Using multidimensional scaling with Duchon splines for reliable finite area smoothing

David Lawrence Miller

Mathematical Sciences
University of Bath

useR! 2011
Warwick University
18 August 2011
Spatial smoothing

- Have \((x, y)\) locations and some response, \(z\).
- Want a smooth map or to explain spatial auto-correlation.
- Thanks to \texttt{mgcv}, \texttt{inla}, \texttt{sp}, etc spatial smoothing now easy.
- \textit{But} still some problems – e.g. leakage.
Smoothing using splines

- Take some set of basis functions, estimate coefficients, penalize based on integrated derivatives (roughness).

\[ \| f(x, y) - z \|^2 + \lambda J_{m,d} \]

\[ J_{2,2} = \int \int \left( \frac{\partial^2 f(x, y)}{\partial x^2} \right)^2 + \left( \frac{\partial^2 f(x, y)}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 f(x, y)}{\partial y^2} \right)^2 \, dx \, dy \]

- Here \( f \) is a thin plate regression spline:

\[ f(x, y) = \sum_{i=1}^{n} \delta_i \eta_{md}(r) + \sum_{j=1}^{M} \alpha_j \phi_j(x, y) \]

- Integrate into bigger models (GAMs/GAMMs/etc).
Solutions to leakage

- Boundary conditions - soap film smoothing (Wood et al., 2008), FELSPLINE (Ramsay, 2002).
- Within-area distance metrics - GLTPS (Wang and Ranalli, 2007).
- Domain transformation - what I’m going to talk about.
Multidimensional scaling + within-area distances = domain transform

- MDS: Take \((n \times n)\) symmetric distance matrix, project into \((< n)\) dimensions.
- Using Euclidean distances \(\Rightarrow\) same point set (up to scale/rotate).
- Use *within-area* distances reflect distances travelled by objects in domain.
MDS+TPRS smooths

Truth

MDS

tprs

soap
Projections in higher dimensions

- Ordering and crowding issues.
- Using rgl, 3-D projections look like manifolds.
- Unreliable smoothing with thin plate in high dimensions.
- Nullspace++ (in size and function complexity)

\[
f(x, y) = \sum_{i=1}^{n} \delta_i \eta_{md}(r) + \sum_{j=1}^{M} \alpha_j \phi_j(x, y)
\]
Nullspace explosion

![Graph showing the nullspace dimension (M) vs. smoothing dimension (d)]
Duchon splines (I)

- Usual thin plate penalty:

\[
J_{m,d} = \int \cdots \int_{\mathbb{R}^d} \sum_{\nu_1 + \cdots + \nu_d = m} \frac{m!}{\nu_1! \cdots \nu_d!} \left( \frac{\partial^m f(x_1, \ldots, x_d)}{\partial x_1^{\nu_1} \cdots \partial x_d^{\nu_d}} \right)^2 \, dx_1 \cdots dx_d
\]

- Take Fourier transform and weight on frequencies.
- Fudge nullspace (radial basis makes up for global polys).
- Penalize the "smoother" parts of the radial functions less.
- Becomes:

\[
\tilde{J}_{m,d} = \int \cdots \int_{\mathbb{R}^d} |\tau|^{2s} \sum_{\nu_1 + \cdots + \nu_d = m} \frac{m!}{\nu_1! \cdots \nu_d!} \left( \frac{\partial^m f}{\partial x_1^{\nu_1} \cdots \partial x_d^{\nu_d}}(\tau) \right)^2 \, d\tau
\]
Duchon splines (II)

- Smooth in very high dimensions without huge nullspaces.
- Projection dimension selection by GCV score.
Chlorophyll levels in the Aral sea

raw data

raw data

tprs

soap

mds 5D
Generalized distance smoothing

- Distance matrix could be any set of disparities:
  - MP voting records.
  - Distance between patient’s gene expressions.
  - Socio-economic indicators.
- Most variation $\neq$ best predictors.
- Column-wise variance non-constant.
- Issue may be in the metric used.
msg - Multidimensional Scaling for Gams

- Implemented in R as an extra basis in \texttt{mgcv}.
- If you know how to use \texttt{mgcv}, you know how to use \texttt{msg}.
  \[
  \texttt{b<-gam(z~s(x,y,bs="msg",xt=list(bnd=boundary,mds.dim=4)),data=data)}
  \]
- GCV dimension selection coming \texttt{soon}.
Conclusion

- \texttt{msg} performs at least as well as soap film in simulation.
- Duchon splines very useful for high dimensional smoothing.
- Can do smoothing of general distance matrices.
- \textit{But} no killer examples (yet!).
- Do you have any interesting (distance) data?
- Package \texttt{msg} available at \url{http://github.com/dill/msg}. 
Calculating within-area distances