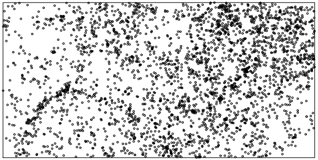
Decomposition of variance for spatial Cox processes

Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

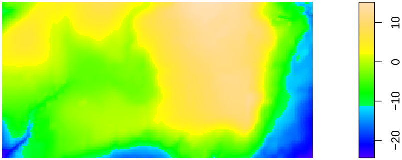
Joint work with Abdollah Jalilian and Yongtao Guan

December 13, 2010

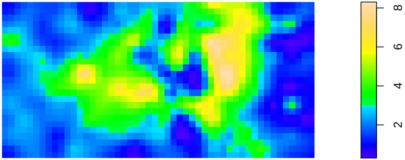
Tropical rain forest example: *Capparis Frondosa*



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



Elevation



Potassium content in soil.

How much variation due to environmental variables and how much due to seed dispersal ?

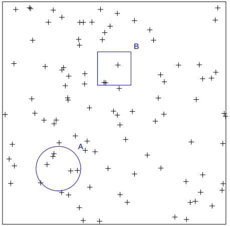
Framework: spatial Cox point processes.

Poisson and Cox processes

Poisson process with intensity function $\rho(\cdot)$:

counts $N(B)$ independent
and Poisson distributed with mean

$$\mathbb{E}N(B) = \int_B \rho(u)\mathrm{d}u$$



Cox process $\mathbf{X}$ with random intensity function Λ: conditional on $\Lambda = \lambda$, $\mathbf{X}$ Poisson process with intensity function λ.

Additive model for Λ

Additive model:

$$\Lambda(u) = \tilde{Z}(u) + \Lambda_0(u)$$

$\tilde{Z}$ contribution from environment:

$$\tilde{Z}(u) = \beta Z(u)^{\mathsf{T}} \quad Z(u) = (Z_1(u), \ldots, Z_p(u))$$

Λ_0: structured variation (e.g. seed dispersal) not due to environment.

Cox process **X** union of independent Cox processes with random intensity functions $\tilde{Z}(u)$ and $\Lambda_0(u)$.

Assume $\tilde{Z}$ and Λ_0 non-negative and stationary.

Log-linear model

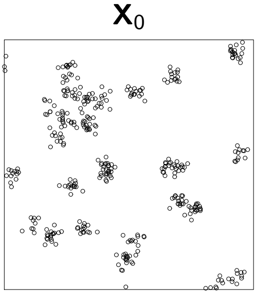
$$\Lambda(u) = \Lambda_0(u) \exp[\tilde{Z}(u)]$$

Interpretation in terms of survival of seedlings:

$\mathbf{X}_0$ seedlings: stationary Cox process with random intensity function Λ_0.

$\mathbf{X}$ thinning of $\mathbf{X}_0$ with survival depending on environment $\tilde{Z}$.

Example: Cox/cluster process: Inhomogeneous Thomas process

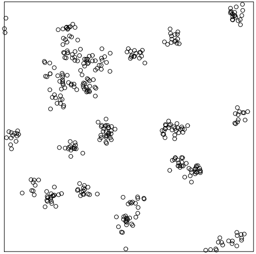


Λ_0 shot-noise process $\Rightarrow$ $\mathbf{X}_0$ cluster process:

Offspring distributed around Poisson parents mothers according to Gaussian density

Example: Cox/cluster process: Inhomogeneous Thomas process

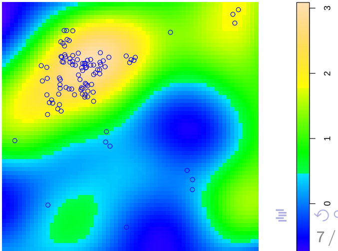
$\mathbf{X}_0$



Λ_0 shot-noise process $\Rightarrow$ $\mathbf{X}_0$ cluster process:

Offspring distributed around Poisson parents mothers according to Gaussian density

$\mathbf{X}$



Inhomogeneity: offspring in $\mathbf{X}_0$ survive according to probability

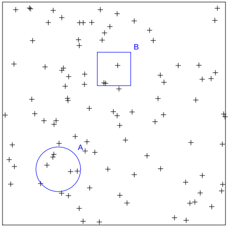
$$p(u) \propto \exp[\beta Z(u)^{\mathsf{T}}]$$

depending on covariates (independent thinning).

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

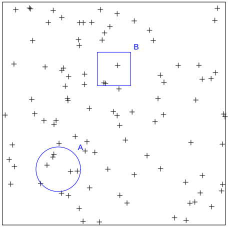
$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$



Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$



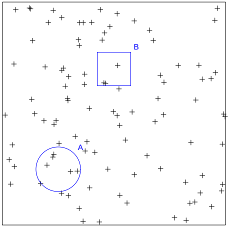
$\mathbb{V}\mathrm{ar} N(B) =$

variation $=$

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$



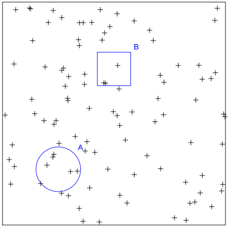
$$\mathbb{V}\mathrm{ar} N(B) = \mathbb{V}\mathrm{ar} \hat{N}(B)$$

variation = variation Λ

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$



$$\mathbb{V}\mathrm{ar} N(B) = \mathbb{V}\mathrm{ar}\hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

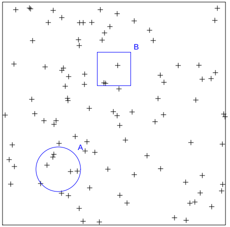
variation = variation Λ + 'Poisson noise'

Further, $\hat{\Lambda}(u) = \mathbb{E}[\Lambda(u)|Z]$

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$



$$\mathbb{V}\mathrm{ar}N(B) = \mathbb{V}\mathrm{ar}\hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

variation $=$ variation $\Lambda +$ 'Poisson noise'

Further, $\hat{\Lambda}(u) = \mathbb{E}[\Lambda(u)|Z]$

$\mathbb{V}\mathrm{ar}\Lambda(u) =$

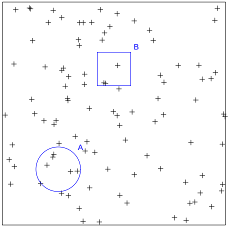
structured variation $=$

SST $=$

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$



$$\mathbb{V}\mathrm{ar} N(B) = \mathbb{V}\mathrm{ar}\hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

variation = variation Λ + 'Poisson noise'

Further, $\hat{\Lambda}(u) = \mathbb{E}[\Lambda(u)|Z]$

$$\mathbb{V}\mathrm{ar}\Lambda(u) = \mathbb{V}\mathrm{ar}\hat{\Lambda}(u)$$

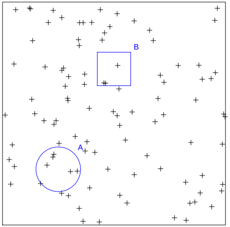
structured variation = variation due to environment

SST = SSR

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$



$$\mathbb{V}\mathrm{ar} N(B) = \mathbb{V}\mathrm{ar} \hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

variation = variation Λ + 'Poisson noise'

Further, $\hat{\Lambda}(u) = \mathbb{E}[\Lambda(u)|Z]$

$$\mathbb{V}\mathrm{ar}\Lambda(u) = \mathbb{V}\mathrm{ar}\hat{\Lambda}(u) + \mathbb{V}\mathrm{ar}\left[\Lambda(u) - \hat{\Lambda}(u)\right]$$

structured variation = variation due to environment + other sources

SST = SSR + SSE

Measure of influence of environmental covariates Z:

$$R^2 = \frac{SSR}{SST} = \frac{\mathbb{V}\mathrm{ar}\mathbb{E}[\Lambda(u)|Z]}{\mathbb{V}\mathrm{ar}\Lambda(u)}$$

(right hand side does not depend on u in case of stationary environment)

R^2 for additive and log-linear models

Additive:

$$R^2 = \frac{\sigma^2_{\tilde{Z}}}{\sigma^2_{\tilde{Z}} + \sigma^2_0}$$

$$\sigma^2_{\tilde{Z}} = \mathbb{V}\mathrm{ar} \tilde{Z}(u) \quad \text{and } \sigma^2_0 = \mathbb{V}\mathrm{ar} \Lambda_0(u)$$

Log-linear:

$$R^2 = \frac{\sigma^2_{\exp \tilde{Z}}}{\sigma^2_{\exp \tilde{Z}} + \sigma^2_0 \left[\sigma^2_{\exp \tilde{Z}} + \rho^2_{\exp \tilde{Z}}\right]}$$

$$\sigma^2_{\exp \tilde{Z}} = \mathbb{V}\mathrm{ar} \exp[\tilde{Z}(u)] \quad \text{and } \rho_{\exp \tilde{Z}} = \mathbb{E} \exp[\tilde{Z}(u)]$$

Estimation: environmental variances

Z observed on grid $G = \{u_i\}_{i=1,\ldots,M}$

$$\widehat{\sigma}^2_{\widetilde{Z}} = \frac{1}{M} \sum_{u \in G} \widehat{\widetilde{Z}}(u)^2$$

and

$$\widehat{\sigma}^2_{\exp \widetilde{Z}} = \frac{1}{M} \sum_{u \in G} \left\{ \exp\left[\widehat{\widetilde{Z}}(u)\right] - \widehat{\rho}_{\exp \widetilde{Z}} \right\}^2$$

where

$$\widehat{\rho}_{\exp \widetilde{Z}} = \frac{1}{M} \sum_{u \in G} \exp\left[\widehat{\widetilde{Z}}(u)\right]$$

and $\widehat{\widetilde{Z}}(u) = \widehat{\beta} Z(u)^{\mathsf{T}}$.

Estimation: β

Estimate β conditioning on Z.

First-order log composite likelihood:

$$\mathrm{CL}_1(\beta) = \sum_{u \in \mathbf{X}} \log \rho(u|Z;\beta) - \int_W \rho(u|Z;\beta)\mathrm{d}u$$

$\rho(\cdot|Z,\beta) = \mathbb{E}[\Lambda(u)|Z]$ intensity function for $\mathbf{X}|Z$.

For additive model:

$$\rho(u|Z,\beta) = \rho_0 + \beta Z(u)^{\mathsf{T}} \quad \rho_0 = \mathbb{E}\Lambda_0(u)$$

Parametric models for $c_0(u-v) = \mathbb{C}\mathrm{ov}[\Lambda_0(u), \Lambda_0(v)]$

Any positive definite function is a covariance function but not necessarily for a non-negative random process Λ_0. Use covariance functions from explicit constructions of Λ_0.

Log-Gaussian:

$$\Lambda_0(u) = \exp[Y(u)] \quad c_0(u-v) = \rho_0^2 \{ \exp[\mathbb{C}\mathrm{ov}(Y(u), Y(v))] - 1 \}$$

where Y Gaussian field and $\rho_0 = \mathbb{E}\Lambda_0(u)$.

Shot-noise:

$$\Lambda_0(u) = \sum_{v \in C} \alpha k(u-v) \quad c_0(u-v) = \kappa\alpha^2 \int_{\mathbb{R}^2} k(w)k(w+v-u)\mathrm{d}w$$

where C homogeneous Poisson with intensity κ and $k(\cdot)$ probability density.

Bessel shot-noise/Matérn covariance

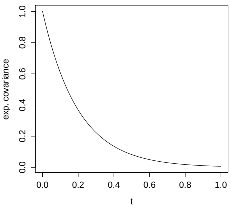
Λ_0 Bessel shot-noise process: sum of $K_{(\nu-1)/2}$ Bessel densities centered around points of homogeneous Poisson process.

Matérn covariance function:

$$c_0(h) = \sigma_0^2 \frac{(\|h\|/\eta)^\nu K_\nu(\|h\|/\eta)}{2^{\nu-1}\Gamma(\nu)}$$

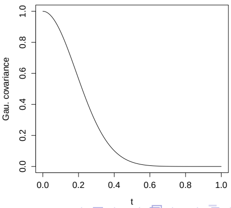
$\nu = 1/2$: exponential model

$$c_0(h) = \sigma_0^2 \exp(-\|h\|/\eta)$$



'$\nu = \infty$':'Gaussian'

$$c_0(h) = \sigma_0^2 \exp[-(\|h\|/\eta)^2]$$



Estimation of ψ

Second-order log composite likelihood (given $\hat{\beta}$, conditioning on Z):

$$\mathrm{CL}_2(\psi|\hat{\beta}) = \sum_{\substack{u,v\in\mathbf{X} \\ \|u-v\|\le R}}^{\neq} \log \rho^{(2)}(u,v|Z;\hat{\beta},\psi) - \iint_{\|u-v\|\le R} \rho^{(2)}(u,v|Z;\hat{\beta},\psi)\mathrm{d}u\mathrm{d}v$$

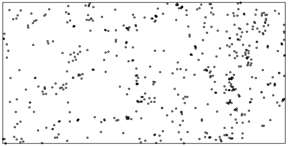
$\rho^{(2)}(u,v|Z;\beta,\psi) = \mathbb{E}[\Lambda(u)\Lambda(v)|Z]$ second-order product density for $\mathbf{X}|Z$.

For additive model:

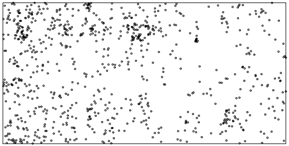
$$\rho^{(2)}(u,v|Z;\beta,\psi) = \rho(u|Z,\beta)\rho(v|Z,\beta) + c_0(u-v;\psi)$$

Three species with different modes of seed dispersal:

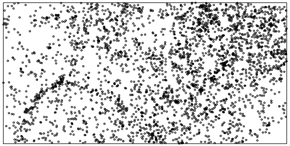
Acalypha Diversifolia explosive capsules



Loncocharpus Heptaphyllus wind



Capparis Frondosa bird/mammal

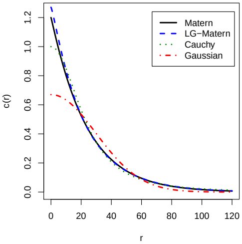


Results for rain forest data

Species	model for Λ	c_0	$CL_2(\widehat{\psi}\vert\widehat{\beta})$	R^2
Acalypha	log-linear	'Gaussian'	-1239.8	0.01
		Matérn	-1221.0	0.02
	additive	'Gaussian'	-1641.7	0.01
		Matérn	-1623.4	0.01
Loncocharpus	log-linear	'Gaussian'	-3204	0.17
		Matérn	-3156	0.10
	additive	'Gaussian'	-4081	0.11
		Matérn	-4055	0.07
Capparis	log-linear	'Gaussian'	-76657	0.37
		Matérn	-76325	0.19
	additive	'Gaussian'	-81139	0.23
		Matérn	-80685	0.16

Some conclusions

Covariance functions for loncocharpus



Best fit with Matérn (heavy tails for covariance/cluster density).

Best fit with log-linear model (interpretation in terms of survival).

Largest R^2 for Capparis (bird/mammal seed dispersal), smallest for Acalypha (explosive capsules).

Advertisement

Ph.D.-scholarship in Spatial Statistics

Department of Mathematics, Aalborg University, Denmark

Info: Rasmus Waagepetersen (`rw@math.aau.dk`), Jesper Møller (`jm@math.aau.dk`)