Monotonicity-preserving bootstrapped kriging metamodels for expensive simulations

Kleijnen, Jack P.C.; van Beers, W.C.M.

Published in:
The Journal of the Operational Research Society

Publication date:
2013

Online Appendix for "Monotonicity-preserving bootstrapped Kriging metamodels for expensive simulations"

Jack P.C. Kleijnen[1] and Wim C.M. van Beers[2]
[1]Tilburg University, Postbox 90153, 5000 LE Tilburg, Netherlands, Kleijnen@tilburguniversity.edu [2]University of Amsterdam, Postbox 19268, 1000 GG Amsterdam, Netherlands, e-mail: W.C.M.vanBeers@uva.nl

Abstract

This appendix consists of two subappendixes; namely, one with basic Kriging formulas, and one with basic linear regression formulas and the estimated coverages for classic Kriging, monotonicity-preserving bootstrapped Kriging, and polynomial regression metamodels.

Version: August 30, 2011

Basic Kriging Formulas

Ordinary Kriging assumes

$$w(\mathbf{x}) = \mu + \delta(\mathbf{x}) \tag{1}$$

where μ is the simulation output averaged over the experimental area, and $\delta(\mathbf{x})$ is the additive noise that forms a covariance stationary process with zero mean. Ordinary Kriging uses the linear predictor

$$y = \lambda' \mathbf{w} \tag{2}$$

with the optimal weights under the MSE criterion are

$$\lambda_o = \mathbf{\Gamma}^{-1}[\gamma + \mathbf{1}\frac{1 - \mathbf{1}'\mathbf{\Gamma}^{-1}\gamma}{\mathbf{1}'\mathbf{\Gamma}^{-1}\mathbf{1}}] \tag{3}$$

where $\mathbf{\Gamma} = (cov(w_i, w_{i'}))$ with $i, i' = 1, \ldots, n$ is the $n \times n$ symmetric and positive semi-definite matrix with the covariances between the n old outputs, and $\gamma = (cov(w_i, w_0))$ is the n-dimensional vector with the covariances between the n old outputs w_i and w_0, the output of the combination to be predicted—which may be either new or old. The

correlation function for a k-dimensional input vector is assumed to be the product of k one-dimensional functions ρ_j $(j = 1, \ldots, k)$; a popular one-dimensional correlation function is the Gaussian one:

$$\rho_j = \exp[-\theta_j h_j^2] \tag{4}$$

where $h_j = |x_{i;j} - x_{i';j}|$ denotes the Euclidean distance between the values of input j in the two input combinations i and i'; θ_j denotes the importance of input j; i.e., the higher θ_j is, the less effect input j has. In practice, these covariances (or correlations) are unknown so they are estimated, usually by Maximum Likelihood Estimation (MLE), which gives $\widehat{\mu}$, $\widehat{\mathbf{\Gamma}}$ and $\widehat{\gamma}$ (or $\widehat{\theta_j}$). Combining these estimators with (1), (2), and (3) gives the predictor

$$\widehat{y} = \widehat{\mu} + \widehat{\gamma}'\widehat{\mathbf{\Gamma}}^{-1}(\mathbf{w} - \widehat{\mu}\mathbf{1}) \tag{5}$$

where $\widehat{\mu} = (\mathbf{1}'\widehat{\mathbf{\Gamma}}^{-1}\mathbf{1})^{-1}\mathbf{1}'\widehat{\mathbf{\Gamma}}^{-1}\mathbf{w}$; this predictor is nonlinear because it uses estimators for the covariances.

An alternative for Ordinary Kriging is *Universal Kriging*, which replaces the constant μ in (1) by a linear combination of known functions; e.g., a low-order polynomial (Cressie, 1993, p. 151), (Lophaven et al., 2002, p. 13). Ordinary Kriging is recommended by most authors; nevertheless, some authors recommend Blind Kriging (Joseph et al., 2008).

Basic Linear Regression Formulas

In our linear regression we assume

$$y(\mathbf{x}) = \sum_{j=1}^{q} \beta_j x_j + \epsilon(\mathbf{x}) = \mathbf{x}'\beta + \epsilon(\mathbf{x}) \tag{6}$$

where x_j is the j^{th} explanatory regression variable, $\mathbf{x} = (x_1, \ldots, x_q)'$ is the input combination, $\beta = (\beta_1, \ldots, \beta_q)'$ is the vector of regression parameters, and $\epsilon(\mathbf{x})$ is the additive noise with zero mean and variances that may vary with $\mathbf{x}$; because we do not use CRN, the noise terms at different points are independent.

Because the variances of the simulation outputs are unknown, we proceed as follows—but there are alternatives (Kleijnen, 2008, pp. 87-91). We use OLS to estimate β from the bootstrapped outputs $\overline{\mathbf{w}}_b^*$ $(b = 1, \ldots, B)$ with bootstrap sample size B (e.g., $B = 100$ in our experiment):

$$\widehat{\beta_b^*} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{w}}_b^* \ (b = 1, \ldots, B)$$

	lower bound	median	upper bound
KA	0.150	0.192	0.234
BA	0.454	0.503	0.552
PA	0.160	0.173	0.186
KQ	0.204	0.251	0.298
BQ	0.339	0.365	0.391
PQ	0.183	0.194	0.206

Table 1: Coverage in classic Kriging (K) and monotonicity-preserving bootstrapped Kriging (B), and Polynomial regression (P), for the Average (A) and the 90% Quantile (Q),), with n = 5, m = 5, T = 1000

where $\mathbf{X}$ is the $n \times q$ matrix of explanatory variables in the n simulated combinations.The corresponding regression predictor for point $\mathbf{x}_u$ is

$$\widehat{y^*_{u;b}} = \sum_{j=1}^{q} \widehat{\beta^*_{j;b}} x_{j;u} = \mathbf{x}'_u \widehat{\beta^*_b} \ (u = 1, \ldots, \upsilon).$$

Our corresponding $(1 - \alpha)$ confidence interval for the true output ζ_u is

$$\widehat{y^*_{u;(\lfloor 0.05B \rfloor)}} < \zeta_u < \widehat{y^*_{u;(\lceil 0.95B \rceil)}}.$$

This gives Table 1.

Note: The known shape of the polynomial regression model may be preserved through semidefinite programming and real algebraic geometry (Siem et al., 2008).

References

Cressie, N.A.C. (1993), *Statistics for spatial data (revised edition).* Wiley, New York

Joseph, V. R., Y. Hung, and A. Sudjianto, ((2008), Blind Kriging: a new method for developing metamodels. *Journal of Mechanical Design*, 130, no. 3, pp. 31-102

Kleijnen, J.P.C. (2008), *Design and analysis of simulation experiments.* Springer Science + Business Media

Lophaven, S.N., H.B. Nielsen, and J. Sondergaard (2002), DACE: a Matlab Kriging toolbox, version 2.0. IMM Technical University of Denmark, Lyngby

Siem, A.Y.D., E. de Klerk, and D. den Hertog (2008), Discrete least-norm approximation by nonnegative (trigonometric) polynomials and rational functions. *Structural Multidisciplinary Optimization*, 35, pp. 327–339