Optimal first order estimating equations

Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

March 18, 2012

joint work

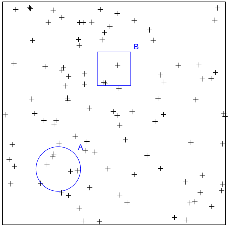
with Yongtao Guan and Abdollah Jalilian

Mean and covariances of counts for spatial point process

Point process $\mathbf{X}$: random point pattern.

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

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



$\rho(\cdot)$: intensity function.

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

$= $ Poisson proc. covariance $+$ extra term due to corr.

$g(u,v)$: pair correlation function

Campbell formulae

$$N(A) = \sum_{u \in \mathbf{X}} 1[u \in A]$$

$$N(A)N(B) = \sum_{u,v \in \mathbf{X}} 1[u \in A, u \in B]$$

Hence by moment formulae, for f function on $\mathbb{R}^2$ or $\mathbb{R}^2 \times \mathbb{R}^2$:

$$\mathbb{E} \sum_{u \in \mathbf{X}} f(u) = \int_{\mathbb{R}^2} f(u)\rho(u)\mathrm{d}u$$

$$\mathbb{E} \sum_{u,v \in \mathbf{X}}^{\neq} f(u,v) = \int_{\mathbb{R}^2 \times \mathbb{R}^2} f(u,v)\rho(u)\rho(v)g(u,v)\mathrm{d}u\mathrm{d}v$$

Starting point for unbiased estimating functions !

Regression model for intensity function

Focus on estimation of parameter in regression model for intensity function.

E.g. log-linear model

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

where

$$Z(u) = (Z_1(u), \ldots, Z_p(u))$$

First-order estimating equations

Campbell $\Rightarrow$ unbiased *first-order* estimating function

$$u_f(\beta) = \sum_{u \in \mathbf{X} \cap W} f_\beta(u) - \int_W f_\beta(u)\rho_\beta(u)\mathrm{d}u$$

Choice

$$f_\beta(u) = \frac{\mathrm{d}}{\mathrm{d}\beta} \log \rho_\beta(u)$$

leads to composite likelihood/Poisson likelihood

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

This is optimal choice for Poisson process (MLE) but what is optimal in the clustered case ?

Asymptotic results - first order estimating function

Let sensitivity

$$S_f = -\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\beta^\mathsf{T}} u_f(\beta)/|W|$$

and

$$\Sigma_f = \frac{\mathbb{V}\mathrm{ar}\, u_f(\beta)}{|W|}$$

Under appropriate mixing conditions,

$$\hat{\beta}_f \approx N(\beta, V_f/|W|)$$

where

$$V_f = S_f^{-1}\Sigma_f S_f^{-1}$$

Optimal first-order estimating equation

Optimal choice of f_β: smallest asymptotic variance

$$V_f = S_f^{-1} \Sigma_f S_f^{-1}$$

.

Optimal choice of f_β is solution of Fredholm equation

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

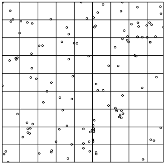
where integral equation kernel is

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

Note: optimal f_β depends on pair correlation !

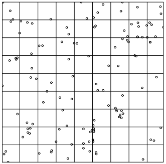
Numerical approximation and quasi-likelihood

Approximate solution of Fredholm equation using numerical quadrature: Riemann sum dividing W into cells C_i with representative points u_i.



Numerical approximation and quasi-likelihood

Approximate solution of Fredholm equation using numerical quadrature: Riemann sum dividing W into cells C_i with representative points u_i.



Resulting estimating function is *quasi-likelihood*

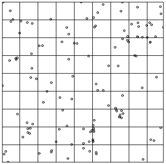
$$(N - \mu)V^{-1}D$$

based on

$$N = (N_1, \ldots, N_m), \quad N_i \text{ count of points in } C_i.$$

Numerical approximation and quasi-likelihood

Approximate solution of Fredholm equation using numerical quadrature: Riemann sum dividing W into cells C_i with representative points u_i.



Resulting estimating function is *quasi-likelihood*

$$(N-\mu)V^{-1}D$$

based on

$$N=(N_1,\ldots,N_m),\quad N_i \text{ count of points in } C_i.$$

μ mean of N:

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

V covariance of N:

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

Practical implementation: IGLS

Pair correlation function inside V estimated by e.g. minimum contrast.

Solve

$$(N - \mu(\beta)V(\beta)^{-1}D(\beta) = 0$$

using iterative generalized least squares:

$$(\beta^{(l+1)} - \beta^{(l)})D(\beta^{(l)})^\mathsf{T} V(\beta^{(l)})^{-1} D(\beta^{(l)}) = (N - \mu(\beta^{(l)}))V(\beta^{(l)})^{-1} D(\beta^{(l)})$$

One issue: use fine discretization (large m) $\Rightarrow$ V highdimensional matrix - e.g. V 10000×10000.

Use tapering and sparse matrix Cholesky from `Matrix` library in R.

Covariance matrix for $\hat{\beta}$:

$$S_{\text{taper}}^{-1} D^\mathsf{T} V_{\text{taper}}^{-1} V V_{\text{taper}}^{-1} D S_{\text{taper}}^{-1}, \quad S_{\text{taper}} = D^\mathsf{T} V_{\text{taper}}^{-1} D$$

Simulation study

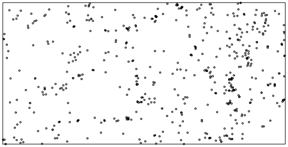
Consider variance of $\hat{\beta}$ obtained from either composite likelihood or quasi-likelihood.

Reduction in variance for quasi-likelihood relative to composite likelihood: 10% to 65%.

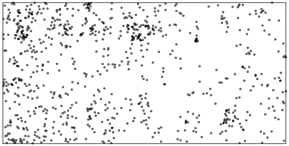
Large reductions when strong clustering and strong inhomogeneity.

Example: three tree species with different modes of seed dispersal:

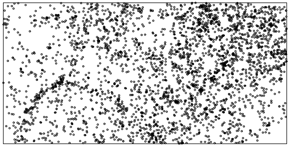
Acalypha Diversifolia

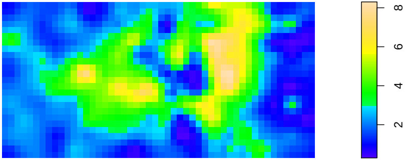


Loncocharpus Heptaphyllus



Capparis Frondosa

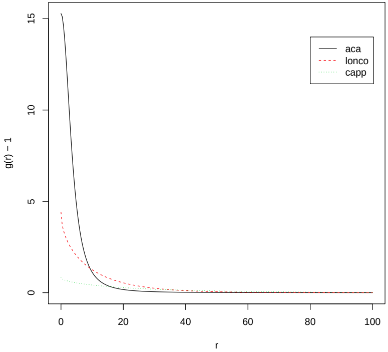




Potassium content in soil.

Covariates pH, elevation, gradient, potassium,...

Fitted pair correlation functions $g(\cdot)$



Acalypha: Cauchy.

Loncocharpus, Capparis: Matérn.

Results with composite likelihood and quasi-likelihood

species		$\widehat{\beta}$
Acalypha	CL	$-6.91 + 0.021\texttt{dem} + 0.0047\texttt{K}$ $(77.34^*, 9.77^*, 1.153^*) \times 10^{-3}$
	QL	$-6.90 + 0.016\texttt{dem} + 0.0047\texttt{K}$ $(77.09^*, 9.54, 1.133^*) \times 10^{-3}$
Loncocharpus	CL	$-6.49 - 0.021\texttt{Nmin} - 0.11\texttt{P} - 0.59\texttt{pH} - 0.11\texttt{twi}$ $(81.06^*, 7.45^*, 58.78, 282.89^*, 53.19^*) \times 10^{-3}$
	QL	$-6.49 - 0.023\texttt{Nmin} - 0.12\texttt{P} - 0.55\texttt{pH} - 0.084\texttt{twi}$ $(80.15^*, 6.95^*, 55.23^*, 266.10^*, 45.47) \times 10^{-3}$
Capparis	CL	$-5.07 + 0.028\texttt{dem} - 1.10\texttt{grad} + 0.0043\texttt{K}$ $(79.54^*, 9.98^*, 1200.36, 1.16^*) \times 10^{-3}$
	QL	$-5.10 + 0.019\texttt{dem} - 2.50\texttt{grad} + 0.0039\texttt{K}$ $(77.77^*, 8.86^*, 935.02^*, 1.02^*) \times 10^{-3}$

Estimated standard errors always smallest for QL. Regression parameters similar except for grad, Capparis.

Thanks for your attention