Kriging formulation with the variogram I

In Kriging, we have

\[
\frac{s(x_0) - \mu(x_0)}{\tilde{s}(x_0)} = \sum_{j=1}^{J} \lambda_j \frac{s(x_j) - \mu(x_j)}{\tilde{s}(x_j)} + \epsilon(x_0)
\]

or

\[
\epsilon(x_0) = \tilde{s}(x_0) - \sum_{j=1}^{J} \lambda_j \tilde{s}(x_j)
\]

with the assumption \(E[\epsilon(x_0)] = 0\).

Kriging formulation with the variogram II

The estimation is performed by:

\[
(\hat{\lambda}_1, \cdots, \hat{\lambda}_J) = \arg \min E[\epsilon(x_0)^2]
\]

solved

- without constraint: Simple Kriging
- with constraints: Universal/Ordinary Kriging

and the estimate is computed using the observations:

\[
\hat{s}_0 = \sum_{j=1}^{J} \hat{\lambda}_j \hat{s}_j^{(1)} \quad \text{or} \quad \hat{s}_0 = \sum_{j=1}^{J} \hat{\lambda}_j \hat{s}_j^{(1)} + \mu (1 - \sum_{j=1}^{J} \hat{\lambda}_j)
\]

Ordinary/Universal Kriging

Simple Kriging

Kriging formulation with the variogram III

Exercise: With the constraint \(\sum_{j=1}^{J} \lambda_j = 1\) (true in Universal/Ordinary Kriging):

- Show that:

\[
\left( \tilde{s}_0 - \sum_{j=1}^{J} \lambda_j \tilde{s}_j \right)^2 = -\frac{1}{2} \sum_{j=1}^{J} \sum_{i=1}^{J} \lambda_i \lambda_j (\tilde{s}_i - \tilde{s}_j)^2 + \sum_{j=1}^{J} \lambda_j (\tilde{s}_0 - \tilde{s}_j)^2
\]

- Show that

\[
E[\epsilon(x_0)^2] = -\sum_{j=1}^{J} \sum_{i=1}^{J} \lambda_i \lambda_j \gamma_{ij} + 2 \sum_{j=1}^{J} \lambda_j \gamma_{0j}
\]

- Deduce what are the assumptions of stationarity (weak or intrinsic) that will be used for Simple, Ordinary and Universal Kriging.

Kriging formulation with the variogram IV

Definition (Kriging variance)

The minimised mean-squared prediction error \(E[\epsilon(x_0)^2]\) is sometimes called the Kriging variance:

\[
\sigma_0^2 = 2 \sum_{j=1}^{J} \hat{\lambda}_j \gamma_{0j} - \sum_{j=1}^{J} \sum_{i=1}^{J} \hat{\lambda}_i \hat{\lambda}_j \gamma_{ij} \quad (\text{Or./Un. Kriging})
\]

or

\[
\sigma_0^2 = c_{00} - 2 \sum_{j=1}^{J} \hat{\lambda}_j c_{0j} + \sum_{i=1}^{J} \sum_{j=1}^{J} \hat{\lambda}_i \hat{\lambda}_j c_{ij} \quad (\text{Simple Kriging})
\]

so prediction interval can be constructed: \(\hat{s}_0 \pm 2\sigma_0\).
Kriging formulation with the variogram V

<table>
<thead>
<tr>
<th>Kriging</th>
<th>Hypotheses $\mu(x)$</th>
<th>Contraints</th>
<th>Prediction at $x_0$</th>
<th>Variance</th>
<th>Stationarity?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Modelling variogram I

We have presented the Kriging approaches to perform prediction. These approaches require the knowledge of the covariance function or variogram of the stochastic process $\tilde{s}(x) = s(x) - \mu(x)$.

**Definition (variogram cloud)**

For any set of data we can compute the semivariance for each pair of points $x_i$ and $x_j$ individually as

$$\hat{\gamma}(x_i, x_j) = \frac{1}{2} (\tilde{s}(x_i) - \tilde{s}(x_j))^2$$

These values are plotted against the difference $h = x_i - x_j$ between the sites. When a variogram is isotropic, it depends only on the distance $||h||$. A scatterplot of this form is called the variogram cloud.

Modelling variogram II

**Exercises:**

- How many points does the variogram cloud contain if the separation distance $h$ is not limited?

- Consider $x_1 = (2, 3)$, $x_2 = (4, 4)$ and $x_3 = (4, 3)$ where we have measured $s_1 = 2$, $s_2 = 3$ and $s_3 = 2.4$. Assuming the variogram is isotropic, plot the variogram cloud.

Modelling variogram III

**Definition (Average semi variance)**

The definition of the semivariance is

$$\gamma(h) = \frac{1}{2} \mathbb{E}[(s(x) - s(x + h))^2]$$

which can be estimated by:

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} (s(x_i) - s(x_i + h))^2$$

$m(h)$ is the number of pairs of sites separated by the particular lag vector $h$. The way it is computed depends on the sampling design.
Modelling variogram IV

Example: In labs, we explore the first tutorial given with the R-package gstat. This tutorial looks at the Meuse data set that gives locations and topsoil heavy metal concentrations, collected in a flood plain of the river Meuse, near the village of Stein (NL). Heavy metal concentrations are from composite samples of an area of approximately 15 m x 15 m.

Modelling variogram V

Figure: Variogram cloud of \( \log(zinc) \) in the Meuse data set.

Modelling variogram VI

Figure: Average semiVariogram of \( \log(zinc) \) in the Meuse data set.

Sampling designs I

Consider the region \( D \subset \mathbb{R}^2 \), the location \( \{x_j\}_{j=1}^N \) can be chosen by:

- Systematic Sampling (e.g. grid)
- Pure random sampling
- Stratified sampling
Conclusion

- We have seen how to estimate a variogram using the observation we have for stochastic processes \( \{s(x_j)\}_{j=1,\ldots,J} \) that are measured at each site \( x_j \).

- Can we estimate the variogram \( \gamma \) in this fashion when Universal Kriging will be used to perform prediction at a new position \( x_0 \)?

- What is the stochastic process for which the variogram is used in Universal Kriging? Ordinary Kriging? Simple Kriging?

Variogram modelling for Universal Kriging I

- If \( \tilde{s}(x_i) - \tilde{s}(x_j) = s(x_i) - s(x_j) \) for simple and ordinary Kriging, this is not the case anymore for universal Kriging where \( \text{E}[s(x)] = \mu(x) \).
- We have collected observations for \( s \) but not for \( \tilde{s} \) and we don’t know the form of \( \mu(x) \).
- For instance, if we assume that \( s(x) = a \, x + b \) (linear drift), then the difference

\[
\tilde{s}(x + h) - \tilde{s}(x) = a \, h
\]

so the estimated variogram (variogram plot) will be affected by this drift and looks like a parabola:

\[
\gamma(h) = \frac{1}{2} (a \, h)^2
\]

Variogram modelling for Universal Kriging II

- The first method proposes to get a quick estimate of the mean to remove it:
  - Specify a linear model \( \mu(x) = \sum_{p=0}^{P} \beta_p \phi_p(x) \) on a chosen basis of functions \( \{\phi_p\}_{p=0,\ldots,P} \).
  - Estimate \( \{\beta_p\}_{p=0,\ldots,P} \) by Least Squares to give an estimate of the mean
    \( \hat{\mu}(x) = \sum_{p=0}^{P} \hat{\beta}_p \phi_p(x) \).
  - Compute the variogram of \( \tilde{s}(x) = s(x) - \hat{\mu}(x) \).

Variogram modelling for Universal Kriging III

- The other approach to estimate the variogram works as follow:
  - Choose a theoretic variogram function.
  - Specify a linear model \( \mu(x) = \sum_{p=0}^{P} \beta_p \phi_p(x) \) on a chosen basis of functions \( \{\phi_p\}_{p=0,\ldots,P} \).
  - Estimate \( \hat{\mu}(x) \) (i.e. find \( \{\hat{\beta}_p\}_{p=0,\ldots,P} \) such that the residuals
    \( \tilde{s}(x) = s(x) - \hat{\mu}(x) \) have a variogram that matches the chosen theoretical one.
Variogram functions I

Several analytical functions have been proposed in the literature to model a variogram. They all follow a few basic rules.

- Limits on variogram function.
  - Behaviour near the origin. The way in which the variogram approaches the origin is determined by the continuity (or lack of continuity) of the variable \( s(x) \) itself. The semivariance at \( |h| = 0 \) is by definition 0. It can happen however that experimental values give a positive value (positive intercept, nugget effect).

  Near 0, the variogram can have a linear approach \( \gamma(h) \approx b|h| \) or quadratic \( \gamma(h) \approx b|h|^2 \) as \( |h| \to 0 \).

Variogram functions II

- Behaviour towards infinity. The variogram is constrained by:

  \[
  \lim_{|h| \to \infty} \frac{\gamma(h)}{|h|^2} = 0 \quad \text{as} \quad |h| \to \infty
  \]

If it does not then the process is not entirely random (and is not compatible with the intrinsic hypothesis).

Variogram functions III

- Examples of models:
  - Unbounded models.
    \[
    \gamma(h) = \omega |h|^\alpha \quad \text{for} \quad 0 < \alpha < 2
    \]
  
  \( \alpha = 1 \) the the variogram is linear .

  - Bounded models. Based on experience, bounded variation is more common than unbounded one.
    - Bounded linear
    - Spherical
    - Exponential
    - ...

Variogram functions IV

Once we have computed the variogram plot, we select the best analytical (theoretic) variogram function by:

- fitting the models to the observed variogram plot,
- choosing the theoretic model that has the smallest RSS (Residual sum of square) and/or AIC (Akaike Information Criterion).

Exercise: Assuming that \( s \) is weakly stationary, explain why the variogram cannot be unbounded.