LA-UR-20-26330

Approved for public release; distribution is unlimited.

Title: Measurement Uncertainty in One-of-a-kind Experiments

Author(s): Makaruk, Hanna Ewa

Intended for: LANNM IEEE Section virtual technical talks

Issued: 2020-08-17

Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Measurement Uncertainty in One-of-a-kind Experiments

Technical Talk for Los Alamos Northern New Mexico IEEE Section

Hanna Makaruk

7/31/2020

P-23, hanna_m@lanl.gov

Managed by Triad National Security, LLC for the U.S. Department of Energy's NNSA

Abstract

A golden standard in science is to repeat an experimental measurement multiple times and calculate the measured value with experimental uncertainty by following well developed statistical procedures.

For various reasons – cost, technical difficulty, international treaties, ethics of dealing with human or animal subjects, ecology - many important experiments and observations can not be repeated. Astronomy, earthquakes, hurricanes produce data from one-of-a-kind events. When information is not available in any other way, it should not be dismissed as qualitative, anecdotal evidence only. Analyzed in a mathematically rigorous way, it produces quantitative experimental data. Analysis of data from one-of-a-kind event differs from analysis of repeated experiments' data.

For repeated experiments, the experimental error includes a range of true values generated by repetitions of the experiment, and measurement uncertainty caused by detectors. They are independent. Repetitions of any experiment, as similar as achievable, always have built-in differences resulting in a range of the true values rather than in a single value. Measurement uncertainty depends on the measurement system only. Digital measurements have very small uncertainty, frequently smaller than the range of true experimental values resulting from built-in differences in the experiment repetitions.

When data from one-of-a kind experiment are analyzed, only the measurement uncertainty can be reported.

*In theory, there is no difference between
theory and practice.*

In practice, there is a difference.

(engineering proverb)

Main Points

- Independent processes create independent results.
- Why does scientific community so frequently mix them up?
- Data from one-of-a-kind experiment.
- Reliable and not-so-reliable data.
- What kind of uncertainty we report from a one-of-a-kind experiment, and what kind we will never be able to report.
- Appendix- Glossary

Measurement system

Experiment

- Experiment determines **Range of the true experimental value.**
- Measurement system and data analysis method determine **Measurement uncertainty.**

Measurement uncertainty and Range of true experimental values originate from separate sources. They are independent. One contains no information about the other.

Range of true experimental values

Range of true experimental values is a range, in which an experimental value changes, when the same experiment is repeated, if measured by a non-existing “perfect” detector. There are built-in differences in the repetitions of an experiment.

- Small, unavoidable variations in experimental conditions cause a non-zero range of true experimental values.
- The range depends exclusively on the physics of the experiment.
- If no measurement is performed the range of true experimental values still exists.

Simple experiments in classical mechanics have a narrow range of true experimental values

Sun corona (NASA)
Turbulence, a stochastic process, causes a broad range of true experimental values

Experimental measurement with no measurement uncertainty does not exist, yet this uncertainty can be incredibly small.

Measurement uncertainty sources include:

- Finite accuracy and precision of a detector.
- Uncertainty causes in data transmission and storage; for example limited bandwidth of a transmission cable, aging of a film on which an image was recorded.
- Digital image pixelization, finite precision of a computer.
- Approximations in data analysis.

Modern digital detectors, fiber optics transmission, advances in data analysis dramatically reduced *measurement uncertainty*.

“Data Uncertainty” – are we all talking about the same thing?

Why Scientific Community too Frequently Mixes Up Independent Objects?

- Textbook example: experiment repeated and measured a statistically meaningful number of times.
 - Experimental error is calculated as a standard deviation (or similar) from a sufficiently large set of data points.
 - It contains everything – Range of the true experimental values and Measurement uncertainty originating from all sources.
 - Problem – repeating an experiment is not always possible.
- Early physics – seventeenth, eighteenth century
 - First success were in non-stochastic experiments with narrow ranges of true experimental values.
 - The earliest measurement methods, and detectors had limited accuracy and precision, resulting in a huge measurement uncertainty, much larger than the range of the true experimental values they measured.
- In quantum mechanics state of the system is changed by the process of measurement.
 - This does not apply to classical, macroscopic systems.

“ $\pm$ ” frequently is used as an upper limit on range of the true experimental values, even if “ $\pm$ ” has nothing to do with this range.

Repetitions of the same event or multiple one-of-a-kind events? All depends on the questions we ask.

Example 1 Recordings from a month of work commuting, including voice from car inside. For calculating road length it is a repeated experiment.

For soundtracks each trip needs to be treated as a one-of-a-kind event. Averaging soundtracks is useless.

Example 2 Hurricanes.

For hurricane frequency as a function of location and season are repeated events.

For analysis of wind patterns, weather dynamic each hurricane is one-of-a-kind.

The same experimental data can be analyzed, depending on the purpose, as a single experiment repeated statistically meaningful number of times, or as a set of unique one-of-a-kind experiments.

Are identical repetitions of the same experiment even possible?

Two repetitions of an experiment, as similar as experimentally achievable. Proton Radiography (PRad) images of shots 0166 and 0183 melting two tin coupons 0.3125 inch-thick by explosion. The major features are preserved; yet size and shape of individual metal blobs is unique for each shot reflecting built-in differences in the internal structure of the metal coupons.

Data analysis for one-of-a kind experiment is different than for a repeated experiment

Repeated experiment

- Golden standard in experimental science.
- Statistical methods of data analysis and experimental error calculation are well developed.
- Data sets with missing information, data affected by an unusual noise can be easily identified and excluded.
- An effect does not need to be identified to be accounted for in the experimental error.
- **Total experimental error is an upper limit for the range of the true experimental values.**

One-of-a-kind experiment (case study)

- In some crucial areas: from nuclear, astronomy, earth sciences, to medicine a lot of data is analyzed in one-of-a-kind event framework.
- Data analysis methods from repeated experiments should not be applied automatically.
- Data sets with missing information, data affected by an unusual noise create a unique problem.
- Each source of the measurement uncertainty needs to be identified, to be accounted for.
- **Total experimental error can not be obtained from one-of-a-kind event analysis.**

One-of-a-kind event provides no information about the range of true experimental values

Three independent problems

**Built-in differences
in experiments**

**Sample/
material
variation**

**External
conditions**

**Geometry of the
experiment**

Stochasticity

**Measurement
uncertainty**

“ ± ”

**Detector
accuracy**

**Detector
precision**

**Data transmission and storage
(cable response, film aging)**

**Data analysis,
computation methods.**

**Incomplete,
& low reliability
data**

Measurement uncertainty and measurement reliability

- **Data reliability** – qualitative judgment by a subject matter expert, if probability of introducing a mistake during the process of obtaining, analyzing the data was high or low.
- Reliability and uncertainty are, in theory, independent.
- In practice, a not-so-robust measurement method is introduced despite a robust one exists when the not-so-robust one significantly lowers the uncertainty.
- When a legacy documentation is ambiguous and/or insufficient it causes low reliability in analysis of a complicated legacy measurement; analysis of a simple legacy measurements is not affected.
- Cross checking a low reliability/ low uncertainty result against an independent high reliability/ large uncertainty result rises reliability of the first one while preserving its low uncertainty.

Data users frequently assume that smaller “ $\pm$ ” translates to a more reliable measurement.

Old simple method: a ruler measures distance with low accuracy and precision creating a large measurement uncertainty. The measurement is very reliable – it is hard to make any mistake.

Modern, sophisticated method: interferometer measures a distance with incredibly high precision, yet this method creates multiple possibilities for obtaining incorrect result.

Low uncertainty/incomplete measurement compared to reliable, high uncertainty result

Example from 2020

1) Low uncertainty, incomplete: “This research was done on **Thursday, the 20-th**”

This is either 02/20/2020 or 08/20/2020 – correct, no unnecessary uncertainty added. Reporting a result with honest use of either/or, despite being the most precise, brings a lot of objections from data users.

Standard reporting: May 20-th 2020 uncertainty 3 months. Formally correct, yet gives unnecessarily huge uncertainty of 6 months, even worse, many data users make completely unjustified Gaussian distribution probability assumption, concluding than end of May is the most probable. For unexplained reason (psychology?) data users prefer it.

2) Adding reliable, high time uncertainty: “At Los Alamos, there was a **snow** outside my window”

Taking into account this rather imprecise information, and weather pattern at Los Alamos NM, makes possible to select the exact date **02/20/2020**

When information about a data point is incomplete, an independent, reliable, high uncertainty result, usually not applicable for a given measurement, may complete it.

2020

- CALENDAR -

JANUARY						
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

FEBRUARY						
S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

MARCH						
S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

APRIL						
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

MAY						
S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

JUNE						
S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

JULY						
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

AUGUST						
S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

SEPTEMBER						
S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

OCTOBER						
S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

NOVEMBER						
S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

DECEMBER						
S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

Summary

- Identical repetitions of an experiment are not achievable. Any macroscopic experiment produces a range of true experimental values, due to built-in differences.
- Range of true experimental values is independent from measurement uncertainty.
- Modern digital equipment frequently reduces measurement uncertainty below the range of true experimental values.
- Data reliability is independent from measurement uncertainty.
- In many important areas of science, data analysis in one-of-a-kind event framework is necessary. It provides valuable quantitative data, which come with measurement uncertainty only.

Data analysis of one-of-a kind event can not give information about the range of true experimental values.

Glossary

Accuracy - description of systematic errors, a measure of statistical bias; *trueness (ISO)*. (Wikipedia)

Precision – description of random errors; a measure of statistical variability (Wikipedia)

“ $\pm$ ” – reported in analysts’ reports without description

Total experimental error – calculated by a statistical method, like a standard deviation, from data obtained in an experiment repeated sufficiently many times. It accounts for all sources of uncertainty, serves as upper limit for the range of true experimental values and for measurement uncertainty.

Range of true experimental values - range, in which the experimental value changes, when the same experiment is repeated, if measured by a non-existing “perfect” detector. It depends exclusively on the experiment, not on the measurement system. It is caused by built-in differences in the experiment.

Built-in differences - sample/material variations, differences in external conditions, stochastic nature of some physical processes, geometry of an experiment– all of them cause that no two repetitions of a physical experiment are exactly identical. Size of the built-in difference depends on specifics of a given experiment.

Measurement uncertainty – uncertainty created by the measurement and data analysis process.

One-of-a-kind experiment – experiment that has not been repeated; experiment that produced data analyzed separately from data produced by other, similar experiments.

Data reliability – qualitative judgment by a subject matter expert, if probability of introducing a mistake during the process of obtaining this data was high or low.

Data point error – difference between the measured value of the data point and its true value if measured by a “perfect” detector. It is always non-zero, and needs to be correctly accounted for and propagated in the data analysis methods

Perfect detector – imaginary detector that would report the measured value without any error.