Quasi-likelihoods for spatial point processes

Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

based on joint work with Yongtao Guan, Abdollah Jalilian, Ganggang Xu, Chong Deng, Emma Zhang

Outline

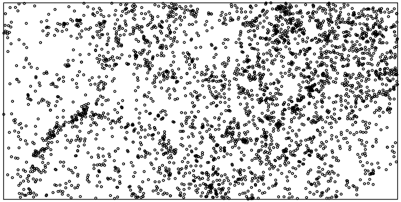
Quasi-likelihood estimating functions for estimation of

- intensity function for spatial point process

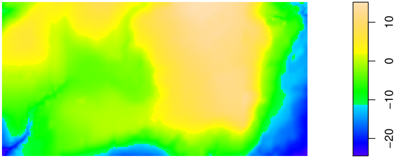
and (if time allows)

- intensity function in case-control setup
- pair correlation function

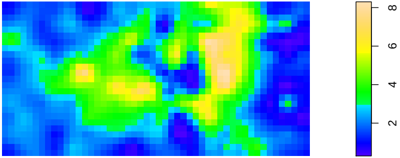
Data example: *Acalypha*



- *observation window W* $= 1000 \text{ m} \times 500 \text{ m}$
- seed dispersal $\Rightarrow$ *clustering*
- environment $\Rightarrow$ *inhomogeneity*



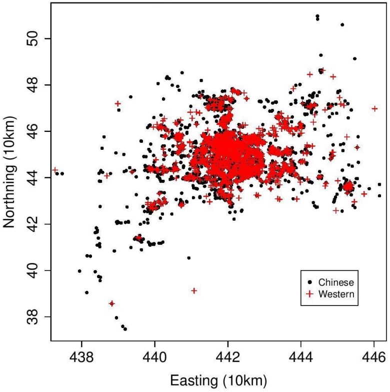
Elevation



Potassium content in soil

Objective: quantify dependence on environmental variables and clustering

Data examples: restaurants in Beijing



- Chinese and western-style restaurants
- Investigate differences in spatial distributions of two types of restaurants
- What is the observation window ?
- How to model intensity function ?

Quasi-likelihood for ordinary numerical data

Gotway and Stroup (1997) suggested to use quasi-likelihood for spatial binary or count data.

Consider data vector

$$N = (N_1, \dots, N_m)^\mathsf{T}$$

with covariance matrix Σ and mean vector (regression model)

$$\mu(\beta) = (\mu_1(\beta), \dots, \mu_m(\beta))^\mathsf{T} \quad \beta \in \mathbb{R}^p$$

Consider all estimating functions

$$A[N - \mu(\beta)] \quad A : p \times m$$

which are *linear* transformations of the residual vector

$$R(\beta) = N - \mu(\beta)$$

Optimal $A = D\Sigma^{-1}$ where $D = \left[\mathrm{d}\mu(u_l)/\mathrm{d}\beta_i\right]_{il}$.

Resulting quasi-likelihood (QL) score function is

$$D\Sigma^{-1}R(\beta)$$

G & S (1997): replace unknown Σ by covariance matrix $V(\psi)$ specified using parametric covariance model from geostatistics.

Advantages of QL:

- only requires specification of mean and covariance
- consistent estimate β even if covariance $V(\psi)$ misspecified
- computationally simple
- statistically efficient (although inferior to MLE)

Questions

- Does there exist underlying distribution over continuous space compatible with a given spatial quasi-likelihood ?
- Want to infer regression models for point pattern data - can we adapt quasi-likelihood to this problem ?
- Computational problem of handling $V(\psi)$?

Point processes and intensity function

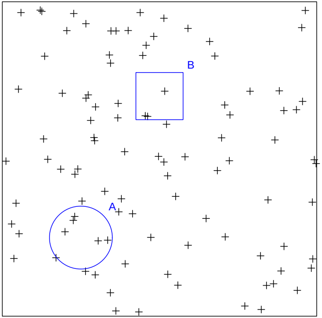
Point process **X**: random subset of the plane

Assume observed in bounded window $W \subset \mathbb{R}^2$. u spatial location in W.

For A subset of the plane, count $N(A)$ is random number of points in A.

$$\mathbb{E}N(A) = \int_A \rho(u)\mathrm{d}u$$

$\rho(\cdot)$: intensity function



Regression model for intensity:

$$\rho(u;\beta) = \exp[\beta z(u)^\mathsf{T}]$$

where $z(u) = (z_1(u), \ldots, z_p(u))^\mathsf{T}$ vector of spatial covariates.

Residual measure

For point process $\mathbf{X}$ and $A \subset \mathbb{R}^2$ *residual measure* is

$$R(A) = N(A) - \mathbb{E}N(A) = \sum_{u \in \mathbf{X}} 1[u \in A] - \int 1[u \in A]\rho(u;\beta)\mathrm{d}u$$

$N(A)$: number of points in A.

In analogy with quasi-likelihood look for optimal linear transformation of the residual measure

$$e_f(\beta) = \int_W f(u)R(\mathrm{d}u) = \sum_{u \in \mathbf{X}} f(u) - \int_W f(u)\rho(u;\beta)\mathrm{d}u$$

where $f : \mathbb{R}^2 \to \mathbb{R}^p$ vector-valued "weight" function.

Estimate $\hat{\beta}_f$ solves estimating equation

$$e_f(\beta) = 0$$

By Campbell formula:

$$\mathbb{E} e_f(\beta) = \mathbb{E} \sum_{u \in \mathbf{X}} f(u) - \int_W f(u) \rho(u; \beta) \mathrm{d}u = 0$$

for true parameter β.

Thus e_f *unbiased* - first step towards consistent parameter estimates.

Example: Poisson/composite likelihood

With

$$f(u) = \frac{\mathrm{d}}{\mathrm{d}\beta} \log \rho(u, \beta) = \frac{\rho'(u; \beta)}{\rho(u; \beta)}$$

we obtain

$$e_f(\beta) = \sum_{u \in \mathbf{X}} \frac{\rho'(u; \beta)}{\rho(u; \beta)} - \int_W \rho'(u; \beta) \mathrm{d}u$$

- for Poisson process this is score of the likelihood function.
- for general **X**: score of composite likelihood function
- computationally easy: just need approximation of integral - Monte Carlo or deterministic quadrature

Estimation variance and optimality

By linear approximation (asymptotically)

$$
\begin{aligned}
e_f(\beta) \approx & e_f(\hat{\beta}_f) + S_f(\hat{\beta}_f - \beta) \\
= & S_f(\hat{\beta}_f - \beta) \Leftrightarrow \\
(\hat{\beta}_f - \beta) \approx & S_f^{-1} e_f(\beta)
\end{aligned}
$$

Variance (sandwich estimator):

$$\mathbb{V}\mathrm{ar}\hat{\beta}_f \approx S_f^{-1} \Sigma_f S_f^{-\mathsf{T}}$$

where (sensitivity):

$$S_f = -\mathbb{E}[\frac{\mathrm{d}}{\mathrm{d}\beta} e_f(\beta)]$$

and

$$\Sigma_f = \mathbb{V}\mathrm{ar} e_f(\beta)$$

Optimality: ϕ is optimal if

$$\mathbb{V}\mathrm{ar}\hat{\beta}_f - \mathbb{V}\mathrm{ar}\hat{\beta}_\phi$$

positive definite for all f.

Sufficient condition: ϕ is optimal if

$$\mathbb{C}\mathrm{ov}[e_\phi, e_f] = S_f$$

for all f.

Covariance can be evaluated using second-order Campbell.

This implies integral equation for ϕ:

$$\phi(u) + \int_W t(u,v)\phi(v)\mathrm{d}v = \frac{\mathrm{d}}{\mathrm{d}\beta} \log \rho(u;\beta) \quad u \in W$$

where integral operator kernel is

$$t(u,v) = \rho(v;\beta)[g(u,v) - 1]$$

and g is so-called pair correlation function.

Pair correlation function

$g(u, v)$ is pair correlation function of **X** if

$$\mathbb{C}\mathrm{ov}[N(A), N(B)] = \mathbb{E}N(A \cap B) + \int_A \int_B \rho(u)\rho(v)[g(u, v) - 1]\mathrm{d}u\mathrm{d}v$$

For Poisson process:

$$g(u, v) = 1$$

so

$$\mathbb{C}\mathrm{ov}[N(A), N(B)] = \mathbb{E}N(A \cap B)$$

Poisson process case

Poisson process case: $g(u,v)=1$ so integral equation simplifies:

$$\phi(u)+\int_W \rho(v;\beta)[g(u,v)-1]\phi(v)\mathrm{d}v=\frac{\mathrm{d}}{\mathrm{d}\beta}\log\rho(u;\beta)\Rightarrow$$

$$\phi(u)=\frac{\mathrm{d}}{\mathrm{d}\beta}\log\rho(u;\beta)=\frac{\rho'(u;\beta)}{\rho(u;\beta)}$$

Hence resulting estimating function is

$$\sum_{u\in\mathbf{X}\cap W}\frac{\rho'(u;\beta)}{\rho(u;\beta)}-\int_W \rho'(u;\beta)\mathrm{d}u$$

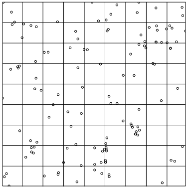
i.e. score of Poisson process log likelihood.

Also applicable if $g(u,v)\neq 1$ (computationally easy but not optimal).

May then be viewed as a *composite likelihood* score.

Quasi-likelihood

Integral equation approximated using Riemann sum dividing W into cells C_i with representative points u_i.



Resulting estimating function is *quasi-likelihood* score

$$DV^{-1}[N - \mu]$$

based on

$$N = (N_1, \dots, N_m)^\mathsf{T}, \quad N_i = 1[\mathbf{X} \text{ has point in } C_i].$$

μ mean of N:

$$\mu_i = \mathbb{E}N_i = \rho(u_i; \beta)|C_i| \text{ and } D = \left[\mathrm{d}\mu(u_l)/\mathrm{d}\beta_i\right]_{il}$$

V covariance of N:

$$V_{ij} = \mathbb{C}\mathrm{ov}[N_i, N_j] = \mu_i 1[i = j] + \mu_i \mu_j [g(u_i, u_j) - 1]$$

Note: back to Gotway and Stroup (1997)'s quasi-likelihood.

QL is consistent with underlying spatial point process model if V_{ij} of the form

$$\mu_i 1[i = j] + \mu_i \mu_j [g(u_i, u_j) - 1]$$

where $g(\cdot, \cdot)$ valid model for point process pair correlation function.

In practice, g must be replaced by preliminary estimate (next talk).

Computation: IGLS

$D(k), V(k), \mu(k)$: D, V, μ at $\beta = \beta^{(k)}$

Iterative generalized least squares: solve

$$D(k)V(k)^{-1}D(k)^{\mathsf{T}}[\beta^{(k+1)} - \beta^{(k)}] = D(k)V(k)^{-1}[Y - \mu(k)]$$

to get new value $\beta^{(k+1)}$ given current $\beta^{(k)}$.

Issue: use fine discretization (large m) $\Rightarrow$ V high dimensional matrix - e.g. V 10000×10000.

Tapering

Structure of $V = \mathbb{C}\mathrm{ov}\, Y$:

$$V = V_\mu^{1/2}[I + G]V_\mu^{1/2}$$

where

$$V_\mu = \mathrm{Diag}(\mu_1, \ldots, \mu_m)$$

and

$$G_{ij} = \mu_i^{1/2}\mu_j^{1/2}[g(u_i, u_j) - 1]$$

Use tapering for G: in IGLS V replaced by

$$V_{\mathrm{taper}} = V_\mu^{1/2}[I + G_{\mathrm{taper}}]V_\mu^{1/2}$$

Efficient computation using sparse matrix Cholesky from `Matrix` library in R.

Application to Capparis

Covariates: elevation (altitude), slope and potassium.

Estimates and standard errors:

	Intcpt.	elevation	slope	potassium
QL	-5.1	2.3	-2.0	4.1
CL	-5.1	2.9	-1.1	4.3
sd QL	0.066	0.80	0.95	1.0
incr. sd CL	1.5%	8.5%	17.1%	7.4%

Slope not significant according to CL.

Simulation study (cluster point process)

Computation depends on grid (discretization) and tapering.

Only estimation:

Grid	ϵ	time (seconds)	estm.
	0.05	1.0	-5.1 2.4 -2.0 4.3
100×50	0.01	2.2	-5.1 2.3 -2.0 4.1
	.002	4.3	-5.1 2.3 -2.0 4.1
	0.05	10.2	-5.1 2.4 -1.9 4.2
150×75	0.01	25.0	-5.1 2.3 -1.9 4.1
	.002	108.2	-5.1 2.3 -1.9 4.0

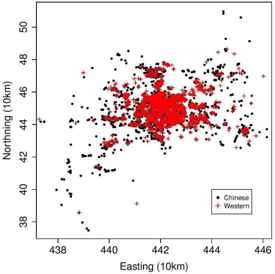
ϵ: level of tapering (smaller means less tapering).

Estimates rather insensitive to choice of grid and tapering.

Further developments

- QL for case-control data/bivariate point processes
- QL for estimation of pair correlation function

Chinese restaurants - case-control QL



Very complicated and heterogeneous intensity functions - depend on city structure (roads, parks, squares...)

Two point processes: **X** (Western) and **Y** (Chinese)

Case-control strategy: study one point process relative to another assuming multiplicative models

$$\rho_{\mathbf{X}}(u) = \rho_{\mathbf{Y}}(u) \exp[\beta z(u)^{\mathsf{T}}]$$

Study how covariate $z(\cdot)$ affects intensity of Western-style restaurants relative to intensity of Chinese restaurants.

Residual measure based on **X** (Western) and **Y** (Chinese):

$$R(A) = \sum_{u \in \mathbf{X}} 1[u \in A] - \sum_{v \in \mathbf{Y}} \exp[\beta z(v)^{\mathsf{T}}] 1[v \in A]$$

Does *not depend* on $\rho_{\mathbf{Y}}$!

Optimal weight function ϕ solution to

$$\phi(u) + \int_W \rho_{\mathbf{Y}}(v) t(u,v) \phi(v) \mathrm{d}v = \frac{z(u)}{1 + \exp[\beta z(u)^{\mathsf{T}}} \quad u \in W$$

where integral operator kernel is

$$t(u,v) = \exp[\beta z(v)^{\mathsf{T}}] \frac{g(u,v) - 1}{1 + \exp[\beta z(u)^{\mathsf{T}}]}$$

Problem: $\rho_{\mathbf{Y}}$ unknown !

Unbiased ('Monte Carlo') estimate of integral:

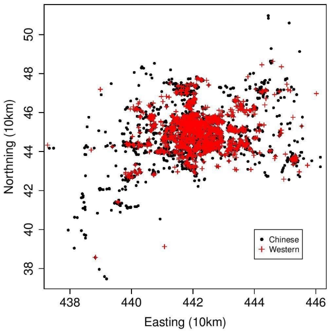
$$\int_W \rho_{\mathbf{Y}}(v)t(u,v)\phi(v)\mathrm{d}v \approx \sum_{v \in \mathbf{Y}} t(u,v)\phi(v)\mathrm{d}v$$

Let $\hat{\phi}(u, \mathbf{Y})$ solution of approximated integral equation.

Resulting approximated QL:

$$\sum_{u \in \mathbf{X}} \hat{\phi}(u, \mathbf{Y}) - \sum_{v \in \mathbf{Y}} \exp[\beta z(v)^{\mathsf{T}}]\hat{\phi}(v, \mathbf{Y} \setminus \{v\})$$

- lots of computational details skipped
- asymptotic results available



Application: district level covariates $z_1(u)$ 'income' and $z_2(u)$ 'tourists'.

'tourists' significant but 'income' not.

Quasi-likelihood for pair correlation function

Second order (zero-mean) residual measure:

$$R(A,B) = \sum_{u,v \in \mathbf{X}}^{\neq} 1[u \in A, v \in B] - \int_A \int_B \rho(u)\rho(v)g(v-u;\psi)\mathrm{d}u\mathrm{d}v$$

Apply weight function:

$$e_f(\psi) = \int_{W^2} f(u,v)R(\mathrm{d}u,\mathrm{d}v) =$$
$$\sum_{u,v \in \mathbf{X}}^{\neq} f(u,v) - \int_{W^2} f(u,v)\rho(u)\rho(v)g(v-u;\psi)\mathrm{d}u\mathrm{d}v$$

Optimal f ?

Can proceed as before but more complicated.

Consider numerical approximation: 'data vector' now consists of all

$$N_{ij} = N_i N_j = 1[\mathbf{X} \text{ has a point in both } C_i \text{ and } C_j]$$

Huge dimension of covariance matrix V for N_{ij}.

Covariances

$$V_{ij,lk} = \mathbb{C}\mathrm{ov}[N_{ij}, N_{lk}] = \mathbb{C}\mathrm{ov}[N_i N_j, N_l N_k]$$

further depend on both third and fourth order point process moments.

Simplifying assumption:

$$f(u, v) = f_0(v - u)$$

and point process stationary $\Rightarrow$ constant intensity ρ and $g(u, v) = g(v - u)$

Further: use only 'close' pairs u, v with $\|v - u\| \leq R$

This leads to simplified estimating function

$$\sum_{\substack{w \in \mathbf{X}^d: \\ \|w\| \leq R}} f_0(w) - \int_{b(0,R)} f_0(w) \rho^d(w) \mathrm{d}w$$

where $\mathbf{X}^d$ point process of differences $v - u$, $v \neq u \in \mathbf{X}$.

$\mathbf{X}^d$ has intensity function

$$\rho^d(w) = \rho^2 g(w) |W \cap W_w|.$$

Thus similar to previous QL for intensity function !

- **Result**: simplified version asymptotically optimal.
- Numerical approximation: 'data' vector now consists of

$$N_i = 1[\mathbf{X}^d \cap C_i \neq 0]$$

  where $b(0, R) = \cup_i C_i$.

  Covariance matrix of N_i's evaluated using simulation.
- Extension to case of nonconstant intensity possible.
- Choice of R: as large as possible (given computational ressources).

  Results quickly stabilize as R grows.

References

1. Deng, C., Guan, Y., Waagepetersen, R. and Zhang, J. (2015) Second-order quasi-likelihood for spatial point processes, *Biometrics*, to appear.
2. Gang, X., Waagepetersen, R. and Guan, Y. (2016) Stochastic Quasi-likelihood for Case-Control Point Pattern Data, submitted.
3. Guan, Y., Jalilian, A. and Waagepetersen, R. (2015) Quasi-likelihood for spatial point processes, *Journal of the Royal Statistical Society, Series B*, 77, 677-697.