

# How Many Monte-Carlo Simulations Are Needed to Adequately Process Interval Uncertainty: An Explanation of the Smart Electric Grid-Related Simulation Results

Afshin Gholamy<sup>1</sup> and Vladik Kreinovich<sup>2</sup>

Departments of <sup>1</sup>Geological Sciences and <sup>2</sup>Computer Science  
University of Texas at El Paso  
500 W. University  
El Paso, TX 79968, USA

Copyright © 2018 Afshin Gholamy and Vladik Kreinovich. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

One of the possible ways of dealing with interval uncertainty is to use Monte-Carlo simulations. A recent study of using this technique for the analysis of different smart electric grid-related algorithms shows that we need approximately 500 simulations to compute the corresponding interval range with 5% accuracy. In this paper, we provide a theoretical explanation for these empirical results.

**Mathematics Subject Classification:** 65G40 65C05 93C95 93D99

**Keywords:** interval uncertainty, Monte-Carlo simulations, electric grid

## 1 Formulation of the Problem

**Need for interval uncertainty.** Data processing means processing measurement results. Measurements are never absolutely accurate: the result  $\tilde{x}$

of measuring a physical quantity is, in general, somewhat different from the actual (unknown) value  $x$  of the corresponding quantity.

In the ideal case, we should know which values of the measurement error  $\Delta x \stackrel{\text{def}}{=} \tilde{x} - x$  are possible, and what is the probability of different possible values. These probabilities can be determined if we have a sufficiently large number of situations in which:

- we know the exact values (to be more precise, we have very good estimates of the exact values) and
- we also have measurement results.

In practice, however, we often do not have enough data to determine the corresponding probabilities. In such situations, often, the only information that we have about the measurement error is the upper bound  $\Delta$  on its absolute value:

$$|\Delta x| \leq \Delta;$$

see, e.g., [6]. Then, once we have the measurement result  $\tilde{x}$ , the only information that we have about the (unknown) actual value  $x$  is that this value belongs to the interval  $[\underline{x}, \bar{x}] = [\tilde{x} - \Delta, \tilde{x} + \Delta]$ . The resulting uncertainty is therefore known as *interval uncertainty*; see, e.g., [1, 4, 5].

**Need to propagate interval uncertainty.** A data processing algorithm  $f$ :

- starts with the results  $\tilde{x}_1, \dots, \tilde{x}_n$  of data processing, and
- uses these results to compute an output  $\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_n)$ .

This output:

- can be an estimate of some difficult-to-measure quantity, or
- it can be an estimate of the future value of some quantity  $y$ .

The corresponding algorithm is usually based on the known relation  $y = f(x_1, \dots, x_n)$  between the actual values of the corresponding quantities. Since, in general, the measurement results  $\tilde{x}_i$  are somewhat different from the actual values  $x_i$ , the result  $\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_n)$  of applying the algorithm  $f$  to the measurement results is, in general, somewhat different from the actual value  $y = f(x_1, \dots, x_n)$ :

$$\Delta y \stackrel{\text{def}}{=} \tilde{y} - y \neq 0.$$

It is therefore desirable not only to produce the estimate  $\tilde{y}$ , but also to find out what the possible values of the corresponding quantity  $y$  are.

We know that each quantity  $x_i$  can take any values within the corresponding interval  $[\underline{x}_i, \bar{x}_i]$ . Thus, the desired range of possible values of  $y$  have the form

$$[\underline{y}, \bar{y}] = \{y = f(x_1, \dots, x_n) : x_1 \in [\underline{x}_1, \bar{x}_1], \dots, x_n \in [\underline{x}_n, \bar{x}_n]\}.$$

The problem of computing this range is known as the main problem of *interval computations*; see, e.g., [1, 4, 5].

**Need for approximate methods.** It is known that, in general, the problem of computing the range  $[\underline{y}, \bar{y}]$  exactly is NP-hard; see, e.g., [2]. This means that, unless P=NP (which most computer scientists believe to be impossible), no feasible algorithm can always compute this range exactly. Thus, we need to use approximate methods for computing the desired range.

**A natural option: Monte-Carlo technique.** One of the natural ways to compute the range is to use *Monte-Carlo* techniques. In this technique, several ( $N$ ) times:

- we generate random numbers  $x_i^{(k)}$  uniformly distributed on the corresponding intervals  $[\underline{x}_i, \bar{x}_i]$ , and then
- we compute  $y^{(k)} = f(x_1^{(k)}, \dots, x_n^{(k)})$ .

When  $N \rightarrow \infty$ , the generated random values  $x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$ ,  $1 \leq k \leq N$ , cover all parts of the box  $[\underline{x}_1, \bar{x}_1] \times \dots \times [\underline{x}_n, \bar{x}_n]$ . Thus, in the same limit, the corresponding values  $y^{(k)} = f(x_1^{(k)}, \dots, x_n^{(k)})$  fill the entire interval  $[\underline{y}, \bar{y}]$ . So, to estimate the desired range, we can use the range formed by the values  $y^{(k)}$  corresponding to a sufficiently large number  $N$ , namely, the range

$$\left[ \min(y^{(1)}, \dots, y^{(N)}), \max(y^{(1)}, \dots, y^{(N)}) \right].$$

**How many simulations do we need?** Which value  $N$  should we choose? Usually,  $N$  is chosen as follows: we repeat the simulations for larger and larger  $N$ , and we stop when a further increase in  $N$  does not change the resulting range.

**Smart electric grid simulations: empirical results.** One of the important application areas is the application to electric grids. Electric grids are known to be unstable: a minor change in supply or demand can potentially cause a serious disruption and a blackout. To avoid such situations, engineers employ complicated (“smart”) control algorithms.

New improvements for such algorithms are being proposed all the time. To make sure that the new algorithm works well, we need to make sure that the resulting characteristics of the electric grids remain within their stable bounds.

Since the parameters of the electric grid are only measured with uncertainty, it is important to make sure that we have stability for all possible combinations of these parameters. One way to do it is to perform Monte-Carlo simulations and to check that the system remains stable for all  $N$  resulting combinations  $(x_1^{(k)}, \dots, x_n^{(k)})$ ,  $1 \leq k \leq N$ . How many combinations should we choose?

An empirical study [3] showed that if we are interested in 5% accuracy – a typical requirement for data analysis – then we need approximately  $N = 500$  simulations to get good results:

- if we have smaller  $N$ , e.g.,  $N = 100$  or  $N = 200$ , we underestimate the range of  $y$ 's;
- on the other hand, if we use a larger  $N$  – e.g.,  $N = 1000$  – we do not achieve any significant improvement in comparison to the case  $N \approx 500$ .

The authors of this study do not have any theoretical explanation for this empirical result.

**What we do in this paper.** In this paper, we provide the desired theoretical explanation.

## 2 Explanation

**Accuracy of Monte-Carlo simulations: reminder.** It is known (see, e.g., [7]) that if we estimate a quantity based on  $m$  measurements, then the relative accuracy of this estimate is

$$\varepsilon \approx \frac{1}{\sqrt{m}}.$$

**Let us apply this general feature to our case study.** Our goal is to reach the accuracy of  $\varepsilon \approx 5\% = 0.05$ .

In view of the above formula, to find the number of simulations needed to reach this accuracy, we must find the value  $m$  for which

$$\frac{1}{\sqrt{m}} \approx 0.05.$$

This approximate equality is equivalent to

$$\sqrt{m} \approx \frac{1}{0.05} = 20.$$

By squaring both sides of this approximate equality, we get

$$m \approx 20^2 = 400.$$

Taking into account that 500 was not the exact optimal value – it was just better than 100, 200, and 1000 – we conclude that  $m = 400$  is a perfect fit for the observed empirical data.

Thus, we provide the desired explanation for the smart electric grid-related simulation results.

**Acknowledgments.** This work was supported in part by the US National Science Foundation grant HRD-1242122.

The authors are thankful to Alfredo Vaccaro for valuable discussions.

## References

- [1] L. Jaulin, M. Kiefer, O. Didrit and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2001. <https://doi.org/10.1007/978-1-4471-0249-6>
- [2] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Springer Kluwer, Dordrecht, 1998. <https://doi.org/10.1007/978-1-4757-2793-7>
- [3] V. Loia, V. Terzija, A. Vaccaro, and P. Wall, An affine-arithmetic-based consensus protocol for smart-grid computing in the presence of data uncertainties, *IEEE Transactions on Industrial Electronics*, **62** (2015), no. 5, 2973–2982. <https://doi.org/10.1109/tie.2014.2363046>
- [4] G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017. <https://doi.org/10.1515/9783110499469>
- [5] R. E. Moore, R. B. Kearfott and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009. <https://doi.org/10.1137/1.9780898717716>
- [6] S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005. <https://doi.org/10.1007/0-387-29143-1>
- [7] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

**Received: January 27, 2018; Published: February 19, 2018**