3-2016

Limitations of Realistic Monte-Carlo Techniques

Andrzej Pownuk
The University of Texas at El Paso, ampownuk@utep.edu

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-16-17
To appear in *Proceedings of the 7th International Workshop on Reliable Engineering Computing REC'2016*, Bochum, Germany, June 15-17, 2016.

Limitations of Realistic Monte-Carlo Techniques

Andrzej Pownuk, Olga Kosheleva, and Vladik Kreinovich
Computational Science Program
University of Texas at El Paso, El Paso, TX 79968, USA
ampownuk@utep.edu, olgak@utep.edu, vladik@utep.edu

Abstract. Because of the measurement errors, the result $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ of processing the measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ is, in general, different from the value $y = f(x_1, \ldots, x_n)$ that we would obtain if we knew the exact values $x_1, \ldots, x_n$ of all the inputs. In the linearized case, we can use numerical differentiation to estimate the resulting difference $\Delta y = \widetilde{y} - y$; however, this requires $> n$ calls to an algorithm computing f, and for complex algorithms and large n this can take too long. In situations when for each input x_i, we know the probability distribution of the measurement error, we can use a faster Monte-Carlo simulation technique to estimate Δy. A similar Monte-Carlo technique is also possible for the case of interval uncertainty, but the resulting simulation is not realistic: while we know that each measurement error $\Delta x_i = \widetilde{x}_i - x_i$ is located within the corresponding interval, the algorithm requires that we use Cauchy distributions which can result in values outside this interval. In this paper, we prove that this non-realistic character of interval Monte-Carlo simulations is inevitable: namely, that no realistic Monte-Carlo simulation can provide a correct bound for Δy.

Keywords: Monte-Carlo techniques, interval uncertainty.

1. Formulation of the Problem

Need to take uncertainty into account when processing data. Data processing means applying some algorithm $f(x_1, \ldots, x_n)$ to the values of the quantities $x_1, \ldots, x_n$, resulting in a value $y = f(x_1, \ldots, x_n)$.

Values x_i usually come from measurements. Measurement are never absolutely accurate; the measurement result $\widetilde{x}_i$ is, in general, different from the actual (unknown) value x_i of the corresponding quantity: $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i \neq 0$; see, e.g., (Rabinovich, 2005).

Because of the this, the computed value $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is, in general, different from the ideal value $y = f(x_1, \ldots, x_n)$.

It is therefore desirable to estimate the accuracy $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$.

Possibility of linearization. In many practical situations, the measurement errors are relatively small. In such cases, we can safely ignore terms which are quadratic or higher order in Δx_i, and conclude that (Rabinovich, 2005)

$$\Delta y = \sum_{i=1}^{n} c_i \cdot \Delta x_i,$$

where

$$c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}.$$

What if we know all the probability distributions. When we know the probability distributions for all Δx_i (and we know that they are independent), then we can use Monte-Carlo techniques:

- several times $k = 1, \ldots, N$, we simulate $\Delta x_i^{(k)}$,
- then the differences

$$\Delta y^{(k)} \stackrel{\text{def}}{=} \widetilde{y} - f(\widetilde{x}_1 - \Delta x_1^{(k)}, \ldots, x_n - \Delta x_n^{(k)})$$

  have the same distribution as Δy.

Monte-Carlo techniques are more computationally efficient than numerical differentiation. Alternatively, we can use numerical differentiation to estimate all the derivatives c_i:

$$c_i \approx \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}}{h_i},$$

and then use the above formula or Δy.

However, this would require $n + 1$ calls to the algorithm f:

- one call to compute $\widetilde{y}$ and
- n calls to compute n partial derivatives $c_1, \ldots, c_n$.

For large n can be too long.

In contrast, the Monte-Carlo method needs $N + 1$ calls, where N is determined only by the accuracy with which we want Δ (and does not depend on n).

Case of interval uncertainty. In many practical situations, we only know the upper bound Δ_i on each measurement error Δx_i: $|\Delta x_i| \le \Delta_i$. In this case, the only information that we have about the unknown (actual) value x_i is that it is in the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ (Jaulin et al., 2001; Moore, Kearfott, and Cloud, 2009; Rabinovich, 2005).

In this case, the value Δy is bounded by

$$\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i,$$

namely, $|\Delta y| \le \Delta$.

How to estimate Δy under interval uncertainty: numerical differentiation and Monte-Carlo techniques. A straightforward computation of Δ requires $n + 1$ calls to f.

However, there is a faster method based on using Cauchy distribution, with probability density function

$$\rho_\Delta(x) = \frac{\Delta}{\pi} \cdot \frac{1}{1 + \dfrac{x^2}{\Delta^2}}.$$

This method is based on the known fact that when each Δx_i is distributed according to Cauchy distribution with parameter Δ_i, then the linear combination $\sum\limits_{i=1}^{n} c_i \cdot \Delta x_i$ is also Cauchy distributed, with parameter $\Delta = \sum\limits_{i=1}^{n} c_i \cdot \Delta x_i$.

Because of this fact, we can perform the following computations:

- first, we simulate $\Delta x_i^{(k)}$ based on Cauchy with parameter Δ_i,
- then the differences

  $$\Delta y^{(k)} \stackrel{\text{def}}{=} \widetilde{y} - f(\widetilde{x}_1 - \Delta x_1^{(k)}, \ldots, x_n - \Delta x_n^{(k)})$$

  are Cauchy distributed with the desired parameter Δ; so, we can find Δ by using, e.g., the Maximum Likelihood approach, which in this case leads to the equation

  $$\sum_{k=1}^{N} \frac{1}{1 + \dfrac{(\Delta y^{(k)})^2}{\Delta^2}} = \frac{N}{2}.$$

The interval-related Monte-Carlo method is not realistic. The Cauchy-based method works, but its simulation is *not realistic*, in the sense that:

- we know that $|\Delta x_i| \le \Delta_i$, but
- a Cauchy distribution goes beyond this bound.

What we do in this paper. In this paper, we show that no realistic Monte-Carlo simulation can lead to the interval estimate, so the use of non-realistic Monte-Carlo techniques is inevitable: simulated values $\Delta x_i^{(k)}$ *have to* go beyond $[-\Delta_i, \Delta_i]$.

Structure of the paper. First, in Section 2, we prove this result under the additional assumption that n distributions used to simulate $\Delta_1^{(k)}, \ldots, \Delta_n^{(k)}$ are independent.

Then, in Section 3, we extend this result to the general case, when we allow dependence between the simulated random variables $\Delta_1^{(k)}, \ldots, \Delta_n^{(k)}$.

2. Proof That Realistic Interval Monte-Carlo Techniques Are Not Possible: Case of Independent Variables

Case study. To prove the desired impossibility result, it is sufficient to prove that we cannot get the correct estimate for one specific function $f(x_1, \ldots, x_n)$.

As such a function, let us consider the simple function $f(x_1, \ldots, x_n) = x_1 + \ldots + x_n$. In this case, all the partial derivatives are equal to 1, i.e., $c_1 = \ldots = c_n = 1$ and thus,

$$\Delta y = \Delta x_1 + \ldots + \Delta x_n.$$

If we assume that each variables Δx_i takes value from the interval $[-\delta, \delta]$, then the range of possible values of the sum is $[-\Delta, \Delta]$, where $\Delta = n \cdot \delta$.

Analysis of the problem. Under Monte-Carlo simulations, we have

$$\Delta y^{(k)} = \Delta x_1^{(k)} + \ldots + \Delta x_n^{(k)}.$$

We assumed that the probability distributions corresponding to all i are independent.

Since the original problem is symmetric with respect to permutations, the corresponding distribution is also symmetric, so all $\Delta_i^{(k)}$ are identically distributed. Thus, the value Δy is the sum of several (n) independent identically distributed random variables.

It is known that due to the Central Limit Theorem (see, e.g., (Sheskin, 2011)), when n increases, the distribution of the sum tends to Gaussian. So, for large n, this distribution is close to Gaussian.

The Gaussian distribution is uniquely determined by its mean μ and variance $V = \sigma^2$. The mean of the sum is equal to the sum of the means, so $\mu = n \cdot \mu_0$, where μ_0 is the mean of the distribution used to simulate each Δx_i. For independent random variables, the variance of the sum is equal to the sum of the variances, so $V = n \cdot V_0$, where V_0 is the variance of the distribution used to simulate each Δx_i. Thus, $\sigma = \sqrt{V} = \sqrt{V_0} \cdot \sqrt{n}$.

It is well known that for a normal distribution, with very high confidence, all the values are contained in a k-sigma interval $[\mu - k \cdot \sigma, \mu + k \cdot \sigma]$:

- with probability $\approx 99.9\%$, the value will be in 3-sigma interval,
- with probability $\approx 1 - 10^{-8}$, the value will be in the 6-sigma interval, etc.

Thus, with high confidence, all the values obtained from simulation are contained in the interval $[\mu - k \cdot \sigma, \mu + k \cdot \sigma]$ of width $2k \cdot \sigma = 2k \cdot \sqrt{V_0} \cdot \sqrt{n}$.

For large n, this interval has the size $\text{const} \cdot \sqrt{n}$. On the other hand, we want the range $[-\Delta, \Delta]$ whose width is $2\Delta = 2\delta \cdot n$. So, when n is large, the simulated values occupy a part of the desired interval that tends to 0:

$$\frac{2k \cdot \sqrt{V_0} \cdot \sqrt{n}}{2\delta \cdot n} = \frac{\text{const}}{\sqrt{n}} \to 0.$$

So, in the independence case, the impossibility is proven.

3. Proof That Realistic Interval Monte-Carlo Techniques Are Not Possible: General Case

Case study. To prove the impossibility result in the general case, it is also sufficient to prove the impossibility for some of the functions. As case studies for this proof, we will consider functions

$$f(x_1, \ldots, x_n) = s_1 \cdot x_1 + \ldots + s_n \cdot x_n,$$

where $s_i \in \{-1, 1\}$.

For each of these functions,

$$\Delta y = s_1 \cdot \Delta x_1 + \ldots + s_n \cdot \Delta x_n,$$

so we have $c_i = s_i$. Similarly to the previous section, we assume that each of the unknowns Δx_i takes value from the interval $[-\delta, \delta]$, for some known value $\delta > 0$.

For each of these functions, $|c_i| = |s_i| = 1$, so the desired range is the same for all these functions and is equal to $[-\Delta, \Delta]$, where

$$\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i = n \cdot \delta.$$

Towards a precise formulation of the problem. Suppose that we want to find the range $[-\Delta, \Delta]$ with some relative accuracy ε. To get the range from simulations, we need to make sure that some of the simulated results are ε-close to Δ, i.e., that

$$\left| \sum_{i=1}^{n} s_i \cdot \Delta x_i^{(k)} - n \cdot \delta \right| \le \varepsilon \cdot n \cdot \delta,$$

or, equivalently,

$$n \cdot \delta \cdot (1 - \varepsilon) \le \sum_{i=1}^{n} s_i \cdot \Delta x_i^{(k)} \le n \cdot \delta \cdot (1 + \varepsilon).$$

We are interested in realistic Monte-Carlo simulations, for which $|\Delta_i^{(k)}| \le \delta$ for all i. Thus, we always have

$$\sum_{i=1}^{n} s_i \cdot \Delta x_i^{(k)} \le n \cdot \delta < n \cdot \delta \cdot (1 + \varepsilon).$$

So, the right-hand inequality is always satisfied, and it is thus sufficient to make sure that we have

$$\sum_{i=1}^{n} s_i \cdot \Delta x_i^{(k)} \ge n \cdot \delta \cdot (1 - \varepsilon)$$

for some simulation k.

For this inequality to be true with some certainty, we need to make sure that the probability of this inequality exceed some constant $p > 0$. Then, if we run $1/p$ simulations, then with high

probability, the inequality will be satisfied for at least one of these simulations. Thus, we arrive at the following condition.

Definition. *Let $\varepsilon > 0$, $\delta > 0$, and $p \in (0,1)$. We say that a probability distribution on the set of all vectors*

$$(\Delta_1 \ldots, \Delta x_n) \in [-\delta, \delta] \times \ldots \times [-\delta, \delta]$$

is a (p,ε)-realistic Monte-Carlo estimation of interval uncertainty *if for every set of values $s_i \in \{-1, 1\}$, we have*

$$\mathrm{Prob}(s_1 \cdot \Delta x_1 + \ldots + s_n \cdot \Delta x_n \geq n \cdot \delta \cdot (1 - \varepsilon)) \geq p.$$

Main Result. *Let $\delta > 0$ and $\varepsilon > 0$. If for every n, we have a (p_n, ε)-realistic Monte-Carlo estimation of interval uncertainty, then $p_n \leq \beta \cdot n \cdot c^n$ for some $\beta > 0$ and $c < 1$..*

Comments.

– As we have mentioned, when the probability is equal to p, we need $1/p$ simulations to get the desired estimates. Due to the Main Result, to get a realistic Monte-Carlo estimate for the interval uncertainty, we thus need

$$\frac{1}{p_n} \sim \frac{c^{-n}}{\beta \cdot n}$$

simulations. For large n, we have

$$\frac{c^{-n}}{\beta \cdot n} \gg n + 1.$$

Thus, the above results shows that realistic Monte-Carlo simulations require even more computational time than numerical differentiation. This defeats the main purpose for using Monte-Carlo techniques, which is – for our problem – to decrease the computation time.

– It is worth mentioning that if we allow p_n to be exponentially decreasing, then a realistic Monte-Carlo estimation of interval uncertainty is possible: e.g., we can take Δx_i to be independent and equal to δ or to $-\delta$ with equal probability 0.5. In this case, with probability 2^{-n}, we get the values $\Delta x_i = s_i \cdot \delta$ for which

$$\sum_{i=1}^{n} s_i \cdot \Delta x_i = \sum_{i=1}^{n} \delta = n \cdot \delta > n \cdot \delta \cdot (1 - \varepsilon).$$

Thus, for this probability distribution, for each combination of signs s_i, we have

$$\mathrm{Prob}(s_1 \cdot \Delta x_1 + \ldots + s_n \cdot \Delta x_n \geq n \cdot \delta \cdot (1 - \varepsilon)) = p_n = 2^{-n}.$$

Proof of the main result. Let us pick some $\alpha \in (0,1)$. Let us denote, by m, the number of indices i or which $s_i \cdot \Delta x_i > \alpha \cdot \delta$. Then, if we have

$$s_1 \cdot \Delta x_1 + \ldots + s_n \cdot \Delta x_n \geq n \cdot \delta \cdot (1 - \varepsilon),$$

then for $n - m$ indices, we have $s_i \cdot \Delta x_i \le \alpha \cdot \delta$ and for the other m indices, we have $s_i \cdot \Delta x_i \le \delta$. Thus,

$$n \cdot \delta \cdot (1 - \varepsilon) \le \sum_{i=1}^{n} s_i \cdot \Delta x_i \le m \cdot \delta + (n - m) \cdot \alpha \cdot \delta.$$

Dividing both sides of this inequality by δ, we get

$$n \cdot (1 - \varepsilon) \le m + (n - m) \cdot \alpha,$$

hence $n \cdot (1 - \alpha - \varepsilon) \le m \cdot (1 - \alpha)$ and thus,

$$m \ge n \cdot \frac{1 - \alpha - \varepsilon}{1 - \alpha}.$$

So, we have at least

$$n \cdot \frac{1 - \alpha - \varepsilon}{1 - \alpha}$$

indices for which Δx_i has the same sign as s_i (and for which $|\Delta x_i| > \alpha \cdot \delta$). This means that for the vector corresponding to a tuple $(s_1, \ldots, s_n)$, at most

$$n \cdot \frac{\varepsilon}{1 - \alpha - \varepsilon}$$

indices have a different sign than s_i.

It is, in principle, possible that the same tuple $(\Delta x_1, \ldots, \Delta x_n)$ can serve two different tuples $s = (s_1, \ldots, s_n)$ and $s' = (s'_1, \ldots, s'_n)$. However, in this case:

- going from s_i to $\mathrm{sign}(\Delta x_i)$ changes at most $n \cdot \dfrac{\varepsilon}{1 - \alpha - \varepsilon}$ signs, and
- going from $\mathrm{sign}(\Delta x_i)$ to s'_i also changes at most $n \cdot \dfrac{\varepsilon}{1 - \alpha - \varepsilon}$ signs.

Thus, between the tuples s and s', at most $2 \cdot \dfrac{\varepsilon}{1 - \alpha - \varepsilon}$ signs are different. In other words, for the Hamming distance

$$d(s, s') \stackrel{\text{def}}{=} \#\{i : s_i \ne s'_i\},$$

we have

$$d(s, s') \le 2 \cdot n \cdot \frac{\varepsilon}{1 - \alpha - \varepsilon}.$$

Thus, if

$$d(s, s') > 2 \cdot n \cdot \frac{\varepsilon}{1 - \alpha - \varepsilon},$$

then no tuples $(\Delta x_1, \ldots, \Delta x_n)$ can serve both sign tuples s and s'. In this case, the corresponding sets of tuples for which

$$s_1 \cdot \Delta x_1 + \ldots + s_n \cdot \Delta x_n \ge n \cdot \delta \cdot (1 - \varepsilon)$$

and

$$s'_1 \cdot \Delta x_1 + \ldots + s'_n \cdot \Delta x_n \geq n \cdot \delta \cdot (1 - \varepsilon)$$

do not intersect. Hence, the probability that the randomly selected tuple belongs to one of these sets is equal to the sum of the corresponding probabilities. Since each of the probabilities is greater than or equal to p, the resulting probability is equal to $2p$.

If we have M sign tuples $s^{(1)}, \ldots, s^{(M)}$ for which

$$d(s^{(i)}, s^{(j)}) > 2 \cdot \frac{\varepsilon}{1 - \alpha - \varepsilon}$$

for all $i \neq j$, then similarly, the probability that the tuple $(\Delta x_1, \ldots, \Delta x_n)$ serves one of these sign tuples is greater than or equal to $M \cdot p$. On the other hand, this probability is ≤ 1, so we conclude that $M \cdot p \leq 1$ and $p \leq \dfrac{1}{M}$.

So, to prove that p_n is exponentially decreasing, it is sufficient to find the sign tuples whose number M is exponentially increasing.

Let us denote $\beta \stackrel{\text{def}}{=} \dfrac{\varepsilon}{1 - \alpha - \varepsilon}$. Then, for each sign tuple s, the number t of all sign tuples s' for which $d(s, s') \leq \beta \cdot n$ is equal to the sum of:

- the number of tuples $\binom{n}{0}$ that differ from s in 0 places,
- the number of tuples $\binom{n}{1}$ that differ from s in 1 place, $\ldots$,
- the number of tuples $\binom{n}{\beta \cdot n}$ that differ from s in $\beta \cdot n$ places,

i.e.,

$$t = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n \cdot \beta}.$$

When $\beta < 0.5$ and $\beta \cdot n < \dfrac{n}{2}$, the number of combinations $\binom{n}{k}$ increases with k, so $t \leq \beta \cdot n \cdot \binom{n}{\beta \cdot n}$. Here,

$$\binom{a}{b} = \frac{a!}{b! \cdot (a - b)!}.$$

Asymptotically,

$$n! \sim \left(\frac{n}{e}\right)^n,$$

so

$$t \leq \beta \cdot n \cdot \frac{\left(\dfrac{n}{e}\right)^n}{\left(\dfrac{\beta \cdot n}{e}\right)^{\beta \cdot n} \cdot \left(\dfrac{(1 - \beta) \cdot n}{e}\right)^{(1 - \beta) \cdot n}}.$$

One can see that the term n^n in the numerator cancels with the term $n^{\beta\cdot n}\cdot n^{(1-\beta)\cdot n}=n^n$ in the denominator. Similarly, the terms e^n and $e^{\beta\cdot n}\cdot e^{(1-\beta)\cdot n}=e^n$ cancel each other, so we conclude that

$$t\le\beta\cdot n\cdot\left(\frac{1}{\beta^\beta\cdot(1-\beta)^{1-\beta}}\right)^n.$$

Here,

$$\gamma\stackrel{\text{def}}{=}\frac{1}{\beta^\beta\cdot(1-\beta)^{1-\beta}}=\exp(S),$$

where

$$S\stackrel{\text{def}}{=}-\beta\cdot\ln(\beta)-(1-\beta)\cdot\ln(1-\beta)$$

is Shannon's entropy. It is well known (and easy to check by differentiation) that its largest possible values is attained when $\beta=0.5$, in which case $S=\ln(2)$ and $\gamma=\exp(S)=2$. When $\beta<0.5$, we have $S<\ln(2)$, thus, $\gamma<2$, and $t\le\beta\cdot n\cdot\gamma^n$ for some $\gamma<2$.

Let us now construct the desired collection of sign tuples $s^{(1)}$, ..., $s^{(M)}$.

- We start with some sign tuple $s^{(1)}$, e.g., $s^{(1)}=(1,\ldots,1)$.
- Then, we dismiss $t\le\gamma^n$ tuples which are $\le\beta$-close to s, and select one of the remaining tuples as $s^{(2)}$.
- We then dismiss $t\le\gamma^n$ tuples which are $\le\beta$-close to $s^{(2)}$. Among the remaining tuples, we select the tuple $s^{(3)}$, etc.

Once we have selected M tuples, we have thus dismissed $t\cdot M\le\beta\cdot n\cdot\gamma^n\cdot M$ sign tuples. So, as long as this number is smaller than the overall number 2^n of sign tuples, we can continue selecting.

This procedure ends when we have selected M tuples for which $\beta\cdot n\cdot\gamma^n\cdot M\ge 2^n$. Thus, we have selected

$$M\ge\left(\frac{2}{\gamma}\right)^n\cdot\frac{1}{\beta\cdot n}$$

tuples. So, we have indeed selected exponentially many tuples.

Hence,

$$p_n\le\frac{1}{M}\le\beta\cdot n\cdot\left(\frac{\gamma}{2}\right)^n,$$

i.e.,

$$p_n\le\beta\cdot n\cdot c^n,$$

where

$$c\stackrel{\text{def}}{=}\frac{\gamma}{2}<1.$$

So, the probability p_n is indeed exponentially decreasing. The main result is proven.

Acknowledgements

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

The authors are thankful to Sergey Shary and to the anonymous referees.

References

Jaulin, L., M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis*, Springer, London, 2001.

Kreinovich, V., and S. Ferson, A new Cauchy-based black-box technique for uncertainty in risk analysis, *Reliability Engineering and Systems Safety*, 85(1–3):267–279, 2004.

Moore, R. E., R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM Press, Philadelphia, Pennsylvania, 2009.

Rabinovich, S. G., *Measurement Errors and Uncertainty. Theory and Practice*, Springer Verlag, Berlin, 2005.

Sheskin, D. J., *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.