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 $i \in \mathcal{O} \subset \mathcal{V} = \{1, \cdots, N\}$

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>5</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>12</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
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 ...
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
We consider a spatial Poisson process in $\mathbb{R}^2$ with intensity $
abla = \{\lambda(x), x \in 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 $\sigma^2$ 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 $\theta = (\beta, \sigma, \alpha)$, $\beta \in \mathbb{R}$, $\alpha, \sigma \in \mathbb{R}_+^*$. 
By straightforward calculations we obtain:

\[ E(Y_i) = |A_i| \exp(\sigma^2/2) \exp(\beta) \]  \hspace{1cm} (1)

\[ \text{Var}(Y_i) = |A_i| \exp(\sigma^2/2) \exp(\beta) \]
\[ + |A_i|^2 \exp(\sigma^2) \exp(2\beta)(\exp(\sigma^2) - 1) \]  \hspace{1cm} (2)

\[ E(Y_i Y_j) = |A|^2 \exp(2\beta) \exp(\sigma^2 (1 + e^{-\alpha r})) \]  \hspace{1cm} (3)

where \( r = |i - j| \)
**β and σ² estimations**

By equations (1) and (2):

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

(4)

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

(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 \(\bar{Y} = \frac{1}{\#O} \sum_{i \in O} Y_i\)

and \(\hat{V}(Y) = \frac{1}{\#O} \sum_{i \in O} (Y_i - \bar{Y})^2\)

\(\Rightarrow \hat{\beta}\) 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).
Using equation (3) we obtain:

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

\(E(Y_iY_j)\) is estimated by using the variogram estimation. Indeed, since:

\[
\gamma(r) = \frac{1}{2} \text{Var}(Y(x) - Y(x + r))
\]

\[= \text{Var}(Y) - \text{Cov}(Y(x), Y(x + r))\]

We easily deduce

\[
\hat{E}(Y_iY_j) = \text{Var}(Y) - \hat{\gamma}(r) + \bar{Y}^2
\]

where:

\[
\hat{\gamma}(r) = \frac{1}{2\#O_r} \sum_{O_r} (Y_i - Y_j)^2
\]
Remarks about the moments method

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

⇒ Maximum likelihood estimators
MCEM algorithm

- Observations $Y_i, i \in \mathcal{O} \subset \mathcal{V} = \{1, \cdots, N\}$.
- Hidden variables $\begin{cases} S_i & i \in \mathcal{V} \\ Y_i & i \in \mathcal{O} = \mathcal{V} \setminus \mathcal{O} \end{cases}$
- Parameters $\theta = (\alpha, \beta, \sigma)$, $\beta \in \mathbb{R}$, $\alpha, \sigma \in \mathbb{R}^*_+$. 
- $\text{Var}(S_1, \cdots, S_N) = \Sigma = \sigma^2 U$
• E-step : Calculation of $p(s, y_{O} | y_{O}, \theta^{(t)})$

• M-step :

$$\arg\max_\theta F(\theta | \theta^{(t)}) = \arg\max_\theta E \left[ \ln p(S, y_{O}, Y_{O} | \theta) | y_{O}, \theta^{(t)} \right]$$
How to simulate from $p(s | y_O; \theta^{(t)})$?

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

1. Simulate from $p(s_O | y_O; \theta^{(t)})$ (Gibbs sampler)
2. Simulate from $p(s_\tilde{O} | s_O)$ (easy using any multivariate Gaussian simulation algorithm)
The Gibbs sampler consists in iterating for $i \in O$:

1. Simulate $S_i | S_{O_{-i}}$. 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:

$$
p_i = P[Y_O = y_O | S_i = s_i, S_{O_{-i}} = s_{O_{-i}}] = \exp(-\Lambda_i) \frac{\Lambda_i^{y_i}}{y_i!} \prod_{j \in O_{-i}} \exp(-\Lambda_j) \frac{\Lambda_j^{y_j}}{y_j!}
$$

$$
p'_i = P[Y_O = y_O | S_i = s'_i, S_{O_{-i}} = s_{O_{-i}}] = \exp(-\Lambda'_i) \frac{\Lambda'_i^{y_i}}{y_i!} \prod_{j \in O_{-i}} \exp(-\Lambda_j) \frac{\Lambda_j^{y_j}}{y_j!}
$$
M step: $\beta$ and $\sigma^2$ update

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

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

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

where:

$$
\hat{y}_i^{(t+1)} = \int y_i p(s, y_{\bar{O}}|y_O, \theta^{(t)})dsdy_{\bar{O}}
$$

$$
= \int y_i p(y_{\bar{O}}|s, y_O, \theta^{(t)})p(s|y_O, \theta^{(t)})dsdy_{\bar{O}}
$$

$$
= \exp(\beta^{(t)}|A_i| \int \exp(s_i)p(s|y_O, \theta^{(t)})ds
$$
\[ \frac{\partial F}{\partial \alpha} = \int p(s, y\bar{O}|y\bar{O}, \theta^{(t)}) \left[ -\frac{1}{2} \frac{\partial \ln |U|}{\partial \alpha} - \frac{1}{2\sigma^2} s \frac{\partial (U^{-1})}{\partial \alpha} s \right] ds dy\bar{O} \]

Following Searle (1982):

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

\[ \frac{\partial U^{-1}}{\partial \alpha} = -U^{-1} \frac{\partial U}{\partial \alpha} U^{-1} \]

So \[ \frac{\partial F}{\partial \alpha} = 0 \iff \text{tr}(U^{-1} \frac{\partial U}{\partial \alpha}) = \frac{1}{\sigma^2} \int s(U^{-1} \frac{\partial U}{\partial \alpha} U^{-1}) s^T p(s|y\bar{O}, \theta^{(t)}) ds \]

solved by a Newton-Raphson algorithm
Some simulations

1. Grid size: $10 \times 40 \ (N = 400)$
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
We can prove that for any distribution $q_s(s)$ on the hidden variables:

$$\ln p(y|\theta) \geq \int q_s(s) \ln p(s, y|\theta) ds - \int q_s(s) \ln q_s(s) ds \equiv F(q_s(s), \theta)$$

The EM algorithm can be written

- **E step**: $q_s^{(t+1)} = \arg\max_{q_s} F(q_s(s), \theta^{(t)})$
- **M step**: $\theta^{(t+1)} = \arg\max_{\theta} F(q_s^{(t+1)}(s), \theta)$

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

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

⇒ choose a family $Q$ of (tractable) distributions

• E step : $q^{(t+1)}_s = \arg\max_{q_s \in Q} \mathcal{F}(q_s(s), \theta^{(t)})$

• M step : $\theta^{(t+1)} = \arg\max_{\theta} \mathcal{F}(q^{(t+1)}_s(s), \theta)$

• Interest : much faster than MCMC solutions
We can prove that for any distribution $q_s(s)$ on the hidden variables:

$$
\ln p(y|\theta) = \ln \int \int p(s, y, \theta) ds d\theta \\
\geq \ln p(y) - KL(q_s, \theta(\cdot)|p_s, \theta(\cdot|y))
$$

- 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:

$$
\ln p(y) \geq \ln p(y) - KL(q_s, \theta(\cdot)|p_s, \theta(\cdot|y)) \\
\equiv F(q_s(s), q_{\theta}(\theta))
$$

- E step : $q_s^{(t+1)} = \arg\max_{q_s \in Q} F(q_s(s), q_{\theta}^{(t)}(\theta))$
- M step : $q_{\theta}^{(t+1)} = \arg\max_{q_{\theta} \in Q'} F(q_s^{(t+1)}(s), q_{\theta}(\theta))$
Result: explicit expression for $q_{S_j}(S_j), j = 1...N$ but not a classical distribution:

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

where $\forall j = 1...N$, $\sigma^2_j = \frac{1}{\mathbb{E}_{q\theta}[(\Sigma^{-1})_{jj}]}$

and $\forall j = 1...N$, $m_j = -\sigma^2_j \sum_{l \neq j} \mathbb{E}_{q\theta}[(\Sigma^{-1})_{jl}] \mathbb{E}_{qS_l}[S_l]$

Problem: how to compute $\mathbb{E}_{qS_j}[S_j], \mathbb{E}_{qS_j}[S^2_j]$ and $\mathbb{E}_{qS_j}[\exp(S_j)]$ needed for the M step?
E step: proposition

∀j = 1...N,

\[ E_{qS_j}[S_j] = \int K \exp(-E_{q\theta}[e^\beta e^{S_j}|A_j| + Y_jS_j]) \exp \left( -\frac{(S_j - m_j)^2}{2\sigma_j^2} \right) = E_{\mathcal{N}(m_j,\sigma_j^2)}[K' \exp(-E_{q\theta}[e^\beta e^{S_j}|A_j| + Y_jS_j])S_j] \]

→ Monte-Carlo estimation

E step

iterate on

1. estimation of \( E_{qS_j}[S_j] \), \( j = 1...N \) from simulations according to \( \mathcal{N}(m_j,\sigma_j^2) \)

2. evaluation of \( m_j \), \( j = 1...N \) from the \( E_{qS_j}[S_j] \), \( j = 1...N \)

evaluation of the other quantities (\( E_{qS_j}[S_j^2] \) and \( E_{qS_j}[\exp(S_j)] \)) 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

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

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

\[
\sum_{l \neq k} (\Sigma^{-1})_{kl} E_{q_{S_l}}[S_l]
\]

\( \rightarrow \) non classical distributions
\( \rightarrow \) again evaluation of the \( E_{q_{\beta}}[e^{\beta}] \) and \( E_{q_{\sigma,\alpha}}[\Sigma^{-1}] \) from simulations according to the \textit{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 $\alpha$, $\beta$ et $\sigma$ 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 $\beta$

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

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

$$\lambda_x = \exp(\beta + S_x), \Sigma_{xx'} = \sigma^2 \exp(-\alpha \| x - x' \|)$$
Example of hidden field $S$

- Simulation
- VBEM
- MCMC
Mean Square Error

\[
\hat{\text{MSE}}(\hat{\beta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} (\hat{\beta}_i - \beta_i)^2
\]

\[
\hat{\text{MSE}}(\hat{S}) = \frac{1}{N \times N_S} \sum_{i=1}^{N_S} \sum_{j=1}^{N} (\hat{S}_{ij} - S_{ij})^2
\]

<table>
<thead>
<tr>
<th></th>
<th>(\beta)</th>
<th>(\sigma)</th>
<th>(\alpha)</th>
<th>(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBEM</td>
<td>0.052</td>
<td>0.012</td>
<td>1.69</td>
<td>0.56</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.034</td>
<td>0.011</td>
<td>0.22</td>
<td>0.52</td>
</tr>
</tbody>
</table>

→ similar MSE except for \(\alpha\)
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 $\alpha$
- Comparison with INLA in a Bayesian framework (Rue et al., 2009)
- Comparison MCEM and VEM in a frequentist framework