Classical Geostatistical Methods

STATMOS Journal Club at NCSU
An-Ting Jhuang
2016.9.8
Overview

- Background
- Geostatistical model
- Provisional estimation of the mean function
- Estimation of the semivariogram
- Modeling the semivariogram
- Re-estimation of the mean function
- Kriging
Objective: use the sampled realization to make inferences about the process

Data type: point-referenced data

Spatial region: usually $D \subset R^2$

Spatial process: $Y(\cdot) \equiv \{Y(s) : s \in D\}$
The model can be expressed as

\[ Y(s) = \mu(s) + e(s) \]

Observation = Mean + Error

where \( Y(s) \) is the response process,
\( \mu(s) \) is the mean process, and
\( e(s) \) is the error process.
Geostatistical model

Two types of stationarity:
1. Second-order stationary
\[
\text{Cov}[e(s), e(t)] = C(s - t) \quad \forall s, t \in D
\]
2. Intrinsically stationary
\[
\frac{1}{2} \text{Var}[e(s) - e(t)] = \gamma(s - t) \quad \forall s, t \in D
\]
\[
\gamma(s - t) : \text{the semivariogram}
\]

2nd-order stationary $\Rightarrow$ intrinsically stationary
Provisional estimation of the mean function

- In regression, we consider $Y = X^T \beta + e$, then
  \[ \hat{\beta} = (X^T X)^{-1} X^T Y . \]

- Now, we take spatial variation into account. Let
  \[ \mu(s; \beta) = X(s)^T \beta, \]
  then
  \[ \hat{\beta}(s) = [X(s)^T X(s)]^{-1} X(s)^T Y(s). \]
Define the lag $h_u = (h_{u1}, h_{u2})^T$, $u=1, \ldots, k$.

The empirical semivariogram is

$$\hat{\gamma}(h_u) = \frac{1}{2N(h_u)} \sum_{s_i-s_j=h_u} [\hat{e}(s_i) - \hat{e}(s_j)]^2$$

$N(h_u)$: no. of times that lag $h_u$ occurs among the data locations.
Estimation of the semivarioogram

- When data locations are irregularly spaced, we can partition the lag space $H$ into lag classes $H_1, \ldots, H_k$.

$$\hat{\gamma}(h_u) = \frac{1}{2N(H_u)} \sum_{s_i - s_j \in H_u} [\hat{e}(s_i) - \hat{e}(s_j)]^2$$

- Common choices: polar partition and rectangular partition
Several issues:

1. How to choose number of bins
   \[ N(H_u) \geq 30, \| h_u \| < 0.5 \max_u \| h_u \|. \]

2. Sensitive to outliers
   \[ \text{Cressie and Hawkins (1980) provided a more robust estimator.} \]
Modeling the semivariogram

- To smooth the empirical semivariogram because
  a. often bumpy
  b. often fail to be conditionally nonpositive definite
  c. need estimates of the semivariogram at lags not included in the lag bins
Three sufficient and necessary conditions for a semivariogram model:
1. Vanishing at 0.
2. Evenness
3. Conditional negative definiteness

Five commonly-used types:
Spherical, exponential, Gaussian, Matern, and power semivariograms.
Modeling the semivariogram

![Graph showing the semivariance with points for range, sill, and nugget.](http://www.intechopen.com/books/advances-in-agrophysical-research/model-averaging-for-semivariogram-model-parameters)
Next, estimate the parameters in a chosen semivariogram.

Weighted least squares (WLS):

\[
\hat{\theta} = \arg\min_{\theta} \sum_{u \in U} \frac{N(h_u)}{\gamma(h_u; \theta)^2} [\hat{\gamma}(h_u) - \gamma(h_u; \theta)]^2
\]

Likelihood–based method will be discussed next time.
Re-estimation of the mean function

- Use estimated generalized least squares (EGLS)

- We have \( \hat{\Sigma} = (\hat{C}(s_i - s_j)) \), then

\[
\hat{\beta}_{EGLS} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} Y
\]
Kriging

- To predict the value $Y(\cdot)$ of at desired locations
- First developed and applied by D. G. Krige

- Two properties of predictor $\hat{Y}(s_0)$
  1. Linearity
  2. Unbiasedness
Universal kriging predictor

\[ \hat{Y}(s_0) = [\gamma + X(X^T \Gamma^{-1} X)^{-1}]^T \Gamma^{-1} Y \]

In practice, two possible modifications:
1. reduce the computing load
2. use the empirical \( \gamma \) and \( \Gamma \)
Thank you!