On the Use of Random Set Theory to Bracket the Results of Monte Carlo Simulations

FULVIO TONON
Department of Geology and Geophysics, University of Utah, 135 South 1460 East, Room 719, Salt Lake City, UT 84112, USA, e-mail: tonon@chpc.utah.edu

(Received: 6 September 2002; accepted: 8 April 2003)

Abstract. Based on Random Set Theory, procedures are presented for bracketing the results of Monte Carlo simulations in two notable cases: (i) the calculation of the entire distribution of the dependent variable; (ii) the calculation of the CDF of a particular value of the dependent variable (e.g. reliability analyses). The presented procedures are not intrusive in that they can be equally applied when the functional relationship between the dependent variable and independent variables is known analytically and when it is a complex computer model (black box). Also, the proposed procedures can handle probabilistic (with any type of input joint PDF), interval-valued, set-valued, and random set-valued input information, as well as any combination thereof.

When exact or outer bounds on the function image can be calculated, the bounds on the CDF of the dependent variable guarantee 100% confidence, and allow for an explicit and exact evaluation of the error involved in the calculation of the CDF. These bounds are often enough to make decisions, and require a minimal amount of functional evaluations. A procedure for effectively approximating the CDF of the dependent variable is also proposed.

An example shows that, compared to Monte Carlo simulations, the number of functional evaluations is reduced by orders of magnitude and that the convergence rate increases tenfold.

1. Introduction

To provide verified bounds on the results of Monte Carlo simulations is not only interesting from a speculative point of view, but also useful from an application point of view, for the calculated bounds may be enough to make a decision.

Consider a deterministic computer model, which computes an output $y$ from a vector input $u$. If the input $u$ is a random vector, one is interested in knowing how uncertainty propagates through the model. If the model can be run inexpensively thousands or millions of times, Monte Carlo methods [1], [28], [34], [42] can be used to make the required inferences on the uncertainty distribution.

In this paper, we consider the case in which the computer model is expensive, to the extent that Monte Carlo methods are impractical. Random Set Theory ([12], [13], and [46], [47] for a list of references) is used to bracket with 100% confidence the results of Monte Carlo simulations: methods are presented for calculating the entire Cumulative Distribution Function (CDF) of $y$, or the CDF of a particular value $y$ of interest. The latter case is of special interest in reliability analysis, where the entire CDF is often not required.
At the outset, the paper introduces some basic definitions of Random Set Theory, and then proceeds to consider the use of Interval Analysis for calculating function ranges. Section 4 deals with the crucial inclusion properties of Random Sets, and Section 5 presents methods for bracketing the results of Monte Carlo simulations. The specialization to reliability analysis is given in Section 6. Finally, the application to a specific example shows in detail how to apply the presented procedures.

It is often maintained that the number of functional evaluations required by Monte Carlo methods does not depend on (or is nearly independent of) the dimension of \( u \). The Appendix presents counterexamples to this belief.

2. Random Set Theory

2.1. Basic Definitions

Let \( U \) be a set of vectors \( u = (u_1, \ldots, u_p) \), and let \( \mathcal{P}(U) \) be the power set of \( U \) (set of all the subsets of \( U \)). Call basic probability assignment, a function \( m : \mathcal{P}(U) \to [0, 1] \) such that [13]:

\[
m(\emptyset) = 0, \quad (2.1)
\]

\[
\sum_{A \in \mathcal{P}(U)} m(A) = 1. \quad (2.2)
\]

If \( m(A_i) > 0 \), \( A_i \) is called a focal element. A random set is the pair \((\mathcal{F}, m)\) where \( \mathcal{F} \) is the family of all focal elements. In Random Set Theory, instead of counting outcomes of singletons \( u \in U \), outcomes of subsets \( A_i \subseteq U \) are counted and an observation of a subset \( A_i \subseteq U \) indicates an event somewhere in \( A_i \), without any specification of the probability that the event be a specific point of \( A_i \). Any distribution of the weight \( m(A_i) \) can be thought of on \( A_i \), as long as the distribution measure is equal to \( m(A_i) \). As a result, it is not possible to calculate the probability, \( \Pr \), of a generic \( u \in U \) or of a generic subset \( E \subseteq U \), but only lower and upper bounds on this probability [4], [8]:

\[
\text{Bel}(E) \leq \Pr(E) \leq \text{Pl}(E), \quad (2.3)
\]

where (let \( E^c \) denote the complement of \( E \)):

\[
\text{Bel}(E) = \sum_{A_i : A_i \subseteq E} m(A_i) = 1 - \text{Pl}(E^c), \quad (2.4a)
\]

\[
\text{Pl}(E) = \sum_{A_i : A_i \cap E \neq \emptyset} m(A_i) = 1 - \text{Bel}(E^c). \quad (2.4b)
\]

\text{Bel}(E) \text{ is called belief measure and is the sum of the frequencies of those focal elements contained in } E \text{ and whose occurrence must then lead to the claim that } u \in E. \]

\text{Pl}(E) \text{ is said plausibility measure and is the sum of the frequencies of those focal elements having some element } u \text{ in common with } E \text{ and whose occurrence } \text{may lead to the claim that } u \in E. \]