Monte-Carlo-Type Techniques for Processing Interval Uncertainty, and Their Geophysical and Engineering Applications

Matthew G. Averill, Kate C. Miller, G. Randy Keller,
Vladik Kreinovich, Jan Beck, Roberto Araiza,
Roberto Torres, and Scott A. Starks

Department of Geological Sciences and
NASA Pan-American Center for Earth
and Environmental Studies (PACES)
University of Texas at El Paso
El Paso, TX 79968, USA
contact email vladik@cs.utep.edu

Abstract

To determine the geophysical structure of a region, we measure seismic travel times and reconstruct velocities at different depths from this data. There are several algorithms for solving this inverse problem, but these algorithms do not tell us how accurate these reconstructions are.

Traditional approach to accuracy estimation assumes that the measurement errors are independently normally distributed. Problem: the resulting accuracies are not in line with geophysical intuition. Reason: a typical error is when we miss the first arrival of the seismic wave; it is not normal (bounded by the wave period Δ) and not independent.

Typically, all we know is the upper bound Δ on the measurement error, so when the measured value is $\widetilde{x}$, we conclude that $x \in [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$. For this interval uncertainty, the resulting velocity accuracy is, qualitatively, in much better accordance with geophysics.

Interval uncertainty naturally appears in other applications as well. In this paper, we describe Monte-Carlo-Type techniques for processing interval uncertainty, and their geophysical and engineering applications.

Keywords: interval uncertainty, inverse problem, geophysics, seismic data

1 Introduction to the Problem

Uncertainty is important. In engineering, decisions about the best design are usually made under uncertainty. The main source of uncertainty is measurement errors. Additional source of uncertainty is that we do not know how exactly the devices will be used. For example, when designing a building, we have limits L_i on the loads l_i in different rooms i, but we do not know how exactly these loads will be distributed, and we want to make sure that our design is safe for all possible $l_i \le L_i$.

General problem: uncertainty of the result of data processing. In data processing, we input the measurement results x_i and output the estimate value of an engineering and/or physical quantity y. We know the algorithm $f(x_1, \ldots, x_n)$, we know the measured values $\widetilde{x}_1, \ldots, \widetilde{x}_n$, and we have some information about the uncertainty $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ of each direct measurement. Based on this information, how can we estimate the uncertainty $\Delta y = \widetilde{y} - y$ of the algorithm's output?

Types of uncertainty. The answer to the above question depends on what we know about the uncertainty of Δx_i. In general, to get a complete information about the uncertainty of each Δx, we must know what are the possible values of Δx, and how often can different possible values occur.

So, in the ideal case of *full* information about uncertainty, we know the cdf $F_i(t) \stackrel{\text{def}}{=} \text{Prob}(x_i \le t)$ for each variable x_i (and we usually know that x_i are independent).

As we have mentioned, often, in applications, we encounter an interval case when we have *no* information about the probabilities; for each input x_i, we only know the interval $[\underline{x}_i, \overline{x}_i]$ of possible values of each x_i. For example, if we know the result $\widetilde{x}_i$ of measuring x_i, and we know the upper bound Δ_i on the measurement error Δx_i of i-th measurement, then we know that the actual value x_i must lie in the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

The general case is when we have *partial* information about the probabilities. Since a full description of a probability distribution means that for every t, we know the exact value of the cdf $F_i(t)$, the partial information means that for every t, we know an interval $[\underline{F}_i(t), \overline{F}_i(t)]$ that contains $F_i(t)$. This interval-valued cdf is called a *p-box*; see, e.g., [2].

An important particular case is Dempster-Shafer uncertainty, when for each variable x_i and for every t, we have finitely many intervals $[\underline{x}_i^{(k)}(t), \overline{x}_i^{(k)}(t)]$ with probabilities $p_i^{(k)}$ attached to them, and we know that $x_i \in [\underline{x}_i^{(k)}(t), \overline{x}_i^{(k)}(t)]$ with probability $p_i^{(k)}$.

In general, we may have different types of information for different inputs x_i, and we may have dependent Δx_i.

Black box. Traditional interval computations approach to solving the above problem is to parse the algorithm f and to apply f step-by-step to the corresponding "uncertain numbers": intervals [4], probability distributions, p-boxes [2], etc.

The problem is that in several practical situations, f is given as a *black box*: we do not know the sequence of steps forming f, we can only plug in different values into f and see the results. For example, commercial software is often given as a black box, to safeguard against competitors; classified security-related software is given as a black box to safeguard against adversary.

An additional problem is that sometimes, applying the algorithm f takes so much time that it is only possible to run it a few times. This happens, for example, in many geophysical applications.

2 Existing Methods for Solving this Problem

Sensitivity analysis: reminder. One of the standard engineering techniques for solving such problems is the technique of sensitivity analysis. This techniques is applicable when the algorithm $f(x_1, \ldots, x_n)$ is monotonic (increasing or decreasing) with respect of each of its variables – when $x_i \in [\underline{x}_i, \overline{x}_i]$. For example, this monotonicity property is true if we can linearizable f, i.e., replace it with a linear function without losing accuracy.

In this case, e.g., if we know the intervals $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ of possible values of x_i, we can find the range $[\underline{y}, \overline{y}]$ of $y = f(x_1, \ldots, x_n)$ as follows:

- First, compute $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- Then, for each i, we determine whether f is increasing or decreasing in x_i by computing $y_i' \stackrel{\text{def}}{=} f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \widetilde{x}_i + h, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n)$ for some $h > 0$.
- Finally, we compute $\underline{y} = f(x_1^-, \ldots, x_n^-)$ and $\overline{y} = f(x_1^+, \ldots, x_n^+)$, where:
  - if $y_i' \geq \widetilde{y}$, then $x_i^- = \underline{x}_i$ and $x_i^+ = \overline{x}_i$;
  - if $y_i' < \widetilde{y}$, then $x_i^- = \overline{x}_i$ and $x_i^+ = \underline{x}_i$.

The only problem with this approach is that we need $n + 3$ calls to f, and in many problems, n is very large: e.g., in ultrasonic testing, we record (= measure) signal values at thousands moments of time. For large n and for complex f, this method is too slow.

Cauchy deviate method. To speed up the computations, in [6, 7], we describe a faster Monte-Carlo-type techniques based on Cauchy distributions. This technique is applicable when the algorithm f is linearizable, i.e., when $\Delta y = c_1 \cdot \Delta x_1 + \ldots + c_n \cdot \Delta x_n$ for some coefficients c_i (namely, for $c_i = \dfrac{\partial f}{\partial x_i}$).

In this case, the range of y can be described as $[\underline{y}, \overline{y}] = [\widetilde{y} - \Delta, \widetilde{y} + \Delta]$, where $\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i$.

Cauchy distribution is a distribution with the probability density $\rho(x) = \dfrac{\Delta}{\pi \cdot (x^2 + \Delta^2)}$. We use this distribution because it is known that if $\xi_1, \ldots, \xi_n$ are independent Cauchy with parameters Δ_i, then $\sum_{i=1}^{n} c_i \cdot \xi_i$ is also Cauchy distributed with the parameter $\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i$. This fact leads to the following algorithm for estimating Δ. We fix the number of simulations N; then:

- for every i and for $k = 1, \ldots, N$, we simulate Cauchy distributed random variables as $\delta x_i^{(k)} = \Delta_i \cdot \tan(\pi \cdot (r_i - 0.5))$, where $r_i = U[0, 1]$ is uniformly distributed over the interval $[0, 1]$;
- compute $\delta y^{(k)} \stackrel{\text{def}}{=} f(\widetilde{x}_1 + \delta x_1^{(k)}, \ldots, \widetilde{x}_n + \delta x_n^{(k)}) - \widetilde{y}$;
- use the Cauchy-distributed sample $\delta y^{(k)}$ to find the parameter Δ by using the Maximum Likelihood Method, which, for Cauchy distribution, leads to the following equation:

$$\frac{1}{1 + \left(\dfrac{\delta y^{(1)}}{\Delta}\right)^2} + \ldots + \frac{1}{1 + \left(\dfrac{\delta y^{(N)}}{\Delta}\right)^2} = \frac{N}{2}.$$

The advantage of this method is that the required the number N of calls to a model f depends only on the desired accuracy ε and not on the number of inputs n.

For example, after $N = 200$ runs, we get 20% accuracy $0.2 \cdot \Delta$ with 95% certainty (corresponding to $2\sigma_e$). So, if $n \geq 200$, the Cauchy deviate method is much faster that sensitivity analysis.

3 Applications: Brief Overview

Cauchy deviate techniques are used in several applications [5, 6, 7]. In environmental and power engineering, the Cauchy deviate method is used to provide safety analysis of complex systems. In civil engineering, it is used to provide building safety; here, f is a Finite Element Method. In petroleum and geotechnical engineering, f is an algorithm that solves inverse problem, i.e., reconstructs the density (or, to be more accurate, the velocity of sound) at different points from the times x_i that it takes a seismic signal to travel from the source to the sensor [1, 3, 8].

As a result, in the environmental and civil engineering, same results as sensitivity analysis, but faster. In geotechnical engineering, the dependence of the accuracy on the location and depth fits much better with the geophysicists' understanding than statistical estimates.

4 Limitations of Cauchy Deviate Techniques

The Cauchy deviate technique is based on the following *assumptions:*

- that the measurement errors are *small*, so we can safely linearize the problem;
- that we only have *interval* information about the uncertainty, and
- that we can actually call the program f 200 times.

In real-life engineering problems, these assumptions are often not satisfied. In this paper, we describe how we can modify the Cauchy techniques to overcome these limitations.

5 What If We Cannot Perform Many Iterations

As we have mentioned, in many real-life engineering problems, we cannot run f 200 times. What can we do?

Our idea is to use Cauchy estimates with the available amount of $N \ll 200$ iterations, but use new formulas for Δ. The possibility of using this idea is based on the fact that, due to the Central Limit Theorem, for reasonable large N, the distribution for $\widetilde{\Delta} - \Delta$ is approximately Gaussian, with relative standard deviation $\sqrt{2/N}$. So, we can conclude that with certainty 95% (corresponding to 2σ), we have $\Delta \le \widetilde{\Delta} \cdot \left(1 + k_0 \cdot \sqrt{\frac{2}{N}}\right)$ (where $k_0 = 2$). To get 99.9% certainty (corresponding to 3σ), we can take $k_0 = 3$.

For example, for $N = 50$, we conclude that $\Delta \le 1.4 \cdot \widetilde{\Delta}$, which is not a bad. estimate.

For smaller N, the difference $\widetilde{\Delta} - \Delta$ is not Gaussian, so we must empirically find the corresponding factor.

6 Dempster-Shafer (DS) Knowledge Bases

In the DS approach, for each i, instead of a single interval $\mathbf{x}_i$, we have several intervals $\mathbf{x}_i^{(k)}$ with probabilities $p_i^{(k)}$. In principle, we could consider all possible combinations of these intervals and apply the above approach to each such

combination. However, even if we have 2 intervals for $n = 50$ inputs, we have an astronomical number of $2^{50} \approx 10^{15}$ output intervals.

To handle the DS situation, let us recall that when $\mathbf{x}_i = [x_i^{\rm mid} - \Delta_i, x_i^{\rm mid} + \Delta_i]$, then $\mathbf{y} = [y^{\rm mid} - \Delta, y^{\rm mid} + \Delta]$, where $y^{\rm mid} = \widetilde{y} + \sum_{i=1}^{n} c_i \cdot (x_i^{\rm mid} - \widetilde{x}_i)$ and $\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i$. In the DS case, we have different pairs $(y^{{\rm mid}(k)}, \Delta_i^{(k)})$ with different probabilities.

Our main idea is that, due to the Central Limit Theorem, $(y^{\rm mid}, \Delta)$ is approximately normally distributed. (It is worth mentioning that the distribution is not *exactly* normal since $\Delta \geq 0$.) How can we use this idea?

In the Cauchy approach, we used Cauchy distribution with given Δ; its characteristic function is $E[\exp(\mathrm{i} \cdot \omega \cdot \xi)] = \exp(-|\omega| \cdot \Delta)$. Now, we have a Gaussian mixture of several Cauchy distributions, with different Δ. The resulting characteristic function has the form:

$$E[\exp(\mathrm{i} \cdot \omega \cdot \xi)] = \int \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp\left(-\frac{(\Delta - \mu)^2}{2\sigma^2}\right) \cdot \exp(-|\omega| \cdot \Delta)\, \mathrm{d}\Delta.$$

Explicitly integrating over Δ, we arrive at the following simplified expression: $E[\exp(\mathrm{i} \cdot \omega \cdot \xi)] = \exp\left(\frac{1}{2} \cdot \sigma^2 \cdot \omega^2 - \mu \cdot |\omega|\right)$. Thus, we can apply the following algorithm:

- For different real values $\omega_1, \ldots, \omega_k > 0$, compute $l(\omega_k) \stackrel{\rm def}{=} -\ln(c(\omega_k))$, where $c(\omega_k) \stackrel{\rm def}{=} \frac{1}{N} \cdot \sum_{k=1}^{N} \cos(\omega \cdot y^{(k)})$.
- Use the Least Squares Method to find the values μ and σ for which

$$\mu \cdot \omega_k - \frac{1}{2}\sigma^2 \cdot \omega_k^2 \approx l(\omega_k).$$

The resulting value μ is the average Δ.

We repeat the above algorithm twice: for samples for which $y^{\rm mid} \leq E[y^{\rm mid}]$, and for samples for which $y^{\rm mid} > E[y^{\rm mid}]$. Based on two μ's, we compute $E[\Delta]$ and $\sigma[\Delta]$.

What about p-boxes? It is known that a p-box can be described as a DS knowledge base. Specifically, a p-box $[\underline{F}(t), \overline{F}(t)]$ can be described by listing, for each p, the interval $[\underline{f}(p), \overline{f}(p)]$ of the possible quantile values:

- the function $\underline{f}(p)$ is an inverse function to $\overline{F}(t)$, and
- the function $\overline{f}(p)$ is an inverse function to $\underline{F}(t)$.

So, whatever method we have for DS knowledge bases, we can apply it to p-boxes as well.

Similarly, if we have different types of uncertainty for different x_i, we can translate them into p-boxes, and apply the above technique.

7 Cauchy Method for Quadratic f

So far, we have considered the method for the case when the algorithm f is linearizable, i.e., when quadratic and higher order terms can be ignored. A natural next case is when linear terms are still prevailing, but quadratic terms can no longer be ignored, i.e.,

$$\Delta y \stackrel{\text{def}}{=} \sum_{\imath=1}^{n} c_i \cdot \Delta x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \cdot \Delta x_i \cdot \Delta x_j.$$

In this case, since linear terms are prevailing, max and min of Δy are attained when $\Delta x_i = \pm\Delta_i$ (depending on $\varepsilon_i \stackrel{\text{def}}{=} \text{sign}(c_i)$):

$$\Delta^+ = \sum_{i=1}^{n} |c_i| \cdot \Delta_i + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \cdot \varepsilon_i \cdot \varepsilon_j \cdot \Delta_i \cdot \Delta_j;$$

$$\Delta^- = \sum_{i=1}^{n} |c_i| \cdot \Delta_i - \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \cdot \varepsilon_i \cdot \varepsilon_j \cdot \Delta_i \cdot \Delta_j.$$

We can explicitly use this formula, but for large n, literal computation takes too long. It is therefore desirable to design a Cauchy-type method for this case.

This algorithm can indeed be designed. It uses the following *auxiliary algorithm* $z = (z_1, \ldots, x_n) \to g(z)$: we apply the linear Cauchy deviate method to the auxiliary function $t \to \dfrac{1}{2} \cdot (f(\widetilde{x} + z + t) - f(\widetilde{x} + z - t))$ and the values $t_i \in [-\Delta_i, \Delta_i]$.

Now, the main algorithm is as follows:

- We apply the algorithm $g(z)$ to the vector $0 = (0, \ldots, 0)$, thus computing the value $g(0)$.
- We apply the linear Cauchy deviate method to the auxiliary function

  $$h(z) = \frac{1}{2} \cdot (g(z) - g(0) + f(\widetilde{x} + z) - f(\widetilde{x} - z));$$

  the result is the desired value Δ^+.
- Finally, we compute Δ^- as $2g(0) - \Delta^+$.

Using the known properties of the Cauchy deviate method, it is easy to check that this algorithm indeed leads to the desired values of Δ^- and Δ^+.

This method uses N calls to g, each of which means n calls to f, so overall, we need $2N^2$ calls to f. Thus, for $N = 200$, this method is more efficient than sensitivity analysis if

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, by Army Research Laboratories grant
DATM-05-02-C-0046, by NIH grant 3T34GM008048-20S1, by Applied Biomathematics, and by a research grant from Sandia National Laboratories as part of the Department of Energy Accelerated Strategic Computing Initiative (ASCI).

References

[1] D. I. Doser, K. D. Crain, M. R. Baker, V. Kreinovich, and M. C. Gerstenberger, "Estimating uncertainties for geophysical tomography", *Reliable Computing*, 1998, Vol. 4, No. 3, pp. 241–268.

[2] S. Ferson, *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*, CRC Press, Boca Raton, Florida, 2002.

[3] J. A. Hole, "Nonlinear High-Resolution Three-Dimensional Seismic Travel Time Tomography", *J. Geophysical Research*, 1992, Vol. 97, No. B5, pp. 6553–6562.

[4] L. Jaulin, M. Keiffer, O. Didrit, and E. Walter, *Applied Interval Analysis*, Springer-Verlag, Berlin, 2001.

[5] V. Kreinovich, J. Beck, C. Ferregut, A. Sanchez, G. R. Keller, M. Averill, and S. A. Starks, "Monte-Carlo-type techniques for processing interval uncertainty, and their engineering applications", *Proceedings of the Workshop on Reliable Engineering Computing*, Savannah, Georgia, September 15–17, 2004, pp. 139–160.

[6] V. Kreinovich and S. Ferson, "A New Cauchy-Based Black-Box Technique for Uncertainty in Risk Analysis", *Reliability Engineering and Systems Safety*, 2004, Vol. 85, No. 1–3, pp. 267-279.

[7] R. Trejo and V. Kreinovich, "Error Estimations for Indirect Measurements", In: S. Rajasekaran et al., eds., *Handbook on Randomized Computing*, Kluwer, 2001, pp. 673–729.

[8] C. A. Zelt and P. G. Barton, "Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin", *J. Geophysical Research*, 1998, Vol. 103, No. B4, pp. 7187–7210.