STP598sta: Spatiotemporal Analysis

Homework 1

Name: Your name; NetID: Your ID

Due 11:59pm Friday Sept 23 2022

Question 1

The *turning bands method* (Christakos, 1984; Stein, 1999a) is a technique for creating stationary covariance functions on $\mathbb{R}^r$. Let $\mathbf{u}$ be a random unit vector on $\mathbb{R}^r$ (by random we mean that the coordinate vector that defines $\mathbf{u}$ is randomly chosen on the surface of the unit sphere in $\mathbb{R}^r$). Let $c(\cdot)$ be a valid stationary covariance function on $\mathbb{R}^1$, and let $W(t)$ be a process on $\mathbb{R}^1$ having $c(\cdot)$ as its covariance function. Then for any location $\mathbf{s} \in \mathbb{R}^r$, define

$$Y(\mathbf{s}) = W(\mathbf{s}^T\mathbf{u}).$$

Note that we can think of the process either conditionally given $\mathbf{u}$, or marginally by integrating with respect to the uniform distribution for $\mathbf{u}$. Note also that $Y(\mathbf{s})$ has the possibly undersirable property that it is constant on planes (i.e. on $\mathbf{s}^T\mathbf{u} = k$).

(a) If W is a Gaussian process, show that, given $\mathbf{u}$, $Y(\mathbf{s})$ is also a Gaussian process and is stationary.

(b) Show that marginally $Y(\mathbf{s})$ is *not* a Gaussian process, but is isotropic. [Hint: Show that $\mathrm{Cov}(Y(\mathbf{s}, Y(\mathbf{s}') = \mathrm{E}_{\mathbf{u}}c((\mathbf{s}-\mathbf{s}')^T\mathbf{u})$.]

Question 2

Consider the `coalash` data frame in the `gstat` package in `R` and available from here. This data comes from the Pittsburgh coal seam on the Robena Mine Property in Greene County, PA (Cressie, 1993, p. 32). This data frame contains 208 coal ash core samples (the variable coal in the data frame) collected on a grid given by x and y planar coordinates (*not* latitude and longitude).

(a) Plot the sampled sites embedded on a map of the region. Add contour lines to the plot.

(b) Provide a descriptive summary (histograms, stems, quantiles, means, range, etc.) of the variable coal in the data frame.

(c) Plot variograms and correlograms of the response and comment on the need for spatial analysis here.

(d) If you think that there is need for spatial analysis, arrive at your best estimates of the range, nugget, and sill.

Question 3

Show that, for W symmetric, $\Sigma_{\mathbf{y}}^{-1} = D_w - \rho W$ is positive definite (thus resolving the impropriety) if $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$, where $\lambda_{(1)} < \lambda_{(2)} < \cdots < \lambda_{(n)}$ are the ordered eigenvalues of $D_w^{-1/2} W D_w^{-1/2}$.

Question 4

Consider the `Columbus` data, available from here, taken from Anselin (1988, p. 189) and also available within the `spdep` `R` package (but with possi- bly different variable names). These data record crime information for 49 neighborhoods in Columbus, OH, during 1980. Variables measured include NEIG, the neighborhood id value (1–49); HOVAL, its mean housing value (in \$1,000); INC, its mean household income (in \$1,000); CRIME, its number of residential burglaries and vehicle thefts per thousand households; OPEN, a measure of the neighborhood's open space; PLUMB, the percentage of housing units without plumbing; DISCBD, the neighborhood centroid's distance from the central business district; X, an x-coordinate for the neighborhood centroid (in arbitrary digitizing units, not polygon coordinates); Y , the same as X for the y-coordinate; AREA, the neighborhood's area; and PERIM, the perimeter of the polygon describing the neighborhood.

(a) Use `spdep` in `R` to construct adjacency matrices for the neighborhoods of Columbus based upon centroid distances less than

i. 25% of the maximum intercentroidal distances;
ii. 75% of the maximum intercentroidal distances.

(b) For each of the two spatial neighborhoods constructed above, use the `spautolm` function to fit SAR models with CRIME as the dependent variable, and HOVAL, INC, OPEN, PLUMB, and DISCBD as the covariates. Compare your results and interpret your parameter estimates in each case.

(c) Repeat part (b) for CAR models. Compare your estimates with those from the SAR model and interpret them.