SAE TECHNICAL PAPER SERIES

2007-01-0741

Considerations for Applying and Interpreting Monte Carlo Simulation Analyses in Accident Reconstruction

Jeffrey K. Ball, David A. Danaher and Richard M. Ziernicki
Knott Laboratory, LLC

Reprinted From: Accident Reconstruction 2007 (SP-2063)

ISBN 0-7680-1631-2

9 780768 016314

2007 World Congress
Detroit, Michigan
April 16-19, 2007

SAE International™

400 Commonwealth Drive, Warrendale, PA 15096-0001 U.S.A. Tel: (724) 776-4841 Fax: (724) 776-0790 Web: www.sae.org

ISSN 0148-7191

Printed in USA

2007-01-0741

Considerations for Applying and Interpreting Monte Carlo Simulation Analyses in Accident Reconstruction

Jeffrey K. Ball, David A. Danaher and Richard M. Ziernicki
Knott Laboratory, LLC

ABSTRACT

In reconstructing any accident, the reconstructionist must properly account for uncertainty in their analysis. One popular method of examining and quantifying the uncertainty within an analysis is the use of Monte Carlo simulation techniques. The methods have been well established and published over the last several years by numerous authors. One of the key factors underlying the Monte Carlo analysis is the assumed probability distribution of the individual factors within the analysis. The literature has examples and recommendations for assuming normal, uniform, or custom distributions for input parameters. However, the literature to date has not examined how the assumption of a distribution affects the resulting probability distribution of the Monte Carlo analysis. This paper attempts to address this issue. Furthermore, with the large number of samples typically considered during a Monte Carlo analysis, the resulting probability distribution tends to be normal and lends itself well to statistical interpretation as to the “most likely” range of the desired parameter. The analysis in this paper was performed with a plug-in to Excel called Crystal Ball®, and the version of Crystal Ball® used for this paper allows the user to selectively filter results which do not agree with physical reality (for example, non-equal force balances within a crush energy analysis). When filtering results, the final probability distribution can be skewed. This paper also examines methods of identifying the “most likely” range from within a skewed probability distribution.

INTRODUCTION

The accident reconstructionist is constantly faced with the problem having to reconstruct an accident on the basis of unknown or ill-defined parameters. The availability of published data and/or testing assists the reconstructionist, but there is always uncertainty when applying the data to the specific accident scenario. Therefore, the reconstructionist must have some means of quantifying the uncertainty or error range in their calculations. Courts also apply standards under Daubert or similar rulings that require the reconstructionist to identify the known or potential rate of error contained in their methodology. [1] Several techniques have been examined in the literature including statistical modeling, upper/lower bounds, differential variations, finite differences, uncertainty analyses, and Monte Carlo analysis. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] In a deterministic analysis, equations are evaluated by inputting appropriate parameters. The result of the analysis is a single numerical value that is obviously dependent on the selected input parameters. The error rate of a deterministic analysis can be estimated through the use of some of the previously mentioned techniques. [9] While these deterministic techniques are certainly valid, they become cumbersome if not impossible to implement when dealing with complex sets of equations. [9] As a result, the Monte Carlo method has become increasingly popular as a means of estimating error rate and/or uncertainty within more complex analyses. [6, 7, 8, 9, 10, 11] The Monte Carlo method evaluates a set of equations by selecting input parameters from user defined input distributions. The evaluation is repeated a statistically valid number of times, and the output of the analysis is a probability distribution instead of a single numerical value. The analyst can then use the probability distribution to estimate both the “most likely” range of values for the solution and the likely error rate. [7, 8, 9] The difficulty with the Monte Carlo method is that the user must not only identify the appropriate mean value for the inputs, but also the shape and variability of the input parameter distribution. Likewise, identifying the “most likely” range for the output calculation becomes more difficult the further the output distribution departs from a normal distribution. The last two of these issues will be addressed in this paper.

CALCULATIONS USED FOR COMPARISON

The objective of this work is to compare the effects of input parameter distribution on the resulting Monte Carlo probability functions and to examine how the analyst can define a “most likely” range when an output is highly skewed. Therefore, two analyses will be used to make the comparisons: a crush-energy analysis and a liner momentum analysis. The analyses follow procedures outlined in previous papers [9, 12] and the reader is directed to those papers for details of the equations used within the current work. Also, the data for the analyses in this work were obtained from actual accident vehicles. For the crush energy analysis, the vehicles involved were a commercial truck that was stopped pre-impact, and an SUV that slid into the commercial vehicle. While crush measurements and stiffnesses were available for the SUV, the stiffness of the commercial vehicle was obtained by using force balance methods described in the literature [13, 14]. The momentum based calculations are based on a low-speed, rear end collision between a small van and a pickup truck. It should be noted that the low speed calculations were selected solely due to the fact that the resulting speed distribution was skewed, and this paper is not attempting to justify any particular low speed models, nor to debate when tire forces can be ignored.

SELECTING INPUT DISTRIBUTIONS

Several researchers have proposed distributions for various parameters within accident reconstruction, including normal [4, 7, 8], uniform [7, 9], and “custom” or user defined distributions [9]. This work will compare the effects of the uniform distribution to those of the normal distribution, as these two distributions are the most commonly used distributions in the available literature. An example of each input distribution is shown in Figure 1.

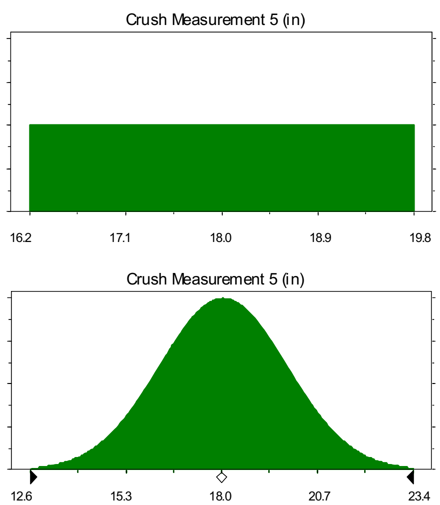

Figure 1. Uniform input distribution (top) and normal input distribution (bottom)

Several researchers have indicated that the uniform distribution is preferred [7, 9] because of the following reasons:

1. Any value within the range is equally likely to occur, and the distribution is therefore most conservative,
2. There is typically not enough data to justify using a normal distribution,
3. Errors in measurement can include systemic errors and random errors, and
4. The analyst must not only select a mean for the distribution but also a standard deviation.

However, the uniform distribution also requires the analyst to select a mean value and a range of variation. Several researchers have also noted that measurements of certain variables, such as coefficient of friction and others, tend to exhibit a normal distribution [4, 9] The difficulty in applying the normal distribution is that it extends to infinity in either direction. Therefore, the analyst runs the risk of using an input value that is not physically possible, for example, a negative coefficient of friction [8].

In the momentum analysis performed in this work, the vehicles involved in the collision were a mini van and a small pickup. The pickup was stopped prior to impact, and the mini van impacted the rear of the pickup. Parameters allowed to vary upon input included vehicle weights, post impact speed based on post impact travel measurements and estimated deceleration rates, and coefficient of restitution. For the crush analysis, the vehicles were a stopped commercial vehicle and an SUV that slid sideways into the commercial vehicle. Parameters allowed to vary upon input included vehicle weights, crush measurements, and stiffness coefficients. For the uniform distributions, the input values were allowed to vary ±10% from the mean [9], while for the uniform distributions, the standard deviation was 10% of the mean (a coefficient of variation of 10%). For each analysis, the calculations included the impact speed (since one vehicle was know to be stopped pre-impact) and delta-V’s for each vehicle. The output range for each calculation was the “most likely” range, which for the purposes of this work was defined as the 51% certainty range. Note that it is beyond the scope of this work to debate whether or not a 51% probability is “more likely than not.” Rather, 51% provides a convenient basis for comparison later in the paper. The results of the analyses are shown in Table 1 below.

Table 1 indicates that for a given analysis, using a uniform input distribution for the variables with ±10% variation produces a narrower range for the “most likely” speeds than do the normal distributions with a 10% coefficient of variation. The primary reason for this is that with the normal distribution, input variables are allowed to vary over a much larger range. Theoretically, the variables can range from negative infinity to positive infinity, although basic statistics shows that the probability of a variable being more than 3 standard deviations away from the mean is less than 0.3%. When dealing with measurements, the research indicates that a more reasonable coefficient of variation is on the order of 1-2% [4]. Therefore, using a coefficient of variation of 10% on the normal distribution usually overestimates the variability of any given input parameter. In short, these results indicate that a reconstructionist should not hesitate to use normal distributions for input variables to a Monte Carlo analysis when published data or logic dictate that the variable tends towards a normal distribution (as in the case of physical measurements or coefficients of friction). If the analyst is in doubt as to the variability of a distribution, then a uniform distribution, where any value in the range is as likely to occur as another, should be used. [7, 9]

Table 1. Results for momentum and crush energy analyses with uniform vs. normal input distributions for input variables.

Momentum	**Uniform Distribution**	**Normal Distribution**
Impact Speed (mph)	13.4 – 14.2	12.9 – 14.9
∆V, vehicle 1 (mph)	7.8 – 8.6	7.5 – 9.0
∆V, vehicle 2 (mph)	8.5 – 9.4	8.3 – 9.8
Crush Energy		
Impact Speed (mph)	34.6 – 36.3	34.2 – 36.9
∆V, vehicle 1 (mph)	28.3 – 30.2	27.9 – 30.8
∆V, vehicle 2 (mph)	2.7 – 3.4	2.8 – 3.2

DETERMINING THE "MOST LIKELY" RANGE FOR A NON-NORMAL DISTRIBUTION

As noted above, when the reconstructionist selects input distributions, care must be taken to minimize the effects of physically impossible values. Likewise, the reconstructionist must use care to avoid physically impossible results. Current versions of Monte Carlo analysis software often allow the user to "filter" out results that are not physically possible. In the current work this feature was used for the crush energy analysis to filter out any results in which the force balance between the vehicles differed from unity by more than 10%. As a result of this filtering, some of the output probability distributions were skewed. An example is the ∆V of the commercial vehicle in the crush energy analysis used in this work. The resulting probability distribution is shown in Figure 2.

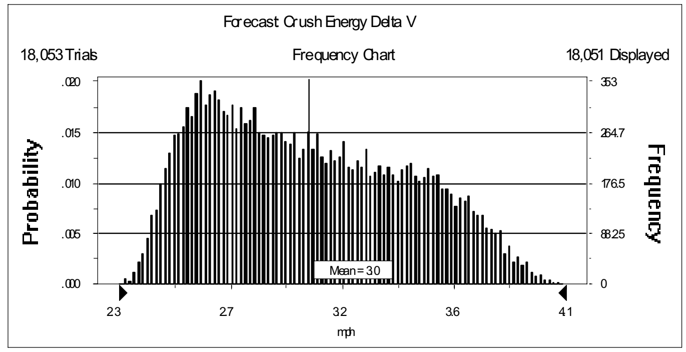

Figure 2. Probability distribution for ∆V for commercial vehicle from Monte Carlo analysis

Since any calculation which resulted in a force balance more than 10% greater or less than 1 caused the entire calculation to be rejected by the software, the resulting probability distribution is skewed right. (Right skewness is defined as the "tail" of the distribution being larger to the right.) The issue with such distributions is determining what is the "most likely" range of values. The software used for this work allows the user to input a desired certainty range, and it also allows the user to move "sliders" in order to define a certainty range. For the distribution shown in Figure 2, the 51% probability range is shown in Figure 3 for the default (i.e., software defined) range, and for arbitrary low and high user defined ranges.

As Figure 3 shows, when the user tells the software to calculate the 51% range, the software calculates 2.7 – 3.4 mph. However, by manipulating the sliders, the 51% range can also be found to be as low as 2.4 – 3.0 mph or as high as 3.0 – 4.1 mph. While the difference in ranges may seem small due to the magnitude of the numbers involved, the ranges vary by 50% – 60% from the mean depending on which method the analyst uses to calculate the 51% range. From a statistical standpoint, each range is "most likely" in its given situation. For example, if one were to calculate the ∆V deterministically based on the input parameters, it is "more likely than not" that the answer would fall in the range of 2.4 – 3.0 mph rather than in the ranges 2.3 – 2.4 mph or 3.1 – 4.1 mph. It is also true that it would be more likely to fall in the range of 3.0 – 4.1 mph than in the range of 2.3 – 3.0 mph. Note also that in each of the three cases shown in Figure 3, the mean value is included within the 51% range. Although selecting a range that did not include the mean value would definitely invalidate the range, the foregoing also shows that just because the mean in included in the range it is not definitively the correct "most likely" range.

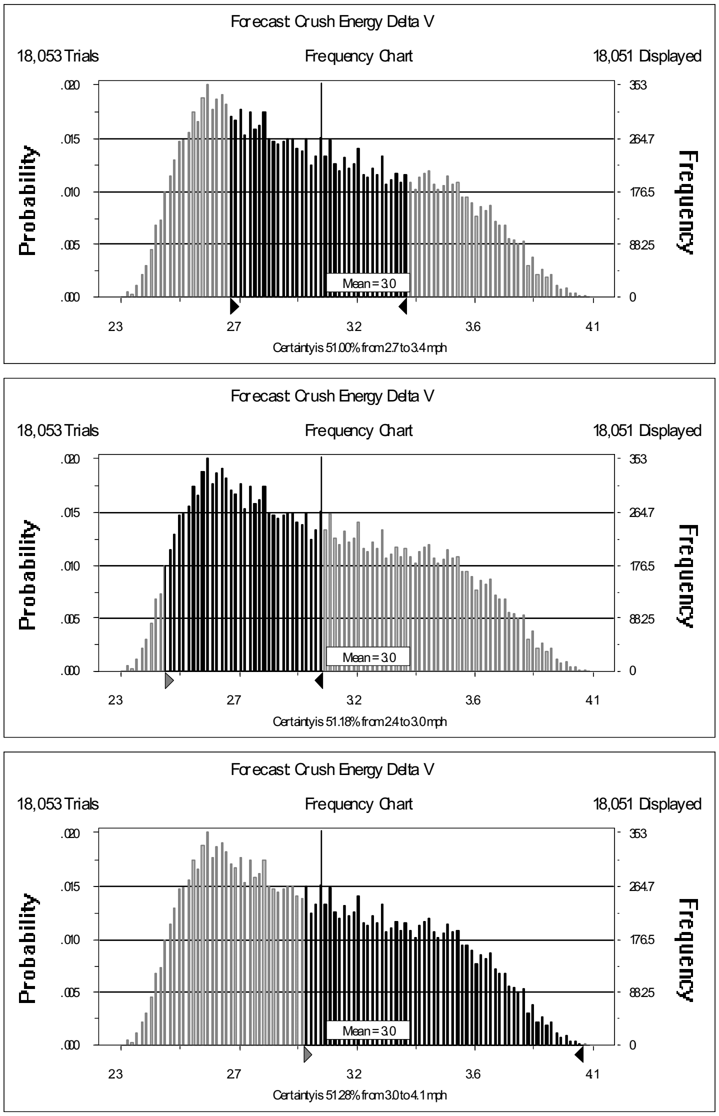

Figure 3. 51% range of distribution shown in Figure 2 for software defined range (top), user defined low range (middle) and user defined high range (bottom).

When the resulting distribution is close to normal, basic statistics would show that the "most likely" range would be centered about the mean, and would encompass the results that fall approximately between the 25th and 75th percentile. This is less obvious for the skewed distribution in Figure 2. Fortunately, techniques exist for calculating percentiles of non-normal distributions. This work relied on a Pearson distribution. Karl Pearson suggested in 1895 that sample estimates of skewness and kurtosis could be used to describe non-normal distributions in cases where the first four moments of the distribution could be calculated. [15] (Kurtosis is a measure of how "flat" or "peaked" a distribution is relative to a normal distribution.) For a sample of size n, the skewness and kurtosis are defined as [16]:

$$skewness = \sqrt{\beta_1} = \frac{m_3}{m_2^{3/2}}$$

$$kurtosis = \beta_2 = \frac{m_4}{m_2^2}$$

Where:

$$m_k = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^k / n$$

And $\overline{X}$ is the sample mean. For a normal distribution, the skewness and kurtosis are 0 and 3 respectively. For the skewed distribution shown in Figure 2, the skewness and kurtosis were calculated by the Monte Carlo software as 0.35 and 2.09 respectively. The software also calculated the mean and variance of the distribution shown in Figure 2. These four values were input into a FORTRAN subroutine [17] which used the Pearson distribution to calculate the 25th and 75th percentiles for the distribution. A comparison is shown in Table 2.

Table 2. Comparison of "most likely" range between Monte Carlo analysis and Pearson distribution.

Monte Carlo (software defined)	Pearson Distribution
2.7 – 3.4 mph	2.7 – 3.3 mph

Table 2 indicates that the Monte Carlo software does calculate the proper 51% range. Note that the slight difference in the ranges can be attributed to the fact that the Pearson subroutine can only calculate the middle 50% range, not the 51% range that the Monte Carlo software computes. Therefore, when a reconstructionist is faced with a skewed or other non-normal distribution, they should allow the Monte Carlo software to calculate the certainty ranges as opposed to using the manual slide bars.

CONCLUSION

The Monte Carlo simulation is a useful tool for defining error rates and "most likely" ranges for the accident reconstructionist. When deciding on the type of input distribution to use, both the uniform distribution and the normal distribution provide good results. The choice of which distribution to use will depend on the data available to the reconstructionist, and whether or not they are comfortable selecting a mean and standard deviation for the normal distribution. Should the results of the Monte Carlo analysis be highly non-normal, then the results depicted here indicate that the user should allow the Monte Carlo software to calculate the "most likely" range of outputs rather than defining the range manually.

ACKNOWLEDGMENTS

The authors would like to thank Ben Railsback and Trevor Buss of Knott Laboratory for their comments and suggestions during the preparation of this paper. Also, the comments provided by the SAE reviewers of this paper were extremely helpful.

REFERENCES

1. Brach, R.M., "Uncertainty in Accident Reconstruction Calculations," SAE Paper 940722.
2. Knott, A.W., "Probable .Speed Analysis: statistical Modeling of Auto Accidents vs. Deterministic Modeling," *Journal of the National Academy of Forensic Engineers*, Vol. X, No. 2, December 1993.
3. Tubergen, R.G., "The Technique of Uncertainty Analysis as Applied to the Momentum Equation for Accident Reconstruction," SAE Paper 950135.
4. Bartlett, W., Schmidt, B., Wright, W., Masory, O., Navin, F., Brach, R., Stanard, T., and Baxter, A., "Evaluating the Uncertainty in Various Measurement Tasks Common to Accident Reconstruction," SAE Paper 2002-01-0546.
5. Bartlett, W., Fonda, A., "Evaluating Uncertainty in Accident Reconstruction with Finite Differences," SAE Paper 2003-01-0489.
6. Kost, G., and Werner, S.M., "Use of Monte Carlo Simulation Techniques in Accident Reconstruction," SAE Paper 940719.
7. Wood, D.P., O'Riordain, S., "Monte Carlo Simulation Methods Applied to Accident Reconstruction and Avoidance Analysis," SAE Paper 940720.
8. Bartlett, W., "Conducting Monte Carlo Analysis with Spreadsheet Programs," SAE Paper 2003-01-0487.
9. Rose, N.A., Fenton, S.J., and Hughes, C.M., "Integrating Monte Carlo Simulation, Momentum-Based Impact Modeling, and Restitution Data to Analyze Crash Severity," SAE Paper 2001-01-3347.
10. Wach, W., and Unarski, J., "Determination of vehicle Velocities and Collision Location by Means of Monte Carlo Simulation Method," SAE Paper 2006-01-0907.
11. Moser, A., Steffan, H., Spek, A., and Makkinga, W., "Application of the Monte Carlo methods for Stability Analysis Within the Accident Reconstuction Software PC-CRASH," SAE Paper 2003-01-0488.
12. Rose, N.A., Fenton, S.J., and Beauchamp, G., "Restitution Modeling for Crush Analysis: Theory and Validation," SAE Paper 2006-01-0908.
13. Neptune, J.A., and Flynn, J.E., "A Method for Determining Accident Specific Crush Stiffness Coefficients," SAE Paper 940913.
14. Grimes, W.D., Heusser, R., Hunter, J., and Neptune, J.A., "Developing a Crush Profile estimate by Balancing Impact Forces," SAE Paper 970942.
15. 15. Pearson, K. " Contributions to the Mathematical theory of Evolution," Philosophical Transactions of the Royal Society of London, 91, pp. 343-358, 1895.
16. D'Agostino, R.B., Belanger, A., and D'Agostino, R.B. Jr., "A Suggestion for Using Powerful and Informative Tests of Normality," The American Statistician, Vol. 44, No. 4, November, 1990.
17. Davis, C.S., and Stephens, M.A., "Approximate Percentage Points using Pearson Curves," *Applied Statistics*, Royal statistical Society, London, 1983.

CONTACT

The authors welcome comments and questions and can be reached at the address, phone number, or email address below.

Knott Laboratory, LLC
7185 South Tucson Way
Englewood, Colorado 80112
(303) 925-1900
(303) 925-1901 Fax
jball@knottlab.com
.