Roser, Christoph, and Masaru Nakano. "Single Simulation Confidence Intervals Using the Delta Method." In International Symposium on Scheduling, edited by H. Fujimoto and M. Kuroda, 63–66. Hamamatsu, Japan, 2002.

SINGLE SIMULATION CONFIDENCE INTERVALS USING THE DELTA METHOD

Christoph Roser
Masaru Nakano
Toyota Central Research and Development Laboratories
Nagakute, Aichi, 480-1192, JAPAN
croser@robotics.tytlabs.co.jp
nakano@robotics.tytlabs.co.jp

ABSTRACT

This paper describes a method for the calculation of confidence intervals of simulation throughputs and utilizations. The method is based on the delta method and uses only a single simulation, where the variation of the underlying means is used to determine the variation of the performance function by using the first derivative of the performance function. While the delta method requires independent distributed data, such data is frequently available for many performance measures, allowing the practical application of the method. In addition, the method can also be used for short simulations or rare event applications, where methods based on batch means fail. This method can easily be implemented into existing simulation software.

INTRODUCTION

Scheduling is the process to arrange a number of tasks in a sequence. A frequent goal is to reduce the overall time for the performance of all tasks, or to ensure the completion of some tasks before a deadline. The time needed for the tasks have to be known to ensure the timely completion of the tasks. Unfortunately, these times are rarely static but vary depending on outside influences and random events. Often, simulation is used to estimate these times, and confidence intervals are used to determine the accuracy of these estimates. The underlying equations to calculate a confidence interval are well known (Devore 1995).

However, there are some complications to calculate confidence intervals in discrete event simulation. One complication is, that independent and identically distributed (i.i.d.) data is required, but simulation data is frequently neither independent (e.g. waiting times) nor identically distributed (e.g. warming up period) (Rinne 1997), (Kleijnen 1987). Additionally, many performance measures in discrete event simulation are a function of one or more means. For example, the throughput is the inverse of the mean time between completions of two parts. This further complicates the calculation of confidence intervals

This paper describes a method to determine the variance of the function of the means of one or more variable using the mean and variance of the variables and the gradient of the function at the mean value.

(Alexopoulos, and Seila 2000), (Law, and Kelton 1991), and (Banks 1998) give good overviews of the currently available methods for confidence interval calculation in simulations, with the most popular method being the batch means method. There are a number of different batch means and related methods developed. (Seila 1992), (Goldsman 1992), (Schmeiser, and Song 1996), (Pawlikowski 1990) give an overview of different batching methods, like overlapping batch means, non overlapping batch means, or fixed number of batches methods. There are a number of problems associated with the batch mean method. First, it is difficult to decide on the number of batches. Secondly, large data sets are needed to achieve valid confidence intervals. Third, different batching methods differ widely in their results. Finally, it is computationally intensive to calculate the confidence intervals, and therefore the confidence intervals are usually not calculated continuously during simulations.

This paper addresses the calculation of confidence intervals of functions of mean values. The presented method avoids most of the above problems for i.i.d. data. Example applications are given for throughputs and utilizations. The method is then validated experimentally using a complex simulation.

THE DELTA METHOD

The delta method calculates the deviation of a function of one or more means based on the mean and deviation of the function variables, using the gradients of the function at the mean values. The following section describes the delta method, leading to the general equation for this method (3).

Assume there is a general performance measure y as a function f of the mean values one or more variables $\bar{x}_i$, as for example the throughput is an inverse of the time between parts, or the utilization is the working time divided by the time between parts. The mean values $\bar{x}_i$ are calculated based on a set of n_i data values $x_{i,j}$, where the mean $\bar{x}_i$ and the standard deviation σ_{xi} is calculated using the well-known equations as shown in (1).

$$\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{i,j} \qquad \sigma_{x_i} = \sqrt{\frac{\sum_{j=1}^{n_i} \left(x_{i,j} - \bar{x}_i\right)^2}{n_i - 1}} \tag{1}$$

Yet, if the mean values $\bar{x}_i$ are applied to the function *f*, only one performance measure *y* is generated. The variation of the performance measure *y* and subsequently the confidence interval is yet unknown. While it is possible to enter the individual values $x_{i,j}$ into the equation *f*, the resulting mean and variation of the performance measure *y* would be incorrect as shown in equation (2) for all non-linear functions, i.e. the function of the mean would differ from the expected value of the function of the individual data values. Only for linear functions *f* will the function of the means and the mean of the function be equal (Papoulis 1991).

$$f(\bar{x}_1, \bar{x}_2 \ldots) \neq E\left[f(x_1, x_2 \ldots)\right] \tag{2}$$

The delta method replaces the function *f* by its tangent f^* at the mean values $\bar{x}_i$. Using this tangent f^*, it is possible to determine the standard deviation σ_f of the function *f* of the means $\bar{x}_i$ based on the deviation of the variables σ_{xi}. Figure 1 visualizes the throughput example for a tangential line f^* replacing the function *f*.

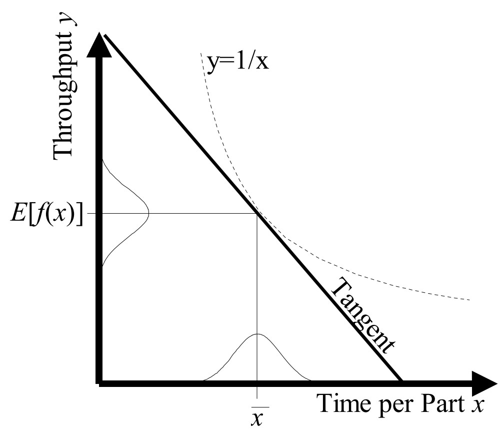

Figure 1: TANGENT AT THE MEAN VALUE

Using the standard deviation and the covariance of the variables $\bar{x}_i$, the standard deviation of the function value *y* can be determined using the delta method as shown in equation (3) (Rao 2001). Equation (3) includes the effect of the correlation between two paired variables, where $cov[x_1, x_2]$ is the unbiased estimate of the covariance as shown in equation (4) (Papoulis 1991).

$$\sigma_y^2 = \left(\left[\frac{df}{dx_1}\right]_{\substack{x_1=\bar{x}_1 \\ x_2=\bar{x}_2}} \cdot \sigma_{x_1}\right)^2 + \left(\left[\frac{df}{dx_2}\right]_{\substack{x_1=\bar{x}_1 \\ x_2=\bar{x}_2}} \cdot \sigma_{x_2}\right)^2 + 2 \cdot \left[\frac{df}{dx_1}\right]_{\substack{x_1=\bar{x}_1 \\ x_2=\bar{x}_2}} \cdot \left[\frac{df}{dx_2}\right]_{\substack{x_1=\bar{x}_1 \\ x_2=\bar{x}_2}} \cdot Cov[x_1, x_2] \tag{3}$$

$$Cov[x_1, x_2] = \frac{\sum_{j=1}^{n_1} x_{1,j} \cdot x_{2,j} - \frac{1}{n_1} \sum_{j=1}^{n_1} x_{1,j} \cdot \sum_{j=1}^{n_1} x_{2,j}}{n_1 - 1} \tag{4}$$

The resulting standard deviation σ_y of the function value *y* can then be used to calculate desired measures of accuracy, as for example a confidence interval as shown in equation (5), where *t* is the student-t distribution and α is the confidence level (Student 1908).

$$CI_y = t_{n_1 - 1, \alpha/2} \cdot \frac{\sigma_y}{\sqrt{n_1}} \tag{5}$$

This approach is valid if the underlying data in x is i.i.d. The independence can be tested using the von Neumann ratio η of the mean squared successive difference to the variation (RMSSDV) (Neumann 1941; Neumann 1942). Equation (6) shows the calculation of the RMSSDV η based on a set of data x of size n, where the mean squared difference between successive data is divided by the variance of the data. Variants of equation (6) can be found in (Kleijnen 1987) or (Steiger, and Wilson 1999).

$$\eta = \frac{n}{(n-1)} \cdot \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{\sum_{i=1}^{n_1} (x_i - \bar{x})^2} \tag{6}$$

If the data x is independent the RMSSDV η has a value of two. Thus this method can be used to determine if the collected data is approximately independent (i.e. with a mean value at or near two) or not (i.e. the mean differs from two).

COMMON PERFORMANCE MEASURES

Frequencies and Throughputs

Frequencies are a measurement of the number of occurrences in a given time. Throughputs are a type of frequencies, measuring the number of parts produced in a given period of time. Other frequencies are for example failure rates, i.e. the number of failures in a given period of time. These frequencies are defined by the number of occurrences of an event in a given period of time. This can also be described as the inverse of the mean time between the occurrences of an event. Subsequently, the frequency can be defined as the inverse of the average time between occurrences $\bar{x}_1$ as shown in equation (7). The advantage of this approach is that the deviation of the time between occurrences can be calculated, and therefore also the deviation of the frequency.

$$y = f(\bar{x}_1) = \frac{1}{\bar{x}_1} \tag{7}$$

Applying equation (3) to the function of the frequency in equation (7), the standard deviation of the frequency can be determined as shown in equation (8). A subsequent confidence interval can be calculated as shown in equation (5).

$$\sigma_y = \sqrt{\left(\frac{-1}{\bar{x}_1^2} \cdot \sigma_{x_1} \right)^2} \tag{8}$$

Percentages

Another common performance measure in discrete event simulation are percentages of times, as for example the percentage of time a machine is working, or the percentage of time a machine is under repair. In general, a percentage can be calculated by dividing the total time a machine is in a certain state by the total simulation time. This can also be represented as the mean duration a machine is in a certain state $\bar{x}_2$ divided by the mean duration between the beginnings of a certain state $\bar{x}_1$. For example, the percentage repair is the mean time to repair divided by the mean time between the beginnings of repairs. The function of the two mean values is shown in equation (9).

$$y = f(\bar{x}_1, \bar{x}_2) = \frac{\bar{x}_2}{\bar{x}_1} \tag{9}$$

Applying equation (3) to the function of the percentage in equation (9), the standard deviation of the percentage can be determined as shown in equation (10). A subsequent confidence interval can be calculated as shown in equation (5).

$$\sigma_y^2 = \left(\frac{\bar{x}_2}{\bar{x}_1^2} \cdot \sigma_{x_1} \right)^2 + \left(\frac{1}{\bar{x}_1} \cdot \sigma_{x_2} \right)^2 + 2 \cdot \frac{\bar{x}_2}{\bar{x}_1^2} \cdot \frac{1}{\bar{x}_1} \cdot Cov[x_1, x_2] \tag{10}$$

Roser, Christoph, and Masaru Nakano. "Single Simulation Confidence Intervals Using the Delta Method." In International Symposium on Scheduling, edited by H. Fujimoto and M. Kuroda, 63–66. Hamamatsu, Japan, 2002.

COMPLEX MANUFACTURING SYSTEM

The presented method was verified using a complex simulation example, consisting of seven machines in a complex setting and a mixture of two different products. The simulation was performed using the GAROPS simulation software as shown in Figure 2 (Kubota, Sato, and Nakano 1999), (Nakano et al. 1994).

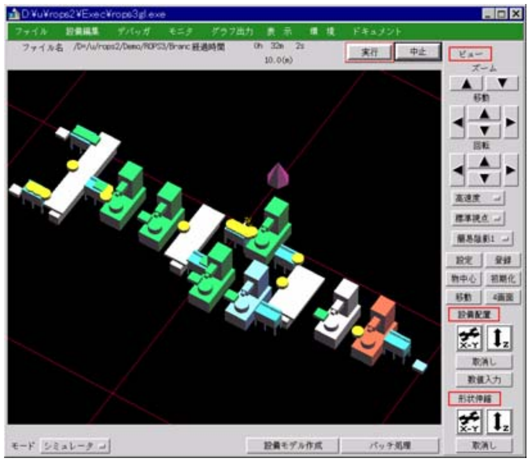

Figure 2: GAROPS Simulation Example

The total simulation time was almost two years of simulation. After removing the warming up period, this data was then split into 101 subsets with a simulation time of 6 days each. In order to calculate valid confidence intervals, the data has to be independent. Therefore, the RMSSDV has been calculated for the data using equation (6) to determine if the data is independent. While simulations are notorious for dependent data, the actual machine performance data was surprisingly often independent or near independent.

Despite the complex interactions of the system, most machine performance measures were independent. In fact, out of 46 measured parameters as for example the working times or the time between failures, all but four were approximately independent with a RMSSDV η between 1.7 and 2.2. This allows the calculation of a valid standard deviation and a confidence interval for these values as described above.

For each of the 101 subsets, the frequencies and the percentages of all machines working, idle, blocked or repaired were measured and the 95% confidence intervals calculated. These confidence intervals were then compared to the overall average, which are very close to the unknown true value. Ideally, for confidence intervals with a confidence level of 95%, 95% of the confidence intervals contain the true value, i.e. the desired coverage is 95%. However, in the real case, the percentage of the confidence intervals containing the true value may differ from the ideal case, i.e. the actual coverage differs from the desired coverage. The closer the actual coverage is to the desired coverage, the more accurate is the confidence interval method. Table 1 shows an overview of the coverage results of the complex simulation.

Table 1: SIMULATION EXAMPLE COVERAGE

Performance Measure	Desired Coverage	Actual Coverage	Too Small	Too Large
Frequency	95%	94.4%	2.9%	2.7%
Percentage	95%	92.9%	4.3%	2.8%

Out of the 6219 frequency confidence intervals with a desired coverage of 95%, the actual coverage was 94.44%. The instances where the long-term average was outside of the confidence interval were also symmetrically distributed with 2.8% under prediction and 2.7% over prediction. This indicates a very good overall fit.

Out of the 6219 percentage confidence intervals with a desired coverage of 95%, the actual coverage was 92.86%. The instances where the long-term average was outside of the confidence interval contained 4.3% under prediction and 2.8% over prediction. While the fit is not as good as for the frequencies, the coverage is still very close to the desired coverage. Overall, the actual coverage is almost identical with the desired coverage. Furthermore, the actual coverage is also nicely centered, with the number of over and under predictions being almost equal.

The presented method has been compared to the batching method, where the confidence interval is based on the batch means. The confidence intervals of the frequencies and percentages have been obtained from 100 simulations, using a fixed number of 30 batches with independent batch means. A total of 2180 confidence intervals for both the frequencies and percentages have been evaluated, of which only 498 and 1503 confidence intervals contained the true mean value. Therefore, the batch means method had coverage of only 22.8% and 68.8% for the frequencies and throughputs respectively, missing the desired coverage of 95 by a wide margin and is clearly inferior to the delta method for independent data.

CONCLUSION

In conclusion, the method provides very accurate results for near independent and identically distributed data. While simulation data is known to be dependent, the machine performance data was actually found to be frequently independent, allowing the calculation of the confidence intervals using the delta method.

Compared to batching, it is very fast to calculate the confidence interval, as it is not necessary to calculate different batch sizes and perform complex statistical tests. Moreover, if additional data becomes available, this data can easily be integrated into the previous calculation, and the confidence interval can be updated. This allows a sequential adding of data while updating the confidence interval.

Furthermore, the method works also with small sets of data. This is extremely useful for example to analyze rare events, where even a long simulation does not have many occurrences of the rare event, and subsequently batch means methods cannot be applied.

In summary, the method provides a preferable alternative to calculate the confidence intervals for approximately independent data.

REFERENCES

Alexopoulos, Christos, and Seila, Andrew F. 2000. Output Analysis for Simulations. In *Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 101-108, Orlando, Florida, USA.

Banks, Jerry 1998. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice.* John Wiley & Sons.

Devore, Jay L. 1995. *Probability and Statistics for Engineering and the Sciences.* Belmont: Duxbury Press, Wadsworth Publishing.

Goldsman, David 1992. Simulation output analysis. In *Winter Simulation Conference*, 97-103, Arlington, VA USA.

Kleijnen, Jack P. C. 1987. *Statistical Tools for Simulation Practitioners.* New York and Basel: Marcel Dekker.

Kubota, Fumiko, Sato, Shuichi, and Nakano, Masaru 1999. Enterprise Modeling and Simulation Platform Integrated Manufacturing System Design and Supply Chain. In *IEEE Conference on Systems, Man, and Cybernetics*, 511-515, Tokyo, Japan.

Law, Averill M., and Kelton, David W. 1991. *Simulation Modeling & Analysis.* McGraw Hill.

Nakano, Masaru, Sugiura, Norio, Tanaka, Minoru, and Kuno, Toshitaka 1994. ROPSII: Agent Oriented Manufacturing Simulator on the basis of Robot Simulator. In *Japan-USA Symposium on Flexible Automation*, 201-208, Kobe, Japan.

Neumann, John von 1941. Distribution of the Ratio of the Mean Square Successive Difference to the Variance. *Annals of Mathematical Statistic*, 12, 367-395.

Neumann, John von 1942. A Further Remark Concerning the Distribution of the Ratio of the Mean Square Successive Difference to the Variance. *Annals of Mathematical Statistic*, 13, 86-88.

Papoulis, Athanasios 1991. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill.

Pawlikowski, Krzysztof 1990. Steady-state simulation of queueing processes: survey of problems and solutions. *ACM Computing Surveys*, 22(2), 123-170.

Rao, C. Radhakrishna 2001. *Linear statistical inference and its applications.* Wiley Press.

Rinne, Horst 1997. *Taschenbuch der Statistik.* Frankfurt am Main: Verlag Harri Deutsch.

Schmeiser, Bruce W., and Song, Wheyming Tina 1996. Batching methods in simulation output analysis: what we know and what we don't. In *Winter Simulation Conference*, J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 122-127, San Diego, CA USA.

Seila, Andrew F. 1992. Advanced output analysis for simulation. In *Winter Simulation Conference*, 190-197, Arlington, VA USA.

Steiger, Natalie Miller, and Wilson, James R. 1999. Improved Batching for Confidence Interval Construction in Steady State Simulation. In *Winter Simulation Conference*, P. A. Farrington, D. T. Nembhard, D. T. Sturrock, and G. W. Evans, 442-451, Phoenix, AZ, USA.

Student 1908. The probable error of a mean. *Biometrika*, 6, 1-25.