BST 6200 Spatial Statistics and Disease Mapping

Section 6.8 Kriging

Steven E. Rigdon

Spring 2020

Section 6.8: Kriging

The model:

Let $z(\boldsymbol{x})$ denote the outcome at location $\boldsymbol{x}$, where $\boldsymbol{x}$ is a vector (i.e., a point in the plane).

Let $f(\boldsymbol{x}_i)$ denote the deterministic (not random) trend function evaluated at $\boldsymbol{x}_i$.

$$z(\boldsymbol{x}_i) = f(\boldsymbol{x}_i) + \nu(\boldsymbol{x}_i) + \epsilon_i$$

where $\nu()$ is a **random function** and $\epsilon_i, \quad i = 1, 2, \ldots$ are i.i.d. $N(0, \sigma^2)$. [More on random functions coming up; this is a fairly deep and slippery concept!]

The Model for Kriging

$$\underbrace{z(\boldsymbol{x}_i)}_{\text{Outcome}} = \underbrace{f(\boldsymbol{x}_i)}_{\text{True Function}} + \underbrace{\nu(\boldsymbol{x}_i)}_{\text{Random Function}} + \underbrace{\epsilon_i}_{N(0,\sigma^2)\text{ error}}$$

Multivariate Normal Distribution

$\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$. $\boldsymbol{X} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma)$ means that $\boldsymbol{X}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ with PDF

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)$$

Mean, Variance, Covariance, and Correlation of Random Variables

$$E(X) = \int_{-\infty}^{\infty} x \; p(x) \; dx = \mu_X$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \; p(x) \; dx$$

$$V(X) \stackrel{\text{def}}{=} E\left[(X - \mu_X)^2\right] = E(X^2) - \mu_X^2 = \sigma_X^2$$

$$\text{covariance } = \sigma_{X,Y} = \operatorname{cov}(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

$$\text{correlation } = \rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Mean and Covariance of Multivariate Normal Distribution

If $\boldsymbol{X} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma)$ then

$$E(\boldsymbol{X}) = \boldsymbol{\mu}$$

and

$$\text{cov}(\boldsymbol{X}) \stackrel{\text{def}}{=} \begin{bmatrix} V(X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_1, X_2) & V(X_2) & \text{cov}(X_2, X_3) \\ \text{cov}(X_1, X_3) & \text{cov}(X_2, X_3) & V(X_3) \end{bmatrix} = \Sigma$$

Random Function in One Variable

Spatial data analysis requires us to consider random functions in two dimensions, because the plane is two dimensional. Let's begin with random functions in one dimension to get the concept down.

Suppose x_1, x_2, x_3 are three points on a (one-dimensional) number line, and let $\nu(x_1), \nu(x_2), \nu(x_3)$ denote the outcome at these three points.

Suppose

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 \end{bmatrix} \right)$$

What about the covariance (or correlation) between the random function at two nearby x-points?

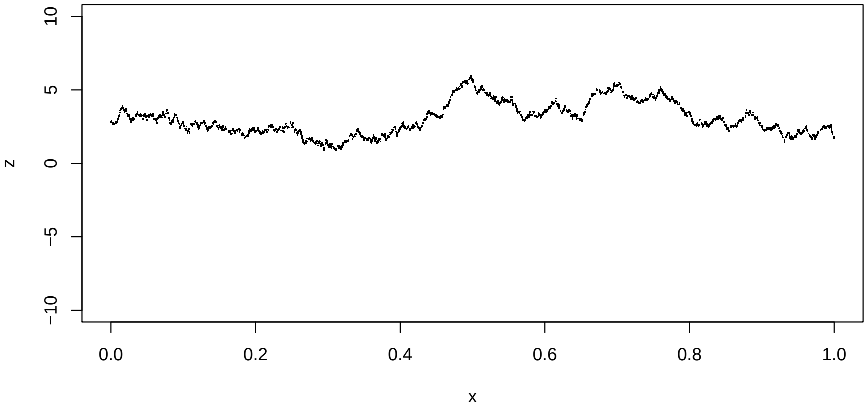
We make the assumption that the correlation between the outcomes ν_1 and ν_2 at points x_1 and x_2, respectively, is a function of the distance between them:

$$\rho_{ij} = \rho(d)$$

where d is the distance between x_1 and x_2.

Note, in the case of a single variable (as we've assumed here), $d = |x_1 - x_2|$. In the case of points in the plane, $d = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$, the distance between the two points.

Using R to Visualize a Random Function



Back to Vectors in the Plane

The correlation between the outcomes ν_1 and ν_2 at points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively, is a function of the distance between them:

$$\rho_{ij} = \rho(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)$$

If the outcomes $[\nu(x_1), \nu(x_2), \ldots, \nu(x_p)]'$ are multivariate normal, then the limit as $n \to \infty$ of this process is said to be a **Gaussian random field**.

A Gaussian random field is **isotropic**, which means that the covariance function between points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ depends only on the distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and not on the direction.

A function that does not have this property is said to be **anisotropic**.

Assumptions Behind Kriging

$$\underbrace{z(\boldsymbol{x}_i)}_{\text{Outcome}} = \underbrace{f(\boldsymbol{x}_i)}_{\text{True Function}} + \underbrace{\nu(\boldsymbol{x}_i)}_{\text{Random Function}} + \underbrace{\epsilon_i}_{N(0,\sigma^2)\text{ error}}$$

Theory behind the variogram and the semi-variogram

A few facts to recall ...

$$V(cX) = c^2 V(X)$$

$$V(X+Y) = V(X) + V(Y) + 2\text{cov}(X,Y)$$

$$V(X-Y) = V(X) + V(Y) - 2\text{cov}(X,Y)$$

The Variogram

Let $\boldsymbol{x}_1$ and $\boldsymbol{x_2}$ be distinct points in the plane.

$$V(\nu(\boldsymbol{x_1}) - \nu(\boldsymbol{x_2})) =$$

$$V(\nu(\boldsymbol{x_1}) - \nu(\boldsymbol{x_2})) = 2\, \underbrace{\sigma^2(1 - \rho(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|))}_{\text{Semi-variogram: } \gamma(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)}$$