

Sensitivity of Monte Carlo Modeling in Crash Reconstruction

2010-01-0071

Published
04/12/2010Garrett Fleck and Jeremy S. Daily
The University of Tulsa

Copyright © 2010 SAE International

Abstract

The Monte Carlo method is a well-known technique for propagating uncertainty in complex systems and has been applied to traffic crash reconstruction analysis. The Monte Carlo method is a probabilistic technique that randomly samples input distributions and then combines these samples according to a deterministic model. However, describing every input variable as a distribution requires knowledge of the distribution, which may or may not be available, and the time and expense of determining the distribution parameters may be prohibitive. Therefore, the most influential parameters from the input data, such as mean values, standard deviations, shape parameters, and correlation coefficients, can be determined using an analytical sensitivity calculation based on the score function. In this paper, some probability concepts are reviewed, the score function methodology is explained and a probabilistic sensitivity analysis is completed on a crush energy determination example specific to traffic crash reconstruction. The computed sensitivity values were obtained using analytical derivatives which only required a single Monte Carlo simulation. The analytical derivative values are compared to a finite difference value to demonstrate numerical accuracy. These sensitivity results provide insight into the extent each input parameter will influence the output distribution, thus allowing the analyst to focus efforts on characterizing the most influential input distribution parameters. Moreover, the sensitivity values help determine the effect of correlation in a probabilistic analysis.

INTRODUCTION

Vehicle crashes are often linked to serious legal consequences and corresponding financial repercussions, making the details surrounding the accidents significant and valuable. Physical evidence and digital records from event data recorders can be gathered and interpreted to determine

the events leading up to a crash; however, the evidence may contain some measurement uncertainty. Therefore, it is necessary to assess the reliability of a traffic crash reconstruction analysis to account for uncertainty in the measured values.

A popular method for assessing reconstruction reliability is to use the so-called classic uncertainty analysis. This process uses the governing equations and linearizes the result about its mean using a partial Taylor series expansion. The truncated series is a linear function for which variances can be summed to determine the variance of the result. This method requires calculating partial derivatives of the result with respect to each input variable. Examples of uncertainty analysis pertaining to traffic crash reconstruction can be found in numerous references [6, 7, 11, 26]. To alleviate the need to compute sensitivity gradients by hand, a finite difference technique can be employed [5].

Another popular technique for determining the reliability of an analysis is to simulate the randomness associated with the evidence, physical measurements, and random phenomena by using the Monte Carlo method. The Monte Carlo method is a well-studied [12] random sampling technique used for the simulation of probabilistic functions and can be summarized in three general steps:

1. The probability distributions for the input variables must be determined. In crash reconstruction it is common to have an estimate rather than an exact value. Typically, variables follow a normal distribution, where the most likely value is the mean and the standard deviation accounts for the spread. Other distributions, such as the log-normal, extreme value, and uniform distribution are also applied.
2. The second step is to repeatedly sample the input distributions, perform a deterministic analysis based on those samples, and store the results. Essentially, a random sample is taken from each of the input distributions and evaluated using

an appropriate formula. Ang and Tang [1] give a formula for evaluating the percent error in the resulting output population mean, $\bar{x}$

$$\%error = 200 \sqrt{\frac{1 - \bar{x}}{N\bar{x}}} \quad (1)$$

which leads to a fairly significant sample size if high precision is needed.

3. Finally, the stored results are used to generate an empirical distribution and perform subsequent interpretation. Given the collection of results one can form an output distribution, from which a range of possible outcomes and their confidence limits can be established, giving significant insight into the nature of the randomness of the problem.

The Monte Carlo method has been successfully applied to crash reconstruction analysis [2,3,4, 13, 15, 16, 18,19,20, 27, 28]. However, the analyst must specify the distribution, as well as the distribution specific parameters such as the mean value, standard deviation, correlation coefficient, upper and lower bound, or other shape parameters, if they exist, for each of the input variables. At times, the importance of each of these input parameters may be unknown. For example, Daily [9] described the effect of correlation in a Monte Carlo analysis of a traffic crash but failed to quantify the relative importance of correlation on the result. Therefore, it is useful to use the so-called score function method for determining the sensitivity of Monte Carlo results with respect to all input parameters.

The goal of this paper is to quantify the potential changes in the output distribution of a Monte Carlo simulation due to the distribution parameters of the input variables. These sensitivities are useful in helping the analyst determine what variables are the most influential in the final result. It should be pointed out that distributions used in Monte Carlo simulations may not be available which may make it unsuitable as a pure reconstruction tool. However, the Monte Carlo method remains useful for analysis.

The paper is organized in the following manner. The next section contains a basic presentation of probabilistic theory and its application to the Monte Carlo technique. Then, the score function method for computing analytical sensitivity estimates from Monte Carlo estimates is presented. A numerical example of determining crush energy shows the application. Finally, a discussion and conclusions are presented.

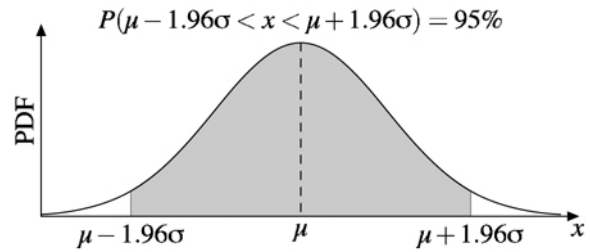


Figure 1. The univariate normal distribution corresponding to Eq. 2. The shaded area represents where 95% of the values are likely between $\mu \pm 1.96\sigma$.

PROBABILITY CONCEPTS

The most common way to articulate the inherent randomness of an event is to use a probability distribution. Basic discussions of probability concepts can be found in many textbooks, so only a few poignant concepts are presented in this section.

DISTRIBUTIONS

When a crash reconstruction analyst uses the Monte Carlo method, he or she must determine what type of distribution to use to describe the uncertainty associated with the various factors and parameters involved in the analysis. This process is largely dependent on the analyst's intuition and knowledge of the variables he or she is dealing with. In many instances, the variables are focused on an assumed or measured value with a level of confidence on either side. This suggests that the Gaussian (normal) distribution is the most widely applicable probability distribution. Furthermore, the sources of randomness for input values are usually numerous, which, when added together, result in a normal distribution by the central limit theorem. It is important to realize that other distributions, such as the uniform or log-normal distribution, can be used as well. However, the theory governing the analysis of measurement uncertainty assumes a normal distribution [10] and the discussion in this paper will be limited as such.

For a continuous random variable X , the Gaussian probability density function (PDF) is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right] \quad (2)$$

The distribution of this function is shown with a 95% precision interval in Fig. 1. If no other information is available to the analyst to suggest otherwise, it is recommended that a normal distribution be used.

In general, more than one random variable is needed in a probabilistic analysis, so the multivariate normal distribution is useful. In vector notation, the multivariate normal distribution is given as

$$\mathbf{f}_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \tilde{\mu})^T \Sigma^{-1} (\mathbf{x} - \tilde{\mu}) \right] \quad (3)$$

where n is the total number of variables, $\mathbf{X}$ is the vector of random variables, $|\Sigma|$ is the determinate of the covariance matrix, and $\tilde{\mu}$ is the vector of mean values. The covariance matrix is given as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,n}\sigma_1\sigma_n \\ \rho_{2,1}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2,n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n,1}\sigma_n\sigma_1 & \rho_{n,2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix} \quad (4)$$

where σ_i is the standard deviation of the i th normal random variable, σ_i^2 is the variance of the i th normal random variable, and $\rho_{i,j}$ is the correlation coefficient between the i th and j th variable (Note: when $i = j$, $\rho = 1$).

For the case of two random normal (Gaussian) variables, X and Y , Eq. 3 simplifies to the bivariate normal distribution:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[\frac{-1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_x^2} - \frac{2\rho_{xy}}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} \right) \right] \quad (5)$$

When two variables are statistically independent, then $\rho = 0$ and the bivariate distribution is simply the product of two univariate normal distributions.

MATHEMATICAL EXPECTATION

The expected value operator is defined over the whole sampling space as

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx. \quad (6)$$

If $g(x) = x$, then the mean can be calculated as

$$\mu_X = E[x] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx. \quad (7)$$

The r th central moment can be determined by

$$E[(x - \mu)^r] = \int_{-\infty}^{\infty} (x - \mu)^r \cdot f_X(x) dx \quad (8)$$

The variance is defined as the second central moment ($r = 2$) and the standard deviation is simply the square root of the variance.

$$\sigma_X = \sqrt{E[(x - \mu)^2]} \quad (9)$$

It can also be shown that

$$\sigma_X = \sqrt{E[x^2] - (E[x])^2} \quad (10)$$

where $E[x]^2 = \mu^2$. The mean and standard deviation can be estimated from samples of size N . An estimate for the population mean is

$$\mu_X = E[x] \approx \frac{1}{N} \sum_{i=1}^N x_i \quad (11)$$

when all occurrences of X are equally likely. An estimate for the population standard deviation is

$$\sigma_X \approx \sqrt{\frac{1}{N} \left(\sum_{i=1}^N x_i^2 \right) - \mu_x^2} \quad (12)$$

or

$$\sigma_X \approx \sqrt{\frac{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2}{N^2}} \quad (13)$$

Recognize that, based on the axiom of total probability,

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (14)$$

PROBABILITY OF EXCESS

In many cases, it is interesting to determine the probability of an analysis result exceeding some value. This is common in structural reliability assessment where the probability of failure is desired, as demonstrated by work from various



researchers (e.g. [23]). In the application of crash reconstruction the probability of an individual exceeding the speed limit may be of interest. Therefore, an indicator function is introduced such that

$$I(x) = \begin{cases} 1 & x \geq x_{limit} \\ 0 & x < x_{limit} \end{cases} \quad (15)$$

The probability of x exceeding the limit then becomes

$$P_{limit} = \int_{-\infty}^{\infty} I(x) f_X(x) dx = \int_{x_{lim}}^{\infty} f_X(x) dx \quad (16)$$

and the estimation of the probability of excess from a sample is simply

$$P_{limit} \approx \frac{1}{N} \sum_{i=1}^N I_i(x). \quad (17)$$

These limits can apply to many aspects of a reconstruction including minimum speed, maximum distance, perception and response time, etc. Determining the mean, standard deviation and probability of excess of the Monte Carlo result is fairly straight forward. However, it can be useful to know the relative sensitivity of those parameters with respect to each of the input variables. To address this, the score function method is discussed to determine the sensitivity of Monte Carlo results with respect to input parameters.

THE SCORE FUNCTION METHOD

Following the work of Rubinstein and others [14, 21, 22], Millwater [17], Sues & Caesar [23], and Wu & Mohanty [29] have developed, refined, and applied the score function method to engineering reliability problems. In essence, the score function method is an analytical approach to determining the sensitivity of Monte Carlo (or other sampling) based results with respect to input values. This section briefly summarizes the formulation of the score function method.

Consider an analysis result, z , of n variables:

$$z = g(\mathbf{x}) = g(x_1, x_2, \dots, x_n) \quad (18)$$

where x_i can be random. If x_i are random, then z is a random variable described by a probability distribution, $f_Z(z)$. The Monte Carlo method provides a means to estimating $f_Z(z)$ though a large number of simulated samples from the input distribution, $f_X(\mathbf{x})$. In general, the input distribution is a joint

distribution with potential correlation among the variables. To determine the sensitivity of the Monte Carlo result, parameters from the distribution of the result are needed. Commonly, the mean, standard deviation, and probability of excess are used to summarize the results.

The primary advantage of the score function method over a finite difference scheme is that the same results from the Monte Carlo analysis can be used to determine the sensitivities. This represents a potentially significant savings in computational cost. In other words, relatively little extra computational effort is needed to extract all the sensitivity gradients along with the original Monte Carlo result.

SENSITIVITY OF THE MEAN

The mean value of the response is computed as

$$\mu_z = E[g(\mathbf{x})] = \int \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} \quad (19)$$

and the sensitivity of the mean with respect to any input parameter θ_i is

$$\frac{\partial \mu_z}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \int \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x}. \quad (20)$$

Since z has been calculated, and the order of operations do not matter for linear operations (integration and differentiation), the sensitivity gradient of the mean can be written as

$$\frac{\partial \mu_z}{\partial \theta_i} = \int \dots \int_{-\infty}^{+\infty} z \frac{\partial f_X(\mathbf{x})}{\partial \theta_i} d\mathbf{x}. \quad (21)$$

Multiplying the integrand by $f_X(\mathbf{x})/f_X(\mathbf{x}) = 1$ and recognizing the form of the expectation operator yields

$$\frac{\partial \mu_z}{\partial \theta_i} = \int \dots \int_{-\infty}^{+\infty} z \left(\frac{\partial f_X(\mathbf{x})}{\partial \theta_i} \frac{1}{f_X(\mathbf{x})} \right) f_X(\mathbf{x}) d\mathbf{x} \quad (22)$$

$$\frac{\partial \mu_z}{\partial \theta_i} = \int \dots \int_{-\infty}^{+\infty} z (\kappa_{\theta_i}) f_X(\mathbf{x}) d\mathbf{x} \quad (23)$$

$$\frac{\partial \mu_z}{\partial \theta_i} = E[z \cdot \kappa_{\theta_i}] \quad (24)$$