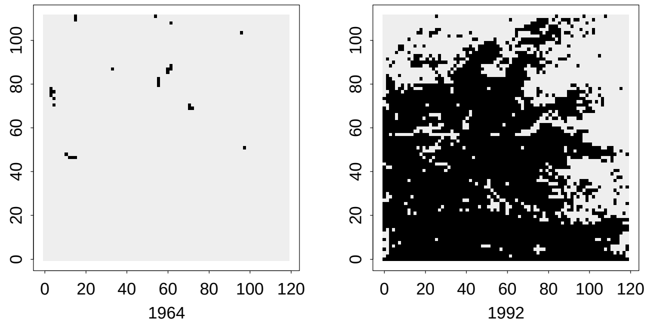
An Alternative to Composite Likelihood for Analyzing Large Sets of Spatially Correlated Binary Data

Samuel D. Oman
Dept of Statistics, Hebrew University of Jerusalem

Joint with

- Victoria Landsman
  Dept of Statistics, Hebrew Univ
- Yohay Carmel
  Faculty of Agricultural Engineering, Technion
- Ronen Kadmon
  Dept of Evolution, Systematics and Ecology, Hebrew Univ

The Problem



- Research area: 1,125 $\times$ 1,200 m^2 region of natural forest in Mt. Meron
- $75 \times 80 = 6{,}000$ pixels (sites) of size $15m \times 15m$
- Black: trees (woody vegetation more than $2.5m$ high) are the predominant growth form at site
- **Goal**: identifying factors affecting changes in tree cover

Variables

- Spatially dependent binary responses

$$Y_i = \begin{cases} 1 & \text{if trees dominant in 1992 at site } i \\ 0 & \text{if not} \end{cases}$$

  $i = 1, \ldots, 6{,}000$

- x_i = vector of covariates:
  - Tree dominance in 1964
  - Environmental factors - height above sea level, aspect & slope (which affect solar radiation), ...
  - Anthropogenic factors - grazing type and intensity, distance to nearest road, ...

Model: Hierarchical GLIM with Probit Link

- $Z^* = (Z_i^*)$: latent, spatially dependent Gaussian field
- $\{Y_i\}|Z^* \sim \text{Ber}(p_i^*)$ independently,

  $p_i^* = \Phi(x_i^t\gamma + Z_i^*) = P(Z_i \leq x_i^t\gamma + Z_i^*)$

- Integrating out Z^*:

$$p_i \equiv P(Y_i = 1) = P(Z_i \leq x_i^t\beta)$$

$$\pi_{ij} \equiv P(Y_i = 1, Y_j = 1) = P(Z_i \leq x_i^t\beta, Z_j \leq x_j^t\beta)$$

  where

$$Z_i \sim N(0, 1)$$

$$\text{cor}(Z_i, Z_j) = \sigma^2\rho^{\|i-j\|} \qquad (\sigma^2 \leq 1 \text{ for possible nugget effect})$$

$$\|i - j\| = \text{distance between sites } s_i \text{ and } s_j$$

Estimation: Composite Likelihood (Heagerty and Lele, 1998 JASA)

- $\theta = (\beta^t, \rho, \sigma^2)^t$: parameter vector
- loglik $\approx \sum_{\|i-j\|\le d} \log p(y_i, y_j) \equiv \ell$
- $\nabla \ell \equiv U = \sum_{\|i-j\|\le d} U_{ij}(\theta)$, relatively simple U_{ij}
- $U(\hat\theta) = 0 \Rightarrow \hat\theta \approx N(\theta, V)$, $\qquad V = H^{-1}\text{cov}(U)H^{-t} \qquad (H = E[(\partial/\partial\theta)U])$

  (assuming regularity; based on Guyon 1995, Lindsay 1988, White 1982)
- "Sandwich Estimator" of V: use window sub-sampling (Sherman, 1996):

  assuming $\quad | R | \text{cov}(\mathbf{U}_R) \to \Sigma \quad$ as region R expands,

  $\widehat{\text{cov}}(U_R) = J^{-1} \sum_j \frac{|R|}{|R_j|} [U_{R_j}(\hat\theta)][U_{R_j}(\hat\theta)]^t$ for J subwindows $\{R_j\}$ of R
- Like GEE2

BUT it took a lonnnnnng time

- At each Newton - Raphson iteration, computing

  $U = \sum_{\|i-j\|<d} U_{ij}(\theta) \qquad$ and $\qquad \nabla U$

  required $\approx 10^6$ evaluations of

  $\pi_{ij} = P(Z_i \le x_i^t\beta, Z_j \le x_j^t\beta)$

  for the correlated Normal Z_i and Z_j.
- S-plus routines $\rightarrow$ C-plus: several hours

Alternative: **I**ndependent **B**lock **E**stimating **E**quations (**IBEE**)

- Let $Y_{N\times 1} = (Y_i), p = (p_i), V = \text{cov}(Y)$
- Want $\hat{\beta} : D^t V^{-1}[Y - p(\beta)] = 0$ where $D = (\partial p)/(\partial \beta)$, but $N = 6{,}000$
- Instead:
  - $V \rightarrow$ "Working" covariance matrix $W = \text{diag}(W_{(i)})$ (independent blocks)
  - $\text{cov}(Y_j, Y_k) \approx \phi(x_j^t\beta)\phi(x_k^t\beta)\,\text{arcsin}(\sigma^2\rho^{\|j-k\|})$ (Pearson, 1901)
  - $\hat{\sigma}^2, \hat{\rho}$ from separate estimating equation based on $\{(Y_j - Y_k)^2\}$
- Result: $\hat{\beta}$ as solution to $U(\beta) \equiv \sum_{\text{blocks } i} G_{(i)}(Y_{(i)} - p_{(i)}) = 0$
  - Estimation of $\text{cov}(\hat{\beta})$, asymptotics: as in Haegerty and Lele (1998).
  - Like IEE, but with independent *blocks*.

Application to Long-Term Vegetation Growth Data

- 75×80 lattice $\rightarrow$ 15×16 - sized blocks; took several minutes
- "Ind" is independence estimator of Haegerty and Lumley (2000)
- Total of 14 explanatory variables

	Coefficient estimate			Standard error			p-value		
Variable	Ind	Cl	Ibee	Ind	Cl	Ibee	Ind	Cl	Ibee
tree.64	1.16	1.35	0.55	0.27	0.27	0.27	0.00	0.00	0.05
alt	2.12	2.20	1.65	0.53	0.38	0.40	0.00	0.00	0.00
rd.dist	0.96	0.85	0.48	0.60	0.49	0.44	0.11	0.09	0.27
g.spor	-0.57	-0.61	-0.06	0.27	0.29	0.43	0.04	0.04	0.89
g.mod	-1.00	-0.96	-0.77	0.32	0.33	0.24	0.00	0.00	0.00
c.spor	0.53	**0.54**	-0.23	0.27	0.28	0.19	0.05	**0.05**	0.23
c.mod	-0.94	-0.92	-0.61	0.28	0.25	0.25	0.00	0.00	0.02
c.heavy	-1.66	-1.70	-0.79	0.44	0.39	0.23	0.00	0.00	0.00

- In general, $\hat{\beta}_i^{\text{ind}} \approx \hat{\beta}_i^{\text{cl}} \neq \hat{\beta}_i^{\text{ibee}}$
- $\text{avg}\{\text{s.e.}(\hat{\beta}_i^{\text{cl}})/\text{s.e.}(\hat{\beta}_i^{\text{ind}})\} = 0.95$; $\text{avg}\{\text{s.e.}(\hat{\beta}_i^{\text{ibee}})/\text{s.e.}(\hat{\beta}_i^{\text{ind}})\} = 0.73$

Asymptotic Efficiencies (one explanatory variable, $N = 24 \times 24$)

- $W = \text{diag}(p_i[1-p_i])^{1/2}, \quad u = \frac{\partial}{\partial\beta}p, \quad R = \text{cor}(Y)$
- $\widehat{\beta}$ solution to $\quad c^t W^{-1}[Y - p(\beta)] = 0 \quad$ for $\quad c \in \Re^N \quad \Rightarrow$

  (asymptotic) $\quad \text{var}(\widehat{\beta}) = c^t Rc/(c^t u)^2$
- $\widehat{\beta}_{\text{opt}} : c = R^{-1}u, \qquad \text{var(opt)} = 1/(u^t R^{-1} u)$
- $\widehat{\beta}_{\text{ind}} : c = u, \qquad \text{var(ind)} = u^t Ru/(u^t u)^2$
- $\widehat{\beta}_{\text{cl}}, \widehat{\beta}_{\text{ibee}} : c = \ldots, \qquad \text{var} = \#$
- Compute eff(est) = var(opt)/var(est), especially when x : eff(ind) is low:

  $\text{eff(ind)} = (u^t u)^2/[(u^t Ru)(u^t R^{-1}u)] \approx \lambda_{\text{min}}/\lambda_{\text{max}} \quad \{\lambda_i = \text{eig}(R)\}$

  if $\quad \lambda_{\text{max}} \gg \lambda_{\text{min}} \quad$ and $\quad u = (v_{\text{max}} + v_{\text{min}}) \equiv \bar{v}$

Numerical Results

$\beta = 1, \text{cov}(Y)$ given (exactly) by Pearson's approximation
C = condition number of R when $\beta = 0$.
$\mathbf{1}$: vector of ones; $\bar{\mathbf{v}}$: unfavorable vector; $\mathbf{r}$: pseudo-random numbers.

Cov matrix				Ind	Ibee(k)			Cl(d)		
σ^2	ρ	C	x		2	3	4	1	2	3
			$\mathbf{1}$	0.81	0.81	0.81	0.81	0.79	0.78	0.77
0.8	0.90	225	$\bar{\mathbf{v}}$	0.05	0.27	0.53	0.75	**0.15**	**0.08**	**0.09**
			$\mathbf{r}$	0.85	0.92	0.96	0.97	0.92	0.92	0.89
			$\mathbf{1}$	0.92	0.92	0.92	0.92	0.90	0.88	0.86
0.8	0.61	21	$\bar{\mathbf{v}}$	0.32	0.69	0.84	0.91	0.52	0.40	0.41
			$\mathbf{r}$	0.86	0.93	0.97	0.98	0.92	0.90	0.88
			$\mathbf{1}$	0.98	0.98	0.98	0.97	0.96	0.94	0.93
0.8	0.37	5	$\bar{\mathbf{v}}$	0.70	0.90	0.95	0.97	0.82	0.75	0.74
			$\mathbf{r}$	0.94	0.97	0.98	0.99	0.96	0.94	0.92

- eff(cl(d)) tends to *decrease* with d (Nott and Ryden, 1999)

Remarks

IBEE: In $D^t V^{-1}[Y - p(\beta)] = 0$, $V \to$ "Working" block-diagonal W

- Although use working cov(Y), correlations are modeled on *latent* scale:
  - cov$(Y_j, Y_k) = \phi(x_j^t\beta)\phi(x_k^t\beta)$ arcsin$(\sigma^2\rho^{\|j-k\|})$, not $\sigma^2\rho^{\|j-k\|}$
  - Preliminary analysis: compute binned sample cov(Y_j, Y_k), invert Pearson approx, fit covariogram on latent scale
- IBEE seems to "work" for exponential autocorrelation structure

  Possible insight: time series

  - cov$(Y) = V$ of AR(1) structure
  - V^{-1} is tri-diagonal $\approx$ block-diagonal
- IBEE "tuning parameters": block size; max distance for σ^2, ρ est eqns; subsampling window size