Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

BOOTSTRAPPING AND CONDITIONAL SIMULATION IN KRIGING: BETTER CONFIDENCE INTERVAL AND OPTIMIZATION?

Ehsan Mehdad
Tilburg University
Warandelaan 2
5037 AB Tilburg, NETHERLANDS

ABSTRACT

This paper investigates two related questions: (1) How to derive a confidence interval for the output of a combination of simulation inputs not yet simulated? (2) How to select the next combination to be simulated when searching for the optimal combination? To answer these questions, the paper uses parametric bootstrapped Kriging and "conditional simulation", whereas classic Kriging estimates the variance of its predictor by plugging-in the estimated GP parameters so this variance is biased. The main conclusion is that classic Kriging seems quite robust; i.e., classic Kriging gives acceptable confidence intervals and estimates of the optimal solution.

1 INTRODUCTION

In this paper we address the following two related questions that arise in simulation, especially when the simulation is "computationally expensive": (i) How to derive a *confidence interval* (CI) for the output of a combination of simulation inputs that is not yet simulated? (ii) How to select the next combination that is to be simulated, when searching for the *optimal* combination? Question 1 (Q1) arises in sensitivity analysis or "what if" analysis. Question 2 (Q2) arises in "simulation optimization", which aims at finding the input combination—also called scenario or point—that gives the minimal simulation output (response); we limit our optimization to unconstrained problems, like many authors do.

To answer these two questions, simulation analysts often use metamodels. A popular type of metamodel is a *Kriging* or *Gaussian process* (GP) model. *Classic Kriging* (CK) estimates the variance of its predictor by plugging-in the estimated parameters (say) $\widehat{\psi}$ of the assumed stationary GP; we assume a GP with covariance matrix Σ_M and parameter vector ψ consisting of the constant mean β_0, the constant variance τ^2, and the correlation vector θ. Unfortunately, plugging-in $\widehat{\psi}$ makes the Kriging predictor nonlinear so $s^2(\mathbf{x})$, the classic variance estimator of the Kriging predictor at point $\mathbf{x}$, is *biased*. The literature has not thoroughly investigated the consequences of this bias. We therefore (empirically) compare the CI of CK and alternative CIs (see Q1).

Moreover, $s^2(\mathbf{x})$ is also used in *efficient global optimization* (EGO), which is a well-known sequential method that balances local and global search; i.e., EGO combines exploitation and exploration. The classic EGO article is Jones et al. (1998).

The main contribution of this paper is the use of conditional simulation (CS) to improve the estimation of the Kriging predictor variance. We detail CS in § 2.

A preliminary version of our research was presented in Kleijnen and Mehdad (2013), comparing the estimated variances of the Kriging predictors in CK, BK (bootstrapped Kriging), and CS and their effects on EGO. Now we investigate the role of this variance in CIs (see again Q1); i.e., what are the coverages and lengths of the CIs when using these three methods? Moreover, for these CIs we use either the classic *parametric* method assuming the predictor is unbiased and normally distributed—even though the predictor is nonlinear—and a *distribution-free* method using the percentile method that is classic in

bootstrapping. Furthermore, we present details on additional examples; i.e., to the detailed one-dimensional example in Kleijnen and Mehdad (2013) we add three well-known higher-dimensional examples; namely Camel-back, Hartmann-3, and Hartmann-6.

We limit our research to *deterministic* simulation, which is popular in engineering, and will be the basis for our future research on random discrete-event simulation.

2 CONDITIONAL SIMULATION

To answer Q1 and Q2 when using Kriging, we compare the following alternative methods with each other and with CK: (i) *B*K originally proposed in Den Hertog et al. (2006) to examine $s^2(\mathbf{x})$ and in Kleijnen et al. (2012) to examine EGO; (ii) CS which is popular in the French literature on Kriging.

We formulate the *basic idea* of CS in Chilès and Delfiner 1999, pp. 465-469 as follows. Let $S(\cdot)$ be a non-conditional simulation (or bootstrap sample) of $Y(\cdot)$ independent of $Y(\cdot)$ and with the same covariance as $Y(\cdot)$. When " conditioning", we pass from $S(\cdot)$ to a simulation $Y_{\text{CS}}(\cdot)$ that equals $Y(\cdot)$ in the old points. Let $\widehat{Y}(\mathbf{x}_0)$ be the Kriging predictor of $Y(\mathbf{x}_0)$ based on the old I/O data $(\mathbf{X}, \mathbf{w})$. Obviously, we have $Y(\mathbf{x}_0) = \widehat{Y}(\mathbf{x}_0) + [Y(\mathbf{x}_0) - \widehat{Y}(\mathbf{x}_0)]$ where the Kriging error $Y(\mathbf{x}_0) - \widehat{Y}(\mathbf{x}_0)$ is unknown because $Y(\mathbf{x}_0)$ is unknown. Analogously, we have $S(\mathbf{x}_0) = \widehat{S}(\mathbf{x}_0) + [S(\mathbf{x}_0) - \widehat{S}(\mathbf{x}_0)]$, but now $S(\mathbf{x}_0)$ is known and so is the error term. Substituting the simulated error into the decomposition of $Y(\mathbf{x}_0)$, we obtain $Y_{\text{CS}}(\mathbf{x}_0) = \widehat{Y}(\mathbf{x}_0) + [S(\mathbf{x}_0) - \widehat{S}(\mathbf{x}_0)]$. Because Kriging is an exact interpolator at an old point $\mathbf{x}$, we have $\widehat{Y}(\mathbf{x}) = Y(\mathbf{x})$ and $\widehat{S}(\mathbf{x}) = S(\mathbf{x})$ so $Y_{\text{CS}}(\mathbf{x}) = Y(\mathbf{x})$.

We detail our CS algorithm as follows.

1. Compute $\widehat{\psi}$ from the old I/O data$(\mathbf{X}, \mathbf{w})$.
2. Given $\widehat{\psi}$ from step 1, use $\text{N}_k\left(\widehat{\beta}_0 \mathbf{1}_k, \widehat{\Sigma}_M\right)$ to sample the k old points $\mathbf{w}_b^*(\mathbf{X}, \widehat{\psi}) = (w_{1;b}^*(\mathbf{X}, \widehat{\psi}), \ldots, w_{k;b}^*(\mathbf{X}, \widehat{\psi}))^\top$.
3. Given the k old points $\mathbf{w}_b^*(\mathbf{X}, \widehat{\psi})$ of step 2, use the conditional normal distribution to sample the new output $w_b^*(\mathbf{x}_0, \widehat{\psi})$.
4. Given $\mathbf{w}_b^*(\mathbf{X}, \widehat{\psi})$ from Step 2 compute $\widehat{\psi}_b^*$.
5. Given $\widehat{\psi}_b^*$ of step 4, calculate the new output $\widehat{Y}(\mathbf{x}_0, \widehat{\psi}_b^*)$.
6. Combining $\widehat{Y}(\mathbf{x}_0, \widehat{\psi}_b^*)$ of step 5, $w_b^*(\mathbf{x}_0, \widehat{\psi})$ of step 2, and CK's $\widehat{Y}(\mathbf{x}_0, \widehat{\psi})$, compute the CS output $\widehat{Y}_{\text{CS}}(\mathbf{x}_0, b) = \widehat{Y}(\mathbf{x}_0, \widehat{\psi}) + [w_b^*(\mathbf{x}_0, \widehat{\psi}) - \widehat{Y}(\mathbf{x}_0, \widehat{\psi}_b^*)]$.
7. Repeat steps 2 through 6 B times (so $b = 1, \ldots, B$), and compute $\widehat{\sigma}^2[\widehat{Y}_{\text{CS}}(\mathbf{x}_0)]$.

3 CONCLUSIONS

Our main conclusion is that CK seems quite robust; i.e., (i) BK and CS give CIs with coverages and lengths that are not significantly better than CK gives; (ii) EGO with BK or CS may or may not perform better in expensive simulation with small samples.

REFERENCES

Chilès, J., and P. Delfiner. 1999. *Geostatistics: Modeling Spatial Uncertainty*. 1 ed. Wiley.

Den Hertog, D., J. Kleijnen, and A. Siem. 2006. "The correct Kriging variance estimated by bootstrapping". *Journal of The Operational Research Society* 57:400–409.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13:455–492.

Kleijnen, J. P. C., and E. Mehdad. 2013. "Conditional simulation for efficient global optimization". In *Proceedings of the 2013 Winter Simulation Conference (WSC)*. (Accepted).

Kleijnen, J. P. C., W. Van Beers, and I. Van Nieuwenhuyse. 2012. "Expected improvement in efficient global optimization through bootstrapped Kriging". *Journal of Global Optimization* 54:59–73.