Limitations of Realistic Monte-Carlo Techniques in Estimating Interval Uncertainty

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

1. Need for Data Processing

- We want to predict the future state of the world, i.e., the future values y of different quantities.
- For this, we need to know how y depends on the current values $x_1, \ldots, x_n$ of the related quantities:
$$y = f(x_1, \ldots, x_n).$$
- Then, we measure x_i and make a prediction
$$\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n).$$
- Weather prediction shows that the data processing algorithm f can be very complex.
- Data processing is also needed if we are interested in a difficult-to-measure quantity y.
- To estimate y, we measure easier-to-measure quantities $x_1, \ldots, x_n$ related to y by a known dependence
$$y = f(x_1, \ldots, x_n).$$

2. Need to Take Uncertainty Into Account When Processing Data

- Measurement are never absolutely accurate: in general,

$$\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i \neq 0.$$

- As a result, the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is, in general, different from the ideal value $y = f(x_1, \ldots, x_n)$.
- To estimate the accuracy $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$, we need to have some information about the measurement errors Δx_i.
- Traditional engineering approach assumes that we know the probability distribution of each Δx_i.
- Often, $\Delta x_i \sim N(0, \sigma_i)$, and different Δx_i are assumed to be independent.
- In such situations, our goal is to find the probability distribution for Δy.

3. Case of Interval Uncertainty

- Often, we only know the upper bound Δ_i: $|\Delta x_i| \le \Delta_i$.
- Then, the only information about the x_i is that
$$x_i \in \mathbf{x}_i \stackrel{\text{def}}{=} [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i].$$
- Different $x_i \in \mathbf{x}_i$ lead, in general, to different
$$y = f(x_1, \ldots, x_n).$$
- We want to find the range $\mathbf{y}$ of possible values of y:
$$\mathbf{y} = \{f(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$
- Often, measurement errors are relatively small.
- We can then only keep terms linear in Δx_i:
$$\Delta y = \sum_{i=1}^{n} c_i \cdot \Delta x_i, \text{ where } c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}.$$
- In this case, $\mathbf{y} = [\widetilde{y} - \Delta, \widetilde{y} + \Delta]$, where $\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i$.

4. How to Compute the Interval Range: Linearized Case

- Sometimes, we have explicit expressions or efficient algorithms for the partial derivatives c_i.
- Often, however, we proprietary software in our computations.
- Then, we cannot use differentiation formulas or automatic differentiation (AD) tools.
- We can use numerical differentiation:

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

- *Problem:* We need $n+1$ calls to f, to compute $\widetilde{y}$ and n values c_i.
- When f is time-consuming and n is large, this takes too long.

5. A Faster Method: Cauchy-Based Monte-Carlo

- *Idea:* use Cauchy distribution $\rho_\Delta(x) = \dfrac{\Delta}{\pi} \cdot \dfrac{1}{1 + x^2/\Delta^2}$.
- *Why:* when $\Delta x_i \sim \rho_{\Delta_i}(x)$ are indep., then $\Delta y = \sum\limits_{i=1}^n c_i \cdot \Delta x_i \sim \rho_\Delta(x)$, with $\Delta = \sum\limits_{i=1}^n |c_i| \cdot \Delta_i$.
- Thus, we simulate $\Delta x_i^{(k)} \sim \rho_{\Delta_i}(x)$; then, $\Delta y^{(k)} \stackrel{\text{def}}{=} \widetilde{y} - f(\widetilde{x}_1 - \Delta x_1^{(k)}, \ldots) \sim \rho_\Delta(x)$.
- Maximum Likelihood method can estimate Δ: $\prod\limits_{k=1}^N \rho_\Delta(\Delta y^{(k)}) \to \max$, so $\sum\limits_{k=1}^N \dfrac{1}{1 + (\Delta y^{(k)})^2/\Delta^2} = \dfrac{N}{2}$.
- To find Δ from this equation, we can use, e.g., the bisection method for $\underline{\Delta} = 0$ and $\overline{\Delta} = \max\limits_{1 \le k \le N} |\Delta y^{(k)}|$.

6. Monte-Carlo: Successes and Limitations

- *Fact:* for Monte-Carlo, accuracy is $\varepsilon \sim 1/\sqrt{N}$.
- *Good news:* the number N of calls to f depends only the desired accuracy ε.
- *Example:* to find Δ with accuracy 20% and certainty 95%, we need $N = 200$ iterations.
- *Limitation:* this method is *not realistic*; indeed:
  - we know that Δx_i is *inside* $[-\Delta_i, \Delta_i]$, but
  - Cauchy-distributed variable has a high probability to be *outside* this interval.
- *Natural question:* is it a limitation of our method, or of a problem itself?
- *Our answer:* for interval uncertainty, a realistic Monte-Carlo method is not possible.

7. Proof : Case of Independent Variables

- It is sufficient to prove that we cannot get the correct estimate for *one* specific function

$$f(x_1, \ldots, x_n) = x_1 + \ldots + x_n, \text{ when } \Delta y = \Delta x_1 + \ldots + \Delta x_n.$$

- When each variables Δx_i is in the interval $[-\delta, \delta]$, then the range of Δy is $[-\Delta, \Delta]$, where $\Delta = n \cdot \delta$.
- In Monte-Carlo, $\Delta y^{(k)} = \Delta x_1^{(k)} + \ldots + \Delta x_n^{(k)}$.
- $\Delta_i^{(k)}$ are i.i.d. Due to the Central Limit Theorem, when $n \to \infty$, the distribution of the sum tends to Gaussian.
- For a normal distribution, with very high confidence, $\Delta y \in [\mu - k \cdot \sigma, \mu + k \cdot \sigma]$.
- Here, $\sigma \sim \sqrt{n}$, so this interval has width $w \sim \sqrt{n}$.
- However, the actual range of Δy is $\sim n \gg w$. Q.E.D.

8. General Case

- Let's take $f(x_1, \ldots, x_n) = s_1 \cdot x_1 + \ldots + s_n \cdot x_n$, where $s_i \in \{-1, 1\}$.
- Then, $\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i = n \cdot \delta$.
- Let $\varepsilon > 0$, $\delta > 0$, and $p \in (0, 1)$. We consider probability distributions P on the set of all vectors

$$(\Delta x_1 \ldots, \Delta x_n) \in [-\delta, \delta] \times \ldots \times [-\delta, \delta].$$

- We say that P is a (p, ε)-*realistic Monte-Carlo estimation (MCE)* if for all $s_i \in \{-1, 1\}$, we have

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

- **Result.** *If for every n, we have a (p_n, ε)-realistic MCE, then $p_n \leq \beta \cdot n \cdot c^n$ for some $\beta > 0$ and $c < 1$.*
- For probability p_n, we need $1/p_n \sim c^{-n}$ simulations – more than $n + 1$ for numerical differentiation.

9. Why Cauchy Distribution: Formulation of the Problem

- We want to find a family of probability distributions with the following property:
  - when independent $X_1, \ldots, X_n$ have distributions from this family with parameters $\Delta_1, \ldots, \Delta_n$,
  - then each $Y = c_1 \cdot X_1 + \ldots + c_n \cdot X_n \sim \Delta \cdot X$, where X corr. to parameter 1, and $\Delta = \sum\limits_{i=1}^{n} |c_i| \cdot \Delta_i$.
- In particular, for $\Delta_1 = \ldots = \Delta_n = 1$, the desired property of this probability distribution is as follows:
  - if we have n independent identically distributed random variables $X_1, \ldots, X_n$,
  - then each $Y = c_1 \cdot X_1 + \ldots + c_n \cdot X_n$ has the same distribution as $\Delta \cdot X_i$, where $\Delta = \sum\limits_{i=1}^{n} |c_i|$.

10. Analysis of the Problem

- For $n = 1$ and $c_1 = -1$, the desired property says that $-X \sim X$, the distribution is even.
- A usual way to describe a probability distribution is to use a probability density function $\rho(x)$.
- Often, it is convenient to use its Fourier transform – the *characteristic function* $\chi_X(\omega) \stackrel{\rm def}{=} E[\exp(\mathrm{i} \cdot \omega \cdot X)]$.
- When X_i are independent, then for $S = X_1 + X_2$:

$$\chi_S(\omega) = E[\exp(\mathrm{i} \cdot \omega \cdot S)] = E[\exp(\mathrm{i} \cdot \omega \cdot (X_1 + X_2)] =$$
$$E[\exp(\mathrm{i} \cdot \omega \cdot X_1 + \mathrm{i} \cdot \omega \cdot X_2)] =$$
$$E[\exp(\mathrm{i} \cdot \omega \cdot X_1) \cdot \exp(\mathrm{i} \cdot \omega \cdot X_2)].$$

- Since X_1 and X_2 are independent,

$$\chi_S(\omega) = E[\exp(\mathrm{i}\cdot\omega\cdot X_1)]\cdot E[\exp(\mathrm{i}\cdot\omega\cdot X_2)] = \chi_{X_1}(\omega)\cdot\chi_{X_2}(\omega).$$

11. Analysis of the Problem (cont-d)

- Similarly, for $Y = \sum\limits_{i=1}^{n} c_i \cdot X_i$, we have

$$\chi_Y(\omega) = E[\exp(\mathrm{i}\cdot\omega\cdot Y)] = E\left[\exp\left(\mathrm{i}\cdot\omega\cdot\sum_{i=1}^{n} c_i \cdot X_i\right)\right] =$$

$$E\left[\prod_{i=1}^{n} \exp\left(\mathrm{i}\cdot\omega\cdot c_i \cdot X_i\right)\right] = \prod_{i=1}^{n} \chi_X(\omega\cdot c_i).$$

- The desired property is $Y \sim \Delta \cdot X$, so

$$\prod_{i=1}^{n} \chi_X(\omega\cdot c_i) = \chi_{\Delta\cdot X}(\omega) = E[\exp(\mathrm{i}\cdot\omega\cdot(\Delta\cdot X))]\chi_X(\omega\cdot\Delta),$$

so $\chi_X(c_1\cdot\omega)\cdot\ldots\cdot\chi_X(c_n\cdot\omega) = \chi_X((|c_1|+\ldots+|c_n|)\cdot\omega)$.

- In particular, for $n = 1$, $c_1 = -1$, we get $\chi_X(-\omega) = \chi_X(\omega)$, so $\chi_X(\omega)$ should be an even function.

12. Analysis of the Problem (cont-d)

- Reminder:

  $$\chi_X(c_1 \cdot \omega) \cdot \ldots \cdot \chi_X(c_n \cdot \omega) = \chi_X((|c_1| + \ldots + |c_n|) \cdot \omega).$$

- For $n = 2$, $c_1 > 0$, $c_2 > 0$, and $\omega = 1$, we get

  $$\chi_X(c_1 + c_2) = \chi_X(c_1) \cdot \chi_X(c_2).$$

- The characteristic function should be measurable.
- *Known:* the only measurable functions with this property are $\chi_X(\omega) = \exp(-k \cdot \omega)$ for some k.
- Due to evenness, for a general ω, we get $\chi_X(\omega) = \exp(-k \cdot |\omega|)$.
- By applying the inverse Fourier transform, we conclude that X is Cauchy distributed.
- *Conclusion:* so, only Cauchy distribution works.

13. Acknowledgments

- 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 from the Prudential Foundation.
- The authors are thankful to Sergey Shary and to the anonymous referees for their valuable suggestions.

14. 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 $s_i \cdot \Delta x_i > \alpha \cdot \delta$.
- 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 $s_i \cdot \Delta x_i \leq \alpha \cdot \delta$ and
  - for the other m indices, we have $s_i \cdot \Delta x_i \leq \delta$.
- Thus, $n \cdot \delta \cdot (1 - \varepsilon) \leq \sum\limits_{i=1}^{n} s_i \cdot \Delta x_i \leq m \cdot \delta + (n - m) \cdot \alpha \cdot \delta$.
- Dividing this inequality by δ, we get
$$n \cdot (1 - \varepsilon) \leq m + (n - m) \cdot \alpha.$$
- So, $n \cdot (1 - \alpha - \varepsilon) \leq m \cdot (1 - \alpha)$ and $m \geq n \cdot \dfrac{1 - \alpha - \varepsilon}{1 - \alpha}$.
- So, we have at least $n \cdot \dfrac{1 - \alpha - \varepsilon}{1 - \alpha}$ indices for which Δx_i has the same sign as s_i (and for which $|\Delta x_i| > \alpha \cdot \delta$).

15. Proof (cont-d)

- So, for Δx_i corr. to $(s_1, \ldots, s_n)$, at most $n \cdot \dfrac{\varepsilon}{1-\alpha-\varepsilon}$ indices have a different sign than s_i.
- It is possible that the same tuple Δx can serve two tuples $s \neq s'$. In this case:
  - going from s_i to $\text{sign}(\Delta x_i)$ changes at most $n \cdot \dfrac{\varepsilon}{1-\alpha-\varepsilon}$ signs, and
  - going from $\text{sign}(\Delta x_i)$ 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 $d(s, s') \stackrel{\text{def}}{=} \#\{i : s_i \neq s'_i\}$, we have $d(s, s') \le 2 \cdot n \cdot \dfrac{\varepsilon}{1-\alpha-\varepsilon}$.

16. Proof (cont-d)

- Thus, if $d(s,s') > 2\cdot n\cdot \dfrac{\varepsilon}{1-\alpha-\varepsilon}$, then no tuples $(\Delta x_1,\ldots,\Delta x_n)$ can serve both sign tuples s and s'.
- In this case, the two sets of tuples Δx do not intersect:
  - tuples s.t. $s_1\cdot\Delta x_1+\ldots+s_n\cdot\Delta x_n\ge n\cdot\delta\cdot(1-\varepsilon)$;
  - tuples s.t. $s'_1\cdot\Delta x_1+\ldots+s'_n\cdot\Delta x_n\ge n\cdot\delta\cdot(1-\varepsilon)$.
- Let's take take M sign tuples $s^{(1)},\ldots,s^{(M)}$ for which $d(s^{(i)},s^{(j)}) > 2\cdot\dfrac{\varepsilon}{1-\alpha-\varepsilon}$ for all $i\ne j$.
- Then the probability P that Δx serves one of these sign tuples is $\ge M\cdot p$.
- Since $P\le 1$, we have $p\le\dfrac{1}{M}$; so:
  - to prove that p_n is exponentially decreasing,
  - it is sufficient to find the sign tuples whose number M is exponentially increasing.

17. Proof (cont-d)

- 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 $d(s,s') \le \beta \cdot n$ 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, ...,
  - the number of tuples $\binom{n}{\beta \cdot n}$ that differ from s in $\beta \cdot n$ places,
- Thus, $t = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n \cdot \beta}$.

18. Proof (cont-d)

- When $\beta < 0.5$ and $\beta \cdot n < \dfrac{n}{2}$, the number of combinations $\dbinom{n}{k}$ increases with k, so $t \le \beta \cdot n \cdot \dbinom{n}{\beta \cdot n}$.
- Here, $\dbinom{a}{b} = \dfrac{a!}{b! \cdot (a-b)!}$. Since $n! \sim \left(\dfrac{n}{e}\right)^n$, we have
$$t \le \beta \cdot n \cdot \left(\frac{1}{\beta^\beta \cdot (1-\beta)^{1-\beta}}\right)^n.$$
- Here, $\gamma \stackrel{\text{def}}{=} \dfrac{1}{\beta^\beta \cdot (1-\beta)^{1-\beta}} = \exp(S)$, where $S \stackrel{\text{def}}{=} -\beta \cdot \ln(\beta) - (1-\beta) \cdot \ln(1-\beta)$ is Shannon's entropy.
- It is known that S attains its largest value 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 $\gamma < 2$.

19. Proof (cont-d)

- Let us now construct the desired collection of sign tuples $s^{(1)}, \ldots, s^{(M)}$.
  - We start with some sign tuple $s^{(1)}$, e.g., $s^{(1)} = (1, \ldots, 1)$.
  - Then, we dismiss $t \leq \gamma^n$ tuples which are $\leq \beta$-close to s, and select one of the remaining tuples as $s^{(2)}$.
  - We then dismiss $t \leq \gamma^n$ tuples which are $\leq \beta$-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 $t \cdot M \leq \beta \cdot n \cdot \gamma^n \cdot M$ sign tuples.
- So, as long as this number is smaller than the overall number 2^n of sign tuples, we can continue selecting.

20. Proof (conclusion9)

- Our 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 $M \ge \left(\frac{2}{\gamma}\right)^n \cdot \dfrac{1}{\beta \cdot n}$ tuples.
- So, we have indeed selected exponentially many tuples.
- Hence, $p_n \le \dfrac{1}{M} \le \beta \cdot n \cdot \left(\dfrac{\gamma}{2}\right)^n$, i.e.,

$$p_n \le \beta \cdot n \cdot c^n, \text{ where } c \stackrel{\text{def}}{=} \frac{\gamma}{2} < 1.$$

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