Reliable Computing 1 (1) 1995, pp. 33–40

Why intervals? A simple limit theorem that is similar to limit theorems from statistics

Vladik Kreinovich

What is the set of possible values of a measurement error? In the majority of practical applications, an error is caused not by a single cause; it is caused by a large number of independent causes, each of which adds a small component to the total error. This fact is widely used in statistics: Namely, since it is known that the distribution of the sum of many independent small random variables is close to one of the so-called infinitely divisible ones (a class that includes the well-known Gaussian distribution), we can safely assume that the distribution of the total error is infinitely divisible. This assumption is used in the majority of the statistical applications.

In this paper, we prove a similar result for the set of possible values of an error. Namely, if an error equals the sum of many small independent components, then its set of possible values is close to an interval; the smaller the components, the closer this set to an interval.

This result provides one more justification for using intervals in data processing.

Почему интервалы? Простая предельная теорема, аналогичная предельным теоремам статистики

В. Крейнович

Что представляет собой множество значений погрешностей измерений? В большинстве практических приложений возникновение погрешностей обусловлено не единственной причиной, а большим количеством независимых друг от друга причин, каждая из которых вносит свой маленький вклад в общую погрешность. Этот факт широко используется в статистике: поскольку известно, что распределение суммы большого числа малозначащих случайных переменных близко к одному из так называемых бесконечно делимых распределений (к которым принадлежит и широко известное нормальное распределение), можно с уверенностью считать, что распределение суммарной погрешности бесконечно делимо. Этот вывод используется в большинстве статистических приложений.

В настоящей работе доказывается аналогичный результат для множества возможных значений погрешности. А именно, если значение погрешности равно сумме большого количества независимых компонент малой величины, то множество ее возможных значений близко к интервалу, причем чем меньше значимость (вклад) компонентов, тем ближе это множество к интервалу.

Данный результат является еще одним аргументом в пользу применения интервалов при обработке данных.

1. Formulation of a real-life problem: What is the set of all possible values of an error? Is it necessarily an interval?

Main question: are all values of error possible? Suppose that we have a measuring device that measures a physical quantity x (e.g., mass or voltage) with a guaranteed precision ε. This precision is usually supplied by the manufacturer of this device. The word "guaranteed" means that if the measurement result is $\tilde{x}$, then we are sure that the actual value x satisfies the inequality $|\tilde{x} - x| \le \varepsilon$.

In other words, possible values of an error $e = \tilde{x} - x$ belong to an interval $[-\varepsilon, \varepsilon]$, and possible values of x belong to an interval $[\tilde{x} - \varepsilon, \tilde{x} + \varepsilon]$.

If this estimate is an "overshoot" in the sense that in practice the errors are always smaller, then we are not using this device at its best: its results are more precise and thus more reliable than we think. So, it is important to get this ε as small as possible.

Now, suppose that this ε is indeed the smallest possible in the sense that:

- no errors are greater than ε but
- there have been cases (during the testing) with the errors pretty close to ε and to $-\varepsilon$.

In other words, both endpoints of the interval $[-\varepsilon, \varepsilon]$ are possible values of the error.

The next question is: *are all internal points of that interval possible values of the error?*

In some exotic cases, the answer is "no". We can imagine realistic situations when both values $-\varepsilon$ and ε are possible values of the error, while some values inside an interval $[-\varepsilon, \varepsilon]$ cannot occur as the values of the error. For example, suppose that we are measuring the electromagnetic field in the close vicinity of a computer memory element. In this situation, the external field caused by this element is the main source of error. This element can be in two possible states (depending on whether it represents bit "1" or bit "0"), so we have two possible values of an error. Crudely speaking, for this situation, the set of possible values of e consists of only two points $\{-\varepsilon, \varepsilon\}$, and does not contain any internal values at all.

If in addition to this main source of error, we take into consideration other possible sources of error, then the resulting set of possible values of total error becomes a union of two small intervals: one close to $-\varepsilon$, and a one close to ε.

In the majority of the cases, all interior values are possible. The case when we have one prevailing cause of error is really exotic. In the majority of cases, an error is arising from the cumulative effect of a large number of independent factors. In these cases, experiments usually show that all the interior values are possible (see, e.g., a survey monograph [5] and references therein). In other words, the set of all possible values of the error forms an interval $[-\varepsilon, \varepsilon]$.

A question. *Why is the set of all possible values of e an interval? Is it an empirical fact or a theoretically justified law?*

What we are planning to do. In this paper, we prove the fact that all values from an interval are possible can be theoretically justified (in the same manner as the normal distribution is).

From the mathematical viewpoint, this result is extremely simple to prove. However, we believe that our result is worth writing down, because it provides one more explanation of why intervals are so widely used in data processing (there are lots of examples starting from the pioneer paper [4]; for a latest survey, see, e.g., [1]).

2. How to explain why the set of all possible values of an error is an interval: the main idea

An analogy with limit theorems of mathematical statistics. A similar situation is analyzed in statistics: We have a random error that is caused by a large number of different factors. Therefore, this error is a sum of the large number n of small independent component random variables. It is known that when $n \to \infty$, the distribution law for such a sum tends to one of the so-called *infinitely divisible* distributions (see, e.g., [3]; for more recent results see, e.g., [2]). This class includes the well-known Gaussian (= normal) distribution.

Therefore, for sufficiently big n, we can use infinitely divisible distributions as a perfect approximation for the error distribution.

Comment. Traditionally, in statistics, mainly Gaussian distribution is used (see, e.g., [8], pp. 2.17, 6.5, 9.8, and references therein). However, other distributions are also necessary because error distribution is often non-Gaussian (see, e.g., [5, 6]).

How we are going to use this analogy. We consider the case when the error e is equal to the sum of small independent components: $e = e_1 + e_2 + \cdots + e_n$. To make this a mathematical statement, we must somehow explain what "small" means, and what "independent" means.

Comment. In this section, we will try not only to give the definitions, but to provide *motivations* for these definitions. The resulting definitions will be then formally stated (or repeated) in Section 3 (so, a reader who is interested in the mathematical result only, can skip this section, and go to Section 3).

Let us denote the set of all possible values of a component e_i by E_i.

What does "small" mean? If a number $\delta > 0$ is fixed, we say that a component is *δ-small* if all its possible values do not exceed δ, i.e., if $|a| \le \delta$ for all $a \in E_i$.

What does "independent" mean? This is easy to explain. For example, let's consider the case when the components e_i and e_j are not independent; e.g., they are mainly caused by the same factor and must therefore be α-close for some small α. Then, for a given value of e_i, the corresponding set of possible values of e_j is equal to $[e_i - \alpha, e_i + \alpha]$, and is thus different for different e_i.

Components e_i and e_j are *independent* if the set of possible values of e_i does not depend on the value of e_j.

In other words, this means that all pairs (e_i, e_j), where $e_i \in E_i$ and $e_j \in E_j$, are possible. Therefore, the set of all possible values of the sum $e_i + e_j$ coincides with the set $\{e_i + e_j : \ e_i \in E_i, e_j \in E_j\}$, i.e., with the sum $E_i + E_j$ of the two sets E_i and E_j.

Before we turn to formal definitions, we need to make one more remark.

We will consider closed sets of possible values. Our point is that if the set of all possible values of an error is not closed, we will never be able to find that out. Indeed, suppose that E is not closed. This means that there exists a value e that belongs to the closure of E, but does not belong to E itself.

Let us show that in every test measurement, we could get this value e as the measured value of error. Indeed, in every test measurement, we measure error with some accuracy δ.

Since e belongs to the closure of E, there exists a value $e' \in E$ such that $|e' - e| \le \delta$. So, if the actual error is e' (and $e' \in E$, and is thus a possible value of an error), we can get e as a result of measuring that error. So, no matter how precisely we measure errors, e is always possible. Therefore, we will never be able to experimentally distinguish between the cases when e is possible and when it is not.

In view of that, to add e to E or not to add is purely a matter of convenience. Usually, the border values are added. For example, we usually consider closed intervals $[-\varepsilon, \varepsilon]$ as sets of possible values. Following this usual agreement, we will assume that the sets E and E_i are closed.

Now, we are ready for formal definitions.

3. Main result

Let us give some (more or less) standard definitions and denotations.

Remark. In this section, we will consider only 1-dimensional case (i.e., all our sets will be subsets of a real line R).

Definitions and denotations. By a *sum* $A + B$ of two sets $A, B \subseteq R$, we understand the set $\{a + b : a \in A, b \in B\}$. For a given $\delta > 0$, a set A is called *δ-small* if $|a| \le \delta$ for all $a \in A$. By a distance $\rho(A, B)$ between sets A and B, we will understand Hausdorff distance (so, for sets, terms like "δ-close" will mean δ-close in the sense of ρ).

Comment. For reader's convenience, let us reproduce the definition of Hausdorff distance: $\rho(A, B)$ is the smallest real number δ for which the following two statements are true:

- for every $a \in A$, there exists a $b \in B$ such that $|a - b| \le \delta$;
- for every $b \in B$, there exists an $a \in A$ such that $|a - b| \le \delta$.

Proposition 1. *If $E = E_1 + \cdots + E_n$ is a sum of δ-small closed sets from R, then E is δ-close to an interval.*

Comments.

1. This result proves that if e is a sum of a large number of independent small components, then the set of all possible values of e is close to an interval.

2. All the proofs are given in Section 5.

Proposition 2. *If $E \subseteq R$ is a bounded set, and for every $\delta > 0$, E can be represented as a finite sum of δ-small closed sets, then E is an interval.*

Comment. Proposition 2 is similar to the description of infinitely divisible distributions. Namely, it gives the following description of infinitely divisible sets: If a bounded set is infinitely divisible (i.e., representable as a sum of arbitrarily small terms), then this set is an interval.

4. An auxiliary result for a multi-dimensional case

In multi-dimensional case (i.e., for subsets of R^k for $k > 1$), no final result similar to Propositions 1 and 2 is known. What we can prove is the following:

Proposition 3. *If $E = E_1 + \cdots + E_n$ is a sum of δ-small closed sets from R^k $(k \ge 1)$, then E is δ-close to a connected set.*

Proposition 4. *If $E \subseteq R^k$ is a bounded set, and for every $\delta > 0$, E can be represented as a finite sum of δ-small closed sets, then E is connected.*

Comments.

1. The formulations of these results emerged from the suggestion of Sergey P. Shary.

2. These results are not final, because not all connected sets can be thus represented. Our hypothesis is that under the conditions of Proposition 4, E is a convex compact set.

5. Proofs

Proof of Proposition 1. Since each set E_i is δ-small, it is bounded. Since E_i is also closed, it contains its least upper bound $sup\, E_i$, and its greatest lower bound $inf\, E_i$ (see, e.g., [7]). Let us denote $sup\, E_i$ by e_i^+, and $inf\, E_i$ by e_i^-. Then, $\{e_i^-, e_i^+\} \subseteq E_i \subseteq [e_i^-, e_i^+]$. Therefore, $\underline{E} \subseteq E \subseteq \overline{E}$, where we denoted $\underline{E} = \{e_1^-, e_1^+\} + \{e_2^-, e_2^+\} + \cdots + \{e_n^-, e_n^+\}$, $\overline{E} = [e_1^-, e_1^+] + [e_2^-, e_2^+] + \cdots + [e_n^-, e_n^+] = [e^-, e^+]$,

$$e^- = \sum_{i=1}^{n} e_i^-$$

and

$$e^+ = \sum_{i=1}^{n} e_i^+.$$

Let us show that E is δ-close to the interval $\overline{E}$. Since $E \subseteq \overline{E}$, every element $a \in E$ belongs to $\overline{E}$. So, it is sufficient to prove that if $b \in \overline{E}$, then b is δ-close to some $a \in E$.

We will show that b is δ-close to some a from the set $\underline{E}$ (which belongs to E because $\underline{E} \subseteq E$). Indeed, by definition of the sum of the sets, the set $\underline{E}$ contains, in particular, the following points:

$$\begin{aligned}
a_0 &= e_1^- + e_2^- + \cdots + e_n^-,\\
a_1 &= e_1^+ + e_2^- + \cdots + e_n^-,\\
a_2 &= e_1^+ + e_2^+ + e_3^- + \cdots + e_n^-,\\
&\dots\\
a_i &= e_1^+ + e_2^+ + \cdots + e_i^+ + e_{i+1}^- + \cdots + e_n^-,\\
&\dots\\
a_n &= e_1^+ + e_2^+ + \cdots + e_n^+.
\end{aligned}$$

Notice that the values a_0 and a_n coincide with the endpoints e^-, e^+ of the interval $\overline{E}$.

Each value a_i is obtained from the previous one by changing one term in the sum (namely, e_i^-) to another term that is not smaller than e_i^-, namely, to e_i^+. Therefore, $a_0 \le a_1 \le a_2 \le \cdots \le a_n$.

The difference between two consequent terms in this sequence is equal to $a_i - a_{i-1} = e_i^+ - e_i^-$. Since each E_i is δ-small, we have $|e_i^+| \le \delta$, $|e_i^-| \le \delta$, and therefore, $|a_i - a_{i-1}| = |e_i^+ - e_i^-| \le |e_i^+| + |e_i^-| \le 2\delta$. So, the distance between any two consequent numbers in a sequence $a_0 \le a_1 \le \cdots \le a_n$ is $\le 2\delta$.

Now, suppose that we are given a number $b \in \overline{E} = [a_0, a_n]$. If $b = a_i$ for some i, then we can take $a = a_i = b$. So, it is sufficient to consider the case when $b \ne a_i$ for all i. In particular, in this case, $a_0 < b < a_n$. The value $a_0 - b$ is negative, the value $a_n - b$ is positive, so the sign of $a_i - b$ must change from $-$ to $+$ somewhere. Let us denote by i the value where it changes, i.e., the value for which $a_i - b < 0$ and $a_{i+1} - b > 0$. For this i, $a_i < b < a_{i+1}$. Therefore,

$$|a_i - b| + |a_{i+1} - b| = (b - a_i) + (a_{i+1} - b) = a_{i+1} - a_i \leq 2\delta.$$

The sum of two positive numbers $|a_i - b|$ and $|a_{i+1} - b|$ does not exceed 2δ. Hence, the smallest of these two numbers cannot exceed the half of 2δ, i.e., cannot exceed δ. So, either for $a = a_i$, or for $a = a_{i+1}$, we get $|a - b| \leq \delta$. Hence, E is δ-close to the interval $\overline{E}$. □

Proof of Proposition 2. Let E be a set that satisfies the condition of this proposition. Since E is a sum of finitely many closed sets, it is itself closed. Since E is bounded and close, it contains $inf\, E$ and $sup\, E$. So, $E \subseteq [inf\, E, sup\, E]$. Let us prove that $E = [inf\, E, sup\, E]$.

Indeed, let e be an arbitrary point from an interval $[inf\, E, sup\, E]$. Let us prove that $e \in E$. Indeed, for every natural k, we can take $\delta_k = 2^{-k}$. Since $\delta_k > 0$, E is a sum of closed δ_k-small sets. Therefore, according to Proposition 1, there exists a $e_k \in E$ such that $|e_k - e| \leq \delta_k = 2^{-k}$. So, $e = lim\, e_k$, where $e_k \in E$, and e is thus a limit point for E. Since E is closed, $e \in E$. □

Proof of Proposition 3. Let's denote the Euclidean norm on R^k by $\|\cdot\|$.

1°. Let us first prove that for every two points $e, e' \in E$, there exists a finite sequence $e^{(0)} = e, e^{(1)}, \ldots, e^{(n)} = e'$ of elements of E for which $\|e^{(i)} - e^{(i-1)}\| \leq \delta$ for all i.

Indeed, since $E = E_1 + \cdots + E_n$, we have $e = e_1 + \cdots + e_n$ for some $e_i \in E_i$. Similarly, we have $e' = e'_1 + \cdots + e'_n$ for some $e'_i \in E_i$. Let us take $e^{(i)} = (e'_1 + \cdots + e'_i) + (e_{i+1} + \cdots + e_n)$. Then, by definition, $e^{(0)} = e$, $e^{(n)} = e'$. Now, $e^{(i)} - e^{(i-1)} = e'_i - e_i$. Since both elements e_i and e'_i belong to E_i, and E_i is δ-small, we conclude that $\|e_i\| \leq \delta$, $\|e'_i\| \leq \delta$, and therefore, $\|e_i - e'_i\| \leq \|e_i\| + \|e'_i\| \leq 2\delta$. So, $\|e^{(i)} - e^{(i-1)}\| \leq 2\delta$.

2°. Let us now construct the following set C: we take all pairs $(e, \tilde{e})$ of 2δ-close points from E, and connect each pair by a straight line segment $[e, \tilde{e}]$ (in particular, we allow the degenerate pairs (e, e) for which $\tilde{e} = e$; for such pairs, the connecting segment consists of this very point e). The union of all these segments is our C.

3°. Let us show that C is a connected set.

Indeed, assume that $c \in C$ and $c' \in C$. By definition of C, this means that $c \in [e, \tilde{e}] \subseteq C$, and $c' \in [e', \tilde{e}'] \subseteq C$ for some $e, e', \tilde{e}, \tilde{e}' \in E$. In particular, the segments $[e, c]$ and $[c', e']$ (that are subsegments of $[e, \tilde{e}]$ and $[e', \tilde{e}']$) are subsets of C.

Due to 1°, there exists a sequence $e^{(0)} = e, e^{(1)}, \ldots, e^{(n)} = e'$ for which $e^{(i)} \in E$ and $\|e^{(i)} - e^{(i-1)}\| \leq 2\delta$. By construction of C, it means that the segment $[e^{(i)}, e^{(i-1)}]$ belongs to C. So, the points c and c' can be connected by a finite sequence of adjacent segments: $[c, e] = [c, e^{(0)}], [e^{(0)}, e^{(1)}], \ldots, [e^{(n-1}, e^{(n)}] = [e^{(n-1)}, e'], [e', c']$. Therefore, C is connected.

4°. To complete the proof, let us show that $\rho(E, C) \leq \delta$.

Indeed, by definition of C, every point $e \in E$ belongs to C (because it belongs to a degenerate segment $[e, e]$). So, we can take $c = e$ and thus guarantee that $\|c - e\| = 0 \leq \delta$.

Now, if $c \in C$, this means that $c \in [e, e']$ for some $e, e' \in E$ for which $\|e - e'\| \leq 2\delta$. The point c is either closer to e, or it is closer to c', or it exactly in the middle between e and e'. Let's consider all three cases.

- If c is closer to e, then the distance $\|c - e\|$ is not greater than half of the segment's length (hence, not greater than δ).
- If c is closer to e', then similarly, $\|c - e'\| \leq \delta$.

- If c is in the middle, then $\|c - e\| = \|c - e'\| \leq \delta$.

In all three cases, there exists an $e \in E$ for which $\|c - e\| \leq \delta$. So, $\rho(C, E) \leq \delta$. □

Comments.

1. This proof is similar to the proof of Proposition 1.

2. For $k = 1$, intervals are the only bounded conected closed sets. So, Proposition 1 follows from Proposition 3.

Proof of Proposition 4. Similarly to the proof of Proposition 2, one can prove that E is bounded and close and therefore, E is compact. Let's prove by reduction to a contradiction that C cannot be disconnected. Indeed, if E is disconnected, this means that it can be represented as a union of two disjoint closed subsets E' and E''. Since E is bounded, both E' and E'' are compact sets. Therefore, there exists points $e' \in E'$ and $e'' \in E''$ for which the distance $\|e' - e''\|$ is the smallest possible. This smallest distance cannot be equal to 0, because then $e' = e''$ would be a common point of disjoint sets. Therefore, this distance is positive.

Let's take $1/3$ of this distance as δ. According to the condition of the proposition, E can be represented as a finite sum of δ-small sets. Therefore, due to statement 1° from the proof of Proposition 3, we can conclude that there exists a sequence $e^{(0)} = e, e^{(1)}, \ldots, e^{(n)} = e'$ of elements of E for which $e^{(0)} = e'$, $e^{(n)} = e''$, and $\|e^{(i)} - e^{(i-1)}\| \leq \delta$ for all i. The first elements of this sequence belongs to E', the last one belongs to E''. If by i we denote the index of the first element of this sequence that belongs to E'', then $e^{(i-1)} \in E'$, $e^{(i)} \in E''$, and $\|e^{(i)} - e^{(i-1)}\| \leq 2\delta = (2/3)\|e' - e''\| < \|e' - e''\|$. This inequality contradicts to our choice of e' and e'' as the pair with the smallest possible distance. This contradiction shows that our assumption is false, and E cannot be disconnected. So, E is connected. □

6. Conclusions

We prove that if an error e is a sum of a large number of independent small component errors, then the set E of its possible values is close to an interval. The smaller the components, the closer E to an interval. This result justifies the use of intervals in data processing.

This limit theorem is similar to limit theorems of mathematical statistics that justify (in a similar manner) the use of infinitely divisible distributions (in particular, the use of the Gaussian distribution).

Acknowledgments. This work was supported by an NSF grant No. CDA−9015006, NASA Grant No. 9−482, and Grant No. PF90−018 from the General Services Administration (GSA), administered by the Materials Research Institute and the Institute of Manufacturing and Materials Management. The author is grateful to Sergey P. Shary for valuable improvements.

References

[1] Aberth, O. *Precise numerical analysis*. Wm. C. Brown Publishers, Dubuque, Iowa, 1988.

[2] Arak, T. V. and Zaitsev, A. Yu. *Uniform limit theorems for sums of independent random variables*. In: "Proceedings of the Steklov Institute of Mathematics, Vol. 174", American Mathematical Society, Providence, RI, 1988.

[3] Gnedenko, B. V. and Kolmogorov, A. N. *Limit distributions for sums of independent random variables*. Addison-Wesley, Cambridge, 1954.

[4] Moore, R. E. *The automatic analysis and control of error in digital computation based on the use of interval numbers*. In: Rall, L. B. (ed.) "Proceedings of the Advanced Seminar", Math. Research Center, Wiley, N.Y., 1964, pp. 61–130.

[5] Novitskii, P. V. and Zograph, I. A. *Estimating the measurement errors*. Energoatomizdat, Leningrad, 1991 (in Russian).

[6] Orlov, A. I. *How often are the observations normal?* Industrial Laboratory **57** (7) (1991), pp. 770–772.

[7] Sprecher, D. A. *Elements of real analysis*. Dover, N.Y., 1987.

[8] Wadsworth Jr, H. M. (ed.) *Handbook of statistical methods for engineers and scientists*. McGraw-Hill Publishing Co., N.Y., 1990.

Received: May 22, 1993
Revised version: January 5, 1994

Computer Science Department
University of Texas at El Paso
El Paso, TX 79968
USA