Uncertainty associated with numerical computation

Alistair Forbes
Mathematics and Scientific Computing
National Physical Laboratory
University of Huddersfield

alistair.forbes@npl.co.uk

NOBUGS 24–26 September, 2012, ISIS/Diamond

Funded by European Metrology Research Programme (EMRP) Joint Research Project NEW06, Traceability in Computationally-Intensive Metrology. The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

National Physical Laboratory

Overview

Metrology context: traceability, uncertainty

Digital context: importance of computation to metrology (science and technology)

Use of numerical artefacts to test numerical software

Nullspace data generation

Measures of performance

European project: provision of benchmarking, validation and customised data suites

National Physical Laboratory

NPL: UK's National Metrology Institute

Founded in 1900

Located in Teddington, South West London, near Richmond, Hampton Court

450+ scientists: physicists, material scientists, chemists, mathematicians

One of a international network of NMIs: NIST (US), PTB (DE), LNE (FR), INRIM (IT), NMIJ (JP), etc.

Overseen by the Bureau International des Poids et Mesures (BIPM), Paris

Government owned (BIS), contractor operated (Serco), contract renewal 2014

NPL's mission (ABF version)

Develop the scientific and technical infrastructure to provide standards

- Equitable
- Independent
- Accessible
- Based on fundamental science: not proprietary, not parochial, not arbitrary

to the benefit of the UK

- Trade, industry, enterprise
- Science and innovation
- Wellbeing of the citizen, now and in the future

Metrology

Base units: kilogram, metre, second, ampere, kelvin, candela, mole

Traceability: property of a measurement result whereby the result can be related to a reference through a documented, unbroken chain of calibrations, each contributing to the measurement uncertainty (VIM)

Measurement uncertainty: non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used (VIM)

The GUM

Guide to Expression of Uncertainty in Measurement: if

$$Y = f(\boldsymbol{X}), \quad X_i \sim \mathsf{N}(x_i, u_i^2)$$

then

$$Y \sim \mathsf{N}(y, u^2), \quad y = f(\boldsymbol{x}), \quad u^2 = \sum_i c_i^2 u_i^2, \quad c_i = \frac{\partial f}{\partial X_i}(\boldsymbol{x}).$$

Bottom-up, model-based approach.

ISO 5725 Accuracy: trueness and precision of measuring systems

Observational approach, repeatability and reproducibility studies

Cannot account for systematic effects that do not contribute to variation in observations

Computation links in the traceability chain

Much of metrology depends on complex computations involving the analysis of data, e.g., to determine the best-fit model to a set of data according to a specified criteria.

In order that traceability of metrology is maintained it is necessary to assess the computational links in the chain.

There are many reasons why numerical software will not give the mathematically exact answer:

- bugs (!?)
- finite precision (IEEE 1.11×10^{-16})
- convergence tolerances for iterative algorithms for nonlinear problems
- approximate algorithms

Regard numerical computation as an influence factor, whose uncertainty has to be evaluated.

EMRP project TraCIM

Traceability in Computationally Intensive Metology

European Metrology Research Programme project (EMRP) involving European labs: CMI (CZ), INRIM (IT), NPL (UK, coordinator), PTB (DE), UM (SL) and VSL (NE)

Industrial partners: Hexagon, Mitutoyo, Werth, Zeiss, leading metrology companies in dimensional metrology

Collaborators: Rolls-Royce, Airbus et al.

Three years, started June 2012, 3 M euros

Traceability of computation at the point of use, through test data (numerical artefacts) available over the internet

Data analysis context

Measurement data $\boldsymbol{y}$ gathered according to the model

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{a}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \in \mathsf{N}(\mathbf{0}, V_{\boldsymbol{y}}).$$

Estimation:

$$\hat{\boldsymbol{a}} = \boldsymbol{g}(\boldsymbol{y}), \quad V_{\boldsymbol{a}} = G V_{\boldsymbol{y}} G^{\mathsf{T}}, \quad G_{ij} = \frac{\partial g_j}{\partial y_i}$$

Bayesian inference:

$$p(\boldsymbol{a}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{a}) p(\boldsymbol{a})$$

Maximise $p(\boldsymbol{a}|\boldsymbol{y})$, approximate with a Gaussian with same mode and curvature at the mode.

Sample from $p(\boldsymbol{a}|\boldsymbol{y})$ using Markov chain Monte Carlo methods.

Uncertainty contributions

Measurement uncertainty associated with the data $\boldsymbol{y}$

Likelihood/statistical model $\mathrm{N}(\mathbf{0}, V_{\boldsymbol{y}})$ associated with the data

Functional model $\boldsymbol{f}$

Computation

Numerical artefacts

Computational task $\boldsymbol{a} = \mathcal{C}(\boldsymbol{y})$ efines an m-dimensional surface in $\mathcal{R}^m \times \mathcal{R}^n$

$\boldsymbol{y}$ input data
$\boldsymbol{a}$ exact mathematical solution

The relation $\boldsymbol{a} = \mathcal{C}(\boldsymbol{y})$ defines an m-dimensional surface in $\mathcal{R}^m \times \mathcal{R}^n$.

Numerical artefact $\langle \boldsymbol{y}, \boldsymbol{a} \rangle$ such that $\langle \boldsymbol{y}, \boldsymbol{a} \rangle$ is very close to the surface.

$$\boldsymbol{a} + \Delta\boldsymbol{a} = \mathcal{C}(\boldsymbol{y}), \quad \boldsymbol{a} = \mathcal{C}(\boldsymbol{y} + \Delta\boldsymbol{y}), \quad \boldsymbol{a} + \Delta\boldsymbol{a}' = \mathcal{C}(\boldsymbol{y} + \Delta\boldsymbol{y}').$$

$\Delta\boldsymbol{a}$, $\Delta\boldsymbol{y}$ are measures of the numerical uncertainty associated with the numerical artefact.

Software/algorithm S under test.

Compare $\hat{\boldsymbol{a}} = S(\boldsymbol{y})$ with $\boldsymbol{a}$.

Nullspace data generation

Optimality conditions $O(\boldsymbol{y}, \boldsymbol{a}) \geq \mathbf{0}$ associated with a computational task, e.g.,

$$\frac{\partial p}{\partial a_j}(\boldsymbol{a}|\boldsymbol{y}) = 0, \quad \lambda_j \geq 0, \quad j = 1, \ldots, n.$$

In the context of fitting a model $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{a})$ to data $\boldsymbol{y}$, let $\boldsymbol{y}^* = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{a})$ so that $O(\boldsymbol{y}^*, \boldsymbol{a}) \geq \mathbf{0}$.

Require $\boldsymbol{y} = \boldsymbol{y}^* + \Delta \boldsymbol{y}$ such that $O(\boldsymbol{y}, \boldsymbol{a}) \geq \mathbf{0}$, also.

$$O(\boldsymbol{y}, \boldsymbol{a}) \approx O(\boldsymbol{y}^*, \boldsymbol{a}) + J_Y \Delta \boldsymbol{y}, \quad \Delta \boldsymbol{y} = \boldsymbol{y} - \boldsymbol{y}^*.$$

If $O(\boldsymbol{y}, \boldsymbol{a})$ is linear in $\boldsymbol{y}$, then we require

$$O(\boldsymbol{y}, \boldsymbol{a}) = [O(\boldsymbol{y}^*, \boldsymbol{a}) = \mathbf{0}] + J_Y \Delta \boldsymbol{y} \Rightarrow J_Y \Delta \boldsymbol{y} = \mathbf{0},$$

i.e., $\Delta \boldsymbol{y}$ is in the null space of J_Y.

Easy to generate null space vectors of a matrix using orthogonal factorisation techniques.

Advantages of the null space approach

The data generation problem is usually much simpler to solve that the associated computational problem.

Avoids the need to produce ‘reference implementations’ for the computational problem. (Who tests the reference implementation?)

Multiple test data sets can be generated for almost no additional effort.

It is easy to generate data that simulates real world data.

Example: errors-in-variables

Standard model

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{a}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \in \mathsf{N}(\mathbf{0}, \sigma^2 I).$$

Maximum likelihood model fit $\boldsymbol{a}_0 = S_0(\boldsymbol{y})$:

$$\min_{\boldsymbol{a}} \sum_i (y_i - f_i(\boldsymbol{x}, \boldsymbol{a}))^2.$$

More realistic model:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x} + \boldsymbol{\delta}, \boldsymbol{a}) + \boldsymbol{e} + \boldsymbol{\epsilon}, \quad \boldsymbol{\delta} \in \mathsf{N}(\mathbf{0}, V_{\boldsymbol{x}}), \quad \boldsymbol{\epsilon} \in \mathsf{N}(\mathbf{0}, V_{\boldsymbol{y}}),$$

and

$$\boldsymbol{e} \in \mathsf{N}(\mathbf{0}, V_{\boldsymbol{e}}), \quad \mathsf{cov}(e, e') = k(\boldsymbol{x}, \boldsymbol{x}').$$

Maximum likelihood model fit $\boldsymbol{a} = S(\boldsymbol{x}, \boldsymbol{y})$.

Easy to generate numerical artefacts for the more comprehensive model.

Measures of information loss

Approximate solution and comprehensive solution

$$\mathrm{N}(\boldsymbol{a}_0, V_0), \quad \mathrm{N}(\boldsymbol{a}, V)$$

Relative entropy:

$$D_{\mathrm{KL}}(p \| p_0) = D_1 + D_2,$$

where

$$D_1 = \frac{1}{2}\left[\log\left|V_0 V^{-1}\right| + \mathrm{Tr}\left(V V_0^{-1}\right) - n\right]$$

and

$$D_2 = \frac{1}{2}(\boldsymbol{a} - \boldsymbol{a}_0)^\top V_0^{-1}(\boldsymbol{a} - \boldsymbol{a}_0).$$

Numerical uncertainty

$$\langle \boldsymbol{y}, \boldsymbol{a} \rangle, \quad \boldsymbol{a} \in \mathsf{N}(\boldsymbol{a}^*, V_\sharp), \quad \boldsymbol{a}^* = \mathcal{C}(\boldsymbol{y})$$

Simulated measurement uncertainty

$$\langle \boldsymbol{y}, \boldsymbol{a} \rangle, \quad \boldsymbol{y}^* = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{a}^*), \quad \boldsymbol{y} \in \mathsf{N}(\boldsymbol{y}^*, V_{\boldsymbol{y}}), \quad V_\flat = G V_{\boldsymbol{y}} G^\top$$

Measures of performance

Numerical artefact $\langle \boldsymbol{y}, \boldsymbol{a} \rangle$, $V_\sharp$, $V_{\boldsymbol{y}}$, $V_\flat$.

Software/algorithm under test $\hat{\boldsymbol{a}} = S(\boldsymbol{y})$.

Numerical accuracy:

$$D_\sharp^2 = (\hat{\boldsymbol{a}} - \boldsymbol{a})^\top V_\sharp^{-1} (\hat{\boldsymbol{a}} - \boldsymbol{a})$$

Fitness for purpose (forward):

$$D_\flat^2 = (\hat{\boldsymbol{a}} - \boldsymbol{a})^\top V_\flat^{-1} (\hat{\boldsymbol{a}} - \boldsymbol{a})$$

Fitness for purpose (inverse):

$$E_\flat^2 = (\breve{\boldsymbol{y}} - \boldsymbol{y})^\top V_{\boldsymbol{y}}^{-1} (\breve{\boldsymbol{y}} - \boldsymbol{y}), \quad \breve{\boldsymbol{y}} = \mathcal{C}(\hat{\boldsymbol{a}})$$

Data suites

Formalised computational aim $\mathcal{C}_k$ defined in documentary standards, e.g., ISO.

Public domain benchmarking data suites for that aim, $B_{k,q} = \langle \boldsymbol{y}_q, \boldsymbol{a}_q \rangle$ and associated measures of performance M_k.

Confidential validation data suites, $V_{k,q} = \langle \boldsymbol{y}_q, \boldsymbol{a}_q \rangle$ and associated measures of performance.

Certified test results, $\langle \boldsymbol{y}_q, \widehat{\boldsymbol{a}}_q \rangle$.

Customised data sets based on user supplied data:

$$\langle \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{a}} \rangle \mapsto \langle \boldsymbol{y}, \boldsymbol{a} \rangle, \quad \boldsymbol{y} \approx \tilde{\boldsymbol{y}}, \quad \boldsymbol{a} \approx \tilde{\boldsymbol{a}}, \quad \boldsymbol{a} = \mathcal{C}(\boldsymbol{y}).$$

Traceable computation at point of use

Data analysis software S with associated computational aim $\mathcal{C}_k$, last validation date d_S, could be $-\infty$.

User calls S.

Compare validation date d_S with date d_V associated with the latest validation suite V_k.

If $d_S < d_V$, user is alerted.

Can run the latest validation suite now or later.

Software can be kept in constant 'calibration'.

Traceable computation at point of use and application

User calls S for a particular set of data and supplies $\langle \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{a}} \rangle$.

Data generator produces

$$\langle \boldsymbol{y}, \boldsymbol{a} \rangle, \quad \boldsymbol{y} \approx \tilde{\boldsymbol{y}}, \quad \boldsymbol{a} \approx \tilde{\boldsymbol{a}}, \quad \boldsymbol{a} = \mathcal{C}(\boldsymbol{y}).$$

Runs S on $\boldsymbol{y}$ to produce $\hat{\boldsymbol{a}} = S(\boldsymbol{y})$ and compares $\hat{\boldsymbol{a}}$ with $\boldsymbol{a}$.

Comparator principle: if $\hat{\boldsymbol{a}} \approx \mathcal{C}(\boldsymbol{y})$ then $\tilde{\boldsymbol{a}} \approx \mathcal{C}(\tilde{\boldsymbol{y}})$ (?)

Software is ‘calibrated’ for that particular instance of the computational task.

Summary

Metrology in the digital age has to take into account computational links in the traceability chain.

New European project to address the computational links.

Standardised computational tasks $\mathcal{C}_k$.

Data generators to produce $\langle \boldsymbol{y}, \boldsymbol{a} \rangle$ using nullspace methods, if possible.

Use of web services to implement traceability at point of use.

Major instrument manufacturers to implement online system in the field of dimensional metrology.

Extend to other metrology/science domains.