

# Notes - UQ for Seismic Hazard

---

Ibrahim SEYDI

---

## Table des matières

---

<b>DP for spatial modeling of earthquake</b>	<b>3</b>
<b>Hawkes for Seismic Hazard</b>	<b>32</b>
<b>Bibliographie</b>	<b>88</b>

DP for spatial modeling of earthquake

## 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 :

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

where :

- $\mathcal{N}(\cdot \mid \mu_i, \Sigma_i)$  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  $\theta_i$ .

Liste de papiers pour méthodes zoneless (à filtrer) :

- Woessner et al (2015) The 2013 European Seismic hazard model : key components and results.
- Petersen MD, Harmsen SC, Jaiswal KS, Rukstales KS, Luco N, Haller KM, Mueller CS, Shumway AM (2018) Seismic hazard, risk, and design for south America.
- Helmstetter A, Werner MJ (2012) Adaptive spatiotemporal smoothing of seismicity for long-term earthquake forecasts in California.
- Woo G (1996) Kernel estimation methods for seismic hazard area source modeling.
- S. Molina, C. Lindholm, H. Bungum (2001) Probabilistic seismic hazard analysis : zoning free versus zoning methodology.
- Chethanamba Kempanna Ramanna, G. Dodagoudar (2012) Probabilistic seismic hazard analysis using kernel density estimation technique for Chennai, India.
- S. Lasocki (2021) Kernel Density Estimation in Seismology
- M. Danese, M. Lazzari, B. Murgante (2008) Kernel Density Estimation Methods for a Geostatistical Approach in Seismic Risk Analysis : The Case Study of Potenza Hilltop Town (Southern Italy)
- C. Stock, Euan Smith (2002) Adaptive Kernel Estimation and Continuous Probability Representation of Historical Earthquake Catalogs
- C. Stock, Euan Smith (2002) Comparison of Seismicity Models Generated by Different Kernel Estimations
- G. Estévez-Pérez, H. L. Cimadevila, A. Quintela-del-Río (2002) Nonparametric analysis of the time structure of seismicity in a geographic region
- M. Crespo, F. Martínez, J. Martí (2014) Seismic hazard of the Iberian Peninsula : evaluation with kernel functions
- Francis Tong, Stanisław Lasocki, Beata Orlecka-Sikora (2025) Non-parametric kernel density estimation of magnitude distribution for the analysis of seismic hazard posed by anthropogenic seismicity
- Karaburun, A. ; Demirci, A. (2016) Spatio-temporal cluster analysis of the earthquake epicenters in Turkey and its surrounding area between 1900 and 2014
- Kernel Density Estimation for the Interpretation of Seismic Big Data in Tectonics Using QGIS : The Türkiye–Syria Earthquakes (2023)
-

# 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 :

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

- Comme  $G_0$  encode les connaissances a priori globales. On pourrait utiliser un mélange sur  $G_0$  :

$$G_0(\cdot) = \sum_{j=1}^J \omega_j \cdot G_{0,j}(\cdot), \quad \text{avec } \omega_j \propto \text{connaissance sur zone } j$$

- Rendre  $G$  ou  $G_0$  dépendant de la zone géographique :

$$G \sim DP(\alpha(s), G_0(s))$$

# 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>

On cherche à approximer la distance  $L^2$  entre deux densités  $f$  et  $g$  de  $\mathbb{R}^2$ , définie par :

$$\|f - g\|_{L^2} = \left( \int_{\Omega} (f(x, y) - g(x, y))^2 dx dy \right)^{1/2}$$

où  $\Omega$  est domaine d'étude.

On peut utiliser l'approche des sommes discrètes pour approximer l'intégrale. On utilise la formule d'approximation :

$$\int_{\Omega} (f(x, y) - g(x, y))^2 dx dy \approx \sum_{i,j} (f(x_i, y_j) - g(x_i, y_j))^2 \cdot \Delta x \cdot \Delta y$$

où  $\Delta x, \Delta y$  sont les pas de grille.

- Deux tâches pour l'inférence :
  1. Inférer  $G$  (combien de clusters et caractéristiques des clusters)
  2. Inférer les observations  $x_i$  (dans quel cluster)
- Modèle conjugué → Algo 1, 2 et 3 de Neal (2000) [Du Gibbs]

On cherche à approximer la densité moyenne d'un DPMM avec un prior informatif. L'approximation se fait par moyenne de Monte Carlo sur  $N$  densités générées. Chaque composante  $k$  du mélange est tirée sous une NIW :

$$(\mu_k, \Sigma_k) \sim \text{NIW}(\mu_0^{(j)}, \lambda_0, \Psi_0, \nu_0) \quad \text{où} \quad \mu_0^{(j)} \in \{[0.5, 0.5], [1.5, 0.5], [0.5, 1.5], [1.5, 1.5]\},$$

$$\Psi_0 = \begin{pmatrix} 0.26 & 0 \\ 0 & 0.26 \end{pmatrix}, \quad \lambda_0 = 50.0, \quad \text{et} \quad \nu_0 = 4.$$

Les poids du mélange sont générés via stick-breaking avec paramètre de concentration  $\alpha$  et seuil de troncature  $\tau$  :

$$v_k \sim \text{Beta}(1, \alpha), \quad w_k = v_k \prod_{i=1}^{k-1} (1 - v_i)$$

$$\text{On arrête quand } \prod_{i=1}^k (1 - v_i) < \tau$$

Les poids sont normalisés ensuite :  $\sum_{k=1}^K w_k = 1$

Chaque densité générée s'écrit comme un mélange de normales :

$$f^{(i)}(\cdot) = \sum_{k=1}^{K^{(i)}} w_k^{(i)} \cdot \mathcal{N}(\cdot | \mu_k^{(i)}, \Sigma_k^{(i)})$$

avec les composantes  $(\mu_k^{(i)}, \Sigma_k^{(i)}) \sim \text{NIW}(\mu_0^{(j)}, \lambda_0, \Psi_0, \nu_0)$

Enfin, on calcule la moyenne empirique des densités :

$$\bar{f}_N(\cdot) = \frac{1}{N} \sum_{i=1}^N f^{(i)}(\cdot) = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^{K^{(i)}} w_k^{(i)} \cdot \mathcal{N}(\cdot | \mu_k^{(i)}, \Sigma_k^{(i)})$$

Formule complète de la densité d'un DPMM tronqué sur le carré  $[0, 2]^2$  :

$$f(\cdot) = \sum_{k=1}^K w_k \cdot \mathcal{N}(\cdot | \mu_k, \Sigma_k) \cdot \frac{\mathbb{1}_{[0,2]^2}(\cdot)}{Z_k}$$

où :  $Z_k = \int_{[0,2]^2} \mathcal{N}(u | \mu_k, \Sigma_k) du$

- ‘define\_zonage\_grid( $n\_rows$ ,  $n\_cols$ ,  $x\_range$ ,  $y\_range$ )’ : Crée une partition régulière de l'espace  $\Omega = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$  en  $J = n_{\text{rows}} \times n_{\text{cols}}$  sous-zones. On définit des rectangles  $S_j = [x_j^{(0)}, x_j^{(1)}] \times [y_j^{(0)}, y_j^{(1)}] \subset \Omega$ , pour  $j = 1, \dots, J$ , tels que :

$$\bigcup_{j=1}^J S_j = \Omega \quad \text{et} \quad S_j \cap S_k = \emptyset \text{ pour } j \neq k$$

- ‘compute\_f0\_density( $x$ ,  $y$ ,  $zones$ ,  $weights$ ,  $areas$ )’ : Définit une densité par morceaux  $f_0(x, y)$ , constante sur chaque zone. Soient :

- $w_j$  un poids associé à la zone  $S_j$
- $A_j = \text{aire}(S_j)$
- La densité est définie par :

$$f_0(x) = \frac{\sum_{j=1}^J w_j \mathbb{1}_{S_j}(x)}{\sum_{j=1}^J w_j A_j}$$

et  $f_0(x, y) = 0$  sinon.

- ‘sample\_from\_f0( $n$ ,  $zones$ ,  $weights$ ,  $areas$ )’ : Génère des échantillons selon la loi  $f_0$  définie plus haut.

1. Tirer une zone  $S_j$  avec probabilité :

$$\mathbb{P}(S_j) = \frac{w_j A_j}{\sum_{k=1}^J w_k A_k}$$

2. Tirer  $(X, Y) \sim \mathcal{U}(S_j)$ .

- ‘compute\_zone\_gaussian\_parameters( $zones$ )’ : Approxime chaque zone  $S_j$  par une gaussienne centrée sur son centre de gravité avec une covariance isotrope. On a :

$$-\mu_j = \text{centre de } S_j = \left( \frac{x_j^{(0)} + x_j^{(1)}}{2}, \frac{y_j^{(0)} + y_j^{(1)}}{2} \right)$$

- $\Sigma_j = \sigma_j^2 I_2$ , où  $\sigma_j$  est proportionnel au rayon de la zone :

$$\sigma_j = \frac{\text{diam}(S_j)}{2 \cdot 1.96} \Rightarrow \Sigma_j = \sigma_j^2 \cdot I$$

- ‘compute\_f0tilde\_density(x, y, mus, covariances, weights)’ : Calcul la densité d’un mélange de lois normales pondérées. On a :

$$\tilde{f}_0(x, y) = \sum_{j=1}^J w_j \cdot \mathcal{N}((x, y) | \mu_j, \Sigma_j)$$

où  $\mathcal{N}(\cdot | \mu_j, \Sigma_j)$  est la densité de la gaussienne centrée en  $\mu_j$  avec covariance  $\Sigma_j$ .

- ‘sample\_from\_f0tilde(n, mus, covariances, weights, areas)’ : Génère des échantillons suivant la loi  $\tilde{f}_0$ , selon le processus suivant :

1. Tirer un indice  $j \in \{1, \dots, J\}$  avec probabilité :

$$\mathbb{P}(j) = \frac{w_j A_j}{\sum_{k=1}^J w_k A_k}$$

2. Tirer un échantillon  $(X, Y) \sim \mathcal{N}(\mu_j, \Sigma_j)$

## Développement calcul de la matrice de covariance $\Sigma_j$

Soit, chaque zone  $S_j$  est un rectangle de forme :

$$S_j = [x_j^{(0)}, x_j^{(1)}] \times [y_j^{(0)}, y_j^{(1)}]$$

On approxime la densité sur cette zone par une loi normale bidimensionnelle centrée au centroïde de la zone, et avec une matrice de covariance diagonale  $\Sigma_j = \sigma_j^2 I$ , où  $I$  est la matrice identité  $2 \times 2$ .

Le diamètre est la distance maximale entre deux points dans la zone rectangulaire  $S_j$ . Ici, comme la zone est rectangulaire, ce diamètre correspond à la diagonale :

$$\text{diam}(S_j) = \sqrt{(x_j^{(1)} - x_j^{(0)})^2 + (y_j^{(1)} - y_j^{(0)})^2}$$

L’enjeu ici est d’approximer la densité uniforme sur  $S_j$  par une densité normale  $\mathcal{N}(\mu_j, \Sigma_j)$ , et de calibrer  $\Sigma_j$  pour que cette gaussienne couvre 95% de la masse dans la zone. À cet objectif, on choisit un écart-type  $\sigma_j$  tel que le disque de rayon  $\text{diam}(S_j)/2$  corresponde à environ 1.96 écarts-types (quantile pour avoir une confiance à 95% :  $\mathbb{P}(1.96 < Z < 1.96) \approx 0.95$  pour  $Z \sim \mathcal{N}(0, 1)$  ; quantile pour confiance à 99% 2.576).

Ainsi, on a :

$$\sigma_j = \frac{\text{diam}(S_j)}{2 \cdot 1.96}$$

On suppose que la covariance est isotrope, donc diagonale proportionnelle à l'identité :  $\Sigma_j = \sigma_j^2 \cdot I$ .

Autrement dit :

$$\Sigma_j = \left( \frac{\text{diam}(S_j)}{2 \cdot 1.96} \right)^2 \cdot I$$

**Application dans cas jouet :** Soit zone  $S_j = [0, 1] \times [0, 1]$ . On a :

- Le diamètre :  $\text{diam}(S_j) = \sqrt{1^2 + 1^2} = \sqrt{2}$
- L'écart-type :  $\sigma_j = \frac{\sqrt{2}}{2 \cdot 1.96}$
- La matrice de covariance est donc :

$$\Sigma_j = \left( \frac{\sqrt{2}}{2 \cdot 1.96} \right)^2 \cdot I \approx (0.255)^2 \cdot I \approx 0.065 \cdot I$$

$Z$  loi normale standard :  $\mathbb{P}(-1.96 \leq Z \leq 1.96) \approx 0.95$

95% de la masse d'une loi normale univariée est contenue dans l'intervalle  $[-1.96, +1.96]$ .

On veut approximer une densité uniforme sur une zone rectangulaire  $S_j$  par une densité normale centrée au centre de la zone. On veut que cette normale couvre environ 95%. On choisit alors un écart-type  $\sigma_j$  tel que :

$$\text{Rayon} = 1.96 \cdot \sigma_j \quad \Rightarrow \quad \sigma_j = \frac{\text{diam}(S_j)}{2 \cdot 1.96}$$

**Code Julien - Étude de la capacité à bien approcher un pavage uniforme par des gaussiennes :**

- $gmm\_em(X, G, mu\_init, sigma\_init = NULL, max\_iter = 100, tol = 1e-6)$  : Estimation des paramètres d'un mélange de  $G$  lois normales multivariées à partir d'un jeu de données  $X$ , à l'aide de l'algorithme EM.

Paramètres d'entrée :

- $X$  : matrice  $n \times d$  de données (n observations, d variables)
- $G$  : nombre de composantes du mélange
- $mu\_init$  : matrice initiale  $G \times d$  des moyennes

- sigma\_init : tableau  $d \times d \times G$  des matrices de covariance initiales
- max\_iter : nombre maximal d'itérations
- tol : tolérance pour l'arrêt basé sur la variation du log-vraisemblance

Grandes lignes :

1. Initialisation ( $n$  : nombre d'observations ;  $d$  : dim des données ;  $\mu_0$  : initialisation des moyennes ;  $p_{ik}$  : proportions initiales (toutes égales))
2. Initialisation des matrices de covariance (Si sigma\_init est NULL → utiliser la matrice de covariance globale de X pour toutes les composantes)
3. **Algo EM** : Pour chaque itération jusqu'à max\_iter :
  - **E-Step** : Calcule  $\gamma_{ik}$ , i.e. la proba que l'observation  $x_i$  provienne de la composante  $k$  :
$$\gamma_{ik} = \frac{\pi_k \cdot \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_{j=1}^G \pi_j \cdot \mathcal{N}(x_i | \mu_j, \Sigma_j)}$$
  - **M-Step** : Met à jour les paramètres  $\pi_k, \mu_k, \Sigma_k$  à partir de  $\gamma$  :
    - \*  $N_k = \sum_{i=1}^n \gamma_{ik} \rightarrow \pi_k = N_k/n$
    - \*  $\mu_k = \frac{1}{N_k} \sum_{i=1}^n \gamma_{ik} x_i$
    - \*  $\Sigma_k = \frac{1}{N_k} \sum_{i=1}^n \gamma_{ik} (x_i - \mu_k)(x_i - \mu_k)^T$
  - **Calcul log-vraisemblance** :

$$\log L = \sum_{i=1}^n \log \left( \sum_{k=1}^G \pi_k \cdot \mathcal{N}(x_i | \mu_k, \Sigma_k) \right)$$

Si la variation de la log-vraisemblance est inférieure à tol, on stop.

4. La fonction renvoie une liste avec les paramètres estimés :  $G, \pi_k, \mu, \Sigma, \log L$

- *rgmm(n, prob, mean, sigma)* : Génère n points aléatoires selon un mélange normales multivariées
- *dgmm(x, prob, mean, sigma)* : Calcule la densité totale d'un mélange gaussien multivarié
- *d\_mar\_gmm(x, mar, prob, mean, sigma)* : Calcul la densité marginale (1d)
- *stick\_breaking(alpha, tau = 1e-3)* : SB
- *g\_0(prob, mean, sigma, nu\_0, lambda\_0)* : Tire un param de composante  $(\mu, \Sigma)$  à partir d'un DP
- *sample\_f(alpha, prob, mean, sigma, nu\_0, lambda\_0, tau = 1e-3)* : Simule DP

Il est important d'adopter une vision plus large et globale lorsque je code. Mon code doit être à la fois adaptable et facilement débogable. Il paraît pertinent de viser une uniformisation maximale dans la structure des fonctions. Par ailleurs, écrire mathématiquement ce que je veux implémenter, semble être une approche judicieuse (surtout d'un point de vue longtermiste).

---

We are pleased to present the Epos-France database, which includes more than 16,000 high-quality seismic records from seismological stations in mainland France. They correspond to earthquakes with local magnitude between 2 and 5.6 that occurred between 1996 and the end of 2021. (Sur site SIGMA <https://sigma-programs.com/newsletter-1-february-2025-copy-2/>)

---

Un processus de Dirichlet dépendant est une extension du DP qui permet à la distribution aléatoire de varier en fonction d'un ensemble de covariables ou d'indices (par exemple, l'espace, le temps, ou d'autres facteurs). Autrement dit, au lieu d'avoir une unique mesure aléatoire  $G \sim DP(\alpha, G_0)$ , on considère une famille de mesures  $\{G_x : x \in \mathcal{X}\}$  indexées par  $x$ , où chaque  $G_x$  suit une distribution similaire à un DP, mais dépendant de  $x$ . Cela permet de modéliser la dynamique ou la variation locale des distributions. Par exemple, dans un contexte spatio-temporel, on pourrait avoir une distribution de points (ou d'événements) qui varie selon la localisation géographique ou le moment.

Modélisation adaptative de la densité spatiale : un DDP pourrait modéliser la distribution spatiale des événements de manière flexible, permettant à la forme de la densité de changer localement selon la position géographique ou d'autres covariables (ex. caractéristiques géologiques).

---

**'compute\_zone\_gaussian\_parameters'** : Approxime mélange d'uniforme  $X$  par un mélange de gaussiennes initialisé avec k-means, afin d'estimer : les centroïdes et les covariances (et éventuellement les poids) de chaque composante gaussienne.

1. Initialisation centroïdes avec KMeans
2. Initialisation covariances : *GaussianMixture* accepte l'initialisation via précisions (inverses des covariances), d'où utilisation de *pinv* donne version robuste de l'inverse (pseudo-inverse de Moore-Penrose)

3. Entrainement GMM (par algo EM)
  4. Renvoie des centroïdes et covariances estimés
- 

**Objectif** : Construire une loi  $\text{NIW}(\mu_0, \lambda_0, \Psi_0, \nu_0)$  telle que :

- la moyenne a priori de la distribution  $\mu \sim \mathcal{N}(\mu_0, \Sigma/\lambda_0)$  soit proche d'un centroïde estimé  $\mu^*$  (supposé connu à ce stade),
- la valeur centrale de  $\Sigma \sim \text{Inv-Wishart}(\Psi_0, \nu_0)$  soit proche d'une matrice covariance estimée  $\Sigma^*$  (supposée connue à ce stade).

**Caractéristiques de la NIW** : Soit  $(\mu, \Sigma) \sim \text{NIW}(\mu_0, \lambda_0, \Psi_0, \nu_0)$ , alors :

$$\begin{aligned}\Sigma &\sim \mathcal{IW}(\Psi_0, \nu_0) \\ \mu \mid \Sigma &\sim \mathcal{N}(\mu_0, \Sigma/\lambda_0)\end{aligned}$$

Moyennes a priori :

$$\mathbb{E}[\mu] = \mu_0 \quad ; \quad \mathbb{E}[\Sigma] = \frac{\Psi_0}{\nu_0 - d - 1} \quad \text{si } \nu_0 > d + 1 \text{ (en dim d = 2 ici)}$$

**Construction à partir de  $\mu^*$  et  $\Sigma^*$  - Étapes :**

- Choisir  $\mu_0 = \mu^*$ , on place la moyenne de la NIW exactement sur l'estimation empirique du centroïde
- Fixer  $\nu_0 > d + 1$  (e.g.  $\nu_0 = d + 3 = 5$  pour rester modéré)
- Fixer  $\Psi_0 = \Sigma^* \cdot (\nu_0 - d - 1)$ . De sorte que :

$$\mathbb{E}[\Sigma] = \frac{\Psi_0}{\nu_0 - d - 1} = \Sigma^*$$

- Fixer  $\lambda_0$  selon la confiance en la moyenne  $\mu_0$ . Si on veut un prior fortement concentré autour de  $\mu_0$ , on prend un  $\lambda_0$  élevé. Si on veut quelque chose de plus vague :  $\lambda_0 \approx 0.01$ .

Ainsi, on construit un prior NIW centré sur une gaussienne empirique, avec une force de concentration ajustable via  $\alpha$ ,  $\lambda_0$  et  $\nu_0$ .

---

## Objective

Let  $\mathcal{X} \subset \mathbb{R}^2$ . We aim to estimate an unknown density  $f : \mathcal{X} \rightarrow \mathbb{R}_+$  using a nonparametric Bayesian model based on a Dirichlet Process Mixture Model (DPMM), defined as :

$$f(x) = \int \mathcal{N}(x | \mu, \Sigma) dG(\mu, \Sigma)$$

where  $G \sim \text{DP}(\alpha, G_0)$  is a Dirichlet process with concentration parameter  $\alpha > 0$  and base distribution  $G_0$ .

### Stick-Breaking Approximation

The Dirichlet process is approximated using stick-breaking construction :

$$\begin{aligned}\beta_k &\sim \text{Beta}(1, \alpha) \\ \pi_k &= \beta_k \prod_{j=1}^{k-1} (1 - \beta_j), \quad k = 1, 2, \dots \\ \theta_k &\sim G_0, \quad \text{with } \theta_k = (\mu_k, \Sigma_k)\end{aligned}$$

Truncating the process at a small threshold  $\tau \ll 1$ , we obtain :

$$G \approx \sum_{k=1}^K \pi_k \delta_{\theta_k}, \quad \text{with } \sum_{k=1}^K \pi_k \approx 1$$

### Choice of the concentration parameter $\alpha$

Work in progress ...

### Choice of the Prior $G_0$

#### Non-informative Prior

A standard Normal-Inverse-Wishart prior can be used :

$$G_0 = \mathcal{NIW}(\mu_0, \lambda_0, \Psi_0, \nu_0)$$

which implies :

$$\begin{aligned}\Sigma &\sim \mathcal{W}^{-1}(\Psi_0, \nu_0) \\ \mu \mid \Sigma &\sim \mathcal{N}(\mu_0, \Sigma / \lambda_0)\end{aligned}$$

### Informative Prior (Zonage)

In an informative prior setting, we exploit knowledge about the spatial structure of the domain. We divide the input space (in our simplified case we have taken  $[0, 2] \times [0, 2]$ ) into a regular or irregular grid of  $n_{\text{rows}} \times n_{\text{cols}}$  rectangular zones. Let the zones be denoted as  $\{Z_j\}_{j=1}^{J_0}$  with corresponding areas  $A_j$  and weights  $w_j$ . We can define a zonal uniform density  $f_0$  as :

$$f_0(x) = \sum_{j=1}^{J_0} \frac{w_j}{A_j} \mathbb{1}_{x \in Z_j}$$

From this piecewise-constant density  $f_0$ , we generate synthetic data and fit a Gaussian Mixture Model using an EM algo. Initialization is performed via KMeans to provide stable estimates of the components means. The fitted GMM yields a smoothed approximation  $\tilde{f}_0$  of the zonal density :

$$\tilde{f}_0(x) = \sum_{j=1}^J \tilde{w}_j \cdot \mathcal{N}(x | \tilde{\mu}_j, \tilde{\Sigma}_j)$$

We then define an informative prior  $G_0^{\text{inf}}$  using the estimated GMM parameters :

- The component means  $\tilde{\mu}_j$  are used as prior centers  $\mu_j$
- The covariance matrices  $\tilde{\Sigma}_j$  are multiplied by  $\nu_0 - d - 1$  to define the NIW scale matrices  $\Psi_j$
- The estimated weights  $\tilde{w}_j$  are used as discrete probas over the base

Thus,  $G_0^{\text{inf}}$  is a mixture of NIW components written as :

$$G_0^{\text{inf}} = \sum_{j=1}^J \tilde{w}_j \cdot \mathcal{NIW}(\mu_j, \lambda_0, \Psi_j, \nu_0)$$

This informative base measure allows the Dirichlet process to generate more plausible components in regions of high prior mass, effectively incorporating structured domain knowledge into the prior.

### Posterior predictive density

The resulting posterior predictive density becomes :

$$f(x) \approx \sum_{k=1}^K \pi_k \cdot \mathcal{N}(x | \mu_k, \Sigma_k)$$

## Cadre

Soit  $(X_1, X_2, \dots, X_d)$  un vecteur aléatoire de dimension  $d$  défini sur un espace  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_d$  avec loi jointe  $\pi(x_1, \dots, x_d)$  connue à une constante multiplicative près :

$$\pi(x_1, \dots, x_d) = \frac{\tilde{\pi}(x_1, \dots, x_d)}{Z}$$

où  $\tilde{\pi}$  est connue et  $Z = \int_{\mathcal{X}} \tilde{\pi}(x) dx$  est inconnu.

**Objectif :** Générer une suite  $\{X^{(t)}\}_{t \geq 0}$  telle que la loi stationnaire soit  $\pi$ , afin d'estimer des espérances

$$\mathbb{E}_{\pi}[h(X)] \approx \frac{1}{T} \sum_{t=1}^T h(X^{(t)}).$$

**Idée clé du Gibbs sampling :** Plutôt que d'échantillonner directement selon  $\pi$ , on utilise les lois conditionnelles complètes :

$$\pi(x_i | x_{-i}) \quad \text{où} \quad x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d).$$

Hypothèse essentielle :

$\pi(x_i | x_{-i})$  est simulable explicitement pour tout  $i$ .

## Algo de Gibbs

Soit  $X^{(t)} = (X_1^{(t)}, \dots, X_d^{(t)})$  l'état à l'itération  $t$ .

1. Initialisation : choisir  $X^{(0)} \in \mathcal{X}$  arbitrairement (ou selon une loi de départ)
2. Itération : pour  $t = 0, 1, \dots$   
Pour  $i = 1, \dots, d$  :

$$X_i^{(t+1)} \sim \pi(x_i | X_1^{(t+1)}, \dots, X_{i-1}^{(t+1)}, X_{i+1}^{(t)}, \dots, X_d^{(t)})$$

Autrement dit :

- On balaye les coordonnées une par une
- Pour la coordonnée  $i$ , on utilise les valeurs déjà mises à jour pour  $1, \dots, i-1$  et les valeurs de l'itération précédente pour  $i+1, \dots, d$

## Propriétés théoriques

### Chaîne de Markov

La suite  $\{X^{(t)}\}$  est une chaîne de Markov sur  $\mathcal{X}$  avec noyau de transition

$$K(x, dy) = \prod_{i=1}^d \pi(y_i | y_{1:i-1}, x_{i+1:d}) dy_i$$

où  $y_{1:i-1} = (y_1, \dots, y_{i-1})$  et  $x_{i+1:d} = (x_{i+1}, \dots, x_d)$ .

## Invariance

$\pi$  est loi invariante pour  $K$  :

$$\int_{\mathcal{X}} \pi(x) K(x, dy) dx = \pi(dy).$$

Cela vient du fait que l'échantillonnage séquentiel selon les conditionnelles reproduit la loi jointe.

## Réversibilité

Le noyau  $K$  est  $\pi$ -réversible :

$$\pi(dx) K(x, dy) = \pi(dy) K(y, dx).$$

Donc  $\pi$  est bien stationnaire.

## Convergence

Si la chaîne est :

- irréductible (toute région de  $\mathcal{X}$  atteignable)
- apériodique (si la chaîne n'évolue pas en cycles fixes)
- et positive récurrente (récurrente : la proba de revisiter chaque état est 1 ; positive récurrente : le temps moyen de retour de chaque état est fini)

alors :

$$\lim_{t \rightarrow \infty} \mathcal{L}(X^{(t)}) = \pi.$$

On a convergence en loi.

## Version bloc (blocked Gibbs)

On peut regrouper certaines coordonnées en blocs  $X = (Y_1, \dots, Y_m)$  et mettre à jour chaque bloc selon sa loi conditionnelle jointe :

$$Y_j^{(t+1)} \sim \pi(y_j \mid Y_{-j}).$$

Avantage : réduit la corrélation entre itérations quand les variables sont fortement dépendantes.

## Conjugaison de la NIW à gaussienne multivariée

On considère un vecteur aléatoire  $\mathbf{x} \in \mathbb{R}^d$  tel que :

$$\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

On suppose un échantillon i.i.d.  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , et on cherche  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . On place une loi normale-inverse-Wishart a priori :

$$\begin{aligned}\boldsymbol{\Sigma} &\sim \mathcal{IW}(\nu_0, \boldsymbol{\Psi}_0) \\ \boldsymbol{\mu} | \boldsymbol{\Sigma} &\sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\lambda_0)\end{aligned}$$

Les densités sont :

$$\begin{aligned}p(\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-(\nu_0+d+1)/2} \exp\left(-\frac{1}{2} \text{Tr}(\boldsymbol{\Psi}_0 \boldsymbol{\Sigma}^{-1})\right) \\ p(\boldsymbol{\mu} | \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{\lambda_0}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right)\end{aligned}$$

Donc :

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu_0+d+2)/2} \exp\left(-\frac{1}{2} [\text{Tr}(\boldsymbol{\Psi}_0 \boldsymbol{\Sigma}^{-1}) + \lambda_0 (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)]\right)$$

La vraisemblance des données est :

$$p(\mathbf{x}_{1:n} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^n \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})\right)$$

En posant  $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$  et  $\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$ , on réécrit :

$$\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) = \text{Tr}(\mathbf{S} \boldsymbol{\Sigma}^{-1}) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

Ainsi, on a :

$$p(\mathbf{x}_{1:n} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2} [\text{Tr}(\mathbf{S} \boldsymbol{\Sigma}^{-1}) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})]\right)$$

## Calcul postérieur

Le calcul du posterior est :

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{x}_{1:n}) \propto p(\mathbf{x}_{1:n} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \cdot p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu_0+n+d+2)/2} \exp\left(-\frac{1}{2} Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})\right)$$

avec :

$$Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}) = \text{Tr}((\boldsymbol{\Psi}_0 + \mathbf{S}) \boldsymbol{\Sigma}^{-1}) + \lambda_0 (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) + n(\boldsymbol{\mu} - \bar{\mathbf{x}})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{x}})$$

En regroupant les termes quadratiques en  $\boldsymbol{\mu}$ , on a :

$$\lambda_n = \lambda_0 + n, \quad \boldsymbol{\mu}_n = \frac{\lambda_0 \boldsymbol{\mu}_0 + n \bar{\mathbf{x}}}{\lambda_0 + n}$$

$$\lambda_0(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^2 + n(\boldsymbol{\mu} - \bar{\mathbf{x}})^2 = \lambda_n(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^2 + \frac{\lambda_0 n}{\lambda_n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^2$$

En écrivant :

$$\begin{aligned}\boldsymbol{\Psi}_n &= \boldsymbol{\Psi}_0 + \mathbf{S} + \frac{\lambda_0 n}{\lambda_0 + n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \\ \nu_n &= \nu_0 + n\end{aligned}$$

En conclusion, on a :

$$\begin{aligned}p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{x}_{1:n}) &\propto |\boldsymbol{\Sigma}|^{-(\nu_n+d+2)/2} \exp\left(-\frac{1}{2} [\text{Tr}(\boldsymbol{\Psi}_n \boldsymbol{\Sigma}^{-1}) + \lambda_n(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)]\right) \\ &\sim \mathcal{NIW}(\boldsymbol{\mu}_n, \lambda_n, \nu_n, \boldsymbol{\Psi}_n)\end{aligned}$$

Ainsi, la loi NIW est conjuguée à la loi normale multivariée.

## Loi Student

La loi de Student t est une loi de proba continue, symétrique autour de 0, similaire à une gaussienne, mais avec des queues plus épaisses. Elle apparaît naturellement dans le contexte suivant :

- On a un échantillon  $X_1, \dots, X_n$  tiré d'une population normale avec moyenne  $\mu$  et variance  $\sigma^2$ .
- La variance  $\sigma^2$  est inconnue.
- On veut construire une statistique centrée et normalisée :

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

où :

- $\bar{X}$  : moyenne
- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  : variance

Alors, cette statistique  $t$  suit exactement une loi de Student à  $n - 1$  degrés de liberté. La loi de Student t est caractérisée par :

1.  $\nu$  : degrés de liberté ; Typiquement  $\nu = n - 1$  pour un échantillon de taille  $n$ . Plus  $\nu$  est grand, plus la loi t ressemble à une normale standard.

2. Moyenne et variance :  $\mu$  et  $\sigma^2$ ; Moyenne = 0 si  $\nu > 1$  et Variance =  $\nu/(\nu - 2)$  si  $\nu > 2$

Les queues plus épaisses reflètent l'incertitude supplémentaire due à l'estimation de la variance à partir de l'échantillon. La loi t s'élargit pour refléter l'incertitude sur la variance. Si on avait la variance exacte, on aurait juste une normale. À partir d'une certaine valeur de degré de liberté, la student est assimilable à une gaussienne.

La densité de la loi t pour  $x \in \mathbb{R}$  et  $\nu$  degrés de liberté :

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

où  $\Gamma$  est la fonction gamma.

Quand  $\nu \rightarrow \infty$ , t converge vers la normale standard  $N(0, 1)$ .

## Apparition de la Student

On considère :

$$\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \mathcal{NIW}(\boldsymbol{\mu}_0, \lambda_0, \boldsymbol{\Psi}_0, \nu_0)$$

On veut la distribution prédictive marginale  $p(\mathbf{x})$ , c'est-à-dire la distribution de  $\mathbf{x}$  après avoir intégré  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  :

$$p(\mathbf{x}) = \int \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathcal{NIW}(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{\mu}_0, \lambda_0, \boldsymbol{\Psi}_0, \nu_0) d\boldsymbol{\mu} d\boldsymbol{\Sigma}$$

### Intégration sur $\boldsymbol{\mu}$ conditionnellement à $\boldsymbol{\Sigma}$

On a :

$$\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\lambda_0)$$

Du coup, la convolution avec la normale de  $\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}$  donne :

$$\mathbf{x} | \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}_0, (1 + 1/\lambda_0)\boldsymbol{\Sigma}) = \mathcal{N}\left(\boldsymbol{\mu}_0, \frac{\lambda_0 + 1}{\lambda_0} \boldsymbol{\Sigma}\right)$$

### Intégration sur $\boldsymbol{\Sigma}$

On a :

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\nu_0, \boldsymbol{\Psi}_0)$$

La distribution de  $\mathbf{x}$  après avoir intégré  $\Sigma$  devient une Student-t multivariée :

$$\mathbf{x} \sim t_{\nu_0-d+1}(\boldsymbol{\mu}_0, \frac{\lambda_0 + 1}{\lambda_0(\nu_0 - d + 1)} \boldsymbol{\Psi}_0)$$

avec :

- $\nu = \nu_0 - d + 1$  : degrés de liberté
- $\boldsymbol{\mu}_0$  : centre de la Student-t
- $\boldsymbol{\Sigma}_t = \frac{\lambda_0 + 1}{\lambda_0(\nu_0 - d + 1)} \boldsymbol{\Psi}_0$  : matrice de dispersion

Dans l'algorithme Gibbs pour l'inférence du DPMM :

- Pour un nouveau cluster, la predictive est donnée par :

$$\mathbf{x}_i \mid \boldsymbol{\mu}_0, \lambda_0, \boldsymbol{\Phi}_0, \nu_0 \sim t_{\nu_0-d+1}(\boldsymbol{\mu}_0, \frac{\lambda_0 + 1}{\lambda_0(\nu_0 - d + 1)} \boldsymbol{\Psi}_0)$$

C'est ce qu'on utilise pour calculer les poids prédictifs  $t_{\text{new}}$ .

On veut la distribution prédictive  $p(\mathbf{x} \mid G_0^{in})$ , c'est-à-dire la distribution de  $\mathbf{x}$  après avoir intégré  $(\boldsymbol{\mu}, \Sigma)$  :

$$p(\mathbf{x}) = \int \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) \cdot \sum_{j=1}^J \tilde{w}_j \mathcal{NIW}(\boldsymbol{\mu}, \Sigma \mid \boldsymbol{\mu}_j, \lambda_j, \boldsymbol{\Psi}_j, \nu_j) d\boldsymbol{\mu} d\Sigma$$

On décompose :

$$\mathcal{NIW}(\boldsymbol{\mu}, \Sigma \mid \boldsymbol{\mu}_j, \lambda_j, \boldsymbol{\Psi}_j, \nu_j) = \mathcal{N}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_j, \frac{1}{\lambda_j} \Sigma) \cdot \mathcal{IW}(\Sigma \mid \boldsymbol{\Psi}_j, \nu_j).$$

Les densités étant positives par Fubini-Tonelli, on échange somme et intégrale :

$$p(\mathbf{x}) = \sum_{j=1}^J \tilde{w}_j \int \int \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) \mathcal{N}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_j, \frac{1}{\lambda_j} \Sigma) \mathcal{IW}(\Sigma \mid \boldsymbol{\Psi}_j, \nu_j) d\boldsymbol{\mu} d\Sigma.$$

### Intégration sur $\boldsymbol{\mu}$

Conditionnellement à  $\Sigma$ , on a (**résultat à étayer**) :

$$\int \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) \mathcal{N}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_j, \frac{1}{\lambda_j} \Sigma) d\boldsymbol{\mu} = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_j, \frac{\lambda_j + 1}{\lambda_j} \Sigma).$$

Donc :

$$p(\mathbf{x}) = \sum_{j=1}^J \tilde{w}_j \int \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_j, \frac{\lambda_j + 1}{\lambda_j} \Sigma) \mathcal{IW}(\Sigma \mid \boldsymbol{\Psi}_j, \nu_j) d\Sigma$$

## Intégration sur $\Sigma$

L'intégrale d'une normale conditionnelle en  $\Sigma$  contre une inverse-Wishart donne une Student multivariée (**résultat à étayer**) :

$$\int \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_j, \frac{\lambda_j+1}{\lambda_j} \boldsymbol{\Sigma}) \mathcal{IW}(\boldsymbol{\Sigma} | \boldsymbol{\Psi}_j, \nu_j) d\boldsymbol{\Sigma} = t_{\nu_j-d+1}(\mathbf{x} | \boldsymbol{\mu}_j, \frac{\lambda_j+1}{\lambda_j(\nu_j-d+1)} \boldsymbol{\Psi}_j), \quad \nu_j > d-1$$

## Résultat

Au final, on a :

$$p(\mathbf{x}) = \sum_{j=1}^J \tilde{w}_j t_{\nu_j-d+1}(\mathbf{x} | \boldsymbol{\mu}_j, \frac{\lambda_j+1}{\lambda_j(\nu_j-d+1)} \boldsymbol{\Psi}_j).$$

Dans l'algorithme Gibbs pour l'inférence du DPMM :

- Pour un nouveau cluster, la predictive est donnée par :

$$\mathbf{x}_i | \boldsymbol{\mu}_0, \lambda_0, \boldsymbol{\Psi}_0, \nu_0 \sim \sum_{j=1}^J \tilde{w}_j t_{\nu_j-d+1}(\mathbf{x} | \boldsymbol{\mu}_j, \frac{\lambda_j+1}{\lambda_j(\nu_j-d+1)} \boldsymbol{\Psi}_j)$$

C'est ce qu'on utilise pour calculer les poids prédictifs  $t_{\text{new}}$ .

## Modèle d'étude

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

$$\begin{aligned} x_i | \theta_i &\sim F(x_i | \theta_i), \quad F(\cdot | \theta_i) = \sum_{k=1}^{\infty} \pi_k \mathcal{N}(\cdot | \mu_k, \Sigma_k) \\ \theta_i | G &\stackrel{i.i.d.}{\sim} G \\ G &\sim \text{DP}(\alpha, G_0) \\ G_0 &= \mathcal{NIW}(m_0, \kappa_0, \boldsymbol{\Psi}_0, \nu_0) \end{aligned}$$

Dans le cas prior informatif :

$$G_0^{inf} = \sum_{j=1}^J \omega_j \mathcal{NIW}(m_{0j}, \kappa_{0j}, \boldsymbol{\Psi}_{0j}, \nu_{0j})$$

Reformulation avec variables latentes :

$$\begin{aligned}
 x_i \mid \phi, c_i &\sim F(x_i \mid \phi_{c_i}), \quad F(\cdot \mid \phi_{c_i}) = \sum_{k=1}^{\infty} \pi_k f(\cdot \mid \phi_{c_k}) \\
 c_i \mid \boldsymbol{\pi} &\sim \text{Discrete}(\pi_1, \dots, \pi_K) \\
 \phi_c &\sim G_0 \\
 \pi_k &= \beta_k \prod_{l=1}^{k-1} (1 - \beta_l) = \beta_k \left( 1 - \sum_{l=1}^{k-1} \pi_l \right) \\
 \beta_k &\sim \text{Beta}(1, \alpha)
 \end{aligned}$$

On a :

- $c_i \in \{1, 2, \dots\}$  : indice du cluster auquel appartient  $x_i$ .
- $(\mu_k, \Sigma_k)$  : paramètres du cluster  $k$

## Éléments de l'algorithme de Gibbs

### (I) Échantillonnage des clusters $c_i$

- Pour chaque cluster existant  $k$  avec  $i-1$  observations. On pose les stats suffisantes suivantes :  $\bar{x}$  et  $S = \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^\top$ . On a

$$p(c_i = k \mid c_{-i}, \mu_k, \Sigma_k, \alpha, x_i) \propto n_{k,-i} \cdot p(x_i \mid \mu_k, \Sigma_k) \quad (1)$$

où  $p(x_i \mid \mu_k, \Sigma_k)$  est la vraisemblance :

$$p(x_i \mid \mu_k, \Sigma_k) = \mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$

- (Non informatif) Pour un nouveau cluster, on a :

$$p(c_i \neq c_j, \forall j \neq i \mid c_{-i}, \alpha, x_i, \boldsymbol{\mu}_0, \lambda_0, \boldsymbol{\Phi}_0, \nu_0) \propto \alpha \cdot p(x_i \mid G_0) \quad (2)$$

$p(x_i \mid G_0)$  est la prior predictive où :

$$p(x_i \mid G_0) = t_{\nu_0-d+1}(\boldsymbol{\mu}_0, \frac{\lambda_0 + 1}{\lambda_0(\nu_0 - d + 1)} \boldsymbol{\Psi}_0)$$

- (Informatif) Pour un nouveau cluster, on a :

$$p(c_i \neq c_j, \forall j \neq i \mid c_{-i}, \alpha, x_i, \boldsymbol{\mu}_0, \lambda_0, \boldsymbol{\Phi}_0, \nu_0) \propto \alpha \cdot p(x_i \mid G_0) \quad (3)$$

$p(x_i \mid G_0)$  est la prior predictive où :

$$p(x_i \mid G_0) = \sum_{j=1}^J \tilde{w}_j \cdot t_{\nu_j-d+1}(\mathbf{x} \mid \boldsymbol{\mu}_j, \frac{\lambda_j + 1}{\lambda_j(\nu_j - d + 1)} \boldsymbol{\Psi}_j)$$

Par ailleurs, pour  $x \in \mathbb{R}^d$ ,  $\nu > 0$  degrés de liberté, un centre  $\mu \in \mathbb{R}^d$  et une matrice d'échelle  $\Sigma \in \mathbb{R}^{d \times d}$ , la densité de la Student multivariée  $t_\nu(\mu, \Sigma)$  est :

$$t_\nu(x | \mu, \Sigma) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{d/2} |\Sigma|^{1/2}} \left[ 1 + \frac{1}{\nu}(x - \mu)^\top \Sigma^{-1}(x - \mu) \right]^{-\frac{\nu+d}{2}}.$$

où  $\Gamma(\cdot)$  est la fonction gamma,  $\mu$  est le vecteur moyenne et la matrice de variance-covariance est  $\text{Cov}[X] = \frac{\nu}{\nu-2} \Sigma$ .

## (II) Échantillonnage des paramètres de cluster $(\mu_k, \Sigma_k)$

Supposons un cluster  $c_i = k$ . On pose les stats suffisantes suivantes :

- $n_k = |c_i|$
- $\bar{x}_k = \frac{1}{n_k} \sum_{x \in c_i} x$
- $S_k = \sum_{x \in c_i} (x - \bar{x}_k)(x - \bar{x}_k)^\top$ .

En écrivant :

$$\lambda_k = \lambda_0 + n_k \tag{4}$$

$$\boldsymbol{\mu}_k = \frac{\lambda_0 \boldsymbol{\mu}_0 + n_k \bar{x}_k}{\lambda_0 + n_k} \tag{5}$$

$$\nu_k = \nu_0 + n_k \tag{6}$$

$$\Psi_k = \Psi_0 + S_k + \frac{\lambda_0 n_k}{\lambda_0 + n_k} (\bar{x}_k - \boldsymbol{\mu}_0)(\bar{x}_k - \boldsymbol{\mu}_0)^\top \tag{7}$$

Le posterior est donc donné par :

$$\Sigma_k \sim \mathcal{IW}(\nu_k, \boldsymbol{\Psi}_k), \quad \boldsymbol{\mu}_k | \Sigma_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k / \lambda_k) \tag{8}$$

## (III) Échantillonnage du paramètre de concentration $\alpha$

Work in progress ...

---

### Abstract pour Luis et sismos

Probabilistic Seismic Hazard Assessment (PSHA) is traditionally based on piecewise spatially homogeneous and temporally homogeneous Poisson process assumptions. While this framework provides a clear and structured methodology, it tends to oversimplify the complex spatial variability of seismicity. To address these limitations, more recent approaches have turned to so-called

zoneless methods, which promise for greater realism by estimating the spatial density of earthquake characteristics using kernel density estimation (KDE).

However, KDE belongs to a frequentist framework that suffers from a kind of rigidity : parameters have to be fixed in advance by the model constructors, potentially leading to sub-optimal/sup-optimal representations. Although adaptive KDE techniques exist to address this issue, they often come with significant computational costs. Moreover, previous studies have shown that KDE is not so well-suited for regions with low or diffuse seismicity, such as France.

To overcome these limitations, we adopt a Bayesian perspective into. We begin with a simplified setting in which only the spatial distribution of earthquake epicenters is considered. To model this distribution flexibly, we employ a Dirichlet Process — a Bayesian nonparametric tool that defines a probability distribution over probability distributions. This allows the data itself to inform the complexity of the spatial model, without the need to fix the number of seismic zones/clusters.

Grounded in the deep belief that “we feed on the wisdom of our ancestors”, we propose a framework for integrating expert knowledge derived from traditional seismic source zoning. This knowledge, which encapsulates decades of geological, seismological, and geophysical expertise, is incorporated into our model through the construction of informative priors. Our approach thus bridges modern Bayesian learning with historical seismological insights, offering a coherent way to address the limitations of both zoneless and zoned methodologies.

---

**Algorithm 1** Gibbs pour DPMM

---

1: **Input :**

2: Observations  $X = \{x_i\}_{i=1\dots n} \subset \mathbb{R}^2$

3: Hyperparamètres :  $\alpha, (\boldsymbol{\mu}_0, \lambda_0, \boldsymbol{\Psi}_0, \nu_0)$

4: Iterations  $T$ , burn-in  $B$ ?, thinning  $L$ ?

5: **Initialisation :**

6: Assigner chaque  $x_i$  à un cluster

7: Calculer stats suffisantes  $\{n_k, \bar{x}_k, S_k\}$  pour chaque cluster non vide

8:  $K \leftarrow$  nb de clusters non vides

9: **for**  $t = 1$  to  $T$  **do**

10:   (I) Mise à jour des clusters  $c_i = k$  :

11:   **for**  $i = 1$  to  $n$  **do**

12:     Retirer  $x_i$  de son cluster  $k_{\text{old}}$  :  $n_{k_{\text{old}}} \leftarrow n_{k_{\text{old}}} - 1$

13:     **if**  $n_{k_{\text{old}}} = 0$  **then**

14:       Supprimer le cluster  $k_{\text{old}}$  :  $K \leftarrow K - 1$

15:     **end if**

16:     Pour chaque cluster existant  $k = 1, \dots, K$  :

17:       Calculer le poids  $w_k$  en utilisant l'équation 1

18:     Pour un nouveau cluster :

19:       Calculer le poids  $w_{\text{new}}$  en utilisant l'équation 2 ou l'équation 3 (cas informatif)

20:       Normaliser  $w_1, \dots, w_K, w_{\text{new}}$  et échantillonner parmi ces poids

21:       **if** nouveau cluster choisit **then**

22:         Créer cluster  $k < -K + 1$  avec  $c_k = c_{\text{new}}$

23:         Initialiser paramètres du cluster depuis posterior avec  $x_i$

24:       **end if**

25:     **end for**

26:    (II) Echantillonnage des paramètres de cluster  $\boldsymbol{\mu}_k$  et  $\Sigma_k$  :

27:    **for** chaque cluster  $k = 1$  to  $K$  **do**

28:      Calculer  $(\boldsymbol{\mu}_k, \lambda_k, \boldsymbol{\Psi}_k, \nu_k)$  avec toutes les données du cluster en utilisant les équations 4, 5, 6, 7.

29:      Echantillonner  $\Sigma_k$  et  $\boldsymbol{\mu}_k$  en utilisant l'équation 8.

30:    **end for**

31: **end for**

---

---

# Hawkes for Seismic Hazard

## Modèle ETAS (Epidemic Type Aftershock Sequence)

On observe des événements  $\{(t_i, x_i, y_i, m_i)\}_{i=1}^n$  avec  $t_i \in (0, T]$ ,  $x_i, y_i \in S \subset \mathbb{R}^2$  et magnitude  $m_i \geq m_0$ . Notons  $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i) : t_i < t\}$  l'historique jusqu'au temps  $t$ .

L'intensité conditionnelle d'un processus de comptage détermine de façon unique la structure de probabilité du processus (cf **Proposition 7.2.IV** de ?)

Le modèle ETAS est un Hawkes spatio-temporel marqué, la fonction d'intensité conditionnelle du quadriplet  $(t, x, y, m)$  s'écrit :

$$\lambda^*(t, x, y, m | \mathcal{H}_t) = p_M(m) \times \lambda(t, x, y | \mathcal{H}_t)$$

où  $p_M$  est la loi conditionnelle des magnitudes et  $\lambda(t, x, y | \mathcal{H}_t)$  l'intensité d'occurrence temps-espace.

### Loi des magnitudes $p_M$

Les magnitudes suivent la loi Gutenberg–Richter tronquée :

$$p_M(m) = \begin{cases} \frac{\beta \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{\max} - m_c)]} & \text{si } m_c \leq m \leq m_{\max}, \\ 0 & \text{sinon} \end{cases} \quad \text{avec } \beta = \ln(10) b.$$

La loi  $p_M$  est indépendante de  $(t, x, y)$  et identique pour les chocs de fond et déclenchés.

### Intensité spatio-temporelle

Elle est donnée par :

$$\lambda(t, x, y | \mathcal{H}_t) = \mu(x, y) + \sum_{j : t_j \in \mathcal{H}_t} g(t - t_j, x - x_j, y - y_j, m_j)$$

- $\mu(x, y)$  est l'intensité de fond (supposé stationnaire en temps)
- $g$  est le noyau d'excitation/de déclenchement (productivité × temps × espace)

### Modélisation de l'intensité de fond

On écrit souvent  $\mu(x, y) = \tilde{\mu} f(x, y)$  avec  $f \geq 0$ ,  $\int_S f = 1$ . La densité  $f$  peut être estimé en amont (KDE, DP, GP, SB etc. avec un catalogue "déclusterisé"), tandis que  $\tilde{\mu}$  est ajusté conjointement avec les autres paramètres ETAS.

## Modélisation du noyau d'excitation

Le noyau d'excitation dans le cadre du modèle ETAS s'écrit :

$$g(t - t_j, x - x_j, y - y_j, m_j) = g_m(m_j) \times g_t(t - t_j) \times g_s(x - x_j, y - y_j | m_j)$$

### Noyau de productivité (nombre moyen de descendants directs)

$$g_m(m_j) = A e^{\alpha(m_j - m_c)}, \quad A > 0, \alpha \geq 0$$

### Noyau temporelle (Omori–Utsu modifiée)

$$g_t(t - t_j) = (p - 1) c^{p-1} (t - t_j + c)^{-p}, \quad c > 0, p > 1$$

### Noyau spatial (isotrope de type Student)

En posant  $r^2 = (x - x_j)^2 + (y - y_j)^2$  et

$$R(m_j) = d e^{\gamma(m_j - m_c)}, \quad d > 0, \gamma \geq 0,$$

on a :

$$g_s(x - x_j, y - y_j | m_j) = \frac{q - 1}{\pi R(m_j)} \left[ 1 + \frac{r^2}{R(m_j)} \right]^{-q}, \quad q > 1.$$

Ce noyau est normalisé sur  $\mathbb{R}^2$  et est isotrope (-> pas d'information explicite sur les failles dans la région et n'élimine pas la possibilité d'avoir une réplique exactement au même endroit que son parent (peu probable))

## Expression finale de la fonction d'intensité spatio-temporel

Ainsi, on a :

$$\begin{aligned} \lambda^*(t, x, y, m | \mathcal{H}_t) &= \frac{\beta \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{\max} - m_c)]} \mathbb{1}_{\{m_c \leq m \leq m_{\max}\}} \times \\ &\left[ \tilde{\mu} f(x, y) + \sum_{j : t_j \in \mathcal{H}_t} A e^{\alpha(m_j - m_c)} (p - 1) c^{p-1} (t - t_j + c)^{-p} \times \frac{q - 1}{\pi R(m_j)} \left( 1 + \frac{r^2}{R(m_j)} \right)^{-q} \right] \end{aligned}$$

## Fonction de vraisemblance du modèle

**Proposition 1** [Version spatio-temporelle de (*Proposition 7.2.III.* de ?)] Soit  $N(\cdot)$  un processus de Hawkes sur  $[0, T] \times S$  d'intensité  $\lambda(t, x, y | \mathcal{H}_t)$ . Si l'on note  $t_1, \dots, t_n$  les réalisations du processus, alors la fonction de vraisemblance  $L$  de  $N(\cdot)$  est donnée par :

$$L = \prod_{i=1}^n \lambda(t_i, x_i, y_i) \exp \left( - \int_0^T \int_S \lambda(t, x, y) dt dx dy \right).$$

La log-vraisemblance de  $N(\cdot)$  est :

$$\ell = \sum_{i=1}^n \log (\lambda(t_i, x_i, y_i)) - \int_0^T \int_S \lambda(t, x, y) dt dx dy.$$

Ainsi, pour un processus de Hawkes marqué sur  $[0, T] \times S \times [m_c, m_{\max}]$  dont l'intensité conditionnelle marquée est :

$$\lambda^*(t, x, y, m | \mathcal{H}_t) = p_M(m) \times \lambda(t, x, y | \mathcal{H}_t)$$

la log-vraisemblance s'écrit (selon la proposition précédente) :

$$\ell(\theta) = \sum_{i=1}^n \log (\lambda^*(t_i, x_i, y_i, m_i | \mathcal{H}_{t_i})) - \int_0^T \int_S \int_{m_c}^{m_{\max}} \lambda^*(t, x, y, m | \mathcal{H}_t) dm dx dy dt$$

Comme  $p_M$  est indépendante de  $(t, x, y)$  et normalisée sur  $[m_c, m_{\max}]$ , on a  $\int_{m_c}^{m_{\max}} p_M(m) dm = 1$ . D'où la décomposition :

$$\ell(\theta) = \underbrace{\sum_{i=1}^n \log (p_M(m_i))}_{\text{termes magnitudes}} + \underbrace{\left[ \sum_{i=1}^n \log (\lambda(t_i, x_i, y_i | \mathcal{H}_{t_i})) - \int_0^T \int_S \lambda(t, x, y | \mathcal{H}_t) dx dy dt \right]}_{\text{termes spatio-temporel}}$$

### Termes magnitudes

On a :

$$\begin{aligned} \sum_{i=1}^n \log (p_M(m_i)) &= n \log(\beta) - \beta \sum_{i=1}^n (m_i - m_c) - n \log (1 - \exp[-\beta(m_{\max} - m_c)]) \\ &= n \log \left( \frac{\beta}{1 - \exp[-\beta(m_{\max} - m_c)]} \right) - \beta \sum_{i=1}^n (m_i - m_c) \end{aligned}$$

avec  $\beta = \ln(10) b$ .

## Termes spatio-temporel - Somme des logs des intensités spatio-temporeles

On a :

$$\begin{aligned} \sum_{i=1}^n \log \left( \lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i}) \right) &= \sum_{i=1}^n \log \left[ \tilde{\mu} f(x_i, y_i) + \sum_{j: t_j \in \mathcal{H}_{t_i}} A e^{\alpha(m_j - m_c)} (p-1) c^{p-1} \right. \\ &\quad \times \left. (t_i - t_j + c)^{-p} g_s(x_i - x_j, y_i - y_j \mid m_j) \right]. \end{aligned}$$

## Termes spatio-temporel - Intégrale de l'intensité spatio-temporel

On sépare l'intensité de fond du Hawkes de la fonction d'excitation. On a :

- **Intensité de fond** : En utilisant  $\mu(x, y) = \tilde{\mu} f(x, y)$  et  $\int_S f = 1$ , on a :

$$\int_0^T \int_S \mu(x, y) dx dy dt = T \tilde{\mu}.$$

- **Fonction d'excitation** : Pour chaque événement  $j$ , par Fubini on a :

$$\begin{aligned} &\int_0^T \int_S \sum_{j: t_j \in \mathcal{H}_t} g_m(m_j) g_t(t - t_j) g_s(x - x_i, y - y_i \mid m_j) dx dy dt \\ &= \int_0^T \int_S \sum_{j=1}^n \mathbb{1}_{t_j < t} g_m(m_j) g_t(t - t_j) g_s(x - x_i, y - y_i \mid m_j) dx dy dt \\ &= \sum_{j=1}^n g_m(m_j) \times \underbrace{\int_{t_j}^T g_t(t - t_j) dt}_{T_j} \times \underbrace{\int_S g_s(x - x_i, y - y_i \mid m_j) dx dy}_{S_j} \end{aligned}$$

1. Calcul de  $T_j$  :

$$T_j := \int_{t_j}^T g_t(t - t_j) dt = \int_0^{T-t_j} (p-1) c^{p-1} (u + c)^{-p} du = 1 - \left( \frac{c}{T - t_j + c} \right)^{p-1}$$

2. Calcul de  $S_j$  :

$$S_j := \int_S g_s(x - x_j, y - y_j \mid m_j) dx dy = \int_S \frac{q-1}{\pi R(m_j)} \left[ 1 + \frac{r^2}{R(m_j)} \right]^{-q} dx dy$$

$S_j$  est la probabilité qu'une réplique née au point  $(x_j, y_j)$  tombe à l'intérieur de  $S$ . Si  $S = \mathbb{R}^2$  ou si on suppose que notre domaine  $S$  couvre tous les séismes (chocs principaux et répliques), alors  $S_j = 1$  (student).

Ainsi, en rassemblant ce qui est plus haut, on a :

$$\int_0^T \int_S \lambda(t, x, y \mid \mathcal{H}_t) dx dy dt = T \tilde{\mu} + \sum_{j=1}^n A e^{\alpha(m_j - m_c)} \times \left[ 1 - \left( \frac{c}{T - t_j + c} \right)^{p-1} \right] \times S_j$$

## Expression finale de la log-vraisemblance

En agrégeant le travail fait plus haut, on a :

$$\ell(\theta) = \underbrace{n \log \left( \frac{\beta}{1 - \exp[-\beta(m_{\max} - m_c)]} \right) - \beta \sum_{i=1}^n (m_i - m_c)}_{\text{magnitudes}} + \sum_{i=1}^n \log \left[ \tilde{\mu} f(x_i, y_i) + \sum_{j : t_j \in \mathcal{H}_{t_i}} A e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \frac{q-1}{\pi R(m_j)} \left(1 + \frac{r^2}{R(m_j)}\right)^{-q} \right] - \left( T \tilde{\mu} + \sum_{j=1}^n A e^{\alpha(m_j - m_c)} \times \left[ 1 - \left( \frac{c}{T - t_j + c} \right)^{p-1} \right] \times S_j \right).$$

avec  $S_j = \int_S \frac{q-1}{\pi R(m_j)} \left[ 1 + \frac{r^2}{R(m_j)} \right]^{-q} dx dy$  qui serait à évaluer numériquement a priori.

## Vecteur de paramètres à estimer

Le vecteur de paramètres à estimer pour le modèle ETAS est :

$$\theta = (\beta, \tilde{\mu}, A, \alpha, c, p, d, q, \gamma)$$

avec :

- $\tilde{\mu}$  : niveau du fond
- $A, \alpha$  : paramètres de productivité et dépendance à la magnitude
- $c, p$  : paramètres du noyau temporel d'Omori–Utsu
- $d, q, \gamma$  : paramètres du noyau spatial (dépendance à la magnitude via  $R(m) = d e^{\gamma(m-m_c)}$ )

## Propriétés du modèle

WORK IN PROGRES ...

## GP-ETAS : semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model ([Molkenthin et al. \[2022a\]](#))

### I - Modèle GP-ETAS

**Objectif** Améliorer inférence du ETAS en introduisant un cadre bayésien, afin de permettre une *quantification complète de l'incertitude*. Alors que les méthodes classiques reposent sur le max de vraisemblance, le cadre bayésien fournit une estimation plus robuste et une évaluation explicite des incertitudes.

**Principe du GP-ETAS** Le modèle conserve la structure générale d'un ETAS, mais estime l'intensité de fond  $\mu(x)$  de manière non paramétrique avec un processus gaussien (GP). La fonction de déclenchement  $\phi(\cdot)$  garde une forme paramétrique classique avec paramètres  $\theta_\phi$ .

#### Formulation du taux de fond

$$\mu(x) = \bar{\lambda} \sigma(f(x)) = \frac{\bar{\lambda}}{1 + e^{-f(x)}},$$

où  $\sigma(\cdot)$  est la fonction sigmoïde logistique,  $\bar{\lambda}$  un scalaire positif qui borne  $\mu(x) \in [0, \bar{\lambda}]$ , et  $f$  une fonction réelle sur  $x \in \mathcal{X}$ .

**Prior sur  $f(x)$**  Le modèle suppose un GP a priori :

$$f \sim \mathcal{GP}(0, k(x, x'|\nu)),$$

avec une fonction de covariance gaussienne :

$$k(x, x'|\nu) = \nu_0^2 \exp\left(-\frac{(x_1 - x'_1)^2}{2\nu_1^2} - \frac{(x_2 - x'_2)^2}{2\nu_2^2}\right),$$

où  $\nu_0$  est l'amplitude et  $(\nu_1, \nu_2)$  les longueurs d'échelle spatiales. Les hyperparamètres  $\nu$  et  $\bar{\lambda}$  sont estimés avec les données.

#### Spécification des priors

$$\begin{aligned} \nu &\sim p_\nu \quad (\text{loi exponentielle}) \\ f &\sim \mathcal{GP}(0, k(\cdot, \cdot|\nu)) \\ \bar{\lambda} &\sim p_{\bar{\lambda}} \quad (\text{loi gamma}) \\ \mu(x) &= \bar{\lambda} \sigma(f(x)) \\ \theta_\phi &\sim p_{\theta_\phi}, \text{ prior sur params de } \phi \text{ (loi uniforme)} \end{aligned}$$

Le modèle d'observation est alors un Hawkes avec intensité conditionnelle dépendant de  $\mu$  et  $\phi$  :

$$\mathcal{D} \mid \mu, \theta_\phi \sim \text{Hawkes process with GP-ETAS intensity.}$$

Sans terme d'excitation  $\phi(\cdot)$ , le modèle se réduit à un **SGCP** (Sigmoid Gaussian Cox Process), un processus de Poisson inhomogène à intensité "guidée" par un GP.

**Simulation du modèle GP-ETAS** La génération de données  $\mathcal{D} = \{(t_i, x_i, y_i, m_i)\}_{i=1}^{N_D}$  se fait en deux étapes selon la structure de branchement latente du processus.

1. Génération des événements de fond  $\mathcal{D}_0 = \{(t_i, x_i, y_i, m_i)\}_{i=1}^{N_{\mathcal{D}_0}}$  depuis un SGCP comme indiqué dans ?
2. Génération des événements déclenchés  $D_\phi$  à partir des événements de fond, selon la fonction de déclenchement  $\phi(\cdot)$

L'ensemble final est :  $\mathcal{D} = \mathcal{D}_0 \cup D_\phi$ . La procédure ci-dessus est réalisée en utilisant un algo de thinning pour processus poissoniens (une sorte d'acceptation-rejet).

## II - Inférence du GP-ETAS

**Objectif** L'objectif est d'estimer la posterior

$$p(\mu, \theta_\phi \mid D),$$

**Difficultés de l'inférence** L'inférence directe est intractable à cause de trois obstacles majeurs :

1. Couplage de  $\mu$  et  $\phi$  dans la vraisemblance, empêchant leur traitement séparé
2. Présence d'une intégrale intractable dans le terme exponentiel due au GP sur  $f \Rightarrow$  la postérieure est "doublement intractable"
3. La vraisemblance du processus ponctuel est non gaussienne, donc non conjuguée au prior GP

**solution : inférence par data augmentation** Pour contourner ces difficultés, trois variables latentes sont introduites pour transformer le problème en un modèle "hiérarchique" tractable et échantillonnable par un Gibbs.

1. *Structure de branchement latente Z* : Chaque événent  $i$  reçoit un indicateur  $z_i$  identifiant son parent :

$$z_i = \begin{cases} 0 & \text{si l'événent est de fond} \\ j & \text{si l'événent est déclenché par l'événent } j \end{cases}$$

Cette structure de branchement *décorrèle* les composantes de fond et de déclenchement, et la vraisemblance se factorise comme :

$$p(D, Z \mid \mu, \theta_\phi) = p(D_0 \mid \mu, Z) p(D_\phi \mid \theta_\phi, Z).$$

**Développé plus bas**

2. *Processus de Poisson latent* : ?

**Développé plus bas**

3. *Variables latentes de Pólya-Gamma* ? : Inspiré de la solution de ? au problème de classification logistique GP.

**Développé plus bas**

**Échantillonnage Gibbs** Le modèle complet est inféré par un sorte de blocked Gibbs :

1. Échantillonnage de la structure de branchement  $Z$
2. Inférence de l'intensité de fond  $(f, \bar{\lambda}, \nu)$  (les postérieurs se calculent bien)
3. Inférence des params de déclenchement  $\theta_\phi$

Dans les étapes 2 et 3, on y retrouve du Metropolis-Hastings.

### III - Résultats numériques

...

## Augmentation de données par structure de branchement latente

La vraisemblance d'un Hawkes spatio-temporel s'écrit :

$$p(D \mid \mu, \theta_\phi) = \left\{ \prod_{i=1}^{N_D} \lambda(t_i, x_i \mid \mu, \theta_\phi, H_{t_i}) \right\} \exp \left( - \int_0^{|\mathcal{T}|} \int_{\mathcal{X}} \lambda(t, x \mid \mu, \theta_\phi, H_t) dx dt \right). \quad (9)$$

On introduit des variables latentes  $z_i \in \{0, 1, \dots, i-1\}$  pour chaque événement  $i$  :  $z_i = 0$  si que l'événement est un mainshock, et  $z_i = j$  ( $1 \leq j \leq i-1$ ) si  $i$  est déclenché par  $j$ . On note  $Z = (z_1, \dots, z_{N_D})$ . On pose,

$$\phi_{ij}(\theta_\phi) := \phi(t_i - t_j, x_i - x_j \mid m_j, \theta_\phi), \quad \phi_i(\theta_\phi) := \phi(t - t_i, x - x_i \mid m_i, \theta_\phi),$$

et  $T_i := [t_i, |\mathcal{T}|] \subset \mathcal{T}$ .

On suppose un a priori uniforme sur  $Z$ , i.e.  $p(Z) = \text{cste}$ . La vraisemblance augmentée s'obtient en factorisant, pour chaque  $i$  :

$$\lambda(t_i, x_i \mid \mu, \theta_\phi, H_{t_i}) = \mu(x_i) + \sum_{j=1}^{i-1} \phi_{ij}(\theta_\phi) = \mu(x_i)^{I(z_i=0)} \prod_{j=1}^{i-1} \phi_{ij}(\theta_\phi)^{I(z_i=j)}.$$

Le terme exponentiel de (9) se décompose, par linéarité, comme :

$$\int_0^{|\mathcal{T}|} \int_{\mathcal{X}} \lambda(t, x \mid \mu, \theta_\phi, H_t) dx dt = |\mathcal{T}| \int_{\mathcal{X}} \mu(x) dx + \sum_{i=1}^{N_D} \int_{T_i} \int_{\mathcal{X}} \phi_i(\theta_\phi) dx dt.$$

Ainsi, la vraisemblance jointe augmentée s'écrit :

$$p(D, Z \mid \mu, \theta_\phi) = \left\{ \prod_{i=1}^{N_D} \mu(x_i)^{I(z_i=0)} \right\} \exp \left( -|\mathcal{T}| \int_{\mathcal{X}} \mu(x) dx \right) \\ \times \left\{ \prod_{i=1}^{N_D} \prod_{j=1}^{i-1} \phi_{ij}(\theta_\phi)^{I(z_i=j)} \right\} \exp \left( -\sum_{i=1}^{N_D} \int_{T_i} \int_{\mathcal{X}} \phi_i(\theta_\phi; t, x) dx dt \right) \times p(Z). \quad (10)$$

En notant  $D_0 = \{x_i : z_i = 0\}$  l'ensemble des mainshocks, on reconnaît dans (10) le produit :

$$p(D, Z \mid \mu, \theta_\phi) = \underbrace{\left[ \prod_{i=1}^{N_D} \mu(x_i)^{I(z_i=0)} \right]}_{p(D_0 \mid Z, \mu)} \exp \left( -|\mathcal{T}| \int_{\mathcal{X}} \mu(x) dx \right) \quad (11)$$

$$\times \underbrace{\left[ \prod_{i=1}^{N_D} \prod_{j=1}^{i-1} \phi_{ij}(\theta_\phi)^{I(z_i=j)} \right]}_{p(D \mid Z, \theta_\phi)} \exp \left( -\sum_{i=1}^{N_D} \int_{T_i} \int_{\mathcal{X}} \phi_i(\theta_\phi) dx dt \right) \times p(Z). \quad (12)$$

Autrement dit,

$$p(D, Z \mid \mu, \theta_\phi) = p(D_0 \mid Z, \mu) p(D \mid Z, \theta_\phi) p(Z).$$

**Loi conditionnelle de  $z_i$ .** La loi conditionnelle d'un  $z_i$ , sachant  $(D, \mu, \theta_\phi)$ , s'obtient en ne retenant dans (10) que les termes dépendant de  $z_i$  :

$$p(z_i = j \mid D, \mu, \theta_\phi) \propto \begin{cases} \mu(x_i), & \text{si } j = 0, \\ \phi_{ij}(\theta_\phi), & \text{si } j \in \{1, \dots, i-1\}. \end{cases}$$

En normalisant par  $\lambda(t_i, x_i \mid \mu, \theta_\phi, H_{t_i})$ , on a une loi catégorielle sur  $\{0, \dots, i-1\}$  :

$$p(z_i = 0 \mid D, \mu, \theta_\phi) = \frac{\mu(x_i)}{\lambda(t_i, x_i \mid \mu, \theta_\phi, H_{t_i})}, \quad p(z_i = j \mid D, \mu, \theta_\phi) = \frac{\phi_{ij}(\theta_\phi)}{\lambda(t_i, x_i \mid \mu, \theta_\phi, H_{t_i})}, \quad j = 1, \dots, i-1.$$

En notant  $p_i = (p_{i0}, p_{i1}, \dots, p_{i,i-1})$  le vecteur de ces proba, on a :

$$z_i \mid D, \mu, \theta_\phi \sim \text{Categorical}(p_i), \quad \sum_{j=0}^{i-1} p_{ij} = 1.$$

# Tractable Nonparametric Bayesian Inference in Poisson Processes with Gaussian Process Intensities Adams et al. [2009]

## Le modèle SGCP

Le Sigmoidal Gaussian Cox Process (SGCP) est un modèle non paramétrique qui permet de modéliser des processus de Poisson inhomogènes. L'idée clé est l'utilisation d'un processus gaussien (GP) pour définir une loi a priori sur la fonction d'intensité du processus.

**Le GP comme a priori** Les auteurs introduisent une fonction aléatoire  $g(s) \sim \text{GP}(m(s), C(s, s'))$ , avec :

- $m(s)$  est la moyenne
- $C(s, s')$  est le noyau de covariance (paramétré par des hyperparamètres  $\theta$ )

Cette fonction  $g(s)$  sert à générer une intensité via la transformation :

$$\lambda(s) = \lambda^* \sigma(g(s))$$

avec :

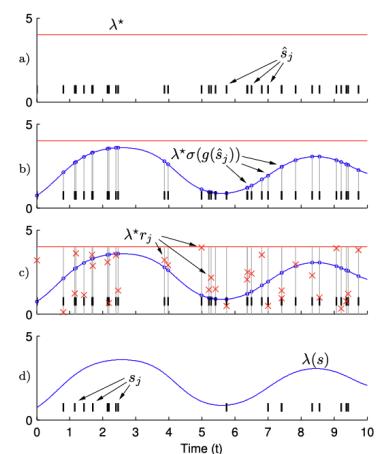
- $\sigma(\cdot)$  est la fonction sigmoïde logistique :  $\sigma(z) = \frac{1}{1+e^{-z}}$
- $\lambda^*$  est une borne supérieure sur l'intensité

Ainsi,  $\lambda(s)$  est une fonction continue, positive et vivant entre 0 et  $\lambda^*$ , c'est bien une intensité.

**Simulation exacte de données de Poisson** Les auteurs développent une construction à travers laquelle ils peuvent simuler des données de Poisson exactes à partir du GP, sans approximation num. Ça se fait par la méthode du "thinning", une forme d'échantillonnage par acceptation-rejet adaptée aux processus ponctuels. Les steps sont les suivantes :

1. Générer  $J$  points  $\hat{s}_j$  dans un domaine selon un processus de Poisson homogène d'intensité  $\lambda^*$
2. Ces points  $\hat{s}_j$  deviennent les points d'entrée de la réalisation d'un GP  $g(s)$  et on a un ensemble  $\{g(\hat{s}_j)\}$
3. Tirer pour chaque point un nombre aléatoire  $r_j \in (0, 1)$
4. Conserver le point si  $r_j < \sigma(g(\hat{s}_j))$

Les points conservés forment un échantillon issu d'un processus de Poisson inhomogène d'intensité  $\lambda(s)$ .



## Inférence

**Problème :** La vraisemblance du processus de Poisson implique une intégrale intractable et la loi a posteriori est doublement intractable (numérateur également intractable), ce qui rend l'application de MCMC impossible en l'état.

**Solution :** L'idée est d'augmenter la loi a posteriori avec des variables latentes :

- $M$  : nombre total de points rejetés (les événements “thinned”)
- $\{\tilde{s}_m\}$  : positions de ces points rejetés
- $g_M = g(\tilde{s}_m)$  et respectivement  $g_K = g(s_k)$  : valeurs de la fonction  $g(s)$  aux points rejetés et respectivement observés

En travaillant avec ces variables finies, on évite d'avoir à intégrer sur une fonction infinie.

**MCMC :** Leur algo alterne trois mises à jour :

1. Échantillonnage de  $M$  : Utilise Metropolis–Hastings pour proposer des insertions ou suppressions de points
2. Échantillonnage des  $\tilde{s}_m$  : M–H aussi
3. Échantillonnage de  $g(s)$  : Utilise Hamiltonian Monte Carlo (HMC)

Les hyperparamètres  $\theta$  du processus gaussien peuvent aussi être estimés directement par HMC. Le paramètre  $\lambda^*$  peut également être estimé conjointement, en posant une loi Gamma comme a priori, permettant une mise à jour analytique pendant le MCMC.

**Predictive :** La loi prédictive est accessible, on pourrait échantillonner des ensembles d'événements selon un certaine procédure, si souhaité.

## Efficient Bayesian inference of sigmoidal Gaussian cox processes Donner and Opper

### Problème d'inférence pour le SGCP

Cadre général : On se place dans le cadre du modèle SGCP décrit plus haut. On observe un ensemble d'événements

$$\mathcal{D} = \{x_n\}_{n=1}^N, \quad x_n \in \mathcal{X} \subset \mathbb{R}^d,$$

supposés issus d'un processus de Poisson sur le compact  $\mathcal{X}$ , de fonction d'intensité inconnue  $\Lambda : \mathcal{X} \rightarrow \mathbb{R}_+$ . La vraisemblance est donnée par :

$$L(D | \Lambda) = \exp\left(-\int_{\mathcal{X}} \Lambda(x) dx\right) \prod_{n=1}^N \Lambda(x_n). \quad (13)$$

Dans cadre bayésien,  $\Lambda$  devient un processus aléatoire positif :

$$\Lambda(x) \geq 0, \quad x \in \mathcal{X}.$$

Un processus de Poisson dont l'intensité elle-même est aléatoire est appelé **processus de Cox** ([?](#)).

Pour assurer la positivité de  $\Lambda$ , on introduit la reparamétrisation sigmoïde :

$$\Lambda(x) = \lambda \sigma(g(x)), \quad \sigma(z) = \frac{1}{1 + e^{-z}}, \quad (14)$$

où  $\lambda > 0$  est le taux d'intensité maximal et  $g : \mathcal{X} \rightarrow \mathbb{R}$  est une fonction latente non bornée.

On place sur  $g$  un GP :  $g \sim \mathcal{GP}(0, k(x, x'))$ , et on choisit pour  $\lambda$  un a priori Gamma :  $\lambda \sim \text{Gamma}(\alpha_0, \beta_0)$ , de densité  $p(\lambda) \propto \lambda^{\alpha_0-1} e^{-\beta_0 \lambda}$ .

**Vraisemblance et postérieure :** La vraisemblance induite pour  $g$  et  $\lambda$  est :

$$L(D | g, \lambda) = \exp\left(-\int_{\mathcal{X}} \lambda \sigma(g(x)) dx\right) \prod_{n=1}^N \lambda \sigma(g_n), \quad g_n := g(x_n). \quad (15)$$

Le modèle complet admet la mesure a priori produit :

$$dP_{\text{prior}} = dP_{\text{GP}}(g) p(\lambda) d\lambda.$$

Le posterior sur  $(g, \lambda)$ , au sens des mesures infinies, s'écrit alors comme dérivée de Radon–Nikodym :

$$\frac{dP_{\text{post}}}{dP_{\text{prior}}}(g, \lambda) = \frac{L(D | g, \lambda)}{\mathbb{E}_{P_{\text{prior}}}[L(D | g, \lambda)]}. \quad (16)$$

**Problème intractable :** Le calcul du postérieur (16) est doublement intractable (Murray et al., 2006) pcq :

- le terme exponentiel de (15) contient l'intégrale  $\int \sigma(g(x)) dx$ ,
- la normalisation du dénominateur exige une intégration sur l'espace fonctionnel du GP,
- la dépendance non-linéaire de  $\sigma(g(x))$  empêche la conjugaison avec le prior gaussien.

Les auteurs introduisent alors une double augmentation de données (processus de Poisson latent (marqué) + Pólya–Gamma) qui rend l'inférence calculable.

**Théorème 1 [Campbell]** Pour un ensemble aléatoire de points  $\Pi_Z = z_m, , z_m \in Z$  défini sur un espace  $Z$ , et pour une fonction  $h : Z \rightarrow \mathbb{R}$ , on définit la fonctionnelle caractéristique du processus de Poisson par :

$$\mathbb{E}_{P_{\Lambda}} \left[ \prod_{z_m \in \Pi_Z} e^{h(z_m)} \right] = \exp\left(-\int_Z [1 - e^{h(z)}] \Lambda(z) dz\right), \quad (17)$$

où  $P_{\Lambda}$  désigne la mesure de probabilité d'un processus de Poisson d'intensité  $\Lambda(z)$ . Cette relation caractérise de manière unique un Poisson d'intensité  $\Lambda(z)$ .

## Augmentation I : processus de Poisson latent :

L'équation suivante donne la vraisemblance du background conditionnellement à la structure de branchement :

$$p(D_0 \mid f, \bar{\lambda}, Z) = \bar{\lambda}^{N_D} \prod_{i:z_i=0} \sigma(f_i) \exp\left(-\int_T \int_X \bar{\lambda} \sigma(f(x)) dx dt\right)$$

Le problème est le terme intégral dans l'exponentielle, difficile à manipuler dans une inférence MCMC. L'objectif de l'augmentation par un processus de Poisson latent est d'obtenir une vraisemblance. L'idée clé repose sur la superposition de processus de Poisson. On introduit un processus de Poisson latent  $\Pi$  avec intensité :

$$\hat{\lambda}(x) = \bar{\lambda}(1 - \sigma(f(x))) = \bar{\lambda}\sigma(-f(x))$$

pour que sa superposition avec le processus réel d'intensité

$$\bar{\lambda}\sigma(f(x))$$

donne un processus homogène de taux constant :

$$\bar{\lambda} = \bar{\lambda}\sigma(f(x)) + \bar{\lambda}\sigma(-f(x)).$$

Donc, dans la région d'observation  $\mathcal{X} \times \mathcal{T}$ , le nombre total de points dans  $D_0 \cup \Pi$ , suit un Poisson de paramètre  $\bar{\lambda} \mathcal{X} \mathcal{T}$ .

La vraisemblance jointe d'un processus de Poisson homogène d'intensité  $\bar{\lambda}$  sur un domaine de volume  $\mathcal{X} \mathcal{T}$  est :

$$p(D_0, \Pi \mid \bar{\lambda}, f) = \bar{\lambda}^{N_{D_0 \cup \Pi}} \exp(\bar{\lambda} \mathcal{X} \mathcal{T}) \prod_{x \in D_0 \cup \Pi} p(x \mid f).$$

Le facteur  $\bar{\lambda}^{N_{D_0 \cup \Pi}} \exp(-\bar{\lambda} \mathcal{X} \mathcal{T})$  vient du PP homogène. Le terme important est donc la probabilité qu'un point appartienne au processus connu ou latent :

- pour un point réel (background), la probabilité d'être dans  $D_0$  est  $\sigma(f(x))$ ,
- pour un point latent, la probabilité d'être dans  $\Pi$  est  $\sigma(-f(x))$ .

D'où la factorisation :

$$\prod_{i:z_i=0} \sigma(f_i) \prod_{l \in \Pi} \sigma(-f_l).$$

On combine :

- la partie du PP homogène :  $\bar{\lambda}^{N_{D_0 \cup \Pi}} \exp(-\bar{\lambda} \mathcal{X} \mathcal{T})$ ,
- les poids  $\sigma(f_i)$  pour les points observés,
- les poids  $\sigma(-f_l)$  pour les points latents.

On réécrit :

$$\bar{\lambda}^{N_{D_0 \cup \Pi}} = \bar{\lambda}^{N_D} \bar{\lambda}^{N_\Pi}.$$

Ce qui donne :

$$p(D_0, \Pi \mid f, \bar{\lambda}, Z) = \left( \prod_{i:z_i=0} \bar{\lambda}\sigma(f_i) \right) \left( \prod_{l \in \Pi} \bar{\lambda}\sigma(-f_l) \right) \exp(-\bar{\lambda}|X||T|).$$

## Augmentation II : variables de Pólya–Gamma et processus de Poisson marqué

On démarre de la représentation :

$$\frac{1}{\cosh^b(z/2)} = \int_0^\infty e^{-\frac{z^2}{2}\omega} p_{\text{PG}}(\omega | b, 0) d\omega, \quad (18)$$

où  $p_{\text{PG}}$  est densité de la loi de Pólya–Gamma.

Étant donné que :

$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} = \frac{e^{z/2}}{e^{z/2} + e^{-z/2}} = \frac{e^{z/2}}{2 \cosh(z/2)} = \frac{e^{z/2}}{2} \cosh^{-1}\left(\frac{z}{2}\right) \quad (19)$$

D'où pour  $b = 1$  :

$$\sigma(z) = \frac{e^{z/2}}{2} \int_0^\infty e^{-\frac{z^2}{2}\omega} p_{\text{PG}}(\omega | 1, 0) d\omega \quad (20)$$

$$= \int_0^\infty \exp\left(\frac{z}{2} - \frac{z^2}{2}\omega - \ln 2\right) p_{\text{PG}}(\omega | 1, 0) d\omega. \quad (21)$$

En posant  $z = -g(x)$ , on en déduit :

$$\sigma(-g(x)) = \int_0^\infty \exp\left(\frac{-g(x)}{2} - \frac{(-g(x))^2}{2}\omega - \ln 2\right) p_{\text{PG}}(\omega | 1, 0) d\omega. \quad (22)$$

En écrivant

$$f(\omega, z) := \frac{z}{2} - \frac{z^2}{2}\omega - \ln 2, \quad \text{on a : } \sigma(-g(x)) = \int_0^\infty e^{f(\omega, -g(x))} p_{\text{PG}}(\omega | 1, 0) d\omega. \quad (23)$$

Ainsi, on a :

$$\begin{aligned} \exp\left(-\int_{\mathcal{X}} \lambda [1 - \sigma(-g(x))] dx\right) &= \exp\left(-\lambda \int_{\mathcal{X}} \left[1 - \int_0^\infty e^{f(\omega, -g(x))} p_{\text{PG}}(\omega | 1, 0) d\omega\right] dx\right) \\ &= \exp\left(-\lambda \int_{\mathcal{X} \times \mathbb{R}_+} \left[1 - e^{f(\omega, -g(x))}\right] p_{\text{PG}}(\omega | 1, 0) d\omega dx\right). \end{aligned}$$

Le théorème de Campbell donne alors :

$$\exp\left(-\int_{\widehat{\mathcal{X}}} \left[1 - e^{f(\omega, -g(x))}\right] \Lambda(x, \omega) d\omega dx\right) = \mathbb{E}_{P_\Lambda} \left[ \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{X}}}} e^{f(\omega_m, -g(x_m))} \right], \quad (24)$$

où  $\Pi_{\widehat{\mathcal{X}}}$  est un processus de Poisson marqué sur l'espace produit  $\widehat{\mathcal{X}} = \mathcal{X} \times \mathbb{R}_+$ , et de mesure d'intensité  $\Lambda(x, \omega) = \lambda p_{\text{PG}}(\omega | 1, 0)$ , où  $p_{\text{PG}}(\omega | 1, 0)$  est la densité de la loi de Pólya–Gamma. Cette formulation permet de :

1. générer les positions  $\{x_m\}$  selon un processus de Poisson inhomogène sur  $\mathcal{X}$ ,
2. associer ensuite à chaque  $x_m$  un marquage indépendant  $\omega_m \sim p_{\text{PG}}(\cdot | 1, 0)$ .

Pour chaque facteur discret de (15), on applique :

$$\sigma(g(x_n)) = \int_0^\infty e^{f(\omega_n, g(x_n))} p_{\text{PG}}(\omega_n | 1, 0) d\omega_n.$$

D'où :

$$\prod_{n=1}^N \lambda \sigma(g(x_n)) = \int_{\mathbb{R}_+} \left[ \prod_{n=1}^N \lambda e^{f(\omega_n, g(x_n))} \right] \prod_{n=1}^N p_{\text{PG}}(\omega_n | 1, 0) d\omega_n. \quad (25)$$

En combinant (24) et (25) dans (15), on obtient la représentation de la vraisemblance jointe suivante :

$$\begin{aligned} L(D | g, \lambda) &= \mathbb{E}_{P_\Lambda} \left[ \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{X}}}} e^{f(\omega_m, -g(x_m))} \right] \times \int_{\mathbb{R}_+} \left[ \prod_{n=1}^N \lambda e^{f(\omega_n, g(x_n))} \right] \prod_{n=1}^N p_{\text{PG}}(\omega_n | 1, 0) d\omega_n \\ &= \int \left( \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{X}}}} e^{f(\omega_m, -g(x_m))} \right) P_\Lambda(d\Pi_{\widehat{\mathcal{X}}}) \times \int \left( \prod_{n=1}^N \lambda e^{f(\omega_n, g(x_n))} \right) P_{\omega_N}(d\omega_N) \quad (26) \end{aligned}$$

$$= \int \left( \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{X}}}} e^{f(\omega_m, -g(x_m))} \right) \left( \prod_{n=1}^N \lambda e^{f(\omega_n, g(x_n))} \right) dP_{\text{aug}}(d\Pi_{\widehat{\mathcal{X}}}, d\omega_N), \quad (27)$$

$$= \int \frac{dP_{\text{joint}}}{dP_{\text{aug}}} dP_{\text{aug}} \quad (28)$$

où la mesure a priori des variables augmentées est :

$$P_{\text{aug}} = P_\Lambda \times P_{\omega_N}, \quad \text{et} \quad dP_{\omega_N} = \prod_{n=1}^N p_{\text{PG}}(\omega_n | 1, 0) d\omega_n.$$

Comme les deux intégrations se font sur des variables indépendantes dans (26) (l'une sur le processus latent, l'autre sur les variables PG (les variables PG sont les marques, donc sont indépendants du processus)), on peut combiner dans (27) les deux intégrales en une seule sur la mesure produit  $P_{\text{aug}}$  (à revérifier mais il me semble que ça découle de Fubini).

Or, on peut écrire :

$$L(D | g, \lambda) = \int L(D, \omega_N, \Pi_{\widehat{\mathcal{X}}} | g, \lambda) dP_{\text{aug}}(d\Pi_{\widehat{\mathcal{X}}}, d\omega_N) \quad (29)$$

Donc par identification, on a :

$$\frac{dP_{\text{joint}}}{dP_{\text{aug}}} = L(D, \omega_N, \Pi_{\widehat{\mathcal{X}}} | g, \lambda) = \left( \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{X}}}} e^{f(\omega_m, -g(x_m))} \right) \left( \prod_{n=1}^N \lambda e^{f(\omega_n, g(x_n))} \right) \quad (30)$$

Par ailleurs, la vraisemblance originelle se retrouve par marginalisation sur les variables latentes :

$$L(D | g, \lambda) = \mathbb{E}_{P_{\text{aug}}} [L(D, \omega_N, \Pi_{\widehat{\mathcal{X}}} | g, \lambda)]$$

Enfin, la forme exponentielle-quadratique de  $f(\omega, z)$  implique que  $L(D, \omega_N, \Pi_{\widehat{\mathcal{X}}} | g, \lambda)$  est une fonction gaussienne en  $g$  : elle ne dépend de  $g(\cdot)$  que de manière linéaire et quadratique. Le modèle augmenté est donc conjugué au prior gaussien (GP), ce qui permet inférence bayésienne.

---

## Abstract pour symposium SIGMA 3 (04/12/2025)

Probabilistic Seismic Hazard Assessment (PSHA) is traditionally based on piecewise spatially and temporally homogeneous Poisson process assumptions. While this framework provides a clear and structured methodology, it tends to oversimplify the complex spatial variability of seismicity. To address these limitations, more recent approaches have moved toward zoneless methods, which aim for greater realism by estimating the spatial density of earthquake occurrences using kernel density estimation (KDE).

However, KDE belongs to a frequentist framework that suffers from a kind of rigidity : parameters have to be fixed in advance by the model constructors, potentially leading to sub-optimal/sup-optimal representations. Although adaptive KDE techniques exist to address this issue, they often come with significant computational costs. Moreover, previous studies have shown that KDE is not so well-suited for regions with low or diffuse seismicity, such as France.

To overcome these limitations, we focus on marked spatio-temporal point process models, such as Hawkes processes (ETAS), where earthquake occurrences are modeled through a conditional intensity function combining a background (or baseline) term and self-exciting components. In our approach, the background intensity is learned using Bayesian nonparametric methods — notably Dirichlet Processes, Gaussian Processes, etc. — allowing the model to flexibly adapt its spatial structure and complexity directly from the data.

Grounded in the belief that “we feed on the wisdom of our ancestors”, we want to propose a Bayesian framework for integrating expert knowledge derived from traditional seismic source zoning. This knowledge, which encapsulates decades of geological, seismological, and geophysical expertise, is incorporated into our model through the construction of informative priors. Our approach thus bridges modern Bayesian learning with historical seismological insights, offering a coherent way to address the limitations and unify zoned and zoneless methodologies.

---

## Modèle GP-ETAS informatif

Soient des événements observés

$$D = \{(t_i, x_i, y_i, m_i)\}_{i=1}^N, \quad (x_i, y_i) \in (\mathcal{X}, \mathcal{Y}) \subset \mathbb{R}^2, \quad t_i \in (0, T], \quad m_i \in \mathbb{R}_+.$$

Le modèle GP-ETAS est un processus de Hawkes spatio-temporel marqué dont la fonction d'intensité conditionnelle du quadriplet  $(t, x, y, m)$  s'écrit :

$$\boxed{\lambda^*(t, x, y, m \mid \mathcal{H}_t) = p_M(m) \times \lambda(t, x, y \mid \mathcal{H}_t)} \quad (31)$$

où  $p_M$  est la loi conditionnelle des magnitudes et  $\lambda(t, x, y \mid \mathcal{H}_t)$  l'intensité d'occurrence temps-espace.

### I – Loi des magnitudes $p_M$

Les magnitudes suivent la loi Gutenberg–Richter tronquée :

$$p_M(m) = \begin{cases} \frac{\beta \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{\max} - m_c)]} & \text{si } m_c \leq m \leq m_{\max}, \\ 0 & \text{sinon} \end{cases} \quad \text{avec } \beta = \ln(10)b.$$

La loi  $p_M$  est indépendante de  $(t, x, y)$  et identique pour les chocs de fond et déclenchés.

### II – Intensité spatio-temporelle

Elle est donnée par :

$$\lambda(t, x, y \mid \mathcal{H}_t) = \mu(x, y) + \sum_{j: t_j \in \mathcal{H}_t} \phi(t - t_j, x - x_j, y - y_j, m_j)$$

- $\mu(x, y)$  est l'intensité de fond (supposé homogène en temps)
- $\phi$  est le noyau d'excitation/de déclenchement (productivité × temps × espace)

#### Modélisation du noyau d'excitation

Le noyau d'excitation dans le cadre du modèle GP-ETAS s'écrit :

$$\phi(t - t_j, x - x_j, y - y_j, m_j) = \phi_m(m_j) \times \phi_t(t - t_j) \times \phi_s(x - x_j, y - y_j \mid m_j)$$

Noyau de productivité (nombre moyen de descendants directs) :

$$\phi_m(m_j) = A e^{\alpha(m_j - m_c)}, \quad A > 0, \quad \alpha \geq 0$$

Noyau temporelle (Omori–Utsu modifiée) :

$$\phi_t(t - t_j) = (p - 1) c^{p-1} (t - t_j + c)^{-p}, \quad c > 0, p > 1$$

Noyau spatial (isotrope de type Student) : En posant  $r^2 = (x - x_j)^2 + (y - y_j)^2$  et

$$R(m_j) = d e^{\gamma(m_j - m_c)}, \quad d > 0, \gamma \geq 0,$$

on a :

$$\phi_s(x - x_j, y - y_j | m_j) = \frac{q-1}{\pi R(m_j)} \left[ 1 + \frac{r^2}{R(m_j)} \right]^{-q}, \quad q > 1.$$

Ce noyau est normalisé sur  $\mathbb{R}^2$  et est isotrope (-> pas d'information explicite sur les failles dans la région et n'élimine pas la possibilité d'avoir une réplique exactement au même endroit que son parent (peu probable))

### Modélisation de l'intensité de fond

Le taux de fond est défini par une transformation sigmoïdale d'un processus latent  $f(x, y)$  :

$$\mu(x, y) = \tilde{\mu} \sigma(f(x, y)) = \frac{\tilde{\mu}}{1 + e^{-f(x, y)}}, \quad \tilde{\mu} > 0. \quad (32)$$

Moyenne du processus gaussien :

Contrairement au GP standard, la moyenne du processus n'est pas nulle, elle est définie comme une somme finie de composantes uniformes pondérées par des intensités non connues :

$$m(x, y) = \sum_{j=1}^J \epsilon_j u_j(x, y), \quad (33)$$

où  $u_j(x, y)$  est la fonction indicatrice de la sous-région  $S_j \subset \mathcal{X}$  :

$$u_j(x, y) = \begin{cases} 1 & \text{si } (x, y) \in S_j, \\ 0 & \text{sinon} \end{cases}$$

Les sous-régions  $\{S_j\}_{j=1}^J$  forment une partition de l'espace d'étude  $\mathcal{X} \times \mathcal{Y}$  :

$$S_j \cap S_{j'} = \emptyset \quad \text{si } j \neq j', \quad \text{et} \quad \bigcup_{j=1}^J S_j = \mathcal{X} \times \mathcal{Y}.$$

Les coefficients d'intensité  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_J)^\top$  sont considérés comme des variables aléatoires suivant un prior gaussien :

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_J, \mathbf{I}_J), \quad (34)$$

où  $I_J$  désigne la matrice identité de taille  $J \times J$ .

Ainsi, la moyenne du GP devient elle-même aléatoire et reflète une incertitude spatiale sur le niveau de fond dans chacune des zones  $S_j$ . D'où, conditionnellement à  $\epsilon$  et aux hyperparamètres  $\nu$ , on a :

$$f \sim \mathcal{GP}(m(x, y), k((x, y), (x', y') | \nu)), \quad (35)$$

$k$  est un noyau gaussien :

$$k((x, y), (x', y') | \nu) = \nu_0^2 \exp\left(-\frac{(x - x')^2}{2\nu_x^2} - \frac{(y - y')^2}{2\nu_y^2}\right), \quad \nu = (\nu_0, \nu_x, \nu_y) \in \mathbb{R}_+^3. \quad (36)$$

Écriture hiérarchique du modèle :

$$\begin{aligned} \nu &\sim p_\nu && \text{(loi exponentielle)} \\ \tilde{\mu} &\sim p_{\tilde{\mu}} && \text{(loi gamma)} \\ \boldsymbol{\theta}_\phi &\sim p_{\boldsymbol{\theta}_\phi} && \text{(loi uniforme)} \\ \epsilon &\sim \mathcal{N}(\mathbf{0}_J, \mathbf{I}_J) \\ f | \epsilon, \nu &\sim \mathcal{GP}(m(x, y), k((x, y), (x', y') | \nu)), \\ \mu(x, y) &= \tilde{\mu} \sigma(f(x, y)), \\ \beta &\sim p_\beta \\ D | \beta, \mu, \boldsymbol{\theta}_\phi &\sim \text{Hawkes}(\lambda^\star(t, x, y, m)), \end{aligned}$$

avec  $\lambda^\star(t, x, y, m)$  défini par (67).

*Remarque :* Si on fixe  $\phi(\cdot)$  à 0, ce modèle se réduit à un SGCP "à moyenne informative", c'est un processus de Poisson inhomogène dont l'intensité spatiale est guidée par un processus gaussien qui incorpore une structure de zonage uniforme sur les régions  $S_j$  (intensités inconnues).

### Expression finale de la fonction d'intensité conditionnelle

Ainsi, on a :

$$\begin{aligned} \lambda^\star(t, x, y, m | \mathcal{H}_t) &= \frac{\beta \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{\max} - m_c)]} \mathbb{1}_{\{m_c \leq m \leq m_{\max}\}} \times \\ &\left[ \tilde{\mu} \sigma(f(x, y)) + \sum_{j: t_j \in \mathcal{H}_t} A e^{\alpha(m_j - m_c)} (p - 1) c^{p-1} (t - t_j + c)^{-p} \times \frac{q - 1}{\pi R(m_j)} \left(1 + \frac{r^2}{R(m_j)}\right)^{-q} \right] \end{aligned} \quad (37)$$

## III – Vraisemblance et posterior du modèle

Le vecteur de paramètres à estimer pour notre modèle est :

$$\boldsymbol{\theta} = (\beta, \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\phi)$$

où :

$$\boldsymbol{\theta}_\mu = (\tilde{\mu}, f, \boldsymbol{\epsilon}, \boldsymbol{\nu}) \quad \text{et} \quad \boldsymbol{\theta}_\phi = (A, \alpha, c, p, d, q, \gamma)$$

avec :

- $\beta$  : paramètre de magnitude
- $\tilde{\mu}$  : niveau du fond
- $A, \alpha$  : paramètres de productivité et dépendance à la magnitude
- $c, p$  : paramètres du noyau temporel d'Omori–Utsu
- $d, q, \gamma$  : paramètres du noyau spatial (dépendance à la magnitude via  $R(m) = d e^{\gamma(m-m_c)}$ )

La vraisemblance complète est :

$$L_N(\boldsymbol{\theta}) = \left[ \prod_{i=1}^N \lambda^*(t_i, x_i, y_i, m_i \mid \mathcal{H}_{t_i}) \right] \exp \left( - \int_0^T \int_{\mathcal{X}, \mathcal{Y}} \int_{m_c}^{m_{\max}} \lambda^*(t, x, y, m \mid \mathcal{H}_t) dm dx dy dt \right).$$

En substituant (67) et en utilisant le fait que  $\int_{m_c}^{m_{\max}} p_M(m) dm = 1$ , on a :

$$L_N(\boldsymbol{\theta}) = \underbrace{\prod_{i=1}^N p_M(m_i)}_{\text{facteurs magnitudes}} \times \underbrace{\left[ \prod_{i=1}^N \lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i}) \right] \exp \left( - \int_0^T \int_{\mathcal{X}, \mathcal{Y}} \lambda(t, x, y \mid \mathcal{H}_t) dx dy dt \right)}_{\text{facteurs spatio-temporels}}.$$

Facteurs magnitudes :

Avec

$$p_M(m) = \frac{\beta e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max}-m_c)}}, \quad \beta = \ln(10) b,$$

on a :

$$\prod_{i=1}^N p_M(m_i) = \left( \frac{\beta}{1 - e^{-\beta(m_{\max}-m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right).$$

Facteurs spatio-temporels - produit des intensités :

On a :

$$\lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i}) = \tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j: t_j < t_i} \phi(t_i - t_j, x_i - x_j, y_i - y_j \mid m_j),$$

d'où :

$$\prod_{i=1}^N \lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i}) = \prod_{i=1}^N \left[ \tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j: t_j < t_i} \phi_{ij} \right],$$

avec :

$$\begin{aligned} \phi_{ij} &= A e^{\alpha(m_j - m_c)} (p - 1) c^{p-1} (t_i - t_j + c)^{-p} \frac{q - 1}{\pi R(m_j)} \left( 1 + \frac{r_{ij}^2}{R(m_j)} \right)^{-q}, \\ R(m_j) &= d e^{\gamma(m_j - m_c)}, \quad r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2. \end{aligned}$$

### Facteurs spatio-temporels - intégrale de l'intensité :

On sépare la contribution du taux de fond et celle du processus d'excitation.

- **Taux de fond** : En intégrant sur le temps et l'espace :

$$\int_0^T \int_{\mathcal{X}, \mathcal{Y}} \mu(x, y) \, dx \, dy \, dt = \tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) \, dx \, dy. \quad (38)$$

Cette intégrale est intractable (car  $\sigma(f(x, y))$  dépend du GP).

- **Composante d'excitation** : Comme précédemment, en appliquant Fubini :

$$\int_0^T \int_{\mathcal{X}, \mathcal{Y}} \sum_{j: t_j < t} \phi(t - t_j, x - x_j, y - y_j, m_j) \, dx \, dy \, dt = \sum_{j=1}^N A e^{\alpha(m_j - m_c)} T_j S_j,$$

avec

$$T_j = 1 - \left( \frac{c}{T - t_j + c} \right)^{p-1}, \quad S_j = \int_{\mathcal{X}, \mathcal{Y}} \frac{q-1}{\pi R(m_j)} \left( 1 + \frac{r^2}{R(m_j)} \right)^{-q} \, dx \, dy.$$

Si  $\mathcal{X}, \mathcal{Y}$  couvre toute la zone d'étude, alors  $S_j \approx 1$ .

### **Expression étendue de la vraisemblance**

En agrégeant les termes précédents, la vraisemblance du modèle GP-ETAS informatif s'écrit :

$$\begin{aligned} L_N(\boldsymbol{\theta}) &= \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right) \times \prod_{i=1}^N \left[ \tilde{\mu} \sigma(f(x_i, y_i)) \right. \\ &\quad \left. + \sum_{j: t_j < t_i} A e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \times \frac{q-1}{\pi R(m_j)} \left( 1 + \frac{r_{ij}^2}{R(m_j)} \right)^{-q} \right] \\ &\quad \times \exp \left( -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) \, dx \, dy - \sum_{j=1}^N A e^{\alpha(m_j - m_c)} T_j S_j \right) \end{aligned}$$

Les priors sont :

$$\begin{aligned} f \mid w, \nu &\sim \mathcal{GP}(m(\cdot; w), k(\cdot, \cdot \mid \nu)), & \epsilon &\sim \mathcal{N}(0, I_J), \\ \boldsymbol{\nu} &\sim p_{\boldsymbol{\nu}}, & \tilde{\mu} &\sim p_{\tilde{\mu}}, & \beta &\sim p_{\beta}, & \boldsymbol{\theta}_{\phi} &\sim p_{\boldsymbol{\theta}_{\phi}}. \end{aligned}$$

### **Expression finale de la postérior**

$p(\boldsymbol{\theta} \mid D) = \underbrace{L_N(\boldsymbol{\theta})}_{\text{vraisemblance}} \times \underbrace{p(f \mid \boldsymbol{\epsilon}, \boldsymbol{\nu})}_{\text{prior GP}} \times \underbrace{p(\boldsymbol{\epsilon})}_{\text{gaussien}} \times \underbrace{p(\boldsymbol{\nu})}_{\text{hyperparams noyau}}$
$\times \underbrace{p(\tilde{\mu})}_{\text{taux de fond}} \times \underbrace{p(\boldsymbol{\theta}_{\phi})}_{\text{noyau auto-excitation}} \times \underbrace{p(\beta)}_{\text{G-R}} \times \underbrace{C}_{\text{const. normalisation}}$

Cette posterior est doublement intractable : une première fois dans la vraisemblance avec l'intégrale (75) et une seconde fois dans la constante de normalisation  $C$  où la même intégrale apparaît.

## IV – Augmentation de données

### Augmentation par processus de branchement

Pour chaque événement  $i$ , on introduit  $z_i \in \{0, 1, \dots, i-1\}$  :  $z_i = 0$  si  $i$  est un choc de fond et  $z_i = j$  si  $i$  est déclenché par  $j$  ( $1 \leq j \leq i-1$ ). On note  $Z = (z_1, \dots, z_N)$  et on pose (pour  $i > j$ )

$$\phi_{ij}(\boldsymbol{\theta}_\phi) := \phi(t_i - t_j, x_i - x_j, y_i - y_j, m_j), \quad \phi_i(\boldsymbol{\theta}_\phi) := \phi(t - t_i, x - x_i, y - y_i, m_i),$$

et  $T_i := [t_i, T]$ .

On suppose  $p(Z) = \text{cste}$  (a priori uniforme). Alors, pour chaque  $i$  :

$$\lambda(t_i, x_i, y_i | \mathcal{H}_{t_i}) = \left( \tilde{\mu} \sigma(f(x_i, y_i)) \right) + \sum_{j=1}^{i-1} \phi_{ij}(\boldsymbol{\theta}_\phi) \quad (39)$$

$$= \left( \tilde{\mu} \sigma(f(x_i, y_i)) \right)^{\mathbb{1}_{(z_i=0)}} \times \left( \prod_{j=1}^{i-1} \phi_{ij}(\boldsymbol{\theta}_\phi) \right)^{\mathbb{1}_{(z_i=j)}}. \quad (40)$$

Log-vraisemblance augmentée :

On a :

$$\begin{aligned} L_N(\boldsymbol{\theta}, Z) &= \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right) \\ &\times \prod_{i=1}^N \left( \tilde{\mu} \sigma(f(x_i, y_i)) \right)^{\mathbb{1}_{(z_i=0)}} \times \exp \left( -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy \right) \\ &\times \prod_{i=1}^N \prod_{j=1}^{i-1} \phi_{ij}(\boldsymbol{\theta}_\phi)^{\mathbb{1}_{(z_i=j)}} \times \exp \left( - \sum_{i=1}^N A e^{\alpha(m_i - m_c)} T_i S_i \right) \\ &\times p(Z). \end{aligned} \quad (41)$$

Loi conditionnelle de  $z_i$  :

En ne gardant que les termes dépendant de  $z_i$  dans (77),

$$p(z_i = j | D, \boldsymbol{\theta}) \propto \begin{cases} \tilde{\mu} \sigma(f(x_i, y_i)) & j = 0, \\ \phi_{ij}(\boldsymbol{\theta}_\phi) & j \in \{1, \dots, i-1\} \end{cases}$$

En normalisant, on obtient une loi catégorielle sur  $\{0, \dots, i-1\}$  :

$$p(z_i = 0 | \cdot) = \frac{\tilde{\mu} \sigma(f(x_i, y_i))}{\tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j < i} \phi_{ij}(\boldsymbol{\theta}_\phi)}, \quad p(z_i = j | \cdot) = \frac{\phi_{ij}(\boldsymbol{\theta}_\phi)}{\tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j < i} \phi_{ij}(\boldsymbol{\theta}_\phi)}.$$

### Augmentation par processus de Poisson marqué et Pólya–Gamma

Pour  $z \in \mathbb{R}$ , on a la représentation :

$$\begin{aligned} \sigma(z) &= \int_0^\infty \exp\left(\frac{z}{2} - \frac{z^2}{2}\omega - \ln 2\right) p_{\text{PG}}(\omega | 1, 0) d\omega, \\ \sigma(-z) &= \int_0^\infty \exp\left(-\frac{z}{2} - \frac{z^2}{2}\omega - \ln 2\right) p_{\text{PG}}(\omega | 1, 0) d\omega. \end{aligned}$$

On pose, pour tout  $\omega \geq 0$  et  $z \in \mathbb{R}$ ,

$$f_{\text{PG}}(\omega, z) := \frac{z}{2} - \frac{z^2}{2}\omega - \ln 2.$$

On note la contribution de l'intensité de fond à la vraisemblance augmentée par processus de branchement (cf. (78)) :

$$L(D_0 | f, \tilde{\mu}) = \prod_{i=1}^N (\tilde{\mu} \sigma(f_i))^{\mathbb{1}_{(z_i=0)}} \times \exp\left(-\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy\right), \quad (42)$$

où  $f_i := f(x_i, y_i)$  et  $D_0 = \{i : z_i = 0\}$  désigne les chocs de fond.

Représentation continue (partie intégrale) :

En utilisant la représentation Pólya–Gamma et l'identité  $1 - \sigma(-z) = \sigma(z)$ , on réécrit :

$$\begin{aligned} -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy &= -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} [1 - \sigma(-f(x, y))] dx dy. \\ &= -\tilde{\mu} T \int_{\mathcal{X} \times \mathcal{Y} \times \mathbb{R}_+} \left[1 - \exp[f_{\text{PG}}(\omega, -f(x, y))]\right] p_{\text{PG}}(\omega | 1, 0) d\omega dx dy. \end{aligned} \quad (43)$$

On introduit alors l'espace produit :

$$\widehat{\mathcal{S}} := \mathcal{X} \times \mathcal{Y} \times \mathbb{R}_+, \quad \Lambda(x, y, \omega) = \tilde{\mu} T p_{\text{PG}}(\omega | 1, 0),$$

et le processus de Poisson marqué  $\Pi_{\widehat{\mathcal{S}}}$  d'intensité  $\Lambda$ . Le théorème de Campbell donne :

$$\exp\left(-\int_{\widehat{\mathcal{S}}} [1 - e^{f_{\text{PG}}(\omega, -f(x, y))}] \Lambda(x, y, \omega) d\omega dx dy\right) = \mathbb{E}_{P_\Lambda} \left[ \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} \right]. \quad (44)$$

Représentation discrète (partie produit) :

Pour la produit dans (79), la représentation PG donne :

$$\begin{aligned} \prod_{i:z_i=0} \tilde{\mu} \sigma(f_i) &= \prod_{i:z_i=0} \tilde{\mu} \int_0^\infty e^{f_{\text{PG}}(\omega_i, f_i)} p_{\text{PG}}(\omega_i | 1, 0) d\omega_i \\ &= \mathbb{E}_{\omega_i \sim PG(1,0)} \left[ \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right] \end{aligned} \quad (45)$$

Vraisemblance augmentée par processus de Poisson marqué et variables Polya–Gamma :

En combinant la partie continue (81) et la partie discrète (82), on a :

$$\begin{aligned} L(D_0 | f, \tilde{\mu}) &= \prod_{i=1}^N (\tilde{\mu} \sigma(f_i))^{\mathbb{1}_{(z_i=0)}} \times \exp \left( -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy \right), \\ &= \mathbb{E}_{P_\Lambda} \left[ \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m))} \right] + \mathbb{E}_{\omega_i \sim PG(1,0)} \left[ \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right] \\ &= \int \left( \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m))} \right) P_\Lambda(d\Pi_{\widehat{\mathcal{S}}}) \times \int \left( \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) P_{\omega_{N_0}}(d\omega_{N_0}) \\ &= \int \left( \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m))} \right) \left( \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) P_{\text{aug}}(d\Pi_{\widehat{\mathcal{S}}}, d\omega_{N_0}) \end{aligned} \quad (46)$$

où la mesure des variables augmentées est :

$$P_{\text{aug}} = P_\Lambda \otimes P_{\omega_{D_0}}, \quad \text{avec} \quad P_{\omega_{D_0}} = \prod_{i:z_i=0} p_{\text{PG}}(\omega_i | 1, 0) d\omega_i.$$

Or, on peut écrire :

$$L(D_0 | f, \tilde{\mu}) = \int L(D_0, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi | f, \tilde{\mu}) d\Pi_{\widehat{\mathcal{S}}} d\omega_{D_0}$$

Donc, par identification on a :

$$L(D_0, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi | f, \tilde{\mu}) = \left( \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} \right) \left( \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) \quad (47)$$

$$\times p(\Pi_{\widehat{\mathcal{S}}}) p(\omega_{D_0}), \quad (48)$$

où :

$$p(\omega_{D_0}) = \prod_{i:z_i=0} p_{\text{PG}}(\omega_i | 1, 0)$$

et où :

$$p(\Pi_{\hat{\mathcal{S}}}) = \exp\left(-\int_{\hat{\mathcal{S}}} \Lambda(x, y, \omega) dx dy d\omega\right) \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} \Lambda(x_m, y_m, \omega_m).$$

En utilisant ici

$$\Lambda(x, y, \omega) = \tilde{\mu} T p_{\text{PG}}(\omega | 1, 0) \quad \text{et} \quad \int_{\hat{\mathcal{S}}} \Lambda(x, y, \omega) = \tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|$$

on a finalement :

$$\begin{aligned} L(D_0, \omega_{D_0}, \Pi_{\hat{\mathcal{S}}}, \omega_{\Pi} | f, \tilde{\mu}) &= \left( \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} \right) \left( \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) \\ &\times \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} \tilde{\mu} T p_{\text{PG}}(\omega_m | 1, 0) \prod_{i:z_i=0} p_{\text{PG}}(\omega_i | 1, 0) \\ &= \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} p_{\text{PG}}(\omega_i | 1, 0) \quad (49) \\ &\quad \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} \tilde{\mu} T e^{f_{\text{PG}}(\omega_m, -f_m)} p_{\text{PG}}(\omega_m | 1, 0) \end{aligned}$$

Finalement, les structures linéaire et quadratique en  $f$  des termes  $f_i/2$  et  $f_i^2 \omega_i/2$  dans le résultat ci-dessus montre que la vraisemblance augmentée est **gaussienne en  $f$** . Ainsi, le modèle augmenté est conjugué au prior Gaussien (GP) et permet inférence bayésienne tractable.

En récrivant entièrement la vraisemblance augmentée finale du modèle :

$$\begin{aligned} L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\hat{\mathcal{S}}}, \omega_{\Pi}, f) &= \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp\left(-\beta \sum_{i=1}^N (m_i - m_c)\right) \times \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \\ &\times \prod_{i:z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} p_{\text{PG}}(\omega_i | 1, 0) \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} \tilde{\mu} T e^{f_{\text{PG}}(\omega_m, -f_m)} p_{\text{PG}}(\omega_m | 1, 0) \\ &\times \prod_{i=1}^N \prod_{j=1}^{i-1} \phi_{ij}(\theta_{\phi})^{\mathbb{1}_{(z_i=j)}} \times \exp\left(-\sum_{i=1}^N A e^{\alpha(m_i - m_c)} T_i S_i\right) \\ &\times p(Z). \quad (50) \end{aligned}$$

Posterior finale augmentée :

On en déduit :

$$p(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f | D) = L_N(\boldsymbol{\theta}, Z, \Pi_{\widehat{\mathcal{S}}}, \omega_{N_0}) \times p(f | \boldsymbol{\epsilon}, \boldsymbol{\nu}) \times p(\boldsymbol{\epsilon}) \\ \times p(\boldsymbol{\nu}) \times p(\tilde{\mu}) \times p(\beta) \times p(\boldsymbol{\theta}_{\phi}) \times C$$

## V – Mise à jour des posteriors conditionnelles

- Comme  $\tilde{\mu} \sim \text{Gamma}(a_{\mu}, b_{\mu})$ , en notant

$$N_0 = |\{i : z_i = 0\}|, \quad N_{\Pi} = |\{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}\}|$$

on a :

$$p(\tilde{\mu} | \boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \propto L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \times p(\tilde{\mu}) \\ \propto \tilde{\mu}^{N_0} \tilde{\mu}^{N_{\Pi}} \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \times \tilde{\mu}^{a_{\mu}-1} \exp(-b_{\mu} \tilde{\mu}) \quad (51) \\ \propto \tilde{\mu}^{N_0 + N_{\Pi} + a_{\mu} - 1} \exp\left[-\tilde{\mu} (T |\mathcal{X}| |\mathcal{Y}| + b_{\mu})\right]$$

Danc la loi conditionnelle est :

$$\tilde{\mu} | D, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f \sim \text{Gamma}\left(a_{\mu} + N_0 + N_{\Pi}, b_{\mu} + T |\mathcal{X}| |\mathcal{Y}|\right)$$

- Pour chaque  $i$  tel que  $z_i = 0$  :

$$p(\omega_i | \boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \propto \exp\left(-\frac{f_i^2}{2} \omega_i\right) p_{\text{PG}}(\omega_i | 1, 0), \quad f_i = f(x_i, y_i).$$

Donc :

$$\omega_i | D, Z, f \sim \text{PG}(1, f(x_i, y_i))$$

et, par indépendance :

$$\omega_{D_0} | D, Z, f \sim \bigotimes_{i:z_i=0} \text{PG}(1, f(x_i, y_i)).$$

- De manière similaire à ce qu'il y a juste au dessus, pour chaque  $m$  tq  $(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}$ , on a :

$$\omega_m | D, \Pi_{\widehat{\mathcal{S}}}, f \sim \text{PG}(1, f(x_m, y_m))$$

Ainsi, par indépendance on en déduit :

$$\omega_{\Pi} | D, \Pi_{\widehat{\mathcal{S}}}, f \sim \bigotimes_{(x_m, y_m) \in \Pi_{\widehat{\mathcal{S}}}} \text{PG}(1, |f(x_m, y_m)|).$$

- On a :

$$p(\Pi_{\hat{\mathcal{S}}} \mid \boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\hat{\mathcal{S}}}, \omega_{\Pi}, f) \propto \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} \tilde{\mu} T e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} p_{\text{PG}}(\omega_m \mid 1, 0)$$

on reconnaît la densité non normalisée d'un processus de Poisson inhomogène marqué d'intensité  $\Lambda_{\text{post}}(x, y, \omega)$  tq :

$$\Lambda_{\text{post}}(x, y, \omega) = \tilde{\mu} T e^{f_{\text{PG}}(\omega, -f(x, y))} p_{\text{PG}}(\omega \mid 1, 0)$$

Ainsi, on a :

$$\boxed{\Pi_{\hat{\mathcal{S}}} \mid D, f, \tilde{\mu} \sim \text{PP}(\Lambda_{\text{post}}(x, y, \omega)), \quad \Lambda_{\text{post}}(x, y, \omega) = \tilde{\mu} T e^{f_{\text{PG}}(\omega, -f(x, y))} p_{\text{PG}}(\omega \mid 1, 0).}$$

En intégrant la fonction d'intensité contre les marques  $\omega$ , on obtient l'intensité marginale suivante :

$$\lambda^{\text{post}}(x, y) = \tilde{\mu} T \int_0^\infty e^{f_{\text{PG}}(\omega, -f(x, y))} p_{\text{PG}}(\omega \mid 1, 0) d\omega \quad (52)$$

$$= \tilde{\mu} T \sigma(-f(x, y)) \quad (53)$$

Ce résultat est intéressant pour l'échantillonnage spatial des évènements, on peut passer par l'algorithme classique de thinning et ajouter la marque de manière indépendante.

- On note

$$D_f = D_0 \cup \Pi_{\hat{\mathcal{S}}},$$

l'union des événements chocs de fond observés et des événements latents du processus de Poisson inhomogène marqué. Le processus gaussien a priori est

$$f \sim \mathcal{GP}(m(x, y), k((x, y), (x', y') \mid \nu))$$

donc on écrit :

$$p(f \mid \boldsymbol{\epsilon}, \boldsymbol{\nu}) = \mathcal{N}(f \mid m_f, K_{f,f}),$$

où

$$m_f = m(x, y \mid D_f, \boldsymbol{\epsilon}), \quad K_{f,f} = [K_{f,f}]_{ij} = k(x_i, x_j \mid \boldsymbol{\nu}, D_f).$$

On a donc :

$$\begin{aligned} p(f \mid \boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\hat{\mathcal{S}}}, \omega_{\Pi}, f, D_f) &\propto L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\hat{\mathcal{S}}}, \omega_{\Pi}, f) \times p(f \mid \boldsymbol{\epsilon}, \boldsymbol{\nu}) \\ &\propto \prod_{i: z_i=0} e^{f_i/2 - f_i^2 \omega_i/2} \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} e^{-f_m/2 - f_m^2 \omega_m/2} \times p(f \mid \boldsymbol{\epsilon}, \boldsymbol{\nu}) \\ &\propto \exp\left(-\frac{1}{2} f^\top \Omega f + u^\top f\right) \times \exp\left(-\frac{1}{2} (f - m_f)^\top K_{f,f}^{-1} (f - m_f)\right) \\ &\propto \exp\left(-\frac{1}{2} f^\top \Omega f + u^\top f\right) \times \exp\left(-\frac{1}{2} f^\top K_{f,f}^{-1} f + f^\top K_{f,f}^{-1} m_f\right) \\ &\propto \exp\left(-\frac{1}{2} f^\top \Omega f - \frac{1}{2} f^\top K_{f,f}^{-1} f + f^\top K_{f,f}^{-1} m_f + u^\top f\right) \\ &\propto \exp\left(-\frac{1}{2} f^\top (\Omega + K_{f,f}^{-1}) f + f^\top (K_{f,f}^{-1} m_f + u)\right). \end{aligned}$$

où :

$$\Omega = \text{diag}(\omega_{N_0}, \omega_\Pi), \quad u_k = \begin{cases} \frac{1}{2}, & \text{si } f_k \text{ vient d'un } i \text{ tq } z_i = 0 \\ -\frac{1}{2}, & \text{si } f_k \text{ vient d'un point de } \Pi_{\widehat{\mathcal{S}}} \end{cases}$$

En complétant le carré, on reconnaît une loi normale tq :

$$f \mid Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi, f, D_f, \tilde{\mu}, \boldsymbol{\epsilon}, \boldsymbol{\nu} \sim \mathcal{N}(f \mid \mu_f, \Sigma_f)$$

où

$$\Sigma_f = \left( \Omega + K_{f,f}^{-1} \right)^{-1} \quad \mu_f = \Sigma_f \left( K_{f,f}^{-1} m_f + u \right).$$

*Remarque :* En fixant  $m_f \geq 0$  (cas du prior non informatif), on retombe bien sur la posterior conditionnelle du GP-ETAS standard.

- On rappelle que la moyenne informative du processus gaussien est définie par

$$m(x, y) = \sum_{j=1}^J \epsilon_j u_j(x, y),$$

où  $u_j(x, y)$  est l'indicatrice de la sous-région  $S_j$ . Sur l'ensemble des points où le GP est évalué ( $D_f = D_0 \cup \Pi_{\widehat{\mathcal{S}}}$ ), on réécrit la moyenne sous forme matricielle :

$$m(D_f \mid \boldsymbol{\epsilon}) = U \boldsymbol{\epsilon},$$

où  $U = [U_{kj}]$  est la matrice de taille  $|D_f| \times J$  définie par

$$U_{kj} = u_j(x_k, y_k) = \begin{cases} 1, & \text{si } (x_k, y_k) \in S_j, \\ 0, & \text{sinon} \end{cases}$$

On a :

$$\begin{aligned} p(\boldsymbol{\epsilon} \mid \boldsymbol{\theta}, Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi, f, D_f) &\propto p(f \mid \boldsymbol{\epsilon}, \boldsymbol{\nu}) p(\boldsymbol{\epsilon}) \\ &\propto \exp\left(-\frac{1}{2}(f - U\boldsymbol{\epsilon})^\top K^{-1}(f - U\boldsymbol{\epsilon})\right) \times \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}\right) \\ &\propto \exp\left(-\frac{1}{2}(f - U\boldsymbol{\epsilon})^\top K^{-1}(f - U\boldsymbol{\epsilon}) - \frac{1}{2}\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}\right) \\ &\propto \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^\top (U^\top K^{-1}U + I_J)\boldsymbol{\epsilon} + \boldsymbol{\epsilon}^\top U^\top K^{-1}f\right) \end{aligned} \tag{54}$$

En complétant le carré, on reconnaît une loi normale tq :

$$\boldsymbol{\epsilon} \mid Z, f, D_f, \boldsymbol{\nu} \sim \mathcal{N}(\mu_{\boldsymbol{\epsilon}}, \Sigma_{\boldsymbol{\epsilon}})$$

où

$$\Sigma_{\boldsymbol{\epsilon}} = (U^\top K^{-1}U + I_J)^{-1}, \quad \mu_{\boldsymbol{\epsilon}} = \Sigma_{\boldsymbol{\epsilon}} U^\top K^{-1}f.$$

- On a :

$$\begin{aligned}
 p(\beta \mid \boldsymbol{\theta}, Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) &\propto L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \times p(\beta) \\
 &\propto \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right) \times p(\beta) \\
 &\propto
 \end{aligned} \tag{55}$$

Pas de forme conjuguée pour  $\beta$ , on va devoir faire appelle à du rejet avec Metropolis-Hastings.

- On a :

$$\begin{aligned}
 p(\boldsymbol{\nu} \mid f, \boldsymbol{\theta} \setminus \boldsymbol{\nu}, Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, S_f) &\propto p(f \mid \boldsymbol{\epsilon}, \boldsymbol{\nu}) \times p_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \\
 &\propto \mathcal{N}(f \mid m_f, K_{f,f}(\boldsymbol{\nu})) \times p_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \\
 &\propto |K_{f,f}|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (f - m_f)^{\top} K_{f,f}^{-1} (f - m_f) - \lambda_{\boldsymbol{\nu}} \boldsymbol{\nu} \right)
 \end{aligned} \tag{56}$$

Ici également la mise à jour doit se faire via une étape de type M-H dans l'algorithme de Gibbs.

-

## AutoCorrelation Function (ACF)

L'ACF mesure à quel point une chaîne MCMC est corrélée avec elle-même, à quel point la chaîne mélange bien (à quel point un échantillon à itération  $n\_iter + k$  apporte de l'info nouvelle par rapport à celui à l'itération  $n\_iter$ ) :

Soit une chaîne stationnaire (se fait post burn\_in) :

$$\{X_t\}_{t=1}^N$$

La fonction d'autocovariance en fonction du "lag"  $k$  s'écrit :

$$\gamma(k) = \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)]$$

où  $\mu = \mathbb{E}[X_t]$ . L'autocorrélation (ACF) est la version normalisée :

$$\text{ACF}(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)]}{\mathbb{E}[(X_t - \mu)^2]}$$

## Version empirique de l'ACF (pratique)

Soit une chaîne  $(x_1, \dots, x_n)$ . L'estimateur empirique de l'ACF au lag  $k$  est donné :

$$\widehat{\text{ACF}}(k) = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2},$$

où  $\bar{x}$  est moyenne empirique.

## Quelques propriétés de l'ACF

- $\text{ACF}(0) = 1$
- $-1 \leq \text{ACF}(k) \leq 1$  pour tout  $k$
- $\text{ACF}(k) \approx 1 \rightarrow$  forte dépendance
- $\text{ACF}(k) \approx 0 \rightarrow$  proche de indépendance
- $\text{ACF}(k) < 0.1$  avant  $k \approx 10 \rightarrow$  excellent mélange
- $\text{ACF}(k) < 0.1$  avant  $k \approx 30 \rightarrow$  mélange correct
- $\text{ACF}(k) > 0.5$  sur de nombreux lags  $\rightarrow$  mauvais mélange
- $\text{ACF}(k) \approx \text{const}$   $\rightarrow$  chaîne bloquée

### Lien avec la taille effective d'échantillon (ESS)

L'autocorrélation détermine le nombre d'échantillons effectivement indépendants :

$$\text{ESS} = \frac{N}{1 + 2 \sum_{k=1}^{\infty} \rho(k)}$$

Une décroissance lente de l'ACF implique un ESS bas, et donc un MCMC peu efficace.

---

Probabilistic Seismic Hazard Assessment (PSHA) is traditionally based on piecewise spatially and temporally homogeneous Poisson process assumptions. While this framework provides a clear and structured methodology, it tends to oversimplify the complex spatial variability of seismicity. To address these limitations, more recent approaches have moved toward zoneless methods, which aim for greater realism by estimating the spatial density of earthquake occurrences using kernel density estimation (KDE). However, KDE belongs to a frequentist framework that suffers from a kind of rigidity : parameters have to be fixed in advance by the model constructors, potentially leading to sub-optimal representations. Although adaptive KDE techniques exist to address this issue, they often come with significant computational costs. Moreover, previous studies have shown that KDE is not so well-suited for regions with low or diffuse seismicity, such as France.

To overcome these limitations, we focus on Bayesian nonparametric methods — notably Dirichlet Processes, Gaussian Processes, etc. — allowing the model to flexibly adapt its spatial structure and complexity directly to the data. The adopted Bayesian framework naturally integrates expert knowledge derived from traditional seismic source zoning. This knowledge, which encapsulates decades of geological, seismological, and geophysical expertise, is incorporated into our model through the construction of informative priors. Our approach thus bridges modern Bayesian learning with historical seismological insights, offering a coherent way to address the limitations and unify zoned and zoneless methodologies.

---

## Quelques éléments sur les GP



## Modèle GP-ETAS informatif (avec bords flexibles ou via partition de Voronoï)

Soient des événements observés

$$D = \{(t_i, x_i, y_i, m_i)\}_{i=1}^N, \quad (x_i, y_i) \in (\mathcal{X}, \mathcal{Y}) \subset \mathbb{R}^2, \quad t_i \in (0, T], \quad m_i \in \mathbb{R}_+.$$

Le modèle GP-ETAS est un processus de Hawkes spatio-temporel marqué dont la fonction d'intensité conditionnelle du quadriplet  $(t, x, y, m)$  s'écrit :

$$\boxed{\lambda^*(t, x, y, m | \mathcal{H}_t) = p_M(m) \times \lambda(t, x, y | \mathcal{H}_t)} \quad (57)$$

où  $p_M$  est la loi conditionnelle des magnitudes et  $\lambda(t, x, y | \mathcal{H}_t)$  l'intensité d'occurrence temps-espace.

### I – Loi des magnitudes $p_M$

Inchangé : les pistes de modification interviennent uniquement via la moyenne du GP.

### II – Intensité spatio-temporelle

Inchangé : les pistes de modification interviennent uniquement via la moyenne du GP.

#### Modélisation du noyau d'excitation

Inchangé : les pistes de modification interviennent uniquement via la moyenne du GP.

#### Modélisation de l'intensité de fond

Le taux de fond est défini par une transformation sigmoïdale d'un processus latent  $f(x, y)$  :

$$\mu(x, y) = \tilde{\mu} \sigma(f(x, y)) = \frac{\tilde{\mu}}{1 + e^{-f(x, y)}}, \quad \tilde{\mu} > 0. \quad (58)$$

Alternative 1 – Moyenne informative à bords flexibles : gaussienne sur les bords

Contrairement au GP standard, la moyenne du processus n'est pas nulle. On part d'une partition de l'espace d'étude  $\mathcal{D} := \mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^2$ , en  $J$  polygones :

$$\{S_j^{(0)}\}_{j=1}^J, \quad \bigcup_{j=1}^J S_j^{(0)} = \mathcal{D}, \quad S_j^{(0)} \cap S_{j'}^{(0)} = \emptyset \quad \text{pour } j \neq j'$$

Cette partition initiale représente notre carte sismotectonique, et l'on veut autoriser des déformations des bords (localement). On représente la partition par une sorte de graphe de sommets tq :

$$V^{(0)} = \{v_\ell^{(0)}\}_{\ell=1}^L, \quad v_\ell^{(0)} \in \mathcal{D} \subset \mathbb{R}^2,$$

et des arêtes entre ces sommets. On aurait donc des paramètres représentants les positions des sommets :

$$\vartheta := (v_1, \dots, v_L) \in (\mathbb{R}^2)^L.$$

Pour rendre les frontières flexibles (sans trop s'écartez de la partition de départ), on pose un prior gaussien sur les déplacements des sommets (en gardant à l'esprit qu'il faut contraindre les frontières à rester dans l'espace d'étude) :

$$v_\ell^{(i+1)} = v_\ell^{(i)} + \delta_\ell, \quad \delta_\ell \sim \mathcal{N}(0, \Sigma_\delta), \quad \ell = 1, \dots, L, \quad (59)$$

Ainsi, les frontières sont des variables latentes.

### Alternative 2 – Moyenne informative à bords flexibles : partition de Voronoï

Contrairement au GP standard, la moyenne du processus n'est pas nulle. On introduit une partition de Voronoï de l'espace d'étude  $\mathcal{D} := \mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^2$ , définie par un ensemble de  $J$  sites

$$\vartheta := (s_1, \dots, s_J), \quad s_j \in \mathcal{D}.$$

La cellule de Voronoï associée à  $s_j$  est

$$S_j(\vartheta) := \left\{ (x, y) \in \mathcal{D} : \| (x, y) - s_j \| \leq \| (x, y) - s_{j'} \|, \forall j' \neq j \right\}.$$

Les cellules convexes  $\{S_j(\vartheta)\}_{j=1}^J$  forment une partition de  $\mathcal{D}$ .

On définit :

$$u_j(x, y | \boldsymbol{\vartheta}) = \mathbb{1}_{\{(x, y) \in S_j(\boldsymbol{\vartheta})\}} \quad \text{et} \quad m(x, y | \boldsymbol{\epsilon}, \boldsymbol{\vartheta}) = \sum_{j=1}^J \epsilon_j u_j(x, y | \boldsymbol{\vartheta}), \quad (60)$$

où les coefficients  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_J)^\top$  contrôlent l'intensité par zone. Pour moduler la force du prior, on peut introduire un "paramètre de shrinkage"  $\tau$  tq :

$$\boldsymbol{\epsilon} | \tau \sim \mathcal{N}(\mathbf{0}_J, \tau^2 \mathbf{I}_J), \quad (61)$$

où  $\tau$  contrôle la confiance dans la structure du zonage :

$$\tau \approx 0 \rightarrow \epsilon_j \approx 0 \rightarrow m(x, y) \approx 0 \rightarrow \text{prior peu informatif}$$

$\tau$  grand  $\rightarrow \epsilon_j$  libres  $\rightarrow$  prior très informatif

Un prior peut-être posé sur  $\tau$ , par exemple une inverse Gamma :

$$\tau \sim \text{IG}(a_\tau, b_\tau), \quad (62)$$

ce qui permet au modèle d'ajuster automatiquement le degré de confiance accordé à la moyenne informative  $m(x, y)$  selon l'adéquation avec les données.

Ainsi, conditionnellement à  $(\boldsymbol{\epsilon}, \boldsymbol{\vartheta}, \boldsymbol{\nu})$ , on a :

$$f \sim \mathcal{GP}\left(m(x, y | \boldsymbol{\epsilon}, \boldsymbol{\vartheta}), k((x, y), (x', y') | \boldsymbol{\nu})\right). \quad (63)$$

$k$  est un noyau gaussien :

$$k((x, y), (x', y') | \boldsymbol{\nu}) = \nu_0^2 \exp\left(-\frac{(x - x')^2}{2\nu_1^2} - \frac{(y - y')^2}{2\nu_2^2}\right), \quad \boldsymbol{\nu} = (\nu_0, \nu_1, \nu_2) \in \mathbb{R}_+^3. \quad (64)$$

Écriture matricielle :

Sur l'ensemble des points où le processus gaussien est évalué ( $D_f = D_0 \cup \Pi_{\hat{S}}$ ), la moyenne s'écrit sous forme matricielle :

$$m(D_f | \boldsymbol{\epsilon}, \boldsymbol{\vartheta}) = U(\boldsymbol{\vartheta}) \boldsymbol{\epsilon},$$

où  $U(\boldsymbol{\vartheta}) = [U_{kj}(\boldsymbol{\vartheta})]$  est la matrice de taille  $|D_f| \times J$  définie par

$$U_{kj}(\boldsymbol{\vartheta}) = u_j(x_k, y_k) = \begin{cases} 1, & \text{si } (x_k, y_k) \in S_j(\boldsymbol{\vartheta}), \\ 0, & \text{sinon.} \end{cases}$$

Écriture hiérarchique du modèle :

$$\begin{aligned} \boldsymbol{\nu} &\sim p_{\boldsymbol{\nu}} && \text{(loi exponentielle)} \\ \tilde{\mu} &\sim p_{\tilde{\mu}} && \text{(loi gamma)} \\ \boldsymbol{\theta}_\phi &\sim p_{\boldsymbol{\theta}_\phi} && \text{(loi uniforme)} \\ \tau &\sim p_\tau && \text{(loi inverse-gamma)} \\ \boldsymbol{\epsilon} | \tau &\sim \mathcal{N}(\mathbf{0}_J, \tau^2 \mathbf{I}_J) \\ \boldsymbol{\vartheta} &\sim p_{\boldsymbol{\vartheta}} && \text{(prior sur les partitions } S_j) \\ f | \boldsymbol{\epsilon}, \boldsymbol{\vartheta}, \boldsymbol{\nu} &\sim \mathcal{GP}\left(m(\cdot | \boldsymbol{\epsilon}, \boldsymbol{\vartheta}), k(\cdot, \cdot | \boldsymbol{\nu})\right), \\ \mu(x, y) &= \tilde{\mu} \sigma(f(x, y)), \\ \beta &\sim p_\beta \\ D | \beta, \mu, \boldsymbol{\theta}_\phi &\sim \text{Hawkes}(\lambda^\star(t, x, y, m)) \end{aligned}$$

*Remarque :* Si on fixe  $\phi(\cdot)$  à 0, ce modèle se réduit à un SGCP "à moyenne informative", dans lequel la structure de zonage est flexible et éventuellement portée par une partition de Voronoï (sites  $\boldsymbol{\vartheta}$ ) et des poids aléatoires  $\boldsymbol{\epsilon}$ .

### Expression finale de la fonction d'intensité conditionnelle

Inchangé : les pistes de modification interviennent uniquement via la moyenne du GP.

## III – Vraisemblance et posterior du modèle

Le vecteur de paramètres à estimer pour notre modèle est :

$$\boldsymbol{\theta} = (\beta, \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\phi)$$

où :

$$\boldsymbol{\theta}_\mu = (\tilde{\mu}, f, \boldsymbol{\epsilon}, \boldsymbol{\nu}, \boldsymbol{\vartheta}) \quad \text{et} \quad \boldsymbol{\theta}_\phi = (A, \alpha, c, p, d, q, \gamma).$$

## IV – Agmentation de données

Inchangé : les pistes de modification interviennent uniquement via la moyenne du GP.

## V – Mise à jour des posteriors conditionnelles

- **Mises à jour inchangées**
- **Mise à jour de  $\boldsymbol{\vartheta}$  (bords flexibles / sites Voronoï)** Conditionnellement à  $(f, \boldsymbol{\epsilon}, \boldsymbol{\nu})$ , on a :

$$p(\boldsymbol{\vartheta} | f, \boldsymbol{\epsilon}, \boldsymbol{\nu}) \propto p(f | \boldsymbol{\epsilon}, \boldsymbol{\vartheta}, \boldsymbol{\nu}) p_\vartheta(\boldsymbol{\vartheta}), \quad (65)$$

où :

$$p(f | \boldsymbol{\epsilon}, \boldsymbol{\vartheta}, \boldsymbol{\nu}) = \mathcal{N}\left(f \mid U(\boldsymbol{\vartheta})\boldsymbol{\epsilon}, K_{f,f}(\boldsymbol{\nu})\right).$$

- **Mise à jour de  $\tau$**  Conditionnellement à  $\boldsymbol{\epsilon}$ , on a :

$$p(\tau | \boldsymbol{\epsilon}) \propto p(\boldsymbol{\epsilon} | D_f, f, \boldsymbol{\nu}, \boldsymbol{\vartheta}) p_\tau(\tau), \quad (66)$$

où :

$$p(\boldsymbol{\epsilon} | Z, f, D_f, \boldsymbol{\nu}, \boldsymbol{\vartheta}) = \mathcal{N}\left((U^\top(\boldsymbol{\vartheta})K^{-1}U(\boldsymbol{\vartheta}) + I_J)^{-1} U^\top(\boldsymbol{\vartheta})K^{-1}f, (U^\top(\boldsymbol{\vartheta})K^{-1}U(\boldsymbol{\vartheta}) + I_J)^{-1}\right)$$

et

$$p_\tau(\tau) = \mathcal{IG}(a_\tau, b_\tau)$$

Limites instrumentales/historiques dans les données : 01 janvier 1965

## Modèle GP-ETAS informatif (**DERNIÈRE VERSION EN DATE**)

Soient des événements observés

$$D = \{(t_i, x_i, y_i, m_i)\}_{i=1}^N, \quad (x_i, y_i) \in (\mathcal{X}, \mathcal{Y}) \subset \mathbb{R}^2, \quad t_i \in (0, T], \quad m_i \in \mathbb{R}_+.$$

Le modèle GP-ETAS est un processus de Hawkes spatio-temporel marqué dont la fonction d'intensité conditionnelle du quadriplet  $(t, x, y, m)$  s'écrit :

$$\boxed{\lambda^*(t, x, y, m | \mathcal{H}_t) = p_M(m) \times \lambda(t, x, y | \mathcal{H}_t)} \quad (67)$$

où  $p_M$  est la loi conditionnelle des magnitudes et  $\lambda(t, x, y | \mathcal{H}_t)$  l'intensité d'occurrence temps-espace.

### I – Loi des magnitudes $p_M$

Les magnitudes suivent la loi Gutenberg–Richter tronquée :

$$p_M(m) = \begin{cases} \frac{\beta \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{\max} - m_c)]} & \text{si } m_c \leq m \leq m_{\max}, \\ 0 & \text{sinon} \end{cases} \quad \text{avec } \beta = \ln(10) b.$$

La loi  $p_M$  est indépendante de  $(t, x, y)$  et identique pour les chocs de fond et déclenchés.

### II – Intensité spatio-temporelle

Elle est donnée par :

$$\lambda(t, x, y | \mathcal{H}_t) = \mu(x, y) + \sum_{j: t_j \in \mathcal{H}_t} \phi(t - t_j, x - x_j, y - y_j, m_j)$$

- $\mu(x, y)$  est l'intensité de fond (supposé homogène en temps)
- $\phi$  est le noyau d'excitation/de déclenchement (productivité × temps × espace)

## Modélisation du noyau d'excitation

Le noyau d'excitation dans le cadre du modèle GP-ETAS s'écrit :

$$\phi(t - t_j, x - x_j, y - y_j, m_j) = \phi_m(m_j) \times \phi_t(t - t_j) \times \phi_s(x - x_j, y - y_j | m_j)$$

Noyau de productivité (nombre moyen de descendants directs) :

$$\phi_m(m_j) = A e^{\alpha(m_j - m_c)}, \quad A > 0, \alpha \geq 0$$

Noyau temporelle (Omori–Utsu modifiée) :

$$\phi_t(t - t_j) = (p - 1) c^{p-1} (t - t_j + c)^{-p}, \quad c > 0, p > 1$$

Noyau spatial (isotrope de type Student) : En posant  $r^2 = (x - x_j)^2 + (y - y_j)^2$  et

$$R(m_j) = d e^{\gamma(m_j - m_c)}, \quad d > 0, \gamma \geq 0,$$

on a :

$$\phi_s(x - x_j, y - y_j | m_j) = \frac{q - 1}{\pi R(m_j)} \left[ 1 + \frac{r^2}{R(m_j)} \right]^{-q}, \quad q > 1.$$

Ce noyau est normalisé sur  $\mathbb{R}^2$  et est isotrope ( $\rightarrow$  pas d'information explicite sur les failles dans la région et n'élimine pas la possibilité d'avoir une réplique exactement au même endroit que son parent (peu probable))

## Modélisation de l'intensité de fond

Le taux de fond est défini par une transformation sigmoïdale d'un processus latent  $f(x, y)$  :

$$\mu(x, y) = \tilde{\mu} \sigma(f(x, y)) = \frac{\tilde{\mu}}{1 + e^{-f(x, y)}}, \quad \tilde{\mu} > 0. \quad (68)$$

Moyenne du processus gaussien :

Contrairement au GP standard, la moyenne du processus n'est pas nulle, elle est définie comme une somme finie de composantes uniformes pondérées par des intensités non connues :

$$m(x, y) = \sum_{j=1}^J \epsilon_j u_j(x, y), \quad (69)$$

où  $u_j(x, y)$  est la fonction indicatrice de la sous-région  $S_j \subset \mathcal{X}$  :

$$u_j(x, y) = \begin{cases} 1 & \text{si } (x, y) \in S_j, \\ 0 & \text{sinon} \end{cases}$$

Les sous-régions  $\{S_j\}_{j=1}^J$  forment une partition de l'espace d'étude  $\mathcal{X} \times \mathcal{Y}$  :

$$S_j \cap S_{j'} = \emptyset \quad \text{si } j \neq j', \quad \text{et} \quad \bigcup_{j=1}^J S_j = \mathcal{X} \times \mathcal{Y}.$$

Les coefficients d'intensité zonaux

$$\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_J)^\top$$

sont modélisés par un prior gaussien centré, avec une covariance dépendant des centres des zones :

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_J, \boldsymbol{\Sigma}_\epsilon). \quad (70)$$

La matrice de covariance  $\boldsymbol{\Sigma}_\epsilon \in \mathbb{R}^{J \times J}$  est définie par un noyau gaussien évalué sur les centres des zones  $\{S_j\}_{j=1}^J$  :

$$(\boldsymbol{\Sigma}_\epsilon)_{jj'} = \delta_0 \exp\left(-\frac{(c_x(S_j) - c_x(S_{j'}))^2 + (c_y(S_j) - c_y(S_{j'}))^2}{2\delta_1^2}\right), \quad 1 \leq j, j' \leq J, \quad (71)$$

où :

- $(c_x(S_j), c_y(S_j))$  désigne le centroïd de la zone  $S_j$
- $\delta_1 > 0$  est un paramètre d'échelle qui module la décroissance spatiale des corrélations et  $\delta_0$  un paramètre de variance

Ainsi, les poids associés à des zones spatialement proches sont plus corrélés, ce qui donne une sorte de régularisation du niveau de fond.

Ainsi, la moyenne du GP devient elle-même aléatoire et reflète une incertitude spatiale sur le niveau de fond dans chacune des zones  $S_j$ . D'où, conditionnellement à  $\boldsymbol{\epsilon}$  et aux hyperparamètres  $\boldsymbol{\nu}$ , on a :

$$f \sim \mathcal{GP}(m(x, y), k((x, y), (x', y') | \boldsymbol{\nu})), \quad (72)$$

$k$  est un noyau gaussien :

$$k((x, y), (x', y') | \boldsymbol{\nu}) = \nu_0^2 \exp\left(-\frac{(x - x')^2}{2\nu_x^2} - \frac{(y - y')^2}{2\nu_y^2}\right), \quad \boldsymbol{\nu} = (\nu_0, \nu_x, \nu_y) \in \mathbb{R}_+^3. \quad (73)$$

Écriture hiérarchique du modèle :

$$\begin{aligned} \boldsymbol{\nu} &\sim p_{\boldsymbol{\nu}} && \text{(loi exponentielle)} \\ \tilde{\mu} &\sim p_{\tilde{\mu}} && \text{(loi gamma)} \\ \boldsymbol{\theta}_\phi &\sim p_{\boldsymbol{\theta}_\phi} && \text{(loi uniforme)} \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(\mathbf{0}_J, \boldsymbol{\Sigma}_\epsilon) \\ f | \boldsymbol{\epsilon}, \boldsymbol{\nu} &\sim \mathcal{GP}(m(x, y), k((x, y), (x', y') | \boldsymbol{\nu})), \\ \mu(x, y) &= \tilde{\mu} \sigma(f(x, y)), \\ \beta &\sim p_\beta \\ D | \beta, \mu, \boldsymbol{\theta}_\phi &\sim \text{Hawkes}(\lambda^*(t, x, y, m)), \end{aligned}$$

avec  $\lambda^*(t, x, y, m)$  défini par (67).

*Remarque :* Si on fixe  $\phi(\cdot)$  à 0, ce modèle se réduit à un SGCP "à moyenne informative", c'est un processus de Poisson inhomogène dont l'intensité spatiale est guidée par un processus gaussien qui incorpore une structure de zonage uniforme sur les régions  $S_j$  (intensités inconnues).

### Expression finale de la fonction d'intensité conditionnelle

Ainsi, on a :

$$\begin{aligned} \lambda^*(t, x, y, m \mid \mathcal{H}_t) &= \frac{\beta \exp[-\beta(m - m_c)]}{1 - \exp[-\beta(m_{\max} - m_c)]} \mathbb{1}_{\{m_c \leq m \leq m_{\max}\}} \times \\ &\left[ \tilde{\mu} \sigma(f(x, y)) + \sum_{j: t_j \in \mathcal{H}_t} A e^{\alpha(m_j - m_c)} (p - 1) c^{p-1} (t - t_j + c)^{-p} \times \frac{q - 1}{\pi R(m_j)} \left(1 + \frac{r^2}{R(m_j)}\right)^{-q} \right] \end{aligned} \quad (74)$$

## III – Vraisemblance et posterior du modèle

Le vecteur de paramètres à estimer pour notre modèle est :

$$\boldsymbol{\theta} = (\beta, \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\phi)$$

où :

$$\boldsymbol{\theta}_\mu = (\tilde{\mu}, f, \boldsymbol{\epsilon}, \boldsymbol{\nu}) \quad \text{et} \quad \boldsymbol{\theta}_\phi = (A, \alpha, c, p, d, q, \gamma)$$

avec :

- $\beta$  : paramètre de magnitude
- $\tilde{\mu}$  : niveau du fond
- $A, \alpha$  : paramètres de productivité et dépendance à la magnitude
- $c, p$  : paramètres du noyau temporel d'Omori–Utsu
- $d, q, \gamma$  : paramètres du noyau spatial (dépendance à la magnitude via  $R(m) = d e^{\gamma(m - m_c)}$ )

La vraisemblance complète est :

$$L_N(\boldsymbol{\theta}) = \left[ \prod_{i=1}^N \lambda^*(t_i, x_i, y_i, m_i \mid \mathcal{H}_{t_i}) \right] \exp \left( - \int_0^T \int_{\mathcal{X}, \mathcal{Y}} \int_{m_c}^{m_{\max}} \lambda^*(t, x, y, m \mid \mathcal{H}_t) dm dx dy dt \right).$$

En substituant (67) et en utilisant le fait que  $\int_{m_c}^{m_{\max}} p_M(m) dm = 1$ , on a :

$$L_N(\boldsymbol{\theta}) = \underbrace{\prod_{i=1}^N p_M(m_i)}_{\text{facteurs magnitudes}} \times \underbrace{\left[ \prod_{i=1}^N \lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i}) \right]}_{\text{facteurs spatio-temporels}} \exp \left( - \int_0^T \int_{\mathcal{X}, \mathcal{Y}} \lambda(t, x, y \mid \mathcal{H}_t) dx dy dt \right).$$

Facteurs magnitudes :

Avec

$$p_M(m) = \frac{\beta e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max}-m_c)}}, \quad \beta = \ln(10) b,$$

on a :

$$\prod_{i=1}^N p_M(m_i) = \left( \frac{\beta}{1 - e^{-\beta(m_{\max}-m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right).$$

Facteurs spatio-temporels - produit des intensités :

On a :

$$\lambda(t_i, x_i, y_i | \mathcal{H}_{t_i}) = \tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j: t_j < t_i} \phi(t_i - t_j, x_i - x_j, y_i - y_j | m_j),$$

d'où :

$$\prod_{i=1}^N \lambda(t_i, x_i, y_i | \mathcal{H}_{t_i}) = \prod_{i=1}^N \left[ \tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j: t_j < t_i} \phi_{ij} \right],$$

avec :

$$\phi_{ij} = A e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \frac{q-1}{\pi R(m_j)} \left( 1 + \frac{r_{ij}^2}{R(m_j)} \right)^{-q},$$

$$R(m_j) = d e^{\gamma(m_j - m_c)}, \quad r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2.$$

Facteurs spatio-temporels - intégrale de l'intensité :

On sépare la contribution du taux de fond et celle du processus d'excitation.

- **Taux de fond :** En intégrant sur le temps et l'espace :

$$\int_0^T \int_{\mathcal{X}, \mathcal{Y}} \mu(x, y) dx dy dt = \tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy. \quad (75)$$

Cette intégrale est intractable (car  $\sigma(f(x, y))$  dépend du GP).

- **Composante d'excitation :** Comme précédemment, en appliquant Fubini :

$$\int_0^T \int_{\mathcal{X}, \mathcal{Y}} \sum_{j: t_j < t} \phi(t - t_j, x - x_j, y - y_j, m_j) dx dy dt = \sum_{j=1}^N A e^{\alpha(m_j - m_c)} T_j S_j,$$

avec

$$T_j = 1 - \left( \frac{c}{T - t_j + c} \right)^{p-1}, \quad S_j = \int_{\mathcal{X}, \mathcal{Y}} \frac{q-1}{\pi R(m_j)} \left( 1 + \frac{r^2}{R(m_j)} \right)^{-q} dx dy.$$

Si  $\mathcal{X}, \mathcal{Y}$  couvre toute la zone d'étude, alors  $S_j \approx 1$ .

## Expression étendue de la vraisemblance

En agrégeant les termes précédents, la vraisemblance du modèle GP-ETAS informatif s'écrit :

$$\begin{aligned}
 L_N(\boldsymbol{\theta}) = & \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right) \times \prod_{i=1}^N \left[ \tilde{\mu} \sigma(f(x_i, y_i)) \right. \\
 & + \sum_{j: t_j < t_i} A e^{\alpha(m_j - m_c)} (p-1) c^{p-1} (t_i - t_j + c)^{-p} \times \frac{q-1}{\pi R(m_j)} \left( 1 + \frac{r_{ij}^2}{R(m_j)} \right)^{-q} \Big] \\
 & \times \exp \left( -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy - \sum_{j=1}^N A e^{\alpha(m_j - m_c)} T_j S_j \right)
 \end{aligned}$$

Les priors sont :

$$\begin{aligned}
 f | w, \nu & \sim \mathcal{GP}(m(\cdot; w), k(\cdot, \cdot | \nu)), & \epsilon & \sim \mathcal{N}(0, I_J), \\
 \boldsymbol{\nu} & \sim p_{\boldsymbol{\nu}}, & \tilde{\mu} & \sim p_{\tilde{\mu}}, & \beta & \sim p_{\beta}, & \boldsymbol{\theta}_{\phi} & \sim p_{\boldsymbol{\theta}_{\phi}}.
 \end{aligned}$$

## Expression finale de la postérior

$$\boxed{
 \begin{aligned}
 p(\boldsymbol{\theta} | D) = & \underbrace{L_N(\boldsymbol{\theta})}_{\text{vraisemblance}} \times \underbrace{p(f | \boldsymbol{\epsilon}, \boldsymbol{\nu})}_{\text{prior GP}} \times \underbrace{p(\boldsymbol{\epsilon})}_{\text{gaussien}} \times \underbrace{p(\boldsymbol{\nu})}_{\text{hyperparams noyau}} \\
 & \times \underbrace{p(\tilde{\mu})}_{\text{taux de fond}} \times \underbrace{p(\boldsymbol{\theta}_{\phi})}_{\text{noyau auto-excitation}} \times \underbrace{p(\beta)}_{\text{G-R}} \times \underbrace{C}_{\text{const. normalisation}}
 \end{aligned}
 }$$

Cette posterior est doublement intractable : une première fois dans la vraisemblance avec l'intégrale (75) et une seconde fois dans la constante de normalisation  $C$  où la même intégrale apparaît.

## IV – Agmentation de données

### Augmentation par processus de branchement

Pour chaque événement  $i$ , on introduit  $z_i \in \{0, 1, \dots, i-1\}$  :  $z_i = 0$  si  $i$  est un choc de fond et  $z_i = j$  si  $i$  est déclenché par  $j$  ( $1 \leq j \leq i-1$ ). On note  $Z = (z_1, \dots, z_N)$  et on pose (pour  $i > j$ )

$$\phi_{ij}(\boldsymbol{\theta}_{\phi}) := \phi(t_i - t_j, x_i - x_j, y_i - y_j, m_j), \quad \phi_i(\boldsymbol{\theta}_{\phi}) := \phi(t - t_i, x - x_i, y - y_i, m_i),$$

et  $T_i := [t_i, T]$ .

On suppose  $p(Z) = \text{cste}$  (a priori uniforme). Alors, pour chaque  $i$  :

$$\lambda(t_i, x_i, y_i | \mathcal{H}_{t_i}) = \left( \tilde{\mu} \sigma(f(x_i, y_i)) \right) + \sum_{j=1}^{i-1} \phi_{ij}(\boldsymbol{\theta}_\phi) \quad (76)$$

$$= \left( \tilde{\mu} \sigma(f(x_i, y_i)) \right)^{\mathbb{1}(z_i=0)} \times \left( \prod_{j=1}^{i-1} \phi_{ij}(\boldsymbol{\theta}_\phi) \right)^{\mathbb{1}(z_i=j)}. \quad (77)$$

Log-vraisemblance augmentée :

On a :

$$\begin{aligned} L_N(\boldsymbol{\theta}, Z) &= \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right) \\ &\times \prod_{i=1}^N \left( \tilde{\mu} \sigma(f(x_i, y_i)) \right)^{\mathbb{1}(z_i=0)} \times \exp \left( - \tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy \right) \\ &\times \prod_{i=1}^N \prod_{j=1}^{i-1} \phi_{ij}(\boldsymbol{\theta}_\phi)^{\mathbb{1}(z_i=j)} \times \exp \left( - \sum_{i=1}^N A e^{\alpha(m_i - m_c)} T_i S_i \right) \\ &\times p(Z). \end{aligned} \quad (78)$$

Loi conditionnelle de  $z_i$  :

En gardant que les termes dépendant de  $z_i$  dans (77),

$$p(z_i = j | D, \boldsymbol{\theta}) \propto \begin{cases} \tilde{\mu} \sigma(f(x_i, y_i)) & j = 0, \\ \phi_{ij}(\boldsymbol{\theta}_\phi) & j \in \{1, \dots, i-1\} \end{cases}$$

En normalisant, on obtient une loi catégorielle sur  $\{0, \dots, i-1\}$  :

$$p(z_i = 0 | \cdot) = \frac{\tilde{\mu} \sigma(f(x_i, y_i))}{\tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j < i} \phi_{ij}(\boldsymbol{\theta}_\phi)}, \quad p(z_i = j | \cdot) = \frac{\phi_{ij}(\boldsymbol{\theta}_\phi)}{\tilde{\mu} \sigma(f(x_i, y_i)) + \sum_{j < i} \phi_{ij}(\boldsymbol{\theta}_\phi)}.$$

## Augmentation par processus de Poisson marqué et Pólya–Gamma

Pour  $z \in \mathbb{R}$ , on a la représentation :

$$\begin{aligned} \sigma(z) &= \int_0^\infty \exp \left( \frac{z}{2} - \frac{z^2}{2} \omega - \ln 2 \right) p_{\text{PG}}(\omega | 1, 0) d\omega, \\ \sigma(-z) &= \int_0^\infty \exp \left( -\frac{z}{2} - \frac{z^2}{2} \omega - \ln 2 \right) p_{\text{PG}}(\omega | 1, 0) d\omega. \end{aligned}$$

On pose, pour tout  $\omega \geq 0$  et  $z \in \mathbb{R}$ ,

$$f_{\text{PG}}(\omega, z) := \frac{z}{2} - \frac{z^2}{2}\omega - \ln 2.$$

On note la contribution de l'intensité de fond à la vraisemblance augmentée par processus de branchement (cf. (78)) :

$$L(D_0 \mid f, \tilde{\mu}) = \prod_{i=1}^N (\tilde{\mu} \sigma(f_i))^{\mathbb{1}_{(z_i=0)}} \times \exp\left(-\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy\right), \quad (79)$$

où  $f_i := f(x_i, y_i)$  et  $D_0 = \{i : z_i = 0\}$  désigne les chocs de fond.

Représentation continue (partie intégrale) :

En utilisant la représentation Pólya–Gamma et l'identité  $1 - \sigma(-z) = \sigma(z)$ , on réécrit :

$$\begin{aligned} -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy &= -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} [1 - \sigma(-f(x, y))] dx dy. \\ &= -\tilde{\mu} T \int_{\mathcal{X} \times \mathcal{Y} \times \mathbb{R}_+} [1 - \exp[f_{\text{PG}}(\omega, -f(x, y))]] p_{\text{PG}}(\omega \mid 1, 0) d\omega dx dy. \end{aligned} \quad (80)$$

On introduit alors l'espace produit :

$$\hat{\mathcal{S}} := \mathcal{X} \times \mathcal{Y} \times \mathbb{R}_+, \quad \Lambda(x, y, \omega) = \tilde{\mu} T p_{\text{PG}}(\omega \mid 1, 0),$$

et le processus de Poisson marqué  $\Pi_{\hat{\mathcal{S}}}$  d'intensité  $\Lambda$ . Le théorème de Campbell donne :

$$\exp\left(-\int_{\hat{\mathcal{S}}} [1 - e^{f_{\text{PG}}(\omega, -f(x, y))}] \Lambda(x, y, \omega) d\omega dx dy\right) = \mathbb{E}_{P_\Lambda} \left[ \prod_{(x_m, y_m, \omega_m) \in \Pi_{\hat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} \right]. \quad (81)$$

Représentation discrète (partie produit) :

Pour la produit dans (79), la représentation PG donne :

$$\begin{aligned} \prod_{i: z_i=0} \tilde{\mu} \sigma(f_i) &= \prod_{i: z_i=0} \tilde{\mu} \int_0^\infty e^{f_{\text{PG}}(\omega_i, f_i)} p_{\text{PG}}(\omega_i \mid 1, 0) d\omega_i \\ &= \mathbb{E}_{\omega_i \sim PG(1, 0)} \left[ \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right] \end{aligned} \quad (82)$$

Vraisemblance augmentée par processus de Poisson marqué et variables Polya–Gamma :

En combinant la partie continue (81) et la partie discrète (82), on a :

$$\begin{aligned}
 L(D_0 \mid f, \tilde{\mu}) &= \prod_{i=1}^N (\tilde{\mu} \sigma(f_i))^{\mathbb{1}_{(z_i=0)}} \times \exp \left( -\tilde{\mu} T \int_{\mathcal{X}, \mathcal{Y}} \sigma(f(x, y)) dx dy \right), \\
 &= \mathbb{E}_{P_\Lambda} \left[ \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m))} \right] + \mathbb{E}_{\omega_i \sim PG(1,0)} \left[ \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right] \\
 &= \int \left( \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m))} \right) P_\Lambda(d\Pi_{\widehat{\mathcal{S}}}) \times \int \left( \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) P_{\omega_{N_0}}(d\omega_{N_0}) \\
 &= \int \left( \prod_{(x_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m))} \right) \left( \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) P_{\text{aug}}(d\Pi_{\widehat{\mathcal{S}}}, d\omega_{N_0})
 \end{aligned} \tag{83}$$

où la mesure des variables augmentées est :

$$P_{\text{aug}} = P_\Lambda \otimes P_{\omega_{D_0}}, \quad \text{avec} \quad P_{\omega_{D_0}} = \prod_{i: z_i=0} p_{\text{PG}}(\omega_i \mid 1, 0) d\omega_i.$$

Or, on peut écrire :

$$L(D_0 \mid f, \tilde{\mu}) = \int L(D_0, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi \mid f, \tilde{\mu}) d\Pi_{\widehat{\mathcal{S}}} d\omega_{D_0}$$

Donc, par identification on a :

$$L(D_0, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi \mid f, \tilde{\mu}) = \left( \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} \right) \left( \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) \tag{84}$$

$$\times p(\Pi_{\widehat{\mathcal{S}}}) p(\omega_{D_0}), \tag{85}$$

où :

$$p(\omega_{D_0}) = \prod_{i: z_i=0} p_{\text{PG}}(\omega_i \mid 1, 0)$$

et où :

$$p(\Pi_{\widehat{\mathcal{S}}}) = \exp \left( - \int_{\widehat{\mathcal{S}}} \Lambda(x, y, \omega) dx dy d\omega \right) \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} \Lambda(x_m, y_m, \omega_m).$$

En utilisant ici

$$\Lambda(x, y, \omega) = \tilde{\mu} T p_{\text{PG}}(\omega \mid 1, 0) \quad \text{et} \quad \int_{\widehat{\mathcal{S}}} \Lambda(x, y, \omega) = \tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|$$

on a finalement :

$$\begin{aligned}
 L(D_0, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi} \mid f, \tilde{\mu}) &= \left( \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} \right) \left( \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} \right) \\
 &\times \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} \tilde{\mu} T p_{\text{PG}}(\omega_m \mid 1, 0) \prod_{i: z_i=0} p_{\text{PG}}(\omega_i \mid 1, 0) \\
 &= \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} p_{\text{PG}}(\omega_i \mid 1, 0) \\
 &\quad \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} \tilde{\mu} T e^{f_{\text{PG}}(\omega_m, -f_m)} p_{\text{PG}}(\omega_m \mid 1, 0)
 \end{aligned} \tag{86}$$

Finalement, les structures linéaire et quadratique en  $f$  des termes  $f_i/2$  et  $f_i^2 \omega_i/2$  dans le résultat ci-dessus montre que la vraisemblance augmentée est **gaussienne en  $f$** . Ainsi, le modèle augmenté est conjugué au prior Gaussien (GP) et permet inférence bayésienne tractable.

En récrivant entièrement la vraisemblance augmentée finale du modèle :

$$\begin{aligned}
 L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) &= \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right) \times \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \\
 &\times \prod_{i: z_i=0} \tilde{\mu} e^{f_{\text{PG}}(\omega_i, f_i)} p_{\text{PG}}(\omega_i \mid 1, 0) \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} \tilde{\mu} T e^{f_{\text{PG}}(\omega_m, -f_m)} p_{\text{PG}}(\omega_m \mid 1, 0) \\
 &\times \prod_{i=1}^N \prod_{j=1}^{i-1} \phi_{ij}(\theta_{\phi})^{\mathbb{1}(z_i=j)} \times \exp \left( - \sum_{i=1}^N A e^{\alpha(m_i - m_c)} T_i S_i \right) \\
 &\times p(Z).
 \end{aligned} \tag{87}$$

**Posterior finale augmentée :**

On en déduit :

$$\boxed{
 \begin{aligned}
 p(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f \mid D) &= L_N(\boldsymbol{\theta}, Z, \Pi_{\widehat{\mathcal{S}}}, \omega_{N_0}) \times p(f \mid \boldsymbol{\epsilon}, \boldsymbol{\nu}) \times p(\boldsymbol{\epsilon}) \\
 &\times p(\boldsymbol{\nu}) \times p(\tilde{\mu}) \times p(\beta) \times p(\boldsymbol{\theta}_{\phi}) \times C
 \end{aligned}}$$

## V – Mise à jour des posteriors conditionnelles

- Comme  $\tilde{\mu} \sim \text{Gamma}(a_{\mu}, b_{\mu})$ , en notant

$$N_0 = |\{i : z_i = 0\}|, \quad N_{\Pi} = |\{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}\}|$$

on a :

$$\begin{aligned}
 p(\tilde{\mu} | \boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) &\propto L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \times p(\tilde{\mu}) \\
 &\propto \tilde{\mu}^{N_0} \tilde{\mu}^{N_{\Pi}} \exp(-\tilde{\mu} T |\mathcal{X}| |\mathcal{Y}|) \times \tilde{\mu}^{a_{\mu}-1} \exp(-b_{\mu} \tilde{\mu}) \quad (88) \\
 &\propto \tilde{\mu}^{N_0 + N_{\Pi} + a_{\mu} - 1} \exp\left[-\tilde{\mu}(T|\mathcal{X}| |\mathcal{Y}| + b_{\mu})\right]
 \end{aligned}$$

Danc la loi conditionnelle est :

$$\tilde{\mu} | D, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f \sim \text{Gamma}\left(a_{\mu} + N_0 + N_{\Pi}, b_{\mu} + T |\mathcal{X}| |\mathcal{Y}|\right)$$

- Pour chaque  $i$  tel que  $z_i = 0$  :

$$p(\omega_i | \boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \propto \exp\left(-\frac{f_i^2}{2} \omega_i\right) p_{\text{PG}}(\omega_i | 1, 0), \quad f_i = f(x_i, y_i).$$

Donc :

$$\omega_i | D, Z, f \sim \text{PG}(1, f(x_i, y_i))$$

et, par indépendance :

$$\omega_{D_0} | D, Z, f \sim \bigotimes_{i:z_i=0} \text{PG}(1, f(x_i, y_i)).$$

- De manière similaire à ce qu'il y a juste au dessus, pour chaque  $m$  tq  $(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}$ , on a :

$$\omega_m | D, \Pi_{\widehat{\mathcal{S}}}, f \sim \text{PG}(1, f(x_m, y_m))$$

Ainsi, par indépendance on en déduit :

$$\omega_{\Pi} | D, \Pi_{\widehat{\mathcal{S}}}, f \sim \bigotimes_{(x_m, y_m) \in \Pi_{\widehat{\mathcal{S}}}} \text{PG}(1, |f(x_m, y_m)|).$$

- On a :

$$p(\Pi_{\widehat{\mathcal{S}}} | \boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \propto \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} \tilde{\mu} T e^{f_{\text{PG}}(\omega_m, -f(x_m, y_m))} p_{\text{PG}}(\omega_m | 1, 0)$$

on reconnaît la densité non normalisée d'un processus de Poisson inhomogène marqué d'intensité  $\Lambda_{\text{post}}(x, y, \omega)$  tq :

$$\Lambda_{\text{post}}(x, y, \omega) = \tilde{\mu} T e^{f_{\text{PG}}(\omega, -f(x, y))} p_{\text{PG}}(\omega | 1, 0)$$

Ainsi, on a :

$$\Pi_{\widehat{\mathcal{S}}} | D, f, \tilde{\mu} \sim \text{PP}(\Lambda_{\text{post}}(x, y, \omega)), \quad \Lambda_{\text{post}}(x, y, \omega) = \tilde{\mu} T e^{f_{\text{PG}}(\omega, -f(x, y))} p_{\text{PG}}(\omega | 1, 0).$$

En intégrant la fonction d'intensité contre les marques  $\omega$ , on obtient l'intensité marginale suivante :

$$\lambda^{\text{post}}(x, y) = \tilde{\mu} T \int_0^\infty e^{f_{\text{PG}}(\omega, -f(x, y))} p_{\text{PG}}(\omega | 1, 0) d\omega \quad (89)$$

$$= \tilde{\mu} T \sigma(-f(x, y)) \quad (90)$$

Ce résultat est intéressant pour l'échantillonnage spatial des événements, on peut passer par l'algorithme classique de thinning et ajouter la marque de manière indépendante.

- On note

$$D_f = D_0 \cup \Pi_{\widehat{\mathcal{S}}},$$

l'union des événements chocs de fond observés et des événements latents du processus de Poisson inhomogène marqué. Le processus gaussien a priori est

$$f \sim \mathcal{GP}(m(x, y), k((x, y), (x', y') | \nu))$$

donc on écrit :

$$p(f | \boldsymbol{\epsilon}, \boldsymbol{\nu}) = \mathcal{N}(f | m_f, K_{f,f}),$$

où

$$m_f = m(x, y | D_f, \boldsymbol{\epsilon}), \quad K_{f,f} = [K_{f,f}]_{ij} = k(x_i, x_j | \boldsymbol{\nu}, D_f).$$

On a donc :

$$\begin{aligned} p(f | \boldsymbol{\theta}, Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi, f, D_f) &\propto L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi, f) \times p(f | \boldsymbol{\epsilon}, \boldsymbol{\nu}) \\ &\propto \prod_{i: z_i=0} e^{f_i/2 - f_i^2 \omega_i/2} \prod_{(x_m, y_m, \omega_m) \in \Pi_{\widehat{\mathcal{S}}}} e^{-f_m/2 - f_m^2 \omega_m/2} \times p(f | \boldsymbol{\epsilon}, \boldsymbol{\nu}) \\ &\propto \exp\left(-\frac{1}{2} f^\top \Omega f + u^\top f\right) \times \exp\left(-\frac{1}{2} (f - m_f)^\top K_{f,f}^{-1} (f - m_f)\right) \\ &\propto \exp\left(-\frac{1}{2} f^\top \Omega f + u^\top f\right) \times \exp\left(-\frac{1}{2} f^\top K_{f,f}^{-1} f + f^\top K_{f,f}^{-1} m_f\right) \\ &\propto \exp\left(-\frac{1}{2} f^\top \Omega f - \frac{1}{2} f^\top K_{f,f}^{-1} f + f^\top K_{f,f}^{-1} m_f + u^\top f\right) \\ &\propto \exp\left(-\frac{1}{2} f^\top (\Omega + K_{f,f}^{-1}) f + f^\top (K_{f,f}^{-1} m_f + u)\right). \end{aligned}$$

où :

$$\Omega = \text{diag}(\omega_{N_0}, \omega_\Pi), \quad u_k = \begin{cases} \frac{1}{2}, & \text{si } f_k \text{ vient d'un } i \text{ tq } z_i = 0 \\ -\frac{1}{2}, & \text{si } f_k \text{ vient d'un point de } \Pi_{\widehat{\mathcal{S}}} \end{cases}$$

En complétant le carré, on reconnaît une loi normale tq :

$$f | Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi, f, D_f, \tilde{\mu}, \boldsymbol{\epsilon}, \boldsymbol{\nu} \sim \mathcal{N}(f | \mu_f, \Sigma_f)$$

où

$$\Sigma_f = \left( \Omega + K_{f,f}^{-1} \right)^{-1} \quad \mu_f = \Sigma_f \left( K_{f,f}^{-1} m_f + u \right).$$

*Remarque :* En fixant  $m_f$  à 0 (cas du prior non informatif), on retombe bien sur la posterior conditionnelle du GP-ETAS standard.

- On rappelle que la moyenne informative du processus gaussien est définie par

$$m(x, y) = \sum_{j=1}^J \epsilon_j u_j(x, y),$$

où  $u_j(x, y)$  est la fonction indicatrice de la sous-région  $S_j$ . Sur l'ensemble des points où le GP est évalué ( $D_f = D_0 \cup \Pi_{\widehat{\mathcal{S}}}$ ), la moyenne peut s'écrire sous forme matricielle :

$$m(D_f | \boldsymbol{\epsilon}) = U \boldsymbol{\epsilon},$$

où  $U = [U_{kj}]$  est la matrice de taille  $|D_f| \times J$  définie par

$$U_{kj} = u_j(x_k, y_k) = \begin{cases} 1 & \text{si } (x_k, y_k) \in S_j, \\ 0 & \text{sinon} \end{cases}$$

Le prior sur les coefficients des zones est :

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon),$$

où  $\boldsymbol{\Sigma}_\epsilon$  encode la structure de dépendance spatiale entre les intensités des différentes zones.

On a donc :

$$\begin{aligned} p(\boldsymbol{\epsilon} | \boldsymbol{\theta}, Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_\Pi, f, D_f) &\propto p(f | \boldsymbol{\epsilon}, \boldsymbol{\nu}) p(\boldsymbol{\epsilon}) \\ &\propto \exp\left(-\frac{1}{2}(f - U\boldsymbol{\epsilon})^\top K^{-1}(f - U\boldsymbol{\epsilon})\right) \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^\top \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\epsilon}\right) \\ &\propto \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^\top (U^\top K^{-1}U + \boldsymbol{\Sigma}_\epsilon^{-1}) \boldsymbol{\epsilon} + \boldsymbol{\epsilon}^\top U^\top K^{-1}f\right). \end{aligned} \tag{91}$$

En complétant le carré, on reconnaît une loi normale multivariée :

$$\boxed{\boldsymbol{\epsilon} | Z, f, D_f, \boldsymbol{\nu} \sim \mathcal{N}(\boldsymbol{\mu}_\epsilon^{\text{post}}, \boldsymbol{\Sigma}_\epsilon^{\text{post}})}$$

où

$$\boldsymbol{\Sigma}_\epsilon^{\text{post}} = (U^\top K^{-1}U + \boldsymbol{\Sigma}_\epsilon^{-1})^{-1}, \quad \boldsymbol{\mu}_\epsilon^{\text{post}} = \boldsymbol{\Sigma}_\epsilon^{\text{post}} U^\top K^{-1}f.$$

- On a :

$$\begin{aligned}
 p(\beta | \boldsymbol{\theta}, Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) &\propto L_N(\boldsymbol{\theta}, Z, \omega_{D_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f) \times p(\beta) \\
 &\propto \left( \frac{\beta}{1 - e^{-\beta(m_{\max} - m_c)}} \right)^N \exp \left( -\beta \sum_{i=1}^N (m_i - m_c) \right) \times p(\beta) \\
 &\propto
 \end{aligned} \tag{92}$$

Pas de forme conjuguée pour  $\beta$ , on va devoir faire appelle à du rejet avec Metropolis-Hastings.

- On rappelle que la moyenne informative du processus gaussien est définie par

$$m(x, y) = \sum_{j=1}^J \epsilon_j u_j(x, y),$$

où  $u_j(x, y)$  est la fonction indicatrice de la sous-région  $S_j$ . Sur l'ensemble des points où le GP est évalué ( $D_f = D_0 \cup \Pi_{\widehat{\mathcal{S}}}$ ), la moyenne peut s'écrire sous forme matricielle :

$$m(D_f | \boldsymbol{\epsilon}) = U \boldsymbol{\epsilon},$$

où  $U = [U_{kj}]$  est la matrice de taille  $|D_f| \times J$  définie par

$$U_{kj} = u_j(x_k, y_k) = \begin{cases} 1 & \text{si } (x_k, y_k) \in S_j, \\ 0 & \text{sinon} \end{cases}$$

Le prior sur les coefficients des zones est :

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon}),$$

où  $\boldsymbol{\Sigma}_{\epsilon}$  encode la structure de dépendance spatiale entre les intensités des différentes zones.

On a donc :

$$\begin{aligned}
 p(\boldsymbol{\epsilon} | \boldsymbol{\theta}, Z, \omega_{N_0}, \Pi_{\widehat{\mathcal{S}}}, \omega_{\Pi}, f, D_f) &\propto p(f | \boldsymbol{\epsilon}, \boldsymbol{\nu}) p(\boldsymbol{\epsilon}) \\
 &\propto \exp \left( -\frac{1}{2} (f - U \boldsymbol{\epsilon})^\top K^{-1} (f - U \boldsymbol{\epsilon}) \right) \exp \left( -\frac{1}{2} \boldsymbol{\epsilon}^\top \boldsymbol{\Sigma}_{\epsilon}^{-1} \boldsymbol{\epsilon} \right) \\
 &\propto \exp \left( -\frac{1}{2} \boldsymbol{\epsilon}^\top (U^\top K^{-1} U + \boldsymbol{\Sigma}_{\epsilon}^{-1}) \boldsymbol{\epsilon} + \boldsymbol{\epsilon}^\top U^\top K^{-1} f \right).
 \end{aligned} \tag{93}$$

En complétant le carré, on reconnaît une loi normale multivariée :

$$\boxed{\boldsymbol{\epsilon} | Z, f, D_f, \boldsymbol{\nu} \sim \mathcal{N}(\boldsymbol{\mu}_{\epsilon}^{\text{post}}, \boldsymbol{\Sigma}_{\epsilon}^{\text{post}})}$$

où

$$\boldsymbol{\Sigma}_{\epsilon}^{\text{post}} = (U^\top K^{-1} U + \boldsymbol{\Sigma}_{\epsilon}^{-1})^{-1}, \quad \boldsymbol{\mu}_{\epsilon}^{\text{post}} = \boldsymbol{\Sigma}_{\epsilon}^{\text{post}} U^\top K^{-1} f.$$

- Pour les hyperparamètres  $\boldsymbol{\nu} = (\nu_0, \nu_1)$  du noyau de covariance  $K_{f,f}$ , la posterior conditionnelle est donnée par :

$$\begin{aligned} p(\boldsymbol{\nu} \mid f, \boldsymbol{\theta} \setminus \boldsymbol{\nu}, Z, \dots) &\propto p(f \mid m_f, K_{f,f}(\boldsymbol{\nu})) \times p(\boldsymbol{\nu}) \\ &\propto \det(K_{f,f}(\boldsymbol{\nu}))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(f - m_f)^\top K_{f,f}(\boldsymbol{\nu})^{-1}(f - m_f)\right) \quad (94) \\ &\quad \times \exp\left(-\lambda_\nu \sum_i \nu_i\right) \end{aligned}$$

Par passage en échelle log, on a :

$$\log(p(\boldsymbol{\nu} \mid f, \boldsymbol{\theta} \setminus \boldsymbol{\nu}, Z, \dots)) \propto -\frac{1}{2} \log|K_{f,f}(\boldsymbol{\nu})| - \frac{1}{2}(f - m_f)^\top K_{f,f}(\boldsymbol{\nu})^{-1}(f - m_f) - \lambda_\nu \sum_i \nu_i \quad (95)$$

On reconnaît aucune loi standard, on procéde donc à une étape de Metropolis-Hastings au sein du Gibbs.

Pour garantir la positivité des hyperparamètres ( $\nu_i > 0$ ), on choisit un lognormal random walk pour la proposition de candidats. À l'itération  $t$ , un candidat  $\boldsymbol{\nu}^*$  est généré en fonction de l'état actuel  $\boldsymbol{\nu}^{(t-1)}$  selon :

$$\begin{aligned} \log(\nu_i^*) &= \log(\nu_i^{(t-1)}) + \xi_i \\ \Leftrightarrow \nu_i^* &= \nu_i^{(t-1)} \cdot e^{\xi_i} \quad \text{avec } \xi_i \sim \mathcal{N}(0, \sigma_{\text{MH}}^2) \end{aligned} \quad (96)$$

La loi de proposition est donc une loi log-normale (donc non symétrique), de densité  $f(\cdot)$  tq :

$$f(x \mid \mu, \sigma) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$$

Si on calcule le ratio de Hastings :

$$\begin{aligned} \prod_{i=0}^1 \frac{q(\nu_i^{(t-1)} \mid \nu_i^*)}{q(\nu_i^* \mid \nu_i^{(t-1)})} &= \prod_{i=0}^1 \frac{\frac{1}{\nu_i^{(t-1)}\sqrt{2\pi}\sigma_{\text{MH}}} \exp\left(-\frac{(\log \nu_i^{(t-1)} - \log \nu_i^*)^2}{2\sigma_{\text{MH}}^2}\right)}{\frac{1}{\nu_i^*\sqrt{2\pi}\sigma_{\text{MH}}} \exp\left(-\frac{(\log \nu_i^* - \log \nu_i^{(t-1)})^2}{2\sigma_{\text{MH}}^2}\right)} = \prod_{i=0}^1 \frac{\nu_i^*}{\nu_i^{(t-1)}} \\ \Leftrightarrow &= \sum_{i=0}^1 \log \nu_i^* - \log \nu_i^{(t-1)} \end{aligned} \quad (97)$$

Le taux d'acceptation  $\alpha(\nu^* \mid \nu^{(t-1)})$  est donné par :

$$\alpha(\nu^* \mid \nu^{(t-1)}) = \min\left(1, \frac{p(\nu^* \mid \dots)}{p(\nu^{(t-1)} \mid \dots)} \times \prod_{i=0}^1 \frac{\nu_i^*}{\nu_i^{(t-1)}}\right) \quad (98)$$

En base logarithmique, le taux d'acceptation  $\log \alpha(\nu^* | \nu^{(t-1)})$  s'écrit :

$$\begin{aligned}\log \alpha(\nu^* | \nu^{(t-1)}) &= \min \left( 0, \left( \log p(\nu^* | \dots) - \log p(\nu^{(t-1)} | \dots) \right) \right. \\ &\quad \left. + \sum_{i=0}^1 \left( \log \nu_i^* - \log \nu_i^{(t-1)} \right) \right)\end{aligned}\tag{99}$$

•  
•

## Pourquoi le modèle GP-ETAS Informatif ?

1. Le modèle non-informatif suppose que nous ne savons rien de la géographie de la région avant d'observer les données, or cette hypothèse est fausse (nous avons accès à un tout récent partitionnement du territoire avec moins de zones que les zonages antérieures).
2. Il est possible d'identifier des zones de la carte avec des taux de sismicité intrinsèquement différents. La connaissance a priori est sur la délimitation de ces zones et non pas sur les taux de sismicité qui sont quaant à elle inconnus à priori, sans passage par les données.
3. On peut considérer que la sismicité change parfois brutalement à la frontière d'une zone (on peut parler de province tectonique et voir les choses sur une échelle bcp plus grande, pas limité à la France et travailler sur l'ensemble du globe).
4. Dans le modèle non-informatif, pour modéliser un saut dans l'activité entre deux zones, le GP a comme marge de manœuvre son noyau de covariance, dont j'imagine qu'il réduit ses longueurs de corrélation pour avoir quelque chose de plus rugueux mais c'est à échelle globale (et dans le cas où c'est un phénomène local, ce n'est pas souhaité d'avoir une diminution marquée de ces paramètres). Dans le modèle informatif, les sauts dans l'activité sismique sont générés principalement par les indicatrices des zones (variation des  $\epsilon_j$ ). Le processus GP ne s'occupera alors que des variations locales à l'intérieur des zones, ce qui permet de garder un noyau "régulier" et plus "réaliste".
5. La transformation  $\mu = \tilde{\mu} \sigma(f)$  peut induire un biais si  $f$  est centré en 0 ( $\sigma(0) = 0.5$ ). Ça incombe au modèle de démarrer avec une hypothèse où l'intensité de fond est à moitié de sa capacité maximale  $\tilde{\mu}$  partout ( $\mu = \tilde{\mu} \sigma(0) = \frac{\tilde{\mu}}{2}$ ). Or, la sismicité est un phénomène "sparse" en France : la majorité de l'espace est vide ou très calme, avec des points chauds. Dans le modèle informatif, le modèle peut apprendre les valeurs  $\epsilon_j$  pour les différentes zones (calmes, moins calmes, actives, très actives). Ça permet au GP de modéliser une distribution asymétrique de l'intensité (beaucoup de valeurs faibles, peu de valeurs fortes) sans tordre  $f$  ou  $\tilde{\mu}$ . Ici aussi, le GP (noyau de covariance) n'agit plus que localement (fluctuations locales autour du niveau des zones).
6. Pour qu'un GP "non-informatif" "découvre" qu'une zone entière est active, il lui faut beaucoup de points de données bien répartis sur la zone. Avec l'approche informative, dès que quelques événements sont observés dans la zone  $S_j$ , le paramètre  $\epsilon_j$  monte, augmentant la probabilité sur toute la zone. Le modèle généralise mieux avec peu de données (?). [Ma compréhension d'un argument soulevé par de Julien]
7. INTERPRÉTABILITÉ : Les valeurs a posteriori des  $\epsilon_j$  quantifient le 'potentiel' sismique de chaque zone relativement aux autres zones (SEMBLE TRÈS APPRÉCIÉ DES SISMOLOGUES). De manière plus local, la partie GP peut montrer les fluctuations locales.

## Identifiabilité du GP-ETAS informatif

L'identifiabilité c'est :

$$\lambda^*(t, x, y, m \mid \mathcal{H}_t, \boldsymbol{\theta}) = \lambda^*(t, x, y, m \mid \mathcal{H}_t, \boldsymbol{\theta}') \Rightarrow \theta = \theta' \quad (100)$$

Soit  $\boldsymbol{\theta}$  et  $\boldsymbol{\theta}'$  deux jeux de paramètres. On suppose  $\lambda^*(t, x, y, m \mid \mathcal{H}_t, \boldsymbol{\theta}) = \lambda^*(t, x, y, m \mid \mathcal{H}_t, \boldsymbol{\theta}')$ . On a :

$$\begin{aligned} \lambda^*(t, x, y, m \mid \mathcal{H}_t, \boldsymbol{\theta}) &= \lambda^*(t, x, y, m \mid \mathcal{H}_t, \boldsymbol{\theta}') \\ p_M(m \mid \beta) \times \lambda(t, x, y \mid \mathcal{H}_t, \boldsymbol{\theta}_{-\beta}) &= p_M(m \mid \beta') \times \lambda(t, x, y \mid \mathcal{H}_t, \boldsymbol{\theta}'_{-\beta'}) \end{aligned}$$

### Partie magnitudes ( $\beta$ )

En intégrant sur le domaine spatio-temporel (dépend pas de  $m$ ), on a :  $\forall m \in [m_c, m_{\max}]$

$$\begin{aligned} p_M(m \mid \beta) = p_M(m \mid \beta') &\Leftrightarrow \frac{\beta e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max}-m_c)}} = \frac{\beta' e^{-\beta'(m-m_c)}}{1 - e^{-\beta'(m_{\max}-m_c)}} \\ &\Leftrightarrow C_\beta \times e^{-\beta(m-m_c)} = C_{\beta'} \times e^{-\beta'(m-m_c)} \\ &\Leftrightarrow (\ln C_\beta + \beta m_c) - \beta m = (\ln C_{\beta'} + \beta' m_c) - \beta' m \\ &\Rightarrow \beta = \beta' \end{aligned}$$

### Partie Spatio-temporelle

En simplifiant par  $p_M$ , on a :

$$\begin{aligned} \lambda(t, x, y \mid \mathcal{H}_t, \boldsymbol{\theta}_{-\beta}) &= \lambda(t, x, y \mid \mathcal{H}_t, \boldsymbol{\theta}'_{-\beta'}) \\ \Leftrightarrow \mu(\cdot \mid \tilde{\mu}, \boldsymbol{\epsilon}, \boldsymbol{\nu}) + \sum_{t_j < t} \phi(\cdot \mid A, \alpha, c, p, d, q, \gamma) &= \mu(\cdot \mid \tilde{\mu}', \boldsymbol{\epsilon}', \boldsymbol{\nu}') + \sum_{t_j < t} \phi(\cdot \mid A', \alpha', c', p', d', q', \gamma') \\ \Leftrightarrow \mu(\cdot \mid \tilde{\mu}, \boldsymbol{\epsilon}, \boldsymbol{\nu}) &= \mu(\cdot \mid \tilde{\mu}', \boldsymbol{\epsilon}', \boldsymbol{\nu}') \quad \text{et} \quad \phi(\cdot \mid A, \alpha, c, p, d, q, \gamma) = \phi(\cdot \mid A', \alpha', c', p', d', q', \gamma') \end{aligned}$$

par identification (on peut considérer exemple avec seulement mainshocks ou exemple avec une seule réplique).

### Paramètres du noyau d'excitation ( $\theta_\phi$ )

En utilisant la décomposition :

$$\phi = \phi_m(m \mid A, \alpha) \phi_t(t \mid c, p) \phi_s(x, y \mid d, q, \gamma)$$

et les propriétés de normalisation, on a :

- a) **Productivité** ( $A, \alpha$ ) : En intégrant  $\phi$  sur  $\mathbb{R}^+ \times \mathbb{R}^2$ , les facteurs  $\phi_t(t)$  et  $\phi_s(x, y)$  disparaissent et on a :

$$\begin{aligned} \phi_m(m \mid A, \alpha) &= \phi_m(m \mid A', \alpha') \Leftrightarrow A e^{\alpha(m-m_c)} = A' e^{\alpha'(m-m_c)} \\ \Rightarrow A &= A' \quad \text{pour } m = m_c \quad \text{et} \quad \alpha = \alpha' \quad \text{pour } m > m_c \end{aligned}$$

- b) **Paramètres temporels** ( $c, p$ ) : En divisant par  $\phi_m$  et en intégrant sur l'espace, on a :

$$\begin{aligned} \phi_t(t \mid c, p) &= \phi_t(t \mid c', p') \Leftrightarrow (p-1) c^{p-1} (t-t_1+c)^{-p} = (p'-1) (c')^{p'-1} (t-t_1+c')^{-p'} \\ \Rightarrow p &= p' \quad \text{pour } t \rightarrow \infty \text{ (identique à } \beta) \quad \text{et} \quad c = c' \quad \text{pour } t \rightarrow t_1 \end{aligned}$$

- c) **Paramètres spatiaux** ( $d, \gamma, q$ ) : En posant

$$r = (x - x_1)^2 + (y - y_1)^2,$$

on a :

$$\begin{aligned} \phi_s(x, y \mid d, q, \gamma) &= \phi_s(x, y \mid d', q', \gamma') \Leftrightarrow \frac{q-1}{\pi R(m)} \left(1 + \frac{r^2}{R(m)}\right)^{-q} = \frac{q'-1}{\pi R'(m)} \left(1 + \frac{r^2}{R'(m)}\right)^{-q'} \\ \Rightarrow q &= q' \quad \text{pour } r \rightarrow \infty \quad \text{et} \quad R(m) = R'(m) \quad \text{pour } r \rightarrow 0. \end{aligned}$$

Puis, en décomposant  $R(m)$  :

$$d e^{\gamma(m-m_c)} = d' e^{\gamma'(m-m_c)} \Rightarrow d = d' \quad \text{pour } m = m_c \quad \text{et} \quad \gamma = \gamma' \quad \text{pour } m > m_c$$

### Paramètres de l'intensité de fond ( $\tilde{\mu}, f, \epsilon$ )

On a :

$$\mu(x, y \mid \tilde{\mu}, \epsilon, \nu, f) = \mu(x, y \mid \tilde{\mu}', \epsilon', \nu', f') \Leftrightarrow \frac{\tilde{\mu}}{1 + e^{-f(x,y)}} = \frac{\tilde{\mu}'}{1 + e^{-f'(x,y)}}$$

WORK IN PROGRESS...



---

## Bibliographie

---

Ryan Prescott Adams, Iain Murray, and David J. C. MacKay. Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 9–16, Montreal Quebec Canada, June 2009. ACM. ISBN 978-1-60558-516-1. doi : 10.1145/1553374.1553376. URL <https://dl.acm.org/doi/10.1145/1553374.1553376>.

M. Antónia Amaral Turkman, Carlos Daniel Paulino, and Peter Müller. *Computational Bayesian Statistics : An Introduction*. Cambridge University Press, 1 edition, February 2019. ISBN 978-1-108-64618-5 978-1-108-48103-8 978-1-108-70374-1. doi : 10.1017/9781108646185. URL <https://www.cambridge.org/core/product/identifier/9781108646185/type/book>.

Arash A. Amini, Marina Paez, Lizhen Lin, and Zahra S. Razaee. Exact slice sampler for Hierarchical Dirichlet Processes, March 2019. URL <http://arxiv.org/abs/1903.08829>. arXiv :1903.08829 [stat].

Taoufik Bdiri, Nizar Bouguila, and Djemel Ziou. Variational Bayesian inference for infinite generalized inverted Dirichlet mixtures with feature selection and its application to clustering. *Applied Intelligence*, 44(3) :507–525, April 2016. ISSN 0924-669X, 1573-7497. doi : 10.1007/s10489-015-0714-6. URL <http://link.springer.com/10.1007/s10489-015-0714-6>.

Mario Beraha, Alessandra Guglielmi, and Fernando A. Quintana. The Semi-Hierarchical Dirichlet Process and Its Application to Clustering Homogeneous Distributions. *Bayesian Analysis*, 16(4), December 2021. ISSN 1936-0975. doi : 10.1214/21-BA1278. URL <https://projecteuclid.org/journals/bayesian-analysis/volume-16/issue-4/The-Semi-Hierarchical-Dirichlet-Process-and-Its-Application-to-Clustering/10.1214/21-BA1278.full>.

Alba Bernabeu, Jianchang Zhuang, and Jorge Mateu. Spatio-Temporal Hawkes Point Processes : A Review. *Journal of Agricultural, Biological and Environmental Statistics*, 30(1) :89–119, March 2025. ISSN 1085-7117, 1537-2693. doi : 10.1007/s13253-024-00653-7. URL <https://link.springer.com/10.1007/s13253-024-00653-7>.

David M. Blei and Michael I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1), March 2006. ISSN 1936-0975. doi : 10.1214/06-BA104.

URL <https://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-1/Variational-inference-for-Dirichlet-process-mixtures/10.1214/06-BA104.full>.

Christopher A. Bush, Juhee Lee, and Steven N. MacEachern. Minimally Informative Prior Distributions for Non-Parametric Bayesian Analysis. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 72(2) :253–268, March 2010. ISSN 1369-7412, 1467-9868. doi : 10.1111/j.1467-9868.2009.00735.x. URL <https://academic.oup.com/rssb/article/72/2/253/7076502>.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte Carlo guided Diffusion for Bayesian linear inverse problems, October 2023. URL <http://arxiv.org/abs/2308.07983>. arXiv :2308.07983 [stat].

Djalil Chafai and Florent Malrieu. *Recueil de Modèles Aléatoires*, volume 78 of *Mathématiques et Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. ISBN 978-3-662-49767-8 978-3-662-49768-5. doi : 10.1007/978-3-662-49768-5. URL <http://link.springer.com/10.1007/978-3-662-49768-5>.

Florencia Chimard and Jean Vaillant. Processus Stick-Breaking et extensions pour le traitement bayésien de processus ponctuels.

Andreas C. Damianou and Neil D. Lawrence. Deep Gaussian Processes, March 2013. URL <http://arxiv.org/abs/1211.0358>. arXiv :1211.0358 [stat].

Guillaume Daniel and Pierre Arroucau. Data-Driven Seismicity Models Based on Voronoi Diagrams. *Seismological Research Letters*, 96(5) :3050–3062, September 2025. ISSN 0895-0695, 1938-2057. doi : 10.1785/0220240428. URL <https://pubs.geoscienceworld.org/srl/article/96/5/3050/654486/Data-Driven-Seismicity-Models-Based-on-Voronoi>.

Snigdha Das, Yabo Niu, Yang Ni, Bani K. Mallick, and Debdeep Pati. Blocked Gibbs sampler for hierarchical Dirichlet processes. *Journal of Computational and Graphical Statistics*, 34(2) : 519–529, April 2025. ISSN 1061-8600, 1537-2715. doi : 10.1080/10618600.2024.2388543. URL <http://arxiv.org/abs/2304.09945>. arXiv :2304.09945 [stat].

Michael Deistler, Jan Boelts, Peter Steinbach, Guy Moss, Thomas Moreau, Manuel Gloeckler, Pedro L. C. Rodrigues, Julia Linhart, Janne K. Lappalainen, Benjamin Kurt Miller, Pedro J. Gonçalves, Jan-Matthis Lueckmann, Cornelius Schröder, and Jakob H. Macke. Simulation-Based Inference : A Practical Guide, August 2025. URL <http://arxiv.org/abs/2508.12939>. arXiv :2508.12939 [stat].

Alex Diana, Eleni Matechou, Jim Griffin, and Alison Johnston. A hierarchical dependent Dirichlet process prior for modelling bird migration patterns in the UK. *The Annals of Applied Statistics*, 14(1), March 2020. ISSN 1932-6157. doi : 10.1214/19-AOAS1315. URL <https://projecteuclid.org/journals/annals-of-applied-statistics/volume-14/issue-1/A-hierarchical-dependent-Dirichlet-process-prior-for-modelling-bird-migration/10.1214/19-AOAS1315.full>.

Christian Donner and Manfred Opper. Efficient Bayesian Inference of Sigmoidal Gaussian Cox Processes.

Stéphane Drouet, Gabriele Ameri, Kristell Le Dortz, Ramon Secanell, and Gloria Senfaute. A probabilistic seismic hazard map for the metropolitan France. *Bulletin of Earthquake Engineering*, 18(5) :1865–1898, March 2020. ISSN 1570-761X, 1573-1456. doi : 10.1007/s10518-020-00790-7. URL <http://link.springer.com/10.1007/s10518-020-00790-7>.

Nan Du, Mehrdad Farajtabar, Amr Ahmed, Alexander J. Smola, and Le Song. Dirichlet-Hawkes Processes with Applications to Clustering Continuous-Time Document Streams. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 219–228, Sydney NSW Australia, August 2015. ACM. ISBN 978-1-4503-3664-2. doi : 10.1145/2783258.2783411. URL <https://dl.acm.org/doi/10.1145/2783258.2783411>.

Vincent Dutordoir, Alan Saul, Zoubin Ghahramani, and Fergus Simpson. Neural Diffusion Processes, June 2023. URL <http://arxiv.org/abs/2206.03992>. arXiv :2206.03992 [stat].

Mame Diarra Fall and Éric Barat. A simple and efficient method for sampling mixture models based on Dirichlet and Pitman-Yor processes. *Computational Statistics*, 40(8) :4675–4716, November 2025. ISSN 0943-4062, 1613-9658. doi : 10.1007/s00180-025-01637-y. URL <https://link.springer.com/10.1007/s00180-025-01637-y>.

Zoubin Ghahramani. Non-parametric Bayesian Methods.

Zoubin Ghahramani. Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 371(1984) :20110553, February 2013. ISSN 1364-503X, 1471-2962. doi : 10.1098/rsta.2011.0553. URL <https://royalsocietypublishing.org/doi/10.1098/rsta.2011.0553>.

Manuel Gloeckler, Michael Deistler, Christian Weilbach, Frank Wood, and Jakob H. Macke. All-in-one simulation-based inference, July 2024. URL <http://arxiv.org/abs/2404.09636>. arXiv :2404.09636 [cs].

Clara Grazian. Spatio-Temporal Stick-Breaking Process. *Bayesian Analysis*, 20(3), September 2025. ISSN 1936-0975. doi : 10.1214/24-BA1419. URL <https://projecteuclid.org/journals/bayesian-analysis/volume-20/issue-3/Spatio-Temporal-Stick-Breaking-Process/10.1214/24-BA1419.full>.

Creighton Heaukulani and Daniel M. Roy. Gibbs-type Indian Buffet Processes. *Bayesian Analysis*, 15(3), September 2020. ISSN 1936-0975. doi : 10.1214/19-BA1166. URL <https://projecteuclid.org/journals/bayesian-analysis/volume-15/issue-3/Gibbs-type-Indian-Buffet-Processes/10.1214/19-BA1166.full>.

M. Holschneider, K. Ferrat, G. Zöller, and Ch Molkenthin. Richter b-value maps from local moments of seismicity, October 2020. URL <http://arxiv.org/abs/2010.12298>. arXiv :2010.12298 [physics].

Weipeng Huang, Nishma Laitonjam, Guangyuan Piao, and Neil Hurley. Inferring Hierarchical Mixture Structures : A Bayesian Nonparametric Approach, May 2021. URL <http://arxiv.org/abs/1905.05022>. arXiv :1905.05022 [stat].

Abdollah Jalilian. **ETAS** : An R Package for Fitting the Space-Time ETAS Model to Earthquake Data. *Journal of Statistical Software*, 88(Code Snippet 1), 2019. ISSN 1548-7660. doi : 10.18637/jss.v088.c01. URL <http://www.jstatsoft.org/v88/c01/>.

Alejandro Jara. Theory and computations for the Dirichlet process and related models : An overview. *International Journal of Approximate Reasoning*, 81 :128–146, February 2017. ISSN 0888613X. doi : 10.1016/j.ijar.2016.11.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S0888613X16302328>.

Maria Kalli, Jim E. Griffin, and Stephen G. Walker. Slice sampling mixture models. *Statistics and Computing*, 21(1) :93–105, January 2011. ISSN 0960-3174, 1573-1375. doi : 10.1007/s11222-009-9150-y. URL <http://link.springer.com/10.1007/s11222-009-9150-y>.

Erhan Karakoca, Güneş Karabulut Kurt, and Ali Görçin. Hierarchical Dirichlet Process Based Gamma Mixture Modelling for Terahertz Band Wireless Communication Channels. *IEEE Access*, 10 :84635–84647, 2022. ISSN 2169-3536. doi : 10.1109/ACCESS.2022.3197603. URL <http://arxiv.org/abs/2205.03812>. arXiv :2205.03812 [eess].

Merlin Keller, Sanaa Zannane, Clara Duverger, and Jessie Mayor. Bayesian estimation and selection of seismic source models.

Aleksandar A. Kolev and Gordon J. Ross. Semiparametric Bayesian Forecasting of Spatial Earthquake Occurrences, February 2020. URL <http://arxiv.org/abs/2002.01706>. arXiv :2002.01706 [stat].

Athanasis Kottas, Jason A. Duan, and Alan E. Gelfand. Modeling Disease Incidence Data with Spatial and Spatio Temporal Dirichlet Process Mixtures. *Biometrical Journal*, 50(1) :29–42, February 2008. ISSN 0323-3847, 1521-4036. doi : 10.1002/bimj.200610375. URL <https://onlinelibrary.wiley.com/doi/10.1002/bimj.200610375>.

Jaehoon Lee, Yasaman Bahri, Roman Novak, Samuel S. Schoenholz, Jeffrey Pennington, and Jascha Sohl-Dickstein. Deep Neural Networks as Gaussian Processes, March 2018. URL <http://arxiv.org/abs/1711.00165>. arXiv :1711.00165 [stat].

Sujeong Lee, Won Chang, Jorge Mateu, Heejin Lee, and Jaewoo Park. A Spatio-Temporal Dirichlet Process Mixture Model on Linear Networks for Crime Data, January 2025. URL <http://arxiv.org/abs/2501.08673>. arXiv :2501.08673 [stat].

Naiqi Li, Wenjie Li, Yong Jiang, and Shu-Tao Xia. Deep Dirichlet Process Mixture Models. a.

Weihan Li, Yu Qi, and Gang Pan. Online Neural Sequence Detection with Hierarchical Dirichlet Point Process. b.

Julia Linhart, Gabriel Victorino Cardoso, Alexandre Gramfort, Sylvain Le Corff, and Pedro Luiz Coelho Rodrigues. Diffusion posterior sampling for simulation-based inference in tall data settings.

Maria Lomeli. General Bayesian inference schemes in infinite mixture models, February 2017. URL <http://arxiv.org/abs/1702.08781>. arXiv :1702.08781 [stat].

Jan-Matthis Lueckmann, Jan Boelts, David S Greenberg, Pedro J Gonçalves, and Jakob H Macke. Benchmarking Simulation-Based Inference.

Steven N. MacEachern. Nonparametric Bayesian methods : a gentle introduction and overview. *Communications for Statistical Applications and Methods*, 23(6) :445–466, November 2016. ISSN 2383-4757. doi : 10.5351/CSAM.2016.23.6.445. URL <http://www.csam.or.kr/journal/view.html?doi=10.5351/CSAM.2016.23.6.445>.

K. Manchuel, P. Traversa, D. Baumont, M. Cara, E. Nayman, and C. Durouchoux. The French seismic CATalogue (FCAT-17). *Bulletin of Earthquake Engineering*, 16(6) :2227–2251, June 2018. ISSN 1570-761X, 1573-1456. doi : 10.1007/s10518-017-0236-1. URL <http://link.springer.com/10.1007/s10518-017-0236-1>.

Jean-Michel Marin, Pierre Pudlo, Christian P. Robert, and Robin J. Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6) :1167–1180, November 2012. ISSN 0960-3174, 1573-1375. doi : 10.1007/s11222-011-9288-2. URL <http://link.springer.com/10.1007/s11222-011-9288-2>.

Jon D. McAuliffe, David M. Blei, and Michael I. Jordan. Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, 16(1) :5–14, January 2006. ISSN 0960-3174, 1573-1375. doi : 10.1007/s11222-006-5196-2. URL <http://link.springer.com/10.1007/s11222-006-5196-2>.

Christian Molkenthin, Christian Donner, Sebastian Reich, Gert Zöller, Sebastian Hainzl, Matthias Holschneider, and Manfred Opper. GP-ETAS : semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Statistics and Computing*, 32(2) :29, April 2022a. ISSN 0960-3174, 1573-1375. doi : 10.1007/s11222-022-10085-3. URL <https://link.springer.com/10.1007/s11222-022-10085-3>.

Christian Molkenthin, Christian Donner, Sebastian Reich, Gert Zöller, Sebastian Hainzl, Matthias Holschneider, and Manfred Opper. GP-ETAS : semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Statistics and Computing*, 32(2) :29, April 2022b. ISSN 0960-3174, 1573-1375. doi : 10.1007/s11222-022-10085-3. URL <https://link.springer.com/10.1007/s11222-022-10085-3>.

Christian Molkenthin, Gert Zöller, Sebastian Hainzl, and Matthias Holschneider. Bayesian Earthquake Forecasting Using Gaussian Process Modeling : GP-ETAS Applications. *Seismological Research Letters*, 95(6) :3532–3544, November 2024. ISSN 0895-0695, 1938-2057. doi : 10.1785/0220240170. URL <https://pubs.geoscienceworld.org/srl/article/95/6/3532/648870/Bayesian-Earthquake-Forecasting-Using-Gaussian>.

Camille M. Moore, Nichole E. Carlson, Samantha MaWhinney, and Sarah Kreidler. A Dirichlet Process Mixture Model for Non-Ignorable Dropout. *Bayesian Analysis*, 15(4), December 2020. ISSN 1936-0975. doi : 10.1214/19-BA1181. URL <https://projecteuclid.org/journals/bayesian-analysis/volume-15/issue-4/A-Dirichlet-Process-Mixture-Model-for-Non-Ignorable-Dropout/10.1214/19-BA1181.full>.

Radford M. Neal. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9(2) :249–265, June 2000. ISSN 1061-8600, 1537-2715. doi : 10.1080/10618600.2000.10474879. URL <http://www.tandfonline.com/doi/abs/10.1080/10618600.2000.10474879>.

Sally Paganin, Amy H. Herring, Andrew F. Olshan, and David B. Dunson. Centered Partition Processes : Informative Priors for Clustering (with Discussion). *Bayesian Analysis*, 16(1), March 2021. ISSN 1936-0975. doi : 10.1214/20-BA1197. URL <https://projecteuclid.org/journals/bayesian-analysis/volume-16/issue-1/Centered-Partition-Processes-Informative-Priors-for-Clustering-with-Discussion/10.1214/20-BA1197.full>.

John Paisley, David M Blei, and Michael I Jordan. Variational Bayesian Inference with Stochastic Search.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing Flows for Probabilistic Modeling and Inference.

Omiros Papaspiliopoulos and Gareth Roberts. Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical model, October 2007. URL <http://arxiv.org/abs/0710.4228>. arXiv :0710.4228 [stat].

Giuseppe Petrillo and Jiancang Zhuang. Bayesian earthquake forecasting approach based on the epidemic type aftershock sequence model. *Earth, Planets and Space*, 76(1) :78, May 2024. ISSN 1880-5981. doi : 10.1186/s40623-024-02021-8. URL <https://earth-planets-space.springeropen.com/articles/10.1186/s40623-024-02021-8>.

Emilia Pompe, Chris Holmes, and Krzysztof Łatuszyński. A framework for adaptive MCMC targeting multimodal distributions. *The Annals of Statistics*, 48(5), October 2020. ISSN 0090-5364. doi : 10.1214/19-AOS1916. URL <https://projecteuclid.org/journals/annals-of-statistics/volume-48/issue-5/A-framework-for-adaptive-MCMC-targeting-multimodal-distributions/10.1214/19-AOS1916.full>.

Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher. Le Processus Powered Dirichlet-Hawkes comme A Priori Flexible pour Clustering Temporel de Textes, January 2022. URL <http://arxiv.org/abs/2201.12568>. arXiv :2201.12568 [cs].

Fernand A. Quintana, Peter Mueller, Alejandro Jara, and Steven N. MacEachern. The Dependent Dirichlet Process and Related Models, July 2020. URL <http://arxiv.org/abs/2007.06129>. arXiv :2007.06129 [stat].

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 3. print edition, 2008. ISBN 978-0-262-18253-9.

Lu Ren, David B. Dunson, and Lawrence Carin. The dynamic hierarchical Dirichlet process. In *Proceedings of the 25th international conference on Machine learning - ICML '08*, pages 824–831, Helsinki, Finland, 2008. ACM Press. ISBN 978-1-60558-205-4. doi : 10.1145/1390156.1390260. URL <http://portal.acm.org/citation.cfm?doid=1390156.1390260>.

Stefano Rinaldi and Walter Del Pozzo. (H)DPGMM : A Hierarchy of Dirichlet Process Gaussian Mixture Models for the inference of the black hole mass function. *Monthly Notices of the Royal Astronomical Society*, 509(4) :5454–5466, December 2021. ISSN 0035-8711, 1365-2966.

doi : 10.1093/mnras/stab3224. URL <http://arxiv.org/abs/2109.05960>. arXiv :2109.05960 [astro-ph].

Gordon J Ross and Dean Markwick. dirichletprocess : An R Package for Fitting Complex Bayesian Nonparametric Models.

Yushu Shi, Michael Martens, Anjishnu Banerjee, and Purushottam Laud. Low Information Omnibus (LIO) Priors for Dirichlet Process Mixture Models. *Bayesian Analysis*, 14(3), September 2019. ISSN 1936-0975. doi : 10.1214/18-BA1119. URL <https://projecteuclid.org/journals/bayesian-analysis/volume-14/issue-3/Low-Information-Omnibus-LIO-Priors-for-Dirichlet-Process-Mixture-Models/10.1214/18-BA1119.full>.

I. Spassiani, S. Gentili, R. Console, M. Murru, M. Taroni, and G. Falcone. Reconciling the irreconcilable : window-based versus stochastic declustering algorithms, August 2024. URL <http://arxiv.org/abs/2408.16491>. arXiv :2408.16491 [physics].

Samuel Stockman, Daniel J. Lawson, and Maximilian J. Werner. SB-ETAS : using simulation based inference for scalable, likelihood-free inference for the ETAS model of earthquake occurrences, May 2024a. URL <http://arxiv.org/abs/2404.16590>. arXiv :2404.16590 [stat].

Samuel Stockman, Daniel J. Lawson, and Maximilian J. Werner. SB-ETAS : using simulation based inference for scalable, likelihood-free inference for the ETAS model of earthquake occurrences, May 2024b. URL <http://arxiv.org/abs/2404.16590>. arXiv :2404.16590 [stat].

Samuel Stockman, Daniel Lawson, and Maximilian Werner. EarthquakeNPP : A Benchmark for Earthquake Forecasting with Neural Point Processes, September 2025. URL <http://arxiv.org/abs/2410.08226>. arXiv :2410.08226 [physics].

Erik B Sudderth. Graphical Models for Visual Object Recognition and Tracking.

Shiliang Sun, John Paisley, and Qiuyang Liu. Location Dependent Dirichlet Processes. In Yi Sun, Huchuan Lu, Lihe Zhang, Jian Yang, and Hua Huang, editors, *Intelligence Science and Big Data Engineering*, volume 10559, pages 64–76. Springer International Publishing, Cham, 2017. ISBN 978-3-319-67776-7 978-3-319-67777-4. doi : 10.1007/978-3-319-67777-4\_6. URL [https://link.springer.com/10.1007/978-3-319-67777-4\\_6](https://link.springer.com/10.1007/978-3-319-67777-4_6). Series Title : Lecture Notes in Computer Science.

Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, 101(476) :1566–1581, December 2006. ISSN 0162-1459, 1537-274X. doi : 10.1198/016214506000000302. URL <https://www.tandfonline.com/doi/full/10.1198/016214506000000302>.

Xinyuan Tian, Maria Ciarleglio, Jiachen Cai, Erich Greene, Denise Esserman, Fan Li, and Yize Zhao. Bayesian semi-parametric inference for clustered recurrent events with zero-inflation and a terminal event/4163305, December 2022. URL <http://arxiv.org/abs/2202.06636>. arXiv :2202.06636 [stat].

Carlo Vicentini and Ian Hyla Jermyn. Prior selection for the precision parameter of Dirichlet Process Mixtures, June 2025. URL <http://arxiv.org/abs/2502.00864>. arXiv :2502.00864 [stat].

Stephen G. Walker. Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics - Simulation and Computation*, 36(1) :45–54, January 2007. ISSN 0361-0918, 1532-4141. doi : 10.1080/03610910601096262. URL <http://www.tandfonline.com/doi/abs/10.1080/03610910601096262>.

Haoxuan Wang. Nonparametric Bayesian Density Estimation with Gaussian Processes. 2024.

Yanyi Wang. Density Estimation Using Nonparametric Bayesian Methods.

Florian Wenzel, Théo Galy-Fajou, Christian Donner, Marius Kloft, and Manfred Opper. Efficient Gaussian Process Classification Using Pólya-Gamma Data Augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01) :5417–5424, July 2019. ISSN 2374-3468, 2159-5399. doi : 10.1609/aaai.v33i01.33015417. URL <https://ojs.aaai.org/index.php/AAAI/article/view/4481>.

Eric P Xing, Hai Phan, Magesh Kanna, and Stephen Tsou. Lecture 23 : Bayesian Nonparametrics : Dirichlet Proccesses.

I Zentner, I Main, J Douglas, and E Viallet. Name Date Name Date Name Date.

Irmela Zentner, Gabriele Ameri, and Emmanuel Viallet. Bayesian Estimation of the Maximum Magnitude mmax Based on the Extreme Value Distribution for Probabilistic Seismic Hazard Analyses. *Pure and Applied Geophysics*, 177(12) :5643–5660, December 2020. ISSN 0033-4553, 1420-9136. doi : 10.1007/s00024-020-02612-y. URL <http://link.springer.com/10.1007/s00024-020-02612-y>.

Haoyuan Zhang, Shuya Ke, Wenqi Liu, and Yongwen Zhang. A combining earthquake forecasting model between deep learning and epidemic-type aftershock sequence (ETAS) model. *Geophysical Journal International*, 239(3) :1545–1556, October 2024a. ISSN 0956-540X, 1365-246X. doi : 10.1093/gji/ggae349. URL <https://academic.oup.com/gji/article/239/3/1545/7775538>.

Ying Zhang, Chengxiang Zhan, Qinghua Huang, and Didier Sornette. Forecasting future earthquakes with deep neural networks : application to California. *Geophysical Journal International*, 240(1) :81–95, November 2024b. ISSN 0956-540X, 1365-246X. doi : 10.1093/gji/ggae373. URL <https://academic.oup.com/gji/article/240/1/81/7826790>.

Oleg Zlydenko, Gal Elidan, Avinatan Hassidim, Doron Kukliansky, Yossi Matias, Brendan Meade, Alexandra Molchanov, Sella Nevo, and Yohai Bar-Sinai. A neural encoder for earthquake rate forecasting. *Scientific Reports*, 13(1) :12350, July 2023. ISSN 2045-2322. doi : 10.1038/s41598-023-38033-9. URL <https://www.nature.com/articles/s41598-023-38033-9>.

Nevena Šipčić, Mohsen Kohrangi, Athanasios N. Papadopoulos, Warner Marzocchi, and Paolo Bazzurro. The Effect of Seismic Sequences in Probabilistic Seismic Hazard Analysis. *Bulletin of the Seismological Society of America*, 112(3) :1694–1709, June 2022. ISSN 0037-1106, 1943-3573. doi : 10.1785/0120210208. URL <https://pubs.geoscienceworld.org/bssa/article/112/3/1694/612260/The-Effect-of-Seismic-Sequences-in-Probabilistic>.