



Robust Prediction Interval Estimation for Gaussian Processes by Cross-Validation Method

NAOUFAL ACHARKI

Ecole Polytechnique - TotalEnergies One Tech

Supervisor(s): Prof. Josselin Garnier (Ecole Polytechnique) and Dr. Antoine Bertoncello (TotalEnergies OneTech)

PhD expected duration: Oct. 2019 - Oct. 2022

Address: Nano-Innov, 2 Bd Thomas Gobert, 91120 Palaiseau

E-mail: naoufal.acharki@polytechnique.edu

Abstract: Gaussian Processes are considered as one of the most important Bayesian Machine Learning methods [5]. They typically use the Maximum Likelihood Estimation or Cross-Validation to fit parameters [2]. Unfortunately, these methods may give an advantage to the solutions that fit observations in the mean square sense, but they do not pay attention to the coverage and the width of Prediction Intervals. This may be inadmissible, especially for systems that require risk management. Indeed, an interval is crucial and offers valuable information that helps better management than just predicting a single value.

In this work [1], we address mainly the problem of model misspecifications when estimating Prediction Intervals for Gaussian Processes Regression and we develop a robust approach for estimating these Prediction Intervals, called *Robust Prediction Interval Estimation* (RPIE).

First, we use the Leave-One-Out Cross-Validation method to define a new loss function with respect to the Gaussian Processes hyperparameters that ensure that the estimated Prediction Intervals the desired level of coverage probability, also known as the optimal type II Coverage Probability. Unfortunately, the defined loss function is non-convex, non-continuous and step-wise constant. Moreover, replacing the loss function by a continuous loss-function converging point-wise to the original one does not simplify the task. Indeed, the optimization is still unfeasible with classical minimization approaches and there would be infinite solutions as we show that, under some appropriate assumptions, the solution's set is non-empty.

The second step of the method consists of reformulating it differently. The idea is to use a similarity measure between Gaussian distributions to pick up the closest pair of hyperparameters among the set of hyperparameters that achieved the desired Coverage Probability to some reference hyperparameters. To do so, we considered hyperparameters determined by a standard Cross-Validation or Maximum Likelihood Estimation method as a reference and the Wasserstein distance [3] between the Gaussian distributions as the optimal similarity measure. Finally, we reduce the complexity and Dimensionality of the problem by applying the *relaxation* method on the length-scale vector.

This method is applied successfully to academic examples and an industrial case that has motivated the approach. In the following example, we consider the Morokoff & Cafilisch [4] function, evaluated on an experimental design \mathbf{X} of $n = 600$ observations with $d = 10$ variables, and we consider the Matérn 5/2 anisotropic geometric correlation model. Table 1 shows the results before and after applying the RPIE method. It highlights the ability to achieve the optimal coverage level (here 90%) of Prediction Intervals by reducing their width and variance.

References

- [1] N. Acharki, A. Bertoncello, and J. Garnier. Robust prediction interval estimation for gaussian processes by cross-validation method, 2021. Submitted to *Computational Statistics & Data Analysis*.

Table 1: Performances of the estimated Prediction Intervals; Confidence level $1 - \alpha = 90\%$

| | Before RPIE | | After RPIE | |
|---------------------------------|------------------------|-----------------------------|--|---|
| | ML $\hat{\theta}_{ML}$ | MSE-CV $\hat{\theta}_{MSE}$ | ML $\hat{\theta}_{ML}$ | MSE-CV $\hat{\theta}_{MSE}$ |
| $\tilde{\mathbb{P}}_{1-\alpha}$ | 94.3 | 99.1 | 90.0 | 90.0 |
| $\text{CP}_{1-\alpha}$ | 95.3 | 98.7 | 88.7 | 92.0 |
| $\text{MPIW}_{1-\alpha}$ | $1.49 \cdot 10^{-1}$ | $1.66 \cdot 10^{-1}$ | $5.57 \cdot 10^{-2}$ (- 62.3%) | $6.18 \cdot 10^{-2}$ (- 62.7%) |
| $\text{SdPIW}_{1-\alpha}$ | $7.30 \cdot 10^{-3}$ | $5.01 \cdot 10^{-3}$ | $1.19 \cdot 10^{-2}$ (+ 63.0%) | $1.53 \cdot 10^{-2}$ (\times 2.05) |

$\text{CP}_{1-\alpha}$: Coverage Probability in %; $\tilde{\mathbb{P}}_{1-\alpha}$: The Leave-One-Out probability in %; MPIW: Mean of Prediction Interval widths and SdPIW: standard deviation of Prediction Interval widths.

- [2] F. Bachoc. Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66:55–69, 2013.
- [3] Valentina Masarotto, Victor M. Panaretos, and Yoav Zemel. Procrustes metrics on covariance operators and optimal transportation of gaussian processes. *Sankhya A*, 81(1):172–213, Feb 2019.
- [4] W. J. Morokoff and R. E. Caflisch. Quasi-monte carlo integration. *Journal of computational physics*, 122:218–230, 1995.
- [5] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

Short biography – Naoufal Acharki obtained his Engineer degree in September 2019 from Ecole des Mines de Saint-Etienne (France). He also holds MSc in applied mathematics from Paris-1 Panthéon Sorbonne University. He started his PhD thesis, funded by TotalEnergies, in October 2019. Its purpose is to develop a data-driven approach, coupling causality with statistical learning, for optimization and decision-making, with a focus on uncertainty quantification.