Méthode variationnelle pour l'estimation d'un modèle spatial de données binaires

Cécile Hardouin
MODAL'X, Université Paris Nanterre

SIXIEMES RECONTRES DE STATISTIQUE AVIGNON-MARSEILLE
2018

Motivation

Données binaires 0, 1
Modèle auto-logistique, modèle hiérachique

Estimation : Vraisemblance complétée, et problème dans l'étape E

- Méthodes de Monte-Carlo
- Importance sampling
- Approximations de Laplace
- Approche variationnelle

Jaakola T., Jordan M., 2000, Bayesian parameter estimation via variational methods. Statistics and Computing (2000) 10, 25–37

Plan de l'exposé

1. Le modèle
2. Estimation
3. Quelques résultats
4. Discussion

Le modèle

Domaine fini $D \equiv \{\mathbf{s}_i : i = 1, \ldots, n\} \subset \mathbf{R}^2$,
avec $\mathbf{s}_i = (s_{i1}, s_{i2})$ pour $i = 1, ..., n$.
Processus $\mathbf{Z} = (Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n))^{\mathrm{T}}$, à valeurs dans $E = \{0, 1\}^n$,

$$Z(\mathbf{s}_i) \mid Y(\mathbf{s}_i) \sim Ber(p(\mathbf{s}_i))$$

Les $Z(\mathbf{s}_i) \mid Y(\mathbf{s}_i)$ sont indépendantes

Le modèle

Domaine fini $D \equiv \{\mathbf{s}_i : i = 1, \ldots, n\} \subset \mathbf{R}^2$,
avec $\mathbf{s}_i = (s_{i1}, s_{i2})$ pour $i = 1, ..., n$.
Processus $\mathbf{Z} = (Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n))^{\mathrm{T}}$, à valeurs dans $E = \{0, 1\}^n$,

$$Z(\mathbf{s}_i) \mid Y(\mathbf{s}_i) \sim Ber(p(\mathbf{s}_i))$$

Les $Z(\mathbf{s}_i) \mid Y(\mathbf{s}_i)$ sont indépendantes

$$p(\mathbf{s}_i) = \frac{e^{Y(\mathbf{s}_i)}}{1 + e^{Y(\mathbf{s}_i)}}.$$

ou encore $\mathrm{logit}(p(\mathbf{s}_i)) = \ln \dfrac{p(\mathbf{s}_i)}{1 - p(\mathbf{s}_i)} = Y(\mathbf{s}_i)$

Le modèle

On pose

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon(\mathbf{s}).$$

Premier terme : $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), \cdots, X_p(\mathbf{s}))^{\mathrm{T}}$; $\boldsymbol{\beta}$ doit être estimé.

Second terme : variation spatiale à petite échelle,

$$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \Sigma),$$

Σ doit être estimée.
Si $\Sigma = \Sigma(\boldsymbol{\theta})$, les paramètres sont alors $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$.

La vraisemblance

Y n'est pas observé,
On considère la vraisemblance complétée,

$$[\mathbf{Z}, \varepsilon \mid \beta, \Sigma] = [\mathbf{Z} \mid \varepsilon, \beta] \times [\varepsilon \mid \Sigma],$$

log de la vrais. complétée,

$$L_c(\mathbf{Z}, \varepsilon \quad | \quad \beta, \Sigma) = -\sum_{\mathbf{s} \in D} \ln(1 + e^{Y(\mathbf{s})}) + \sum_{\mathbf{s} \in D} Y(\mathbf{s}) Z(\mathbf{s})$$
$$-\frac{1}{2} \ln(\det \Sigma) - \frac{1}{2} \varepsilon^{\mathrm{T}} \Sigma^{-1} \varepsilon - \frac{n}{2} \ln 2\pi$$

avec $Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^{\mathrm{T}} \beta + \varepsilon(\mathbf{s})$.

On utilise un algorithme EM.

Algorithme EM

Notons $\boldsymbol{\varphi} = (\boldsymbol{\beta}, \Sigma)$.
On définit

$$q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l)}) = E\left[L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi}) \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l)}\right].$$

- Initialisation $\hat{\boldsymbol{\varphi}}^{(0)}$,

Algorithme EM

Notons $\boldsymbol{\varphi} = (\boldsymbol{\beta}, \Sigma)$.
On définit

$$q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l)}) = E\left[L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi}) \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l)}\right].$$

- Initialisation $\hat{\boldsymbol{\varphi}}^{(0)}$,
- A la $l-$ième étape,
  Etape E, on calcule $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)})$.
  Etape M, on maximise $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)})$ pour obtenir
  $\hat{\boldsymbol{\varphi}}^{(l)} = \arg\max_{\varphi} q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)})$.

Algorithme EM

Notons $\boldsymbol{\varphi} = (\boldsymbol{\beta}, \Sigma)$.
On définit

$$q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l)}) = E\left[L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi}) \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l)}\right].$$

- Initialisation $\hat{\boldsymbol{\varphi}}^{(0)}$,
- A la $l-$ième étape,
  Etape E, on calcule $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)})$. $E[\varepsilon \mid \mathbf{Z}, \hat{\boldsymbol{\phi}}^{(l)}]$ is unknown
  Etape M, on maximise $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)})$ pour obtenir
  $\hat{\boldsymbol{\varphi}}^{(l)} = \arg\max_{\boldsymbol{\varphi}} q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)})$.

Algorithme EM

E step: Monte Carlo

Importance Sampling

Laplace Approximation

On remplace par un VEM

E step: Monte Carlo

1. On simule M fois les ε sous la loi conditionnelle $\left[\varepsilon \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l)}\right]$.

   Cette loi n'étant pas connue, on simule ε avec un algo. de Metropolis

2. On approxime les espérances $E\left[g(\varepsilon(\mathbf{s})) \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l)}\right]$ par

$\overline{g_M(\mathbf{s})} = \frac{1}{M} \sum_{m=1}^{M} g(\varepsilon^{(m)}(\mathbf{s}))$

Algorithme de Metropolis

1. Initialisation avec tirage de ε_0.

Algorithme de Metropolis

1. Initialisation avec tirage de ε_0.
2. On tire ε_1 de loi $h(\varepsilon_0)$. *Choix de h ?*

Algorithme de Metropolis

1. Initialisation avec tirage de ε_0.
2. On tire ε_1 de loi $h(\varepsilon_0)$. *Choix de h ?*
3. On calcule $f(\varepsilon_0, z)$ et $f(\varepsilon_1, z)$ les densités de $[\varepsilon_0, \mathbf{Z} \mid \boldsymbol{\varphi}]$ et $[\varepsilon_1, \mathbf{Z} \mid \boldsymbol{\varphi}]$ (avec $[\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi}] = [\mathbf{Z} \mid \varepsilon, \boldsymbol{\varphi}] \times [\varepsilon \mid \boldsymbol{\varphi}]$).
   On calcule le ratio d'acceptation de Metropolis

$$r_{0,1} = \frac{f(\varepsilon_1, \mathbf{z})h(\varepsilon_0)}{f(\varepsilon_0, \mathbf{z})h(\varepsilon_1)}$$

   Si $f(\varepsilon_1, \mathbf{z})h(\varepsilon_0) \geq f(\varepsilon_0, \mathbf{z})h(\varepsilon_1)$, alors on garde ε_1. Retour en 2. Sinon, on tire u suivant une loi uniforme sur $[0; 1]$; si $u \leq r_{0,1}$ alors on garde ε_1, sinon on conserve la valeur initiale ε_0. Retour en 2.

Algorithme de Metropolis

Choix de h ? Chib and Greenberg 95

$$h: \quad N(\varepsilon_m, -cH(\varepsilon_m)^{-1})$$

c doit être calibré

Algorithme de Metropolis - Inconvénients

- Choix du burn-in
- Calibrage de c
- Calcul du mode ε_m de $f(\varepsilon, z)$ puis recalculer $N(\varepsilon_m, -cH(\varepsilon_m)^{-1})$.
- Le temps: simulation de M champs

Algorithme de Metropolis - Inconvénients

- Choix du burn-in
- Calibrage de c
- Calcul du mode ε_m de $f(\varepsilon, z)$ puis recalculer $N(\varepsilon_m, -cH(\varepsilon_m)^{-1})$.
- Le temps: simulation de M champs
- ET surtout, le $\sum_{\mathbf{s}\in D} \ln(1 + e^{Y(\mathbf{s})})$ pose problème...

Changements de h....Rien à faire

EM avec Importance Sampling

Principe de l'importance sampling
On a $X_1, ..., X_M$ de loi $f(x)$
$\widetilde{X}_1, ..., \widetilde{X}_M$ de loi $h(x)$, h est l'importance density

$$E_f[X] = \int xf(x)dx = \int x\frac{f(x)}{h(x)}h(x)dx = E_h[X\frac{f(X)}{h(X)}]$$

EM avec Importance Sampling

Principe de l'importance sampling
On a $X_1, ..., X_M$ de loi $f(x)$
$\widetilde{X}_1, ..., \widetilde{X}_M$ de loi $h(x)$, h est l'importance density

$$E_f[X] = \int xf(x)dx = \int x\frac{f(x)}{h(x)}h(x)dx = E_h[X\frac{f(X)}{h(X)}]$$

ici, $X = g(\varepsilon)$, $f = f(\varepsilon \mid \mathbf{Z}, \boldsymbol{\varphi})$ et h une densité pour ε.

EM avec Importance Sampling

1. On simule M fois les ε
2. On approxime $E\left[g(\varepsilon(\mathbf{s})) \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l)}\right]$ par

$$\sum_{m=1}^{M} g(\varepsilon^{(m)}(\mathbf{s}))p_m$$

$$p_m = \frac{\frac{1}{M}\pi(Z(\mathbf{s}) \mid \boldsymbol{\varepsilon}^{(m)}, \hat{\boldsymbol{\varphi}}^{(l)})}{\frac{1}{M}\sum_{m=1}^{M}\pi(Z(\mathbf{s}) \mid \boldsymbol{\varepsilon}^{(m)}, \hat{\boldsymbol{\varphi}}^{(l)})}$$

EM avec Importance Sampling

1. On simule M fois les ε sous la loi conditionnelle $\left[\varepsilon \mid \hat{\boldsymbol{\varphi}}^{(l-1)}\right]$:

$$\boldsymbol{\varepsilon} \sim N(0, \Sigma(\hat{\boldsymbol{\varphi}}^{(l-1)})) \textit{ semble judicieux}$$

2. On approxime $E\left[g(\varepsilon(\mathbf{s})) \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l-1)}\right]$ par $\sum_{m=1}^{M} g(\varepsilon^{(m)}(\mathbf{s}))p_m$

A nouveau le $\sum_{\mathbf{s}\in D} \ln(1+e^{Y(\mathbf{s})})$ pose problème...Il ne se simplifie pas.

Changements de lois pour les ε....Rien à faire

EM avec Approximation de Laplace

La densité $f(\varepsilon \mid \mathbf{Z}, \boldsymbol{\varphi})$ est proportionnelle à $\exp\ L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi})$.
On cherche ε_m le mode de $L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi})$ et on écrit un devt de Taylor à l'ordre 2 de $L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi})$ en ε_m:

$$L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi}) = L_c(\mathbf{Z}, \varepsilon_m \mid \boldsymbol{\varphi}) + (\varepsilon - \varepsilon_m)^{\mathrm{T}} \frac{\partial}{\partial \varepsilon} L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi})|_{\varepsilon_m} + \frac{1}{2}(\varepsilon - \varepsilon_m)^{\mathrm{T}}(H(\varepsilon_m))(\varepsilon - \varepsilon_m) + \ldots$$

avec $H(\varepsilon_m) = \dfrac{\partial^2}{\partial \varepsilon \partial \varepsilon^{\mathrm{T}}} L_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi})|_{\varepsilon_m}$.

EM avec Approximation de Laplace

Alors $f(\varepsilon \mid \mathbf{Z}, \boldsymbol{\varphi})$ est approx. proportionnelle à

$$\exp L_c(\mathbf{Z}, \varepsilon_m \mid \boldsymbol{\varphi}) \times \exp\left[-\frac{1}{2}(\varepsilon - \varepsilon_m)^{\mathrm{T}}(-H(\varepsilon_m))(\varepsilon - \varepsilon_m)\right];$$

c.à.d. une distribution gaussienne.

$$\int f(\varepsilon \mid \mathbf{Z}, \boldsymbol{\varphi}) d\varepsilon = 1 \simeq Cnste \times \exp L_c(\mathbf{Z}, \varepsilon_m \mid \boldsymbol{\varphi}) \times |-H(\varepsilon_m)|^{-1/2} \times (2\pi)^{n/2}$$

Finalement on a les approximations

$$\begin{aligned} E\left[\varepsilon \mid \mathbf{Z}, \boldsymbol{\varphi}\right] &\simeq \varepsilon_m \\ Var(\varepsilon \mid \mathbf{Z}, \boldsymbol{\varphi}) &\simeq -H(\varepsilon_m)^{-1}. \end{aligned}$$

EM avec Approximation de Laplace
L'étape E doit calculer $E\left[L_c(\mathbf{Z},\varepsilon \mid \boldsymbol{\varphi}) \mid \mathbf{Z},\boldsymbol{\varphi}\right]$,
avec $L_c(\mathbf{Z},\varepsilon \mid \boldsymbol{\varphi}) =$
$-\sum_{\mathbf{s}\in D} \ln(1+e^{Y(\mathbf{s})}) + \sum_{\mathbf{s}\in D} Y(\mathbf{s})Z(\mathbf{s}) - \frac{1}{2}\ln(\det\Sigma) - \frac{1}{2}\varepsilon^{\mathrm{T}}\Sigma^{-1}\varepsilon - \frac{n}{2}\ln 2\pi$

Il reste à calculer $E\left[\ln(1+e^{Y(\mathbf{s})})\right]$. On suit la même méthode.

EM avec Approximation de Laplace

C'est facile, assez rapide, mais biais négatif.

En fait, c'est $\ln(1 + e^{Y(\mathbf{s})})$ qui nous embête.
L'approche variationnelle va nous débarrasser de cette fonction.

EM Variationnel

On considère la fonction logistique,

$$g(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}},$$

Jaakola et Jordan (2000):

$$\ln g(x) \geq \ln g(\tau) + \frac{x-\tau}{2} - \lambda(\tau)(x^2 - \tau^2) \qquad \text{(JJ)}$$

où $\lambda(\tau) = \frac{1}{4\tau}\tanh(\tau/2) = \frac{g(\tau)-1/2}{2\tau}$.

De plus, on a l'égalité pour $\tau^2 = x^2$.

On applique ceci à $-\ln(1+e^{Y(\mathbf{s})}) = \ln g(-Y(\mathbf{s}))$, pour chaque $\mathbf{s} \in D$.

EM Variationnel
On introduit des *paramètres variationnels* $\boldsymbol{\tau} = (\tau(\mathbf{s}_1), ..., \tau(\mathbf{s}_n))^{\mathrm{T}}$;

On obtient une fonction minorante pour L_c,

$$L_c(\mathbf{Z}, \boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \Sigma) \geq \widetilde{L}_c(\mathbf{Z}, \boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\tau}),$$

Et en plus, on a $L_c(\mathbf{Z}, \boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \Sigma) = \widetilde{L}_c(\mathbf{Z}, \boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\tau})$ pour $\tau(\mathbf{s}_i)^2 = Y(\mathbf{s}_i)^2, \quad i = 1, \cdots n.$

EM Variationnel
On part d'une initialisation des $\tau(\mathbf{s}_i)$, et alternativement, on maximise en paramètres du modèle, puis on met à jour les paramètres variationnels;

$$Init \quad : \quad L_c(\mathbf{Z}, \varepsilon \mid \beta, \Sigma) = \widetilde{L}_c(\mathbf{Z}, \varepsilon \mid \beta, \Sigma, \boldsymbol{\tau})$$

$$\widetilde{L}_c(\mathbf{Z}, \varepsilon \quad | \quad \beta, \Sigma, \boldsymbol{\tau}) \leq \widetilde{L}_c(\mathbf{Z}, \varepsilon \mid \beta', \Sigma', \boldsymbol{\tau}) \leq \widetilde{L}_c(\mathbf{Z}, \varepsilon \mid \beta', \Sigma', \boldsymbol{\tau}')$$

On itère cette procédure pour obtenir à la fin
$\widetilde{L}_c(\mathbf{Z}, \varepsilon \mid \beta_{\max}, \Sigma_{\max}, \boldsymbol{\tau}_{\max}) \simeq L_c(\mathbf{Z}, \varepsilon \mid \beta_{\max}, \Sigma_{\max})$.

EM Variationnel

$$\ln g(x) \geq \ln g(\tau) + \frac{x-\tau}{2} - \lambda(\tau)(x^2 - \tau^2)$$

$$\widetilde{L}_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\tau}) = T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + T_3(\boldsymbol{\tau}, \boldsymbol{\beta}, \Sigma) - \frac{1}{2}\varepsilon'\Sigma^{-1}\varepsilon - \frac{1}{2}\ln(\det\Sigma)$$

$$\begin{aligned} T_1(\boldsymbol{\tau}) &= \sum_{\mathbf{s}\in D} \left\{ \ln g(\tau(\mathbf{s})) - \frac{\tau(\mathbf{s})}{2} + \tau(\mathbf{s})^2 \lambda(\tau(\mathbf{s})) \right\}, \\ T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) &= \sum_{\mathbf{s}\in D} \left\{ -\lambda(\tau(\mathbf{s}))(X(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta})^2 + X(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta}(Z(\mathbf{s}) - \frac{1}{2}) \right\}, \end{aligned}$$

et

$$T_3(\boldsymbol{\tau}, \boldsymbol{\beta}, \Sigma) = \sum_{\mathbf{s}\in D} \left\{ \begin{array}{c} -\varepsilon(\mathbf{s})^2 \lambda(\tau(\mathbf{s})) \\ +\varepsilon(\mathbf{s}) \left[Z(\mathbf{s}) - \frac{1}{2} - 2\lambda(\tau(\mathbf{s}))\mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} \right] \end{array} \right\}.$$

EM Variationnel

On a

$$T_3(\boldsymbol{\tau}, \boldsymbol{\beta},\Sigma) = \sum_{\mathbf{s}\in D} \left\{ \begin{array}{c} -\varepsilon(\mathbf{s})^2\lambda(\tau(\mathbf{s})) \\ +\varepsilon(\mathbf{s})\left[Z(\mathbf{s}) - \frac{1}{2} - 2\lambda(\tau(\mathbf{s}))\mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta}\right] \end{array} \right\}.$$

Notons $\mathbf{M} = (M(\mathbf{s}_1), ..., M(\mathbf{s}_n))^{\mathrm{T}}$, avec

$$M(\mathbf{s}) = Z(\mathbf{s}) - \frac{1}{2} - 2\lambda(\tau(\mathbf{s}))\mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta}.$$

Alors on écrit

$$T_3(\boldsymbol{\tau}, \boldsymbol{\beta},\Sigma) - \frac{1}{2}\varepsilon^{\mathrm{T}}\Sigma^{-1}\varepsilon = \varepsilon^{\mathrm{T}}\mathbf{M} - \frac{1}{2}\varepsilon^{\mathrm{T}}\mathbf{W}^{-1}\varepsilon$$

où

$$\mathbf{W}^{-1} = \Sigma^{-1} + 2\Lambda(\boldsymbol{\tau}),$$

$\Lambda(\boldsymbol{\tau}) = diag(\lambda(\tau(\mathbf{s})))$

EM Variationnel
On obtient

$$\widetilde{L}_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{\tau}) = T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \varepsilon^{\mathrm{T}}\mathbf{M} - \frac{1}{2}\varepsilon^{\mathrm{T}}\mathbf{W}^{-1}\varepsilon - \frac{1}{2}\ln(\det\Sigma) + con$$

Pour $\boldsymbol{\tau}$ fixé,

$$p(\varepsilon \mid \mathbf{Z}, \boldsymbol{\beta}, \Sigma, \boldsymbol{\tau}) \propto \exp\left\{T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \frac{1}{2}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{W}^{-1}\boldsymbol{\mu}\right\} \times \frac{1}{\sqrt{\det\Sigma}}\exp\left\{-\frac{1}{2}(\varepsilon - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{W}^{-1}(\varepsilon - \boldsymbol{\mu})\right\}.$$

Variational EM

A $\boldsymbol{\tau}$ fixé,

$$[\varepsilon \mid \mathbf{Z}, \beta, \Sigma, \boldsymbol{\tau}] = N(\boldsymbol{\mu}, \mathbf{W})$$

$$\begin{aligned} q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l)}; \boldsymbol{\tau}) &= E\left[\widetilde{L}_c(\mathbf{Z}, \varepsilon \mid \boldsymbol{\varphi}, \boldsymbol{\tau}) \mid \mathbf{Z}, \hat{\boldsymbol{\varphi}}^{(l)}\right] \\ &= T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \widehat{\boldsymbol{\mu}}^{(l)\mathrm{T}}\mathbf{M} \\ &\quad -\frac{1}{2} Tr((\widehat{\mathbf{W}}^{(l)} + \widehat{\boldsymbol{\mu}}^{(l)}\widehat{\boldsymbol{\mu}}^{(l)\mathrm{T}})\mathbf{W}^{-1}) - \frac{1}{2}\ln(\det\Sigma) + const. \end{aligned}$$

VEM $l-$ième tour

$$
\begin{aligned}
q(\boldsymbol{\varphi},\hat{\boldsymbol{\varphi}}^{(l)};\boldsymbol{\tau}) &= E\left[\widetilde{L}_c(\mathbf{Z},\varepsilon \mid \boldsymbol{\varphi},\boldsymbol{\tau}) \mid \mathbf{Z},\hat{\boldsymbol{\varphi}}^{(l)}\right] \\
&= T_1(\boldsymbol{\tau})+T_2(\boldsymbol{\tau},\boldsymbol{\beta}) + \widehat{\boldsymbol{\mu}}^{(l)\mathrm{T}}\mathbf{M} \\
&\quad -\frac{1}{2}Tr((\widehat{\mathbf{W}}^{(l)}+\widehat{\boldsymbol{\mu}}^{(l)}\widehat{\boldsymbol{\mu}}^{(l)\mathrm{T}})\mathbf{W}^{-1}) - \frac{1}{2}\ln(\det\Sigma)+const.
\end{aligned}
$$

Paramètres du modèle: Σ et $\boldsymbol{\beta}$ (dans T_2 et $\mathbf{M}$)
Paramètres variationnels: $\boldsymbol{\tau}$, dans T_1, T_2, $\mathbf{M}$ et $\mathbf{W}$
$M(\mathbf{s}) = Z(\mathbf{s}) - \frac{1}{2} - 2\lambda(\tau(\mathbf{s}))\mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta}$.
$\mathbf{W}^{-1} = \Sigma^{-1} + 2\Lambda(\boldsymbol{\tau})$

Algorithme EM

Variational EM

1. E-step. Calcul de $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)}; \widehat{\boldsymbol{\tau}}^{(l-1)})$.

Algorithme EM

Variational EM

1. E-step. Calcul de $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)}; \widehat{\boldsymbol{\tau}}^{(l-1)})$.
2. M-step for the model parameters.

Algorithme EM

Variational EM

1. E-step. Calcul de $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)}; \widehat{\boldsymbol{\tau}}^{(l-1)})$.
2. M-step for the model parameters.

   a Calcul $\widehat{\boldsymbol{\beta}}^{(l)} = \arg\max_{\boldsymbol{\beta}} \left(T_2(\widehat{\boldsymbol{\tau}}^{(l-1)}, \boldsymbol{\beta}) + \widehat{\boldsymbol{\mu}}^{\mathrm{T}}_{\widehat{\boldsymbol{\tau}}^{(\mathbf{l-1})}, \widehat{\boldsymbol{\beta}}^{(\mathbf{l-1})}, \widehat{\circ}^{(\mathbf{l-1})}} \mathbf{M}_{\widehat{\boldsymbol{\tau}}^{(l-1)}, \boldsymbol{\beta}} \right)$

Algorithme EM

Variational EM

1. E-step. Calcul de $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)}; \widehat{\boldsymbol{\tau}}^{(l-1)})$.
2. M-step for the model parameters.

   a Calcul $\widehat{\boldsymbol{\beta}}^{(l)} = \arg\max_{\beta} \left(T_2(\widehat{\boldsymbol{\tau}}^{(l-1)}, \boldsymbol{\beta}) + \widehat{\boldsymbol{\mu}}^{\mathrm{T}}_{\widehat{\boldsymbol{\tau}}^{(\mathbf{l-1})}, \widehat{\boldsymbol{\beta}}^{(\mathbf{l-1})}, \hat{\circ}^{(\mathbf{l-1})}} \mathbf{M}_{\widehat{\boldsymbol{\tau}}^{(l-1)}, \boldsymbol{\beta}} \right)$

   b MAJ $\widehat{\boldsymbol{\beta}}^{(l)}$; calcul de $\widehat{\mathbf{M}}_{\widehat{\boldsymbol{\tau}}^{(l-1)}, \widehat{\boldsymbol{\beta}}^{(l)}}$ et $\widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\tau}}^{(l-1)}, \widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\Sigma}^{(l-1)}}$,
   Puis calcul
   $\widehat{\Sigma}^{(l)} = \arg\max_{\Sigma} \left\{ -\frac{1}{2} Tr((\widehat{\mathbf{W}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}}) \mathbf{W}^{-1}_{\widehat{\tau}^{(l-1)}, \Sigma}) - \frac{1}{2} \ln(\det \Sigma) \right\}$

Algorithme EM

Variational EM

1. E-step. Calcul de $q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l-1)}; \widehat{\boldsymbol{\tau}}^{(l-1)})$.
2. M-step for the model parameters.

   a Calcul $\widehat{\boldsymbol{\beta}}^{(l)} = \arg\max_{\beta} \left(T_2(\widehat{\boldsymbol{\tau}}^{(l-1)}, \boldsymbol{\beta}) + \widehat{\boldsymbol{\mu}}^{\mathrm{T}}_{\widehat{\boldsymbol{\tau}}^{(\mathbf{l-1})}, \widehat{\boldsymbol{\beta}}^{(\mathbf{l-1})}, \hat{\circ}^{(\mathbf{l-1})}} \mathbf{M}_{\widehat{\boldsymbol{\tau}}^{(l-1)}, \boldsymbol{\beta}} \right)$

   b MAJ $\widehat{\boldsymbol{\beta}}^{(l)}$; calcul de $\widehat{\mathbf{M}}_{\widehat{\boldsymbol{\tau}}^{(l-1)}, \widehat{\boldsymbol{\beta}}^{(l)}}$ et $\widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\tau}}^{(l-1)}, \widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\Sigma}^{(l-1)}}$,
   Puis calcul
   $\widehat{\Sigma}^{(l)} = \arg\max_{\Sigma} \left\{ -\frac{1}{2} Tr((\widehat{\mathbf{W}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}})\mathbf{W}^{-1}_{\widehat{\tau}^{(l-1)}, \Sigma}) - \frac{1}{2} \ln(\det \Sigma) \right\}$

3. Variational parameter update:

   Update $\widehat{\Sigma}^{(l)}$ Then compute $\widehat{\boldsymbol{\tau}}^{(l)} = \arg\max_{\tau} \Big\{ T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(l)})$

$$+\widehat{\boldsymbol{\mu}}^{\mathrm{T}} \mathbf{M}_{\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\Sigma}^{(l)}} - \frac{1}{2} Tr((\widehat{\mathbf{W}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}})\mathbf{W}^{-1}_{\boldsymbol{\tau}, \widehat{\Sigma}^{(l)}}) \Big\}$$

Step 2-a

On veut maximiser (à $\boldsymbol{\tau}$ fixé)

$$T(\boldsymbol{\beta};\boldsymbol{\tau}) = \sum_{\mathbf{s}\in D}\left\{-\lambda(\tau(\mathbf{s}))(X(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta})^2 + X(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta}(Z(\mathbf{s}) - \tfrac{1}{2} - 2\lambda(\tau(\mathbf{s}))\widehat{\mu}(\mathbf{s}))\right\}$$

$\frac{\partial}{\partial\boldsymbol{\beta}}T(\boldsymbol{\tau},\boldsymbol{\beta}) = G(\boldsymbol{\tau},\boldsymbol{\beta})$; soit on résoud $G(\boldsymbol{\tau},\boldsymbol{\beta}) = 0$;
Sinon,

$$\widehat{\boldsymbol{\beta}}^{(k)} = \widehat{\boldsymbol{\beta}}^{(k-1)} - \left(\frac{\partial}{\partial\boldsymbol{\beta}}G(\boldsymbol{\tau},\boldsymbol{\beta})\right)^{-1}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k-1)}} G(\boldsymbol{\tau},\widehat{\boldsymbol{\beta}}^{(k-1)})).$$

Step 2-b

Avec $\mathbf{W}^{-1} = \Sigma^{-1} + 2\Lambda$, (et à $\boldsymbol{\tau}, \boldsymbol{\beta}$ fixés), on cherche

$$\widehat{\Sigma}^{(l)} = \arg\max_{\Sigma} \left\{ -\frac{1}{2} Tr((\widehat{\mathbf{W}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}})\Sigma^{-1}) - \frac{1}{2} \ln(\det \Sigma) \right\}$$

Step 2-b
Avec $\mathbf{W}^{-1} = \Sigma^{-1} + 2\Lambda$, (et à $\boldsymbol{\tau}, \boldsymbol{\beta}$ fixés), on cherche
$\widehat{\Sigma}^{(l)} = \arg\max_{\Sigma} \left\{ -\frac{1}{2} Tr((\widehat{\mathbf{W}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}})\Sigma^{-1}) - \frac{1}{2} \ln(\det \Sigma) \right\}$
Si $\Sigma = \sigma_\varepsilon^2 \mathbf{Q}$, on veut minimiser

$$f(\mathbf{Q}, \sigma_\varepsilon^2) = \frac{1}{\sigma_\varepsilon^2} Tr((\widehat{\mathbf{W}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}})\mathbf{Q}^{-1}) + n \ln \sigma_\varepsilon^2 + \ln(\det \mathbf{Q}),$$

On obtient

$$\sigma_\varepsilon^2(\mathbf{Q}) = \frac{1}{n} Tr((\widehat{\mathbf{W}}_{\boldsymbol{\tau}, \widehat{\Sigma}^{(l-1)}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}})\mathbf{Q}^{-1}).$$

Et on minimise
$g(\mathbf{Q}) = n \ln \left[\frac{1}{n} Tr((\widehat{\mathbf{W}} + \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}^{\mathrm{T}})\mathbf{Q}^{-1}) \right] + \ln \det \mathbf{Q}$.

Si de plus $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta})$, on minimise par rapport à $\boldsymbol{\theta}$.

Step 3

$$q(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}^{(l)}; \boldsymbol{\tau}) = A(\hat{\boldsymbol{\varphi}}^{(l)}; \boldsymbol{\tau}) + \text{ autres termes qui ne dépendent pas de } \boldsymbol{\tau}$$

Après calculs...

$$\widehat{\tau}^{(l)}(\mathbf{s})^2 = (X(\mathbf{s})^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(l-1)})^2 + 2\mathbf{X}(\mathbf{s})^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(l-1)}\widehat{\mu}^{(l-1)}(\mathbf{s}) + \widehat{W}_{\mathbf{ss}}^{(l-1)} + \widehat{\mu}^{(l-1)}(\mathbf{s})^2.$$

Ce résultat est attendu.
L'inégalité (JJ) devient égalité si $x^2 = \tau^2$;
on veut donc $\tau(\mathbf{s})^2 \simeq Y(\mathbf{s})^2 = (\mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon(\mathbf{s}))^2$;

On prend $\widehat{\tau}^{(l)}(\mathbf{s})^2 = E[Y(\mathbf{s})^2 \mid \mathbf{Z}, \widehat{\boldsymbol{\varphi}}^{(l-1)}]$..

Simulations

D lattice 40×60, $n = 2400$ sites.

1. On simule $N_n(\mathbf{0}; \Sigma)$ avec $C(\mathbf{h}) = \sigma^2 e^{-\frac{\|\mathbf{h}\|}{\theta}}$, for $\mathbf{h} \in \mathbb{R}^2$.

$\sigma^2 = 1$. $\theta = 5$, 20;

Simulations

D lattice 40×60, $n = 2400$ sites.
1. On simule $N_n(\mathbf{0}; \Sigma)$ avec $C(\mathbf{h}) = \sigma^2 e^{-\frac{\|\mathbf{h}\|}{\theta}}$, for $\mathbf{h} \in \mathbb{R}^2$.
$\sigma^2 = 1$. $\theta = 5$, 20;

Trend:

$$X(\mathbf{s})^{\mathrm{T}} = (1, \mathbf{s}_1 - 20, \mathbf{s}_2 - 30).$$

On définit $Vs = \frac{1}{n} Tr(\Sigma) + \frac{1}{n} \sum_{i=1}^{n} \left(X(\mathbf{s}_i)^{\mathrm{T}} \boldsymbol{\beta} - \mathrm{average}_{\mathbf{s} \in D} (X(\mathbf{s})^{\mathrm{T}} \boldsymbol{\beta}) \right)^2$.
Suivant Aldworth et Cressie, $Vs \simeq 2$.
Ici on a pris $\beta_0 = \frac{1}{10}$, $\beta_1 = \frac{1}{16}$ et $\beta_2 = \frac{1}{25}$

Simulations

D lattice 40×60, $n = 2400$ sites.
1. On simule $N_n(\mathbf{0}; \Sigma)$ avec $C(\mathbf{h}) = \sigma^2 e^{-\frac{\|\mathbf{h}\|}{\theta}}$, for $\mathbf{h} \in \mathbb{R}^2$.
$\sigma^2 = 1$. $\theta = 5$, 20;

Trend:

$$X(\mathbf{s})^{\mathrm{T}} = (1, \mathbf{s}_1 - 20, \mathbf{s}_2 - 30).$$

On définit $Vs = \frac{1}{n} Tr(\Sigma) + \frac{1}{n} \sum_{i=1}^{n} \left(X(\mathbf{s}_i)^{\mathrm{T}} \boldsymbol{\beta} - \mathrm{average}_{\mathbf{s} \in D}(X(\mathbf{s})^{\mathrm{T}} \boldsymbol{\beta}) \right)^2$.
Suivant Aldworth et Cressie, $Vs \simeq 2$.
Ici on a pris $\beta_0 = \frac{1}{10}$, $\beta_1 = \frac{1}{16}$ et $\beta_2 = \frac{1}{25}$

2. $p(s) = \dfrac{e^{Y(\mathbf{s})}}{1 + e^{Y(\mathbf{s})}}$, $\mathbf{s} \in D$.

Estimation

Resultats sur 400 simulations

	β_0	β_1	β_2	θ	σ_ε^2
Target	0.1	0.0625	0.0417	5	1
Mean	0.0838	0.0586	0.0398	5.3867	0.6204
Std. Dev.	0.183	0.014	0.011	1.328	0.138

	β_0	β_1	β_2	θ	σ_ε^2
Target	0.1	0.0625	0.0417	10	1
Mean	0.1033	0.0605	0.0429	8.1660	0.5767
Std. Dev.	0.283	0.018	0.014	2.524	0.191

Discussion and further works

Conclusion

On présente une méthode spécifique aux données binaires

Extensions

SRE model; $\varepsilon(\mathbf{s}) = \mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} + S(\mathbf{s})^{\mathrm{T}}\eta + \nu(\mathbf{s})$

$\eta \sim N_r(0, K), \ \nu \sim N_n(0, \sigma_\varepsilon^2 I) \qquad \Sigma = SKS^{\mathrm{T}} + \sigma_\varepsilon^2 I$

References Variationnel

James H. Albert and Siddhartha Chib. Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association, Vol. 88, No. 422 (Jun., 1993), pp. 669-679

Jordan M., Ghahramani Z., Jaakkola T., Saul L. 1999. An introduction to variational methods in graphical models. In: Jordan M.I. (Ed.), Learning in graphical models. Cambridge, MA, MIT Press.

Jaakola T., Jordan M., 2000, Bayesian parameter estimation via variational methods. Statistics and Computing (2000) 10, 25–37

Rustagi J. 1976. Variational methods in statistics. New York, Academic press.

References EM

Olivier Cappé, Eric Moulines and Tobias Rydén. 2005. Inference in Hidden Markov Models. Springer Series in Statistics. 2005

Ming-Hui Chen, Qi-Man Shao, Joseph G. Ibrahim. 2000. Monte Carlo methods in Bayesian computation. Springer Series in Statistics. Springer.

Chib S. Greenberg E. 1995 Understanding the Metropolis algorithm. The American statistician Vol 49 n4

Chib S. Greenberg E. 1996 Markov Chain Monte Carlo simulation methods in econometrics . Econometric Theory, 12, 1996, 409-431.

Dempster, A.P., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1–38.

McLachlan, G.J., Krishnan, T., 2008. The EM Algorithm and Extensions, second ed. Wiley-Interscience, New York, NY.

Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods. Springer,

References Spatial

- Aldworth J., Cressie N., 1999. Sampling designs and prediction methods for Gaussian spatial processes. In: Ghosh, S. (Ed.), Mulivariate Analysis, Designs of Experiments, and Survey Sampling. Marckel Dekker, Inc., New York, NY, pp. 1-54.
- Sengupta, A., Cressie, N., 2013. Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions, Spatial Statistics, 4, 14-44.