

Bayesian modeling of earthquakes

Ibrahim SEYDI

Restricted case

Spatial data without aftershocks or foreshocks

Let observed seismic events be :

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

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

We set that :

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

where :

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

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 θ_k à θ
 - Incrément $k \leftarrow k + 1$

Output : Approximation d'un DP avec :

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

em!

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 Ω :

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

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 Ω 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 Ω 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 Ω , 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 Ω .

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 μ et la matrice de covariance à cette loi normale multivariée Σ . Cette loi est caractérisée par quatre hyperparamètres :

- μ_0 (vect de dim d) : la moyenne prior sur μ
- λ_0 (scalaire positif) : un facteur d'échelle sur la précision de la moyenne

- Ψ_0 (matrice d x d, sym et def pos) : un paramètre d'échelle pour la matrice de covariance
- ν_0 (degré de liberté > d - 1) : un paramètre qui contrôle la concentration de la loi Inverse Wishart sur Σ

La NIW est donnée ainsi :

$$\Sigma \sim \mathcal{IW}(\Psi_0, \nu_0), \quad \mu | \Sigma \sim \mathcal{N}(\mu_0, \frac{\Sigma}{\lambda_0})$$

où \mathcal{IW} est la loi inverse Wishart et μ suit une normale multivariée avec covariance Σ/λ_0 .

La densité de l'Inverse Wishart est :

$$f(\Sigma | \Psi_0, \nu_0) = \frac{|\Psi_0|^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0 d}{2}} \Gamma_d\left(\frac{\nu_0}{2}\right)} |\Sigma|^{-\frac{\nu_0+d+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Psi_0 \Sigma^{-1})\right)$$

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

La densité de la loi normale conditionnelle est :

$$f\left(\mu | \mu_0, \frac{\Sigma}{\lambda_0}\right) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma/\lambda_0|^{\frac{1}{2}}} \exp\left(-\frac{\lambda_0}{2} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)\right)$$

Donc la densité NIW est :

$$f(\mu, \Sigma) = f(\Sigma | \Psi_0, \nu_0) \cdot f\left(\mu | \mu_0, \frac{\Sigma}{\lambda_0}\right)$$

$$= \frac{|\Psi_0|^{\frac{\nu_0}{2}} \lambda_0^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}} 2^{\frac{\nu_0 d}{2}} \Gamma_d\left(\frac{\nu_0}{2}\right)} |\Sigma|^{-\frac{\nu_0+d+2}{2}}$$

$$\times \exp\left(-\frac{1}{2} \text{tr}(\Psi_0 \Sigma^{-1}) - \frac{\lambda_0}{2} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)\right)$$

$\Gamma_d(\cdot)$ 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
μ_0	<ul style="list-style-type: none"> Moyenne de la loi de μ Plus λ_0 est grand, plus μ est concentré autour de μ_0 Plus λ_0 est faible, μ s'écarte de μ_0
λ_0	<ul style="list-style-type: none"> C'est un facteur d'échelle sur la variance de μ Contrôle l'incertitude a priori sur la moyenne μ $\lambda_0 \rightarrow 0$ signifie une incertitude infinie sur μ
Ψ_0	<ul style="list-style-type: none"> Contrôle la taille et l'orientation moyennes des matrices de covariance Σ Plus les valeurs propres de Ψ_0 sont grandes, plus les réalisations de Σ sont grandes (variances plus larges) et/ou des corrélations plus marquées
ν_0	<ul style="list-style-type: none"> Contrôle la concentration de la loi de Σ Doit être strictement supérieur à $d - 1$ pour que la moyenne existe Plus ν_0 est grand, plus Σ est concentré autour de sa moyenne $\Psi_0/(\nu_0 - d - 1)$ Si ν_0 est faible (proche de d), la dispersion des matrices Σ est grande (forte incertitude)

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 Ω 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.

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 α et seuil de troncature τ :

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

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

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 σ_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 μ_j avec covariance Σ_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 Σ_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 2×2 .

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 Σ_j pour que cette gaussienne couvre 95% de la masse dans la zone. À cet objectif, on choisit un écart-type σ_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 centroïde de la zone. On veut que cette normale couvre environ 95%. On choisit alors un écart-type σ_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'algo 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 : initialisation des moyennes ; pi_k : proportions initiales (toutes égales))
2. Initialisation des matrices de covariance (Si $sigma_init$ est NULL \rightarrow 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 γ_{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 π_k, μ_k, Σ_k à partir de γ :

$$\begin{aligned} * N_k &= \sum_{i=1}^n \gamma_{ik} \quad \longrightarrow \quad \pi_k = N_k/n \\ * \mu_k &= \frac{1}{N_k} \sum_{i=1}^n \gamma_{ik} x_i \\ * \Sigma_k &= \frac{1}{N_k} \sum_{i=1}^n \gamma_{ik} (x_i - \mu_k)(x_i - \mu_k)^T \end{aligned}$$

- **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 (μ, Σ) à partir d’un DP
- *sample_f(alpha, prob, mean, sigma, nu_0, lambda_0, tau = 1e-3)* : Simule DP

