# An unbiased logistic regression estimating function for spatial Gibbs point processes

Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

joint work (in progress !) with

Adrian Baddeley, Jean-Francois Coeurjolly and Ege Rubak

# Gibbs point process and conditional intensity

Point process $X$: random point pattern.

Conditional intensity $\lambda(u, X)$: for small region $A$ and $u \in A$,

$$|A|\lambda(u, X) \approx P(X \text{ has a point in A}|X \setminus A)$$

GNZ-formula:

$$\mathbb{E} \sum_{u \in X} f(u, X \setminus u) = \int \mathbb{E}[f(u, X)\lambda(u, X)]\mathrm{d}u$$

for non-negative functions $f$.

$X$ observed in bounded region $W$.

Parametric model $\lambda_\theta(u, X)$ for conditional intensity.

Strauss:

$$\lambda_\theta(u, X) = \beta \gamma^{n_R(u,X)}, \quad \beta > 0,\ 0 < \gamma \le 1$$

$n_R(u, X) = \sum_{v \in X \setminus u} 1[\|u - v\| \le R]$: number of neighboring points within distance $R$ from $u$.

Log linear:

$$\lambda_\theta(u, X) = \exp[\theta \cdot t(u, X)]$$

for some statistic $t(u, X)$

E.g. (Strauss):

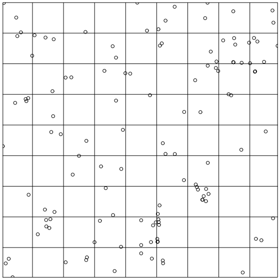
$$t(u, X) = (1, n_R(u, X))$$

# Pseudo-likelihood

Disjoint subdivision $W = \cup_{i=1}^{m} C_i$ in 'cells' $C_i$.

Random indicator variables:
$N_i = 1[X \cap C_i \neq \emptyset]$ (presence/absence of points in $C_i$).

$P(N_i = 1 | X \setminus C_i) \approx |C_i| \lambda_\theta(u_i, X)$,
$u_i \in C_i \setminus X$.



log binary pseudo-likelihood based on $N_i$'s takes logistic regression form:

$$\sum_{i=1}^{m} N_i \log[|C_i|\lambda_\theta(u_i, X)] + (1 - N_i) \log[1 - |C_i|\lambda_\theta(u_i, X)] \approx$$
$$\sum_{i=1}^{m} N_i \log \frac{|C_i|\lambda_\theta(u_i, X)}{1 + |C_i|\lambda_\theta(u_i, X)} + (1 - N_i) \log \frac{1}{1 + |C_i|\lambda_\theta(u_i, X)}$$

Log-linear case

$$\frac{|C_i|\lambda_\theta(u_i, X)}{1 + |C_i|\lambda_\theta(u_i, X)} = \frac{\exp(\theta \cdot t(u_i, X)}{\exp(\theta \cdot t(u_i, X) + |C_i|^{-1}}$$

binary pseudo-likelihood converges to spatial point process pseudo-likelihood ($|C_i| \to 0$):

$$\sum_{u \in X} \log \lambda_\theta(u, X \setminus u) - \int_W \lambda_\theta(u, X) \mathrm{d}u$$

with score

$$\sum_{u \in X} \frac{\lambda'_\theta(X \setminus u)}{\lambda_\theta(u, X \setminus u)} - \int_W \lambda'_\theta(u, X) \mathrm{d}u$$

(unbiased due to GNZ formula)

# Bias problems with pseudo-likelihood

- Binary pseudo-likelihood: biased due to approximation

$$P(N_i = 1 | X \setminus C_i) \approx |C_i| \lambda_\theta(u_i, X)$$

- spatial point process pseudo-likelihood: score requires numerical approximation

$$\sum_{u \in X} \frac{\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u)} - \int_W \lambda'_\theta(u, X) \mathrm{d}u \approx \sum_{u \in X} \frac{\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u)} - \sum_{v \in Q} \lambda'_\theta(v, X) w(v)$$

for quadrature weights and points $w(v), v \in Q \Rightarrow$ bias.

# Unbiased 'logistic regression' estimating function

$$s(\theta) = \sum_{u \in X} \frac{\rho(u)\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u)[\lambda_\theta(u, X \setminus u) + \rho(u)]} - \sum_{u \in D} \frac{\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)}$$

$D$: dummy point process of intensity $\rho(\cdot)$ independent of $X$ (random 'quadrature points').

$s(\theta)$ derivative of 'logistic regression'

$$\sum_{u \in X} \log \frac{\lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} + \sum_{u \in D} \log \frac{1}{\lambda_\theta(u, X) + \rho(u)}$$

Advantages:

- unbiased by GNZ and Campbell formulae (later slide)
- formally logistic regression score - computation easy using `glm` with logistic link function.
- tractable asymptotic distribution of parameter estimate in the stationary case
- fast computation - parametric bootstrap feasible in inhomogeneous case

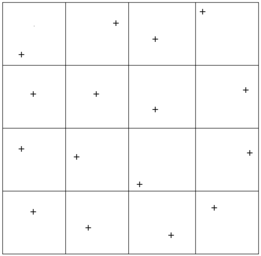
# Dummy point process

Should be easy to simulate and mathematically tractable.

Possibilities:

1. Poisson process
2. binomial point process (fixed number of independent points)
3. stratified binomial point process (`stratrand()` in `spatstat`)

Stratified:



# Relation to pseudo-likelihood

We can rewrite logistic regression score

$$s(\theta) = \sum_{u \in X} \frac{\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u)} - \sum_{u \in (X \cup D)} \frac{\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)}$$

By GNZ and Campbell:

$$\mathbb{E} \sum_{u \in (X \cup D)} \frac{\lambda'_\theta(u, X \setminus u; \theta)}{\lambda_\theta(u, X \setminus u) + \rho(u)} = \mathbb{E} \int_W \lambda'_\theta(u, X) \mathrm{d}u. \qquad (1)$$

Hence

$$\sum_{u \in (X \cup D)} \frac{\lambda'_\theta(u, X \setminus u; \theta)}{\lambda_\theta(u, X \setminus u) + \rho(u)}$$

unbiased Monte Carlo approximation of last term in pseudo-likelihood score:

$$\sum_{u \in X} \frac{\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u)} - \int_W \lambda'_\theta(u, X) \mathrm{d}u$$

# Decomposition of logistic regression score

$$
\begin{aligned}
s(\theta) &= \sum_{u \in X} \frac{\rho(u)\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u)[\lambda_\theta(u, X \setminus u) + \rho(u)]} - \sum_{u \in D} \frac{\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \\
&= \sum_{u \in X} \frac{\rho(u)\lambda'_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u)[\lambda_\theta(u, X \setminus u) + \rho(u)]} - \int_W \frac{\rho(u)\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mathrm{d}u \\
&\quad + \int_W \frac{\rho(u)\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mathrm{d}u - \sum_{u \in D} \frac{\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \\
&= T_1 + T_2
\end{aligned}
$$

$$
\int_W \frac{\rho(u)\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mathrm{d}u = \mathbb{E}[\sum_{u \in D} \frac{\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} | X] \Rightarrow \mathbb{E}[T_2 | X] = 0
$$

By GNZ formula for $X$

$$\mathbb{E}\sum_{u\in X}\frac{\rho(u)\lambda'_\theta(u,X\setminus u)}{\lambda_\theta(u,X\setminus u)[\lambda_\theta(u,X\setminus u)+\rho(u)]}=\mathbb{E}\int_W\frac{\rho(u)\lambda'_\theta(u,X)}{\lambda_\theta(u,X)+\rho(u)}\mathrm{d}u$$

so

$$\mathbb{E}\,T_1=$$

$$\mathbb{E}[\sum_{u\in X}\frac{\rho(u)\lambda'_\theta(u,X\setminus u)}{\lambda_\theta(u,X\setminus u)[\lambda_\theta(u,X\setminus u)+\rho(u)]}-\int_W\frac{\rho(u)\lambda'_\theta(u,X)}{\lambda_\theta(u,X)+\rho(u)}\mathrm{d}u]=0$$

$T_1$ only depends on $X$ and $\mathbb{E}[T_2|X] = 0 \Rightarrow T_1$ and $T_2$ uncorrelated:

$$\mathbb{C}\mathrm{ov}[T_1, T_2] = \mathbb{E}\mathbb{C}\mathrm{ov}[T_1, T_2|X] + \mathbb{C}\mathrm{ov}[\mathbb{E}[T_1|X], \mathbb{E}[T_2|X]] = 0$$

# Approximate distribution of parameter estimate

Parameter estimate $\hat{\theta}$ solution of

$$s(\theta) = 0$$

First order Taylor approximation:

$$s(\theta) \approx S(\hat{\theta} - \theta) \Leftrightarrow \hat{\theta} \approx \theta + S^{-1} s(\theta)$$

where

$$S = -\mathbb{E}[\frac{\mathrm{d}}{\mathrm{d}\theta} s(\theta)]$$

# Approximate distribution of parameter estimate

Parameter estimate $\hat{\theta}$ solution of

$$s(\theta) = 0$$

First order Taylor approximation:

$$s(\theta) \approx S(\hat{\theta} - \theta) \Leftrightarrow \hat{\theta} \approx \theta + S^{-1} s(\theta)$$

where

$$S = -\mathbb{E}[\frac{\mathrm{d}}{\mathrm{d}\theta} s(\theta)]$$

Thus

$$\mathbb{V}\mathrm{ar}\, \hat{\theta} \approx S^{-1} \mathbb{V}\mathrm{ar}\, s(\theta) S^{-1} = S^{-1} \mathbb{V}\mathrm{ar}\, T_1 S^{-1} + S^{-1} \mathbb{V}\mathrm{ar}\, T_2 S^{-1} = \Sigma_1 + \Sigma_2$$

NB: $T_1$ depends only on $X$ while $T_2$ involves both $X$ and $D$.

$\mathbb{V}\mathrm{ar}\, T_2 \to 0$ if $\rho(\cdot) \to \infty$ (dense $D$)

Hence $\Sigma_2$ extra variance due to $D$.

# Asymptotic normality

Restrict attention to stationary $X$ and increasing observation window $W$.

$T_1$ asymptotically $N(0, \Sigma_1)$ by CLT for Gibbs point process innovations (Coeurjolly *et al.*, 2012).

$T_2|X$ asymptotically normal $N(0, \Sigma_2)$ by CLT for independent but not identically distributed random variables.

NB: $\Sigma_2$ does not depend on $X$ !

# Asymptotic normality

Restrict attention to stationary $X$ and increasing observation window $W$.

$T_1$ asymptotically $N(0, \Sigma_1)$ by CLT for Gibbs point process innovations (Coeurjolly *et al.*, 2012).

$T_2|X$ asymptotically normal $N(0, \Sigma_2)$ by CLT for independent but not identically distributed random variables.

NB: $\Sigma_2$ does not depend on $X$ !

Case of stratified points:

$$T_2 = -\sum_{i=1}^{m}[\frac{\lambda_\theta'(U_i, X)}{\lambda_\theta(U_i, X) + \rho(U_i)} - \int_{C_i} \frac{\rho(u)\lambda_\theta'(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mathrm{d}u]$$

where $W = \cup_i C_i$ and $U_i$ uniform on $C_i$.

# Asymptotic normality

Restrict attention to stationary $X$ and increasing observation window $W$.

$T_1$ asymptotically $N(0, \Sigma_1)$ by CLT for Gibbs point process innovations (Coeurjolly *et al.*, 2012).

$T_2|X$ asymptotically normal $N(0, \Sigma_2)$ by CLT for independent but not identically distributed random variables.

NB: $\Sigma_2$ does not depend on $X$ !

Case of stratified points:

$$T_2 = -\sum_{i=1}^{m}[\frac{\lambda'_\theta(U_i, X)}{\lambda_\theta(U_i, X) + \rho(U_i)} - \int_{C_i} \frac{\rho(u)\lambda'_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mathrm{d}u]$$

where $W = \cup_i C_i$ and $U_i$ uniform on $C_i$.

Conclusion

$$\hat{\theta} \approx \theta + S^{-1}T_1 + S^{-1}T_2 \approx N(\theta, S^{-1}[\Sigma_1 + \Sigma_2]S^{-1})$$

# Preliminary simulation results: Strauss process

Strauss process with $\beta = 200$ $\gamma = 0.5$ $R = 0.05$ on unit square.

- Compare numerically approximated pseudo-likelihood (implementation in `spatstat`) with logistic regression score.
- variance decomposition

Mean number of Poisson quadrature/dummy-points:
$625, 25000, 10000, 40000$.

# Bias and variance

Mean and std. dev. of parameter estimates and proportion of variance due to $D$ in % for logistic regression score.

Numbers in [] are with `spatstat` implementation of pseudo-likelihood.

| $\rho$ | $\beta = 200$ $\mathbb{E}\hat{\beta}$ | sd $\hat{\beta}$ | % | $\gamma = 0.5$ $\mathbb{E}\hat{\gamma}$ | sd $\hat{\gamma}$ | % |
|---|---|---|---|---|---|---|
| 625 | 204.2 [168.9] | 37.8 | 13 | 0.502 [0.62] | 0.114 | 5.8 |
| 2500 | 203.4 [188.9] | 35.9 | 4 | 0.502 [0.55] | 0.112 | 1.5 |
| $100^2$ | 203.4 [202.1] | 35.3 | 1 | 0.502 [0.51] | 0.111 | 0.4 |
| $200^2$ | 203.3 [205.5] | 35.2 | 0 | 0.502 [0.505] | 0.111 | 0.1 |

Bias small and does not depend on intensity of dummy points for logistic regression score.

Even smaller variance contributions from $D$ if stratified dummy points used !

Bias problems with `spatstat` !

To be done/work in progress:

- further simulation studies
- applications to data examples
- implementation of estimation procedure and computation of asymptotic covariance matrix in `spatstat` is on the way !

Thanks for your attention !