> *Since there are some problems about mathjax, I also upload the pdf version [../Spatial / 1 Statistics for Spatial Data].*

Introduction

Three Main Examples

Geostatistical Data

Geostatistics recognizes spatial variability at both the large scale (Spatial Trend) and the small scale (Spatial Correlation). The nearby data tend to be similar.

- Trend-Surface method include only large-scale variation, i.e. independent errors.
- Small variation is important, typically exhibiting strong positive correlation between data at nearby spatial locations.

Target:

- Predict the value at a new region from observed samples

Lattice Data

A *lattice* of locations evokes an idea of regularly spaced points in $\mathbb{R}^d$, linked to nearest neighbors, second-nearest neighbors. The regular lattice is analogue to a time series observed at equally spaced time points.

Comparison with geostatistical data:

- In contrast to geostatistical problems, data from lattice problems may be exhaustive of the phenomenon.

Point Patterns

Point patterns arise when the important variable to be analyzed is the location of "events".

Question:

- Whether the pattern is exhibiting complete spatial randomness, clustering, or regularility.

The *mark variable* is the covariate associated with each point and the whole process is then called a *marked spatial point process*.

Statistics for Spatial Data

Why consider the correlation structure is important?

Estimation

Suppose the data $Z(1), \cdots, Z(n)$ are *positively correlated* with a correlation that decreases as the separation between data increases:

$$\operatorname{cov}(Z(i), Z(j)) = \sigma_0^2 \cdot \rho^{|i-j|}, \quad i, j = 1, \ldots, n, 0 < \rho < 1$$

The covariance of $\bar{Z}$ is:

$$\begin{aligned} \operatorname{var}(\bar{Z}) =& n^{-2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}(Z(i), Z(j))\right\} \\ =& \left\{\sigma_0^2/n\right\}[1 + 2\{\rho/(1-\rho)\}\{1-(1/n)\} \\ & -2\{\rho/(1-\rho)\}^2\left(1-\rho^{n-1}\right)/n] \end{aligned}$$

Compared with the variance of $\bar{Z}$ with iid observations, *equivalent number of independent observations* is proposed:

$$n' \equiv n/[1 + 2\{\rho/(1-\rho)\}\{1-(1/n)\} - 2\{\rho/(1-\rho)\}^2\left(1-\rho^{n-1}\right)/n]$$

Prediction

The MSE of prediction with positive correlated data is

$$\begin{aligned} E(Z(n+1)-\bar{Z})^2 = \sigma_0^2\{1 + (1/n)[1 + 2\{\rho/(1-\rho)\}\{\rho^n - (1/n)\} \\ -2\{\rho/(1-\rho)\}^2\left(1-\rho^{n-1}\right)/n]\} \end{aligned}$$

Experimental Design

Suppose that experimental units are laid out in a $t \times b$ array made up of b blocks (columns), each with t units in them. Any experiment on calender treatments has to take this into account, as well as the possible spatial correlation running the length of the sheet.

Goal of Experimental Design:

- Find the allocation of treatments to units that will give the most precise estimates of (estimable) treatment effects.

Linear Models with Spatially Dependent Error

Suppose $Z(s)$ are generated by the random process

$$Z(s) = \sum_{l=1}^{q} \beta_l x_l(s) + \delta(s), \quad s \in D \subset \mathbb{R}^d$$

where $var(\delta) = \Sigma$.

Goal:

- Efficient estimation of β.
- $\hat{\boldsymbol{\beta}}_{\text{gls}} = \left(X' \Sigma^{-1} X\right)^{-1} X' \Sigma^{-1} \mathbf{Z}$

Summary

- **Universality**: All data have a more-or-less precise spatial and temporal label associated with them.
- **Otherness**: Whether the spatial labels are thought to be an important part of the modeling and analysis of the data is a concern that should be addressed problem by problem.
- **Relativity to Spatial**: Data that are close together in space (and time) are often more alike than those that are far apart.
- **Generality**: Nonspatial model is a special case of a spatial model.
- **Importance of Explanatory variables**: All explanatory variables should be included in the mean structure first. A missed spatial-relevant variable will contribute to the spatial dependence and cause misspecification.