Numerical aspects in the evaluation of measurement uncertainty

Maurice Cox, Alistair Forbes, Peter Harris and Clare Matthews
National Physical Laboratory, UK
maurice.cox@npl.co.uk

Uncertainty Quantification in Scientific Computing
Boulder, Colorado, 1–4 August 2011

V110803


The Community Foundation
Boulder County
1877
Willamette Arnett builds house for $4,000, plants 10 cottonwood trees
1883
Boulder's Victorian-era county courthouse is completed
1894
The 100-year flood
1898
Boulder Chautauqua opens
1900
Willamette Arnett dies in the Klondike seeking to reverse his fortunes
1914
Eliza Jane, widow of Hiram Fullen, purchases house
1913
The Big Snow, 3 feet in two days
1918
Spanish influenza kills 41 in Nederland; Boulder quarantined
1932
Old courthouse is gutted by fire
1954
President Eisenhower visits Boulder to open NBS labs
1994
Historic Boulder purchases Arnett-Fullen House


1954
PRESIDENT EISENHOWER
VISITS BOULDER TO
OPEN NBS LABS

Summary

Numerical quantification of results from a measurement uncertainty (MU) computation in terms of computational inputs

Primary output often an approximation to the PDF (probability density function) for the (univariate or multivariate) measurand (quantity intended to be measured)

From this PDF results of interest can be derived

Many metrology problems small-scale; important exceptions

Main driver: production of guidance for metrologists on MU evaluation

Specifics

Account for available knowledge of input quantities

Propagation of distributions through a computational model

Numerical quality and representation of MC results

Sensitivity issues

Concluding remarks and speculations

Traditional approach to MU evaluation

National Metrology Institutes (NMIs) and industrial laboratories routinely propagate uncertainties related to input quantities through computational models to provide uncertainties related to output quantities

Computational model $\equiv$ mathematical model of measurement

Relevant guidance available and supporting software exists

Joint Committee for Guides in Metrology (JCGM)

"To maintain and promote the use of the Guide to the Expression of Uncertainty in Measurement (GUM) and the International Vocabulary of Basic and General Terms in Metrology (VIM)"

JCGM Member Organizations (up to 3 reps from each)

BIPM	Bureau International des Poids et Mesures
IEC	International Electrotechnical Commission
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
ILAC	International Laboratory Accreditation Cooperation
ISO	International Organization for Standardization
IUPAC	International Union for Pure and Applied Chemistry
IUPAP	International Union for Pure and Applied Physics
OIML	International Organization of Legal Metrology

GUM

In a very practical sense, has served metrology well since 1993

Basis: Linearization of model $[Y = f(\boldsymbol{X})]$, normality assumption, mix of frequentist and Bayesian statistics

JCGM revising GUM because of limitations and inconsistencies

JCGM view: characterize input quantities by PDFs, which are propagated through the model to obtain PDF for output quantities

Best estimate, standard uncertainty and coverage intervals for the measurand (all used by metrologists) then readily obtained

JCGM 100	Guide to the expression of uncertainty in measurement (GUM)	2008†
JCGM 101	Propagation of distributions using a Monte Carlo method	2008
JCGM 102	Extension to any number of output quantities	2011*
JCGM 103	Developing and using measurement models	Draft
JCGM 104	An introduction to the GUM and related documents	2009
JCGM 105	Concepts, principles, and methods for the evaluation of MU	Draft
JCGM 106	Conformity assessment	2011*
JCGM 107	Least squares adjustment	‡
JCGM 200	Vocabulary of metrology (VIM)	2008

www.bipm.org †Under revision *Expected ‡Pending

Main considerations

MU evaluation

Some principles apply more widely (not all inputs always relate to measurement)

The result of a computation represents the effect of uncertainty from all sources considered

Numerical methods of solution, especially MC and MC-like methods, used

Knowledge and PDFs

Computational model $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$: measurement equation (ME)

$\boldsymbol{X}$: input quantities (N in number)
$\boldsymbol{Y}$: output quantities

$\boldsymbol{f}$: given function, specified by a computational model

Given knowledge about $\boldsymbol{X}$, knowledge is required about $\boldsymbol{Y}$

Prior knowledge of $\boldsymbol{Y}$ may be available

Components of $\boldsymbol{X}$ are characterized by random variables, and in all cases we encode available knowledge about $\boldsymbol{X}$ as a PDF

More boring notation

$p(\boldsymbol{Z})$: PDF for quantity $\boldsymbol{Z}$ by

$\boldsymbol{z}$: estimate of $\boldsymbol{Z}$, taken as $\boldsymbol{E}(\boldsymbol{Z})$

$\boldsymbol{U}_{\boldsymbol{z}}$: associated covariance matrix, taken as $\boldsymbol{V}(\boldsymbol{Z})$

$u(z_i)$: standard uncertainty associated with ith component of $\boldsymbol{z}$

Different forms of uncertainty?

Consider (A) aleatory uncertainties (due to random effects) and (B) epistemic uncertainties (due to other effects)

Some authors treat (A) as random variables with PDFs, and (B) as intervals with no assumed PDFs

In metrology we encode knowledge of any quantity by a PDF, as advocated by the GUM

The rules of probability calculus can then be employed

In contrast, the two types of uncertainty are propagated separately and results combined, with nesting of A within B, e.g., Roy and Oberkampf (2011), DAKOTA

Numerical analysis

Numerical analysis has long history in uncertainty quantification (UQ) when computing in finite arithmetic

Two principal techniques for carrying out error propagation: interval analysis and floating-point (FP) error analysis

We recognize value of FP error analysis: analysis of numerical stability of algorithms used within the computational model

We distinguish between errors and uncertainties

Error: difference between the value of a quantity and the true value for that quantity

Uncertainty: measure of dispersion (such as the standard deviation of the PDF) for that quantity

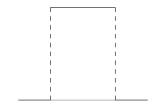
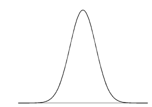
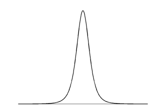
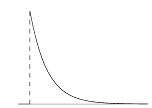
Accounting for available knowledge of input quantities

Encoding of knowledge of input quantity by PDF for that quantity

Use MAXENT, the maximum entropy principle, or

Bayes' theorem when repeated observations of a quantity are available

Knowledge-based PDFs (GUM Supplement 1)

Lower and upper limits	Uniform	
Best estimate, standard uncertainty	Gaussian	
Independent Gaussian observations (unknown expectation, variance)	Scaled, shifted t	
Best estimate of non-negative quantity	Exponential	

More in GUM Supplement 1

Example: Has a horse been doped?

Hibbert et al (2011) apply MAXENT and Bayesian model selection to decision-making problems in horse-doping:

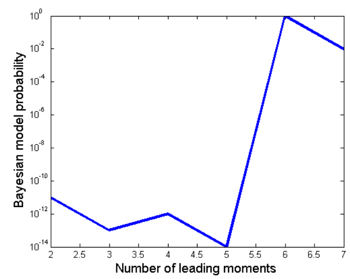
From large mass of historical data construct PDFs for TCO2 concentration in pre-race samples of plasma

Obtain separate PDFs for 'clean' horses and horses that were subsequently tested positive

Using q leading moments of data, apply MAXENT to deliver PDF based on set of Lagrangian parameters

Use Bayesian model selection to obtain q that maximizes Bayesian model probability: avoids model over-fitting

For clean horses, probabilities of models (log scale):



Conclude: Bayesian model selection strongly settles for moderately complex model of form $\exp(a_1 X + \cdots + a_6 X^6)$

Compared with simple model such as $\exp(a_1 X + a_2 X^2)$, which, for $a_2 < 0$, is Gaussian

Measured data for further horse compared with these PDFs and decision made on whether horse has been doped

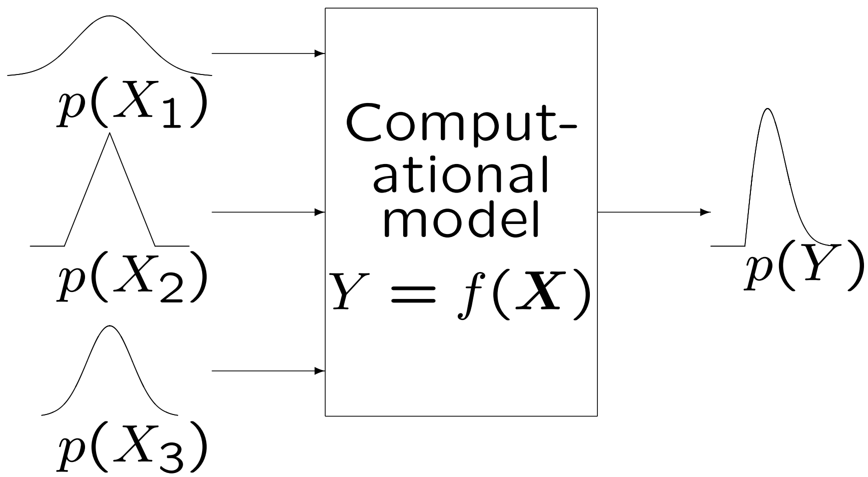
Candidate example for JCGM conformity assessment document

Propagation of distributions and Markov's theorem

Obtaining $p(\boldsymbol{Y})$ given PDF $p(\boldsymbol{X})$

Formally, apply Markov's theorem (see Cox & Siebert, 2006)

$$p(\boldsymbol{Y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\boldsymbol{X})\delta(\boldsymbol{\eta} - f(\boldsymbol{X}))\,\mathrm{d}\boldsymbol{X}$$



$p(Y)$ often asymmetric for non-linear $f(\boldsymbol{X})$ or asymmetric $p(X_i)$

Approach

Quadrature rule can be applied, but inefficiently, to evaluate integral, so as to provide approximation $\widehat{p}(\boldsymbol{Y})$ to $p(\boldsymbol{Y})$

Approach commonly used to obtain a $\widehat{p}(\boldsymbol{Y})$ is an MC method as in GUM Supplement 1

Make random draws from $p(\boldsymbol{X})$, evaluate $\boldsymbol{f}$ in each case, and use resulting set of values to form $\widehat{p}(\boldsymbol{Y})$

Applies when $p(\boldsymbol{X})$ does not depend on measurand

Bayes

When observations of an X_i are available, observation equation (OE) approach is appropriate, and Bayes' rule can be used to determine $p(\boldsymbol{Y})$

Let $\boldsymbol{X}$ denote the original $\boldsymbol{X}$ *less* W, one of the X_i that is observable

Re-express ME as $\boldsymbol{Y} = \boldsymbol{f}(W, \boldsymbol{X})$ and consider the OE $W = \phi(\boldsymbol{Y}, \boldsymbol{X})$ (Possolo & Toman 2007, Forbes and Sousa 2011) and observations $W_i \in \mathrm{N}(W, \sigma^2)$

Bayes' rule used to update prior knowledge of Y, $\boldsymbol{X}$ and σ^2 (regarded as random variables) with observations W_i to give posterior distribution, with $p(\boldsymbol{Y})$ obtained by marginalization

MCMC

An MCMC algorithm can be used to obtain $\widehat{p}(\boldsymbol{Y})$

Generates sequence $\{\boldsymbol{y}_k\}$ in which $\boldsymbol{y}_k$ is obtained from $\boldsymbol{y}_{k-1}$

Asymptotically generates draws from $p(\boldsymbol{Y})$

Metropolis-Hastings algorithm: MCMC algorithm that allows $p(\boldsymbol{Y})$ to be specified straightforwardly

Numerical quality of Monte Carlo results

When making draws from $p(\boldsymbol{X})$, that PDF can often be decomposed into univariate PDFs or joint PDFs involving smaller number of variables

Procedures for sampling from variety of PDFs commonly occurring in metrology such as normal, multinormal, t and arcsine summarized in JCGM Supplement 1

Rely on quality of uniform RNG: high-quality generators available that pass extensive tests of statistical properties

RNG on distributed computing systems: Wichmann & Hill (2006)

Monte Carlo convergence

Suppose M random draws made from joint PDF for $\boldsymbol{Y}$ and corresponding model values $\boldsymbol{f}$ calculated

Closeness of agreement between average of these values and $\mathrm{E}(Y_j)$ expected to be proportional to $M^{-1/2}$

'Convergence rate' can be improved for certain classes of problem by using schemes such as Latin Hypercube sampling (LHS)

Adaptive schemes

Above approach necessitates specifying M in advance

Thus, numerical accuracy of results obtained unknown a priori

An adaptive scheme, designed to meet a specified numerical tolerance δ, provides information required by metrologists:

1. estimate $\boldsymbol{y}$ of $\boldsymbol{Y}$
2. associated covariance matrix $\boldsymbol{U}_{\boldsymbol{y}}$
3. coverage region for $\boldsymbol{Y}$ for a stipulated coverage probability p

Approach

An approach, involving carrying out sequence of applications of MCM operates in terms of δ and a sequence of batches of say $M_0 = 10^4$ MC trials:

1. Carry out a batch of MC trials and use model values to calculate batch results (averages, standard deviations, etc.)

2. Use updating techniques to calculate results for all batches

3. Regard computation as having stabilized when standard deviations of average of batch results $\leq \delta$

Can be tailored to other sampling procedures such as LHS

Representation of MC results for ...

(a) visualization purposes (e.g., surfaces, contours),
(b) subsequent MU evaluation

Regarding (b) the output of one MU evaluation should be transferable, i.e., usable as input to further evaluation (GUM)

In particular, not always convenient to retain the $M = 10^6$, say, (vector) values produced by MC and use them subsequently

But, MC output ideal in that it automatically conveys covariance information, which can subsequently be sampled

Methods such as kernel density estimation (or, better, approximation) can be used for (a) and (b)

Memory, time and other considerations

Possible problems when M large, say $O(10^7)$. At NPL we

1. Perform a modest number, say $M_0 = 10^4$, of MC trials
2. Establish set of bins based on these trials
3. Make further $M - M_0$ trials, allocating to bins or, when outside bins, saving individually

Bins, bin frequencies and further values used to obtain results

Advantages in providing coverage intervals and regions, which depend crucially on tail information

Density approximation

Kernel density approximation (KDA) to obtain a $\widehat{p}(\boldsymbol{Y})$ from sampled values (not yet used greatly in metrology)

KDA to a univariate PDF $p(Y)$:

$$\widehat{p}_h(Y) = \frac{1}{Mh} \sum_{r=1}^{M} K\left(\frac{Y - y_r}{h}\right)$$

$y_1, \ldots, y_M$ are sampled values with underlying density p, and K is a kernel function with unit area

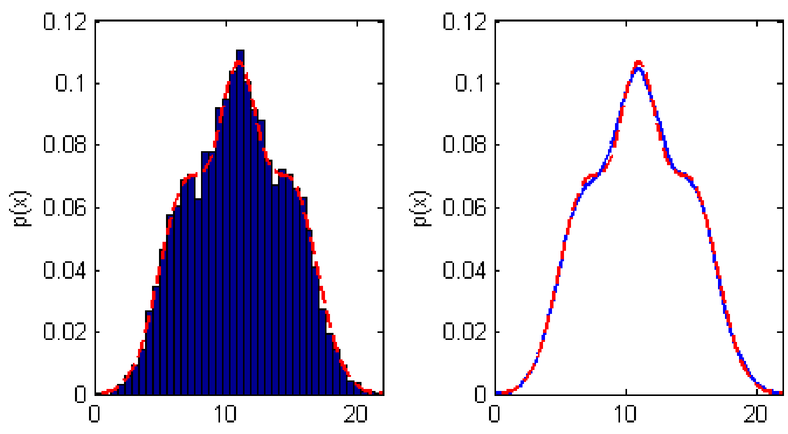
Common kernel functions are Gaussian and B-spline

B-splines have appreciable speed advantages when sampling from $\widehat{p}_h$

Smoothing parameter

h is a smoothing parameter, the bandwidth, and plays a similar role to that of bin width in a histogram

Too small an h: spurious behaviour
Too large an h: over-smoothing, losing local detail
Determination of h: Silverman (1986), Sheather (2004)



KDAs and parametric forms

A (conventional) KDA has the same information content as the data it represents

With M often $O(10^7)$, possible to produce KDAs with many fewer terms

Also possible to describe the x_r by some parametric form with adjustable parameters, e.g., Willink (2009) uses an asymmetric form of 'lambda distribution'

Distribution defined by quantile function (inverse distribution function): only 4 parameters, so sampling straightforward

Should this approximation be inadequate in any particular case, a KDA can be used

Contouring

Bivariate PDF sometimes represented by set of contour lines

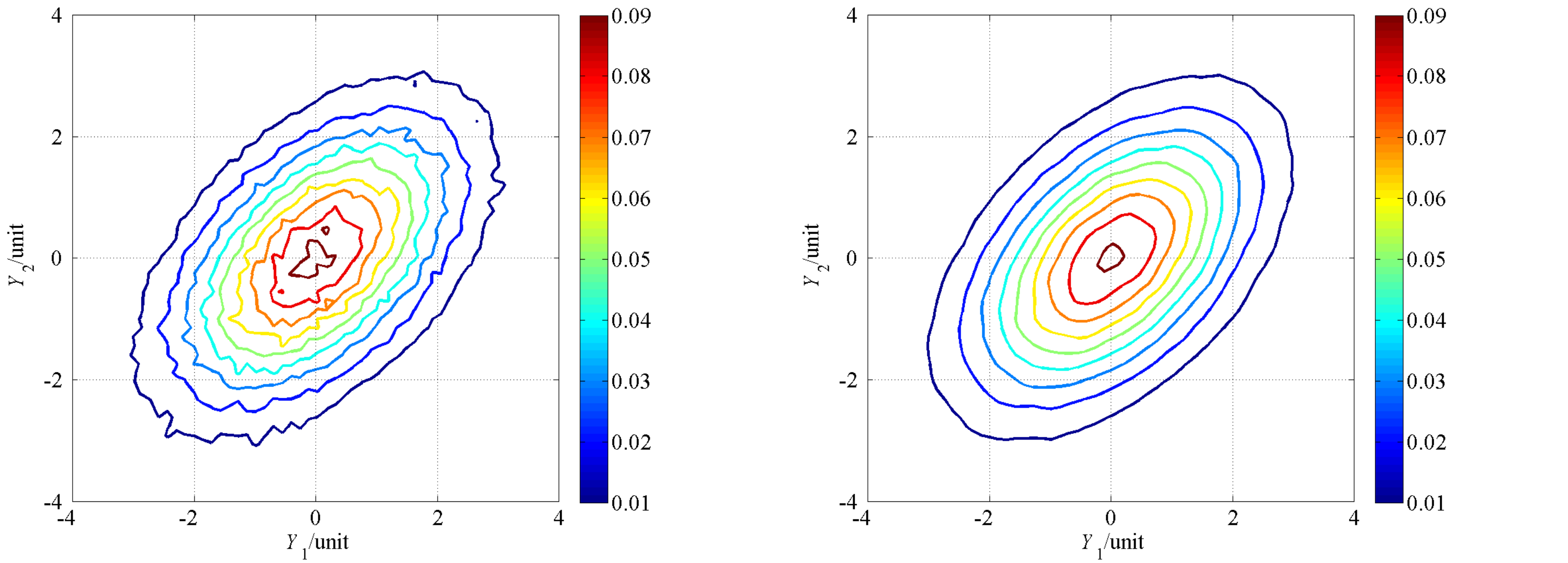
Contour lines should be faithfully reproduced: as $M \to \infty$, they converge to the contours of the corresponding PDF

Appropriate smoothing needed (Silverman 1986, Scott et al 2004)

Some contour diagrams can be computed directly from a KDA (or some other approximation to the corresponding PDF)

For others an appropriate smoothing algorithm can be applied to the MC results and the resulting smoothed contours drawn

Example: without and with smoothing



Coverage regions

In metrology coverage intervals and coverage regions frequently required to accompany measurement results

A procedure (Possolo, 2010), provides an approximation to the smallest $100p\,\%$ coverage region

Included in GUM Supplement 2

1. Construct a rectangular region in space of output quantities

2. Subdivide this region into mesh of small rectangles

3. Assign each output quantity value to rectangle containing it

4. Use the fraction of the values assigned to each rectangle as the approximate probability that $\boldsymbol{Y}$ lies in that rectangle

5. List the rectangles in terms of decreasing probability

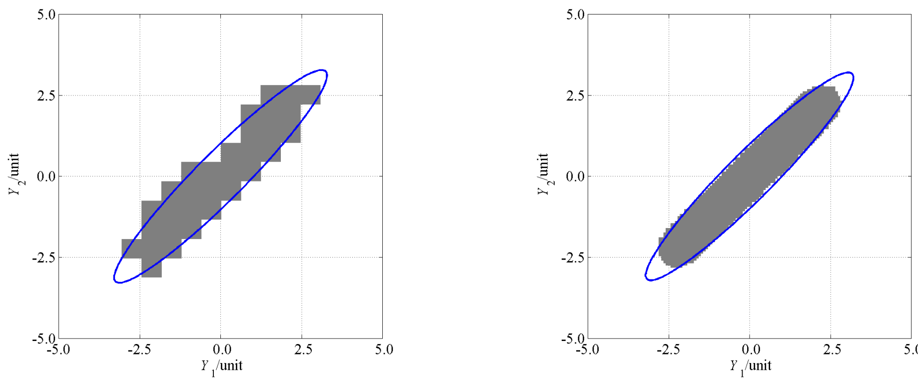
6. Form cumulative sum of probabilities for these listed rectangles: stop when sum $\geq p$, taking chosen rectangles as defining smallest coverage region

Example of smallest coverage region

Model $Y_1 = X_1 + X_3$, $Y_2 = X_2 + X_3$, with independent $X_1 \sim N(0, 0.1)$, $X_2 \sim N(0, 0.1)$, $X_3 \sim \mathrm{R}(-(5.7)^{1/2}, (5.7)^{1/2})$

Approximations to smallest 95 % coverage region, obtained using procedure, based on a 10×10 and a 100×100 mesh

95 % elliptical coverage region (shown by solid line) for $\boldsymbol{Y}$ based on Gaussian parameters estimated from model values



Uncertainty budget and sensitivity coefficients

Uncertainty budget: quantifies uncertainty contributions

Sensitivity: $c_i = \frac{\partial f}{\partial X_i}$ evaluated at $\boldsymbol{x}$

(First-order) uncertainty contribution: $u_i(y) = |c_i| u(x_i)$

Complex-step method (Lyness and Moler, 1967) to obtain c_i: deserves greater recognition

Applicable when real types can be replaced by complex types

'Non-linear' sensitivity coefficient: Carry out MU evaluation by MC, holding all inputs but one at their estimates

Might be unreliable if there are interaction terms

Complex-step method

Provides numerically accurate first derivatives (does not extend to higher-order)

Uses Taylor expansion of function f of a complex variable:

$$f(z+w) = \sum_{r=0}^{\infty} \frac{w^r}{r!} f^{(r)}(z)$$

Setting $z = x$ and $w = \mathrm{i}h$ where x is real and h is real and small, and taking real and imaginary parts:

$$\Re f(x+\mathrm{i}h) = f(x) - \frac{h^2}{2} f''(x) + \ldots, \quad \Im f(x+\mathrm{i}h) = hf'(x) - \frac{h^3}{6} f'''(x) + \ldots$$

from which, with truncation errors $O(h^2)$

$$f(x) = \Re f(x+\mathrm{i}h), \qquad f'(x) = \frac{1}{h} \Im f(x+\mathrm{i}h)$$

Practicalities

Unlike use of finite-difference formula for $f'(x)$, h chosen to be very small

No concern about loss of significant digits through cancellation since no subtraction is involved

$h = 10^{-100}$ in NPL's software (Higham et al, 2010), suitable for all but pathologically-scaled problems

NPL routinely applies the complex-step method

Concluding remarks and forward look

Intensive computation beyond GUM

Infrastructure in place to deal with MU when expressed as $u(y)$ or $U(y)$ for some coverage probability p

GUM goals—universal, internally consistent and transferable framework—largely achieved

When MU expressed using PDFs, MU evaluation needs (possibly intensive) computation, and generates data to be represented suitably

Many measurement models can only be treated numerically

Example: radiation-transport calculation, itself an MC calculation

Efficiency

With MU evaluation, need for more efficient techniques for propagating PDFs

MC naturally highly parallelizable (NPL uses a grid of PCs to treat complex computational models)

As all such techniques are based on MC, it is a matter of tuning those techniques appropriately

For some problems the basic technique can hardly be bettered

Approaches such as LHS can give appreciable gains for certain classes of problem

Embracing UQ

In future, MU evaluation to embrace more strongly concepts used in UQ

Model uncertainty recognized, being termed definitional uncertainty (VIM)

Elicitation so far hardly treated in metrology

Numerical uncertainty considered when computational models constitute FE solvers, e.g., or in using adaptive schemes (GUM Supplement 1)

Embracing UQ

As with general UQ, main aim to provide probabilistic statements about quantities of interest to inform decision makers

A politician or manufacturing production manager considers evidence and decides course of action to achieve some goal

Decisions: A process remaining under statistical control?
An athlete (or animal) regarded as using banned substance?

Tools requiring development for MU evaluation

Use of MAXENT

Other metrological applications could benefit from the approach to the horse-doping problem (Hibbert, 2011)

Numerical difficulties, though, can give rise to ill-determined PDFs or prevent a PDF from being obtained at all when MAXENT is applied to moments (Lira, 2002)

These difficulties arise when metrological problem ‘unrealistic’ in that it relates to a poor measurement of a quantity

Useful to have a characterization of such problems

Efficiency of KDA forms

Compact form for $\widehat{p}(\boldsymbol{Y})$ desirable when used as input to subsequent MU evaluation

Either (a) assemble MC results as histogram and use KDA, or (b) represent ordered MC results by suitable monotonic approximating CDF

KDA choice: Gaussians, B-splines, . . .

For (a), evaluation efficiency of inverse CDF, when generating draws from distribution using B-spline representation, because of compact support of B-splines

For (b), CDF can be differentiated to form corresponding PDF

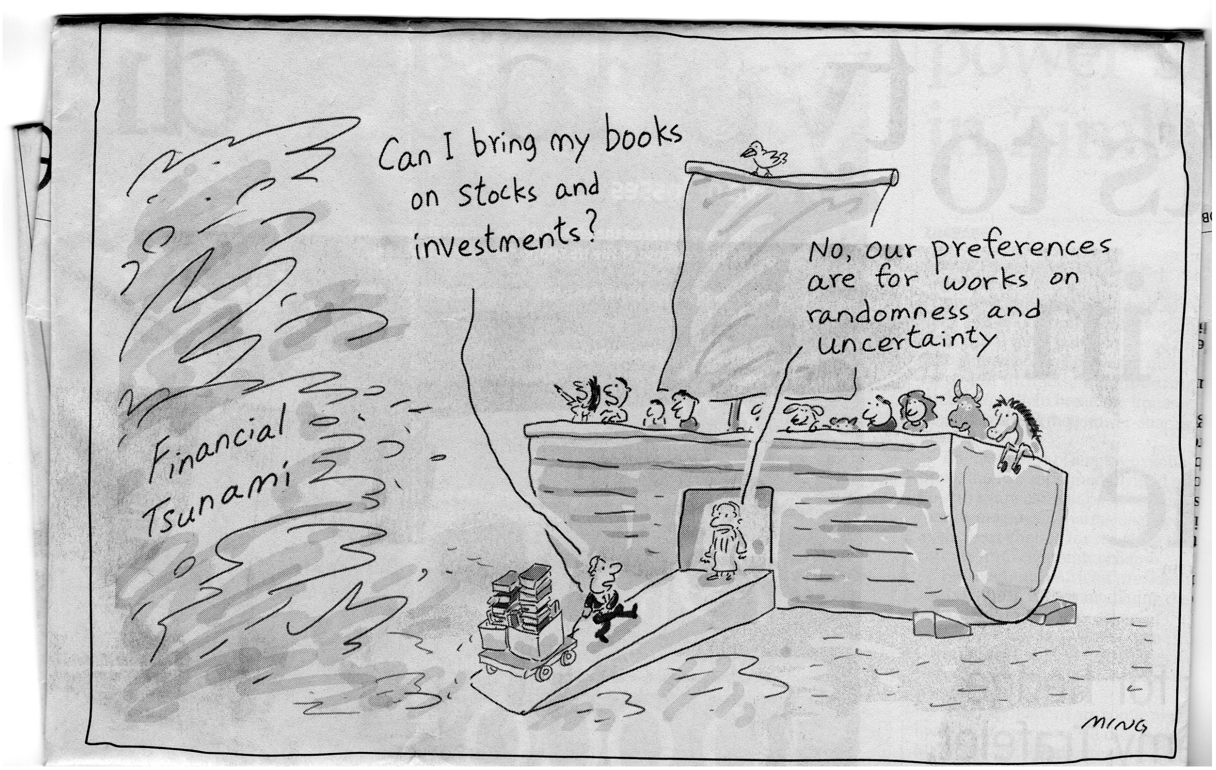
Calibration certificates in the future

Conveying results from MC calculation, particularly in the presence of asymmetric PDFs, etc.

Summarizing

Advice to NMIs and industry

Acceptance by accreditation bodies



Acknowledgment: South China Morning Post, 19 November 2008

Thank you