Algorithmes MCEM VEM et VBEM pour l'estimation d'un processus de Cox log-Gaussien

Céline Delmas - Julia Radoszycki - Nathalie Peyrard - Régis Sabbadin

Unité de Mathématiques et Informatique Appliquées, INRA, Toulouse

Motivation

Modeling of a spatial phenomenon when data are sampled counts

- A regular grid of quadrats $A_1, \cdots A_N$ on a $\mathcal{D} \subset \mathbb{R}^2$
- We observe Y_i the count in quadrat $i \in \mathcal{O} \subset \mathcal{V} = \{1, \cdots, N\}$

3					11	
10	20		5	15		
		1	12		9	
		50				30

Motivation

Such type of data (counts associated with spatial point processes) are encountered in various fields of applications :

- forestry (counts of trees of a given species)
- ecology (sightings of wild animals)
- epidemiology (disease mapping based on reported infection cases)
- environmental sciences (radioactivity counts)
- agronomy (counts of weeds)
- etc ...

Objectives

The log-Gaussian Cox process is often used for modeling this type of data.

1. We derive the parameters estimations by a moments method
2. We propose a MCEM algorithm
3. We present a preliminary comparison
4. We propose a VEM and a VBEM algorithm
5. We present some simulation results

Poisson process

We consider a spatial Poisson process in $\mathbb{R}^2$ with intensity $\lambda = \{\lambda(x), x \in \mathcal{D}\}$.

- $Y_i \sim \mathcal{P}(\Lambda_i)$ with $\Lambda_i = \int_{A_i} \lambda(x)dx$
- Non stochastic $\Lambda_i \Rightarrow Y_i \perp Y_j$ for $i \neq j \Rightarrow$ No statistical correlation between Y_i and Y_j for $i \neq j$
- In practice
    - there may exist a stochastic dependence between the numbers of points observed in non-overlapping domains
    - the intensity of the point process is often uncertain in areas without data

$\Rightarrow$ More convenient and more parsimonious to use a stochastic modeling of this intensity

Cox process

We consider a spatial Poisson process in $\mathbb{R}^2$ with stochastic intensity $\lambda = \{\lambda(x), x \in \mathcal{D}\}$.

- $\lambda(x) = \exp(\beta)\exp(S(x))$
- $S(\cdot)$ is a Gaussian random field centered with variance σ^2 and exponential covariance function

$$\text{Cov}(S(x_1), S(x_2)) = \sigma^2 \exp(-\alpha||x_1 - x_2||)$$

- $Y_i|\Lambda_i \sim \mathcal{P}(\Lambda_i)$ with $\Lambda_i = \int_{A_i} \lambda(x)dx$ approximated by $\Lambda_i = |A_i| \exp(\beta)\exp(S_i)$ where S_i is the value of S at the center of A_i.
- $Y_i|\Lambda_i \perp Y_j|\Lambda_j$ for $i \neq j$.
- 3 parameters in the model $\theta = (\beta, \sigma, \alpha)$, $\beta \in \mathbb{R}$, $\alpha, \sigma \in \mathbb{R}_+^\star$.

Moments method

By straightforward calculations we obtain :

$$
\begin{aligned}
E(Y_i) &= |A_i| \exp(\sigma^2/2) \exp(\beta) && (1)\\
\mathrm{Var}(Y_i) &= |A_i| \exp(\sigma^2/2) \exp(\beta) \\
&+ |A_i|^2 \exp(\sigma^2) \exp(2\beta)(\exp(\sigma^2) - 1) && (2)\\
E(Y_i Y_j) &= |A|^2 \exp(2\beta) \exp(\sigma^2(1 + e^{-\alpha r})) && (3)
\end{aligned}
$$

where $r = |i - j|$

β and σ^2 estimations

By equations (1) and (2) :

$$\sigma^2 = \ln\left[\frac{\text{Var}(Y) - E(Y) + E(Y)^2}{E(Y)^2}\right] \quad (4)$$

$$\beta = \ln\left[\frac{E(Y)}{|A|\sqrt{\frac{\text{Var}(Y)-E(Y)+E(Y)^2}{E(Y)^2}}}\right] \quad (5)$$

Y_i, $\forall i$, are identically distributed. By the weak law of large number, $E(Y_i)$ and $\text{Var}(Y_i)$ are estimated by $\bar{Y} = \frac{1}{\#\mathcal{O}}\sum_{i\in\mathcal{O}} Y_i$ and $\hat{V}(Y) = \frac{1}{\#\mathcal{O}}\sum_{i\in\mathcal{O}}(Y_i - \bar{Y})^2$

$\Rightarrow$ $\hat{\beta}$ and $\hat{\sigma}^2$ are obtained by replacing $E(Y_i)$ and $\text{Var}(Y_i)$ by $\bar{Y}$ and $\hat{V}(Y)$ in (4) and (5).

α estimation

Using equation (3) we obtain :

$$\hat{\alpha} = -\frac{1}{r} \ln \left[\frac{1}{\hat{\sigma}^2} \ln \left[\frac{\hat{E}(Y_i Y_j)}{|A_i|^2 \exp(2\hat{\beta})} \right] - 1 \right]$$

$E(Y_i Y_j)$ is estimated by using the variogram estimation. Indeed, since :

$$\begin{aligned} \gamma(r) &= \frac{1}{2} \mathsf{Var}(Y(x) - Y(x+r)) \\ &= \mathsf{Var}(Y) - \mathsf{Cov}(Y(x), Y(x+r)) \end{aligned}$$

We easily deduce

$$\hat{E}(Y_i Y_j) = \widehat{\mathsf{Var}}(Y) - \hat{\gamma}(r) + \bar{Y}^2$$

where :

$$\hat{\gamma}(r) = \frac{1}{2\#\mathcal{O}_r} \sum_{\mathcal{O}_r} (Y_i - Y_j)^2$$

Remarks about the moments method

- These estimators are easily calculated
- But these estimators are not optimal

$\Longrightarrow$ Maximum likelihood estimators

MCEM algorithm

- Observations Y_i, $i \in \mathcal{O} \subset \mathcal{V} = \{1, \cdots, N\}$.
- Hidden variables $\left\{ \begin{array}{ll} S_i & i \in \mathcal{V} \\ Y_i & i \in \bar{\mathcal{O}} = \mathcal{V} \setminus \mathcal{O} \end{array} \right.$
- Parameters $\theta = (\alpha, \beta, \sigma)$, $\beta \in \mathbb{R}$, α, $\sigma \in \mathbb{R}_+^\star$.
- $\text{Var}(S_1, \cdots, S_N) = \Sigma = \sigma^2 U$

EM algorithm

- E-step : Calculation of $p(s, y_{\bar{\mathcal{O}}} | y_{\mathcal{O}}, \theta^{(t)})$
- M-step :

$$\text{argmax}_\theta F(\theta|\theta^{(t)}) = \text{argmax}_\theta E\left[\ln p(S, y_{\mathcal{O}}, Y_{\bar{\mathcal{O}}}|\theta) | y_{\mathcal{O}}, \theta^{(t)}\right]$$

How to simulate from $p(s|y_{\mathcal{O}};\theta^{(t)})$?

How to simulate from $p(s|y_{\mathcal{O}};\theta^{(t)})$?

We use the algorithm given by Emery and Hernandez (Computers & Geosciences , 2010)

1. Simulate from $p(s_{\mathcal{O}}|y_{\mathcal{O}};\theta^{(t)})$ (Gibbs sampler)
2. Simulate from $p(s_{\bar{\mathcal{O}}}|s_{\mathcal{O}})$ (easy using any multivariate Gaussian simulation algorithm)

The Gibbs sampler

The Gibbs sampler consists in iterating for $i \in \mathcal{O}$:

1. Simulate $S_i | S_{\mathcal{O}_{-i}}$. Let s'_i denote the new simulated value of S_i
2. Simulate a uniform random variable U on $[0,1]$.
3. If $p_i U < p'_i$ substitute s'_i for s_i

where :

$$
\begin{aligned}
p_i &= P[Y_{\mathcal{O}} = y_{\mathcal{O}} | S_i = s_i, S_{\mathcal{O}_{-i}} = s_{\mathcal{O}_{-i}}] \\
&= \exp(-\Lambda_i) \frac{\Lambda_i^{y_i}}{y_i!} \prod_{j \in \mathcal{O}_{-i}} \exp(-\Lambda_j) \frac{\Lambda_j^{y_j}}{y_j!} \\
p'_i &= P[Y_{\mathcal{O}} = y_{\mathcal{O}} | S_i = s'_i, S_{\mathcal{O}_{-i}} = s_{\mathcal{O}_{-i}}] \\
&= \exp(-\Lambda'_i) \frac{\Lambda_i'^{y_i}}{y_i!} \prod_{j \in \mathcal{O}_{-i}} \exp(-\Lambda_j) \frac{\Lambda_j^{y_j}}{y_j!}
\end{aligned}
$$

M step : β and σ^2 update

Solving $\frac{\partial F}{\partial \beta} = 0$ and $\frac{\partial F}{\partial \sigma^2} = 0$ leads to :

$$\beta^{t+1} = \ln \left[\frac{\sum_{i \in \mathcal{O}} y_i + \sum_{i \in \bar{\mathcal{O}}} \hat{y}_i^{(t+1)}}{\sum_{i=1}^{N} |A_i| \int \exp(s_i) p(s|y_{\mathcal{O}}, \theta^{(t)}) ds} \right]$$

$$\sigma^{2(t+1)} = \frac{1}{N} \int s U^{-1} s^T p(s|y_{\mathcal{O}}, \theta^{(t)}) ds$$

where :

$$\begin{aligned} \hat{y}_i^{(t+1)} &= \int y_i p(s, y_{\bar{\mathcal{O}}} | y_{\mathcal{O}}, \theta^{(t)}) ds dy_{\bar{\mathcal{O}}} \\ &= \int y_i p(y_{\bar{\mathcal{O}}} | s, y_{\mathcal{O}}, \theta^{(t)}) p(s|y_{\mathcal{O}}, \theta^{(t)}) ds dy_{\bar{\mathcal{O}}} \\ &= \exp(\beta^{(t)}) |A_i| \int \exp(s_i) p(s|y_{\mathcal{O}}, \theta^{(t)}) ds \end{aligned}$$

M step : α update

$$\frac{\partial F}{\partial \alpha} = \int p(s, y_{\bar{\mathcal{O}}} | y_{\mathcal{O}}, \theta^{(t)}) \left[-\frac{1}{2} \frac{\partial \ln |U|}{\partial \alpha} - \frac{1}{2\sigma^2} s \frac{\partial (U^{-1})}{\partial \alpha} s \right] ds dy_{\bar{\mathcal{O}}}$$

Following Searle (1982) :

$$\frac{\partial \ln |U|}{\partial \alpha} = \mathrm{tr}[U^{-1} \frac{\partial U}{\partial \alpha}]$$

$$\frac{\partial U^{-1}}{\partial \alpha} = -U^{-1} \frac{\partial U}{\partial \alpha} U^{-1}$$

So $\frac{\partial F}{\partial \alpha} = 0 \Leftrightarrow$

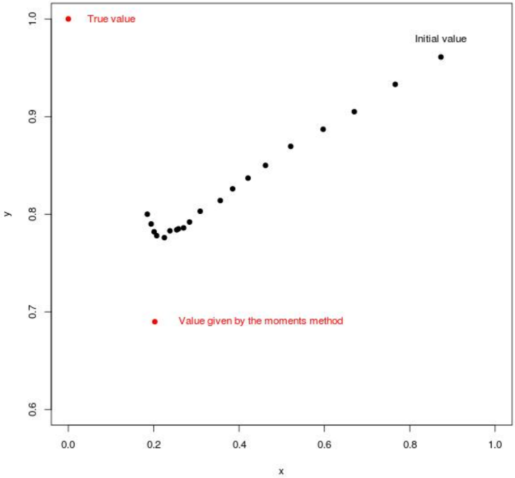
$$\mathrm{tr}(U^{-1} \frac{\partial U}{\partial \alpha}) = \frac{1}{\sigma^2} \int s (U^{-1} \frac{\partial U}{\partial \alpha} U^{-1}) s^T p(s | y_{\mathcal{O}}, \theta^{(t)}) ds$$

solved by a Newton-Raphson algorithm

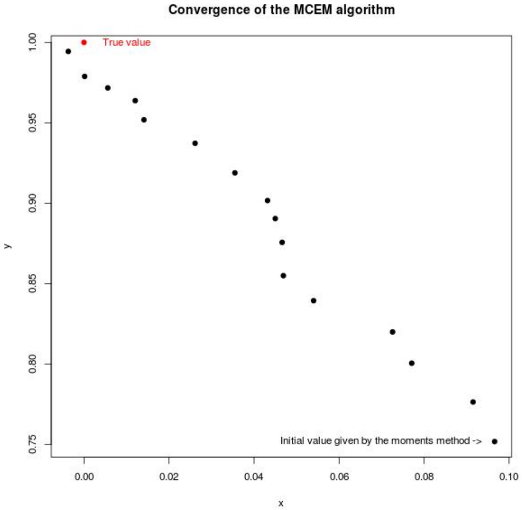
Some simulations

1. Grid size : 10×40 ($N = 400$)
2. Quadrat size : 0.6 ($|A_i| = 0.36$)
3. True values of the parameters $\beta = 0$, $\sigma^2 = 1$, $\alpha = 1$

Some simulations



Some simulations



Variational methods for estimation

We can prove that for any distribution $q_s(s)$ on the hidden variables :

$$\begin{aligned} \ln p(y|\theta) &\geq \int q_s(s) \ln p(s, y|\theta) ds - \int q_s(s) \ln q_s(s) ds \\ &\equiv \mathcal{F}(q_s(s), \theta) \end{aligned}$$

The EM algorithm can be written

- E step : $q_s^{(t+1)} = \text{argmax}_{q_s} \mathcal{F}(q_s(s), \theta^{(t)})$
- M step : $\theta^{(t+1)} = \text{argmax}_{\theta} \mathcal{F}(q_s^{(t+1)}(s), \theta)$

For the E step the exact solution is $q_s^{(t+1)} = p(s|y, \theta^{(t)})$.
For the M step the solution is
$\theta^{(t+1)} = \text{argmax}_{\theta} E[\ln p(S, y|\theta)|y, \theta^{(t)}]$

VEM algorithm

- Too difficult to evaluate the distribution $p(s|y,\theta)$

$\Longrightarrow$ choose a family $\mathcal{Q}$ of (tractable) distributions

- E step : $q_s^{(t+1)} = \text{argmax}_{q_s \in \mathcal{Q}} \mathcal{F}(q_s(s), \theta^{(t)})$
- M step : $\theta^{(t+1)} = \text{argmax}_{\theta} \mathcal{F}(q_s^{(t+1)}(s), \theta)$

- Interest : much faster than MCMC solutions

VBEM algorithm

We can prove that for any distribution $q_s(s)$ on the hidden variables :

$$\begin{aligned}\ln p(y|\theta) &= \ln \int\int p(s,y,\theta)dsd\theta \\ &\geq \ln p(y) - KL\left(q_{s,\theta}(\cdot)|p_{s,\theta}(\cdot|y)\right)\end{aligned}$$

- We choose separable distributions $q_{s,\theta}(\cdot) = q_s(s)q_\theta(\theta)$
- $q_s(s)$ is an approximation of $p(s|y)$
- $q_\theta(\theta)$ is an approximation of $p(\theta|y)$ We note :

$$\begin{aligned}\ln p(y) &\geq \ln p(y) - KL\left(q_{s,\theta}(\cdot)|p_{s,\theta}(\cdot|y)\right) \\ &\equiv \mathcal{F}(q_s(s), q_\theta(\theta))\end{aligned}$$

- E step : $q_s^{(t+1)} = \text{argmax}_{q_s\in\mathcal{Q}}\mathcal{F}(q_s(s), q_\theta^{(t)}(\theta))$
- M step : $q_\theta^{(t+1)} = \text{argmax}_{q_\theta\in\mathcal{Q}'}\mathcal{F}(q_s^{(t+1)}(s), q_\theta(\theta))$

E step

Result : explicit expression for $q_{S_j}(S_j), j = 1...N$ but not a classical distribution :

$q_{S_j}(S_j) \propto \exp(-\mathsf{E}_{q_\theta}[e^{\beta}]e^{S_j}|A_j| + Y_j S_j)\exp\left(-\frac{(S_j - m_j)^2}{2\sigma_j^2}\right)$

where $\forall j = 1...N,\ \sigma_j^2 = \frac{1}{\mathsf{E}_{q_\theta}[(\Sigma^{-1})_{jj}]}$

and $\forall j = 1...N,\ m_j = -\sigma_j^2 \sum_{l \neq j} \mathsf{E}_{q_\theta}[(\Sigma^{-1})_{jl}]\mathsf{E}_{q_{S_l}}[S_l]$

Problem : how to compute $\mathsf{E}_{q_{S_j}}[S_j]$, $\mathsf{E}_{q_{S_j}}[S_j^2]$ and $\mathsf{E}_{q_{S_j}}[\exp(S_j)]$ needed for the M step ?

E step : proposition

$\forall j = 1...N,$

$$\mathsf{E}_{q_{S_j}}[S_j] = \int K \exp(-\mathsf{E}_{q_\theta}[e^\beta]e^{S_j}|A_j| + Y_j S_j) \exp\left(-\frac{(S_j - m_j)^2}{2\sigma_j^2}\right)$$
$$= \mathsf{E}_{\mathcal{N}(m_j,\sigma_j^2)}[K' \exp(-\mathsf{E}_{q_\theta}[e^\beta]e^{S_j}|A_j| + Y_j S_j) S_j]$$

$\rightarrow$ Monte-Carlo estimation

E step

iterate on

1. estimation of $\mathsf{E}_{q_{S_j}}[S_j], j = 1...N$ from simulations according to $\mathcal{N}(m_j, \sigma_j^2)$
2. evaluation of $m_j, j = 1...N$ from the $\mathsf{E}_{q_{S_j}}[S_j], j = 1...N$

evaluation of the other quantities ($\mathsf{E}_{q_{S_j}}[S_j^2]$ and $\mathsf{E}_{q_{S_j}}[\exp(S_j)]$) again from simulations according to $\mathcal{N}(m_j, \sigma_j^2)$

M step : proposition

Result : $q_\theta(\theta) \propto q_\beta(\beta) q_{\sigma,\alpha}(\sigma,\alpha)$ where

$$q_\beta(\beta) \propto \exp\left(\sum_{k=1\ldots N}\left(-e^{\beta}|A_k|\mathsf{E}_{q_{S_k}}[e^{S_k}] + \beta y_k\right)\right) p(\beta)$$

$$q_{\sigma,\alpha}(\sigma,\alpha) \propto \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\sum_{k=1\ldots N}\left((\Sigma^{-1})_{kk q_{S_k}}[S_k^2] +_{q_{S_k}} [S_k] \sum_{l\neq k}(\Sigma^{-1})_{kl}\mathsf{E}_{q_{S_l}}[S_l]\right)\right) p(\sigma)p(\alpha)$$

→ non classical distributions

→ again evaluation of the $\mathsf{E}_{q_\beta}[e^{\beta}]$ and $\mathsf{E}_{q_{\sigma,\alpha}}[\Sigma^{-1}]$ from simulations according to the *a priori* laws

Evaluation

- grid size : $10x40$ $(N = 400)$
- quadrat size : $0.6m$ $(\forall j, |A_j| = 0.36m^2)$

a priori laws

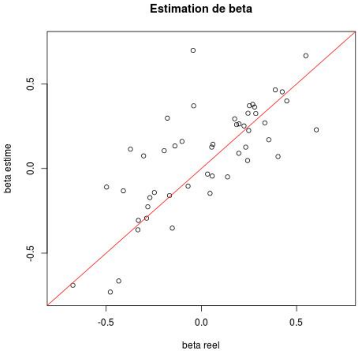
$\beta \sim \mathcal{N}(0, 0.1), \; \ln \sigma \sim \mathcal{N}(0, 0.05), \; \ln \alpha \sim \mathcal{N}(0, 0.2)$

one experiment ($\times N_S = 50$)

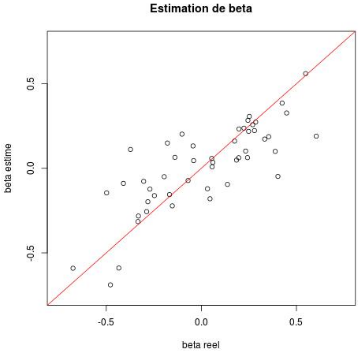
1. generate parameters α, β et σ from the *a priori* law
2. generate the hidden Gaussian field $S = \{S_j, j = 1...N\}$ at each quadrat center
3. generate counts $Y = \{Y_j, j = 1...N\}$ at each quadrat
4. estimation of S and parameters from counts using VBEM and MCMC

Estimation of β

$$\lambda_x = \exp(\beta + S_x),\ \Sigma_{xx'} = \sigma^2 \exp(-\alpha \parallel x - x' \parallel)$$



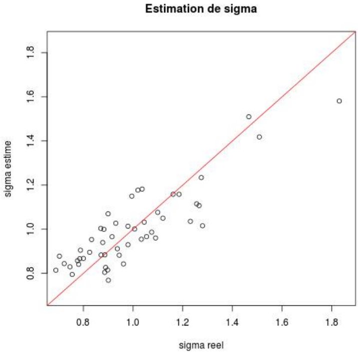
VBEM



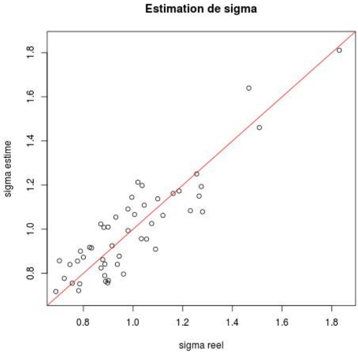
MCMC

Estimation of σ

$$\lambda_x = \exp(\beta + S_x),\ \Sigma_{xx'} = \sigma^2 \exp(-\alpha \parallel x - x' \parallel)$$



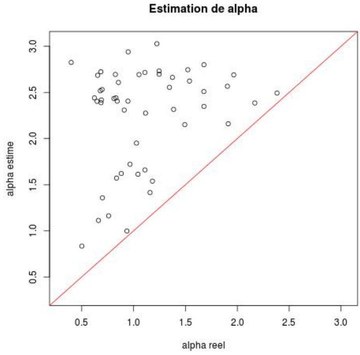
VBEM



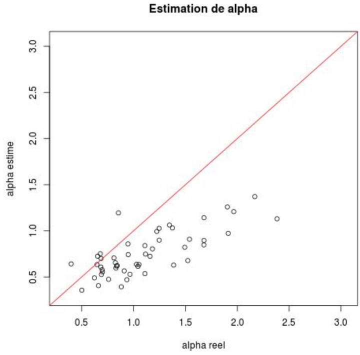
MCMC

Estimation of α

$$\lambda_x = \exp(\beta + S_x),\ \Sigma_{xx'} = \sigma^2 \exp(-\alpha \parallel x - x' \parallel)$$

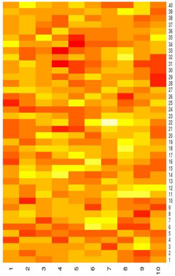


VBEM

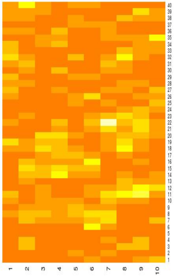


MCMC

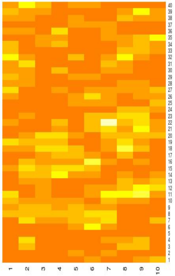
Example of hidden field S



simulation



VBEM



MCMC

Mean Square Error

$$\hat{\mathrm{MSE}}(\hat{\beta}) = \frac{1}{N_S}\sum_{i=1}^{N_S}(\hat{\beta}_i - \beta_i)^2$$
$$\hat{\mathrm{MSE}}(\hat{S}) = \frac{1}{N \times N_S}\sum_{i=1}^{N_S}\sum_{j=1}^{N}(\hat{S}_j^i - S_j^i)^2$$

	β	σ	α	S
VBEM	0.052	0.012	1.69	0.56
MCMC	0.034	0.011	0.22	0.52

$\rightarrow$ similar MSE except for α

Conclusions and Perspectives

- New methods for parameters estimation in a log-Gaussian Cox process
- MCEM more precise but more time-consuming than the moments method
- VBEM much faster than MCMC for bayesian estimation of LGCP (24h vs 12 min)
- Similar estimation quality for VBEM and MCMC except for α
- Comparison with INLA in a Bayesian framework (Rue et al., 2009)
- Comparison MCEM and VEM in a frequentist framework