Optimal Uncertainty Quantification
Distributional Robustness versus Bayesian Brittleness

Tim Sullivan

University of Warwick

How Good is Good Enough?
ASME 2013 Frontiers in Medical Devices
Washington, D.C., U.S.A.
11–13 September 2013

THE UNIVERSITY OF
WARWICK

Credits

Joint work with

- Michael McKerns, Lan Huong Nguyen, Michael Ortiz, and Houman Owhadi (Caltech)

- Clint Scovel (ex-Los Alamos National Laboratory, now Caltech)

- Dominik Meyer (ex-T. U. München)

- Florian Theil (Warwick)

CALTECH
PSAAP

National Nuclear Security Administration

the
UQ foundation

Overview

Overview

Certification Problem

- Many fields now face a need to combine often complex physical modelling with its associated uncertainties with statistical modelling and its uncertainties.
  - engineering, medicine / epidemiology, insurance, meteorology / climate / environment, . . .
- A revealing example problem is certification:
- Certify that a device/procedure/treatment will work satisfactorily with high enough probability ($+$ not fail too badly when it does fail).
- When dealing with commercial, legal or ethical issues, this can be a very high consequence assessment about a rare event.
- Seemingly obvious first steps: users, engineering designers, policy-makers and UQ practitioners must agree upon meanings for "work satisfactorily" and "high enough probability" — and be prepared to perturb them!

Certification Problem

- Introduce $X \sim \mu_0$, a random variable describing the clinical picture, the proposed treatment regimen, &c. — all the inputs to the system.
- Let $g_0(X)$ denote the corresponding outputs on applying the proposed device/procedure.
- Let $q(X, g_0(X))$ be the quantity of interest.
- We want to know

$$\mathbb{E}_{X \sim \mu_0}\big[q(X, g_0(X))\big],$$

  e.g. $q = 1$ if the body temperature rise in an MRI scan is outside FDA-prescribed limits, and $q = 0$ otherwise. Hope $\mathbb{E}[q]$ is small!
- Many UQ methods focus on efficiently evaluating this expected value (i.e. integral), often with as few model evaluations as possible.
- A pressing concern is that of epistemic uncertainty, i.e. that we never know reality's μ_0 and g_0 precisely.

Overview

Distributional and Functional Robustness

Challenge

What should we do in a high-consequence setting when there is significant epistemic uncertainty about what the "correct" model and probability distributions are?

- If it were a question of a single real parameter p_0 known to lie in some range $\mathcal{A} \subset \mathbb{R}$, we would simply optimize (minimize and maximize) the QoI with respect to $p \in \mathcal{A}$.
- So why not do the same with respect to $p = (g, \mu)$? Questions of
    - problem formulation;
    - problem execution (i.e. computability and computation); and
    - payoff — what do we get from the exercise?

Optimal UQ — Formulation

- Just as in the single-parameter case, specify a feasible set of admissible scenarios (g, μ) that could be (g_0, μ_0) according to the available information:

$$\mathcal{A} = \left\{ (g, \mu) \middle| \begin{array}{c} (g, \mu) \text{ is consistent with the current} \\ \text{information about } (g_0, \mu_0) \\ \text{(e.g. legacy data, models, theory, expert judgement)} \end{array} \right\}.$$

- A priori, all we know about reality is that $(g_0, \mu_0) \in \mathcal{A}$; we have no idea exactly which (g, μ) in $\mathcal{A}$ is actually (g_0, μ_0).
  - ▶ No $(g, \mu) \in \mathcal{A}$ is "more likely" or "less likely" to be (g_0, μ_0).
  - ▶ Particularly in high-consequence settings, it makes sense to adopt a posture of healthy conservatism and determine the best and worst outcomes consistent with the information encoded in $\mathcal{A}$.

- Dialogue between UQ practitioners and the domain experts is essential in formulating — and revising — $\mathcal{A}$.

Optimal UQ — Formulation

$$\mathcal{A} = \left\{ (g, \mu) \;\middle|\; \begin{array}{c} (g, \mu) \text{ is consistent with the current} \\ \text{information about (i.e. could be) } (g_0, \mu_0) \end{array} \right\}$$

- Optimal bounds (w.r.t. the information encoded in $\mathcal{A}$) on the quantity of interest $\mathbb{E}_{\mu_0}[q(X, g_0(X))]$ are found by minimizing/maximizing $\mathbb{E}_\mu[q(X, g(X))]$ over all admissible scenarios $(g, \mu) \in \mathcal{A}$:

$$\underline{Q}(\mathcal{A}) \leq \mathbb{E}_{\mu_0}\big[q(X, g_0(X))\big] \leq \overline{Q}(\mathcal{A}),$$

where $\underline{Q}(\mathcal{A})$ and $\overline{Q}(\mathcal{A})$ are defined by the optimization problems

$$\underline{Q}(\mathcal{A}) := \min_{(g,\mu)\in\mathcal{A}} \mathbb{E}_\mu\big[q(X, g(X))\big],$$

$$\overline{Q}(\mathcal{A}) := \max_{(g,\mu)\in\mathcal{A}} \mathbb{E}_\mu\big[q(X, g(X))\big].$$

- Cf. generalized Chebyshev inequalities in decision analysis (**Smith** (1995)), imprecise probability (**Boole** (1854)), distributionally robust optimization, robust Bayesian inference (surv. **Berger** (1984)).

Optimal UQ — Examples

Example: Balancing a Seesaw

You are given 1kg of sand to arrange however you wish on a seesaw ($=$ the real line). Your challenge is to make the region $x \geq t$, $t \geq 0$, as heavy as possible subject to two constraints:

- the centre of mass of the sand (and seesaw) must be at $x = 0$; and
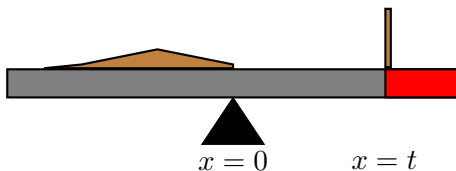- all the sand must be contained in a region of length $\leq L$ (with $L \geq t$).

$$x = 0 \qquad x = t$$
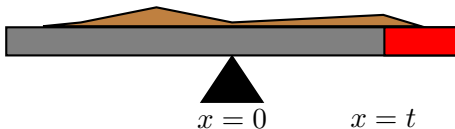
- Optimal UQ can be seen as the extension of the same basic idea to complicated settings: no hope of a pen-and-paper solution, but can compute a numerical solution.

Example: Balancing a Seesaw

You are given 1kg of sand to arrange however you wish on a seesaw (= the real line). Your challenge is to make the region $x \geq t$, $t \geq 0$, as heavy as possible subject to two constraints:

- the centre of mass of the sand (and seesaw) must be at $x = 0$; and
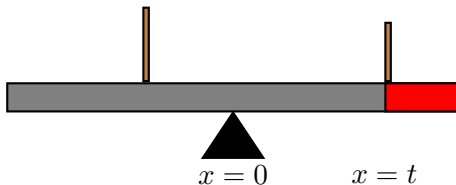- all the sand must be contained in a region of length $\leq L$ (with $L \geq t$).

$x = 0$ $x = t$

- Optimal UQ can be seen as the extension of the same basic idea to complicated settings: no hope of a pen-and-paper solution, but can compute a numerical solution.

Optimal UQ — Examples

Example: Balancing a Seesaw

You are given 1kg of sand to arrange however you wish on a seesaw ($=$ the real line). Your challenge is to make the region $x \geq t$, $t \geq 0$, as heavy as possible subject to two constraints:

- the centre of mass of the sand (and seesaw) must be at $x = 0$; and
- all the sand must be contained in a region of length $\leq L$ (with $L \geq t$).

$$x = 0 \qquad x = t$$

- Optimal UQ can be seen as the extension of the same basic idea to complicated settings: no hope of a pen-and-paper solution, but can compute a numerical solution.

Optimal UQ — Examples

Example: Balancing a Seesaw

You are given 1kg of sand to arrange however you wish on a seesaw (= the real line). Your challenge is to make the region $x \geq t$, $t \geq 0$, as heavy as possible subject to two constraints:

- the centre of mass of the sand (and seesaw) must be at $x = 0$; and
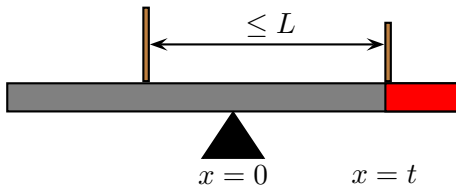- all the sand must be contained in a region of length $\leq L$ (with $L \geq t$).

- Optimal UQ can be seen as the extension of the same basic idea to complicated settings: no hope of a pen-and-paper solution, but can compute a numerical solution.

Optimal UQ — Examples

Example: Balancing a Seesaw

You are given 1kg of sand to arrange however you wish on a seesaw (= the real line). Your challenge is to make the region $x \geq t$, $t \geq 0$, as heavy as possible subject to two constraints:

- the centre of mass of the sand (and seesaw) must be at $x = 0$; and
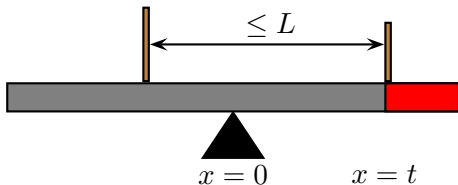- all the sand must be contained in a region of length $\leq L$ (with $L \geq t$).

$$\max = 1 - \frac{t}{L}$$

- Optimal UQ can be seen as the extension of the same basic idea to complicated settings: no hope of a pen-and-paper solution, but can compute a numerical solution.

Reduction of OUQ Problems — LP Analogy

Dimensional Reduction

- A priori, OUQ problems are infinite-dimensional, non-convex*, highly-constrained, global optimization problems.

- However, they can be reduced to equivalent finite-dimensional problems in which the optimization is over the extremal scenarios of $\mathcal{A}$.

- The dimension of the reduced problem is proportional to the number of pieces of information used to describe $\mathcal{A}$.
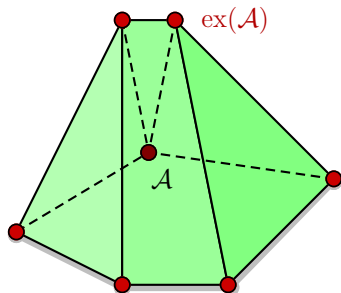

Figure : A linear functional on a convex domain in $\mathbb{R}^n$ finds its extreme value at the extremal points of the domain; similarly, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.

*But see e.g. **Bertsimas & Popescu** (2005) and **Smith** (1995) for convex special cases.

Example: Chebyshev's Inequality in OUQ Form

X has known mean, known variance:

$$\mathcal{A}_{\mathsf{Ch}} := \left\{ \mu \in \mathcal{P}(\mathbb{R}) \,\middle|\, \mathbb{E}_\mu[X] = 0 \text{ and } \mathbb{E}_\mu[X^2] \le \sigma^2 \right\}$$

Least upper bound on probability of deviations of X larger than t:

$$\overline{P}(\mathcal{A}_{\mathsf{Ch}}) := \sup_{\mu \in \mathcal{A}_{\mathsf{Ch}}} \mathbb{P}_\mu\big[|X| \ge t\big] = \frac{\sigma^2}{t^2}.$$

- Can prove this bound using pen and paper or by computing using the reduction theorems (2 constraints $\implies$ enough to consider 3 distinct values for X).
- Optimal UQ can be seen as the extension of the same basic idea to complicated settings: no hope of a pen-and-paper solution, but can compute a numerical solution.

Optimal UQ — Examples

Examples past. . .

- Hypervelocity impact (e.g. micrometeorite on satellite)
  - with simple surrogate model;
  - with full-physics model validated against experiment;
  - with inextensible legacy data set from experiment.
- Seismic safety of an electrical transmission tower
  - with random earthquake in time domain;
  - with random earthquake in frequency domain (Matsuda–Asano mean power spectrum)
- Energy storage placement in power grids (wind generation data).
- Optimal statistical inequalities

. . . and present / near future. . .

- Multilayer graphene composites for photovoltaic applications.
- Performance certification for power generation turbines.
- Assessment of catastrophe risks for insurance and reinsurance.

- Obvious results of OUQ analyses are rigorous bounds on output uncertainties that are sharp with respect to the specified information.
- If you don't like the results, then you need more information!
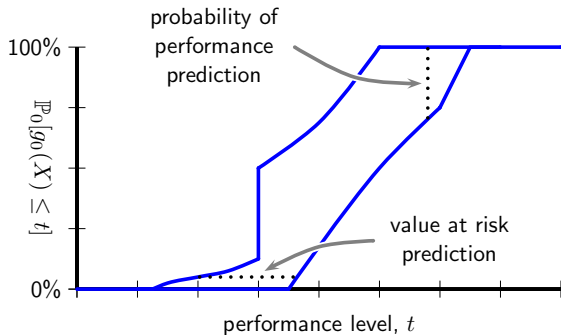
- Abrupt jumps or kinks are common in actual plots of this type: they indicate changeovers among the pieces of information that control the bounds — i.e. the binding constraints — for given t.
- But there are more subtle benefits to this approach...

- Note that the problem formulation and the details of the numerical optimization are separate issues.
- The pieces of information about the unknowns (the constraints) are the central objects.
- By placing information centre-stage, and requiring all assumptions to be stated explicitly, we
  - identify key pieces of information — the ones that control the solution of the optimization problem and make the difference between verdicts of "safe" and "unsafe";
  - are forced to design UQ/optimization frameworks in which constraints are easily perturbed and swapped in/out, as opposed to hard-coded (cf. keynote talk by Mike Hawkins on Thu.);
  - foster the reproducibility of results and open science (cf. keynote talk by Will Schroeder on Wed.).

Overview

Application to Bayesian Inference

- What does the OUQ perspective have to say about Bayesian methods, which are increasingly popular?

- Here, the unknown/variable probability distribution is the prior.

- Are posterior conclusions robust with respect to changes of the prior?

Most statisticians would acknowledge that an analysis is not complete unless the sensitivity of the conclusions to the assumptions is investigated. Yet, in practice, such sensitivity analyses are rarely used. This is because sensitivity analyses involve difficult computations that must often be tailored to the specific problem. This is especially true in Bayesian inference where the computations are already quite difficult.

— L. Wasserman, M. Lavine, and R. L. Wolpert. *J. Statist. Plann. Inference*, **37**(3):307–316, 1993.

Bayesian Models

The ingredients of a Bayesian model:

- The prior distribution π on a parameter space Θ.
- The model class (or likelihood), a function $L \colon \Theta \to \mathcal{P}(\mathcal{D})$.
- Given data $d \in \mathcal{D}$, we update the prior to a posterior by conditioning using Bayes' rule:
$$\pi(\theta|d) \propto L(d|\theta)\pi(\theta).$$
- The application of Bayes' rule to scientific contexts has generated 250 years of controversy, with philosophical and practical objections:

  Twice it has soared to celebrity, twice it has crashed, and it is currently enjoying another boom.
      — B. Efron. *Science* **340**:1177–1178, 2013.

- Notable successes have included the location of the wrecks of *USS Scorpion*, Air France 447, and Nate Silver's correct prediction of the 2012 presidential vote in all 50 states.

Bayesian Misspecification

In (frequentist analyses of) Bayesian statistics, misspecification arises when there is no member of the model class $\{L(\,\cdot\,|\theta) \mid \theta \in \Theta\}$ that agrees precisely with the real data-generating distribution.

> *In practice, Bayesian inference is employed under misspecification*
> *<u>all the time</u>, particularly so in machine learning applications.*
> *While sometimes it works quite well under misspecification, there*
> *are also cases where it does not, so it seems important to*
> *determine precise conditions under which misspecification is*
> *harmful — even if such an analysis is based on frequentist*
> *assumptions.*
>   — P. D. Grünwald. "Bayesian Inconsistency under Misspecification."
> (Emphasis in original.)

- Misspecification has profound consequences for Bayesian robustness
  — in fact, Bayesian inferences become extremely brittle as a function
  of measurement resolution δ.
- If the model is misspecified, and there are possible observed data that
  are arbitrarily unlikely under the model, then under fine enough
  measurement resolution the posterior predictions of nearby priors differ
  as much as possible *regardless of the number of samples observed*.

Figure. As measurement
resolution $\delta \to 0$, the
smooth dependence of the
prior value on the prior
(top-left) shatters into a
patchwork of diametrically
opposed posterior values.
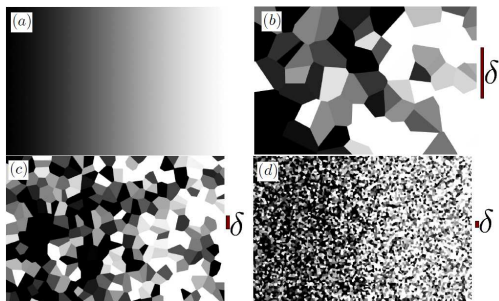
What does this mean in plain terms?

- A regulator specifies some "rules" for device assessors to play by — features that any OK prior must have.
- Consider two rules-compliant prior distributions, e.g. yours and mine.
- The two corresponding posteriors can give arbitrarily different values for the quantity of interest. E.g. "the proposed device is safe", versus "it is unsafe". Whom should we believe?
- This phenomenon persists no matter how close the regulator demands the two priors to be, regardless of the amount of data, and is exacerbated by high-precision data.

What does this mean in plain terms?

- A regulator specifies some "rules" for device assessors to play by — features that any OK prior must have.

- Consider two rules-compliant prior distributions, e.g. yours and mine.

- The two corresponding posteriors can give arbitrarily different values for the quantity of interest. E.g. "the proposed device is safe", versus "it is unsafe". Whom should we believe?

- This phenomenon persists no matter how close the regulator demands the two priors to be, regardless of the amount of data, and is exacerbated by high-precision data.

- The good news is that Bayesian inference is robust when applied to finite and discrete systems.

- However, we should be cautious about applications to continuum systems without further evidence of well-specification, supporting (frequentist) accuracy analysis, and confession of all assumptions.

Overview

How Good is Good Enough?

On a shallow level, "good enough" (or "not too bad") means an understanding of the typical case and the extreme cases.

How Good is Good Enough?

On a shallow level, "good enough" (or "not too bad") means an understanding of the typical case and the extreme cases.

What does the OUQ perspective say about what "good enough" means?

- Explicit statement of requirements / metrics.
- Explicit statement of assumptions and information.
- Implementations that respect the assumptions and are open to adjustment.

How Good is Good Enough?

On a shallow level, "good enough" (or "not too bad") means an understanding of the typical case and the extreme cases.

What does the OUQ perspective say about what "good enough" means?

- Explicit statement of requirements / metrics.
- Explicit statement of assumptions and information.
- Implementations that respect the assumptions and are open to adjustment.

Such frameworks are good because they support

- open, reproducible science;
- meaningful dialogue among experts in the various knowledge domains; and
- robust procedural workflows and software frameworks that are processors converting assumptions/information into conclusions with high (verified, validated, quantified) degree of optimality.

Thank You

- S. Han, U. Topcu, M. Tao, H. Owhadi & R. M. Murray, *Proc. Amer. Control Conf.*, 2013

- H. Owhadi, C. Scovel & T. J. Sullivan, arXiv:1304.6772

- H. Owhadi, C. Scovel & T. J. Sullivan, arXiv:1308.6306

- H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns & M. Ortiz, *SIAM Review* **55**(2):271–345, 2013. arXiv:1009.0679

- T. J. Sullivan, M. McKerns, D. Meyer, F. Theil, H. Owhadi & M. Ortiz, *Math. Model. Numer. Anal.* **47**(6):1657–1689, 2013. arXiv:1202.1928

- OUQ implementation in open-source Mystic optimization framework: http://pythonhosted.org/mystic