Gibbs Reference Posterior for Robust Gaussian Process Emulation

Joseph Muré
Université Paris Diderot and EDF R&D

Supervisor(s): Prof. J. Garnier (École Polytechnique), Dr. L. Le Gratiet (EDF R&D) and Dr. A. Dutfoy (EDF R&D)

Ph.D. expected duration: 2015-2018

Adress:
EDF Recherche et Développement - Dpt. PRISME
6 quai Watier
78401 Chatou
France

Email: joseph.mure@edf.fr

Abstract:

Gaussian Processes are widely used to model the spatial distribution of some real-valued quantity when said quantity is only observed at a few locations. This emulation technique is a convenient way to represent the uncertainty of the value of the quantity at unobserved points [SWN03]. In this work, we focus on Universal Kriging, a metamodel that emulates an unknown function through the sum of some "mean function" and of a stationary Gaussian Process, a framework that is frequently used in both the geostatistical and the machine learning literature. The exact probability distribution of a Gaussian Process depends not only on its mean function (supposed here to belong to some finite-dimensional vectorial space), but also on a variance parameter and a correlation function (also known as "correlation kernel") which itself depends on parameters.

We propose an "objective" posterior distribution on these parameters derived through techniques related to the Bernardo reference posterior, which we call "Gibbs reference posterior".

The need for a Bayesian treatment of the parameters of Kriging models arises from the lack of robustness of the Maximum Likelihood Estimator (MLE) in dealing with parameters of correlation kernels. Indeed, what makes estimating them "notoriously difficult", as [KO01] put it, is that the likelihood function may often be quite flat [LS05]. To tackle this problem, one may stabilize the MLE by adding a nugget to the covariance kernel, namely adding a covariance component concentrated on the diagonal. However, as was noted by [AC12], "the presence of a nugget is equivalent to the assumption that the simulator contains some variability that is not explainable by its inputs". Alternatively, [LS05] proposed penalizing the likelihood function, which may also be interpreted as using a prior distribution and then choosing the Maximum A Posteriori (MAP) estimate instead of the MLE. Of course, using a full-Bayesian approach obviates the problem of robustness of the estimator of the parameters, as one may simply use the integrated predictive distribution.

Whether one wishes to use a prior distribution as a penalizing function for the likelihood or to deploy the whole Bayesian machinery, one often faces the problem of a lack of *a priori* information. This is where Objective Bayes, which was first introduced in this context by [BDOS01], is helpful. Their work on deriving the reference prior in this context and establishing posterior propriety was then successively extended by [Pau05], [RSS13] and [Gu16]. However, the above cited works all make a restrictive assumption in order to guarantee posterior propriety. It turns out this assumption is not satisfied by twice differentiable correlation kernels. Therefore, to the author's

knowledge, there exists currently no proof of the propriety of the reference posterior for such standard covariance kernels as Matérn covariance functions with smoothness parameter $\nu > 1$. This, along with the wish to make the full-Bayesian process tractable in practice, led us to consider a different but similar "objective" posterior distribution. As it is defined through conditional densities and thus is well suited to Gibbs sampling, we call it the "Gibbs reference posterior".

We offer theoretical guarantees of existence and propriety of the Gibbs reference posterior and provide an MCMC algorithm for sampling of the Gibbs reference posterior in the case where a Matérn class covariance kernel with known smoothness parameter is used. Beyond parameter inference, what matters to us is how well we are able to account for uncertainty on values of the Gaussian Process at unobserved points. We show in simulation examples that a posteriori Predictive intervals at unobserved points have effective coverage close to their theoretical level, while predictive intervals produced by plugging-in the MLE have substantially lower coverage.

This new theory for the inference of Kriging parameters is then applied in a Trans-Gaussian Kriging example provided by an industrial problem concerning non-destructive tests of Steam Generator tubes in nuclear power plants. The data that was gathered cannot be convincingly modeled by a Gaussian process, so a suitable transformation must be applied in order to make them Gaussian. Because the transformation is chosen within a parametric family, the complete inference problem contains a transformation parameter in addition to the usual Gaussian parameters. This is where our Bayesian theory for Kriging applies, because it allows us to integrate the Gaussian parameters out of the model, leaving only the transformation parameter to be estimated.

References

[AC12] I. Andrianakis and P. G. Challenor. The effect of the nugget on gaussian process emulators of computer models. *Computational Statistics & Data Analysis*, 56(12):4215–4228, 2012.

[BDOS01] J. O. Berger, V. De Oliveira, and B. Sansó. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374, 2001.

[Gu16] M. Gu. *Robust Uncertainty Quantification and Scalable Computation for Computer Models with Massive Output*. PhD thesis, Duke University, 2016.

[KO01] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[LS05] R. Li and A. Sudjianto. Analysis of computer experiments using penalized likelihood in gaussian kriging models. *Technometrics*, 47(2):111–120, 2005.

[Pau05] R. Paulo. Default priors for Gaussian processes. *Annals of Statistics*, 33(2):556–582, 2005.

[RSS13] C. Ren, D. Sun, and S. K. Sahu. Objective bayesian analysis of spatial models with separable correlation functions. *Canadian Journal of Statistics*, 41(3):488–507, 2013.

[SWN03] T. J. Santner, B. J. Williams, and W. I. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York, 2003.

Short biography – Joseph Muré is a third year PhD student who has a Master's degree of Probability and Random models of the Université Pierre et Marie Curie. The results of his PhD thesis, which is financed by EDF, are applied to improving current metamodels regarding Probability of Defect Detection (POD) in nuclear power plants.