

---

# Trade-Off Between Sample Size and Accuracy: Case of Static Measurements under Interval Uncertainty

Hung T. Nguyen and Vladik Kreinovich

<sup>1</sup> New Mexico State University, Las Cruces, NM 88003, USA [hunguyen@nmsu.edu](mailto:hunguyen@nmsu.edu)

<sup>2</sup> University of Texas at El Paso, El Paso, TX 79968, USA [vladik@utep.edu](mailto:vladik@utep.edu)

**Summary.** In many practical situations, we are not satisfied with the accuracy of the existing measurements. There are two possible ways to improve the measurement accuracy:

- first, instead of a *single* measurement, we can make *repeated* measurements; the additional information coming from these additional measurements can improve the accuracy of the result of this series of measurements;
- second, we can replace the *current* measuring instrument with a *more accurate* one; correspondingly, we can use a more accurate (and more expensive) measurement procedure provided by a measuring lab – e.g., a procedure that includes the use of a higher quality reagent.

In general, we can combine these two ways, and make *repeated* measurements with a *more accurate* measuring instrument. What is the appropriate trade-off between sample size and accuracy? This is the general problem that we address in this paper.

## 1 General formulation of the problem

**We often need more accurate measurement procedures.** Measurements are never 100% accurate, there is always a measurement inaccuracy.

Manufacturers of a measuring instrument usually provide the information about the accuracy of the corresponding measurements. In some practical situations, however, we want to know the value of the measured quantity with the accuracy which is higher than the guaranteed accuracy of a single measurement.

*Comment.* Measurements are provided either by a *measuring instrument* or, in situations like measuring level of pollutants in a given water sample, by a *measuring lab*. Most problems related to measurement accuracy are the same, whether we have an automatic device (measuring instrument) or operator-supervised procedure (measuring lab). In view of this similarity, in the following text, we will consider the term “measuring instrument” in the general

sense, so that the measuring lab is viewed as a particular case of such (general) measuring instrument.

**Two ways to improve the measurement accuracy: increasing sample size and improving accuracy.** There are two possible ways to improve the measurement accuracy:

- first, instead of a *single* measurement, we can make *repeated* measurements; the additional information coming from these additional measurements can improve the accuracy of the result of this series of measurements;
- second, we can replace the *current* measuring instrument with a *more accurate* one; correspondingly, we can use a more accurate (and more expensive) measurement procedure provided by a measuring lab – e.g., the procedure that includes the use of a higher quality reagent.

In general, we can combine these two ways, and make *repeated* measurements with a *more accurate* measuring instrument.

**Problem: finding the best trade-off between sample size and accuracy.** What guidance shall we give to an engineer in this situation? Shall she make repeated measurements with the original instrument? shall she instead purchase a more accurate measuring instrument and make repeated measurements with this new instrument? How more accurate? how many measurement should we perform? In other words, what is the appropriate trade-off between sample size and accuracy?

This is the general problem that we address in this paper.

## 2 In different practical situations, this general problem can take different forms

There are two different situations which, crudely speaking, correspond to engineering and to science.

In most practical situations – in *engineering*, *ecology*, etc. – we know what accuracy we want to achieve. In *engineering*, this accuracy comes, e.g., from the tolerance with which we need to guarantee some parameters of the manufactured object. To make sure that these parameters fit into the tolerance intervals, we must measure them with the accuracy that is as good as the tolerance. For example, if we want to guarantee, e.g., the resistance of a certain wire does not deviate from its nominal value by more than 3%, then we must measure this resistance with an accuracy of at least 3% (or better).

In *ecological* measurements, we want to make sure that the measured quantity does not exceed the required limit. For example, if we want to guarantee that the concentration of a pollutant does not exceed 0.1 units, then we must be able to measure this concentration with an accuracy somewhat higher than 0.1. In such situations, our objective is to minimize the cost of achieving this accuracy.

In *science*, we often face a different objective:

- we have a certain amount of funding allocated for measuring the value of a certain quantity;
- within the given funding limits, we would like to determine the value of the measured quantity as accurately as possible.

In other words:

- In engineering situations, we have a fixed accuracy, and we want to minimize the measurement cost.
- In scientific situations, we have a fixed cost, and we want to maximally improve the measurement accuracy.

### 3 A realistic formulation of the trade-off problem

**Traditional engineering approach.** The traditional engineering approach to solving the above problem is based on the following assumptions – often made when processing uncertainty in engineering:

- that all the measurement errors are normally (Gaussian) distributed known standard deviations  $\sigma$ ;
- that the measurement errors corresponding to different measurement are independent random variables; and
- that the mean value  $\Delta_s$  of the measurement error is 0.

Under these assumptions, if we repeat a measurement  $n$  times and compute the arithmetic average of  $n$  results, then this average approximates the actual value with a standard deviation  $\frac{\sigma}{\sqrt{n}}$ . So, under the above assumptions, by selecting appropriate large number of iterations  $n$ , we can get make measurement errors as small as we want.

This approach – and more general statistical approach – has been actively used in many applications to science in engineering problems; see, e.g., [1, 2, 6, 8].

**Limitations of the traditional approach.** In practice, the distributions are often Gaussian and independent; however, the mean (= systematic error)  $\Delta_s$  is not necessarily 0. Let us show this if we do not take systematic error into account, we will underestimate the resulting measurement inaccuracy.

Indeed, suppose that we have a measuring instrument about which we know that its measurement error cannot exceed 0.1:  $|\Delta x| \leq 0.1$ . This means, e.g., that if, as a result of the measurement, we got the value  $\tilde{x} = 1.0$ , then the actual (unknown) value  $x (= \tilde{x} - \Delta x)$  of the measured quantity can take any value from the interval  $[1.0 - 0.1, 1.0 + 0.1] = [0.9, 1.1]$ .

If the mean of the measurement error (i.e., the systematic error component) is 0, then we can repeat the measurement many times and, as a result, get more and more accurate estimates of  $x$ . However, if – as is often

the case – we do not have any information about the systematic error, it is quite possible that the systematic error is actually equal to 0.07 (and the random error is negligible in comparison with this systematic error). In this case, the measured value 1.0 means that the actual value of the measured quantity was  $x = 1.0 - 0.07 = 0.93$ . In this case, we can repeat the measurement many times, and every time, the measurement result will be equal to  $\approx x + \Delta_s = 0.93 + 0.07 = 1.0$ . The average of these values will still be approximately equal to 1.0 – so, no matter how many times we repeat the measurement, we will get the exact same measurement error 0.07.

In other words, when we are looking for a trade-off between sample size and accuracy, the traditional engineering assumptions can result in misleading conclusions.

**A more realistic description of measurement errors.** We do not know the actual value of the systematic error  $\Delta_s$  – if we knew this value, we could simply re-calibrate the measuring instrument and thus eliminate this systematic error.

What we do know are the bounds on the systematic error. Specifically, in measurement standards (see, e.g., [7]), we are usually provided with the upper bound  $\Delta$  on the systematic error – i.e., with a value  $\Delta$  for which  $|\Delta_s| \leq \Delta$ . In other words, the only information that we have about the systematic error  $\Delta_s$  is that it belongs to the *interval*  $[-\Delta, \Delta]$ .

**Resulting formulas for the measurement accuracy.** Under these assumptions, what is the guaranteed accuracy of a single measurement made by the measuring instrument?

Although formally, a normally distributed random variable can take any value from  $-\infty$  to  $+\infty$ , in reality, the probability of value which are too far away from the average is practically negligible. In practice, it is usually assumed that the values which differ from the average  $a$  by more than  $k_0 \cdot \sigma$  are impossible – where the value  $k_0$  is determined by how confident we want to be:

- 95% confidence corresponds to  $k_0 = 2$ ,
- 99.9% corresponds to  $k_0 = 3$ , and
- confidence 100% –  $10^{-6}$ % corresponds to  $k_0 = 6$ .

Thus, with selected confidence, we know that the measurement error is between  $\Delta_s - k_0 \cdot \sigma$  and  $\Delta_s + k_0 \cdot \sigma$ . Since the systematic error  $\Delta_s$  can take any value from  $-\Delta$  to  $+\Delta$ , the smallest possible value of the overall error is  $-\Delta - k_0 \cdot \sigma$ , and the largest possible value of the overall error is  $\Delta + k_0 \cdot \sigma$ .

Thus, for a measuring instrument with a standard deviation  $\sigma$  of the random error component and an upper bound  $\Delta$  on the systematic error component, the overall error is bounded by the value  $\Delta + k_0 \cdot \sigma$ , where the value  $k_0$  is determined by the desired confidence level.

**Resulting formulas for the accuracy of a repeated measurement.** When we repeat the same measurement  $n$  times and take the average of  $n$

measurement results, the systematic error remains the same, while the standard deviation of the random error decreases  $\sqrt{n}$  times. Thus, after  $n$  measurements, the overall error is bounded by the value  $\Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}}$ .

So, we arrive at the following formulation of the trade-off problem.

**Trade-off problem for engineering.** In the situation when we know the overall accuracy  $\Delta_0$ , and we want to minimize the cost of the resulting measurement, the trade-off problem takes the following form:

$$\text{Minimize } n \cdot F(\Delta, \sigma) \text{ under the constraint } \Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} \leq \Delta_0, \quad (1)$$

where  $F(\Delta, \sigma)$  is the cost of a single measurement performed by a measuring instrument whose systematic error is bounded by  $\Delta$  and whose random error has a standard deviation  $\sigma$ .

**Trade-off problem for science.** In the situation when we are given the limit  $F_0$  on the cost, and the problem is to achieve the highest possible accuracy within this cost, we arrive at the following problem

$$\text{Minimize } \Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} \text{ under the constraint } n \cdot F(\Delta, \sigma) \leq F_0. \quad (2)$$

## 4 Solving the trade-off problem in the general case

*Mathematical comment.* The number of measurement  $n$  is a discrete variable. In general, optimization with respect to discrete variables requires much more computations than continuous optimization (see, e.g., [4]). Since our formulation is approximate anyway, we will treat  $n$  as a real-valued variable – with the idea that in a practical implementation, we should take, as the actual sample size, the closest integer to the corresponding real number solution  $n_{\text{opt}}$ .

**Towards resulting formulas.** For both constraint optimization problems, the Lagrange multiplier method leads to the following unconstrained optimization problem:

$$n \cdot F(\Delta, \sigma) + \lambda \cdot \left( \Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} - \Delta_0 \right) \rightarrow \min_{\Delta, \sigma, n}, \quad (3)$$

where  $\lambda$  can be determined by one of the formulas

$$\Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} = \Delta_0, \quad n \cdot F(\Delta, \sigma) = F_0. \quad (4)$$

Equating the derivatives of the objective function (with respect to the unknowns  $\Delta$ ,  $\sigma$ , and  $n$ ) to 0, we conclude that

$$n \cdot \frac{\partial F}{\partial \Delta} + \lambda = 0; \quad n \cdot \frac{\partial F}{\partial \sigma} + \lambda \cdot \frac{k_0}{\sqrt{n}} = 0; \quad F - \frac{1}{2} \cdot \lambda \cdot k_0 \cdot \frac{\sigma}{n^{3/2}} = 0. \quad (5)$$

Substituting the expression for  $\lambda$  from the first equation into the second one, we conclude that

$$n = k_0^2 \cdot \frac{(\partial F/\partial \Delta)^2}{(\partial F/\partial \sigma)^2}. \quad (6)$$

Substituting these expression into the other equations from (5) and into the equations (4), we get the following non-linear equations with two unknowns  $\Delta$  and  $\sigma$ :

$$F + \frac{1}{2} \cdot \sigma \cdot \frac{\partial F}{\partial \sigma} = 0; \quad (7)$$

$$\Delta + \frac{\sigma \cdot (\partial F/\partial \sigma)}{\partial F/\partial \Delta} = \Delta_0; \quad k_0^2 \cdot \frac{(\partial F/\partial \Delta)^2}{(\partial F/\partial \sigma)^2} \cdot F = F_0. \quad (8)$$

So, we arrive at the following algorithm:

**General formulas: results.** For each of the optimization problems (1) and (2), to find the optimal accuracy values  $\Delta$  and  $\sigma$  and the optimal sample size  $n$ , we do the following:

- First, we determine the optimal accuracy, i.e., the optimal values of  $\Delta$  and  $\sigma$ , by solving a system of two non-linear equations with two unknowns  $\Delta$  and  $\sigma$ : the equation (7) and one of the equations (8) (depending on what problem we are solving).
- After that, we determine the optimal sample size  $n$  by using the formula (6).

**For practical engineering problems, we need more explicit and easy-to-use recommendations.** The above formulas provide a general theoretical solution to the trade-off problem, but to use them in practice, we need more easy-to-use recommendations. In practice, however, we do not have the explicit formula  $F(\Delta, \sigma)$  that determines how the cost of the measurement depends on its accuracy. Therefore, to make our recommendations more practically useful, we must also provide some guidance on how to determine this dependence – and then use the recommended dependence to simplify the above recommendations.

## 5 How Does the Cost of a Measurement Depend on Its Accuracy?

**Two characteristics of uncertainty:  $\Delta$  and  $\sigma$ .** In our description, we use two parameters to characterize the measurement's accuracy: the upper bound  $\Delta$  on the systematic error component and the standard deviation  $\sigma$  of the random error component.

**It is difficult to describe how the cost of a measurement depends on  $\sigma$ .** The standard deviation  $\sigma$  is determined by the noise level, so decreasing  $\sigma$

requires a serious re-design of the measuring instrument. For example, to get a standard measuring instrument, one thing designers usually do is place the instrument in liquid helium so as to eliminate the thermal noise as much as possible; another idea is to place the measuring instrument into a metal cage, to eliminate the effect of the outside electromagnetic fields on the measuring instrument's electronics.

Once we have eliminated the obvious sources of noise, eliminating a new source of noise is a creative problem, requiring a lot of ingenuity, and it is difficult to estimate how the cost of such decrease depends on  $\sigma$ .

**The inability to easily describe the dependence of cost on  $\sigma$  may not be that crucial.** The inability to easily handle the characteristic  $\sigma$  of the random error component may not be so bad because, as we have mentioned, the random error component is the one that can be drastically decreased by increasing the sample size – in full accordance with the traditionally used simplifying engineering assumptions about uncertainty.

As we have mentioned, in terms of decreasing the overall accuracy, it is much more important to decrease the systematic error component, i.e., to decrease the value  $\Delta$ . Let us therefore analyze how the cost of a measurement depends on  $\Delta$ .

**How we can reduce  $\Delta$ : reminder.** As we have mentioned, we can decrease the characteristic  $\Delta$  of the systematic error component by calibrating our measuring instrument against the standard one.

After  $N$  repeated measurements, we get a systematic error  $\Delta_s$  whose standard deviation is  $\approx \sigma/\sqrt{N}$  (and whose distribution, due to the Central Limit Theorem, is close to Gaussian). Thus, with the same confidence level as we use to bound the overall measurement error, we can conclude that  $|\Delta_s| \leq k_0 \cdot \sigma/\sqrt{N}$ .

**Calibration is not a one-time procedure.** To properly take calibration into account, it is important to recall that calibration is not a one-time procedure. Indeed, most devices deteriorate with time. In particular, measuring instruments, if not periodically maintained, become less and less accurate. Because of this, in measurement practices, calibration is not a one-time procedure, it needs to be done periodically.

How frequently do we need to calibrate a device? The change of  $\Delta_s$  with time  $t$  is slow and smooth. A smooth dependence can be represented by a Taylor series  $\Delta_s(t) = \Delta_s(0) + k \cdot t + c \cdot t^2 + \dots$ . In the first approximation, we can restrict ourselves to the main – linear – term (linear trend) in this expansion, and thus, in effect, assume that the change of  $\Delta_s$  with time  $t$  is linear.

Thus, if by calibrating the instrument, we guaranteed that  $|\Delta_s| \leq \Delta$ , then after time  $t$ , we can only guarantee that  $|\Delta_s| + k \cdot t \leq \Delta$ . Once the upper bound on  $\Delta_s$  reaches the level that we want not to exceed, this means that a new calibration is in order. Usually (see, e.g., [7]), to guarantee the bound

$\Delta$  throughout the entire calibration cycle, we, e.g., initially calibrate it to be below  $\Delta/2$ , and then re-calibrate at a time  $t_0$  when  $\Delta/2 + k \cdot t_0 = \Delta$ . In such a situation, the time  $t_0$  between calibrations is equal to  $t_0 = \Delta/(2 \cdot k)$ .

**How the calibration-based reduction procedure translates into the cost of a measurement: the main case.** As we have just mentioned, the way to decrease  $\Delta$  is to calibrate the measuring instrument. Thus, the resulting additional cost of a measurement comes from the cost of this calibration (spread over all the measurement performed between calibrations).

Each calibration procedure consists of two stages:

- first, we transport the measuring instrument to the location of a standard – e.g., to the National Institute of Standard and Technology (NIST) or one of the regional standardization centers – and set up the comparison measurements by the tested and the standard instruments;
- second, the we perform the measurements themselves.

Correspondingly, the cost of calibration can be estimated as the sum of the costs of there two stages.

The standard measuring instrument is usually a very expensive operation. So, setting it up for comparison with different measuring instruments requires a lot of time and a lot of adjustment. Once the set-up is done, the second stage is fast and automatic – and therefore not that expensive.

As a result, usually, the cost of the first stage is the dominating factor. So, we can reasonably assume that the cost of the calibration is just the cost of the set-up – i.e., the cost of the first stage of the calibration procedure.

By definition, the set-up does not depend on how many times  $N$  we perform the comparison measurements. Thus, in the first approximation, we can simply assume that each calibration requires a flat rate  $f_0$ .

The interval between time calibrations is  $t_0 = \Delta/(2 \cdot k)$ , then during a fixed period of time  $T_0$  (e.g., 10 years), we need

$$\frac{T_0}{t_0} = \frac{T_0}{\Delta/(2 \cdot k)} = \frac{2 \cdot k \cdot T_0}{\Delta}$$

calibrations. Multiplying this number by the cost  $f_0$  of each calibration, we get the overall cost of all the calibrations performed during the fixed time  $T_0$  as  $\frac{2 \cdot k \cdot T_0 \cdot f_0}{\Delta}$ . Finally, dividing this cost by the estimated number  $N_0$  of measurements performed during the period of time  $T_0$ , we estimate the cost  $F(\Delta)$  of an individual measurement as

$$F(\Delta) = \frac{c}{\Delta}, \tag{9}$$

where we denoted

$$c \stackrel{\text{def}}{=} \frac{2 \cdot k \cdot T_0 \cdot f_0}{N_0}. \tag{10}$$



*Comment.* The above formula was first described, in a somewhat simplified form, in [3].

**This formula is in good accordance with chemistry-related measurements.** It is worth mentioning that the dependence  $c \sim 1/\Delta$  also occurs in measurements related to chemical analysis. Indeed, in these measurements, the accuracy of the measurement result is largely determined by the quality of the reagents, i.e., mainly, by the concentration level  $\delta$  of the unwanted chemicals (pollutants) in a reagent mix. Specifically, the maximum possible error  $\Delta$  is proportional to this concentration  $\delta$ , i.e.,  $\Delta \approx c_0 \cdot \delta$ .

According to [9], the cost of reducing pollutants to a level  $\delta$  is proportional to  $1/\delta$ . Since the accuracy  $\Delta$  is proportional to  $\delta$ , the dependence of the cost of the accuracy is also inverse proportional to  $\Delta$ , i.e.,  $F(\Delta) = c/\Delta$  for some constant  $c$ .

**This formula is in good accordance with actual prices of different measurements.** This dependence is in good agreement by the experimental data on the cost of measurements of chemical-related measurements. For example, in a typical pollution measurement, a measurement with the 25% accuracy costs  $\approx$  \$200, while if we want to get 7% accuracy, then we have to use a better reagent grade in our measurements which costs between \$500 and \$1,000. Here, the 3–4 times increase in accuracy (i.e., 3–4 times decrease in measurement error) leads to approximately the same (4–5) times increase in cost – which is indeed in good accordance with the dependence  $F(\Delta) \approx c/\Delta$ .

**How the calibration-based reduction procedure translates into the cost of a measurement: cases of more accurate measurements.** In deriving the formula  $F(\Delta) \approx c/\Delta$ , we assumed that the cost of actually performing the measurements with the standard instrument is much smaller than the cost of setting up the calibration experiment. This is a reasonable assumption if the overall number of calibration-related measurement  $N$  is not too large.

How many measurement do we need? After  $N$  measurements, we get the accuracy  $\Delta = k_0 \cdot \sigma/\sqrt{N}$ . Thus, for a measuring instrument with standard deviation  $\sigma$ , if we want to achieve the systematic error level  $\Delta$ , we must use

$$N = k_0 \cdot \frac{\sigma^2}{\Delta^2} \quad (11)$$

measurements.

So, if we want to use the calibration procedure to achieve higher and higher accuracy – i.e., smaller and smaller values of  $\Delta$  – we need to perform more and more calibration-related measurements. For large  $N$ , the duration of the calibration-related measurements exceeds the duration of the set-up. Since the most expensive part of the calibration procedure is the use of the standard measuring instrument, the cost of this procedure is proportional to the overall time during which we use this instrument. When  $N$  is large, this time is roughly proportional to  $N$ .

In this case, instead of a flat fee  $f_0$ , the cost of each calibration becomes proportional to  $N$ , i.e., equal to  $f_1 \cdot N$ , where  $f_1$  is the cost per time of using the standard measuring instrument multiplied by the time of each calibration measurement. Due to the formula (11), the resulting cost of each calibration is equal to  $f_1 \cdot k_0 \cdot \frac{\sigma^2}{\Delta^2}$ . To get the cost of a single measurement, we must multiply this cost by the number of calibrations  $\frac{2 \cdot k \cdot T_0}{\Delta}$  required during the time period  $T_0$ , and then divide by the typical number of measurements performed during this period of time. As a result, the cost of a single measurement becomes  $\frac{\text{const}}{\Delta^3}$ .

**The cost of measurements beyond calibration: general discussion.** In many scientific cutting-edge experiments, we want to achieve higher accuracy than was possible before. In such situations, we cannot simply use the existing standard measuring instrument to calibrate the new one, because we want to achieve the accuracy that no standard measuring instrument has achieved earlier.

In this case, how we can increase the accuracy depends on the specific quantity that we want to measure.

**The cost of measurements beyond calibration: example.** For example, in radioastrometry – the art of determining the locations of celestial objects from radioastronomical observation – the accuracy of a measurement by a single radio telescope is  $\Delta \approx \lambda/D$ , where  $\lambda$  is the wavelength of the radio-waves on which we are observing the source, and  $D$  is the diameter of the telescope; see, e.g., [10]. For a telescope of a linear size  $D$ , just the amount of material is proportional to its volume, i.e., to  $D^3$ ; the cost  $F$  of designing a telescope is even higher – it is proportional to  $D^4$ . Since  $D \approx \text{const}/\Delta$ , in this case, we have  $F(\Delta) \approx \text{const}/\Delta^4$ .

**The cost of measurements beyond calibration: power laws.** The above dependence is a particular case of the *power law*  $F(\Delta) \approx \text{const}/\Delta^\alpha$ . Power laws are, actually, rather typical descriptions of the dependence of the cost of an individual measurement on its accuracy.

In [5], we explain why in the general case, power laws are indeed reasonable approximation: crudely speaking, in the absence of a preferred value of the measured quantity, it is reasonable to assume that the dependence does not change if we change the measuring unit (i.e., that it is scale invariant), and power laws are the only scale-invariant dependencies.

*Comment.* The same arguments about scale invariance apply when we try to find out how the cost of a measurement depends on the standard deviation. So, it is reasonable to assume that this dependence is also described by a power law  $F(\sigma) \approx \text{const}/\sigma^\beta$  for some constant  $\beta$ .

## 6 Trade-off between accuracy and sample size in different cost models

Let us plug in the above cost models into the above general solution for the tradeoff problem and find out what is the optimal trade-off between accuracy and sample size in the above cost models.

Since the above cost models only describe the dependence of the cost of  $\Delta$  and  $n$ , we will assume that the characteristic  $\sigma$  of the random error component is fixed, so we can only select the accuracy characteristic  $\Delta$  and the sample size  $n$ .

**Basic cost model: engineering situation.** Let us start with the basic cost model, according to which  $F(\Delta) = c/\Delta$ . Within this model, we can explicitly solve the above system of equations. As a result, for the engineering situation, we conclude that

$$n_{\text{opt}} = \frac{9 \cdot k_0^2 \cdot \sigma^2}{4 \cdot \Delta_0^2}; \quad \Delta_{\text{opt}} = \frac{1}{3} \cdot \Delta_0. \quad (12)$$

**Observation.** In this case, the overall error bound  $\Delta_0$  is the sum of the bounds coming from two error components:

- the bound  $\Delta_0$  that comes from the systematic error component, and
- the bound  $k_0 \cdot \frac{\sigma}{\sqrt{n}}$  that comes from the random error component.

In the optimal trade-off, the first component is equal to 1/3 of the overall error bound, and therefore, the second component is equal to 2/3 of the overall error bound. As a result, we conclude that when the error comes from several error components, in the optimal trade-off, these error components are of approximately the same size.

**Heuristic consequence of this observation.** As a result of this qualitative idea, it is reasonable to use the following heuristic rule when looking for a good (not necessarily optimal) trade-off: split the overall error into equal parts.

In the above example, this would mean taking  $\Delta = (1/2) \cdot \Delta_0$  (and, correspondingly,  $k_0 \cdot \frac{\sigma}{\sqrt{n}} = (1/2) \cdot \Delta_0$ ) instead of the optimal value  $\Delta = (1/3) \cdot \Delta_0$ .

How non-optimal is this heuristic solution?

For the optimal solution  $\Delta = (1/3) \cdot \Delta_0$ , the resulting value of the objective function (1) (representing the overall measurement cost) is  $\frac{27}{4} \cdot \frac{k_0^2 \cdot \sigma^2 \cdot c}{\Delta_0^2}$ ,

while for  $\Delta = (1/2) \cdot \Delta_0$ , the cost is  $8 \cdot \frac{k_0^2 \cdot \sigma^2 \cdot c}{\Delta_0^2}$  – only  $\approx 20\%$  larger.

If we take into account that all our models are approximate, this means that the heuristic trade-off solution is practically as good as the optimal one.

**Basic cost model: science situation.** In the science situation (2), we get

$$n_{\text{opt}} = \left( \frac{F_0 \cdot k_0 \cdot \sigma}{2 \cdot c} \right)^{2/3}; \quad \Delta_{\text{opt}} = \frac{n_{\text{opt}} \cdot c}{F_0}. \quad (13)$$

**Cases of more accurate and cutting-edge measurements.** When  $F(\Delta) = c/\Delta^\alpha$ , for the engineering case, we get

$$n_{\text{opt}} = \frac{(\alpha + 2)^2 \cdot k_0^2 \cdot \sigma^2}{4 \cdot \Delta_0^2}; \quad \Delta_0 = \frac{\alpha}{2 + \alpha} \cdot \Delta_0.$$

For the science case,

$$n_{\text{opt}} = \left( \frac{F_0}{c} \right)^{2/(2+\alpha)} \cdot \left( \frac{k_0 \cdot \alpha}{2} \right)^{(2\alpha)/(2+\alpha)}; \quad \Delta_{\text{opt}} = \frac{\alpha}{2} \cdot k_0 \cdot \frac{\sigma}{\sqrt{n_{\text{opt}}}}.$$

In both cases, the error bound coming from the systematic error component is approximately equal to the error bound coming from the random error component.

## Conclusion

In many practical situations, we are not satisfied with the accuracy of the existing measurements. There are two possible ways to improve the measurement accuracy. First, instead of a *single* measurement, we can make *repeated* measurements; the additional information coming from these additional measurements can improve the accuracy of the result of this series of measurements. Second, we can replace the *current* measuring instrument with a *more accurate* one; correspondingly, we can use a more accurate (and more expensive) measurement procedure provided by a measuring lab – e.g., a procedure that includes the use of a higher quality reagent. In general, we can combine these two ways, and make *repeated* measurements with a *more accurate* measuring instrument.

What is the appropriate trade-off between sample size and accuracy? Traditional engineering approach to this problem assumes that we know the exact probability distribution of all the measurement errors. In many practical situations, however, we do not know the exact distributions. For example, we often only know the upper bound on the corresponding measurement (or estimation) error; in this case, after the measurements, we only know the interval of possible values of the quantity of interest. In the first part of this paper, we show in such situations, traditional engineering approach can sometimes be misleading, so for interval uncertainty, new techniques are needed. In the remainder of this paper, we describe proper techniques for achieving optimal trade-off between sample size and accuracy under interval uncertainty.

**Acknowledgments.** This work was supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, by the Max Planck Institut für Mathematik, and

by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08.

The authors are very thankful to the anonymous referees for valuable suggestions.

## References

1. J. Eng, “Sample Size Estimation: How Many Individuals Should Be Studied?”, *Radiology*, 2003, Vol. 227, pp. 309–313.
2. J. J. Hox, *Multilevel Analysis: Techniques and Applications*, Lawrence Erlbaum Associates, 2002.
3. V. Kreinovich, *How to compute the price of a measuring instrument?*, Leningrad Center for New Information Technology “Informatika”, Technical Report, Leningrad, 1989 (in Russian).
4. V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997.
5. H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.
6. N. A. Obuchowski and D. K. Mcclish, “Sample Size Determination For Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices”, *Statistics in Medicine*, 1997, Vol. 16, No. 13, pp. 1529–1542.
7. S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.
8. A. K. Skidmore, “Accuracy assessment of spatial information”, In: *Spatial Statistics for Remote Sensing*, Springer Netherlands, 2002, pp. 197–209.
9. D. Stevens, “Analysis of biological systems”, *Proceedings of the NIH MARC Winter Institute on Undergraduate Education in Biology*, Santa Cruz, California, January 7–11, 2005.
10. G. L. Verschuur and K. I. Kellerman, “Galactic and Extragalactic Radio Astronomy”, Springer-Verlag, N.Y., 1988.