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

Hung T. Nguyen[1] and Vladik Kreinovich[2]

[1]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
hunguyen@nmsu.edu

[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

1. Outline

- 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:
  - *repeated* measurements;
  - *more accurate* measurements.
- In general, we can combine these two ways:
  - 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 talk.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

2. General formulation of the problem

- Measurements are never 100% accurate, there is always a measurement inaccuracy.
- Manufacturers usually provide the information about the accuracy of the corresponding measurements.
- In many practical situations, we want the result with higher accuracy.
- Two ways to improve the measurement accuracy:
  - make *repeated* measurements (i.e., increase sample size);
  - use a *more accurate* measuring instrument.
- We can also combine these two approaches.
- Problem: finding the best trade-off between sample size and accuracy.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

3. In different practical situations, this general problem can take different forms

- In most practical situations – in *engineering*, *ecology*, etc. – we know what accuracy we want to achieve.
- In *science*, we often face a different objective:
  - we have a certain amount of funding allocated for measuring the value of a certain quantity;
  - 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 science, we have a fixed cost, and we want to maximally improve the measurement accuracy.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

4. Trade-off problem: traditional engineering approach

- *Assumptions:*
  - all the measurement errors are normally (Gaussian) distributed known standard deviations σ;
  - the measurement errors corresponding to different measurement are independent random variables; and
  - the mean value Δ_s of the measurement error is 0.
- *Result:* the average of n measurements has st. dev. $\dfrac{\sigma}{\sqrt{n}}$.
- *Conclusion:* measurement errors as small as we want.
- *In practice:* the mean (= systematic error) $\Delta_s \neq 0$.
- *Hence:* we underestimate measurement inaccuracy.
- *Example:* $\sigma \approx 0$, $\Delta_s = 0.1$, $x = 1.0$, $\widetilde{x} = 1.1$.
- *Conclusion:* averaging does not improve accuracy.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

5. A realistic formulation of the trade-off problem

- *Usually:* we know the upper bound Δ on the systematic error Δ_s: $|\Delta_s| \le \Delta$.
- *Conclusion:* the largest possible value of the overall error is $\Delta + k_0 \cdot \sigma$, where:
  - 95% confidence corresponds to $k_0 = 2$,
  - 99.9% corresponds to $k_0 = 3$.
- After n measurements, we get $\Delta + k_0 \cdot \dfrac{\sigma}{\sqrt{n}}$.
- Let $F(\Delta, \sigma)$ denote the cost of a single measurement.
- *Trade-off problem for engineering:*

  Min. $n \cdot F(\Delta, \sigma)$ under the constraint $\Delta + k_0 \cdot \dfrac{\sigma}{\sqrt{n}} \le \Delta_0$.
- *Trade-off problem for science:*

  Min. $\Delta + k_0 \cdot \dfrac{\sigma}{\sqrt{n}}$ under the constraint $n \cdot F(\Delta, \sigma) \le F_0$.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

6. Solving the trade-off problem: general case

- *Discussion:*
  - n is a discrete variable;
  - discrete optimization is difficult;
  - the expressions are approximat anyway;
  - so, we treat n as a real-valued variable.
- *Lagrange multiplier method:*

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

where

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

- *Equating derivatives to 0:*

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

General formulation of...
In different practical...
Trade-off problem:...
A realistic formulation...
Solving the trade-off...
General case (cont-d)
How Does the Cost of...
Case of More Accurate...
Trade-off between...
Case of the basic cost...
Cases of more...
Acknowledgments

7. General case (cont-d)

- *Problem:* 4 equations with 4 unknowns λ, n, Δ, and σ.
- *Simplifying:* substituting the expression for λ from the 1st equation, we get the formula

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

- Substituting this expression for n into remaining equations, we get 2 equations with 2 unknowns Δ and σ:

$$F + \frac{1}{2} \cdot \sigma \cdot \frac{\partial F}{\partial \sigma} = 0;$$

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

- After we solve this system, we can use the above expression to find n.
- *Problem:* in practice, we do not have an explicit formula for the measurement cost $F(\Delta, \sigma)$.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

8. How Does the Cost of a Measurement Depend on Its Accuracy?

- *How we can reduce* Δ*:* by calibrating our measuring instrument against the standard one.
- After N repeated measurements, we conclude that
$$|\Delta_s| \le k_0 \cdot \sigma/\sqrt{N}.$$
- *Additional problem:* devices deteriorate with time: after time t, we get $\Delta_s(t) = \Delta_s(0) + k \cdot t + o(t)$.
- *Conclusion:* we take $\Delta_s(0) = \Delta/2$, and re-calibrate when $k \cdot t_0 = \Delta/2$, i.e., after $t_0 = \Delta/(2k)$.
- *Assumption:* each calibration costs the same.
- During a fixed time T_0, we need $T_0/t_0 \sim 1/\Delta$ calibrations, so the cost is $F(\Delta) \sim 1/\Delta$.
- *Empirical analysis:* cost is indeed $\sim 1/\Delta$.

General formulation of...
In different practical...
Trade-off problem:...
A realistic formulation...
Solving the trade-off...
General case (cont-d)
How Does the Cost of...
Case of More Accurate...
Trade-off between...
Case of the basic cost...
Cases of more...
Acknowledgments

9. Case of More Accurate Measurements

- *Reminder:* after N repeated measurements, we conclude that $|\Delta_s| \le k_0 \cdot \sigma/\sqrt{N}$.
- *Conclusion:* to get desired accuracy Δ, we need
$$N \sim \frac{1}{\Delta^2}.$$
- *Reminder:* the necessary number of calibrations is
$$\frac{T_0}{t_0} \sim \frac{1}{\Delta}.$$
- *Conclusion:* the total cost of all $\sim 1/\Delta$ calibrations is
$$F(\Delta) \sim \frac{1}{\Delta^3}.$$
- *General case:* scale invariance leads to $F(\Delta) \approx \text{const}/\Delta^\alpha$ for some constant α.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

10. Trade-off between accuracy and sample size: case of the basic cost model

- *Basic cost model – reminder:* $F(\Delta) = c/\Delta$.
- *Engineering situation – reminder:*

  Min. $n \cdot F(\Delta, \sigma)$ under the constraint $\Delta + k_0 \cdot \dfrac{\sigma}{\sqrt{n}} \le \Delta_0$.
- *Engineering situation – results:*

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

- *Reminder:* the overall error bound Δ_0 is the sum of two error components:
  - the systematic error component bounded by Δ, and
  - random error component bounded by $k_0 \cdot \dfrac{\sigma}{\sqrt{n}}$.
- *Observation:* in the optimal trade-off, these error bounds are of approximately the same size.

11. Case of the basic cost model: cont-d

- *Reminder:* in the optimal trade-off, the systematic and random error components are of $\approx$ the same size.
- *Resulting heuristic:* make sure that all error components are of the same size.
- *Example:* engineering situation under basic cost model.
- *Conclusion:* in this example, the above heuristic is only 20% worse than the optimum.
- *Scientific situation – reminder:*

  Min. $\Delta + k_0 \cdot \dfrac{\sigma}{\sqrt{n}}$ under the constraint $n \cdot F(\Delta, \sigma) \le F_0$.
- *Scientific situation – result:*

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

General formulation of...
In different practical...
Trade-off problem:...
A realistic formulation...
Solving the trade-off...
General case (cont-d)
How Does the Cost of...
Case of More Accurate...
Trade-off between...
Case of the basic cost...
Cases of more...
Acknowledgments

12. Cases of more accurate and cutting-edge measurements

- *Cost model – reminder:* $F(\Delta)=c/\Delta^\alpha$.
- Min. $n\cdot F(\Delta,\sigma)$ under the constraint $\Delta+k_0\cdot\dfrac{\sigma}{\sqrt{n}}\le\Delta_0$.
- *Engineering situation – results:*

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

- Min. $\Delta+k_0\cdot\dfrac{\sigma}{\sqrt{n}}$ under the constraint $n\cdot F(\Delta,\sigma)\le F_0$.
- *Scientific situation – results:*

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

- *Observation:* in both situations, the systematic error component is $\approx$ the random error component.

General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments

13. Acknowledgments

This work was supported in part by:

- by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08,
- by Texas Department of Transportation contract No. 0-5453,
- by National Science Foundation grants HRD-0734825, EAR-0225670, and EIA-0080940, and
- and by the Max Planck Institut für Mathematik.

Outline
General formulation of . . .
In different practical . . .
Trade-off problem: . . .
A realistic formulation . . .
Solving the trade-off . . .
General case (cont-d)
How Does the Cost of . . .
Case of More Accurate . . .
Trade-off between . . .
Case of the basic cost . . .
Cases of more . . .
Acknowledgments