Simulation Modelling Practice and Theory 66 (2016) 166–173

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Estimating the variance of the predictor in stochastic Kriging

Jack P.C. Kleijnen, Ehsan Mehdad*

Tilburg School of Economics and Management, Tilburg University, Tilburg, Netherlands

ARTICLE INFO

Article history:
Received 24 July 2015
Revised 29 March 2016
Accepted 30 March 2016
Available online 4 May 2016

Keywords:
Kriging
Gaussian process
Predictor variance
Plug-in
Bootstrap

ABSTRACT

We study the estimation of the true variance of the predictor in stochastic Kriging (SK). First, we obtain macroreplications for a SK metamodel that approximates a single-server simulation model; these macroreplications give independently and identically distributed predictions. This simulation may use common random numbers (CRN). From these macroreplications we conclude that the usual plug-in estimator of the variance significantly underestimates the true variance. Because macroreplications of practical simulation models are computationally expensive, we next formulate two bootstrap methods that use a single macroreplication: (i) a distribution-free method that resamples simulation replications (within the single macroreplication), and (ii) a parametric method that assumes a Gaussian distribution for the SK predictor, and estimates the (hyper)parameters of that distribution from the single macroreplication. Altogether we recommend distribution-free bootstrapping for the estimation of the SK predictor variance in practical simulation experiments.

1. Introduction

In practice, the final goal of simulation is often sensitivity analysis and optimization of the simulated real system. For this goal, the simulation analysts often use a *metamodel* – also called an emulator or a surrogate – which is a model of the underlying simulation model; i.e., it is an explicit simple approximation of the input/output (I/O) function that is implicitly defined by the simulation model. There are many types of metamodels; see Kleijnen [12, p. 10]. We, however, focus on *Kriging* – also called *Gaussian process* (GP) – metamodels. These metamodels have already acquired a track record in deterministic simulation, and are becoming popular in random (stochastic) simulation – including discrete-event simulation – which transforms a stream of *pseudo-random numbers* (PRN) into an output. Kriging in stochastic simulation is called *stochastic Kriging* (SK); see the classic paper Ankenman et al. [2], recent papers such as Barton et al. [3], Bekki et al. [4], Chen and Kim [6], Plumlee and Tuo [16], Qu and Fu [17], and Sun et al. [21], and the many more references in Kleijnen [12, pp. 206–211].

Common random numbers (CRN) are a popular variance-reduction technique that may improve the statistical analysis of the simulation I/O data in sensitivity analysis and optimization; see Law [13, pp. 586–604]. However, CRN also complicate this analysis when using SK; see Chen et al. [5].

In practice, the (hyper)parameters of the GP are unknown, so they are estimated. Using these estimated parameters makes the linear GP predictor nonlinear (see (8) below), so it is difficult to estimate the variance of this predictor. Many GP publications simply plug-in the estimated parameters into the formulas that assume known parameters. Obviously, this

* Corresponding author. Tel.: +31 134668202.
E-mail addresses: kleijnen@tilburguniversity.edu (J.P.C. Kleijnen), emehdad@gmail.com, E.Mehdad@tilburguniversity.edu (E. Mehdad).

http://dx.doi.org/10.1016/j.simpat.2016.03.008
S1569-190X(16)30167-8/

plugging-in does not give the true variance of the predictor. We focus on the problem of correctly estimating the true variance of the predictor that uses SK with *estimated* Kriging parameters. The fundamental reference Abt [1] proposes a quite simple analytic approximation assuming an exponential correlation function for the so-called extrinsic noise and a constant (homogeneous) variance for the intrinsic noise (these two types of noise are defined below (1)). Recently Thiart et al. [23] compared five estimators of the predictor variance, using Abt's [1] assumptions. We, however, propose well-known bootstrapping methods (for which ample software is available; see Section 2) that allows any type of correlation function for the intrinsic noise (but we present numerical results only for the correlation function that is most often assumed in SK; namely, a Gaussian correlation function) and intrinsic noise with heterogeneous variances and possible correlations caused by CRN. Simamora et al. [20] compares our parametric bootstrapping with semiparametric bootstrapping in deterministic simulation. A more general class of problems including our problem is discussed in Harville [10].

In our paper we investigate the following three methods for estimating the variance of the SK predictor: (i) macroreplications, (ii) distribution-free bootstrapping (DB), and (iii) parametric bootstrapping (PB). We illustrate and evaluate these three methods through experiments with a waiting-time simulation model; namely, a single-server discrete-event simulation model with independent exponential interarrival and service times, known as the M/M/1 model. For details on the simulation of this model we refer to Law [13, pp. 73–77, 102–108].

We organize and summarize the rest of this paper as follows. In Section 2 we summarize SK. In Section 3 we first describe an M/M/1 simulation experiment, and then explain how macroreplications provide an unbiased estimator of the predictor variance; we find that this estimator is significantly higher than the plug-in estimator – without or with CRN. In Section 4 we detail DB, and find that DB is a relatively quick and easy method for estimating the true variance when CRN are applied. In Section 5 we detail PB, and find that PB gives a higher estimate than DB does. In Section 6 we summarize our conclusions; namely, DB gives a fast and relatively good estimator of the predictor variance in practical simulation experiments.

2. Stochastic Kriging

SK assumes that the simulation output (say) w is stochastic. Notice that unlike most authors on SK, we do distinguish between w (simulation output) and y (metamodel output). Furthermore, we use Greek letters to denote unknown parameters, and bold upper-case letters for matrixes and bold lower-case letters for vectors. The simplest type of SK assumes the following metamodel:

$$y_r(\mathbf{x}) = \mu + M(\mathbf{x}) + \varepsilon_r(\mathbf{x}) \text{ with } \mathbf{x} \in \mathbb{R}^k \text{ and } r = 1, \ldots, m_i, \tag{1}$$

where we use the following symbols. We let μ denote the constant mean $E[y(\mathbf{x})]$ where $\mathbf{x}$ is an input combination or 'point' in the given k-dimensional experimental area $\mathbb{R}^k$. Ankenman et al. [2] – the classic paper on SK – defines the so-called *extrinsic noise* $M(\mathbf{x})$ as the additive noise that forms a Gaussian (multivariate normal) stationary process with zero mean and covariance matrix $\mathbf{\Sigma}_M$ (by definition, a stationary process has a constant mean, a constant variance, and covariances that depend only on the distance between the points $\mathbf{x}$ and $\mathbf{x}'$). Because stochastic simulation has noisy outputs, we should obtain *replications*; i.e., *identically independently distributed* (IID) simulation outputs. Ankenman et al. [2] defines the *intrinsic noise* $\varepsilon_r(\mathbf{x})$ as having a Gaussian distribution with zero mean and variance $\text{Var}[\varepsilon(\mathbf{x})]$ and being independent of the extrinsic noise $M(\mathbf{x})$. We let m_i denote the number of replications if $\mathbf{x} = \mathbf{x}_i$ so the subscript r runs from 1 to m_i. We let $\mathbf{\Sigma}_\varepsilon$ denote the covariance matrix of ε. So, if the simulation does not use CRN, then $\mathbf{\Sigma}_\varepsilon$ is diagonal with the elements $\text{Var}[\varepsilon(\mathbf{x})]$ on the main diagonal. If the simulation does use CRN, then $\mathbf{\Sigma}_\varepsilon$ is not diagonal; obviously, $\mathbf{\Sigma}_\varepsilon$ should still be symmetric and positive definite.

Note: The SK model defined in (1) – but without intrinsic noise – is known as *ordinary Kriging* (OK). A more general Kriging model replaces the constant μ in (1) by a prespecified low-order polynomial in $\mathbf{x}$; this Kriging is known as *universal Kriging* (UK). If UK specifies (say) a first-order polynomial in $\mathbf{x}$, then UK requires the estimation of the k effects β_j ($j = 1, \ldots, k$), besides β_0 or μ in OK. The estimation of these k extra parameters may explain why in practice UK often does not give better predictions; also see Tajbakhsh et al. [22]. In this paper we focus on estimating the predictor variance in SK with a constant mean μ.

Averaging the m_i replications gives the average metamodel output $\bar{y}(\mathbf{x}_i)$ and average intrinsic noise $\bar{\varepsilon}(\mathbf{x}_i)$, so (1) is replaced by

$$\bar{y}(\mathbf{x}_i) = \mu + M(\mathbf{x}_i) + \bar{\varepsilon}(\mathbf{x}_i) \text{ with } \mathbf{x}_i \in \mathbb{R}^k \text{ and } i = 1, \ldots, n, \tag{2}$$

where n denotes the number of so-called *old* simulated points. If we do not use CRN, then $\mathbf{\Sigma}_{\bar{\varepsilon}}$ is a diagonal matrix with main-diagonal elements $\text{Var}[\varepsilon(\mathbf{x}_i)]/m_i$. If we do use CRN and m_i is a constant m, then $\mathbf{\Sigma}_{\bar{\varepsilon}} = \mathbf{\Sigma}_\varepsilon/m$. Notice that $\mathbf{x}_i = (x_{i;j})$ with $j = 1, \ldots, k$ and k defined below (1), so $\mathbf{X}' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is a $k \times n$ matrix; $\mathbf{X}$ is the so-called design matrix.

Note: In this paper we focus on estimating the predictor variance in SK – given a design matrix $\mathbf{X}$; i.e., we do not discuss the *optimal* choice of $\mathbf{X}$. For that choice we refer to the literature; e.g., Zimmerman [26] distinguishes between (i) a design that gives an 'optimal' Kriging predictor while assuming that the extrinsic covariance matrix $\mathbf{\Sigma}_M$ is known, (ii) a design that gives an 'optimal' estimator of this matrix $\mathbf{\Sigma}_M$, and (iii) a hybrid design that balances the design objectives (i) and (ii). Zimmerman [26, p. 651] assumes OK, not UK. Our M/M/1 example has a single input (so $k = 1$) – namely, the traffic rate x – and we simulate $n = 11$ equispaced traffic rates x_i between 0.10 and 0.90 (as detailed in Section 3); our design for the M/M/1