N-Mixture Models with Application to Disease Surveillance

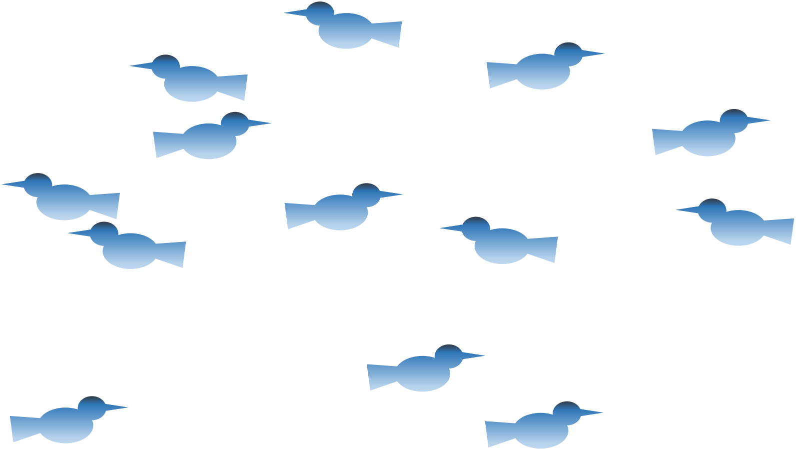
Lisa Madsen [1] Ben Brintz [2] Claudio Fuentes [1]

[1] Oregon State University

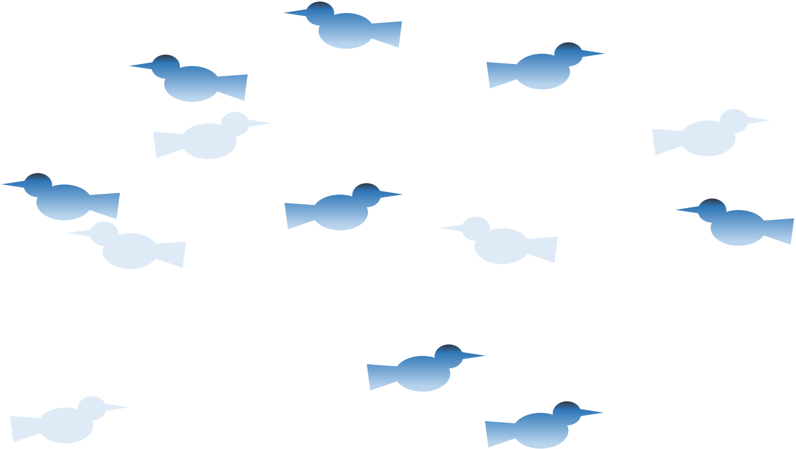
[2]University of Utah

April 8, 2022

Estimating Population Size



Estimating Population Size

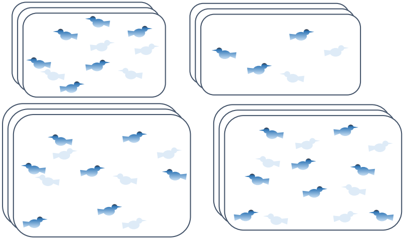


Outline

- N-Mixture Model History
  - Royle’s N-Mixture Model
  - Generalized N-Mixture Model
  - Asymptotic Approximation
- Spatial N-Mixture Model
  - Example
  - Spatial Model
  - Simulations
  - Analysis of Chlamydia Data
- Summary

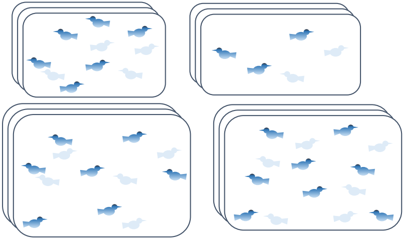
Outline

Royle (2004)



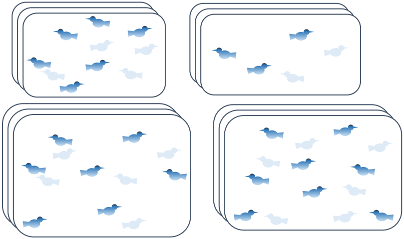
- R independent sites; T visits per site

Royle (2004)



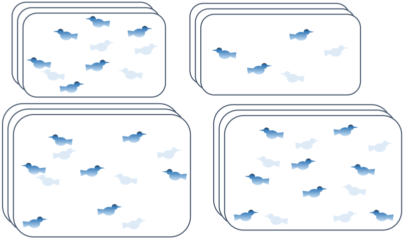
- R independent sites; T visits per site
- Closed populations: N_i animals at ith site

Royle (2004)



- R independent sites; T visits per site
- Closed populations: N_i animals at ith site
- n_{it} animals independently detected with probability p at visit t to site i

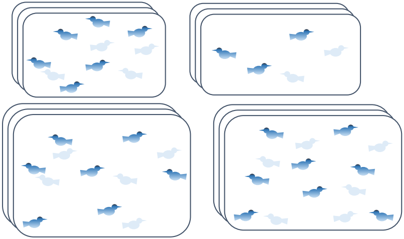
Royle (2004)



- R independent sites; T visits per site
- Closed populations: N_i animals at ith site
- n_{it} animals independently detected with probability p at visit t to site i

Then $n_{it} \sim$ independent Binomial(N_i, p)

Royle (2004)

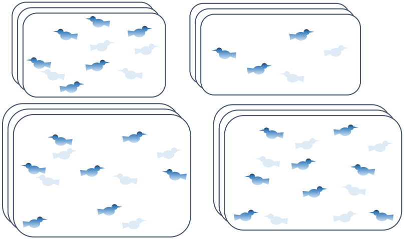


- R independent sites; T visits per site
- Closed populations: N_i animals at ith site
- n_{it} animals independently detected with probability p at visit t to site i

Then $n_{it} \sim$ independent Binomial(N_i, p)

and total abundance is $N = \sum_{i=1}^{R} N_i$.

Royle (2004)



- R independent sites; T visits per site
- Closed populations: N_i animals at ith site
- n_{it} animals independently detected with probability p at visit t to site i

Then $n_{it} \sim$ independent Binomial(N_i, p)

and total abundance is $N = \sum_{i=1}^{R} N_i$.

Prehistory

Carroll and Lombard (1985):

$$n_t | N \sim \text{iid Binomial}(N, p)$$

Prehistory

Carroll and Lombard (1985):

$$n_t|N \sim \text{iid Binomial}(N, p)$$
$$p \sim \text{beta}(\alpha, \beta)$$

Prehistory

Carroll and Lombard (1985):

$$\begin{aligned} n_t | N &\sim \text{iid Binomial}(N, p) \\ p &\sim \text{beta}(\alpha, \beta) \end{aligned}$$

Form joint likelihood $f(n_1, \ldots, n_T | N, p) \cdot f(p | \alpha, \beta)$, then integrate out p and maximize with respect to N, α and β.

Prehistory

Carroll and Lombard (1985):

$$\begin{aligned} n_t | N &\sim \text{iid Binomial}(N, p) \\ p &\sim \text{beta}(\alpha, \beta) \end{aligned}$$

Form joint likelihood $f(n_1, \ldots, n_T | N, p) \cdot f(p | \alpha, \beta)$, then integrate out p and maximize with respect to N, α and β.

Estimator doesn't perform well for small p or small N.

Model and Likelihood

$$n_{it} | N_i \sim \text{Binomial}(N_i, p)$$

Model and Likelihood

$$\begin{aligned} n_{it} | N_i &\sim \text{Binomial}(N_i, p) \\ N_i &\sim \text{iid Poisson}(\lambda) \end{aligned}$$

Model and Likelihood

$$n_{it}|N_i \sim \text{Binomial}(N_i, p)$$
$$N_i \sim \text{iid Poisson}(\lambda)$$

Joint likelihood:

$$L\left(\{N_i\}, p, \lambda | \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \left(\prod_{t=1}^{T} \text{bin}(n_{it}; N_i, p) \right) \text{pois}(N_i; \lambda) \right\},$$

where

$$\text{bin}(n_{it}; N_i, p) = \binom{N_i}{n_{it}} p^{n_{it}} (1-p)^{N_i - n_{it}}$$
$$\text{pois}(N_i; \lambda) = \frac{e^{-\lambda} \lambda^{N_i}}{N_i!}$$

Estimation

Integrated likelihood:

$$L(p, \lambda | \{n_{it}\}) = \prod_{i=1}^{R} \left\{ \sum_{N_i = M_i}^{\infty} \left(\prod_{t=1}^{T} \mathrm{bin}(n_{it}; N_i, p) \right) \mathrm{pois}(N_i; \lambda) \right\},$$

where $M_i = \max_t \{n_{it}\}$.

Estimation

Integrated likelihood:

$$L(p,\lambda|\{n_{it}\}) = \prod_{i=1}^{R}\left\{\sum_{N_i=M_i}^{\infty}\left(\prod_{t=1}^{T}\mathrm{bin}(n_{it};N_i,p)\right)\mathrm{pois}(N_i;\lambda)\right\},$$

where $M_i = \max_t\{n_{it}\}$.

Maximize $\log(L)$ numerically with respect to p and λ.

Estimation

Integrated likelihood:

$$L\left(p, \lambda | \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \sum_{N_i=M_i}^{K} \left(\prod_{t=1}^{T} \text{bin}(n_{it}; N_i, p) \right) \text{pois}(N_i; \lambda) \right\},$$

where $M_i = \max_t \{n_{it}\}$.

Maximize $\log(L)$ numerically with respect to p and λ.

$K >> \max_{it}\{n_{it}\}$

Estimation

Estimating total abundance $N = \sum_{i=1}^{R} N_i$:

Estimation

Estimating total abundance $N = \sum_{i=1}^{R} N_i$:

$$E(N_i) = \lambda$$

Estimation

Estimating total abundance $N = \sum_{i=1}^{R} N_i$:

$$E(N_i) = \lambda \quad \therefore \widehat{N} = R \cdot \widehat{\lambda}$$

Estimation

Estimating total abundance $N = \sum_{i=1}^{R} N_i$:

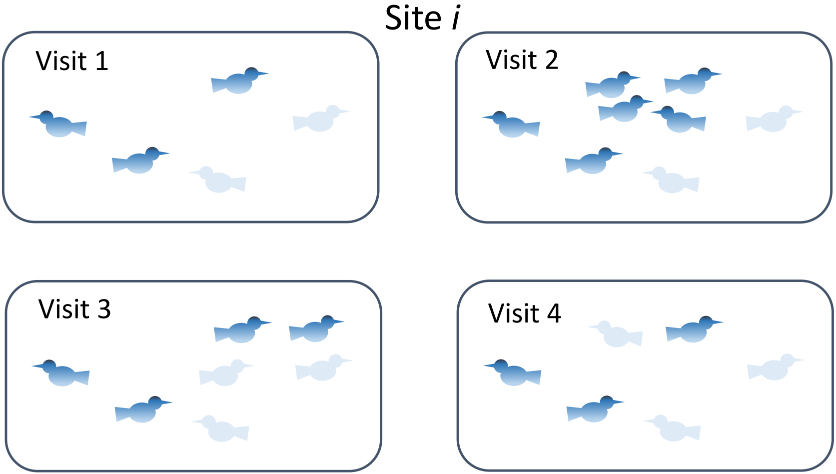
$$E(N_i) = \lambda \quad \therefore \widehat{N} = R \cdot \widehat{\lambda}$$
$$\text{and } \mathrm{SE}(\widehat{N}) = R \cdot \mathrm{SE}(\widehat{\lambda})$$

where $\mathrm{SE}(\widehat{\lambda})$ is from the inverse Hessian evaluated at the MLE.

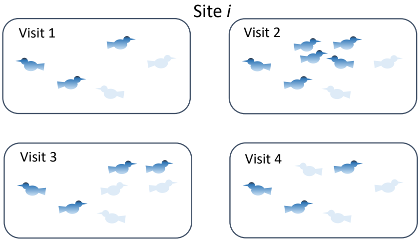
Outline

- N-Mixture Model History
  - Royle's N-Mixture Model
  - Generalized N-Mixture Model
  - Asymptotic Approximation
- Spatial N-Mixture Model
  - Example
  - Spatial Model
  - Simulations
  - Analysis of Chlamydia Data
- Summary

Open Populations

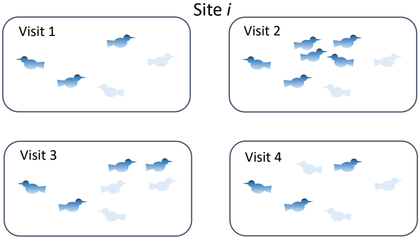


Generalized Setup



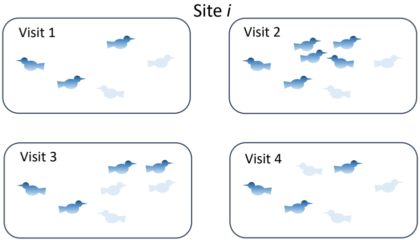
- R independent sites; T visits per site

Generalized Setup



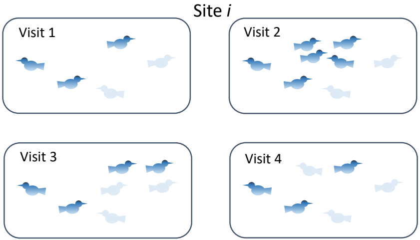
- R independent sites; T visits per site
- **Open** populations: N_{it} animals at ith site on tth visit

Generalized Setup



- R independent sites; T visits per site
- **Open** populations: N_{it} animals at ith site on tth visit
- n_{it} animals independently detected with probability p at visit t to site i

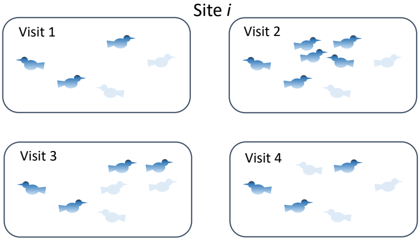
Generalized Setup



- *R* independent sites; *T* visits per site
- **Open** populations: N_{it} animals at ith site on tth visit
- n_{it} animals independently detected with probability p at visit t to site i

Then $n_{it} \sim$ independent Binomial(N_{it}, p)

Generalized Setup



- R independent sites; T visits per site
- **Open** populations: N_{it} animals at ith site on tth visit
- n_{it} animals independently detected with probability p at visit t to site i

Then $n_{it} \sim$ independent Binomial(N_{it}, p)

Goal: Estimate abundance at time t: $N_{\cdot t} \equiv \sum_{i=1}^{R} N_{it}$.

Model

Royle's model:

$$\begin{aligned} n_{it} | N_i &\sim \text{Binomial}(N_i, p) \\ N_i &\sim \text{iid Poisson}(\lambda) \end{aligned}$$

Model

Royle's model:

$$\begin{aligned} n_{it} | N_i &\sim \text{Binomial}(N_i, p) \\ N_i &\sim \text{iid Poisson}(\lambda) \end{aligned}$$

Generalized model:

$$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$$

Model

Royle's model:

$$\begin{aligned} n_{it} | N_i &\sim \text{Binomial}(N_i, p) \\ N_i &\sim \text{iid Poisson}(\lambda) \end{aligned}$$

Generalized model:

$$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$$

$$f(N_{i1}, \cdots, N_{iT}; \boldsymbol{\theta}) = f(N_{i1}; \boldsymbol{\theta}) \prod_{t=2}^{T} f(N_{it} | N_{it-1}; \boldsymbol{\theta})$$

Model

Royle's model:

$$n_{it}|N_i \sim \text{Binomial}(N_i, p)$$
$$N_i \sim \text{iid Poisson}(\lambda)$$

Generalized model:

$$n_{it}|N_{it} \sim \text{Binomial}(N_{it}, p)$$
$$f(N_{i1}, \cdots, N_{iT}; \boldsymbol{\theta}) = f(N_{i1}; \boldsymbol{\theta}) \prod_{t=2}^{T} f(N_{it}|N_{it-1}; \boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ includes $\lambda = E(N_{i1})$ and parameters describing population dynamics.

Generalized Model Population Dynamics

$S_{it} =$ **survivors** at site i from time $t-1$ to time t.

Generalized Model Population Dynamics

$S_{it} =$ **survivors** at site i from time $t-1$ to time t.

$$S_{it} | N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$$

Generalized Model Population Dynamics

$S_{it} =$ **survivors** at site i from time $t-1$ to time t.

$$S_{it}|N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$$

$G_{it} =$ **gains** to site i from time $t-1$ to time t.

Generalized Model Population Dynamics

$S_{it} =$ **survivors** at site i from time $t-1$ to time t.

$$S_{it}|N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$$

$G_{it} =$ **gains** to site i from time $t-1$ to time t.

$$G_{it}|N_{it-1} \sim \text{Poisson}(\gamma)$$

Generalized Model Population Dynamics

$S_{it} =$ **survivors** at site i from time $t-1$ to time t.

$$S_{it}|N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$$

$G_{it} =$ **gains** to site i from time $t-1$ to time t.

$$G_{it}|N_{it-1} \sim \text{Poisson}(\gamma)$$

Population dynamics parameters:

survival rate ω

Generalized Model Population Dynamics

$S_{it} =$ **survivors** at site i from time $t-1$ to time t.

$$S_{it} | N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$$

$G_{it} =$ **gains** to site i from time $t-1$ to time t.

$$G_{it} | N_{it-1} \sim \text{Poisson}(\gamma)$$

Population dynamics parameters:

survival rate ω and **recruitment rate** γ

Joint Distribution of $N_{i1}, \cdots, N_{iT}$

$$N_{i1} \sim \text{Poisson}(\lambda)$$

Joint Distribution of $N_{i1}, \cdots, N_{iT}$

$$N_{i1} \sim \text{Poisson}(\lambda)$$

$$N_{it}|N_{it-1} = S_{it}|N_{it-1} + G_{it}|N_{it-1}, t > 1$$

Joint Distribution of $N_{i1}, \cdots, N_{iT}$

$$\begin{aligned} N_{i1} &\sim \text{Poisson}(\lambda) \\ N_{it}|N_{it-1} &= S_{it}|N_{it-1} + G_{it}|N_{it-1}, t > 1 \\ S_{it}|N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\ G_{it}|N_{it-1} &\sim \text{Poisson}(\gamma) \end{aligned}$$

S_{it} and G_{it} are conditionally independent given N_{it-1}.

Likelihood

Royle's joint likelihood:

$$L\left(\{N_i\}, p, \lambda | \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \left(\prod_{t=1}^{T} \text{bin}(n_{it}; N_i, p) \right) \text{pois}(N_i; \lambda) \right\},$$

Likelihood

Royle's joint likelihood:

$$L\left(\{N_i\}, p, \lambda | \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \left(\prod_{t=1}^{T} \text{bin}(n_{it}; N_i, p) \right) \text{pois}(N_i; \lambda) \right\},$$

Generalized joint likelihood:

$$L\left(\{N_{it}\}, p, \theta | \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \left(\prod_{t=1}^{T} \text{bin}(n_{it}; N_{it}, p) \right) f(\{N_{it}\}; \theta) \right\}$$

Likelihood

Royle's joint likelihood:

$$L\left(\{N_i\}, p, \lambda | \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \left(\prod_{t=1}^{T} \mathrm{bin}(n_{it}; N_i, p) \right) \mathrm{pois}(N_i; \lambda) \right\},$$

Generalized joint likelihood:

$$L\left(\{N_{it}\}, p, \theta | \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \left(\prod_{t=1}^{T} \mathrm{bin}(n_{it}; N_{it}, p) \right) \boxed{f(\{N_{it}\}; \theta)} \right\}$$

Generalized Integrated Likelihood

$$L\left(\{N_{it}\}, p, \theta \middle| \{n_{it}\}\right) = \prod_{i=1}^{R} \left\{ \sum_{N_{i1}=n_{i1}}^{\infty} \cdots \sum_{N_{iT}=n_{iT}}^{\infty} \left(\prod_{t=1}^{T} \mathrm{bin}(n_{it}; N_{it}, p) \right) f(\{N_{it}\}; \theta) \right\}$$

Generalized Integrated Likelihood

$$L(\{N_{it}\}, p, \theta | \{n_{it}\}) =$$

$$\prod_{i=1}^{R} \left\{ \sum_{N_{i1}=n_{i1}}^{K} \cdots \sum_{N_{iT}=n_{iT}}^{K} \left(\prod_{t=1}^{T} \mathrm{bin}(n_{it}; N_{it}, p) \right) f(\{N_{it}\}; \theta) \right\}$$

$$K >> \max_{it}\{n_{it}\}$$

Generalized Integrated Likelihood

$$L(\{N_{it}\}, p, \theta | \{n_{it}\}) =$$

$$\prod_{i=1}^{R} \left\{ \sum_{N_{i1}=n_{i1}}^{K} \cdots \sum_{N_{iT}=n_{iT}}^{K} \left(\prod_{t=1}^{T} \text{bin}(n_{it}; N_{it}, p) \right) f(\{N_{it}\}; \theta) \right\}$$

$K >> \max_{it}\{n_{it}\}$

Maximize $\log(L)$ numerically with respect to $p, \lambda, \omega, \gamma$,

Generalized Integrated Likelihood

$$L\left(\{N_{it}\}, p, \theta | \{n_{it}\}\right) =$$
$$\prod_{i=1}^{R} \left\{ \sum_{N_{i1}=n_{i1}}^{K} \cdots \sum_{N_{iT}=n_{iT}}^{K} \left(\prod_{t=1}^{T} \text{bin}(n_{it}; N_{it}, p) \right) f(\{N_{it}\}; \boldsymbol{\theta}) \right\}$$

$K >> \max_{it}\{n_{it}\}$

Maximize $\log(L)$ numerically with respect to $p, \lambda, \omega, \gamma$, then estimate

$$\widehat{N}_{\cdot 1} = R\widehat{\lambda}$$
$$\widehat{N}_{\cdot t} = \widehat{\omega}\widehat{N}_{\cdot t-1} + R\widehat{\gamma}$$

Generalized Integrated Likelihood

$$L\left(\{N_{it}\}, p, \theta | \{n_{it}\}\right) =$$
$$\prod_{i=1}^{R}\left\{\sum_{N_{i1}=n_{i1}}^{K}\cdots\sum_{N_{iT}=n_{iT}}^{K}\left(\prod_{t=1}^{T}\text{bin}(n_{it}; N_{it}, p)\right) f(\{N_{it}\}; \theta)\right\}$$

$K >> \text{max}_{it}\{n_{it}\}$

Maximize $\log(L)$ numerically with respect to $p, \lambda, \omega, \gamma$, then estimate

$$\widehat{N}_{\cdot 1} = R\widehat{\lambda}$$
$$\widehat{N}_{\cdot t} = \widehat{\omega}\widehat{N}_{\cdot t-1} + R\widehat{\gamma}$$

SEs from inverse Hessian evaluated at MLE and multivariate delta method or parametric bootstrap.

Identifiability

Are the parameters of the generalized model identifiable?

Identifiability

Are the parameters of the generalized model identifiable?

Can we distinguish

> High survival and low recruitment from low survival and high recruitment?

Identifiability

Are the parameters of the generalized model identifiable?

Can we distinguish

- High survival and low recruitment from low survival and high recruitment?
- Low detection probability and high abundance from high detection probability and low abundance?

Limitations of Generalized Model

Limitations of Generalized Model

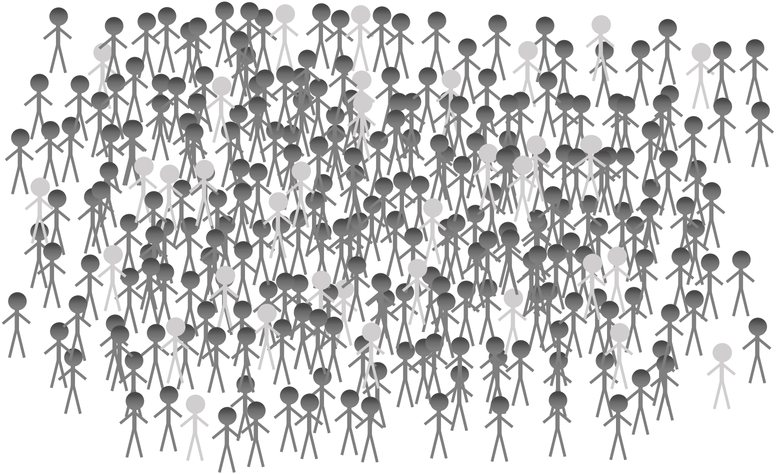
- Potential near non-identifiability

Limitations of Generalized Model

- Potential near non-identifiability
- Approximating infinite sums

Outline

Large Counts



Large Counts

The generalized model was

Large Counts

The generalized model was

$$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$$

$$f(N_{i1}, \cdots, N_{iT}; \theta) = f(N_{i1}; \theta) \prod_{t=2}^{T} f(N_{it} | N_{it-1}; \theta)$$

Large Counts

The generalized model was

$$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$$

$$f(N_{i1}, \cdots, N_{iT}; \theta) = f(N_{i1}; \theta) \prod_{t=2}^{T} f(N_{it} | N_{it-1}; \theta)$$

If the n_{it} are large,

Large Counts

The generalized model was

$$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$$

$$f(N_{i1}, \cdots, N_{iT}; \boldsymbol{\theta}) = f(N_{i1}; \boldsymbol{\theta}) \prod_{t=2}^{T} f(N_{it} | N_{it-1}; \boldsymbol{\theta})$$

If the n_{it} are large, then

$$n_{it} \overset{\text{approx}}{\sim} \text{Normal}(\mu_{it}, \sigma^2_{it}),$$

where $\mu_{it} = E(n_{it})$ and $\sigma^2_{it} = \text{var}(n_{it})$.

Approximate Likelihood

Approximate the joint distribution as multivariate normal:

$$\begin{bmatrix} n_{11} & n_{12} & \dots & n_{RT-1} & n_{RT} \end{bmatrix}' \overset{\text{approx}}{\sim} \text{MVN}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by the generalized model.

Calculating the Mean Vector

$$
\begin{aligned}
n_{it}|N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\lambda) \\
N_{it}|N_{it-1} &= S_{it}|N_{it-1} + G_{it}|N_{it-1}, \quad t > 1 \\
S_{it}|N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\
G_{it}|N_{it-1} &\sim \text{Poisson}(\gamma)
\end{aligned}
$$

Calculating the Mean Vector

$$\begin{aligned}
n_{it}|N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\lambda) \\
N_{it}|N_{it-1} &= S_{it}|N_{it-1} + G_{it}|N_{it-1}, \quad t > 1 \\
S_{it}|N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\
G_{it}|N_{it-1} &\sim \text{Poisson}(\gamma)
\end{aligned}$$

Then

Calculating the Mean Vector

$$\begin{aligned}
n_{it}|N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\lambda) \\
N_{it}|N_{it-1} &= S_{it}|N_{it-1} + G_{it}|N_{it-1}, \quad t > 1 \\
S_{it}|N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\
G_{it}|N_{it-1} &\sim \text{Poisson}(\gamma)
\end{aligned}$$

Then

$$E(n_{it}) = p\,E(N_{it})$$

Calculating the Mean Vector

$$\begin{aligned} n_{it}|N_{it} &\sim \text{Binomial}(N_{it}, p) \\ N_{i1} &\sim \text{Poisson}(\lambda) \\ N_{it}|N_{it-1} &= S_{it}|N_{it-1} + G_{it}|N_{it-1}, \quad t > 1 \\ S_{it}|N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\ G_{it}|N_{it-1} &\sim \text{Poisson}(\gamma) \end{aligned}$$

Then

$$\begin{aligned} E(n_{it}) &= p\,E(N_{it}) \\ E(N_{i1}) &= \lambda \\ E(N_{it}) &= \omega\,E(N_{it-1}) + \gamma, \quad t > 1 \end{aligned}$$

Calculating the Mean Vector

$$\begin{aligned} n_{it}|N_{it} &\sim \text{Binomial}(N_{it}, p) \\ N_{i1} &\sim \text{Poisson}(\lambda) \\ N_{it}|N_{it-1} &= S_{it}|N_{it-1} + G_{it}|N_{it-1}, \quad t > 1 \\ S_{it}|N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\ G_{it}|N_{it-1} &\sim \text{Poisson}(\gamma) \end{aligned}$$

Then

$$\begin{aligned} E(n_{it}) &= p\, E(N_{it}) \\ E(N_{i1}) &= \lambda \\ E(N_{it}) &= \omega\, E(N_{it-1}) + \gamma, \quad t > 1 \end{aligned}$$

Calculate elements of Σ similarly.

Estimation

Maximize log of approximate likelihood

$$L(p, \lambda, \omega, \gamma | \{n_{it}\}) = MVN(\{n_{it}\}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with respect to the parameters, then estimate $N_{\cdot t}$ as before:

Estimation

Maximize log of approximate likelihood

$$L(p, \lambda, \omega, \gamma | \{n_{it}\}) = MVN(\{n_{it}\}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with respect to the parameters, then estimate $N_{\cdot t}$ as before:

$$\begin{aligned}\widehat{N}_{\cdot 1} &= R\widehat{\lambda} \\ \widehat{N}_{\cdot t} &= \widehat{\omega}\widehat{N}_{\cdot t-1} + R\widehat{\gamma}\end{aligned}$$

Identifiability Diagnostic

The $MVN(\boldsymbol{\mu}, \Sigma)$ model has a closed-form expression for *jk*th element of the Fisher Information $\boldsymbol{I}$:

Identifiability Diagnostic

The *MVN*$(\boldsymbol{\mu}, \Sigma)$ model has a closed-form expression for *jk*th element of the Fisher Information $\boldsymbol{I}$:

$$\boldsymbol{I}_{jk} = \frac{\partial \boldsymbol{\mu}'}{\partial \theta_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_k} + \frac{1}{2} \operatorname{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k} \right)$$

Identifiability Diagnostic

The $MVN(\boldsymbol{\mu}, \Sigma)$ model has a closed-form expression for *jk*th element of the Fisher Information $\boldsymbol{I}$:

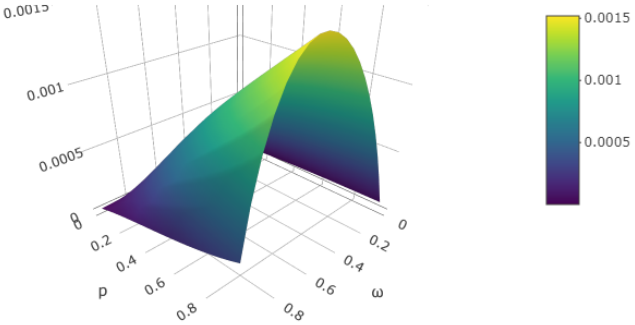
$$\boldsymbol{I}_{jk} = \frac{\partial \boldsymbol{\mu}'}{\partial \theta_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_k} + \frac{1}{2} \operatorname{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k} \right)$$

Use $\boldsymbol{I}$ to diagnose near-non-identifiability.

Identifiability Diagnostic

Example: $R = 1$, $T = 2$, $\lambda = 100$, and $\gamma = 90$

Minimum Eigenvalue of $\boldsymbol{I}$



Outline

Chlamydia in Oregon

Chlamydia in Oregon

Chlamydia is a common sexually-transmitted disease.

- Relatively easy to diagnose and cure

Chlamydia in Oregon

Chlamydia is a common sexually-transmitted disease.

- Relatively easy to diagnose and cure
- Symptoms may not occur for several weeks

Chlamydia in Oregon

Chlamydia is a common sexually-transmitted disease.

- Relatively easy to diagnose and cure
- Symptoms may not occur for several weeks
- Can cause permanent damage if left untreated

Chlamydia in Oregon

Chlamydia is a common sexually-transmitted disease.

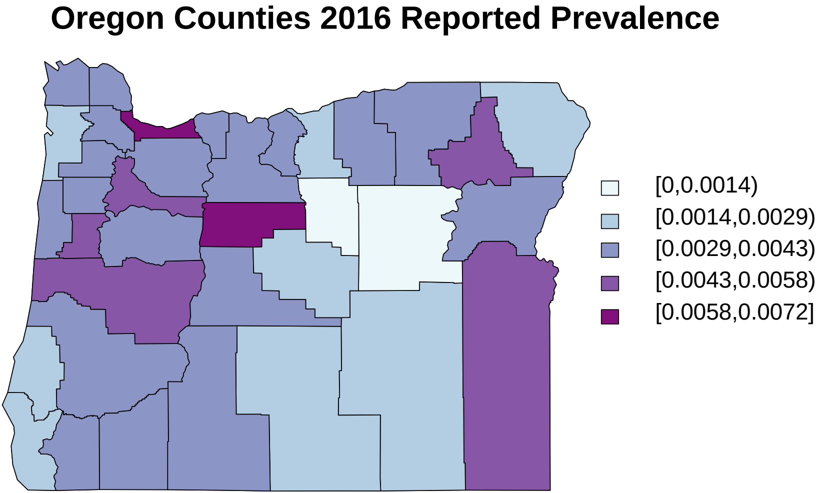
- Relatively easy to diagnose and cure
- Symptoms may not occur for several weeks
- Can cause permanent damage if left untreated

Goal: Use N-Mixture model to estimate chlamydia case counts.

Oregon Population and Cases by County

County	Population			Observed Cases		
	2010	...	2018	2010	...	2018
Multnomah	737,291	...	811,880	3296	...	5459
Washington	531,645	...	597,695	1390	...	2404
Clackamas	376,790	...	416,075	945	...	1394
Lane	351,923	...	379,611	1276	...	1844
Marion	315,951	...	346,868	1395	...	1887
⋮	⋮		⋮	⋮		⋮
Grant	7464	...	7176	14	...	24
Wallowa	7012	...	7081	9	...	9
Gilliam	1882	...	1894	3	...	3
Sherman	1779	...	1708	4	...	4
Wheeler	1447	...	1366	3	...	3

Chlamydia in Oregon



Outline

N-Mixture Model History
- Royle's N-Mixture Model
- Generalized N-Mixture Model
- Asymptotic Approximation

Spatial N-Mixture Model
- Example
- Spatial Model
- Simulations
- Analysis of Chlamydia Data

Summary

Notation and Terminology for Spatial Model

Notation and Terminology for Spatial Model

$$n_{it} = \text{observed case count in county } i\text{, year } t$$
$$N_{it} = \text{true case count in county } i\text{, year } t$$

Notation and Terminology for Spatial Model

$$\begin{aligned} n_{it} &= \text{observed case count in county } i\text{, year } t \\ N_{it} &= \text{true case count in county } i\text{, year } t \\ \text{pop}_{it} &= \text{population in county } i\text{, year } t \end{aligned}$$

Notation and Terminology for Spatial Model

$$
\begin{aligned}
n_{it} &= \text{observed case count in county } i\text{, year } t \\
N_{it} &= \text{true case count in county } i\text{, year } t \\
\text{pop}_{it} &= \text{population in county } i\text{, year } t \\
N_{it}/\text{pop}_{it} &= \textbf{prevalence}
\end{aligned}
$$

Notation and Terminology for Spatial Model

$$\begin{aligned} n_{it} &= \text{observed case count in county } i\text{, year } t \\ N_{it} &= \text{true case count in county } i\text{, year } t \\ \text{pop}_{it} &= \text{population in county } i\text{, year } t \\ N_{it}/\text{pop}_{it} &= \textbf{prevalence} \\ \{N_{\alpha,\tau}\} &= \{N_{it} : i = 1, \ldots, R \text{ and } t = 1, \ldots, \tau\} \end{aligned}$$

Notation and Terminology for Spatial Model

$$
\begin{aligned}
n_{it} &= \text{observed case count in county } i\text{, year } t \\
N_{it} &= \text{true case count in county } i\text{, year } t \\
\text{pop}_{it} &= \text{population in county } i\text{, year } t \\
N_{it}/\text{pop}_{it} &= \textbf{prevalence} \\
\{N_{\alpha,\tau}\} &= \{N_{it} : i = 1, \ldots, R \text{ and } t = 1, \ldots, \tau\}
\end{aligned}
$$

$R \times R$ adjacency matrix:

$$
A_{ij} = \begin{cases} 1, & \text{county } j \text{ borders county } i \text{ or } i = j \\ 0, & \text{otherwise} \end{cases}
$$

Notation and Terminology for Spatial Model

$$\begin{aligned} n_{it} &= \text{observed case count in county } i\text{, year } t \\ N_{it} &= \text{true case count in county } i\text{, year } t \\ \text{pop}_{it} &= \text{population in county } i\text{, year } t \\ N_{it}/\text{pop}_{it} &= \textbf{prevalence} \\ \{N_{\alpha,\tau}\} &= \{N_{it} : i = 1, \ldots, R \text{ and } t = 1, \ldots, \tau\} \end{aligned}$$

$R \times R$ adjacency matrix:

$$A_{ij} = \begin{cases} 1, & \text{county } j \text{ borders county } i \text{ or } i = j \\ 0, & \text{otherwise} \end{cases}$$

Size of neighborhood of county i: $A_{i\cdot} = \sum_{j=1}^{R} A_{ij}$.

Spatial Model

$$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$$

Spatial Model

$$
\begin{aligned}
n_{it} | N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\beta \cdot \text{pop}_{i1})
\end{aligned}
$$

Spatial Model

$$
\begin{aligned}
n_{it} | N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\beta \cdot \text{pop}_{i1}) \\
N_{it} | N_{it-1} &= S_{it} | N_{it-1} + G_{it} | N_{it-1}, \quad t > 1
\end{aligned}
$$

Spatial Model

$$\begin{aligned} n_{it} | N_{it} &\sim \text{Binomial}(N_{it}, p) \\ N_{i1} &\sim \text{Poisson}(\beta \cdot \text{pop}_{i1}) \\ N_{it} | N_{it-1} &= S_{it} | N_{it-1} + G_{it} | N_{it-1}, \quad t > 1 \\ S_{it} | N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \end{aligned}$$

Spatial Model

$$
\begin{aligned}
n_{it} | N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\beta \cdot \text{pop}_{i1}) \\
N_{it} | N_{it-1} &= S_{it} | N_{it-1} + G_{it} | N_{it-1}, \quad t > 1 \\
S_{it} | N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\
G_{it} | \{N_{\alpha,t-1}\} &\sim \text{Poisson}(\gamma \cdot \widetilde{N}_{it-1})
\end{aligned}
$$

where

$$\widetilde{N}_{it-1} = pop_{it-1} \cdot \sum_{r=1}^{R} \frac{A_{ir} N_{rt-1} / pop_{rt-1}}{A_{i\cdot}}.$$

Spatial Model

$$
\begin{aligned}
n_{it} | N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\beta \cdot \text{pop}_{i1}) \\
N_{it} | N_{it-1} &= S_{it} | N_{it-1} + G_{it} | N_{it-1}, \quad t > 1 \\
S_{it} | N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\
G_{it} | \{N_{\alpha,t-1}\} &\sim \text{Poisson}(\gamma \cdot \widetilde{N}_{it-1})
\end{aligned}
$$

where

$$
\widetilde{N}_{it-1} = pop_{it-1} \cdot \sum_{r=1}^{R} \frac{A_{ir} N_{rt-1} / pop_{rt-1}}{A_{i\cdot}}.
$$

Spatial Model

$$\begin{aligned}
n_{it} | N_{it} &\sim \text{Binomial}(N_{it}, p) \\
N_{i1} &\sim \text{Poisson}(\beta \cdot \text{pop}_{i1}) \\
N_{it} | N_{it-1} &= S_{it} | N_{it-1} + G_{it} | N_{it-1}, \quad t > 1 \\
S_{it} | N_{it-1} &\sim \text{Binomial}(N_{it-1}, \omega) \\
G_{it} | \{N_{\alpha,t-1}\} &\sim \text{Poisson}(\gamma \cdot \widetilde{N}_{it-1})
\end{aligned}$$

where

$$\widetilde{N}_{it-1} = pop_{it-1} \cdot \sum_{r=1}^{R} \frac{A_{ir} \boxed{N_{rt-1}/pop_{rt-1}}}{A_{i\cdot}}.$$

$\widetilde{N}_{it-1}$ is the population in county i at time $t-1$ times the average prevalence in its neighborhood.

Interpreting Spatial Model Parameters

$$n_{it}|N_{it} \sim \text{Binomial}(N_{it}, p)$$

Interpreting Spatial Model Parameters

$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p) \quad \rightarrow$ p is detection probability

Interpreting Spatial Model Parameters

$n_{it}|N_{it} \sim \text{Binomial}(N_{it}, p) \quad \rightarrow \quad$ p is detection probability

$N_{i1} \sim \text{Poisson}(\beta \cdot \text{pop}_{i1})$

Interpreting Spatial Model Parameters

$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$ → p is detection probability

$N_{i1} \sim \text{Poisson}(\beta \cdot \text{pop}_{i1})$ → β is expected initial prevalence

Interpreting Spatial Model Parameters

$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$ → p is detection probability

$N_{i1} \sim \text{Poisson}(\beta \cdot \text{pop}_{i1})$ → β is expected initial prevalence

$S_{it} | N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$

Interpreting Spatial Model Parameters

$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$ → p is detection probability

$N_{i1} \sim \text{Poisson}(\beta \cdot \text{pop}_{i1})$ → β is expected initial prevalence

$S_{it} | N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$ → ω is persistence rate

Interpreting Spatial Model Parameters

$n_{it} | N_{it} \sim \text{Binomial}(N_{it}, p)$ → p is detection probability

$N_{i1} \sim \text{Poisson}(\beta \cdot \text{pop}_{i1})$ → β is expected initial prevalence

$S_{it} | N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$ → ω is persistence rate

$G_{it} | \{N_{\alpha,t-1}\} \sim \text{Poisson}(\gamma \cdot \widetilde{N}_{it-1})$

Interpreting Spatial Model Parameters

$n_{it}|N_{it} \sim \text{Binomial}(N_{it}, p)$ → p is detection probability

$N_{i1} \sim \text{Poisson}(\beta \cdot \text{pop}_{i1})$ → β is expected initial prevalence

$S_{it}|N_{it-1} \sim \text{Binomial}(N_{it-1}, \omega)$ → ω is persistence rate

$G_{it}|\{N_{\alpha,t-1}\} \sim \text{Poisson}(\gamma \cdot \widetilde{N}_{it-1})$ → γ is infection rate

Estimation

As with the non-spatial normal-approximation model,

- Write MVN $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in terms of $\boldsymbol{\theta} = (p, \beta, \omega, \gamma)$.
- Maximize log of approximate likelihood

$$L(\boldsymbol{\theta}|\{n_{it}\}) = MVN(\{n_{it}\}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with respect to $\boldsymbol{\theta}$.

Estimation

As with the non-spatial normal-approximation model,

- Write MVN $\boldsymbol{\mu}$ and Σ in terms of $\boldsymbol{\theta} = (p, \beta, \omega, \gamma)$.
- Maximize log of approximate likelihood

$$L(\boldsymbol{\theta} | \{n_{it}\}) = MVN(\{n_{it}\}; \boldsymbol{\mu}, \Sigma)$$

  with respect to $\boldsymbol{\theta}$.
- Estimate N_{it} sequentially:

$$\widehat{N}_{i1} = \widehat{\beta}\,\text{pop}_{i1}$$

$$\widehat{N}_{it} = \widehat{\omega}\,\widehat{N}_{it-1} + \widehat{\gamma}\,\widehat{\widetilde{N}}_{it-1}$$

Estimation

As with the non-spatial normal-approximation model,

- Write MVN $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in terms of $\boldsymbol{\theta} = (p, \beta, \omega, \gamma)$.
- Maximize log of approximate likelihood

$$L(\boldsymbol{\theta} | \{n_{it}\}) = MVN(\{n_{it}\}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

  with respect to $\boldsymbol{\theta}$.
- Estimate N_{it} sequentially:

$$\widehat{N}_{i1} = \widehat{\beta}\,\text{pop}_{i1}$$

$$\widehat{N}_{it} = \widehat{\omega}\,\widehat{N}_{it-1} + \widehat{\gamma}\,\widehat{\widetilde{N}}_{it-1}$$

- For each t, $\widehat{N}_{\cdot t} = \sum_{i=1}^{R} \widehat{N}_{it}$.

Confidence Intervals

Account for uncertainty in $\widehat{\theta}$ with a parametric bootstrap:

- Given $\widehat{\theta}$, calculate Fisher Information $\boldsymbol{I}(\widehat{\theta})$.

Confidence Intervals

Account for uncertainty in $\widehat{\boldsymbol{\theta}}$ with a parametric bootstrap:

- Given $\widehat{\boldsymbol{\theta}}$, calculate Fisher Information $\boldsymbol{I}(\widehat{\boldsymbol{\theta}})$.
- Simulate a large number S of bootstrap samples from MVN$[\widehat{\boldsymbol{\theta}}, \boldsymbol{I}^{-1}(\widehat{\boldsymbol{\theta}})]$.

Confidence Intervals

Account for uncertainty in $\widehat{\theta}$ with a parametric bootstrap:

- Given $\widehat{\theta}$, calculate Fisher Information $\boldsymbol{I}(\widehat{\theta})$.
- Simulate a large number S of bootstrap samples from MVN$[\widehat{\theta}, \boldsymbol{I}^{-1}(\widehat{\theta})]$.

Bootstrap sample $\widehat{\theta}^{(1)}, \ldots, \widehat{\theta}^{(S)}$ represents the sampling distribution of $\widehat{\theta}$.

Confidence Intervals

Similarly, account for the sampling variability in the N_{it}.

Confidence Intervals

Similarly, account for the sampling variability in the N_{it}.

Given $\widehat{\theta}^{(s)}$, for each site i, generate $\widehat{N}_{i1}^{(s)}, \ldots, \widehat{N}_{iT}^{(s)}$ according to the model:

$$\widehat{N}_{i1}^{(s)} \sim \text{Poisson}(\widehat{\beta}^{(s)} \text{pop}_{i1})$$

Confidence Intervals

Similarly, account for the sampling variability in the N_{it}.

Given $\widehat{\theta}^{(s)}$, for each site i, generate $\widehat{N}_{i1}^{(s)}, \ldots, \widehat{N}_{iT}^{(s)}$ according to the model:

$$\widehat{N}_{i1}^{(s)} \sim \text{Poisson}(\widehat{\beta}^{(s)} \text{pop}_{i1})$$
$$\widehat{N}_{it}^{(s)} | \widehat{N}_{it-1}^{(s)} = S_{it}^{(s)} | \widehat{N}_{it-1}^{(s)} + G_{it}^{(s)} | \widehat{N}_{it-1}^{(s)}, \quad t > 1$$

Confidence Intervals

Similarly, account for the sampling variability in the N_{it}.

Given $\widehat{\theta}^{(s)}$, for each site i, generate $\widehat{N}_{i1}^{(s)}, \ldots, \widehat{N}_{iT}^{(s)}$ according to the model:

$$\begin{aligned}
\widehat{N}_{i1}^{(s)} &\sim \text{Poisson}(\widehat{\beta}^{(s)}\text{pop}_{i1}) \\
\widehat{N}_{it}^{(s)}|\widehat{N}_{it-1}^{(s)} &= S_{it}^{(s)}|\widehat{N}_{it-1}^{(s)} + G_{it}^{(s)}|\widehat{N}_{it-1}^{(s)}, \quad t > 1 \\
S_{it}^{(s)}|\widehat{N}_{it-1}^{(s)} &\sim \text{Binomial}(\widehat{N}_{it-1}^{(s)}, \widehat{\omega}^{(s)}) \\
G_{it}^{(s)}|\widehat{N}_{\alpha,t-1}^{(s)} &\sim \text{Poisson}(\widehat{\gamma}^{(s)} \cdot \widehat{\widetilde{N}}_{it-1}^{(s)})
\end{aligned}$$

Confidence Intervals

For $s = 1, \ldots, S$ we have generated:

$$\begin{array}{ccc} \widehat{N}_{11}^{(s)} & ,\ldots, & \widehat{N}_{1T}^{(s)} \\ \vdots & & \vdots \\ \widehat{N}_{R1}^{(s)} & ,\ldots, & \widehat{N}_{RT}^{(s)} \end{array}$$

Confidence Intervals

For $s = 1, \ldots, S$ we have generated:

$$\begin{matrix} \widehat{N}_{11}^{(s)} & , \ldots , & \widehat{N}_{1T}^{(s)} \\ \vdots & & \vdots \\ \widehat{N}_{R1}^{(s)} & , \ldots , & \widehat{N}_{RT}^{(s)} \end{matrix}$$

Then calculate $\widehat{N}_{\cdot t}^{(s)} = \sum_{i=1}^{R} \widehat{N}_{it}^{(s)}$ to get $\widehat{N}_{\cdot t}^{(1)}, \ldots, \widehat{N}_{\cdot t}^{(S)}$.

Confidence Intervals

For $s = 1, \ldots, S$ we have generated:

$$\begin{array}{ccc} \widehat{N}_{11}^{(s)} & , \ldots , & \widehat{N}_{1T}^{(s)} \\ \vdots & & \vdots \\ \widehat{N}_{R1}^{(s)} & , \ldots , & \widehat{N}_{RT}^{(s)} \end{array}$$

Then calculate $\widehat{N}_{\cdot t}^{(s)} = \sum_{i=1}^{R} \widehat{N}_{it}^{(s)}$ to get $\widehat{N}_{\cdot t}^{(1)}, \ldots, \widehat{N}_{\cdot t}^{(S)}$.

$1 - \alpha$ confidence bounds for $N_{\cdot t}$ are the $\alpha/2$ and $1 - \alpha/2$ quantiles of the bootstrapped distribution of $\widehat{N}_{\cdot t}$.

Outline

- N-Mixture Model History
  - Royle's N-Mixture Model
  - Generalized N-Mixture Model
  - Asymptotic Approximation
- Spatial N-Mixture Model
  - Example
  - Spatial Model
  - Simulations
  - Analysis of Chlamydia Data
- Summary

Scenarios

Using Oregon's populations and map, we ran 1000 simulations for each of 24 scenarios.

	Parameter	Values
β	initial expected prevalence	$0.005, 0.05$
p	detection probability	$0.4, 0.7, 0.9$
ω	persistence rate	$0.5, 0.8$
γ	infection rate	$0.3, 0.6$

Procedure for *k*th Simulation

- Simulate $N_{it}^{(k)}$ for $t = 1, \ldots, 9$, with Oregon's populations and map, and draw $n_{it}^{(k)} \sim \text{Bin}(N_{it}^{(k)}, p)$.

Procedure for *k*th Simulation

- Simulate $N_{it}^{(k)}$ for $t = 1, \ldots, 9$, with Oregon's populations and map, and draw $n_{it}^{(k)} \sim \text{Bin}(N_{it}^{(k)}, p)$.
- Estimate $\boldsymbol{\theta} = p, \beta, \omega, \gamma$ by maximizing $L(\boldsymbol{\theta} | \{n_{it}\})$.

Procedure for *k*th Simulation

- Simulate $N_{it}^{(k)}$ for $t = 1, \ldots, 9$, with Oregon's populations and map, and draw $n_{it}^{(k)} \sim \text{Bin}(N_{it}^{(k)}, p)$.
- Estimate $\boldsymbol{\theta} = p, \beta, \omega, \gamma$ by maximizing $L(\boldsymbol{\theta} | \{n_{it}\})$.
- Calculate the $R \cdot T$ estimates $\widehat{N}_{it}^{(k)}$ recursively using the model and $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$.

Procedure for *k*th Simulation

- Simulate $N_{it}^{(k)}$ for $t = 1, \ldots, 9$, with Oregon's populations and map, and draw $n_{it}^{(k)} \sim \text{Bin}(N_{it}^{(k)}, p)$.
- Estimate $\boldsymbol{\theta} = p, \beta, \omega, \gamma$ by maximizing $L(\boldsymbol{\theta}|\{n_{it}\})$.
- Calculate the $R \cdot T$ estimates $\widehat{N}_{it}^{(k)}$ recursively using the model and $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$.
- Record the absolute relative error of $\widehat{N}_{it}^{(k)}$:

$$\text{ARE}_{it}^{(k)} = \frac{|N_{it}^{(k)} - \widehat{N}_{it}^{(k)}|}{E(N_{it}^{(k)})}.$$

Procedure for *k*th Simulation

- Simulate $N_{it}^{(k)}$ for $t = 1, \ldots, 9$, with Oregon's populations and map, and draw $n_{it}^{(k)} \sim \text{Bin}(N_{it}^{(k)}, p)$.
- Estimate $\boldsymbol{\theta} = p, \beta, \omega, \gamma$ by maximizing $L(\boldsymbol{\theta} | \{n_{it}\})$.
- Calculate the $R \cdot T$ estimates $\widehat{N}_{it}^{(k)}$ recursively using the model and $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$.
- Record the absolute relative error of $\widehat{N}_{it}^{(k)}$:

$$\text{ARE}_{it}^{(k)} = \frac{|N_{it}^{(k)} - \widehat{N}_{it}^{(k)}|}{E(N_{it}^{(k)})}.$$

Procedure for *k*th Simulation

- Simulate $N_{it}^{(k)}$ for $t = 1, \ldots, 9$, with Oregon's populations and map, and draw $n_{it}^{(k)} \sim \text{Bin}(N_{it}^{(k)}, p)$.
- Estimate $\boldsymbol{\theta} = p, \beta, \omega, \gamma$ by maximizing $L(\boldsymbol{\theta} | \{n_{it}\})$.
- Calculate the $R \cdot T$ estimates $\widehat{N}_{it}^{(k)}$ recursively using the model and $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$.
- Record the absolute relative error of $\widehat{N}_{it}^{(k)}$:

$$\text{ARE}_{it}^{(k)} = \frac{|N_{it}^{(k)} - \widehat{N}_{it}^{(k)}|}{E(N_{it}^{(k)})}.$$

- Perform parametric bootstrap to simulate sampling distribution of $\widehat{N}_{\cdot 9}$ and calculate 90% interval estimate of $N_{\cdot 9}^{(k)}$.

Simulation Results

For each of the 24 simulation scenarios,

- Average parameters estimates $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$ over the 1000 simulations.

Simulation Results

For each of the 24 simulation scenarios,

- Average parameters estimates $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$ over the 1000 simulations.
- Summarize absolute relative errors:

$$\text{ARE} = \frac{1}{RT} \sum_{i=1}^{R} \sum_{t=1}^{T} \left[\frac{1}{1000} \sum_{k=1}^{1000} \frac{|N_{it}^{(k)} - \hat{N}_{it}^{(k)}|}{E(N_{it}^{(k)})} \right]$$

Simulation Results

For each of the 24 simulation scenarios,

- Average parameters estimates $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$ over the 1000 simulations.
- Summarize absolute relative errors:

$$\text{ARE} = \frac{1}{RT} \sum_{i=1}^{R} \sum_{t=1}^{T} \left[\frac{1}{1000} \sum_{k=1}^{1000} \frac{|N_{it}^{(k)} - \hat{N}_{it}^{(k)}|}{E(N_{it}^{(k)})} \right]$$

- Calculate percentage of successful 90% interval estimates.

Simulation Results

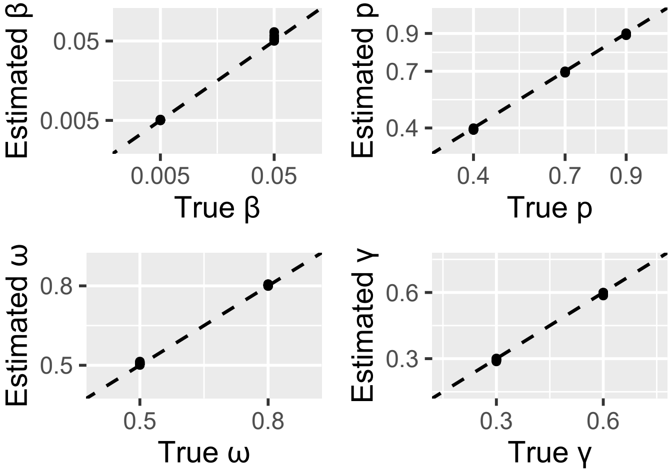
For each of the 24 simulation scenarios,

- Average parameters estimates $\widehat{p}, \widehat{\beta}, \widehat{\omega}, \widehat{\gamma}$ over the 1000 simulations.
- Summarize absolute relative errors:

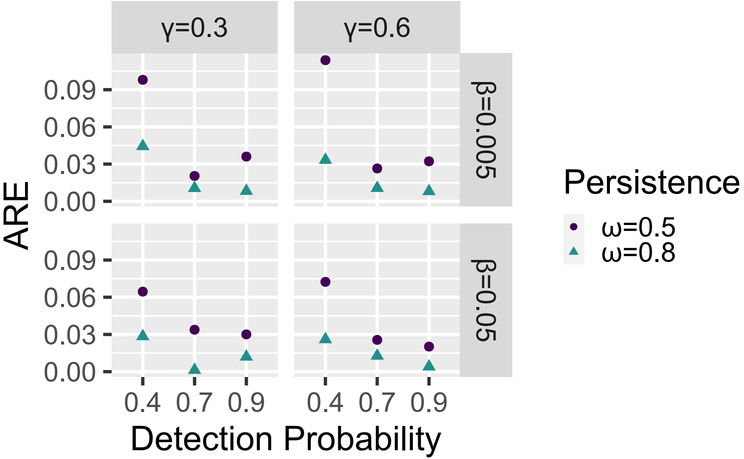
$$\text{ARE} = \frac{1}{RT} \sum_{i=1}^{R} \sum_{t=1}^{T} \left[\frac{1}{1000} \sum_{k=1}^{1000} \frac{|N_{it}^{(k)} - \hat{N}_{it}^{(k)}|}{E(N_{it}^{(k)})} \right]$$

- Calculate percentage of successful 90% interval estimates.
- Calculate average width of 90% interval estimates.

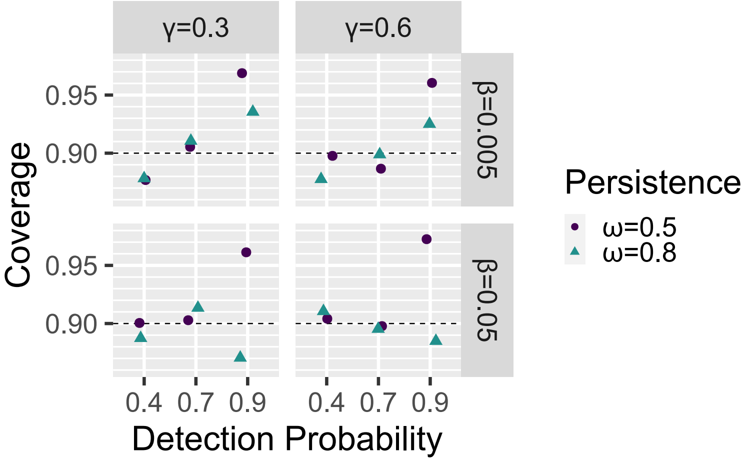
Results: Parameter Estimates



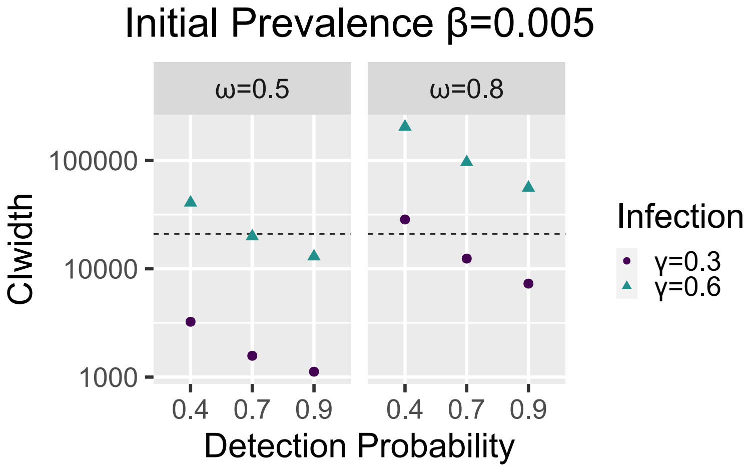
Results: Absolute Relative Error



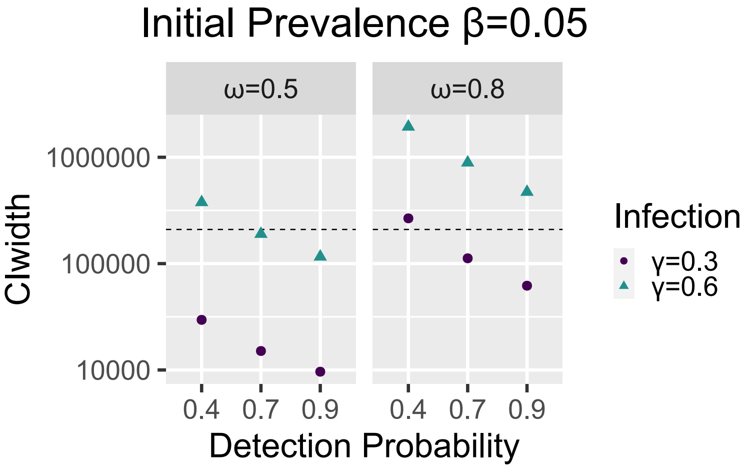
Results: Confidence Coverage



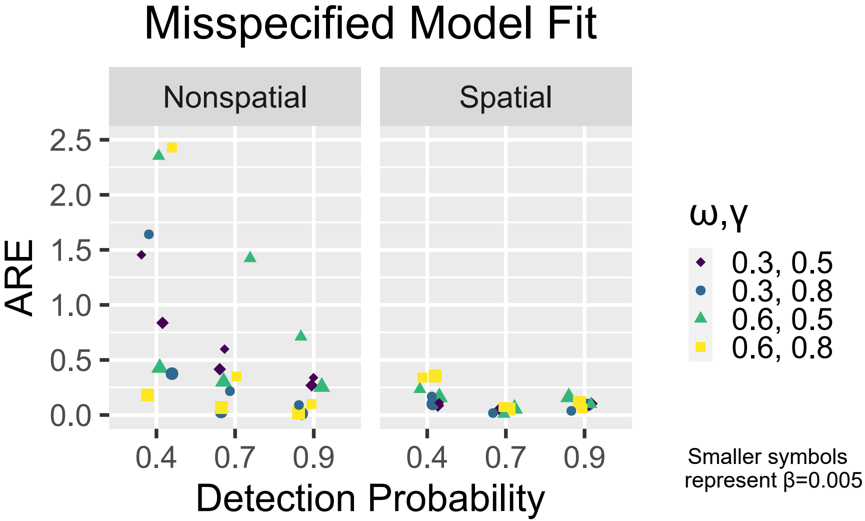
Results: Confidence Interval Width



Results: Confidence Interval Width

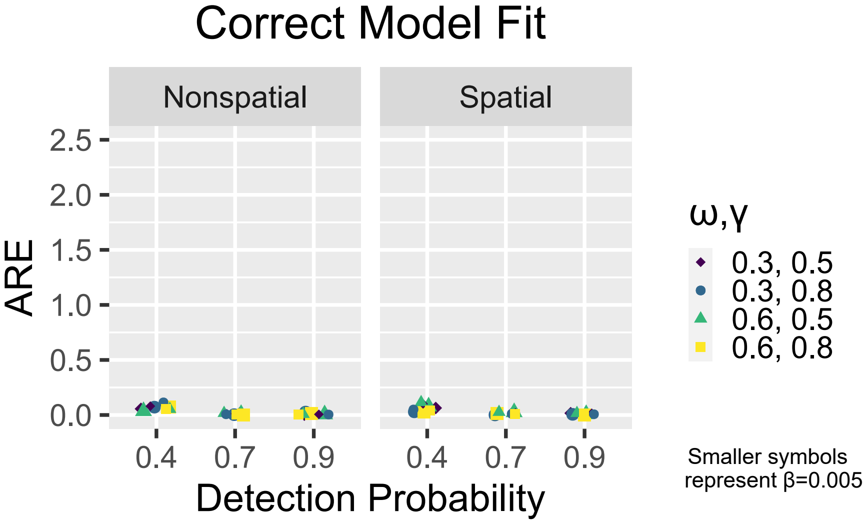


Robustness



Smaller symbols represent β=0.005

Robustness


Correct Model Fit
Nonspatial
Spatial
ARE
2.5
2.0
1.5
1.0
0.5
0.0
0.4
0.7
0.9
0.4
0.7
0.9
Detection Probability
ω,γ
0.3, 0.5
0.3, 0.8
0.6, 0.5
0.6, 0.8
Smaller symbols represent β=0.005

Outline

N-Mixture Model History
- Royle’s N-Mixture Model
- Generalized N-Mixture Model
- Asymptotic Approximation

Spatial N-Mixture Model
- Example
- Spatial Model
- Simulations
- Analysis of Chlamydia Data

Summary

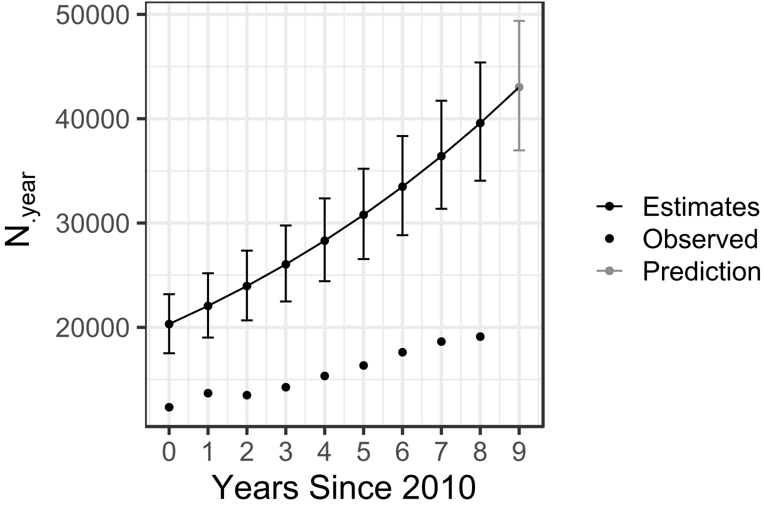
Oregon Chlamydia Data 2010-2018

County	Population			Observed Cases		
	2010	...	2018	2010	...	2018
Multnomah	737,291	...	811,880	3296	...	5459
Washington	531,645	...	597,695	1390	...	2404
Clackamas	376,790	...	416,075	945	...	1394
Lane	351,923	...	379,611	1276	...	1844
Marion	315,951	...	346,868	1395	...	1887
⋮	⋮		⋮	⋮		⋮
Grant	7464	...	7176	14	...	24
Wallowa	7012	...	7081	9	...	9
Gilliam	1882	...	1894	3	...	3
Sherman	1779	...	1708	4	...	4
Wheeler	1447	...	1366	3	...	3

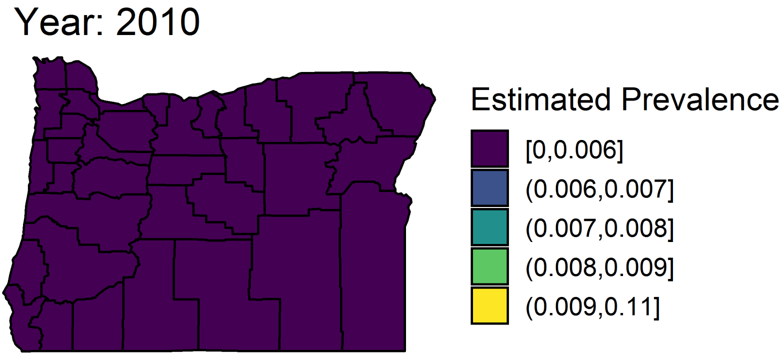
Parameter Estimates (SEs)

Initial expected prevalence	$\widehat{\beta} = 0.0053\ (0.000000014)$
Detection probability	$\widehat{p} = 0.669\ (0.047)$
Persistence rate	$\widehat{\omega} = 0.917\ (0.032)$
Infection rate	$\widehat{\gamma} = 0.169\ (0.032)$

State-wide Interval Estimates/Prediction

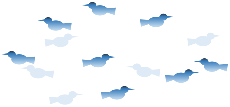


County	2018 Population	Estimated Case Count	95% CL
Multnomah	811,880	7651	(6696, 8574)
Washington	597,695	5558	(4863, 6227)
Clackamas	416,075	3896	(3405, 4373)
Lane	379,611	3607	(3155, 4050)
Marion	346,868	3244	(2829, 3646)
⋮	⋮	⋮	⋮
Grant	7176	75	(56, 95)
Wallowa	7081	71	(52, 91)
Gilliam	1894	20	(11, 29)
Sherman	1708	18	(9, 28)
Wheeler	1366	14	(7, 22)


Year: 2010
Estimated Prevalence
[0,0.006]
(0.006,0.007]
(0.007,0.008]
(0.008,0.009]
(0.009,0.11]

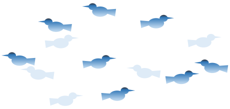
Summary

N-Mixture models provide estimates of N when $p < 1$.

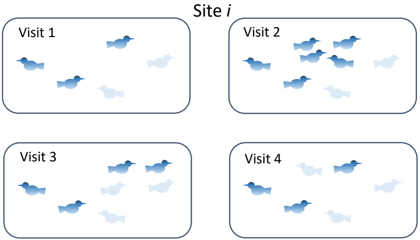


Summary

N-Mixture models provide estimates of N when $p < 1$.

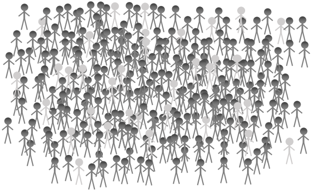


Generalized model allows open populations.



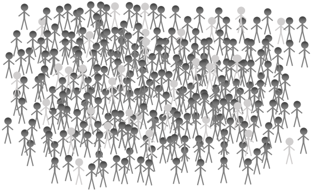
Summary

Asymptotic approximation allows large counts.

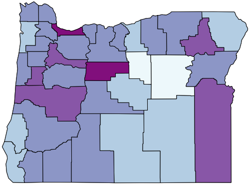


Summary

Asymptotic approximation allows large counts.



Spatial model accounts for spatial dependence.



Future Work

- Further model testing and refinement
- Develop identifiability diagnostic
- Adapt model for other diseases (susceptible/infected/recovered)

Thanks to my coauthors, Ben and Claudio.

Thanks to my coauthors, Ben and Claudio.

Thank you!

References

B Brintz, C Fuentes, and L Madsen. An asymptotic approximation to the N-mixture model for the estimation of disease prevalence. *Biometrics*, 74(4):1512–1518, 2018.

B Brintz, L Madsen, and C Fuentes. A spatially explicit N-mixture model for the estimation of disease prevalence. *Statistical Modelling*, 2021. doi: https://doi.org/10.1177/1471082X211020872.

Raymond J Carroll and F Lombard. A note on N estimators for the binomial distribution. *Journal of the American Statistical Association*, 80(390):423–426, 1985.

D Dail and L Madsen. Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, 67(2): 577–587, 2011.

J Andrew Royle. N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1):108–115, 2004.