Advances in the Analysis of Spatially Aggregated Data

Julia Schedler

11/25/19

Overview

- Motivation
- GLM for areal data
- (extended) Hausdorff Distance
- Background on case-crossover
- STARMA models
- Case-crossover in a STARMA model context

Motivation

- How do we model spatially-referenced, aggregated count data?
- How can we include popular epidemiological methodology within this framework?
- How can we account for characteristics like zero inflation or hierarchical structure?
- How can we provide tools to make this modeling framework easily useable?

Generalized Linear Regression

- (Nelder and Wedderburn 1972) extended Gaussian linear regression models to encompass all (one parameter) exponential family dependent variables
- non-normal linear means modeled using a link function
- later extensions allowed for both fixed and random effects (Gaussian) -(Raudenbush and Bryk 2002) develop Hierarchical GLMs which can have non-Gaussian error distributions

GLM's for Spatial Count Data

- necessarily associated with lattice data
- Early methodology arose as adaptations of methods for time series of counts (for example Liang and Zeger 1986,S. L. Zeger (1988)).
- (Albert and McShane 1995) develop a model for spatially correlated binary count data (neuroimaging); (Gotway and Stroup 1997) generalize these to cateogrical/discrete spatial data
- Huge explosion since then, see (Anselin 2002), (Ward and Gleditsch 2008) , and (De Oliveira 2012)

Flexibility of GLMs

GLMs allow for all sorts of dependent variables:

- Count Data models: Poisson, Binomial, Negative Binomial
- Zero-Inflated models
- Hurdle Models

GLM's for Spatial Data- technical details

Consider the following data model and process model:

$$Z(s_i)|Y(s_i) \sim \textit{ind.exponentialfamily}(\exp(Y(s_i)))$$
$$\mathbf{Y}|\beta, \tau^2, \phi \sim N(\mathbf{X}\beta, \tau^2(\mathbf{I} - \phi\mathbf{H})^{-1})$$

- The conditional distribution of the data (Z) given the process (Y) could be normal, poisson, binomial, etc.
- **W** is a spatial weight matrix
- β is the vector of regression coefficients
- τ is an overdispersion parameter
- ϕ is the spatial autocorrelation parameter

CAR models in the GLM context

If $Z(s_i)$ follows a Gaussian distribution, where the s_i's form a lattice, the CAR model can be written

$$Y(s_i)|Y(N(s_i)) \sim N(X\beta, (I - \rho W)^{-1} M)$$

Where M is a diagonal matrix (e.g. $M = diag(|N(s_1)|^{-1}, \ldots, |N(s_n)|^{-1})$.

Specification of W

- For geostatistical data: **W** is specified by choosing an appropriate covariance model via the empirical variogram
- For lattice data: **W** encodes conditional independence structure (zeroes on diagonals and all entries (i, j) where s_i is not a neighbor of s_j)
- must be row-standardized
- can be binary, or weights
- How to choose neighborhood structure and their weights?

Popular neighborhood structures for lattices

- contiguity: two regions are neighbors if they share at least one (queen) or more than one (rook) boundary point
- can lead to vastly differing numbers of neighbors for different regions (e.g. larger regions will have more neighbors)
- k nearest neighbors: calculate the distances between two regions as the distance between a single point in each
- e.g. geometric centroid, population-weighted centroid, or other meaningful location

How important is the choice of neighborhood structure?

- Wall (2004) found counterintuitive implied correlations from SAR/CAR models fit using various neighborhood schemes
- LeSage (2008) compare the log-likelihood values of models using contiguity matrices and nearest neighbor matrices for varying numbers of neighbors
- nearest neighbor performs better than contiguity. They recommend comparing different values of k to assess sensitivity of results to the number of neighbors.
- Underlying distance metric need not be Euclidean. (Shahid et al. 2009) explore different distance metrics which capture road distance

Hausdorff Distance

The Hausdorff distance measures the distance between two sets:

$$\begin{aligned} H(A,B) &= \max\{h(A,B), h(B,A)\} \\ &= \max\{\max_{p_a \in A} \min_{p_b \in B} d(p_a, p_b), \max_{p_b \in B} \min_{p_a \in A} d(p_a, p_b)\} \end{aligned}$$

- The directional Hausdorff distance $h(A, B)$ from a set A to a set B is the largest possible distance between any point in A and the closest point in B.
- The Hausdorff distance between A and B is then the larger of two two directional Hausdorff distances.
- can use any underlying distance metric d

Two Ideas for Hausdorff Distance

1. use Hausdorff distance as a way to generate spatial weight matrices for lattice data
2. use Hausdorff distance as a way to generate spatial covariates

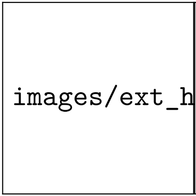
Hausdorff distance for Spatial Weight Matrices

- K nearest neighbors using Hausdorff Distance instead of centroid-based distance
- Inverse distance weighting using Hausdorff Distance

Hausdorff distance for spatial covariates

For example, use the hausdorff distance to generate “distance to” type variables, e.g. the distance between a superneighborhood and the closest highway (rather than centroid distance or closest boundary point)

Hausdorff Distance for irregular geometries



Extended Hausdorff Distance

- The extended Hausdorff distance (Min, Zhilin, and Xiaoyong 2007) allows for a characterization of the distribution of distances between two objects.

$$H^{f_1 f_2}(X, Y) = \max\left\{ k^{th}_{p_a \in A} \min_{p_b \in B} \{d(p_a, p_b)\}, k^{th}_{p_b \in B} \min_{p_a \in A} \{d(p_b, p_a)\} \right\}$$

- $k^{th}_{x \in X} f(x)$ is the k^{th} q-quantile of $f(x)$ over X
- f_1 is the ratio k/q for the first term and f_2 is the ratio for the second term

Extended Hausdorff Distance – illustration

images/ext_haus_example.png

Figure 1:

Extended Hausdorff- real example

Images/ext_haus_50.png

Figure 2: Median Hausdorff Distance from Texas

- notice how buffer width changes as width of target shape changes
- Accounts for the fact that parts of Nebraska are closer to Texas than parts of Arkansas

Calculating Extended Hausdorff distance: The ϵ buffer method

The following is the ϵ buffer method suggested by Min, Zhilin, and Xiaoyong (2007) to calculate extended Hausdorff distance

1. Generate N_B points in/on B
2. Calculate the distance $d(p_b, B)$ for all the points
3. Rank the distances, the k^{th} quantile will be the directional extended Hausdorff distance from A to B.

ϵ buffer method visualized

images/epsilon_buffer.png

images/epsilon_buffer_2.png

Implementation in R

```r
# generate points
n = 10000
a.coords <- sp::spsample(A, n = n,type = "regular")

## points from A to B
dists <- rgeos::gDistance(a.coords, B, byid = T)
## find desired quantile of distances
eps <- quantile(dists[1,], f1)
```

Next Steps for Extended Hausdorff

- Write a function to calculate the extended Hausdorff distance using any underlying distance metric
- will replace `gDistance` function in current code
- include option for user-defined distances
- Create an R package with extended Hausdorff capabilities for Spatial objects in R (sp package)

Case-crossover

- (Maclure 1991) introduced the case-crossover design as a way to assess the effect of a transient exposure on an accute outcome
- Similar to case-control designs, but use subjects at previous time points as controls

Images/cc_diagram.PNG

Figure 3: Case-crossover v. Case-control (Maclure and Mittleman 2000)

Case-crossover model

The case-crossover design uses conditional logistic regression to fit the following model:

$$\lambda_i(t, X_{it}) = \lambda_{0it} \exp(\beta X_{it}) = \lambda_{0i} \exp(\beta X_{it} + \gamma_{it})$$

- individual, time-varying nuisance factors drop out of the model

Relative Risk model, continued

The case-crossover assumption is important in the estimation of the probability that subject i fails at time t, given that t is in a pre-specified reference window R

$$
\begin{aligned}
p_{it} &= P(T_i, \sum_{m=1}^{N_T} Y_{im} = 1 = t | X, R(t)) \\
&= \frac{\lambda_{0i} \exp(\beta X_{it} + \gamma_{it})}{\sum_{j \in R(t)} \lambda_{0i} \exp(\beta X_{ij} + \gamma_{ij})}
\end{aligned}
$$

The case-crossover assumption

Images/cc_assump_edit.PNG

Figure 4: The case-crossover assumption, visualized

$$
\begin{aligned}
p_{it} &= \frac{\lambda_{0i} \exp(\beta X_{it} + \gamma_{it})}{\sum_{j \in R(t)} \lambda_{0i} \exp(\beta X_{ij} + \gamma_{ij})} \\
&= \frac{\exp(\beta X_{it})}{\sum_{j \in R(t)} \exp(\beta X_{ij})}
\end{aligned}
$$

Choice of Reference Window

Two popular choices

- Time-stratified: divides study period into pre-specified reference windows
- leads to unbiased estimates
- has issues when trends are present in outcome variable
- partitions the study period– no overlap bias
- Symmetric bi-directional:
- leads to biased estimates
- does not partition the study period, leading to overlap bias
- adjustments exist (semi-symmetric bi-directional), but are complicated to implement

Choice of Reference Window

M. A. Mittleman (2005) calls the choice of referent window design a “settled" issue and recommends the time-stratified design. This advice seems mostly heeded, though a large number of case-crossover studies do not mention the particular referent window scheme at all.

Figure 5: Publications with “Case-Crossover" in Title, Keywords, or Abstract, 1990-2017

Equivalence with Poisson Regression

Lu and Zeger (2007) generalize the equivalence of case-crossover estimated using conditional logistic regression with Poisson regression

- previously noted by Levy et al. (2001) and Janes, Sheppard, and Lumley (2005)
- Time-stratified design: equivalent to Poisson regression with dummy variables indicating the strata (prespecified reference windows)
- Symmetric bi-directional: equivalent to using a weighted running mean smoother to estimate the nuisance term in the Poisson regression

Equivalence with Poisson Regression

Equivalence is demonstrated by showing the CLR and Poisson regression estimating equations are the same (given a particular reference window design)

Let Y_{it} indicate whether subject i experiences the event of interest at time t.

Then $Y_t = \sum_i Y_{it}$ represents the number of events observed at time t. The expected number of events at time t is given by:

$$\mu_t = \sum_i \lambda_i(t, X_t) = \sum_i \lambda_{0i} \exp(\beta X_t + \gamma_{it}) = \exp(\beta X_t + S_t),$$

where $S_t = \sum_i \lambda_{0i} \exp(\gamma_{it}$ is the sum over all individual nuisance factors.

A spatial case-crossover?

Why should we include a case-crossover component

- widespread use for common epidemiological questions
- encourage more wholistic approach– it's not case crossover OR glm, it can be both (they are equivalent)
- hasn't been done spatially
- What would the case-crossover assumption look like in a spatial model?
- An individual's spatially varying nuisance factor in a given region is the same as it is in neighboring regions ("close" regions)

A spatial case-crossover?

- Motivated by equivalence with Poisson regression (a glm)
- The “spatial" relative risk model is:

$$\lambda_i(s, X_{is}) = \lambda_{0is} \exp(\beta X_{is}) = \lambda_{0i} \exp(\beta X_{is} + \gamma_{is})$$

Does this make sense? It says the relative risk of subject i experiencing the event in region s is a function of their risk of experiencing the event in $R(s)$, the set of reference regions for s.

- But you can’t be in more than one place at once
- This type of spatial dependence works in aggregate, but not at the individual level
- When analyzing the impact of transient effects on acute outcomes, time is a necessary component.
- How can we include a case-crossover component in a spatiotemporal model?

Spatiotemporal Autoregressie Moving Average (STARMA) Models

Consider a spatiotemporal process $== (Z_t(s_1), Z_t(s_2), \ldots, Z_t(s_N))'$ defined by

$$(t) = \sum_{k=0}^{p} \sum_{j=1}^{\lambda_k} \xi_{kj} W_{kj}(t-k) - \sum_{l=0}^{q} \sum_{j=1}^{\mu_l} \phi_{lj} V_{lj}(t-l) + (t)$$

- p, λ_k are the temporal and spatial autoregressive lags
- q, μ_l are the temporal and spatial moving average parts lags
- λ_k is the order of the spatial lag in the
- ξ_{kj} and ϕ_{lj} are the AR and MA parameters to be estimated
- W_{kj} and V_{lj} are spatial weight matrices for AR time lag k and space lag j and MA time lag l and spatial lag
- (t) are i.i.d. mean zero error terms
- note there are no exogenous variables

Regression Models with STARMA errors

Following Wells and SenGupta (2011), consider the following regression model with STARMA errors:

$$= g(, \beta) +$$

$$= \sum_{k=0}^{p} \sum_{j=1}^{\lambda_k} \xi_{kj} W_{kj\,t-k} - \sum_{l=0}^{q} \sum_{j=1}^{\mu_l} \phi_{lj} V_{lj\,t-l} +_{t}$$

For simplicity, we will consider a single spatial weight matrix $W = W_{kj} = V_{lj}$ and set $p = q = 1$. The model simplifies to:

$$= \beta + \xi_{10\,t-1} + \xi_{11} W_{t-1} + \phi_{10\,t-1} + \phi_{11} W_{t-1} +_{t}$$

Indexing in Space

Rather than considering a collection of spatial processes indexed in time, for the purposes of considering a case-crossover component we will consider a collection of temporal processes indexed in space, that is, $== (Y_s(t_1), Y_s(t_2), \ldots, Y_s(t_T))'$. The the STARMA model can be written:

$$
\begin{aligned}
&= g(\beta+) \\
&= \sum_{k=0}^{p}\sum_{j=1}^{n} \xi_{k1} w_{sj} B_j^{(k)} - \sum_{l=0}^{q}\sum_{j=1}^{n} \phi_{l1} w_{sj} B_j^{(l)} +_s
\end{aligned}
$$

- B is the backwards shift operator
- Note that the model as written above assumes:
- the order of the spatial lag is 1 for both the autoregressive and moving average parts

Indexing in Space

Assuming the order of the temporal lag is 1 for both parts, the model simplifies to:

$$\begin{aligned} Y_s(t_i) = X_s(t_i)\beta + \xi_{10} Z_s(t_{i-1}) + \xi_{11} \sum_{j=1}^{n} w_{sj} Z_s(t_{i-1}) \\ + \phi_{10} \epsilon_s(t_{i-1}) + \phi_{11} \sum_{j=1}^{n} w_{sj} \epsilon_s(t_{i-1}) + \epsilon_s(t_i) \end{aligned}$$

Case-crossover in STARMA model context

- The case-crossover model corresponds to a STAR model (no MA part)
- In the case crossover model, the risk of subject k experiencing the event of interest in region s at time t is a function of the risk at times in the reference window of their event time, $R(t)$
- Rather than using the temporal (unidirectional) backwards shift operator B we will consider the temporal shift operator to be omnidirectional
- The shift operator for a symmetric bi-directional design which uses the time immediately prior and immediately after to estimate the relative risk can be written as follows for 5 time points:

$$B^{SBD} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

STAR model with omnidirectional temporal shift operator

Letting $g(\cdot) \equiv \exp(\cdot)$

$$= \exp(\beta+)$$

$$= \xi_{01} \sum_{j=1}^{n} w_{sj} Z_j(t) + \xi_{10} B^{SBD} +_s (t)$$

Written element-wise, this simplifies to:

$$Y_s(t_i) = \exp(X_s(t_i)\beta + Z_s(t_i))$$

$$Z_s(t_i) = \sum_{j=1}^{n} w_{sj} Z_j(t_i) + \xi_{10}(Z_s(t_{i-1}) + Z_s(t_i) + Z_s(t_{i+1})) + \epsilon_s(t_i)$$

Structure of $Z_s(t)$

Following the construction for the temporal case crossover, let $Y_s(t_i) = \sum_k Y_s(t_i, k)$, where $Y_s(t_i, k)$ is 1 if subject k experiences the event in region s at time t. Suppose this probability is given by the relative risk model:

$$\lambda_s(t_i, k) = \lambda_{0st_ik} \exp(X_s(t_i)\beta) = \lambda_{0sk} \exp(X_s(t_i)\beta + \gamma_s(t_i, k))$$

It follows that the expected number of events in region s at time t is the sum over the population of individuals:

$$\begin{aligned}\mu_{st_i} &= \sum_k \lambda_s(t_i, k) = \sum_k \lambda_{0sk} \exp(X_s(t_i)\beta + \gamma_s(t_i, k)) \\ &= \exp(X_s(t_i)\beta + Z_s(t_i))\end{aligned}$$

Where $\exp(Z_s(t_i)) = \sum_k \lambda_{0sk} \exp(\gamma_s(t_i, k))$

STAR Case-crossover model

The case-crossover assumption is that $\gamma_s(t_i, k) = \gamma_s(t^*, k)$ for all $t^* \in R(t_i)$. - then we have that $Z_s(t_i) = Z_s(t^*)$ for all $t^* \in R(t_i)$.

Applying this to the STAR model with SBD, we have:

$$\begin{aligned} Y_s(t_i) &= \exp(X_s(t_i)\beta + Z_s(t_i)) \\ Z_s(t_i) &= \sum_{j=1}^{n} w_{sj} Z_j(t_i) + \xi_{10}(Z_s(t_{i-1}) + Z_s(t_i) + Z_s(t_{i+1})) + \epsilon_s(t_i) \\ &= \sum_{j=1}^{n} w_{sj} Z_j(t_i) + \xi_{10}(|R(t_i)| Z_s(t_i)) + \epsilon_s(t_i) \end{aligned}$$

The term $|R(t_i)|$ replaces B^{SBD}. In fact, this will work independent of referent window design.

Next steps for STAR case-crossover

- Remove the term $|R(t_i)|$ and allow ξ_{10} to scale the effect from the case-crossover assumption
- Allow the term ξ_{10} to vary in space, that is, replace it with ξ_{10s}.
- Explore the ability of this model to account for spatial nonstationarity via differencing
- Estimation and prediction

References

Albert, Paul S, and Lisa M McShane. 1995. "A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data." *Biometrics* 51: 627–38. `https://www.jstor.org/stable/pdf/2532950.pdf`.

Anselin, Luc. 2002. "Under the hood:Issues in the specification and interpretation of spatial regression models." *Agricultural Economics* 27: 247–67. doi:10.1111/j.1574-0862.2002.tb00120.x.

De Oliveira, Victor. 2012. "Bayesian analysis of conditional autoregressive models." *Annals of the Institute of Statistical Mathematics* 64 (1): 107–33. doi:10.1007/s10463-010-0298-1.

Gotway, C A, and W W Stroup. 1997. "A Generalized Linear Model Approach to Spatial Data Analysis and Prediction." *Source: Journal of Agricultural, Biological, and Environmental Statistics Journal of Agricultural, Biological, and Environmental Statistics* 24223640 (18): 157–17826. http://www.jstor.org/stable/1400401 http://about.jstor.org/terms http://www.jstor.org/stable/1400401{\%}0Ahttp://about.jstor.org/terms.