Introduction to geostatistics II

Applications:

- hydrological data,
- mining applications,
- air quality studies
- soil science data
- biological applications
- economic housing data
- etc.

Geostatistics for Environmental Scientists , R. Webster & M. A. Olvier, Wiley 2001.

Introduction to geostatistics III

Lets consider:

- J physical locations $\{x_j\}_{j=1,\cdots,J}$,
- Some information of interest (e.g. radioactivity levels) is modeled as a stochastic process at these locations $\{s_j = s(x_j)\}_{j=1,\cdots,J}$,
- One observation (or measurement) is available for each site $\{s_j^{(1)}\}_{j=1,\cdots,J}$.
- At a new location x_0, we want to predict s_0.

Introduction to geostatistics IV

Content:

1. Adhoc methods: spatial interpolation
2. Statistical modeling with Kriging
    1. simple, ordinary and universal Kriging
    2. modelling with covariance and/or variogram
    3. Stationarity assumptions

Spatial interpolation I

Prediction of s at a new site x_0 can be expressed as a weighted averages of data:

$$s(x_0) = \sum_{j=1}^{J} \lambda_j \; s(x_j)$$

with the constraints:

$$(\lambda_j \geq 0, \; \forall j) \;\wedge\; \left(\sum_{j=1}^{J} \lambda_j = 1 \right)$$

Spatial interpolation II

1. Thiessen polygones (Voronoi polygons, Dirichlet tesselations)
2. Triangulation
3. Natural Neighbour interpolation
4. Inverse function of distance
5. Trend surface

Spatial interpolation III

1. Thiessen polygones (Voronoi polygons, Dirichlet tesselations)

$$s(x_0) = s(x_j) \quad \text{with} \quad x_j = \arg \min_{i=1,\cdots,J} \|x_i - x_0\|$$

so we have binary weights:

$$\lambda_j = 1, \quad \text{and} \quad \lambda_i = 0 \; \forall i \neq j$$

Spatial interpolation IV

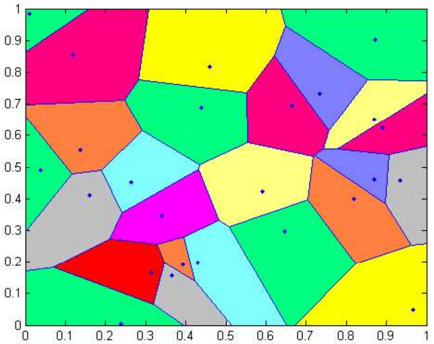


Figure: Voronoi polygons

Spatial interpolation V

2. Triangulation. Sampling points are linked to their neighbours by straight lines to create triangles that do not contain any of the points. Having a new position $x_0 = (u_0, v_0)$ in one of the triangle, let says the one defined by (x_1, x_2, x_3), then

$$\lambda_1 = \frac{|x_0 - x_3 \; ; \; x_2 - x_3|}{|x_1 - x_3 \; ; \; x_2 - x_3|}$$

with the notation

$$|x_0 - x_3 \; ; \; x_2 - x_3| = \begin{vmatrix} u_0 - u_3 & u_2 - u_3 \\ v_0 - v_3 & v_2 - v_3 \end{vmatrix} = \det\left(\begin{bmatrix} u_0 - u_3 & u_2 - u_3 \\ v_0 - v_3 & v_2 - v_3 \end{bmatrix}\right)$$

λ_2 and λ_3 are defined in a similar fashion and all the other λs are 0s. Unlike Thiessen method, the resulting surface is continuous but yet has abrupt changes in gradient at the margins of the triangles.

Spatial interpolation VI

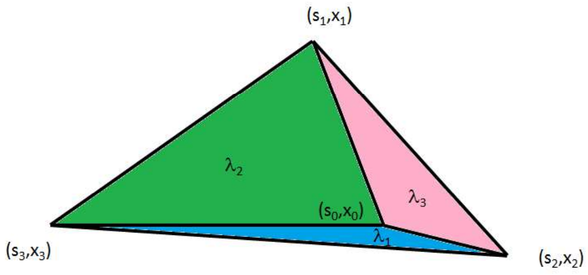
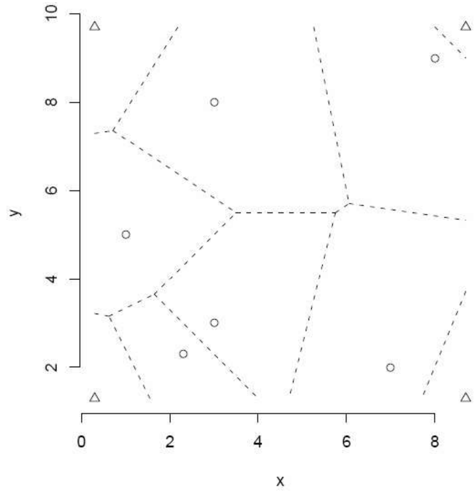


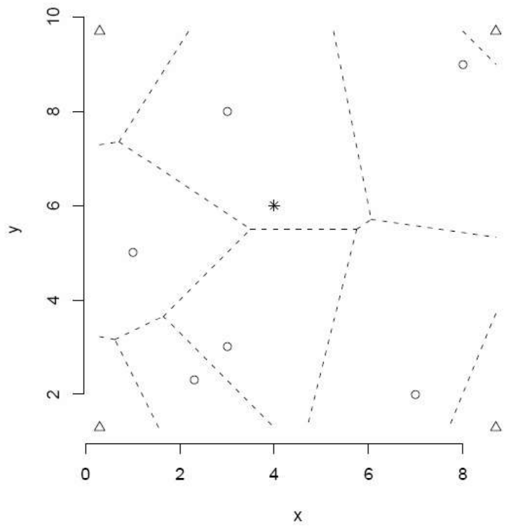
Figure: Triangulation: The weights λ_1 corresponds to the blue area divided by the area of the triangle (x_1, x_2, x_3). Similarly λ_2 corresponds to the green area divided by the area of the triangle (x_1, x_2, x_3) and λ_3 corresponds to the pink area divided by the area of the triangle (x_1, x_2, x_3).

Spatial interpolation VII

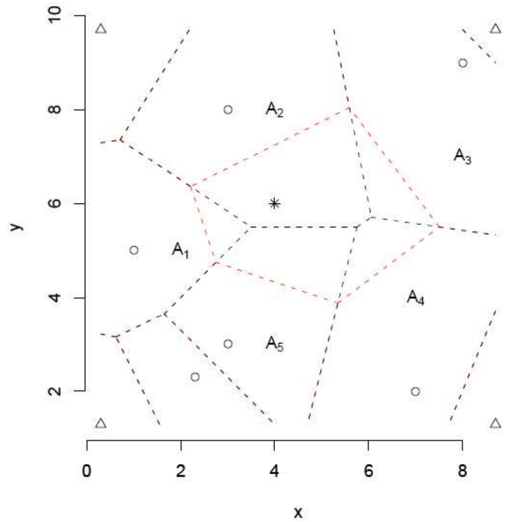
3. Natural Neighbour interpolation



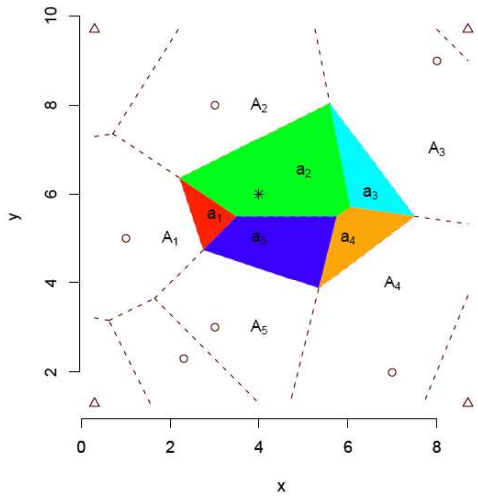
Spatial interpolation VIII



Spatial interpolation IX



Spatial interpolation X



Spatial interpolation XI

$$\lambda_j = \frac{a_j}{\sum_{j=1}^{J} a_j}$$

with $a_j = 0$ if x_j is not a natural neighbour to x_0.

Spatial interpolation XII

4. Inverse function of distance

$$\lambda_j \propto \frac{1}{\|x_j - x_0\|^\beta}, \quad \beta > 0$$

- The weights $\{\lambda_j\}_{j=1,\cdots,J}$ are scaled such that they sum up to 1.
- Usually, $\beta = 2$ (Euclidian distance).
- If $x_0 = x_j$, then $s(x_0) = s(x_j)$.
- There are no discontinuities in the map s.
- There is no measure of the error.

Spatial interpolation XIII

5. Trend Surface. This method proposes to do regression:

$$s(x) = \mu(x) + \epsilon$$

with the error term $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The function μ is a parametric function such as planes or quadratics e.g.

$$\mu(x = (u, v)) = b_0 + b_1\ u + b_2\ v$$

Coefficients $\mathbf{b} = (b_0, b_1, b_2)^T$ can then be estimated by Least Squares using the J observations.
Once $\hat{\mathbf{b}}$ is estimated, the prediction at the new location x_0 is computed by:

$$\hat{s}(x_0) = \hat{b}_0 + \hat{b}_1\ u_0 + \hat{b}_2\ v_0$$

Spatial interpolation XIV

Limits of Interpolation for prediction:

- Some interpolators give a crude prediction and the spatial variation is displayed poorly.
- The interpolators fail to provide any estimates of the error on the prediction.
- With the exception of trend surface, these methods were deterministic. However the processes are stochastic by nature.
- In practice the modelling with trend surface is too simplistic to perform well and the uncertainty is the same everywhere.

Introduction: Kriging I

- The aim of Kriging is to estimate the value of a random variable s at one or more unsampled points or locations, from more or less sparse sample data on a given support say $\{s(x_1),\cdots,s(x_J)\}$ at $\{x_1,\cdots,x_J\}$.
- Different kinds of kriging methods exist, which pertains to the assumptions about the mean structure of the model:

$$\mathbb{E}[s(x)] = \mu(x) \quad \text{or} \quad \mathbb{E}[\underbrace{s(x)-\mu(x)}_{\tilde{s}(x)}] = 0$$

Introduction: Kriging II

- Different Kriging methods:
  - Ordinary Kriging :
    $$\mathbb{E}[s(x)] = \mu \quad (\mu \text{ is unknown})$$
  - Simple Kriging :
    $$\mathbb{E}[s(x)] = \mu \quad (\mu \text{ is known})$$
  - Universal Kriging : the mean is unknown and depends on a linear model:
    $$\mu(x) = \sum_{p=0}^{P} \beta_p\, \phi_p(x)$$
    and coefficients $\{\beta_p\}$ need to be estimated.

Ordinary kriging I

- Ordinary kriging is the most common type of kriging.
- The underlying model assumption in ordinary kriging is:
  $$\mathbb{E}[s(x)] = \mu$$
  with μ unknown.
- The stochastic process s has been observed at J sites (the r.v. $s(x_j) = s_j$ has one observation $s_j^{(1)}$ associated with it).

Ordinary kriging II

- The model for $s(x_0)$ is:

$$s(x_0) - \mu = \sum_{j=1}^{J} \lambda_j \, (s(x_j) - \mu) + \epsilon(x_0)$$

or

$$s(x_0) = \sum_{j=1}^{J} \lambda_j \, s(x_j) + \mu \, (1 - \sum_{j=1}^{J} \lambda_j) + \epsilon(x_0)$$

We filter the unknown mean by requiring that the kriging weights sum to 1, leading to the ordinary kriging estimator :

$$s(x_0) = \sum_{j=1}^{J} \lambda_j \, s(x_j) + \epsilon(x_0) \quad \text{subject to} \sum_{j=1}^{J} \lambda_j = 1$$

Ordinary kriging III

- $\epsilon(x_0)$ is the noise at position x_0 such that:

$$\mathbb{E}[\epsilon(x_0)] = 0$$

- We want to estimate $\hat{s}(x_0)$. In other words we need to get the appropriate $\{\lambda_j\}_{j=1,\cdots,J}$.
- Estimation by Mean square errors subject to a constraint:

$$(\hat{\lambda}_1, \cdots, \hat{\lambda}_J) = \arg \min_{\lambda_1,\cdots,\lambda_J} \{\mathbb{E}[\epsilon(x_0)^2]\} \quad \text{subject to} \sum_{j=1}^{J} \lambda_j = 1$$

Ordinary kriging IV

- This is solved using Lagrange multipliers. We define the energy $\mathcal{J}$ that depends both on $\{\lambda_j\}_{j=1,\cdots,J}$ and ψ:

$$(\{\hat{\lambda}_j\}_{j=1,\cdots,J}, \hat{\psi}) = \arg \min_{\psi,\lambda_1,\cdots,\lambda_J} \left\{ \mathcal{J}(\lambda_1, \cdots, \lambda_J, \psi) = \mathbb{E}[\epsilon(x_0)^2] + 2\,\psi \left(\sum_{j=1}^{J} \lambda_j - 1 \right) \right\}$$

Ordinary kriging V

1. First we express the expectation of the error:

$$\begin{aligned}
\mathbb{E}[\epsilon(x_0)^2] &= \mathbb{E}\left[\left(s(x_0) - \textstyle\sum_{j=1}^{j} \lambda_j \, s(x_j) \right)^2 \right] \\
&= \mathbb{E}\left[\left(s(x_0) - \mu + \mu - \textstyle\sum_{j=1}^{J} \lambda_j \, s(x_j) \right)^2 \right] \\
&= \mathbb{E}\left[(s(x_0) - \mu)^2 \right] - 2 \textstyle\sum_{j=1}^{J} \lambda_j \mathbb{E}\left[(s(x_0) - \mu)\,(s(x_j) - \mu) \right] \\
&\quad + \textstyle\sum_{i=1}^{J} \sum_{j=1}^{J} \lambda_i \lambda_j \mathbb{E}\left[(s(x_j) - \mu)\,(s(x_i) - \mu) \right]
\end{aligned}$$

Remember that the covariance is defined as

$$\mathbb{C}\mathrm{ov}(s(x_j); s(x_i)) = c_{ij} = \mathbb{E}\left[(s(x_j) - \mu)\,(s(x_i) - \mu) \right]$$

So the energy to minimize:

$$\mathcal{J}(\lambda_1, \cdots, \lambda_J, \psi) = c_{00} - 2 \sum_{j=1}^{J} \lambda_j c_{0j} + \sum_{i=1}^{J} \sum_{j=1}^{J} \lambda_i \lambda_j c_{ij} + 2\,\psi \left(\sum_{j=1}^{J} \lambda_j - 1 \right)$$

Ordinary kriging VI

2. Second, we differentiate $\mathcal{J}$ w.r.t. $\lambda_k,\ k = 1, \cdots, J$ and ψ, and the minimum of $\mathcal{J}$ is found when all the derivatives are equal to zeros.

$$\begin{cases} \frac{\partial \mathcal{J}}{\partial \psi} = 0 \\ \\ \frac{\partial \mathcal{J}}{\partial \lambda_k} = 0, \quad \forall k = 1, \cdots, J \end{cases}$$

The derivative w.r.t. ψ is:

$$\frac{\partial \mathcal{J}}{\partial \psi} = \sum_{j=1}^{J} \lambda_j - 1 = 0$$

The derivative w.r.t. λ_k is :

$$\frac{\partial \mathcal{J}}{\partial \lambda_k} = 2\,\psi - 2\,c_{0k} + 2 \sum_{j=1}^{J} \lambda_j\ c_{ik} = 0$$

Ordinary kriging VII

3. The solution is:

$$\underbrace{\begin{bmatrix} c_{11} & \cdots & c_{1J} & 1 \\ \vdots & & & 1 \\ c_{J1} & \cdots & c_{JJ} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}_{\mathrm{A}} \underbrace{\begin{pmatrix} \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_J \\ \hat{\psi} \end{pmatrix}}_{\boldsymbol{\lambda}} = \underbrace{\begin{pmatrix} c_{10} \\ \vdots \\ c_{J0} \\ 1 \end{pmatrix}}_{\mathbf{b}}$$

or

$$\hat{\boldsymbol{\lambda}} = \mathrm{A}^{-1}\mathbf{b}$$

Ordinary kriging VIII

- Once you have the estimate $\hat{\boldsymbol{\lambda}}$, then you can predict (using the observations):

$$\hat{s}(x_0) = \sum_{j=1}^{J} \hat{\lambda}_j\ s_j^{(1)}$$

Simple Kriging I

Assumption for Simple Kriging:

- The mean $\mathbb{E}[s(x)] = \mu$ is known.

We estimate $s(x_0)$ using the relation (same as Ordinary Kriging):

$$s(x_0) - \mu = \sum_{j=1}^{J} \lambda_j\ (s(x_j) - \mu) + \epsilon(x_0)$$

or

$$s(x_0) = \sum_{j=1}^{J} \lambda_j\ s(x_j) + \mu \left(1 - \sum_{j=1}^{J} \lambda_j\right) + \epsilon(x_0)$$

where μ is known. The λ_j do not need to be constrained to sum to 1 anymore and the second term insured that $\mathbb{E}[s(x)] = \mu,\ \forall x$.

Simple Kriging II

The hypothesis for the error is $\mathbb{E}[\epsilon(x_0)] = 0$ and we estimate $\{\lambda_j\}_{j=1,\cdots,J}$ such that the Mean Square Error $\mathbb{E}[\epsilon^2(x_0)]$ is minimised.
The solution is then:

$$\begin{pmatrix} \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_J \end{pmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1J} \\ \vdots & & \\ c_{J1} & \cdots & c_{JJ} \end{bmatrix}^{-1} \begin{pmatrix} c_{10} \\ \vdots \\ c_{J0} \end{pmatrix}$$

Once you have the estimate $\hat{\boldsymbol{\lambda}}$, then you can predict (using the observations):

$$\hat{s}(x_0) = \sum_{j=1}^{J} \hat{\lambda}_j \ s_j^{(1)} + \mu \left(1 - \sum_{j=1}^{J} \hat{\lambda}_j\right)$$

Universal Kriging I

For a new location x_0, we have the following model

$$s(x_0) - \mu(x_0) = \sum_{j=1}^{J} \lambda_j \ (s(x_j) - \mu(x_j)) + \epsilon_{x_0}$$

or

$$s(x_0) = \sum_{j=1}^{J} \lambda_j \ s(x_j) + \mu(x_0) - \sum_{j=1}^{J} \lambda_j \ \mu(x_j) + \epsilon_{x_0}$$

In the Universal Kriging, the mean of s depends on the position x:

$$\mu(x) = \sum_{p=0}^{P} \beta_p \ \phi_p(x)$$

Universal Kriging II

Example of choice of the functions $\{\phi_p\}_{p=1,\cdots,P}$ of $x = (u, v) \in \mathbb{R}^2$:

- Linear trend ($P = 2$):

$$\phi_0(x) = 1, \quad \phi_1(x) = u, \quad \phi_2(x) = v$$

- Quadratic trend ($P = 5$):

$$\phi_3(x) = u^2, \quad \phi_4(x) = uv, \quad \phi_5(x) = v^2$$

Universal Kriging III

In a similar fashion as ordinary kriging (we dont know the β_p, so μ is unknown), we rewrite:

$$s(x_0) = \sum_{j=1}^{J} \lambda_j \ s(x_j) + \mu(x_0) - \sum_{j=1}^{J} \lambda_j \ \mu(x_j) + \epsilon_{x_0}$$

as

$$s(x_0) = \sum_{j=1}^{J} \lambda_j \ s(x_j) + \epsilon_{x_0} \quad \text{subject to} \quad \underbrace{\mu(x_0) - \sum_{j=1}^{J} \lambda_j \ \mu(x_j) = 0}_{\text{constraint}}$$

Universal Kriging IV

Having $\mu(x) = \sum_{p=0}^{P} \beta_p \, \phi_p(x)$, the constraint is equivalent to:

$$\sum_{p=0}^{P} \beta_p \, \phi_p(x_0) = \sum_{p=0}^{P} \beta_p \sum_{j=1}^{J} \lambda_j \, \phi_p(x_j)$$

This is true for any combination of β_p. Hence we have in fact $P+1$ constraints:

$$\left(\phi_p(x_0) = \sum_{j=1}^{J} \lambda_j \, \phi_p(x_j) \right) \quad \forall p = 0, \cdots, P$$

Universal Kriging V

- Note that at $p = 0$, using $\phi_0(x) = 1$, we recover the constraint $\sum_{j=1}^{J} \lambda_j = 1$.
- This minimisation is solved by introducing $P+1$ Lagrange multipliers:

$$\left(\hat{\lambda}_1, \cdots, \hat{\lambda}_J, \hat{m}_0, \cdots, \hat{m}_P \right) = \arg\min \left\{ \mathbb{E}[\epsilon_{x_0}^2] + \sum_{p=0}^{P} m_p \left(\phi_p(x_0) - \sum_{j=1}^{J} \lambda_j \, \phi_p(x_j) \right) \right\}$$

Universal Kriging VI

The solution of Universal Kriging is:

$$\begin{bmatrix} \mathrm{C} & F^T \\ F & 0 \end{bmatrix} \begin{pmatrix} \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_J \\ \hat{m}_0 \\ \hat{m}_1 \\ \vdots \\ \hat{m}_P \end{pmatrix} = \begin{pmatrix} c_{01} \\ \vdots \\ c_{0J} \\ \phi_0(x_0) \\ \phi_1(x_0) \\ \vdots \\ \phi_P(x_0) \end{pmatrix}$$

Universal Kriging VII

with

$$F = \begin{bmatrix} \phi_0(x_1) & \phi_0(x_2) & \cdots & \phi_0(x_J) \\ \phi_1(x_1) & \phi_1(x_2) & \cdots & \phi_1(x_J) \\ \phi_2(x_1) & \phi_2(x_2) & \cdots & \phi_2(x_J) \\ \vdots & & & \\ \phi_P(x_1) & \phi_P(x_2) & \cdots & \phi_P(x_J) \end{bmatrix}$$

and

$$\mathrm{C} = \begin{bmatrix} c_{11} & \cdots & c_{1J} \\ \vdots & & \\ c_{J1} & \cdots & c_{JJ} \end{bmatrix}$$

The prediction at x_0 is computed by:

$$\hat{s}_0 = \sum_{j=1}^{J} \hat{\lambda}_j \, s_j^{(1)}$$

Remarks about Kriging I

- The covariance function $\mathbb{C}[s(x), s(x')]$ is a function assumed to be known in all the solutions proposed here for Kriging.
- In practice such a function is not known and need to be estimated.
- Some hypotheses will be used about the process to ease this estimation.
- We define the concept of variogram as an alternative to covariance and we see next what are the assumptions about the process that will be used for Kriging.

Remarks about Kriging II

Definition (variogram)

The variogram is defined as:

$$\gamma(s(x_i), s(x_j)) = \gamma_{ij} = \frac{1}{2}\mathbb{E}[(s(x_i) - s(x_j))^2]$$

When $\mathbb{E}[(s(x_i) - s(x_j))] = 0$ then the variogram is linked to the covariance as follow:

$$\gamma_{ij} = \frac{1}{2}\left(c_{ii} + c_{jj} - 2\, c_{ij}\right)$$