinate drawn independently from a density  $f$  which is unknown.

We set that :

$$\begin{aligned} x_i | \theta_i &\sim \mathcal{N}(x_i | \mu_i, \Sigma_i), \quad i = 1, \dots, n \\ \theta_i = (\mu_i, \Sigma_i) | G &\sim G \\ G | \alpha, G_0 &\sim \text{DP}(\alpha, G_0) \\ G_0 | m_0, \Lambda_0, \psi_0, \nu_0 &= \mathcal{NIW}(m_0, \Lambda_0, \psi_0, \nu_0) \end{aligned}$$

where :

- $\mathcal{N}(\cdot | \mu_i, \Sigma_i)$  is a bivariate normal distribution
- $\theta_i = (\mu_i, \Sigma_i) \in \mathbb{R}^2 \times \mathcal{S}_+^2$ , with  $\mathcal{S}_+^2$  the set of symmetric positive definite  $2 \times 2$  covariance matrices
- $G$  is a random probability measure over the parameter space  $\theta$ , drawn from a **Dirichlet process**
- $G_0$  is the **base measure** following a normal-inverse-Wishart on  $\theta_i$ .

# Simulation de processus de Dirichlet

## Simulation par Stick-Breaking

Générer une (approximation) de la densité sous la forme :

$$f(x) = \sum_{k=1}^K w_k \delta_{\theta_k}(x)$$

où :  $\theta_k \sim G_0$ .

Tronquer le modèle à un nb K fixé de composantes du mélange.

### Input

- Nombre de composantes  $K$
- Param de concentration  $\alpha > 0$

### Étapes :

1. Initialisation : Créer liste vide **poids** = [] et le reste du bâton :  $r = 1.0$  ; Créer liste vide **θ**
2. Pour  $k = 1$  à  $K - 1$  :
  - Tirer  $v_k \sim \text{Beta}(1, \alpha)$
  - Calculer  $w_k = v_k \cdot r$
  - Ajouter  $w_k$  à **poids**
  - Mise à jour du baton  $r = r \cdot (1 - v_k)$
  - Simuler  $\theta_k \sim G_0$
  - Ajouter  $\theta_k$  à **θ**
3. Ajouter  $w_K = r$  à **poids**
4. Simuler  $\theta_K \sim G_0$  et l'ajouter à **θ**

**Output** : Approximation d'un DP avec :  $\mathcal{P} = \sum_{k=1}^K w_k \delta_{\theta_k}$

## Simulation par Stick-Breaking (avec seuil $\tau$ )

Générer une approximation de la densité sous la forme :

$$f(x) = \sum_{k=1}^{\infty} w_k \delta_{\theta_k}(x)$$

où  $\theta_k \sim G_0$ , et les poids sont générés par le procédé de Stick-Breaking.

### Input :

- Param de concentration  $\alpha > 0$
- Seuil  $\tau > 0$

**Étapes :**

1. Initialisation :
  - Liste vide des poids : `poids = []`
  - Liste vide des paramètres :  $\theta = []$
  - Reste du bâton :  $r \leftarrow 1.0$
  - Indice :  $k \leftarrow 1$
2. Tant que  $r > \tau$ , faire :
  - Tirer  $v_k \sim \text{Beta}(1, \alpha)$
  - Calculer  $w_k = v_k \cdot r$
  - Ajouter  $w_k$  à `poids`
  - Mettre à jour :  $r \leftarrow r \cdot (1 - v_k)$
  - Simuler  $\theta_k \sim G_0$
  - Ajouter  $\theta_k$  à  $\theta$
  - Incrément  $k \leftarrow k + 1$

**Output :** Approximation d'un DP avec :

$$\mathcal{P} = \sum_{k=1}^K w_k \delta_{\theta_k}(x), \quad \text{où } K \text{ est déterminé en fonction de } \tau$$

# Intégrer l'information des zonages sismotectoniques sous forme de prior informatif

Nous avons accès à un nombre  $n$  de positions de séismes sur un lieu  $\Omega$  :

$$x_1, \dots, x_n \sim f \quad (f \text{ densité})$$

où  $x_i = (x_i^{(1)}, x_i^{(2)})$  pour tout  $i \in [1, n]$ .

Notre approche : Estimation bayésienne non paramétrique de  $f$

$$\begin{cases} f(x) = \int \mathcal{N}(\mu, \Sigma) dG(\mu, \Sigma) \\ G \sim \text{DP}(\alpha, G_0) \end{cases}$$

Autre formulation :

$$\begin{aligned} f(x) &= \sum_{k=1}^{\infty} w_k \mathcal{N}(\mu_k, \Sigma_k) \\ (w_k)_k &\sim \text{SB}(\alpha) \\ (\mu_k, \Sigma_k)_k &\sim G_0 = \mathcal{N}(\mu_k \mid \mu_0, \frac{\Sigma_k}{\lambda_0}) \mathcal{IW}(\Sigma_k \mid \psi_0, \nu_0) \end{aligned}$$

## But : Intégrer prior informatif du zonage sismotectonique

On note  $f_0$  la densité de la distribution du zonage. on a :

$$f_0(x) = \frac{\sum_{j=1}^J w_{0,j} \mathbb{1}_{S_{0,j}}(x)}{\sum_{j=1}^J w_{0,j} A_{0,j}}$$

où  $S_{0,1}, \dots, S_{0,J}$  est une partition de  $\Omega$  et représente les zones d'un zonage sismotectonique et chaque  $A_{0,j}$  est la surface de  $S_{0,j}$ .

Une idée serait d'utiliser des gaussiennes pour approcher les découpages du zonage avec  $\mu_{0,j}$  des centroides des zones  $S_{0,j}$  et  $\Sigma_{0,j}$  des diamètres d'ellipses (?). On aurait :

$$\tilde{f}_0(x) = \frac{\sum_{j=1}^J w_{0,j} \mathcal{N}(\mu_{0,j}, \Sigma_{0,j})}{\sum_{j=1}^J w_{0,j} A_{0,j}}$$

Ainsi, on aurait la mesure de base a priori informative suivante :

$$G_0^{\text{inf}}(\cdot) = \sum_{j=1}^J w_{0,j} \mathcal{N}(\cdot \mid \mu_{0,j}, \frac{\Sigma_j}{\lambda_0}) \mathcal{IW}(\cdot \mid \Sigma_{0,j}, \nu_0)$$

Première étape : Évaluation de la qualité de la version informative

On cherche à construire une densité spatiale sur la carte de France  $\Omega$  à partir d'un zonage sismo, c'est-à-dire :

$$\int_{\text{France}} f_0(x, y) dx dy = 1$$

où  $f_0(x, y)$  est constante sur chaque zone  $S_{0,j}$ .

Soit  $\{S_{0,1}, \dots, S_{0,J}\}$  un zonage sismotectonique de la France  $\Omega$ , tel que  $\Omega = \bigcup_{j=1}^J S_{0,j}$ , avec  $S_{0,j} \cap S_{0,i} = \emptyset$  si  $i \neq j$ . Chaque  $S_{0,j}$  a :

- une surface  $A_{0,j} = \text{Surf}(S_{0,j})$
- une poids associé  $w_{0,j} \geq 0$ , avec :  $\sum_{j=1}^J w_{0,j} = 1$

On peut définir :

$$f_0(x, y) = \sum_{j=1}^J \frac{w_{0,j}}{A_{0,j}} \cdot \mathbb{1}_{S_{0,j}}(x, y)$$

On a bien une densité spatiale sur la France.

Si on a une catégorisation de chaque zone selon un niveau de sismicité, on peut attribuer un facteur  $\lambda_{0,j}$  proportionnel à chaque caté pour obtenir un poids normalisé comme suit :

$$w_{0,j} = \frac{\lambda_{0,j}}{\sum_j \lambda_{0,j}}$$

On a :

$$\begin{aligned} \int_{\Omega} f_0(x, y) dx dy &= \int_{\Omega} \sum_{j=1}^J \frac{w_{0,j}}{A_{0,j}} \cdot \mathbb{1}_{S_{0,j}}(x, y) dx dy = \sum_{j=1}^J \frac{w_{0,j}}{A_{0,j}} \cdot \int_{\Omega} \mathbb{1}_{S_{0,j}}(x, y) dx dy \\ &= \sum_{j=1}^J \frac{w_{0,j}}{A_{0,j}} \cdot A_{0,j} = \sum_{j=1}^J w_{0,j} = 1 \end{aligned}$$

Donc  $f_0(x, y)$  est bien une densité sur  $\Omega$ .

La loi associée  $\mathbb{P}_{X,Y} : \mathcal{B}(\mathbb{R}^2) \rightarrow [0, 1]$  associée à cette densité serait donnée par :

$$\begin{aligned} \mathbb{P}_{X,Y}(B) &= \mathbb{P}((X, Y) \in B) = \int_B f_0(x, y) dx dy \\ &= \sum_{j=1}^J \frac{w_{0,j}}{A_{0,j}} \cdot \int_{B \cap S_{0,j}} dx dy \quad \text{pour tout borélien } B \end{aligned}$$

La loi Normale-Inverse Wishart est une loi jointe sur : la moyenne d'une loi normale multivariée  $\boldsymbol{\mu}$  et la matrice de covariance à cette loi normale multivariée  $\boldsymbol{\Sigma}$ . Cette loi est caractérisée par quatre hyperparamètres :

- $\boldsymbol{\mu}_0$  (vect de dim d) : la moyenne prior sur  $\boldsymbol{\mu}$
- $\lambda_0$  (scalaire positif) : un facteur d'échelle sur la précision de la moyenne
- $\boldsymbol{\Psi}_0$  (matrice d x d, sym et def pos) : un paramètre d'échelle pour la matrice de covariance
- $\nu_0$  (degré de liberté  $> d - 1$ ) : un paramètre qui contrôle la concentration de la loi Inverse Wishart sur  $\boldsymbol{\Sigma}$

La NIW est donnée ainsi :

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{\Psi}_0, \nu_0), \quad \boldsymbol{\mu} | \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}}{\lambda_0})$$

où  $\mathcal{IW}$  est la loi inverse Wishart et  $\boldsymbol{\mu}$  suit une normale multivariée avec covariance  $\boldsymbol{\Sigma}/\lambda_0$ .

La densité de l'Inverse Wishart est :

$$f(\boldsymbol{\Sigma} | \boldsymbol{\Psi}_0, \nu_0) = \frac{|\boldsymbol{\Psi}_0|^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0 d}{2}} \Gamma_d\left(\frac{\nu_0}{2}\right)} |\boldsymbol{\Sigma}|^{-\frac{\nu_0+d+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\Psi}_0 \boldsymbol{\Sigma}^{-1})\right)$$

avec  $\Gamma_d(\cdot)$  la fonction gamma multivariée dim  $d$ ,  $|\cdot|$  le déterminant et  $\text{tr}$  la trace.

La densité de la loi normale conditionnelle est :

$$f\left(\boldsymbol{\mu} | \boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}}{\lambda_0}\right) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}/\lambda_0|^{\frac{1}{2}}} \exp\left(-\frac{\lambda_0}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right)$$

Donc la densité NIW est :

$$\begin{aligned} f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= f(\boldsymbol{\Sigma} | \boldsymbol{\Psi}_0, \nu_0) \cdot f\left(\boldsymbol{\mu} | \boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}}{\lambda_0}\right) \\ &= \frac{|\boldsymbol{\Psi}_0|^{\frac{\nu_0}{2}} \lambda_0^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}} 2^{\frac{\nu_0 d}{2}} \Gamma_d\left(\frac{\nu_0}{2}\right)} |\boldsymbol{\Sigma}|^{-\frac{\nu_0+d+2}{2}} \\ &\quad \times \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\Psi}_0 \boldsymbol{\Sigma}^{-1}) - \frac{\lambda_0}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right) \end{aligned}$$

$\Gamma_d(\cdot)$  est la fonction gamma multivariée dim  $d$ , définie par :

$$\Gamma_d(a) = \pi^{\frac{d(d-1)}{4}} \prod_{i=1}^d \Gamma\left(a + \frac{1-i}{2}\right)$$

Paramètre	Effet/Interprétation
$\mu_0$	<ul style="list-style-type: none"> <li>Moyenne de la loi de <math>\mu</math></li> <li>Plus <math>\lambda_0</math> est grand, plus <math>\mu</math> est concentré autour de <math>\mu_0</math></li> <li>Plus <math>\lambda_0</math> est faible, <math>\mu</math> s'écarte de <math>\mu_0</math></li> </ul>
$\lambda_0$	<ul style="list-style-type: none"> <li>C'est un facteur d'échelle sur la variance de <math>\mu</math></li> <li>Contrôle l'incertitude a priori sur la moyenne <math>\mu</math></li> <li><math>\lambda_0 \rightarrow 0</math> signifie une incertitude infinie sur <math>\mu</math></li> </ul>
$\Psi_0$	<ul style="list-style-type: none"> <li>Contrôle la taille et l'orientation moyennes des matrices de covariance <math>\Sigma</math></li> <li>Plus les valeurs propres de <math>\Psi_0</math> sont grandes, plus les réalisations de <math>\Sigma</math> sont grandes (variances plus larges) et/ou des corrélations plus marquées</li> </ul>
$\nu_0$	<ul style="list-style-type: none"> <li>Contrôle la concentration de la loi de <math>\Sigma</math></li> <li>Doit être strictement supérieur à <math>d - 1</math> pour que la moyenne existe</li> <li>Plus <math>\nu_0</math> est grand, plus <math>\Sigma</math> est concentré autour de sa moyenne <math>\Psi_0/(\nu_0 - d - 1)</math></li> <li>Si <math>\nu_0</math> est faible (proche de <math>d</math>), la dispersion des matrices <math>\Sigma</math> est grande (forte incertitude)</li> </ul>

---

## Bibliographie

---