Lecture 7
Autoregressive Processes in Space

Dennis Sun
Stanford University
Stats 253

July 8, 2015

1. Last Time
2. Autoregressive Processes in Space
3. Estimating Parameters
4. Testing for Spatial Autocorrelation
5. Application to the SIDS Data

AR Processes in Time

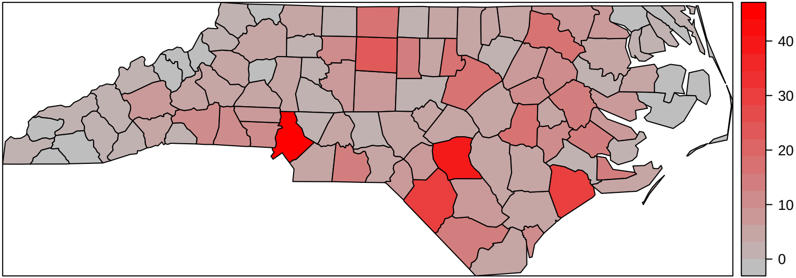
- Rather than model the covariance between errors explicitly, we assumed that the errors followed an AR(p) process:

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \ldots + \phi_p \epsilon_{t-p} + \delta_t.$$

- This induced a covariance structure between the errors.
- Estimation of ϕ is easy:
  - Under the "hack" approach, you will have estimates $\hat{\epsilon}_t$ of the errors, and you can estimate ϕ by regressing $\hat{\boldsymbol{\epsilon}}$ on lagged versions of itself.
  - If you follow the model-based approach, optimization over ϕ is not difficult because $\Sigma_{\boldsymbol{\phi}}^{-1}$ is banded.

Sudden Infant Death Syndrome (SIDS) Data

```
library(spdep)
example(nc.sids)
gr.colors <- colorRampPalette(c("gray", "red"))
spplot(nc.sids, "SID74", col.regions=gr.colors(100))
```



AR processes have traditionally been used to model **lattice data** (or **areal data**), like this.

Generalizing AR Processes to Space

There are two equivalent ways to specify a temporal AR process:

- By defining the variables in terms of each other:

$$\epsilon_t = \sum_{j=1}^{p} \phi_j \epsilon_{t-j} + \delta_t,$$

  where $\delta_t \overset{iid}{\sim} N(0, \tau^2)$.

- By specifying the conditional distribution:

$$p(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, ...) \propto \exp\left\{ -\frac{1}{2\tau^2} \left(\epsilon_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j} \right)^2 \right\}.$$

Both can be generalized naturally to space.

Generalizing AR Processes to Space

- **Simultaneous Autoregression** (SAR):

$$\epsilon_s = \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} \epsilon_{s'} + \delta_s,$$

where $N(s)$ denotes the neighbors of s and $\delta_s \overset{iid}{\sim} N(0, \tau^2)$.

- **Conditional Autoregression** (CAR):

$$p(\epsilon_s | \boldsymbol{\epsilon}_{-s}) \propto \exp\Big\{ -\frac{1}{2\tau^2}\Big(\epsilon_s - \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} \epsilon_{s'}\Big)^2 \Big\}.$$

Are they the same?

Simultaneous Autoregression (SAR)

$$\epsilon_s = \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} \epsilon_{s'} + \delta_s.$$

Let W be the matrix where $W_{ij} = 1/|N(i)|$ if $j \in N(i)$ and 0 otherwise. Then, we can write the SAR model as

$$\boldsymbol{\epsilon} = \phi W \boldsymbol{\epsilon} + \boldsymbol{\delta},$$

where $\boldsymbol{\delta} \sim N(\mathbf{0}, \tau^2 I)$, or equivalently

$$(I - \phi W)\boldsymbol{\epsilon} = \boldsymbol{\delta}.$$

Therefore, for SAR, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau^2 (I - \phi W)^{-1} (I - \phi W)^{-T})$.

Conditional Autoregression (CAR)

$$p(\epsilon_s|\boldsymbol{\epsilon}_{-s}) \propto \exp\Big\{ -\frac{1}{2\tau^2}\Big(\epsilon_s - \phi\frac{1}{|N(s)|}\sum_{s' \in N(s)} \epsilon_{s'}\Big)^2\Big\}.$$

CAR is a bit trickier.

For time series, we could order the data and obtain the joint distribution from the conditionals:

$$p(\epsilon_1, ..., \epsilon_n) = p(\epsilon_1) \cdot p(\epsilon_2|\epsilon_1) \cdot p(\epsilon_3|\epsilon_1, \epsilon_2) \cdot ... \cdot p(\epsilon_n|\epsilon_1, ..., \epsilon_{n-1}).$$

This trick doesn't work here because spatial data don't have a natural ordering.

Conditional Autoregression (CAR)

The following result gives us the joint distribution in terms of the conditionals, up to a normalizing constant:

Theorem (Brook's Lemma)

Let $p(\boldsymbol{\epsilon}) > 0$ *for all* $\boldsymbol{\epsilon}$*. Then, for any* $\boldsymbol{\epsilon}$ *and* $\boldsymbol{\epsilon}'$*:*

$$\frac{p(\boldsymbol{\epsilon})}{p(\boldsymbol{\epsilon}')} = \prod_{i=1}^{n} \frac{p(\epsilon_i | \epsilon_1, ..., \epsilon_{i-1}, \epsilon'_{i+1}, ..., \epsilon'_n)}{p(\epsilon'_i | \epsilon_1, ..., \epsilon_{i-1}, \epsilon'_{i+1}, ..., \epsilon'_n)}$$

Proof.

$$\begin{aligned}
\frac{p(\boldsymbol{\epsilon})}{p(\boldsymbol{\epsilon}')} &= \frac{p(\epsilon_1 | \epsilon'_2, ..., \epsilon'_n)}{p(\epsilon'_1 | \epsilon'_2, ..., \epsilon'_n)} \cdot \frac{p(\epsilon_2, ..., \epsilon_n | \epsilon_1)}{p(\epsilon'_2, ..., \epsilon'_n | \epsilon_1)} \\
&= \frac{p(\epsilon_1 | \epsilon'_2, ..., \epsilon'_n)}{p(\epsilon'_1 | \epsilon'_2, ..., \epsilon'_n)} \cdot \frac{p(\epsilon_2 | \epsilon_1, \epsilon'_3, ..., \epsilon'_n)}{p(\epsilon'_2 | \epsilon_1, \epsilon'_3, ..., \epsilon'_n)} \cdot \frac{p(\epsilon_3, ..., \epsilon_n | \epsilon_1, \epsilon_2)}{p(\epsilon'_3, ..., \epsilon'_n | \epsilon_1, \epsilon_2)} \\
&= \frac{p(\epsilon_1 | \epsilon'_2, ..., \epsilon'_n)}{p(\epsilon'_1 | \epsilon'_3, ..., \epsilon'_n)} \cdot \frac{p(\epsilon_2 | \epsilon_1, \epsilon'_3, ..., \epsilon'_n)}{p(\epsilon'_2 | \epsilon_1, \epsilon'_3, ..., \epsilon'_n)} \cdot ...
\end{aligned}$$

Conditional Autoregression (CAR)

Apply Brook's lemma to obtain $p(\boldsymbol{\epsilon})/p(\mathbf{0})$ for the CAR model:

$$\frac{p(\boldsymbol{\epsilon})}{p(\mathbf{0})} = \prod_{i=1}^{n} \frac{p(\epsilon_i | \epsilon_1, ..., \epsilon_{i-1}, 0_{i+1}, ..., 0_n)}{p(0_i | \epsilon_1, ..., \epsilon_{i-1}, 0_{i+1}, ..., 0_n)}$$

$$= \prod_{i=1}^{n} \frac{\exp\left\{ -\frac{1}{2\tau^2}\left(\epsilon_i - \phi \sum_{j<i} W_{ij}\epsilon_j - \phi \sum_{j>i} 0_j\right)^2 \right\}}{\exp\left\{ -\frac{1}{2\tau^2}\left(0_i - \phi \sum_{j<i} W_{ij}\epsilon_j - \phi \sum_{j>i} 0_j\right)^2 \right\}}$$

$$= \exp\left\{ -\frac{1}{2\tau^2} \sum_{i=1}^{n} \left(\epsilon_i - \phi \sum_{j<i} W_{ij}\epsilon_j\right)^2 + \frac{1}{2\tau^2} \sum_{i=1}^{n} \left(\phi \sum_{j<i} W_{ij}\epsilon_j\right)^2 \right\}$$

$$= \exp\left\{ -\frac{1}{2\tau^2} \sum_{i=1}^{n} \left(\epsilon_i^2 - 2\phi\epsilon_i \sum_{j<i} W_{ij}\epsilon_j\right) \right\}$$

If W is symmetric, then $2\sum_i \sum_{j<i} \epsilon_i W_{ij} \epsilon_j = \sum_i \sum_j \epsilon_i W_{ij} \epsilon_j$, so:

$$= \exp\left\{ -\frac{1}{2\tau^2} \boldsymbol{\epsilon}^T (I - \phi W) \boldsymbol{\epsilon} \right\}, \text{ so } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau^2 (I - \phi W)^{-1}).$$

Comparison of SAR and CAR

- Simultaneous Autoregression (SAR):

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau^2 (I - \phi W)^{-1} (I - \phi W)^{-T}).$$

- Conditional Autoregression (CAR). W must be symmetric,

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau^2 (I - \phi W)^{-1}).$$

Unlike with time series, the two specifications yield different models!

Extensions

- W can be any weight matrix in general. For example, we might...
  - give immediate neighbors more weight than two-hop neighbors.
  - weight pairs depending on the distance between them.
- $\mathrm{Var}[\boldsymbol{\delta}]$ does not have to be $\tau^2 I$.
  - It is common to assume that it is diagonal with different variances τ_i^2.
  - This is important when analyzing data aggregated by county/state, since each data point is based on a different sample size n_i.
  - In this case, we typically assume $\tau_i^2 \propto \frac{1}{n_i}$.
  - If $D = \mathrm{diag}(\tau_i^2)$, then the variance of SAR and CAR become $(I - W)^{-1} D (I - W)^{-T}$ and $(I - W)^{-1} D$, respectively.
  - For CAR, the requirement that W is symmetric needs to be changed accordingly.

Estimating ϕ

- Can we estimate ϕ by regressing ϵ_s on its neighbors?
- No! First, each observation may have a different number of neighbors.
- Even if we had regularly-spaced data where every observation has the same number of neighbors, Whittle (1954) showed that this estimator is inconsistent.

Maximum Likelihood

The log-likelihood for the CAR model is

$$-\log\det(\tau^2(I-\phi W)^{-1}) - \frac{1}{2\tau^2}(\mathbf{y}-X\boldsymbol{\beta})^T(I-\phi W)(\mathbf{y}-X\boldsymbol{\beta}).$$

It is possible to reduce to this to a partial likelihood in just ϕ by substituting the optimal values for β and τ^2:

$$\begin{aligned}\boldsymbol{\beta}(\phi) &= (X^T(I-\phi W)X)^{-1}X^T(I-\phi W)\mathbf{y}\\ \tau^2(\phi) &= \frac{1}{n}(\mathbf{y}-X\boldsymbol{\beta}(\phi))^T(I-\phi W)(\mathbf{y}-X\boldsymbol{\beta}(\phi)).\end{aligned}$$

Optimizing over ϕ is a one-dimensional problem that can be solved by grid search. Note that $\phi < 1/\lambda_1(W)$ is necessary to ensure that the covariance matrix is positive-definite.

Maximum Likelihood

The log-partial likelihood is

$$-\log\det(\tau(\phi)^2(I-\phi W)^{-1})-\frac{1}{2\tau(\phi)^2}(\mathbf{y}-X\boldsymbol{\beta}(\phi))^T(I-\phi W)(\mathbf{y}-X\boldsymbol{\beta}(\phi)).$$

W is usually sparse. The second term can be evaluated with just a few matrix-vector multiplications involving $(I-\phi W)$, which is easy to do.

The real challenge is evaluating $\log\det(I-\phi W)$. This matrix is no longer banded. But notice that

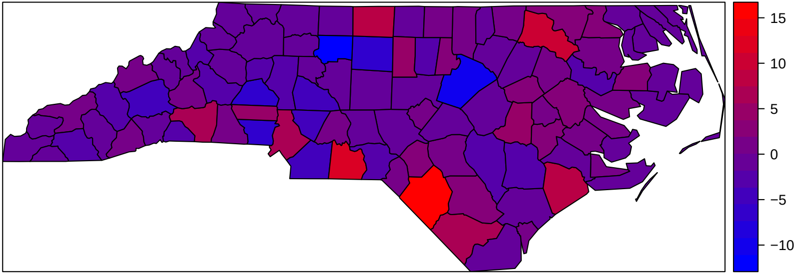
$$\log\det(I-\phi W)=\sum_{i=1}^{n}\log(1-\phi\lambda_i(W)),$$

so we do not need to evaluate the determinant for each ϕ we test. We just have to find the eigenvalues of W once.

Likelihood ratio test for testing $H_0 : \phi = 0$.

Model with Number of Births

Residuals

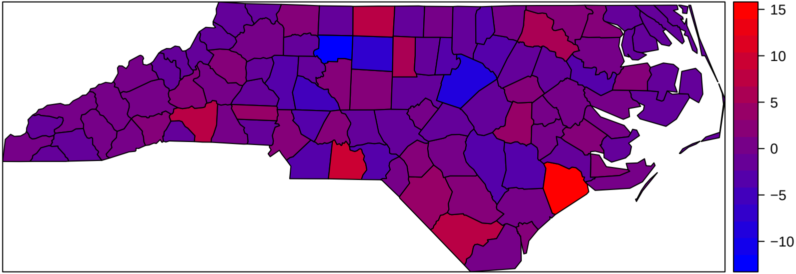


```
Call: spautolm(formula = SID74 ~ BIR74, data = nc.sids,
               listw = nb2listw(ncCR85_nb), family = "SAR")

Residuals:
      Min        1Q    Median        3Q       Max
-11.10079  -1.64522  -0.60629   1.24220  14.89254

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.96393971 0.66719077  1.4448   0.1485
BIR74       0.00173979 0.00010181 17.0890   <2e-16

Lambda: 0.3494 LR test value: 7.4243 p-value: 0.006435
Numerical Hessian standard error of lambda: 0.12092

Log likelihood: -276.4861
ML residual variance (sigma squared): 14.344, (sigma: 3.7874)
Number of observations: 100
Number of parameters estimated: 4
AIC: 560.97
```

Note that what they call “Lambda” is what we have called ϕ above.

Model with Numbers of Births and Nonwhite Births

Residuals



```
Call: spautolm(formula = SID74 ~ BIR74 + NWBIR74, data = nc.sids,
               listw = nb2listw(ncCR85_nb), family = "SAR")

Residuals:
     Min       1Q   Median       3Q      Max
-11.4951  -1.6394  -0.5963   1.3032  14.0163

Coefficients:
              Estimate Std. Error z value  Pr(>|z|)
(Intercept) 1.15912054 0.46252142  2.5061  0.012207
BIR74       0.00053403 0.00020572  2.5959  0.009433
NWBIR74     0.00357220 0.00055472  6.4396 1.198e-10

Lambda: 0.091006 LR test value: 0.38216 p-value: 0.53645
Numerical Hessian standard error of lambda: 0.14599

Log likelihood: -261.2314
ML residual variance (sigma squared): 10.859, (sigma: 3.2953)
Number of observations: 100
Number of parameters estimated: 5
AIC: 532.46
```