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

Hung T. Nguyen[1], Olga Kosheleva[2], Vladik Kreinovich[2], and Scot Ferson[3]

[1]New Mexico State University
Las Cruces, NM 88003, USA

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

[3]Applied Biomathematics
Setauket, NY 11733, USA

1. Formulation of the problem

- For dynamic quantities, we may have two different objectives:
  - We may be interested in knowing the *average* value of the measured quantity.
  - We may want to know the actual dependence of the measured quantity on space location and/or time.
- Meteorological example:
  - to study general weather patterns, we need average wind speed;
  - to guide airplanes, we need to know the exact wind speed at different locations.

2. Measuring the average value: case of ideal measuring instrument

- *Case description:* measurement errors are negligible.
- *Details:* we measure the values $x_1, \ldots, x_n$ at n different locations, and estimate
$$E \stackrel{\text{def}}{=} \frac{x_1 + \ldots + x_n}{n}.$$
- *Notation:* let σ_0 be the standard deviation of the difference $x_i - x_0$.
- *Conclusion:* for the average, standard deviation is $\sigma_0/\sqrt{n}$, so error bound is $\Delta = k_0 \cdot \sigma_0/\sqrt{n}$, where:
  - 95% confidence corresponds to $k_0 = 2$,
  - 99.9% corresponds to $k_0 = 3$.
- *Recommendation:* to get error $\le \Delta_0$, we need
$$n \approx \frac{k_0^2 \cdot \sigma_0^2}{\Delta_0^2}.$$

3. Measuring the average value: case of realistic measuring instrument

- *Case description:* we have random error with st. dev. σ and a systematic error Δ_s bounded by Δ: $|\Delta_s| \le \Delta$.
- *Resulting random error component in measuring average:* st. dev. $\sigma_t \stackrel{\rm def}{=} \sqrt{\sigma^2 + \sigma_0^2}$.
- *Resulting overall error bound:*

$$\Delta_0 = \Delta + k_0 \cdot \frac{\sigma_t}{\sqrt{n}}.$$

- *Observation:* situation is similar to static measurements.
- *Conclusion:* we have use the recommendations from the static case.
- *Example:* to achieve accuracy Δ_0 with the minimal cost, take $\Delta \approx (1/3) \cdot \Delta_0$.

4. Measuring the actual dependence: formulation of the problem.

- We are often interested in the actual dependence of a quantity on space and/or time.
- *Possible situations:*
  - A quantity that only depends on time t.
  - *Example:* temperature at a given location.
  - A quantity that only depends on the spatial location $t = (t_1, t_2)$ or $t = (t_1, t_2, t_3)$.
  - *Example:* density inside the Earth.
  - A quantity that depends both on time t_1 and on the spatial location $(t_2, \ldots)$.
  - *Example:* temperature in the atmosphere.
- *General description:* we measure $x(t)$ for different

$$t = (t_1, \ldots, t_d).$$

5. Approximation inaccuracy

- *Fact:* we only measure $x(t)$ at n different values $t^{(1)}, \ldots, t^{(n)}$.
- To get values $x(t)$ for $t \neq t^{(i)}$, we use interpolation.
- *Main assumptions:* we assume that: $x(t)$ is smooth, and we know the bound g on the rate of change:
$$|x(t) - x(t')| \leq g \cdot \|t - t'\|.$$
- *Conclusion:* to minimize the approximation error $|x(t)-x(t^{(i)})|$, we must minimize the distance $\|t-t^{(i)}\|$.
- If each value t is within distance ρ from one of $t^{(i)}$, then n balls centered in $t^{(i)}$ cover the domain of volume V.
- Hence, $V \leq n \cdot c \cdot \rho^d$, and $\rho \geq c \cdot (V/n)^{1/d}$.
- For a grid: $\rho \approx c_1 \cdot (V/n)^{1/d}$.
- *Conclusion:* approximation error is $d \cdot c_1 \cdot (V/n)^{1/d}$.
- *Overall error:* $\Delta + d \cdot c_1 \cdot (V/n)^{1/d}$.

6. Trade-off problems for engineering and science: formulation

- In engineering applications:
  - we know the overall accuracy Δ_0, and
  - we want to minimize the cost of the resulting measurement:

$$\text{Minimize } n\cdot F(\Delta) \to \min_{\Delta,n} \text{ under the constraint } \Delta + \frac{g_0}{n^{1/d}} = \Delta_0.$$

- In scientific applications:
  - we are given the cost F_0, and
  - the problem is to achieve the highest possible accuracy within this cost:

$$\text{Minimize } \Delta + \frac{g_0}{n^{1/d}} \to \min_{\Delta,n} \text{ under the constraint } n\cdot F(\Delta) = F_0.$$

7. Solutions

- *Reminder:* basic cost model $F(\Delta) = c/\Delta$.

Minimize $n \cdot F(\Delta) \to \min\limits_{\Delta, n}$ under the constraint $\Delta + \dfrac{g_0}{n^{1/d}} = \Delta_0$.

- *Solution for engineering situations:*

$$\Delta_{\text{opt}} = \frac{1}{d+1} \cdot \Delta_0; \quad n_{\text{opt}} = \left(\frac{g_0}{\Delta_0} \cdot \frac{d+1}{d}\right)^d.$$

Minimize $\Delta + \dfrac{g_0}{n^{1/d}} \to \min\limits_{\Delta, n}$ under the constraint $n \cdot F(\Delta) = F_0$.

- *Solution for science situations:*

$$n_{\text{opt}} = \left(\frac{F_0}{c} \cdot \frac{g_0}{d}\right)^{d/(d+1)}; \quad \Delta_{\text{opt}} = \frac{n_{\text{opt}} \cdot c}{F_0}.$$

- *Observation:* the optimal trade-off is when both error components are of approximately the same size.

- *Comment:* this is similar to the static case.

8. Case of non-smooth processes

- *Example of a non-smooth process:* Brownian motion:

$$|x(t) - x(t')| \le \|t - t'\|^{1/2}.$$

- *General (fractal) case:* $|x(t) - x(t')| \le \|t - t'\|^{\beta}$.
- For n measurements, distance ρ is $\approx (V/n)^{1/d}$, so approximation error is $\sim (V/n)^{\beta/d}$.
- *Overall error:* $\Delta + \dfrac{g_\beta}{n^{\beta/d}}$, where $g_\beta \stackrel{\text{def}}{=} C \cdot d^{\beta/2} \cdot \dfrac{1}{2^{\beta/d}} \cdot V^{\beta/d}$.
- *Trade-off problems for engineering:*

$$\text{Min } n \cdot F(\Delta) \text{ under the constraint } \Delta + \frac{g_\beta}{n^{\beta/d}} = \Delta_0.$$

- *Trade-off problems for science:*

$$\text{Min } \Delta + \frac{g_\beta}{n^{\beta/d}} \text{ under the constraint } n \cdot F(\Delta) = F_0.$$

- *Observation:* formulas same as in the smooth case, with $d' \stackrel{\text{def}}{=} d/\beta$ instead of d.

9. Case of non-smooth processes: solutions

- *Case:* basic cost model $F(\Delta) = c/\Delta$.
- *Engineering problem – reminder:*

$$\text{Min } n \cdot F(\Delta) \text{ under the constraint } \Delta + \frac{g_\beta}{n^{\beta/d}} = \Delta_0.$$

- *Engineering problem – solution:*

$$\Delta_{\text{opt}} = \frac{\beta}{d+\beta} \cdot \Delta_0; \quad n_{\text{opt}} = \left(\frac{g_\beta}{\Delta_0} \cdot \frac{d+\beta}{d}\right)^d.$$

- Min $\Delta + \dfrac{g_\beta}{n^{\beta/d}}$ under the constraint $n \cdot F(\Delta) = F_0$.
- *Science problem – solution:*

$$n_{\text{opt}} = \left(\frac{F_0}{c} \cdot \frac{g_\beta}{d}\right)^{d/(d+\beta)}; \quad \Delta_{\text{opt}} = \frac{n_{\text{opt}} \cdot c}{F_0}.$$

- *Comment:* in the optimal trade-off, both error components are of approximately the same value.

10. Case of more accurate measuring instruments

- *Reminder:* the cost $F(\Delta)$ of a measurement depends on its accuracy as
$$F(\Delta) = \frac{c}{\Delta^\alpha}.$$
- Once we go beyond the basic cost model $\alpha = 1$, we get $\alpha = 3$.
- Then, as we increase accuracy, we switch to a different value α.
- *Solution:* in the engineering case, the optimal accuracy is
$$\Delta_{\rm opt} = \frac{\alpha}{\alpha + 2} \cdot \Delta_0.$$
- *Example:* for $\alpha = 3$, we have
$$\Delta_{\rm opt} = \frac{3}{5} \cdot \Delta_0.$$

11. Conclusions

- *General heuristic:* the optimal is when error components are approximately of the same size.
- To measure a single quantity, take $\Delta = \frac{1}{3} \cdot \Delta_0$.
- To reconstruct all the values $x(t)$ of a smooth quantity x depending on d parameters, take $\Delta = \frac{1}{d+1} \cdot \Delta_0$.
- To reconstruct all the values $x(t)$ of a non-smooth quantity x depending on d parameters, take $\Delta = \frac{\beta}{d+\beta} \cdot \Delta_0$.
- Here β is the exponent of the power law that describes how the difference $x(t+\Delta t) - x(t)$ changes with $\|\Delta t\|$.
- For more accurate measuring instruments, when $F(\Delta) = \frac{c}{\Delta^3}$, we should take $\Delta = \frac{3}{5} \cdot \Delta_0$.
- In general, if $F(\Delta) = \frac{c}{\Delta^\alpha}$, take $\Delta = \frac{\alpha}{\alpha+2} \cdot \Delta_0$.

12. 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.