Asymptotic analysis of the role of spatial sampling for hyper-parameter estimation of Gaussian processes

François Bachoc
Josselin Garnier
Jean-Marc Martinez

CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France
LPMA, Université Paris 7

January 2013

The PhD

Two components of the PhD

- Use of Kriging model for code validation
  - Bachoc F, Bois G, and Martinez J.M, Gaussian process computer model validation method, *Submitted*.
- Work on the problem of the covariance function estimation
  - Bachoc F, Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification, *In minor revision for CSDA*.
  - Bachoc F, Asymptotic analysis of the role of spatial sampling for hyper-parameter estimation of Gaussian processes, *In preparation*.

Spatial sampling for hyper-parameter estimation

- Hyper-parameters : The parameters that define the covariance function within a pre-determined set. E.g the correlation lengths
- Spatial sampling : Initial design of experiment for Kriging
- It has been shown that irregular spatial sampling is often an advantage for hyper-parameter estimation

Stein M, Interpolation of Spatial Data : Some Theory for Kriging, *Springer, New York, 1999. Ch.6.9*.

Zhu Z, Zhang H, Spatial sampling design under the infill asymptotics framework, *Environmetrics 17 (2006) 323-337*.

- Our question : Is irregular sampling always better than regular sampling for hyper-parameter estimation ?

Asymptotics for hyper-parameters estimation

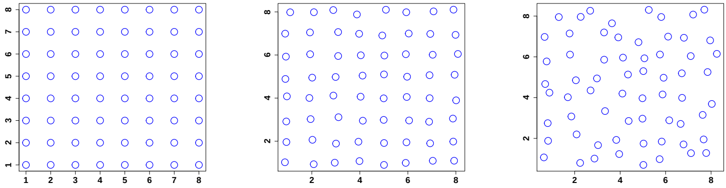
- Finite sample results for hyper-parameter estimation
  - must generally be obtained by simulation
  - are specific to the situation
- Asymptotics (number of observations $n \to +\infty$) is an area of active research (Maximum-Likelihood estimator)
- Two main asymptotic frameworks
- fixed-domain asymptotics : The observations are dense in a bounded domain
  From 80'-90' and onwards. Fruitful theory
  However, when convergence in distribution is proved, the asymptotic distribution does not depend on the spatial sampling $\longrightarrow$ Impossible to compare sampling techniques for estimation in this context
- increasing-domain asymptotics : A minimum spacing exists between the observations $\longrightarrow$ infinite observation domain.
  Asymptotic normality proved for Maximum-Likelihood under general conditions

Mardia K, Marshall R, Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika 71 (1984) 135-146*.

Randomly perturbed regular grid

- Our sampling model : regular square grid of step one in dimension d, $(v_i)_{i \in \mathbb{N}^*}$. The observation points are the $v_i + \epsilon X_i$. The $(X_i)_{i \in \mathbb{N}^*}$ are *iid* and uniform on $[-1, 1]^d$
- $\epsilon \in]-\frac{1}{2}, \frac{1}{2}[$ is the regularity parameter. $\epsilon = 0 \longrightarrow$ regular grid. $|\epsilon|$ close to $\frac{1}{2} \longrightarrow$ irregularity is maximal

Illustration with $\epsilon = 0, \frac{1}{8}, \frac{3}{8}$



The estimators we study

$\{\sigma^2 K_\theta, \sigma^2 \geq 0, \theta \in \Theta\}$ is the set of covariance functions for the GP Y, with K_θ a correlation function

We study the Maximum Likelihood (ML) and Cross Validation (CV) estimators.

Leave-One-Out criteria we study

With

- $\hat{y}_{\theta,i,-i} = \mathbb{E}_{\sigma^2,\theta}(Y(x_i)|y_1, ..., y_{i-1}, y_{i+1}, ..., y_n)$
- $\sigma^2 c^2_{\theta,i,-i} = \mathbb{E}_{\sigma^2,\theta}((Y(x_i) - \hat{y}_{\theta,i,-i})^2|y_1, ..., y_{i-1}, y_{i+1}, ..., y_n)$,

The CV estimator we study is

$$\hat{\theta}_{CV} \in \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \hat{y}_{\theta,i,-i})^2$$

and

$$\hat{\sigma}^2_{CV} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_{\hat{\theta}_{CV}i,-i})^2}{c^2_{\hat{\theta}_{CV},i,-i}}$$

Computational cost similar to ML thanks to the virtual Leave One Out formulas

The spatial sampling model

Main result

Analysis of the asymptotic variance

Conclusion

Main result

Under general conditions

For ML

- a.s convergence of the random Fisher information : The random trace $\frac{1}{n} Tr \left(R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1} \frac{\partial R}{\partial \theta_j} \right)$ converges a.s to the element $(\mathbf{I}_{ML})_{i,j}$ of a $p \times p$ deterministic matrix $\mathbf{I}_{ML}$ as $n \to +\infty$
- asymptotic normality : With $\Sigma_{ML} = 2\mathbf{I}_{ML}^{-1}$

$$\sqrt{n} \left(\hat{\theta}_{ML} - \theta_0 \right) \to \mathcal{N} \left(0, \Sigma_{ML} \right)$$

For CV

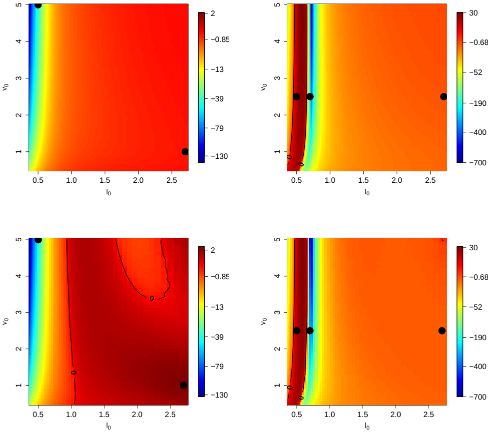
Same result with more complex random traces for asymptotic covariance matrix Σ_{CV}

$\Sigma_{ML,CV}$ depend only on the regularity parameter ϵ.
$\boxed{\longrightarrow}$ in the sequel, we study the functions $\epsilon \to \Sigma_{ML,CV}$

Small random perturbations of the regular grid

Matèrn model. Dimension one. One estimated hyper-parameter.
Levels plot of $(\partial_\epsilon^2 \Sigma_{ML,CV})/\Sigma_{ML,CV}$ in $\ell_0 \times \nu_0$

Top : ML
Bot : CV
Left : $\hat{\ell}$ (ν_0 known)
Right : $\hat{\nu}$ (ℓ_0 known)

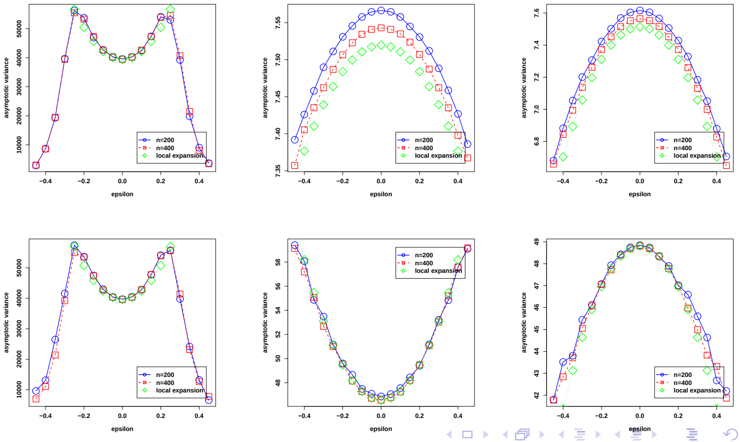


There exist cases of degradation of the estimation for small perturbation for ML and CV. Not easy to interpret

Large random perturbations of the regular grid

Plot of $\Sigma_{ML,CV}$. Top : ML. Bot : CV.

From left to right : $(\hat{\nu}, \ell_0 = 0.5, \nu_0 = 2.5)$, $(\hat{\ell}, \ell_0 = 2.7, \nu_0 = 1)$, $(\hat{\nu}, \ell_0 = 2.7, \nu_0 = 2.5)$



Conclusion

Conclusion

- Our answer to the question : Irregularity in the sampling is generally an advantage for the estimation, but not necessarily
- With ML, irregular sampling is more often an advantage than with CV
- Large perturbations of the regular grid are often better than small ones for estimation
- Keep in mind that hyper-parameter estimation and Kriging prediction are strongly different criteria for a spatial sampling

Perspectives

- Designing other CV procedures (cf talk on CV in DiceKriging)