Gibbs Reference Prior for Robust Gaussian Process Emulation

J. Muré
Université Paris Diderot

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. de Management des Risques Industriels
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. In this work, we focus on stationary Gaussian Processes with null mean function, a framework that is referred to as "Simple Kriging" in the geostatistical literature and is also frequently used in the context of computer experiments and machine learning. The exact probability distribution of a Gaussian Process depends not only on its mean function (supposed here to be null), but also on a variance parameter and a correlation kernel which itself depends on parameters.

We propose an "objective" Jeffreys-like prior distribution on these parameters, which we call "Gibbs reference prior".

The need for a prior distribution on 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, 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. The authors' 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, which essentially implies that for any twice differentiable correlation kernel, the number of observation points must not exceed the spatial dimension by more than 2. Standard covariance kernels such as the Matérn covariance function with regularity parameter $\nu > 1$ thus cannot be used. This, along with the wish to make the full-Bayesian process tractable in practice, led us to consider a

different but similar "objective" prior distribution. As it is defined through conditional densities and thus lends itself well to Gibbs sampling, we call it the "Gibbs reference prior".

Because the Gibbs reference prior is not the reference prior, we need to establish what we mean with the notion of an "objective" prior distribution. The following rule-of-thumb lays out our intent. An "objective" prior distribution should fulfill two requirements. First, it should not require any user input, that is, it should depend on no user-specified metaparameter. Second, it should rely on the information one may expect from the observations given by the design set. For instance, like the reference prior, the Gibbs reference prior relies on the Fisher information.

It should be noted that the second requirement implies an "objective" prior distribution necessarily depends on the design set, because where we observe the Gaussian Process bears heavily on the amount of information we may expect from the observations. In particular, it means the Lebesgue measure cannot be used as an "objective" prior distribution.

We offer theoretical guarantees of propriety for the Gibbs reference posterior on the vector of correlation lengths of Matérn class tensorized and anisotropic geometric covariance kernels. Moreover, we provide a framework which could be used to prove this result for other classes of correlation kernels. We sample from the posterior distribution on parameters using a Metropolis-within-Gibbs approach and contrast two ways of using it, namely the full-Bayesian method and the MAP plug-in approach.

The MAP plug-in approach has the intrinsic disadvantage of requiring parameter estimation. However, comparisons show it to be significantly more robust than the MLE with respect to the variability of the realizations of the Gaussian Process.

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. Examples show that predictive intervals at unobserved points produced by the full-Bayesian method have effective coverage close to their theoretical level, while predictive intervals produced by estimator plug-in methods (MAP and *a fortiori* MLE) have substantially lower effective coverage.

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.

[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.

Short biography – Joseph Muré 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, will be applied to improving current metamodels regarding Probability of Defect Detection (POD) in nuclear power plants.