Curve Fitting via Kriging

Michael J. Sasena

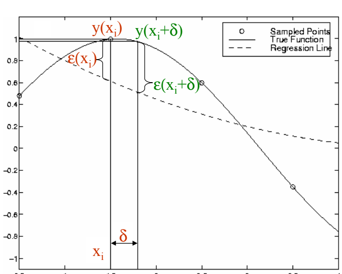
Outline

- What is Kriging?
- Kriging Model Formulation
- Other Issues
- Computer Demos

Kriging Overview

$y(x) = f(x) + Z(x)$

- $f(x)$ = polynomial
  - Global Trend
- $Z(x)$ = functional departure from $f(x)$
  - Local Variation
  - Has $\mu=0$, $\sigma^2>0$
  - Determines how $\varepsilon(x_i)$ and $\varepsilon(x_i+\delta)$ are related



The Spatial Correlation Function

$Z(x)$ is described by a Spatial Correlation Function (SCF)

- SCF quantifies the influence of a sampled point on its neighbors
- SCF => 0 as $|\mathbf{x}|$ => ∞
- the parameter θ effects how quickly the SCF vanishes
  - Matlab demo

Mathematical Model

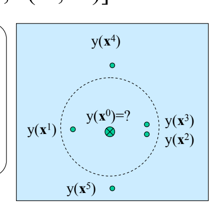
- One choice of SCF is:

  $R(x^{(i)},x^{(j)}) = \exp\{-\theta(x^{(i)}-x^{(j)})^2\}$

- As we expand to multidimensions, we have:
  - scalar θ ⇒ $R(\mathbf{x}^{(i)},\mathbf{x}^{(j)}) = \exp\{-\theta * \Sigma[(x_k^{(i)}-x_k^{(j)})^2]\}$
  - vector θ ⇒ $R(\mathbf{x}^{(i)},\mathbf{x}^{(j)}) = \exp\{-\Sigma\,[\theta_k(x_k^{(i)}-x_k^{(j)})^2]\}$

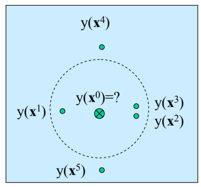
Nomenclature

- $\hat{y} = y(\mathbf{x}^0)$
- $\mathbf{y} = [y(\mathbf{x}^1), y(\mathbf{x}^2), \ldots, y(\mathbf{x}^n)]^T$
- $\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}^0,\mathbf{x}^1), R(\mathbf{x}^0,\mathbf{x}^2),\ldots,R(\mathbf{x}^0,\mathbf{x}^{ns})]^T$
- $\mathbf{R} = \begin{pmatrix} R(\mathbf{x}^1,\mathbf{x}^1) & \ldots & R(\mathbf{x}^1,\mathbf{x}^{ns}) \\ R(\mathbf{x}^2,\mathbf{x}^1) & \ldots & R(\mathbf{x}^2,\mathbf{x}^{ns}) \\ \vdots & & \vdots \\ R(\mathbf{x}^{ns},\mathbf{x}^1) & \ldots & R(\mathbf{x}^{ns},\mathbf{x}^{ns}) \end{pmatrix}$



Kriging as a Weighted Average

$$\begin{pmatrix} 1.00 & 0.06 & 0.05 & 0.00 & 0.00 \\ 0.06 & 1.00 & 0.85 & 0.00 & 0.00 \\ 0.05 & 0.85 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} = \begin{pmatrix} 0.43 \\ 0.43 \\ 0.42 \\ 0.00 \\ 0.00 \end{pmatrix} \Rightarrow \mathbf{R}\,\lambda = \mathbf{r}$$



<u>Solution</u>: $\lambda = \begin{pmatrix} 0.408 \\ 0.264 \\ 0.170 \\ 0.000 \\ 0.000 \end{pmatrix}$

The Flexibility of Kriging

- Pattern of Spatial Variability
  - Zero weights beyond a certain range
- Proximity of Data to $\mathbf{x}^0$
  - Weight decreases for distant data
- Data Configuration
  - Redundancy of $\mathbf{x}^2$ and $\mathbf{x}^3$

Prediction for f(x) = β

- $\mathbf{y} = \mathbf{1}\beta + \mathbf{z}$
- $\mathbf{z} = \mathbf{y} - \mathbf{1}\beta$
- $\hat{z} = \lambda^T \mathbf{z}$, where $\mathbf{R}\,\lambda = \mathbf{r}$
- Rearranging, $\lambda = \mathbf{R}^{-1}\,\mathbf{r}$
- Plugging back into $z = \lambda^T \mathbf{z}$,

  $z = (\mathbf{R}^{-1}\mathbf{r})^T\mathbf{z} = \mathbf{r}^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{1}\hat{\beta})$
- Backtransform leaves:

$$\boxed{\hat{y} = \hat{\beta} + \mathbf{r}^T\,\mathbf{R}^{-1}\,(\mathbf{y} - \mathbf{1}\hat{\beta})}$$

Fitting the Model

- We find θ by maximizing the likelihood

  $\underset{0<\theta<\infty}{\mathrm{MLE}(\theta)} = \{-\,[n_s{*}\ln(\hat{\sigma}^2) + \ln|\mathbf{R}|]\;/\;2\}$

- The underlying variance in the data is

  $\hat{\sigma}^2 = [(\mathbf{y}-\hat{\beta}\mathbf{1})^T\mathbf{R}^{-1}(\mathbf{y}-\hat{\beta}\mathbf{1})]\;/\;n_s$, where

  $\hat{\beta} = (\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1})^{-1}\;\mathbf{1}^T\;\mathbf{R}^{-1}\;\mathbf{y}$

- Therefore, MLE is a function of θ only

Model Validation

- Cross Validation
  - Leave out the point $\mathbf{x}^i$ and fit the model
  - Predict the output $y(\mathbf{x}^i)$ and find the error
  - Repeat for each point and find the RMS error
- Jack-knifing
  - Leave several data points off to the side
  - Fit the model with the remaining points
  - Measure the error at these points and compute the RMS error

Additional Questions

- What is the effect of n_s of the modeling error?
  - one would assume a better model is fit with a larger sample size
- What is the effect of having a vector-valued θ vs. a scalar-valued θ?
  - one would expect better results

Other Strengths

- Can easily compute the variance at any predicted point. ($\hat{\sigma}^2 = 0$ at data points)
- Can account for anisotropy in SCF
  - May be more continuous in one direction (not necessarily along axes)
- Can incorporate secondary information
  - Rainfall prediction augmented by elevation data
  - Expensive simulation augmented by a faster, less accurate simulation (e.g. crash vs. lumped)

Conclusions

- Kriging is an interpolating function
- Kriging becomes more accurate with increased sampling (n_s)
  - Polynomials don't necessarily become more accurate with increasing order (instability)
- Vector-valued θ can greatly increase cost
  - benefits are problem dependent
- Effect of polynomial order in kriging is often small

References

- CEE 570 - Pierre Goovaerts
- P. Goovaerts, 1997. *Geostatistics for Natural Resource Evaluation*, New York, Oxford University Press.
- G. Matheron, 1970. *La Theorie des Variables Regionalisees et ses Applications*, Fascicule 5, Les Cahiers du Centre de Morphologie Mathematique, Ecole des Mines de Paris, Fontainebleau, p.212.
- N. Cressie, 1993. *Statistics for Spatial Data*, New York, J. Wiley.
- J. Sacks, W. Welch, W. Mitchell, H. Wynn, 1989. *Design and Analysis of Computer Experiments*, Statistical Science, vol. 4, no. 4, pp.409-435 (with discussion).