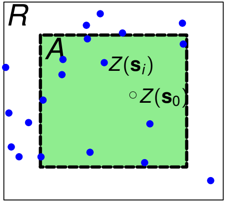
Estimating Abundance from Counts in Large Data Sets of Irregularly-Spaced Plots using Spatial Basis Functions

Jay Ver Hoef

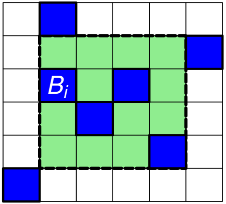
NOAA National Marine Mammal Lab
NOAA Fisheries
International Arctic Research Center
Fairbanks, Alaska, USA

Introduction

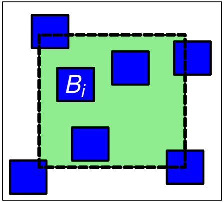
1) Block Kriging



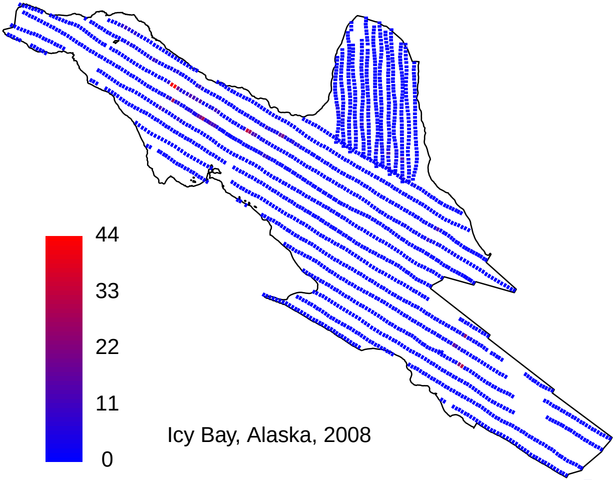
2)Block Prediction for Finite Populations on a Grid



3)Block Prediction for Finite Populations Irregularly Spaced



Motivating Example



Goals

An estimator that is:

- fast to compute, robust, and requires few modeling decisions, similar to classical survey methods,
- based only on counts within plots; actual spatial locations of animals are unknown,
- for the actual number of seals, not the mean of some assumed process that generated the data,
- have a variance estimator with a population correction factor that shrinks to zero as the proportion of the study area that gets sampled goes to one,
- unbiased with valid confidence intervals,
- able to accommodate nonstationary variance throughout the area

Inhomogeneous Spatial Point Processes

$T(V)$ is the total number of points in planar region V

$$\lambda(\mathbf{s}) = \lim_{|dx| \to 0} \frac{E(T(dx))}{|dx|}$$

Expected abundance in $A \subseteq R$:

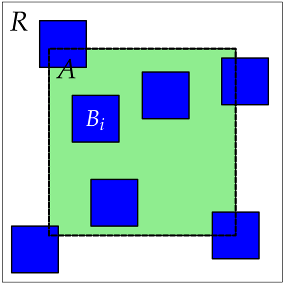
$$\mu(A) = \int_A \lambda(\mathbf{u}|\boldsymbol{\theta}) d\mathbf{u}$$

Abundance is assumed random

$$T(A) \sim \mathrm{Poi}(\mu(A))$$

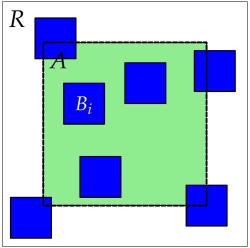
Resulting in an observed pattern $\mathcal{S}^+ = (\mathbf{s}_1, \ldots, \mathbf{s}_N)$

Outline of an Estimator



- $\mathcal{B} = \cup_{i=1}^{n}(B_i \cap A)$
- $\mathcal{U} \equiv \overline{\mathcal{B}} \cap A$
- $T(A) = T(\mathcal{B}) + T(\mathcal{U})$
- $T(\mathcal{U}) \sim \text{Poi}(\mu(\mathcal{U}))$
- $\mu(\mathcal{U}) = \int_{\mathcal{U}} \lambda(\mathbf{u}|\boldsymbol{\theta})d\mathbf{u}$
- $\widehat{T}(A) = T(\mathcal{B}) + \widehat{T}(\mathcal{U})$
- $T(\mathcal{B}) \rightarrow T(A) \Rightarrow \widehat{T}(A) \rightarrow T(A)$

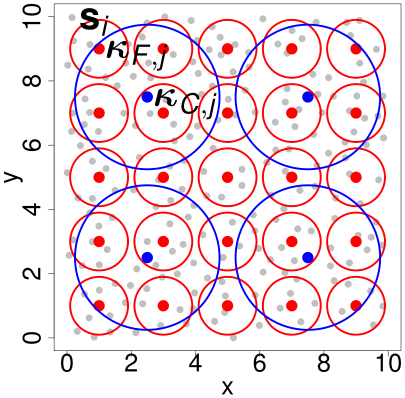
From IPP to Poisson Regression



- $Y(B_i) \sim \text{Poi}(\mu(B_i))$
- $\mu(B_i) = \int_{B_i} \lambda(\mathbf{u}|\boldsymbol{\theta})d\mathbf{u}$
- Let $\mathbf{s}_i$ be centroid of B_i
- $\mu(B_i) \approx |B_i|\lambda(\mathbf{s}_i|\boldsymbol{\theta})$
- $\log(\mu(B_i)) = \log(|B_i|) + \log(\lambda(\mathbf{s}_i|\boldsymbol{\theta}))$
- $\log(\lambda(\mathbf{s}_i|\boldsymbol{\theta})) = \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta}$

Now us spatial basis functions to generate $\mathbf{x}(\mathbf{s}_i)$

Spatial Basis Functions



- $C(h;\rho) = \exp(-h^2/\rho)$
- $\mathbf{X}_{i,j} = C(\|\mathbf{s}_i - \boldsymbol{\kappa}_{F,j})\|;\rho_F)$; $j = 2,\ldots,K_F+1$
- $\mathbf{X}_{i,j} = C(\|\mathbf{s}_i - \boldsymbol{\kappa}_{C,j})\|;\rho_C)$; $j = K_F+2,\ldots,K_F+K_C+1$

knot location: k-means clustering of dense grid of spatial coordinates

Fitting the Model

minimize minus the log-likelihood:

$$-\ell(\boldsymbol{\rho}, \boldsymbol{\beta}; \mathbf{y}) \propto \sum_{i=1}^{n} |B_i| \exp(\mathbf{x}_{\boldsymbol{\rho}}(\mathbf{s}_i)'\boldsymbol{\beta}) - y_i \log|B_i| - y_i \mathbf{x}_{\boldsymbol{\rho}}(\mathbf{s}_i)'\boldsymbol{\beta}$$

Two-part algorithm:

- Condition on ρ and use IWLS to estimate β (with offset for $|B_i|$, ala GLMs)
- optimize for ρ numerically

Back to the Estimator

- $\widehat{T}(A) = T(\mathcal{B}) + \widehat{T}(\mathcal{U})$
- $\widehat{T}(\mathcal{U}) = \mu(\mathcal{U}) = \int_{\mathcal{U}} \lambda(\mathbf{u}|\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}) d\mathbf{u}$
- $\lambda(\mathbf{u}|\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}) = \exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u})'\hat{\boldsymbol{\beta}})$

Approximate integral with dense grid of n_p points within $\mathbf{u}_j \in \mathcal{U}$.

$$\widehat{T}(A) = T(\mathcal{B}) + \sum_{j=1}^{n_p} |U_i| \exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u}_j)'\hat{\boldsymbol{\beta}})$$

where $|U_i|$ is a small area around each $\mathbf{u}_j$

Variance

$$\mathrm{MSPE}(\hat{T}(A)) = E[(\hat{T}(A) - T(A))^2; \boldsymbol{\beta}] = E[(\hat{T}(\mathcal{U}) - T(\mathcal{U}))^2; \boldsymbol{\beta}]$$

Note: as $\mathcal{U} \cap A \to \varnothing \Rightarrow \mathrm{MSPE}(\widehat{T}(A)) \to 0$

From IPP assumption: $\hat{T}(\mathcal{U})$ independent from $T(\mathcal{U})$.

Assuming unbiasedness, $E[(\hat{T}(\mathcal{U})] = E[T(\mathcal{U})]$,

$$\begin{aligned}\mathrm{MSPE} &= \mathrm{var}[T(\mathcal{U}); \boldsymbol{\beta}] + \mathrm{var}[\hat{T}(\mathcal{U}); \boldsymbol{\beta}] \\ &= \mu(\mathcal{U}; \boldsymbol{\beta}) + \mathrm{var}[\hat{T}(\mathcal{U}); \boldsymbol{\beta}]\end{aligned}$$

Now, what about $\mathrm{var}[\hat{T}(\mathcal{U}); \boldsymbol{\beta}]$?

Variance

Recall delta method result: $\mathrm{var}(f(\mathbf{y})) \approx \mathbf{d}'\boldsymbol{\Sigma}\mathbf{d}$

Jay M. Ver Hoef (2012) Who Invented the Delta Method? The American Statistician, 66:2, 124-127

where $\mathrm{var}(\mathbf{y}) = \boldsymbol{\Sigma}$ and $d_i = \partial f(\mathbf{y})/\partial y_i$

$$d_i = \frac{\partial \hat{T}(\mathcal{U})}{\partial \beta_i} = \int_{\mathcal{U}} x_i(\mathbf{u}) \exp(\mathbf{x}(\mathbf{u})'\hat{\boldsymbol{\beta}})d\mathbf{u} \approx \frac{|\mathcal{U}|}{n_p} \sum_{i=1}^{n_p} x_i(\mathbf{s}_i) \exp(\mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}})$$

From Rathbun and Cressie, (1994), if $\hat{\beta}$ is MLE,

$$\hat{\boldsymbol{\Sigma}} = \left[\sum_{i=1}^{n} \int_{B_i} \mathbf{x}(\mathbf{s})\mathbf{x}(\mathbf{s})' \exp(\mathbf{x}(\mathbf{s})'\hat{\boldsymbol{\beta}})d\mathbf{s}\right]^{-1} \approx \left[|B| \sum_{i=1}^{n} \mathbf{x}(\mathbf{s}_i)\mathbf{x}(\mathbf{s}_i)' \exp(\mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}})\right]^{-1}$$

if $|B_i| = |B| \; \forall \; i$.

Rathbun, S. L. and Cressie, N. (1994), “Asymptotic Properties of Estimators for the Parameters of Spatial Inhomogeneous Poisson Point Processes,” Advances in Applied Probability, 26, 122-154.

Summary

$$\widehat{T}(A) = T(\mathcal{B}) + \frac{|\mathcal{U}|}{n_p} \sum_{j=1}^{n_p} \exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u}_j)'\hat{\boldsymbol{\beta}})$$

$$\begin{aligned}\widehat{\mathrm{var}}(\widehat{T}(A)) = \quad & \frac{|\mathcal{U}|}{n_p} \sum_{j=1}^{n_p} \exp(\mathbf{x}_{\hat{\boldsymbol{\rho}}}(\mathbf{u}_j)'\hat{\boldsymbol{\beta}}) + \\ & \mathbf{d}' \left[|B| \sum_{i=1}^{n} \mathbf{x}(\mathbf{s}_i)\mathbf{x}(\mathbf{s}_i)' \exp(\mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}) \right]^{-1} \mathbf{d}\end{aligned}$$

where

$$d_i = \frac{|\mathcal{U}|}{n_p} \sum_{i=1}^{n_p} x_i(\mathbf{s}_i) \exp(\mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}})$$

Simulation