Available online at www.sciencedirect.com
ScienceDirect
Procedia Social and Behavioral Sciences 2 (2010) 7728–7729
Procedia
Social and Behavioral Sciences

Sixth International Conference on Sensitivity Analysis of Model Output

Modeling uncertainties in complex systems

Nabil Rachdi[a]*, Jean-Claude Fort[b], Thierry Klein[c], Fabien Mangeant[a]

[a]EADS IW, 12 rue Pasteur, Suresnes 92152, France
[b]Université Paris V, 12 rue de l'école de médecine, Paris 75006, France
[c]Université Toulouse III, 118 route de Narbonne, Toulouse 31062, France

Abstract

The aim of this work is to develop quantitative approaches to manage uncertainty of variables or functions modeled by complex systems where both experimental and simulation data are available. We present the case of variables, where in current practices, an uncertainty study could be understood as a quantity of interest study of the model output: mean, quantile, threshold probability, etc…This situation is very common in engineering, where complex models exist and where the experimental data are difficult to obtain. First, we propose a method for model calibration based on experimental and simulation data, then we prove the consistency of this calibration procedure. The main tool used here is the empirical processes theory. Our final purpose is to incoporate simulated data from a complex model into an estimator (of a quantity of interest) based on experimental data, and then to compare the performance of our estimators to the classical estimators based on experimental data only.

Keywords: Model calibration; experimental and simulation data; M-estimation; empirical processes; consistency.

1. Introduction

We model a complex phenomenon by a random variable $Y \in \mathbf{R}$ with unknown probability density f, the associated probability measure is noted $\boldsymbol{Q}$. We assume that we have *n* experimental data of this phenomenon, $Y_1,..., Y_n$, considered as a sample of independent and identically-distributed realizations of Y. In practice, the number of experimental data is limited and it doesn't provide enough information on the variable Y. To overcome this lack of knowledge, we consider a numerical model (for example) which represents the complex phenomenon Y. We consider a function $h: R^d \times R^k \to R$ such that $Y \approx h(X,\theta)$ where $X = (X^1,..., X^d)^T$ is a random vector with probability distribution P_X (can be either known or not), and $\theta = (\theta^1,..., \theta^k)^T$ is the vector of parameters supposed to be in a compact set $\Theta \subset R^k$. The vector X represents uncertain variables of the model h and the propagation of this uncertainty through the model h under the parameter θ gives the random output $h(X,\theta)$. Our first goal is to compute a parameter $\theta_0(h) \in \Theta$ such that the numerical model h, under the uncertainty of X and $\theta_0(h)$, behaves like the complex phenomenon Y according to the quantity of interest considered. For this, we suppose that we have a sample of size *m* of the model input X : $X_1,..., X_m$ where X_i, $i = 1,..., m$ are supposed to be drawn from the distribution P_X. It provides *m* simulated data for each fixed $\theta \in \Theta$: $h(X_1,\theta),..., h(X_m,\theta)$, and we propose a

* Tel.: +33 (0)1 46 97 34 85; fax: +33 (0)1 46 97 31 12.
E-mail address: nabil.rachdi@eads.net.

doi:10.1016/j.sbspro.2010.05.200

method for estimating $\theta_0(h)$. It amounts to minimizing a criterion $M_{n,m}(h,\theta)$ built from an objective function $M(h,\theta)$ supposed to be minimum in (h^*,θ^*). This criterion depends on experimental and simulated data and takes the form:

$$M_{n,m}(h,\theta) = \frac{1}{n}\sum_{i=1}^{n}\gamma_{h,\theta}^{m}(Y_i)$$

where the functions $\gamma_{h,\theta}^{m}(.)$ indexed by θ depend on $X_1,\ldots,X_m$. We denote by $\hat{\theta}_{n,m}(h)$ the minimum of $M_{n,m}(h,\theta)$ over Θ for a fixed model *h*.

In this study, we investigate the consistency of this calibration procedure when *n* and *m* go to infinity, and we prove it under some conditions in terms of model complexity through entropy measure. An important point is to control the following *risk excess* , we show the inequality:

$$M(h,\hat{\theta}_{n,m}(h)) - M(h^*,\theta^*) \leq \frac{1}{\sqrt{n}}\sup_{\theta\in\Theta}\left|G_n\gamma_{h,\theta}^{m}\right| + \sup_{\theta\in\Theta}\left\|\gamma_{h,\theta}^{m} - \gamma_{h,\theta}\right\|_{1,Q} + \Delta_h$$

where $G_n = \sqrt{n}(Q_n - Q)$ is the *Q-empirical process* and $\|u\|_{1,Q} = \int_R |u(y)| f(y)dy$. The first term $\frac{1}{\sqrt{n}}\sup_{\theta\in\Theta}\left|G_n\gamma_{h,\theta}^{m}\right|$ is considered as an ***estimation error***, the second $\sup_{\theta\in\Theta}\left\|\gamma_{h,\theta}^{m} - \gamma_{h,\theta}\right\|_{1,Q}$ as a ***simulation error***, and Δ_h is defined as the ***approximation error***.

We also obtain results about the effect of the number of experimental data and simulation data on the quantity of interest estimation based on both experimental and simulated data. Comparisons are done with the estimation using experimental data only.
These results will be illustrated by some numerical examples.

2. References

Aad W. van der Vaart and Jon A. Wellner, 1996: “Weak convergence and empirical processes”, *Springer Series in Statistics.*

Aad W. van der Vaart, 1998: “Asymptotics Statistics”,*Cambridge Series in Statistical and Probabilistic Mathematics.*

David W. Scott, 1992: “Multivariate Density Estimation”,*Wiley Series in Probability and Mathematical Statistics.*

P. Billingsley, 1999: “Convergence of Probability Measures”,*Wiley Series in Probability and Statistics.*

Sara van de Geer, 2000: “Empirical Processes in M-Estimation”,*Cambridge Series in Statistical and Probabilistic Mathematics.*

Galen R. Shorack and Jon A. Wellner, 1986: “Empirical Processes with Applications to Statistics”,*Classics in Applied Mathematics.*