**Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification**

François Bachoc1, Josselin Garnier1 and Jean-Marc Martinez1

1CEA-Saclay, DEN, DM2S, STMP, LGLS, F-91919 Gif-sur-Yvette, France.

| Laboratory de Probabilités et Modèles Aléatoires, Université Paris VII |

**Introduction**

- **Estimation of hyper-parameters in Kriging in case of model misspecification.**
- **Goal:** Comparison of Maximum Likelihood (ML) and Cross Validation (CV).

**Framework**

- Observation of a centered, unit variance, stationary Gaussian process \( Y \) on \( X \) with covariance function \( C \).
- Kriging metamodel \( x_i \rightarrow (Y_i, \hat{y}_i, \hat{\sigma}_i^2) \) given by the set \( C \) of covariance functions:
  \[
  C = \left\{ \sigma^2 C_{p, a} \in \mathbb{R}^+: \sigma \in \mathbb{R} \right\}
  \]
  with \( C \) a stationary correlation function, \( C \) not \( \mathbb{R} \).

**Maximum Likelihood:**

- \( \theta_{ML} \in \arg\min_{\theta \in \mathbb{R}} \left( \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \right) \)
- \( \theta_{ML} \) is the Kriging mean and variance of \( Y_i \), with covariance function \( C \), based on \((x_1, \ldots, x_n, y_1, \ldots, y_n)\):
  \[
  \hat{\theta}_{ML} \approx \arg\min_{\theta \in \mathbb{R}} \left( \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \right) \]

- **Cross-Validation:**
  \[
  \hat{\theta}_{CV} \approx \arg\min_{\theta \in \mathbb{R}} \left( \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \right) \]

- Thanks to the virtual Leave One Out Formulas [Delah83] we have:
  \[
  \bar{\theta}_{CV} \approx \arg\min_{\theta \in \mathbb{R}} \left( \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \right) \]

**Outline:**

- **First step** of the estimation of the variance hyper-parameter \( \sigma^2 \): Closed form expression of \( \hat{\sigma}^2 \), allows for a detailed and quantitative finite sample comparison.
- **Second step** general case of the estimation of the hyper-parameter \( \theta \). Numerical studies on analytical functions.

**Step 1: Estimation of the variance hyper-parameter**

- In this case \( C_0 = C_2, C_2 \neq C_1 \).
- Quantity of interest for \( \sigma^2 \): The risk at \( x_0 \):
  \[
  R_{\sigma^2, x_0} = \mathbb{E} \left[ \left( \sigma^2 (Y_0 - \bar{y}_0)^2 \right)^2 \right].
  \]
  - The risk increases when the predictive variance is wrong.
- **Analytical expression of the risk for an estimator \( \sigma^2 \) as a function of the mean \( \mu_y \) and covariance function \( C \):**
  \[
  R_{\sigma^2, x_0} = f(A, B) = tr(AB(A) \otimes \sigma) + 2tr(AB) \otimes \sigma + \sigma^2 tr(M) \otimes M
  \]

  - With:
    \[
    f(A, B) = tr(AB(A) \otimes \sigma) + 2tr(AB) \otimes \sigma + \sigma^2 tr(M) \otimes M, M = M_1 - M_2,
    \]
  - Case \( C_1 = C_2 \). ML reaches the Chained Ros bound \( \frac{\sigma^2}{\sigma^2} \).
  - Case \( C_1 \neq C_2 \). Numerical evaluation of the risk formulas.

**Step 2: Estimation of the correlation hyper-parameters**

**Procedure**

- **Function f on \([0,1]^d\):**
  - Building of a Kriging Model with training sample \((x_1, \ldots, x_n, y_1, \ldots, y_n)\), with the exponential, Gaussian and Matérn covariance function, and with two different cases for the hyper-parameters estimation:
    - Case 2a: Estimation of an isotropic correlation length, and of the regularity parameter for the Matérn case.
    - Case 2b: Estimation of correlation lengths, and of the regularity parameter for the Matérn case.
- **Quantities of interest on a Monte Carlo test sample \((x_1, \ldots, x_n, y_1, \ldots, y_n)\):**
  - **Mean Square Error (MSE):**
    \[
    \sigma^2 = \mathbb{E} \left[ (\hat{y}_0 - \bar{y}_0)^2 \right]^2
    \]
  - **Predictive Variance Adequation (PVA):**
    \[
    \mathbb{P} \left( \hat{y}_0 < \mathbb{E}(\hat{y}_0) \right) \]

  - **Quantities of interest are averaged over 100 LHS Maximin designs.**

**Results**

- We consider the Ishigami \((d = 1)\) and Morris \((d = 10)\) functions:
  - Ishigami \(\sin(2\pi (x_2 + 1)) + 7 \sin(2\pi x_1) \times 10^{6} + 0.1 \sin(2\pi x_1) \times (2\pi x_2 - 1)^{4}\)
  - Morris: An anisotropic function.

**References**


**Table of the quantities of interest:**

<table>
<thead>
<tr>
<th>Quantity of interest</th>
<th>Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk on Target Ratio (RTR)</td>
<td>( R_{\sigma^2, x_0} = \sqrt{R_{\sigma^2, x_0}} )</td>
</tr>
<tr>
<td>Integrated Risk on Target Ratio (IRTR)</td>
<td>( R_{\sigma^2, x_0} = \int_{RTR(x_0)}^{RTR(x_0)} d\sigma )</td>
</tr>
<tr>
<td>Bias on Target Ratio (BTR)</td>
<td>( R_{\sigma^2, x_0} = \int_{BTR(x_0)}^{BTR(x_0)} d\sigma )</td>
</tr>
</tbody>
</table>

**Procedure**

- We take \( X = \{0,1\}^d \) with uniform measure. We generate \( n_p \) designs \((x_1, \ldots, x_n)\) using the LHS-Maximin technique, compute each time the four criteria above (analytical formulation and Monte Carlo for integration) and plot the average.
- Setting for the figures.

**Conclusion**

- In our studies: When the model misspecification becomes important, CV performs better than ML.
- Possible extension: Studying other Cross Validation estimation methods.