An Uncertainty Quantification Framework

Liam Doherty

Drexel University SIAM Chapter

April 20, 2022

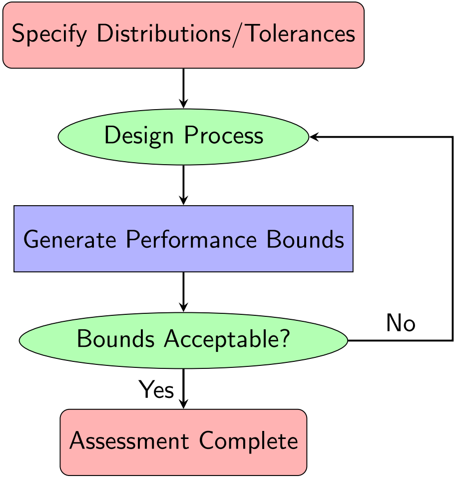
Outline

The Goal

Introduce a framework that:

- Allows for a mathematical separation of epistemic and aleatoric uncertainties when assessing sensitivities of a system,
- Produces bounds on performance measures of a system when determination of the exact parameters of the system is not possible, and
- Is computationally tractable.

An Example: Manufacturing



The Options for Uncertainty Modeling

- Uncertainty
  - Aleatoric
    - Known Distribution (e.g., $X \sim \mathcal{N}(0, 1)$)
  - Epistemic
    - Bounds Only (e.g., $X \in [0, 1]$)
  - Mixed
    - Aleatoric Family (e.g., $X \sim \mathcal{N}(\mu, \sigma)$)

Setup

- Random variables take values in a Polish (complete, separable) metric space $\mathcal{X}$ (often a closed subset of $\mathbb{R}^d$).
- The σ-algebra is the usual Borel σ-algebra.
- The set of probability measures on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$.

Risk-Sensitive Form

We first define the relative entropy (or Kullback-Leibler (KL) divergence):

Definition 1

The *relative entropy* of $\nu \in \mathcal{P}(\mathcal{X})$ with respect to $\mu \in \mathcal{P}(\mathcal{X})$ is given by

$$R(\nu||\mu) = \int_{\mathcal{X}} \log\left(\frac{d\nu}{d\mu}(x)\right)\nu(dx)$$

whenever $\nu \ll \mu$. Otherwise, $R(\nu||\mu) = +\infty$.

This can be interpreted as the information lost when approximating ν with μ.

Risk-Sensitive Form

Now, we define our first risk-sensitive form:

Definition 2

Define for a bounded, continuous function $F : \mathcal{X} \to \mathbb{R}$ and any $c \in (0, \infty)$

$$\Lambda_c \doteq \frac{1}{c} \log \int_{\mathcal{X}} e^{cF(x)} \mu(dx). \tag{1}$$

We also have the variational characterization (given without proof)

$$\Lambda_c = \sup_{\nu \in \mathcal{P}(\mathcal{X})} \left[-\frac{1}{c} R(\nu || \mu) + \int_{\mathcal{X}} F(x) \nu(dx) \right].$$

Interpretation

- $F(x)$ is a performance measure (e.g., a variance or error probability). If F is large, the integral in Λ_c will amplify that (hence "risk-sensitive").
- μ is our *nominal model* ("best guess") at the true distribution ν.

Then the variational characterization gives

$$\int_{\mathcal{X}} F(x)\nu(dx) \leq \Lambda_c + \frac{1}{c}R(\nu||\mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

In plain English, this says that the expected performance has an upper bound depending on μ and the "distance" between ν and μ.

Multiple Types of Uncertainties

- $\mathcal{X}$ - space for random variables with known distribution
- $\mathcal{Y}$ - space for random variables with unknown distribution

The generic performance measure is assumed to be of the form

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x, y)\lambda(dy)\nu(dx),$$

so that, i.e., the known and unknown variables are independent.

Multiple Types of Uncertainties

Letting μ and γ be the nominal distributions of X and Y, we get using the nominal measure $\mu \times \gamma$ and true measure $\nu \times \lambda$ the bound

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x, y)\lambda(dy)\nu(dx) \leq \frac{1}{c} R(\lambda || \gamma) + \frac{1}{c} R(\nu || \mu) + \Lambda_c,$$

where

$$\Lambda_c = \frac{1}{c} \log \int_{\mathcal{X}} \int_{\mathcal{Y}} e^{cF(x,y)} \gamma(dy)\mu(dx). \tag{2}$$

Since $\mathcal{X}$ has known distributions, we can set $\nu = \mu$.

Hybrid Forms

The above performance measure does not distinguish by type of uncertainty, motivating the definition of hybrid forms.

Definition 3

Define

$$\Lambda_c^1 \doteq \frac{1}{c} \log \int_{\mathcal{Y}} e^{\int_{\mathcal{X}} cF(x,y)\mu(dx)} \gamma(dy). \tag{3}$$

Notice by Jensen's inequality, $\Lambda_c^1 \leq \Lambda_c$ in general. The following bound applies (using an appropriate variational characterization):

$$\int_{\mathcal{Y}} \int_{\mathcal{X}} F(x,y)\mu(dx)\theta(dy) \leq \frac{1}{c} R(\theta(dy)||\gamma(dy)) + \Lambda_c^1.$$

Interpretation

Consider the performance measure on epistemic variations

$$G(y) = \int_{\mathcal{X}} F(x, y)\mu(dx)$$

(i.e., replacing F in the original Λ_c definition). Then, we are essentially exploiting the accessibility of μ in this modified performance measure to obtain sharper performance bounds, and the epistemic variation (i.e., random variables in $\mathcal{Y}$) is the primary influence on the bound.

Hybrid Forms

The other hybrid form we consider is useful when epistemic variables are dependent on the values taken by the aleatoric variables.

Definition 4

Define

$$\Lambda_c^2 \doteq \frac{1}{c} \int_{\mathcal{X}} \left[\log \int_{\mathcal{Y}} e^{cF(x,y)} \gamma(dy) \right] \mu(dx). \tag{4}$$

Again by Jensen's inequality, we obtain $\Lambda_c^2 \leq \Lambda_c$. Furthermore, we get under independence that

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x,y) \theta(dy) \mu(dx) \leq \frac{1}{c} R(\theta(dy) || \gamma(dy)) + \Lambda_c^2.$$

Hybrid Forms

We can also modify Λ_c^1 and Λ_c^2 to account for dependence of X on Y or vice versa:

$$\bar{\Lambda}_c^1 = \frac{1}{c} \log \int_{\mathcal{Y}} e^{\int_{\mathcal{X}} cF(x,y)\mu(dx|y)} \gamma(dy)$$

where $\mu(dx|y)$ is the conditonal distribution on X given $Y = y$ (i.e., a *stochastic kernel*), or

$$\bar{\Lambda}_c^2 = \frac{1}{c} \int_{\mathcal{X}} \left[\log \int_{\mathcal{Y}} e^{cF(x,y)} \gamma(dy|x) \right] \mu(dx).$$

Relationship Between Forms

In general, we have

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x, y)\theta(dy)\mu(dx) \leq \Lambda_c^1 \leq \Lambda_c^2 \leq \Lambda_c.$$

There is some technical detail to prove that $\Lambda_c^1 \leq \Lambda_c^2$ in general, but the main takeaway is:
For the sharpest performance measures in situations where the epistemic variations are independent of the aleatoric variations, Λ_c^1 is the best bound.

A Bird's Eye View

The following three results say, essentially, that

- We can obtain bounds for performance over a family of distributions within a prescribed ball of the truth;
- The bounds we obtain are, in some sense, sharp;
- If all we know about the epistemic variables are bounds, then $\Lambda^1_\infty = \lim_{c\to\infty} \Lambda^1_c$ is the tightest possible bound.

Bounding Performance

Theorem 5

Consider the functionals (2) - (4) and let $D = \{c : \Lambda_c < \infty\}$ *(resp.,* $D^i = \{c : \Lambda_c^i < \infty\}$, $i = 1, 2$*). Assume that the interior of* D *(resp.,* D^i*) is nonempty. Then* Λ_c *(resp.,* Λ_c^i*) is differentiable on the interior of* D *(resp.,* D^i*). Assume that* $F(x) \geq 0$*. Then* $c \mapsto \Lambda_c$ *(resp.,* $c \mapsto \Lambda_c^i$*) is nondecreasing for* $c \geq 0$*. Let* $B > 0$ *be given. Then there is a unique* $c \in (0, \infty]$ *at which*

$$c \mapsto \frac{1}{c}B + \Lambda_c \quad \left(\text{resp., } c \mapsto \frac{1}{c}B + \Lambda_c^i\right)$$

attains a local minimum, where the statement that the minimum occurs at $c = \infty$ *means that* $\Lambda_c + B/c > \Lambda_\infty$ *for a well-defined limit* Λ_∞ *and all* $c < \infty$*.*

Interpretation

- All of this can be wrapped up into a nice package: *Within a given tolerance of the truth, there is a unique, optimal bound on the expected performance.*

We get to choose our relative entropy radius B. This allows us to reflect in our computation the degree of certainty we are about "closeness" in our model. In essence, with little data we can still produce bounds by relaxing our radius. As $B \searrow 0$ (i.e., in the "data rich" regime), we are optimizing Λ_c, which is nondecreasing. The optimum is then simply the performance under the nominal model!

Sharpness of the Bound

Theorem 6

Consider the functional (1) and assume the conditions of Theorem 5. Suppose that $L \in \mathbb{R}$ *and* $B \in [0, \infty)$ *are given, and that* c^* *minimizes*

$$c \mapsto \frac{1}{c} B + \Lambda_c$$

so that $\frac{1}{c^*} B + \Lambda_{c^*} < \infty$. *Define* $\mathcal{F}_B \doteq \{\gamma \in \mathbb{P}(\mathcal{X}) : R(\gamma || \mu) \leq B\}$, *the family of alternative distributions on* $\mathcal{X}$. *Then for all* $\nu \in \mathcal{F}_B$,

$$\int_{\mathcal{X}} F(x) \nu(dx) \leq L \iff \frac{1}{c^*} B + \Lambda_{c^*} \leq L.$$

Interpretation

Here, L plays the role of a criterion for sufficiency in the underlying process (e.g., if the expected performance is worse than L, we must re-engineer our manufacturing pipeline to meet the regulation). This theorem says that
The true expected performance will satisfy the criterion if and only if the upper bound for the entire family of alternative measures does as well.
In essence, if the upper bound fails the criteria, we are, in some sense, "close" to failing the criteria and either need to improve our process or relax our tolerances.

The Best Bound for Least Information

Theorem 7

Let $\mathcal{X}$ and $\mathcal{Y}$ be subsets of a finite dimensional Euclidean space. Suppose that $A \subset \mathcal{Y}$ is bounded and the closure of its interior and that γ is the uniform measure on A. Assume that F is lower semicontinuous in y for each $x \in \mathcal{X}$ and bounded from below. Consider the risk-sensitive functional Λ_c^1 and let $\Lambda_\infty^1 = \lim_{c\to\infty} \Lambda_c^1$. Then

$$\sup_{y\in A} \int_{\mathcal{X}} F(x,y)\mu(dx) = \Lambda_\infty^1.$$

Interpretation

- The set A being bounded is essentially saying that all we know about the epistemic variables (i.e., those in $\mathcal{Y}$) are bounds, and the uniform distribution is to say that we are not imposing any further structure on the uncertainties.
- This theorem could be extended to unbounded A as long as the accompanying distribution γ is appropriate (e.g., $A = [0, \infty)$ with γ exponential).

Toy Problem Setup

Consider the dynamical system

$$\frac{d}{dt}u(t) = -z_1 u(t), \quad u(0) = z_2,$$

where $Z_1 \sim U[0,1]$ and $Z_2 \sim U[0,1]$. Our goal is to obtain bounds on the second moment of the solution at time $t = 1$; i.e.,

$$F(Z_1, Z_2) = (u(1; z_1, z_2))^2.$$

Results

Approximations are using gPC representations of F with basis degree 7. "True" solutions are generated with direct numerical quadrature.

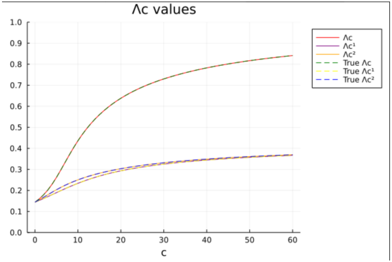


Figure: Approximated Λ_c

Results

Based on a true distribution $\theta(dz_2) \sim \text{beta}(\alpha = 1.5, \beta = 1.5)$. This implies $B = R(\theta || U[0,1]) \approx 0.0484$. Under this true distribution for z_2, with $Z_1 \sim U[0,1]$ the truth, the actual performance measure is approximately 0.13541.

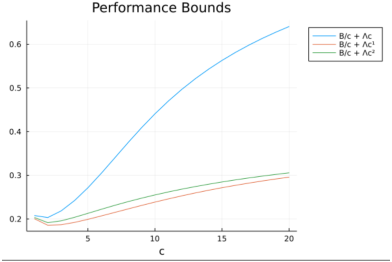


Figure: Approximate Bounds on Performance

Interpretations

- Indeed, we get the desired bound $\Lambda_c^1 \leq \Lambda_c^2 \leq \Lambda_c$.
- It is clear from the second plot that there is a minimizer of $\frac{1}{c}B + \Lambda_c^i$ for each i, consistent with the statement of Theorem 5.
- We can combine this bound of our second moment with an estimate of the mean solution at time $t = 1$ to obtain a bound on the variance of the solution.

Future Directions for Research

A small modification on the variational characterization of (1) gives a bound of the form

$$\mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \frac{1}{c} \log \mathbb{E}_P\left[e^{c(f - \mathbb{E}_P[f])}\right] + \frac{1}{c} R(Q||P) \doteq \Xi_+(c)$$

A similar form for a lower bound Ξ_- holds.
Question: How can we obtain tight bounds over the performance using an approximating distribution for a particular QoI?
Essentially, how can we *practically* estimate and optimize Ξ_+?

- Weighted Ensemble (WE) for variance reduction
- Qualitative properties of Ξ_+: Can we reduce search space?

Thank you!