DISCRETE SPATIAL AUTOREGRESSIVE MODELS

The standard logistic, binomial and Poisson regression models of discrete counting data have natural spatial generalizations in a manner similar to the conditional autoregressive (CAR) model. We consider each in turn.

1. Autologistic Model

Suppose one considers the voting outcomes for Pennsylvania counties during a given Presidential election. Let $v_i = 1$ if county i voted Democratic and $v_i = 0$ otherwise. To model this voting behavior, one may consider a number of county covariates, $x_i = (x_{i1},..,x_{ik})$, say with $x_{i1} =$ "per capita income" in county i, $x_{i2} =$ "average years of education" for voters in county i, and so on. A *standard logistic model* of voting behavior would then take the form:

$$(1) \qquad \Pr(v_i = 1) = \frac{\mu_i}{1+\mu_i}$$

where

$$(2) \qquad \mu_i = \exp\left(\alpha + \sum_{j=1}^{k} \beta_j x_{ij}\right)$$

However, it may well be that there are voting similarities among adjacent counties. Hence, if we let $w_{ij} = 1$ if county j is a contiguous *neighbor* of county i (with $w_{ii} = 0$) and if we let $v_{-i} = (v_1,..,v_{i-1},v_{i+1},..,v_n)$ denote the voting outcomes of all counties other than i, then a natural spatial generalization of this model would be to hypothesize that the *conditional distribution* of v_i given the voting outcomes, v_{-i}, of all other counties depends only on the voting behavior of its neighbors, say,

$$(3) \qquad \Pr(v_i = 1 \mid v_{-i}) = \frac{\mu_i(v_{-i})}{1+\mu_i(v_{-i})}$$

with

$$(4) \qquad \mu_i(v_{-i}) = \exp\left(\alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \rho \sum_{j \neq i} w_{ij} v_j\right)$$

To estimate this model, we start with the standard logistic model that has a likelihood function of the form:

$$(5) \qquad L(\alpha, \beta \mid v) = \prod_{i=1}^{n} \Pr(v_i = 1)^{v_i} [1 - \Pr(v_i = 1)]^{1-v_i}$$

$$\Rightarrow \log L(\alpha,\beta\,|\,v) = \sum_{i=1}^{n}\left\{v_i \log \Pr(v_i=1) + (1-v_i)\log[1-\Pr(v_i=1)]\right\}$$

where $v=(v_1,..,v_n)$ and where $\Pr(v_i=1)$ is given by (1) and (2). This function is quite easy to maximize, and yields well behaved maximum-likelihood estimates of α and β.

But for the spatial generalization in (3) and (4) this is not nearly so simple. First, it is not even clear whether this family of conditional probability distributions is actually consistent with a common joint probability distribution. Here it turns out that they are indeed consistent, and that this joint distribution has the form [see for example in Cressie (1993, section 6.5.1)]:

$$(6) \qquad \Pr(v) = \frac{\exp[Q(v)]}{\sum_s \exp[Q(s)]}$$

where:

$$(7) \qquad Q(v) = \sum_{i=1}^{n} v_i\left(\alpha + \sum_{j=1}^{k}\beta_j x_{ij}\right) + \rho \sum_{1\le i\le j\le n} w_{ij} v_i v_j$$

Hence by definition

$$(8) \qquad L(\alpha,\beta,\rho\,|\,v) = \frac{\exp[Q(v)]}{\sum_s \exp[Q(s)]}$$

with $Q(\cdot)$ given by (7). But notice that the denominator is summed over all possible realizations of $v=(v_1,..,v_n)$. So for example if there were only 20 counties, then the denominator would contain 2^{20} $(=1{,}048{,}576)$ terms. Hence even for small numbers of areal units this function is *completely intractable*.

Several alternative procedures have been suggested. The simplest is to ignore interdependencies and consider maximizing the *pseudo-likelihood function*

$$(9) \qquad L^*(\alpha,\beta,\rho\,|\,v) = \prod_{i=1}^{n}\left\{\Pr(v_i=1\,|\,v_{-i})^{v_i}[1-\Pr(v_i=1\,|\,v_{-i})]^{1-v_i}\right\}$$

in which the marginal probabilities in (5) are simply replaced by the conditional probabilities in (3). If spatial dependencies are not too great, then it can be shown that this procedure still yields consistent estimates. However, the procedure turns out to be extremely inefficient in that the variances of these estimates (especially $\hat{\rho}$) can extremely large.

An alternative procedure is to simulate values of the joint distribution, as proposed by Geyer and Thompson (1992). The idea is very simple. If we let $\theta = (\alpha, \beta, \rho)$ and write (6) as

$$\text{(10)} \qquad \Pr_\theta(v) = \Pr(v) = \frac{\exp[Q(v)]}{\sum_s \exp[Q(s)]} = \frac{\exp[Q(v)]}{c(\theta)}$$

then for any given set of parameter values, say $\theta_0 = (\alpha_0, \beta_0, \rho_0)$, it follows that

$$\text{(11)} \qquad 1 = \sum_v \Pr_\theta(v) = \sum_v \frac{\exp[Q(v)]}{c(\theta)} \;\Rightarrow\; c(\theta) = \sum_v \exp[Q(v)]$$

$$c(\theta) = \sum_v \frac{\exp[Q_\theta(v)]}{\exp[Q_{\theta_0}(v)]} \exp[Q_{\theta_0}(v)] = c(\theta_0) \sum_v \left[\frac{\exp[Q_\theta(v)]}{\exp[Q_{\theta_0}(v)]}\right] \frac{\exp[Q_{\theta_0}(v)]}{c(\theta_0)}$$

$$= c(\theta_0) \cdot E_{\theta_0}\left[\frac{\exp[Q_\theta(v)]}{\exp[Q_{\theta_0}(v)]}\right]$$

But since θ_0 is known, it follows that if we can simulate this distribution, and obtain sample values, $v_{(s)}$, $s = 1,..,N$, then we can obtain corresponding sample estimates,

$$\text{(12)} \qquad \hat{E}_{\theta_0}(\theta) = \frac{1}{N} \sum_{s=1}^{N} \frac{\exp[Q_\theta(v_{(s)})]}{\exp[Q_{\theta_0}(v_{(s)})]}$$

of the mean value in (11). This in turn implies that

$$\text{(13)} \qquad \hat{c}(\theta) = c(\theta_0) \cdot \hat{E}_{\theta_0}(\theta)$$

and thus that we can approximate the function $c(\theta)$ up to an unknown constant of proportionality, $c(\theta_0)$. Hence this approximation can be used to obtain approximations the likelihood function in (10). Notice that while this simulation procedure is time consuming, it only need be done *once*, since the same samples $v_{(s)}$, $s = 1,..,N$ can be used in every round of the iterative maximization procedure to obtain estimates $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\rho})$.

2. Auto Binomial Model

Next suppose that one is looking at housing abandonments in Philadelphia, and has data $h = (h_i : i = 1,..,n)$ on the number of abandoned houses in each block group $i = 1,..,n$. If the abandonment probabilities depend on a number of housing attributes $x_i = (x_{i1},..,x_{ik})$ as well as the frequency of abandonments in neighboring block groups, and if the total number of housing units in i is denoted by N_i, then one might consider *conditional binomial probabilities* of the form:

$$(14) \qquad \Pr(h_i \mid h_{-i}) = \binom{N_i}{h_i} p_i(h_{-i})^{h_i}[1 - p_i(h_{-i})]^{N_i - h_i}$$

where

$$(15) \qquad p_i(h_{-i}) = \frac{\mu_i(h_{-i})}{1 + \mu_i(h_{-i})}$$

with

$$(16) \qquad \mu_i(h_{-i}) = \exp\left(\alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \rho \sum_{j \neq i} w_{ij} h_j\right)$$

Here these conditional binomials can again be shown to be consistent with a joint distribution of the form

$$(17) \qquad \Pr(h) = \frac{\exp[Q(h)]}{\sum_s \exp[Q(s)]}$$

where:

$$(18) \qquad Q(h) = \sum_{i=1}^{n} h_i \left(\alpha + \sum_{j=1}^{k} \beta_j x_{ij}\right) + \rho \sum_{1 \le i \le j \le n} w_{ij} h_i h_j + \sum_{i=1}^{n} \log \binom{N_i}{h_i}$$

Hence the Geyer-Thompson procedure can again be used for estimation.

3. Auto Poisson Model

Finally, if the N_i's are quite large, then it is reasonable to approximate the auto binomial model by an *auto Poisson model* of the form:

$$(19) \qquad \Pr(h_i \mid h_{-i}) = \frac{[\lambda_i(h_{-i})]^{h_i}}{h_i!} \exp[-\lambda_i(h_{-i})]$$

with

$$(20) \qquad \lambda_i(h_{-i}) = \exp\left(\alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \rho \sum_{j \neq i} w_{ij} h_j\right)$$

For the Poisson approximation to the auto binomial, $\lambda_i(h_{-i})$ replaces the binomial mean values, $N_i p_i(h_{-i})$. In all cases, the appropriate joint distribution takes the form (17) with

$$(21) \qquad Q(h) = \sum\nolimits_{i=1}^{n} h_i\left(\alpha + \sum\nolimits_{j=1}^{k} \beta_j x_{ij}\right) + \rho \sum_{1 \le i \le j \le n} w_{ij} h_i h_j - \sum\nolimits_{i=1}^{n} \log(h_i!)$$

However, it turns out that this model *is only well defined if* $\rho < 0$! The problem can be seen from (20) where the influence of neighboring counts *always increases the expected number of counts* at i when $\rho > 0$. This mutual inflation procedure can easily be shown to drive counts to infinity (since there is no upper bound on counts in the auto Poisson model). Some efforts have been made to rectify this [see Augustin, N.H., et al. (2004)] by considering *truncated auto Poissons* of the form,

$$(22) \qquad \Pr(h_i \mid h_{-i}) = \frac{\left\{[\lambda_i(h_{-i})]^{h_i} / h_i!\right\} \exp[-\lambda_i(h_{-i})]}{\sum_{k=0}^{N} \left\{[\lambda_i(h_{-i})]^{k} / k!\right\} \exp[-\lambda_i(h_{-i})]} ; \; h_i = 0,1,..,N$$

Notices that if N is replaced by N_i for each i then this starts to look very much like the auto binomial model above. Hence this truncated model is most useful in cases where there are no reference populations (as for example when looking at the number of traffic accidents in each areal unit).

Finally, if we let

$$(23) \qquad Q(h) = \sum\nolimits_{i=1}^{n} h_i\left(\alpha + \sum\nolimits_{j=1}^{k} \beta_j x_{ij}\right) + \rho \sum_{1 \le i \le j \le n} w_{ij} h_i h_j + \sum\nolimits_{i=1}^{n} D_i(h_{-i})$$

with

$$(24) \qquad D_i(h_{-i}) = \sum\nolimits_{i=1}^{n} \log\binom{N_i}{h_i} - \sum\nolimits_{k=0}^{N} \left\{[\lambda_i(h_{-i})]^{k} / k!\right\} \exp[-\lambda_i(h_{-i})]$$

then the joint distribution consistent with these truncated conditional Poissons is again given by (6). So Geyer and Thompson (1992) can again be used for maximum likelihood estimation of the parameters.

References:

Cressie, N. (1993) *Statistics for Spatial Data*, New York: Wiley.

Geyer, C.J. and E.A. Thompson (1992) “Constrained Monte Carlo Maximum Likelihood for Dependent Data”, *Journal of the Royal Statistical Society B*, 54: 657-699.

Augustin, N.H., et al. (2004) "Using the truncated auto-Poisson model for spatially correlated counts", http://www.stats.gla.ac.uk/Research/TechRep2004/04-14.pdf