Spatial adaptation of the covariance function

Noha Youssef, Henry Wynn

London School of Economics
H.Wynn@lse.ac.uk

LSE, June 2011

Introduction, summary

- We want to update covariance functions in emulation adaptively
- In basic kriging the conditional variance is not constant, but conditional correlation remains the same
- In basics kriging the conditional covariance only depends on past input, x (design)
- Need models in which the conditional covariance depend on past output x and Y
- Full Bayes approach: put priors on covariance
- Difficult integration: suggest a “cheap” approach
- Discussion of a “direct” approach using wavelet regression
- Kernels and metrics

Hierarchical Bayesian analysis

Assigns hyper-priors to the model parameters. Assume the output $Y|\beta, \Lambda, \sigma^2$ has a multivariate normal distribution

$$Y|\beta, \Lambda, \sigma^2 \sim N(X\beta, \delta\sigma^2 I) \tag{1}$$

where X is $n \times p$ (first stage) Λ is the prior covariance for β and δ represents the nugget

$$\begin{aligned}
\beta|\Lambda, \sigma^2 &\sim N(\mu, \sigma^2\Lambda) \\
\Lambda_{p\times p}|\sigma^2 &\sim IW(\Psi, m) \\
\sigma^2 &\sim IG\left(\frac{a}{2}, \frac{b}{2}\right)
\end{aligned}$$

And note that $\Psi = E(\Lambda)$.

Discussion

- Want to find the marginal predictive distribution $f(Y_r|Y_n)$ where Y_n is the vector of n already observed outputs and Y_r is the vector of r unobserved outputs that correspond to a future design of r points.
- Finding the marginal posterior distributions for the parameters or the predictive distribution is computationally hard.
- Eg Integrating over β to obtain the marginal conditional distribution of $\Lambda|Y$ requires numerical methods.
- Suggest an empirical Bayes method, which (i) is computationally cheap (ii) has an a intuitive appeal

Updating

- The predictive distribution $(Y_r|Y_n)$ involves integrating over Λ, β and σ^2.
- Concentrate on the posterior distribution of $\Lambda|\beta, Y_n, \sigma^2$

$$\pi(\Lambda|\beta,\sigma^2,Y_n) \approx \frac{\pi(\Lambda,\beta,\sigma^2|Y_n)}{\pi(\beta,\sigma^2|Y_n)}$$

$$\approx \left(\frac{1}{\sigma^2}\right)^{\frac{p+m+1}{2}-1} \det(\Lambda)^{-\frac{p+m+1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\mathrm{trace}(\Psi^*)\Lambda^{-1})\right),$$

where

$$\Psi^* = \Psi + (\beta-\mu)(\beta-\mu)^T$$

Contd.

Ψ^* is the updated *IW* distribution parameter Ψ. The conditional expectation, $\mathrm{E}(\Lambda|\beta, \sigma^2, Y_n)$, is given by

$$\mathrm{E}(\Lambda|\beta, \sigma^2, Y_n) = \frac{\Psi + (\beta - \mu)(\beta - \mu)^T}{m - p - 1} = \frac{\Psi^*}{m - p - 1} \tag{2}$$

where m is the degrees of freedom of the *IW* distribution and $p + q$ is the number of mean parameters in the model.

Empirical Bayes step

Instead of carrying out the integration we replace β by an estimate. The first option is to replace β by its OLS estimator

$$\hat{\beta}_{OLS} = (X_n^T X_n)^{-1} X_n^T Y_n \tag{3}$$

X_n is the design matrix corresponds to the already selected design Y_n is the observed output at the chosen design points, at stage 1. The second option is to find the posterior mode Bayes estimator by maximising the joint posterior distribution with respect to β. In this case $\hat{\beta}$ is given by

$$\hat{\beta}_{Bayes} = \left(X_n^T \delta^{-1} X_n + \Lambda^{-1} \right)^{-1} \left(X_n^T \delta^{-1} Y_n + \Lambda^{-1} \mu \right). \tag{4}$$

The approximate posterior (predictive) covariance at r non-sample design points is

$$\sigma^2 X_r E(\Lambda|\beta, Y_n) X_r^T + \sigma^2 \delta I,$$

and (ignoring σ^2) this becomes

$$X_r \left[\frac{(\Psi + (\hat{\beta} - \mu)(\hat{\beta} - \mu)^T)}{m - p - 1} \right] X_r^T + \delta I$$

Thus our approximate update is adding a term proportional to

$$S = X_r(\hat{\beta} - \mu)(\hat{\beta} - \mu)^T X_r^T$$

Prior : $\hat{Y}_0 = X_r^T \mu$
Posterior, after stage 1 : $\hat{Y}_1 = X_r^T \hat{\beta}$

$$S = ss^T = (\hat{Y}_1 - \hat{Y}_0)(\hat{Y}_1 - \hat{Y}_0)^T$$

One point optimal design

In this case $X = x^T$ (single row vector at unsampled point) and

$$S = x^T \Psi x + s^2 = v + s^2,$$

Prior variance at design point plus squared deviation after stage 1.

- This adapts to where the largest "errors" are predicted to be on the unsampled design points after stage 1.
- Continue, this process, incorporating the update into Ψ at every stage.
- Could take a weighted version

$$(1 - \alpha)v + \alpha s^2$$

Direct method

- Select a basis $\{f_j(x)\}$
- Fit to data: $Y(x) = \sum_j \theta_j f_j(x)$
- Criteria: Entropy, IMSE, AIC etc
- Model: Keep f_j which have biggest contribution to criteria eg by wavelet thresholding (Bayes or non-Bayes)
- Design: at next stage choose unsampled design point where contribution is large
- Repeat

Haar, sinc, SSM et al

- Need to express local smoothness: wavelets
- Mother wavelet then shift and scale to form wavelets
- Whittaker-Shannon cardinal interpolation: sinc approximates to polynomial interpolation on integers $\{-n, \ldots, n\}$ as $n \to \infty$
- SSM approximates to spline
- sinc et al used in time/frequency analysis: “harmonic wavelets”.
- There are (approximate) wavelet expansions to the covariance function: “wavelet spectrum”.
- Noha: first compute K-L, then expand each eigenfunction into Haar
- Suggestion is to *go directly for wavelet model* and use
- Full Bayes, empirical Bayes or “direct” method
- Even if not good for modeling, good for design?

Kernels and metrics

$$d(x,y) = \langle x,x\rangle + \langle y,y\rangle - 2\langle x,y\rangle$$

- Can transform the metric and still embed in Hilbert space
- eg take real line: $d(x,y) = |x-y|$. Change to $\sqrt{|x-y|}$: obtain a "screw line" in Hilbert space (paper on the general case with Wicher Bergsma, to come)
- Define the new RKHS, new covariance
- Can we adapt the covariance via adaptation of the metric: space dilation?
- Link with wavelet spectrum? Yes, some papers on "time dilation using wavelets"
- Riemannian manifolds
- Does this help the Bayesification to get adaptation?