Week 5.

Stat 222, Spatial Statistics. Lecture MWF 9am, Math-Sci 5203.
Professor: Rick Paik Schoenberg, frederic@ucla.edu, www.stat.ucla.edu/~frederic

DAY TEN. Monday, 4/30/01.

1) Parametric estimation of $2\gamma(h)$.

Assume we have some parametric model for $2\gamma(h)$, e.g. linear or spherical, and we want to estimate the parameters. We will let θ denote the vector of parameters to be estimated. For instance, in the linear variogram model $\theta = \{c_0, b_l\}$.

a) MLE.

(Maximum Likelihood Estimation.) Assume data (or rescaled data) $Z(s)$ are multivariate normal, with some means and covariance matrix Σ, where Σ is some function of θ. Then you can fit θ to the data by MLE. In practice this is done by writing the negative log-likelihood as a function of θ (the formula is on the top of p92 of Cressie), and finding the value of θ that minimizes this function.

(Actually the formula on p92 is a bit more general, since it includes the case where the mean of $Z(s)$ is a linear function of some covariates, $X(s)$. In our scenario we have some mean function but no covariates: to do MLE you'd simply plug in the mean wherever it says $X\beta$.)

Problem: the resulting estimates will be biased. Asymptotically unbiased, but for small samples, the bias can be substantial. Also, fitting parameters by MLE is computationally not trivial, requiring an iterative procedure.

b) VARIATIONS ON MLE.

(i) Restricted MLE. Instead of the negative loglikelihood of the data, look at the negative loglikelihood of the differences, e.g. $Z(i,j) - Z(i-1,j)$. Same problems as MLE.

(ii) MINQ. (Minimim Quadratic.) Instead of minimizing the negative loglikelihood function, minimize the quadratic norm $E[(\hat{\theta} - \theta)^2]$. We won't get into this, in this course.

c) OLS.

(Ordinary Least Squares.) Start with some nonparametric estimate of 2γ, like $2\hat{\gamma}$ or $2\bar{\gamma}$. Call it $2\gamma^{\#}$, to be general.

Given a value of θ, we know the variogram: call this $2\gamma^{\theta}(h)$.

Find θ minimizing $\sum_h [2\gamma^{\#}(h) - 2\gamma^{\theta}(h)]^2$. (Regression.)

Can sum over h's in all different directions.

Problem: values of $\gamma^{\#}(h)$ for h near 0 will tend to be less variable than those for h far from 0, so you might want to be sure you fit small h's betterm and not worry as much about how well you fit the large h's. So....

d) VARIATIONS ON OLS.

(i) GLS. (Generalize Least Squares.) You can find, for each θ, the covariance matrix V of all the $2\gamma^{\#}(h)$'s. The variances and covariances for the classical variogram estimator $2\hat{\gamma}(h)$ are given on p96, equations (2.6.10) and (2.6.11). Then you can minimize

$$(2\gamma^{\#} - 2\gamma^{\theta})^T V^{-1} (2\gamma^{\#} - 2\gamma^{\theta}), \tag{1}$$

where by $2\gamma^{\#}$ I mean the vector $2\gamma^{\#}(1), 2\gamma^{\#}(2), \ldots, 2\gamma^{\#}(k)$.

The nice thing about GLS is that the resulting estimates of θ will be weighted by the appropraite variances (and covariances) of the variogram estimates. The problem, though, is that it is VERY difficult computationally to minimize this function in (1). For each possible value of θ, not only does γ^{θ} depend on θ but also V^{-1} depends on θ!

(ii) WLS. (Weighted Least Squares.) Instead of V being the matrix with all variances and covariances of $2\gamma^{\#}$, just let V be the identity matrix, except instead of having 1s on the diagonal, let V have the variances of $2\gamma^{\#}$ on the diagonal. For the classical variogram estimator, these variances can be obtained from (2.6.11) on p96 of Cressie.

So, in WLS we ignore the covariances between our variogram estimates. This simplifies things a bit computationally.

(iii) APPROXIMATE WLS. What simplifies things **enormously** is the fact that the WLS estimates are very similar to those obtained by minimizing

$$\sum_h N(h)[\frac{\gamma^{\#}(h)}{\gamma^{\theta}(h)} - 1]^2.$$

Here $N(h)$ means the number of pairs of observations that are h apart.

Approximate WLS estimates are easy to compute. If you're worried about covariances between variogram estimates, you can plug in $\hat{\theta}$ obtained by approximate WLS into the formula for V^{-1}, and then do GLS. This is described in in the 2nd paragraph on p97 of Cressie. But generally speaking, approximate WLS is recommended because it performs well and is easy to compute.

DAY ELEVEN. Wednesday, 5/2/01.

1) Variogram estimates for coal-ash data.

p98 shows North-South variogram estimates (nonparametric and parametric) for the coal-ash data. The parametric model is the spherical variogram model.

How many lags should be looked at? A rule of thumb is to use only up to 1/2 the maximum possible lag, and also ensure for each h that $N(h) > 30$. Here, this means going up to lag 10.

2) Comparison of parametric variogram estimators.

(Read top of p100 of Cressie.)
a) WLS generally performs well.
b) Likelihood methods depend on the assumption of Gaussian data.
c) WLS does not depend highly on the Gaussian assumption.

3) Standard errors for variogram estimates.

For nonparametric variogram estimates, to approximate the variance of $2\gamma^{\#}(h)$, use

$$Var[\gamma^{\#}(h)] \approx \frac{2[2\gamma^{\#}(h)]^2}{N(h)}, \tag{2}$$

see (2.6.11), p96. This works for $2\hat{\gamma}$ and $2\bar{\gamma}$.

What about parametric estimates? One option is to plug in $\gamma^{\theta}(h)$ for $\gamma^{\#}(h)$ in (2).

It has been shown that estimates of θ are asymptotically normal for GLS, WLS, and OLS under various conditions.

Also, can use the bootstrap. Difficult for spatial data, because the data aren't independent, but sometimes you can take sub-regions that are thought to be far enough away from one another that they might be independent of each other, and re-sample these sub-regions to obtain simulated datasets. Then you can re-estimate the variogram using the simulated datasets and look at the sample variance of your estimates. Problems with this method: simulated data will have unrealistic edges where the sub-regions meet, and the small-scale variances for the simulated data will be identical to those for the original data.

4) Cross-validation.

To check the fit of the model for $2\gamma(h)$.
a) Delete one of your observations.
b) Calculate your estimate of θ, compute $2\gamma^{\theta}(h)$, and use this to predict the missing observation. (This prediction step is called Kriging; we'll learn how to do this later.) When you predict your missing observation, you obtain an estimate of σ^2, the prediction error assuming the model is correct.

c) Repeat a) and b), and at the end see how far off your predictions are, and compare with σ^2.

That is, look at the collection of standardized prediction errors $\{\frac{Z(s)-\hat{Z(s)}}{\sigma(s)}\}$, and examine the mean, RMS, and histogram (or stem-leaf plot) of these values. If the model fits well, the mean should be close to 0, the RMS should be close to 1, and the histogram should be roughly symmetric around 0. The results for the spherical N-S variogram for the coal-ash data are on p103. When the observation of 17.61 is omitted, the spherical model seems to fit well: the mean decreases from .147 to .081 and the RMS goes down from 1.167 to .957.

Note p104, 2nd paragraph: "[Cross-validation] cannot prove that the fitted model is correct, merely that it is not grossly incorrect."