

Chapitre 9

Exemples d'applications

9.1 Estimation d'un facteur de l'asymétrie

La notion d'asymétrie est une notion très générale en physique. Il s'agit de mesurer la disparité entre deux objets ou deux phénomènes ou deux concepts. Limitons nous ici à l'étude de la disparité entre deux grandeurs comparables et non-négatives X et Y . Idéalement, une mesure de l'asymétrie doit satisfaire un certain nombre de propriétés intuitives, parmi elles, et la plus importante, est l'invariance d'échelle. En Physique, une mesure de la disparité entre deux grandeurs comparables et non-négatives X et Y très utilisée est

$$R = \frac{X - Y}{X + Y},$$

qui est une grandeur sans dimension, invariant par changement d'échelle et limité entre -1 et +1. L'application concernée ici est l'étude de la micro-structure des couches magnétique des unité de stockage informatique comme la bande, les disquettes ou les disques magnétiques en examinant l'étendue et l'alignement des spins des électrons sur des surfaces magnétiques concernées. La propriété magnétique des matériaux étant directement reliée à l'alignement plus ou moins parfaite de ces électrons polarisés. La polarisation des spins de ces électrons est proportionnelle à l'amplitude de la magnétisation et peut être examinée par un microscope électronique à balayage. Un faisceau d'électrons est envoyé sur la surface de l'objet examiné qui cause l'émission des électrons de la surface de l'objet qui gardent leur orientations de spins. L'analyse de l'asymétrie se fait en déterminant les taux d'émissions des électrons de spins différentes λ_1 et λ_2 et en définissant

$$R = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}.$$

Balayant ainsi la surface d'un objet point par point on peut fournir une image de l'asymétrie de la magnétisation qui représente, par exemple, une image des bits sur le disque. L'estimation de λ_1 et de λ_2 et donc celle de R nécessite une modélisation statistique car ce que l'appareil mesure est le nombre d'électrons avec les spins 0 ou 1. Le problème est encore un peu plus délicat du à l'existence du bruit.

En pratique, on peut disposer de trois mesures n_1 , n_2 et n et le problème d'un point de vue d'estimation devient: estimer $R = (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$ connaissant n_1 , n_2 et n en faisant l'hypothèse que ces trois valeurs sont des réalisations de trois processus de Poisson avec les paramètres respectifs $\lambda_1 + \lambda$ et $\lambda_2 + \lambda$ et λ . Examinons maintenant

les différentes solutions classiquement proposées.

– Maximum de vraisemblance :

$$L(\lambda_1, \lambda_2, \lambda | n_1, n_2, n) \propto (\lambda_1 + \lambda)^{n_1} (\lambda_2 + \lambda)^{n_2} \lambda^n \exp[-\lambda_1 - \lambda_2 - 3\lambda]$$

$$\hat{\lambda}_1 = \max(0, n_1 - n), \quad \hat{\lambda}_2 = \max(0, n_2 - n), \quad \hat{\lambda} = \max(0, n)$$

$$\hat{R} = \frac{n_1 - n_2}{n_1 + n_2 - 2n}.$$

– Bootstrap : On fait l'hypothèse que $n_1 \sim Pn(n_1 | \lambda_1)$, $n_2 \sim Pn(n_2 | \lambda_2)$ et $n \sim Pn(n | \lambda)$. On estime alors séparément λ_1 , λ_2 et λ (par MV) à partir de n_1 , n_2 et n . Le résultat est

$$\hat{\lambda}_1 = n_1, \quad \hat{\lambda}_2 = n_2, \quad \hat{\lambda} = n.$$

On génère alors N échantillons $(n_1^i, n_2^i, n^i), i = 1, \dots, N$ de trois variables aléatoires avec des lois de Poisson et des paramètres respectives $\hat{\lambda}_1 = n_1$, $\hat{\lambda}_2 = n_2$, $\hat{\lambda} = n$. Parmi ces échantillons on écarte ceux qui ne satisfont pas les conditions suivantes

$$n^i \leq n_1^i, \quad n^i \leq n_2^i, \quad 2n^i < n_1^i + n_2^i,$$

et pour chaque ensemble (n_1^i, n_2^i, n^i) qui satisfait ces contraintes on calcule $R^i = \frac{n_1^i - n_2^i}{n_1^i + n_2^i - 2n^i}$ et finalement on estime R par la moyenne empirique de ces R^i .

– Approche bayésienne :

$$L(\lambda_1, \lambda_2, \lambda | n_1, n_2, n) = \frac{1}{n_1! n_2! n!} (\lambda_1 + \lambda)^{n_1} (\lambda_2 + \lambda)^{n_2} \lambda^n \exp[-\lambda_1 - \lambda_2 - 3\lambda]$$

$$\lambda_1 \sim \mathbf{Gam}(\lambda_1 | \zeta, \beta_1), \quad \lambda_2 \sim \mathbf{Gam}(\lambda_2 | \zeta, \beta_2), \quad \lambda \sim \mathbf{Gam}(\lambda | \zeta, \beta)$$

Il est intéressant de voir qu'avec ces choix on obtient une loi Béta pour R sur l'intervalle $[-1, 1]$ ce qui est une bonne surprise :

$$\pi_R(r) = \frac{\Gamma(\beta_1 + \beta_2)}{2^{\beta_1 + \beta_2 - 1} \Gamma(\beta_1) \Gamma(\beta_2)} (1+r)^{\beta_1 - 1} (1-r)^{\beta_2 - 1}$$

$$E[R | \beta_1 + \beta_2] = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}$$

$$\text{Var}[R | \beta_1 + \beta_2] = \frac{4\beta_1\beta_2}{(\beta_1 + \beta_2)^2 (1 + \beta_1 + \beta_2)}$$

$$M[R | \beta_1 + \beta_2] = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2 - 2}$$

Utilisation du règle de Bayes nous amène à :

$$\pi_R(r | n_1, n_2, n) \propto (1-r)^{\beta_2 - 1} (1+r)^{b + \beta_1 - 1} \int_0^\infty \int_0^\infty f(\lambda_1 + \lambda) d\lambda_1 d\lambda,$$

avec

$$f(\lambda_1 + \lambda) = \frac{\lambda_1^{\beta_1 + \beta_2 - 1} \lambda^{n + \beta - 1} (\lambda_1 + \lambda)^{n_1} \left(\frac{1-r}{1+r} \lambda_1 + \lambda\right)^{n_2}}{(2\lambda_1 + a(1+r))^{b + \beta_1 + \beta_2}} \exp\left[-\frac{2}{1+r} \lambda_1 - (\alpha + 3)\lambda\right]$$

Simplification :

$$\alpha = 0, \quad \beta_1 = \beta_2 = \beta = 1$$

$$\pi_R(r|n_1, n_2, n) \propto (1+r)^{b+\beta_1-1} \int_0^\infty \int_0^\infty f(\lambda_1 + \lambda) d\lambda_1 d\lambda,$$

with

$$f(\lambda_1 + \lambda) = \frac{\lambda_1^{\beta_1+\beta_2-1} \lambda^{n+\beta-1} (\lambda_1 + \lambda)^{n_1} \left(\frac{1-r}{1+r} \lambda_1 + \lambda\right)^{n_2}}{(2\lambda_1 + a(1+r))^{b+\beta_1+\beta_2}} \exp \left[-\frac{2}{1+r} \lambda_1 - (\alpha + 3)\lambda \right]$$

Complication :

$$\lambda_1 \sim \mathbf{Gam}(\lambda_1|\zeta, \beta_1), \quad \lambda_2 \sim \mathbf{Gam}(\lambda_2|\zeta, \beta_2), \quad \lambda \sim \mathbf{Gam}(\lambda|\alpha, \beta)$$

$$\zeta \sim \mathbf{Gam}(\alpha|a, b)$$

Il est intéressant de voir qu'avec ces choix on obtient une loi Béta pour R sur l'intervalle $[-1, 1]$ ce qui est une bonne surprise :

$$\pi_R(r) = \frac{\Gamma(\beta_1 + \beta_2)}{2^{\beta_1+\beta_2-1} \Gamma(\beta_1) \Gamma(\beta_2)} (1+r)^{\beta_1-1} (1-r)^{\beta_2-1}$$

$$\mathbb{E}[R|\beta_1 + \beta_2] = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}$$

$$\text{Var}[R|\beta_1 + \beta_2] = \frac{4\beta_1\beta_2}{(\beta_1 + \beta_2)^2(1 + \beta_1 + \beta_2)}$$

$$\mathbb{M}[R|\beta_1 + \beta_2] = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2 - 2}$$

La loi conjointe de (λ_1, λ_2) indépendamment à ζ

$$\pi(\lambda_1, \lambda_2|a, b, \beta_1, \beta_2) = \frac{a^b \Gamma(b + \beta_1 + \beta_2)}{\Gamma(b) \Gamma(\beta_1) \Gamma(\beta_2)} \frac{\lambda_1^{\beta_1-1} \lambda_2^{\beta_2-1}}{(a + \lambda_1 + \lambda_2)^{b+\beta_1+\beta_2}}$$

$$\mathbb{E}[\lambda_1|a, b, \beta_1, \beta_2] = \frac{\beta_1 a}{b-1}, \quad \mathbb{E}[\lambda_2|a, b, \beta_1, \beta_2] = \frac{\beta_2 a}{b-1}.$$

Utilisation du règle de Bayes nous amène à :

$$\pi_R(r|n_1, n_2, n) \propto (1-r)^{\beta_2-1} (1+r)^{b+\beta_1-1} \int_0^\infty \int_0^\infty f(\lambda_1 + \lambda) d\lambda_1 d\lambda,$$

avec

$$f(\lambda_1 + \lambda) = \frac{\lambda_1^{\beta_1+\beta_2-1} \lambda^{n+\beta-1} (\lambda_1 + \lambda)^{n_1} \left(\frac{1-r}{1+r} \lambda_1 + \lambda\right)^{n_2}}{(2\lambda_1 + a(1+r))^{b+\beta_1+\beta_2}} \exp \left[-\frac{2}{1+r} \lambda_1 - (\alpha + 3)\lambda \right]$$

9.2 Segmentation d'image

$$X_i | \theta_i \sim \mathbf{N}(\theta_i, \sigma_i^2), \quad \sigma_i^2 \text{ connue}$$

$$\theta_i | \mu_i \sim \mathbf{N}(\mu_i, r_i^2), \quad r_i^2 \text{ connue}$$

$$\mu_i \sim \mathbf{N}(\bar{\mu}_i, \tau^2), \quad m_i \text{ et } \tau^2 \text{ connues}$$

$$\boldsymbol{\mu} \sim \mathbf{N}_n(\bar{\boldsymbol{\mu}}, \mathbf{T}), \quad \mathbf{T} = \tau^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}], \quad \tau > 0, \quad -\frac{1}{n-1} < \rho < 1,$$

avec \mathbf{I} une matrice d'identité et \mathbf{J} une matrice avec tous les éléments égaux à un.

$$\mathbb{E}[\mu_i] = \bar{\mu}_i, \quad \text{Var}[\mu_i] = \tau^2, \quad \text{Cov}[\mu_i, \mu_j] = \rho, \quad (i \neq j).$$

Considérons maintenant une partition g de l'ensemble des pixels $S = \{1, \dots, n\}$:

$$S = \{1, \dots, n\} = \{S_1(g), \dots, S_d(g)\}$$

et supposons que tous les variables μ_i se trouvant dans un sousensemble S_k

$$\{\mu_i : i \in S_k(g)\}$$

sont égaux.

Exemple :

$$S = \{1, 2, 3, 4, 5\},$$

$$g_1 = \{S_1 = \{1, 3, 5\}, S_2 = \{2, 4\}\}, \quad |S_1| = 3, |S_2| = 2,$$

$$g_2 = \{S_1 = \{1, 2\}, S_2 = \{3, 4, 5\}\}, \quad |S_1| = 2, |S_2| = 3,$$

Dans la partition g_1 on suppose $\mu_1 = \mu_3 = \mu_5 = \mu_1(g)$, et $\mu_2 = \mu_4 = \mu_2(g)$.

Pour une partition g fixée on peut déduire

$$\mathbb{E}[\mu_k(g)] = \frac{1}{|S_k(g)|} \sum_{i \in S_k(g)} \bar{\mu}_i = m_k(g), \quad \text{Var}[\mu_k(g)] = \tau^2 \delta_k(g),$$

$$\text{Cov}[\mu_k(g), \mu_l(g)] = \frac{\rho}{\sqrt{\delta_k(g)\delta_l(g)}}, \quad (k \neq l),$$

où

$$\delta_k(g) = \frac{1 + \rho(|S_k(g)| - 1)}{|S_k(g)|}.$$

Notons que dans cette modélisation, plus le nombre d'éléments dans un sousensemble augmente plus la variance $\text{Var}[\mu_k(g)]$ diminue.

En notant $\mu_k(g) = \{\mu_1(g), \dots, \mu_d(g)\}$ on a

$$X_i | \theta_i \sim \mathbf{N}(\theta_i, \sigma_i^2), \quad \sigma_i^2 \text{ connue}$$

$$\theta_i | \mu_k(g) \sim \mathbf{N}(\mu_k(g), r_i^2), \quad r_i^2 \text{ connue}$$

$$\mu_k(g) \sim \mathbf{N}(m_k(g), \tau^2), \quad m_i \text{ et } \tau^2 \text{ connues}$$

$$\boldsymbol{\mu}(g) \sim \mathbf{N}_d(\mathbf{m}(g), \mathbf{T}(g)),$$

avec

$$\mathbb{E}[\mu_k(g)] = m_k(g), \quad T_{k,k} = \text{Var}[\mu_k(g)] = \tau^2 \delta_k(g), \quad T_{k,l} = \text{Cov}[\mu_k(g), \mu_l(g)] = \rho \frac{\rho}{\sqrt{\delta_k(g)\delta_l(g)}}, \quad (i \neq j).$$

On peut maintenant calculer la loi *a posteriori* de θ étant donnée les données \mathbf{x} pour chaque partition g et choisir la partition qui a la probabilité la plus grande.

Calculons d'abord

$$\mathbb{E}[\theta_i|\mu, \mathbf{x}] = b_i\mu_k + (1 - b_i)x_i, \quad \text{Var}[\theta_i|\mu, \mathbf{x}] = b_i r_k^2, \quad b_i = \frac{\sigma_i^2}{\sigma_i^2 + r_k^2}$$

On en déduit ensuite

$$\begin{aligned} \mathbb{E}[\mu_k|\mathbf{x}] &= \hat{\mu}_k + \sum_{j=1}^d \lambda_{kj}(m_j - \hat{\mu}_j) \\ \text{Var}[\mu_k|\mathbf{x}] &= \tau^2(1 - \rho)\alpha_k \left[\frac{1}{|S_k|} + \gamma \frac{\rho\alpha_k}{1 + \rho(L - 1)} \right] \\ \text{Cov}[\mu_k, \mu_l|\mathbf{x}] &= \gamma \frac{\tau^2\rho(1 - \rho)}{1 + \rho(L - 1)} \alpha_k \alpha_l, \quad (k \neq l), \end{aligned}$$

avec

$$\begin{aligned} \hat{\mu}_k &= \frac{\sum_{i \in S_k} (\sigma_i^2 + r_k^2)^{-1} x_i}{\sum_{i \in S_k} (\sigma_i^2 + r_k^2)^{-1}} \\ \lambda_{kk} &= \alpha_k \left[1 - \frac{\rho|S_k|}{1 + \rho(L - 1)} \beta_k \right] \\ \lambda_{kj} &= -\alpha_k \frac{\rho|S_j|}{1 + \rho(L - 1)} \beta_j, \quad (k \neq j), \\ \alpha_k &= \frac{1}{1 + \frac{\tau^2(1 - \rho)}{|S_k|} \sum_{i \in S_k} (\sigma_i^2 + r_k^2)^{-1}} \\ \beta_k &= 1 + \frac{\frac{\rho}{1 + \rho(L - 1)} \sum_{j=1}^d \alpha_j |S_j| - \alpha_k}{1 - \frac{\rho}{1 + \rho(L - 1)} \sum_{j=1}^d \alpha_j |S_j|} \\ \gamma &= \frac{1}{1 - \frac{\rho}{1 + \rho(L - 1)} \sum_{j=1}^d \alpha_j |S_j|} \end{aligned}$$

On peut aussi en déduire

$$\mathbb{E}[\theta_i|\mathbf{x}] = b_i \mathbb{E}[\mu_k|\mathbf{x}] + (1 - b_i)x_i = b_i \left[\hat{\mu}_k + \sum_{j=1}^d \lambda_{kj}(m_j - \hat{\mu}_j) \right] + (1 - b_i)x_i$$

$$\text{Var}[\theta_i|\mathbf{x}] = b_i \left[b_i \text{Var}[\mu_k|\mathbf{x}] + r_k^2 \right]$$

$$\text{Cov}[\theta_i, \theta_j|\mathbf{x}] = \begin{cases} b_i b_j [b_i \text{Var}[\mu_k|\mathbf{x}]], & j \in S_k(g) \quad (i \neq j) \\ b_i b_j [b_i \text{Cov}[\mu_k, \mu_l|\mathbf{x}]], & j \in S_l(g) \quad (i \neq j), (k \neq l) \end{cases}$$

Interprétons ces résultats :

$\mathbb{E}[\theta_i|\mathbf{x}]$ est une combinaison linéaire de x_i et de $\mathbb{E}[\mu_k|\mathbf{x}]$ qui est elle-même une combinaison linéaire de $\hat{\mu}_k(g)$ et les un facteur d'ajustement qui est une combinaison linéaire des différences entre $m_k(g)$ et $\hat{\mu}_k(g)$.

si $\sigma_i^2 = \sigma^2$ alors $\hat{\mu}_k(g) = \frac{\sum_{i \in S_k} x_i}{|S - k|}$.

si $\rho = 0$ alors $\lambda_{kj} = 0$ ($k \neq j$), $\lambda_{kk} = \alpha_k$ et

$$E[\theta_i|x] = b_i \hat{\mu}_k + \alpha_k (m_k - \hat{\mu}_k) + (1 - b_i) x_i$$

si $\rho = 1$ alors $\alpha_k = \beta_k = 1$, $\lambda_{kj} = -\frac{|S_j|}{L}$, $\lambda_{kk} = 1 - \frac{|S_k|}{L}$ et $E[\mu_k|x]$ ne dépendra de k qu'à travers r_k^2 .

si $r_k^2 \rightarrow \infty$ alors $b_i \rightarrow 0$ et par conséquent $E[\theta_i|x] \rightarrow x_i$ pour toute partition g .

si $\tau^2 \rightarrow 0$ alors $E[\mu_k|x] \rightarrow m_k$: on donne un maximum de rôle à l'*a priori* si $r_k^2 \rightarrow 0$ alors $b_i \rightarrow 1$ et par conséquent $E[\theta_i|x] \rightarrow E[\mu_k|x]$.

Le but final qui consiste à inférer sur une partition g peut se faire par la loi *a posteriori*

$$p(g|x) \propto p(x|g)p(g)$$

La loi $p(x|g)$ est une loi gaussienne avec

$$E[x_i|g] = m_k(g),$$

$$\text{Var}[x_i|g] = \sigma_i^2 + r_k^2(g) + \tau^2 \delta_k(g),$$

$$\text{Cov}[x_i, x_j|g] = \begin{cases} \tau^2 \delta_k(g), & j \in S_k(g) \quad (i \neq j) \\ \rho \tau^2, & j \in S_l(g) \quad (k \neq l) \end{cases}$$

Si *a priori* on choisit une loi uniforme pour g alors on peut calculer

$$\frac{p(g_1|x)}{p(g_2|x)} = \frac{p(x|g_1)}{p(x|g_2)} = \left[\frac{|\Sigma(g_2)|}{|\Sigma(g_1)|} \right]^{1/2} \exp \left[-\frac{1}{\tau^2} [Q(g_1) - Q(g_2)] \right]$$

où $\tau^2 \Sigma(g_k)$ représentent les matrices de covariances de $p(x|g_k)$ et

$$Q(g_k) = [x - m(g_k)]^t \Sigma^{-1}(g_k) [x - m(g_k)]$$

9.3 Reconstruction d'image avec modèle makovien

Considérons le problème de la reconstruction d'une image \mathbf{x} à partir des mesures \mathbf{y} en supposant que la relation entre les mesures et l'image est une relation linéaire. Supposons aussi que l'information *a priori* sur \mathbf{x} est qu'il s'agit des images constantes par régions et on a quelques informations statistiques concernant les niveaux de gris et la taille de ses régions. Il s'agit alors de proposer une méthode de reconstruction qui puissent utiliser ces informations.

Par la suite nous utiliserons les notations suivantes :

$$\mathbf{x} = \{x_{i,j}\}, \quad \mathbf{y} = \{y_{r,\phi}\}, \quad \mathbf{b} = \{b_{r,\phi}\},$$

où \mathbf{x} représentent les pixels de l'image, \mathbf{y} les mesures, par exemple les projections suivant différentes angles :

$$y_{r,\phi} = \sum_{i,j} H_{r,\phi,i,j} x_{i,j} + y_{r,\phi} \longrightarrow \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}.$$

\mathbf{b} représentent les erreurs et \mathbf{A} est une matrice dont les éléments sont définis par la réponse du système de mesure $H_{r,\phi,i,j}$.

Pour simplifier les notations la position d'un pixel (i, j) est notée

$$s = (i, j), \quad s' = (i', j'), \quad t = (i, j)$$

L'hypothèse du bruit de mesure gaussien nous définit :

$$\mathbf{y}|\mathbf{x} \sim \mathbf{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \sigma^2 \mathbf{I})$$

Pour traduire notre information *a priori* sur l'image on peut utiliser un modèle markovien de label :

$$\mathbf{x}|\lambda \sim \frac{1}{Z(\lambda)} \exp \left[\lambda \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) \right],$$

où

$$V_c(\mathbf{x}) = \begin{cases} 0 & \text{si } |C| = 1 \\ 1 & \text{si } x_s = x_t, (s, t) \in C \\ -1 & \text{si } x_s \neq x_t, (s, t) \in C \end{cases}$$

est la fonction potentiel et

$$U(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x})$$

est l'énergie associée. Dans ce modèle $U(\mathbf{x})$ mesure la différence entre le nombre de pixels voisins ayant le même niveau de gris et le nombre de pixels voisins ayant des niveaux de gris différents. Avec cette modélisation on a

$$x_t|x_s \sim \frac{1}{Z_t(\lambda)} \exp \left[\lambda \sum_{s' \in v_t} V_{s',t}(\mathbf{x}) \right]$$

Notons

$$Q(\lambda) = \log Z(\lambda),$$

on peut alors vérifier les relations suivantes :

$$\mathbb{E}[U(\mathbf{x})|\lambda] = \frac{dQ(\lambda)}{d\lambda}, \quad \text{Var}[U(\mathbf{x})|\lambda] = \frac{d^2Q(\lambda)}{d\lambda^2}.$$

Étant donnée les lois $p(\mathbf{y}|\mathbf{x})$ et $p(\mathbf{x})$ on peut calculer la loi *a posteriori*

$$\pi(\mathbf{x}|\lambda, \sigma^2, \mathbf{y}) = \frac{1}{Z(\lambda, \sigma^2, \mathbf{y})} \exp \left[\lambda \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \right],$$

où

$$Z(\lambda, \sigma^2, \mathbf{y}) = \int \exp \left[\lambda \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \right] d\mathbf{x}$$

On peut remarquer que cette loi est aussi de la même forme que la loi *a priori* mais avec une fonction d'énergie différente, ce qui nous permet d'en déduire

$$\pi(x_t|\lambda, \sigma^2, \mathbf{y}, x_s) = \frac{1}{Z_t(\lambda, \sigma^2, \mathbf{y})} \exp \left[\lambda \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \right]$$

où

$$Z(\lambda, \sigma^2, \mathbf{y}) = \int \exp \left[\lambda \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \right] d\mathbf{x}$$

À ce stade, il reste à définir des lois *a priori* pour les hyperparamètres σ^2 et λ . Une approche pragmatique, comme nous l'avons vu dans le chapitre 5, consiste à choisir des lois conjuguées, qui au moins permettent facilement de calculer les lois *a posteriori*.

Ainsi pour σ^2 on choisit une loi gamma inverse

$$\sigma^2|\alpha, \beta \sim \mathbf{IGam}(\sigma^2|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha(\sigma^2)^{\alpha+1}} \exp \left[-\frac{1}{\sigma^2\beta} \right],$$

avec les propriétés suivantes :

$$\mathbb{E}[\sigma^2] = \frac{1}{\beta(\alpha-1)}, \quad \text{Var}[\sigma^2] = \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)}$$

et pour λ on choisit la loi conjuguée correspondante qui est

$$\lambda|n_0, u_0 \sim \frac{1}{h(n_0, u_0)} \exp [n_0 u_0 \lambda - n_0 Q(\lambda)], \quad n_0 > 0, \quad u_0 \in (\min U, \max U),$$

où

$$h(n_0, u_0) = \int \left[\frac{\exp[u_0 a]}{Z(\lambda)} \right]^{n_0} da$$

et

$$\mathbb{E}[U(\mathbf{x})] = u_0.$$

Ce choix nous permet d'avoir

$$\lambda|\mathbf{x} \sim \frac{1}{h(n'_0, u'_0)} \exp [n'_0 u'_0 \lambda - n'_0 Q(\lambda)], \quad n'_0 = n_0 + 1, \quad u'_0 = \frac{n_0 u_0 + U(\mathbf{x})}{n_0 + 1}$$

La loi conjointe devient

$$\pi(\mathbf{x}, \lambda, \sigma^2|\mathbf{y}) \propto (\sigma^2)^{-\frac{rc}{2}-\alpha-1} \exp \left[\lambda(n_0 u_0 + U(\mathbf{x})) - \frac{1}{\sigma^2} \left(\frac{1}{\beta} + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \right) - (n_0 + 1)Q(\lambda) \right]$$

et les différentes lois *a posteriori* marginales deviennent :

$$\pi(\lambda|\mathbf{x}, \sigma^2, \mathbf{y}) \propto \exp [\lambda(n_0 u_0 + U(\mathbf{x})) - (n_0 + 1)Q(\lambda)]$$

$$\pi(\sigma^2 | \mathbf{x}, \mathbf{y}) \propto \mathbf{IGam}(\sigma^2 | \alpha', \beta')$$

avec

$$\alpha' = \alpha + rc/2 = \alpha + \frac{1}{2} \sum_{i,j} |x_{i,j}|, \quad \beta' = \left[\frac{1}{\beta} + \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 \right]^{-1}$$

Ces relations dépendent d'une quantité $Q(\lambda)$ qui ne peut être calculé que numériquement.

Mise en œuvre par échantillonneur de Gibbs :

1. On tire σ^2 avec la loi

$$\pi(\sigma^2 | \mathbf{x}, \mathbf{y}) \propto \mathbf{IGam}(\sigma^2 | \alpha', \beta')$$

2. On tire λ avec la loi

$$\pi(\lambda | \mathbf{x}, \sigma^2, \mathbf{y}) \propto \exp [\lambda(n_0 u_0 + U(\mathbf{x})) - (n_0 + 1)Q(\lambda)]$$

3. On tire $x_{i,s}$ avec la loi

$$x_t | x_s \sim \frac{1}{Z_t(\lambda)} \exp \left[\lambda \sum_{s' \in v_t} V_{s',t}(\mathbf{x}) \right]$$

