Methods for Model Approximation and Optimization in the Presence of Model Uncertainty Using Information Divergences

Paul Dupuis

Division of Applied Mathematics
Brown University

R. Atar (Technion), K. Chowdhary (Sandia NL), M. Katsoulakis (Massachusetts), Y. Pantazis (Crete), P. Plecháč (Delaware), L. Rey-Bellet (Massachusetts)

Supported in part by NSF, DOE, DARPA, AFOSR

3rd GAMM ACUQ Workshop on UQ, March 2018

Example

Chemical reaction network: Population numbers of distinct species which interact as reactants to form products, modeled as continuous time jump Markov process.

Example

Chemical reaction network: Population numbers of distinct species which interact as reactants to form products, modeled as continuous time jump Markov process. Small example:

Process	transition rate	population change
Inflow species A	κ_1	$X_A \to X_A + 1$
Outflow species A	$\kappa_2 X_A$	$X_A \to X_A - 1$
Inflow species B	κ_3	$X_B \to X_B + 1$
Outflow species B	$\kappa_4 X_B$	$X_B \to X_B - 1$
Reaction $A + 2B \to 3B$	$\kappa_5 X_A X_B (X_B - 1)$	$X_A \to X_A - 1$ $X_B \to X_B + 1$

Example

Chemical reaction network: Population numbers of distinct species which interact as reactants to form products, modeled as continuous time jump Markov process. Small example:

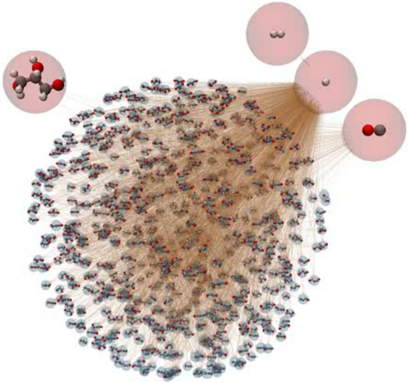
Process	transition rate	population change
Inflow species A	κ_1	$X_A \to X_A + 1$
Outflow species A	$\kappa_2 X_A$	$X_A \to X_A - 1$
Inflow species B	κ_3	$X_B \to X_B + 1$
Outflow species B	$\kappa_4 X_B$	$X_B \to X_B - 1$
Reaction $A + 2B \to 3B$	$\kappa_5 X_A X_B (X_B - 1)$	$X_A \to X_A - 1$ $X_B \to X_B + 1$

Model determined by reaction rates κ_i. Generator

$$\mathcal{L} f(x) = \sum_{i=1}^{5} r_i(x) \left[f(x + v_i) - f(x) \right]$$

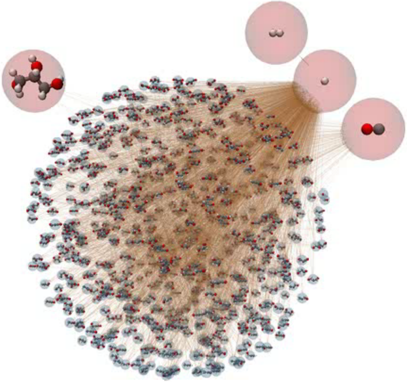
Example

Glycerol reforming CRN. Network visualization:



Example

Glycerol reforming CRN. Network visualization:



780 species, 3898 reactions. Quantities of interest (performance measures) are functionals of stationary distribution (ergodic averages).

Problems we want to solve

- *Sensitivity bounds* (*insensitivity* analysis, model simplification): identify parameters that are *not* important,

Problems we want to solve

- *Sensitivity bounds* (*insensitivity* analysis, model simplification): identify parameters that are *not* important,
- *Predictive uncertainty quantification* (error guarantees): tight bounds on performance measures for $O(1)$ size model errors,

Problems we want to solve

- *Sensitivity bounds* (*insensitivity* analysis, model simplification): identify parameters that are *not* important,
- *Predictive uncertainty quantification* (error guarantees): tight bounds on performance measures for $O(1)$ size model errors,
- *Optimization and control for uncertain systems*: optimize a performance measure for a family of models, a min/max problem

Problems we want to solve

- *Sensitivity bounds* (*insensitivity* analysis, model simplification): identify parameters that are *not* important,
- *Predictive uncertainty quantification* (error guarantees): tight bounds on performance measures for $O(1)$ size model errors,
- *Optimization and control for uncertain systems*: optimize a performance measure for a family of models, a min/max problem
- *Model reduction* (e.g., course graining).

Abstract formulation

Elements of the framework

- *Probability models*, on space $\mathcal{S}$, often a path space

  $P =$ nominal (computational, design) vs $Q =$ true (impractical)

Abstract formulation

Elements of the framework

- *Probability models*, on space $\mathcal{S}$, often a path space

  $P =$ nominal (computational, design) vs $Q =$ true (impractical)

- *Quantities of interest* (QoI, also called observables, costs, performance measures, ...), and methods to bound QoI above and below. E.g.,

$E_Q[f]$	expected value of QoI f under true
$\log E_Q[e^{cf}]$	risk-sensitive QoI
$\mathrm{Var}_Q[f]$	variance of QoI f

Abstract formulation

Elements of the framework

- *Probability models*, on space $\mathcal{S}$, often a path space

  $P =$ nominal (computational, design) vs $Q =$ true (impractical)

- *Quantities of interest* (QoI, also called observables, costs, performance measures, ...), and methods to bound QoI above and below. E.g.,

$E_Q[f]$	expected value of QoI f under true
$\log E_Q[e^{cf}]$	risk-sensitive QoI
$\mathrm{Var}_Q[f]$	variance of QoI f

  Here f may combine a cost with dynamics that take random variables under Q (or P) into the system state:

$$f(w) = \int_0^T c(\mathcal{G}[w](t))dt,$$

$$\mathcal{G} : W \to X, \quad dX(t) = b(X(t))dt + dW(t).$$

Abstract formulation

Ingredients in the framework

- *Divergence or distance* between nominal and true models. Most important (but not only) example is relative entropy, aka Kullback-Leibler divergence:

$$R(Q \| P) = \begin{cases} E_Q\left[\log \frac{dQ}{dP}\right] & \text{if } Q \ll P \\ \infty & \text{else.} \end{cases}$$

Abstract formulation

Ingredients in the framework

- *Divergence or distance* between nominal and true models. Most important (but not only) example is relative entropy, aka Kullback-Leibler divergence:

$$R(Q\|P) = \begin{cases} E_Q\left[\log \frac{dQ}{dP}\right] & \text{if } Q \ll P \\ \infty & \text{else.} \end{cases}$$

  $R(\cdot\|\cdot)$ is jointly convex and lsc, $R(Q\|P) \geq 0$ and $= 0$ iff $Q = P$.

Abstract formulation

Ingredients in the framework

- *Divergence or distance* between nominal and true models. Most important (but not only) example is relative entropy, aka Kullback-Leibler divergence:

$$R(Q \| P) = \begin{cases} E_Q\left[\log \frac{dQ}{dP}\right] & \text{if } Q \ll P \\ \infty & \text{else.} \end{cases}$$

  $R(\cdot \| \cdot)$ is jointly convex and lsc, $R(Q \| P) \geq 0$ and $= 0$ iff $Q = P$.
- *Variational formula* relating QoI under Q with (hopefully tractable) functional of P. E.g.,

$$\log E_P\left[e^{cf}\right] = \sup_{Q \ll P} \left[cE_Q[f] - R(Q \| P)\right],$$

Abstract formulation

Ingredients in the framework

- *Divergence or distance* between nominal and true models. Most important (but not only) example is relative entropy, aka Kullback-Leibler divergence:

$$R(Q \| P) = \begin{cases} E_Q\left[\log \frac{dQ}{dP}\right] & \text{if } Q \ll P \\ \infty & \text{else.} \end{cases}$$

  $R(\cdot \| \cdot)$ is jointly convex and lsc, $R(Q \| P) \geq 0$ and $= 0$ iff $Q = P$.
- *Variational formula* relating QoI under Q with (hopefully tractable) functional of P. E.g.,

$$\log E_P\left[e^{cf}\right] = \sup_{Q \ll P}\left[cE_Q[f] - R(Q \| P)\right],$$

  hence whenever $Q \ll P$,

$$cE_Q[f] \leq R(Q \| P) + \log E_P\left[e^{cf}\right].$$

  Minimizing Q^* is $dQ^* = e^{cf}dP / \int e^{cf}dP$.

Abstract formulation

We will use such variational formulas for sensitivity bounds, optimization of uncertain systems, etc.

Abstract formulation

We will use such variational formulas for sensitivity bounds, optimization of uncertain systems, etc.

Example. Suppose $f = f_\alpha$ with $\alpha \in A$ and we want to solve "optimally robust optimization":

$$\min_{\alpha \in A} \max_{Q : R(Q \| P) \leq r} E_Q[f_\alpha].$$

Abstract formulation

We will use such variational formulas for sensitivity bounds, optimization of uncertain systems, etc.

Example. Suppose $f = f_\alpha$ with $\alpha \in A$ and we want to solve "optimally robust optimization":

$$\min_{\alpha \in A} \max_{Q : R(Q \| P) \le r} E_Q[f_\alpha].$$

Then using Lagrange multipliers ($\lambda = 1/c$)

$$\begin{aligned}
&\min_{\alpha \in A} \left[\max_{Q} \min_{c>0} \left(E_Q[f_\alpha] + \frac{1}{c}[r - R(Q \| P)] \right) \right] \\
&\quad = \min_{\alpha \in A} \left[\min_{c>0} \max_{Q} \left(E_Q[f_\alpha] - \frac{1}{c} R(Q \| P) \right) + \frac{1}{c} r \right] \\
&\quad = \min_{\alpha \in A} \min_{c>0} \frac{1}{c} \left(r + \log E_P \left[e^{c f_\alpha} \right] \right).
\end{aligned}$$

Final problem phrased purely in terms of the design model.

Important properties of a good divergence

Recall variational bound based on KL divergence:

$$cE_Q[f] \leq R\left(Q \,\|\, P\right) + \log E_P\left[e^{cf}\right]$$

Important properties of a good divergence

Recall variational bound based on KL divergence:

$$cE_Q[f] \leq R\left(Q \,\|\, P\right) + \log E_P\left[e^{cf}\right]$$

Important properties of a divergence variational bound pair

- Good scaling properties, e.g.:
  - the bounds remain meaningful for systems over long times, ergodic QoI
  - also remain useful for systems with many degrees of freedom (e.g., number of particles in an interacting particle system)

Important properties of a good divergence

Recall variational bound based on KL divergence:

$$cE_Q[f] \leq R\left(Q \,\|\, P\right) + \log E_P\left[e^{cf}\right]$$

Important properties of a divergence variational bound pair

- Good scaling properties, e.g.:
  - the bounds remain meaningful for systems over long times, ergodic QoI
  - also remain useful for systems with many degrees of freedom (e.g., number of particles in an interacting particle system)
- Good computational properties of the divergence and the dual functional under the design P.

Important properties of a good divergence

Recall variational bound based on KL divergence:

$$cE_Q[f] \leq R\left(Q \,\|\, P\right) + \log E_P\left[e^{cf}\right]$$

Important properties of a divergence variational bound pair

- Good scaling properties, e.g.:
  - the bounds remain meaningful for systems over long times, ergodic QoI
  - also remain useful for systems with many degrees of freedom (e.g., number of particles in an interacting particle system)
- Good computational properties of the divergence and the dual functional under the design P.
- Bounds should not depend on the underlying probability space (e.g., for $dX = b(X)dt + dW$ can have P is Wiener measure vs P measure induced by X).

A little history

First applications of relative entropy/exponential integral duality in problems of model uncertainty:

- Robust properties of risk–sensitive control (D, James and Petersen), *Math. of Control, Signals and Systems*, **13,** (2000), pp. 318–332.

This paper introduces the general framework in the setting of diffusions with uncertain drift, and makes connections with H^∞ control.

- Minimax optimal control of stochastic uncertain systems with relative entropy constraints (Petersen, James and D), *IEEE Trans. on Auto. Control.,* **45,** (2000), pp. 398–412.

Solution for particular classes of models including uncertain linear/quadratic.

A little history

First applications of relative entropy/exponential integral duality in problems of model uncertainty:

- Robust properties of risk–sensitive control (D, James and Petersen), *Math. of Control, Signals and Systems*, **13,** (2000), pp. 318–332.

This paper introduces the general framework in the setting of diffusions with uncertain drift, and makes connections with H^∞ control.

- Minimax optimal control of stochastic uncertain systems with relative entropy constraints (Petersen, James and D), *IEEE Trans. on Auto. Control.,* **45,** (2000), pp. 398–412.

Solution for particular classes of models including uncertain linear/quadratic.

Ideas had substantial impact in economics:

- Robust control and model uncertainty (Hansen and Sargent), *The American Economic Review*, **91**, (2001), pp. 60-66.
- *Robustness*, (Hansen and Sargent), Wiley, 2008.
- Many subsequent papers.

Outline

- Uses of KL divergence and standard QoI
    - Why bounds do not depend on how system is represented
    - Performance bounds for static problem
    - Chain rule for KL divergence
    - Performance bounds for large time (ergodic) problem
    - Sensitivity bounds for ergodic problem
    - Robust optimization for ergodic problem

Outline

- Uses of KL divergence and standard QoI
  - Why bounds do not depend on how system is represented
  - Performance bounds for static problem
  - Chain rule for KL divergence
  - Performance bounds for large time (ergodic) problem
  - Sensitivity bounds for ergodic problem
  - Robust optimization for ergodic problem
- Nonstandard QoI and other distance/variational representation pairs
  - Situations that require new ideas/open problems
  - Rare event performance measures and Rényi divergence

Why bounds do not depend on how system is represented

Relative entropy under mapping. Suppose $\psi : \mathcal{S} \to \mathcal{T}$ and

$$\lambda(A) = P(\{x : \psi(x) \in A\}), \quad \sigma(A) = Q(\{x : \psi(x) \in A\}).$$

Then $R(Q \| P) \geq R(\sigma \| \lambda)$. Given feasible σ there is always a Q^* such that $\psi \sim \sigma$ under Q^*, and $R(Q^* \| P) = R(\sigma \| \lambda)$. In particular, take

$$\frac{dQ^*}{dP}(x) = \frac{d\sigma}{d\lambda}(y) \text{ on } \{x : \psi(x) = y\}.$$

Why bounds do not depend on how system is represented

Relative entropy under mapping. Suppose $\psi : \mathcal{S} \to \mathcal{T}$ and

$$\lambda(A) = P(\{x : \psi(x) \in A\}), \quad \sigma(A) = Q(\{x : \psi(x) \in A\}).$$

Then $R(Q \,\|P) \geq R(\sigma \,\|\lambda)$. Given feasible σ there is always a Q^* such that $\psi \sim \sigma$ under Q^*, and $R(Q^* \,\|P) = R(\sigma \,\|\lambda)$. In particular, take

$$\frac{dQ^*}{dP}(x) = \frac{d\sigma}{d\lambda}(y) \text{ on } \{x : \psi(x) = y\}.$$

Now suppose view $f(y)$ under σ or $f(\psi(x))$ under Q. Performance bounds on f are *same* for a given allowed relative entropy distance:

$$\sup_{Q:R(Q\|P)\leq r} E_Q[f \circ \psi] = \sup_{\sigma:R(\sigma\|\lambda)\leq r} E_\sigma[f].$$

Performance bounds for static problem

For a given QoI f, define centered cumulant generating function

$$\tilde{\Lambda}_{P,f}(c) = \log \int_{\mathcal{S}} e^{c(f(x) - E_P[f])} P(dx).$$

Performance bounds for static problem

For a given QoI f, define centered cumulant generating function

$$\tilde{\Lambda}_{P,f}(c) = \log \int_{\mathcal{S}} e^{c(f(x) - E_P[f])} P(dx).$$

Then the variational inequality becomes

$$E_Q[f] - E_P[f] \leq \frac{1}{c} R\left(Q \,\|\, P\right) + \frac{1}{c} \tilde{\Lambda}_{P,f}(c),$$

Performance bounds for static problem

For a given QoI f, define centered cumulant generating function

$$\tilde{\Lambda}_{P,f}(c) = \log \int_{\mathcal{S}} e^{c(f(x)-E_P[f])} P(dx).$$

Then the variational inequality becomes

$$E_Q[f] - E_P[f] \leq \frac{1}{c} R\left(Q \,\|\, P\right) + \frac{1}{c} \tilde{\Lambda}_{P,f}(c),$$

Also if $c > 0$ then

$$-cE_Q[f] \leq R\left(Q \,\|\, P\right) + \log E_P\left[e^{-cf}\right]$$

so for any $c > 0$

$$-\frac{1}{c}\tilde{\Lambda}_{P,f}(-c) - \frac{1}{c} R\left(Q \,\|\, P\right) \leq E_Q[f] - E_P[f] \leq \frac{1}{c}\tilde{\Lambda}_{P,f}(c) + \frac{1}{c} R\left(Q \,\|\, P\right).$$

Performance bounds for static problem

Optimizing the two bounds gives

$$\sup_{c>0}\left\{-\frac{1}{c}\tilde{\Lambda}_{P,f}(-c)-\frac{1}{c}R\left(Q\,\|P\right)\right\} \leq E_Q[f]-E_P[f]$$
$$\leq \inf_{c>0}\left\{\frac{1}{c}\tilde{\Lambda}_{P,f}(c)+\frac{1}{c}R\left(Q\,\|P\right)\right\}.$$

Performance bounds for static problem

Optimizing the two bounds gives

$$\begin{aligned}\sup_{c>0}\left\{-\frac{1}{c}\tilde{\Lambda}_{P,f}(-c)-\frac{1}{c}R\left(Q\,\|P\right)\right\} &\leq E_Q[f]-E_P[f]\\ &\leq \inf_{c>0}\left\{\frac{1}{c}\tilde{\Lambda}_{P,f}(c)+\frac{1}{c}R\left(Q\,\|P\right)\right\}.\end{aligned}$$

If f is not a constant the quantities

$$\Xi_{\pm}(Q\,\|P\,;f)=\inf_{c>0}\left\{\frac{R\left(Q\,\|P\right)+\tilde{\Lambda}_{P,f}(\pm c)}{c}\right\}$$

behave themselves like divergences: $\Xi_{\pm}(Q\,\|P\,;f)\geq 0$ and $\Xi_{\pm}(Q\,\|P\,;f)=0$ iff $Q=P$. Optimization problem has good properties.

Performance bounds for static problem

Optimizing the two bounds gives

$$\sup_{c>0}\left\{-\frac{1}{c}\tilde{\Lambda}_{P,f}(-c)-\frac{1}{c}R\left(Q\,\|P\right)\right\} \le E_Q[f]-E_P[f]$$
$$\le \inf_{c>0}\left\{\frac{1}{c}\tilde{\Lambda}_{P,f}(c)+\frac{1}{c}R\left(Q\,\|P\right)\right\}.$$

If f is not a constant the quantities

$$\Xi_{\pm}(Q\,\|P\,;f)=\inf_{c>0}\left\{\frac{R\left(Q\,\|P\right)+\tilde{\Lambda}_{P,f}(\pm c)}{c}\right\}$$

behave themselves like divergences: $\Xi_{\pm}(Q\,\|P\,;f)\ge 0$ and $\Xi_{\pm}(Q\,\|P\,;f)=0$ iff $Q=P$. Optimization problem has good properties.

How to scale up to ergodic problem?

Properties of relative entropy/KL divergence

Chain rule and Markov models. For the bounds to be useful, the quantities involved should scale and take concrete form. For Markov systems, done (in discrete time) via the *chain rule*. (In continuous time, for systems driven by Brownian and Poisson noise, see*.)

---

* *Variational representations for continuous time processes*, Budhiraja, Dupuis and Maroulas, 2011.

Properties of relative entropy/KL divergence

Chain rule and Markov models. For the bounds to be useful, the quantities involved should scale and take concrete form. For Markov systems, done (in discrete time) via the *chain rule*. (In continuous time, for systems driven by Brownian and Poisson noise, see*.) Suppose $Q, P \in \mathcal{P}(\mathcal{X}^2)$ with decompositions

$$\begin{aligned} Q(dx_0 \times dx_1) &= \nu(dx_0) q(x_0, dx_1), \\ P(dx_0 \times dx_1) &= \mu(dx_0) p(x_0, dx_1). \end{aligned}$$

---

* *Variational representations for continuous time processes*, Budhiraja, Dupuis and Maroulas, 2011.

Properties of relative entropy/KL divergence

Chain rule and Markov models. For the bounds to be useful, the quantities involved should scale and take concrete form. For Markov systems, done (in discrete time) via the *chain rule*. (In continuous time, for systems driven by Brownian and Poisson noise, see*.) Suppose $Q, P \in \mathcal{P}(\mathcal{X}^2)$ with decompositions

$$\begin{aligned} Q(dx_0 \times dx_1) &= \nu(dx_0)q(x_0, dx_1), \\ P(dx_0 \times dx_1) &= \mu(dx_0)p(x_0, dx_1). \end{aligned}$$

Then relative entropy decomposes as

$$\begin{aligned} R\left(Q \,\|P\right) &= \int_{\mathcal{X}^2} \left[\log\left(\frac{d\nu}{d\mu}(x_0)\right) + \log\left(\frac{dq(x_0,\cdot)}{dp(x_0,\cdot)}(x_1)\right)\right] \nu(dx_0)q(x_0, dx_1) \\ &= R\left(\nu \,\|\mu\right) + \int_{\mathcal{X}} R\left(q(x_0,\cdot) \,\|p(x_0,\cdot)\right) \nu(dx_0) \\ &= R\left(\nu \,\|\mu\right) + E_Q R\left(q(X_0,\cdot) \,\|p(X_0,\cdot)\right) \end{aligned}$$

---

* *Variational representations for continuous time processes*, Budhiraja, Dupuis and Maroulas, 2011.

Properties of relative entropy/KL divergence

Chain rule and Markov models. Now let P_T be a Markov measure on $\{0, 1, \ldots, T\}$ with initial measure μ and transition kernel $p(x, dy)$, and similarly for Q_T with ν and $q(x, dy)$. Then repeated use of chain rule gives

$$R\left(Q_T \,\|P_T\right) = R\left(\nu \,\|\mu\right) + \sum_{t=0}^{T-1} E_{Q_T} R\left(q(X_t, \cdot) \,\|p(X_t, \cdot)\right),$$

where E_{Q_T} means expectation where $(X_0, X_1, \ldots, X_T)$ has distribution Q_T.

Properties of relative entropy/KL divergence

Chain rule and Markov models. Now let P_T be a Markov measure on $\{0, 1, \ldots, T\}$ with initial measure μ and transition kernel $p(x, dy)$, and similarly for Q_T with ν and $q(x, dy)$. Then repeated use of chain rule gives

$$R\left(Q_T \,\|\, P_T\right) = R\left(\nu \,\|\, \mu\right) + \sum_{t=0}^{T-1} E_{Q_T} R\left(q(X_t, \cdot) \,\|\, p(X_t, \cdot)\right),$$

where E_{Q_T} means expectation where $(X_0, X_1, \ldots, X_T)$ has distribution Q_T. If $p(x, dy)$ is ergodic with stationary distribution π_p and $q(x, dy)$ with π_q, then

$$r(q \,\|\, p) = \lim_{T \to \infty} \frac{1}{T} R\left(Q_T \,\|\, P_T\right) = \int_{\mathcal{X}} R\left(q(x, \cdot) \,\|\, p(x, \cdot)\right) \pi_q(dx)$$

is called the *relative entropy rate (RER)*.

Performance bounds for large time (ergodic) problem

Suppose the QoI is a functional of the empirical measure, e.g.,

$$f(\{X_t\}_{t=0}^{T-1}) = \frac{1}{T}\sum_{t=0}^{T-1} g(X_t).$$

Performance bounds for large time (ergodic) problem

Suppose the QoI is a functional of the empirical measure, e.g.,

$$f(\{X_t\}_{t=0}^{T-1}) = \frac{1}{T}\sum_{t=0}^{T-1} g(X_t).$$

Then

$$\begin{aligned} -\frac{1}{c}\frac{1}{T}\tilde{\Lambda}_{P_T,f}(-c) - \frac{1}{c}\frac{1}{T}R\left(Q_T \,\|\, P_T\right) &\leq \frac{1}{T}E_{Q_T}\sum_{t=0}^{T-1} g(X_t) - \frac{1}{T}E_{P_T}\sum_{t=0}^{T-1} g(X_t) \\ &\leq \frac{1}{c}\frac{1}{T}\tilde{\Lambda}_{P_T,f}(c) + \frac{1}{c}\frac{1}{T}R\left(Q_T \,\|\, P_T\right). \end{aligned}$$

Performance bounds for large time (ergodic) problem

Suppose the QoI is a functional of the empirical measure, e.g.,

$$f(\{X_t\}_{t=0}^{T-1}) = \frac{1}{T}\sum_{t=0}^{T-1} g(X_t).$$

Then

$$\begin{aligned}
-\frac{1}{c}\frac{1}{T}\tilde{\Lambda}_{P_T,f}(-c) - \frac{1}{c}\frac{1}{T}R\left(Q_T \,\|\, P_T\right) &\leq \frac{1}{T}E_{Q_T}\sum_{t=0}^{T-1} g(X_t) - \frac{1}{T}E_{P_T}\sum_{t=0}^{T-1} g(X_t) \\
&\leq \frac{1}{c}\frac{1}{T}\tilde{\Lambda}_{P_T,f}(c) + \frac{1}{c}\frac{1}{T}R\left(Q_T \,\|\, P_T\right).
\end{aligned}$$

By chain rule all terms scale $\propto T$. Send $T \to \infty$ to get

$$-\frac{1}{c}\Lambda(-c) - \frac{1}{c}r(q\,\|\,p) \leq E_{\pi_q}[g] - E_{\pi_p}[g] \leq \frac{1}{c}\Lambda(c) + \frac{1}{c}r(q\,\|\,p),$$

$$\Lambda(c) = \lim_{T\to\infty}\frac{1}{T}\log E_{P_T}\left[e^{c\sum_{t=0}^{T-1} g(X_t) - E_{\pi_p}[g]}\right]$$

Performance bounds for large time (ergodic) problem

$$\Lambda(c) = \lim_{T\to\infty} \frac{1}{T} \log E_{P_T}\left[e^{c\sum_{t=0}^{T-1} g(X_t) - E_{\pi_p}[g]}\right]$$

is a centered cumulant generating function on path space.

Performance bounds for large time (ergodic) problem

$$\Lambda(c) = \lim_{T\to\infty} \frac{1}{T} \log E_{P_T} \left[e^{c \sum_{t=0}^{T-1} g(X_t) - E_{\pi_p}[g]} \right]$$

is a centered cumulant generating function on path space.
"Small model error" version of bounds useful and frequently tractable.
Optimizing bounds over $c > 0$ (separately LHS and RHS) in

$$-\frac{1}{c}\Lambda(-c) - \frac{1}{c} r(q \,\|\, p) \leq E_{\pi_q}[g] - E_{\pi_p}[g] \leq \frac{1}{c}\Lambda(c) + \frac{1}{c} r(q \,\|\, p)$$

gives

$$\left| E_{\pi_q}[g] - E_{\pi_p}[g] \right| \leq \sqrt{v_p[g] 2 r(q \,\|\, p)} + O(r(q \,\|\, p))$$

$$v_p[g] = 2 \sum_{t=0}^{\infty} E_p[\left(g(X_t) - E_{\pi_p}[g]\right)\left(g(X_0) - E_{\pi_p}[g]\right)].$$

Sensitivity bounds for ergodic problem

For sensitivity bounds we assume q is a small perturbation of p. Thus

$$p(x, dy) = p(x, dy; \theta) \text{ and } q(x, dy) = p(x, dy; \theta + \varepsilon v),$$

where the parameter θ takes values in a subset of some Euclidean space.

Sensitivity bounds for ergodic problem

For sensitivity bounds we assume q is a small perturbation of p. Thus

$$p(x, dy) = p(x, dy; \theta) \text{ and } q(x, dy) = p(x, dy; \theta + \varepsilon v),$$

where the parameter θ takes values in a subset of some Euclidean space. Assume reference measure so

$$p(x, dy; \theta) = p(x, y; \theta) R(dy).$$

Sensitivity bounds for ergodic problem

For sensitivity bounds we assume q is a small perturbation of p. Thus

$$p(x, dy) = p(x, dy; \theta) \text{ and } q(x, dy) = p(x, dy; \theta + \varepsilon v),$$

where the parameter θ takes values in a subset of some Euclidean space. Assume reference measure so

$$p(x, dy; \theta) = p(x, y; \theta) R(dy).$$

Then one can expand the relative entropy rate:

$$\begin{aligned} r(q \,\|\, p) &= \int_{\mathcal{X}} R\left(q(x, \cdot) \,\|\, p(x, \cdot)\right) \pi_q(dx) \\ &= \int_{\mathcal{X}} \int_{\mathcal{X}} \log \frac{p(x, y; \theta + \varepsilon v)}{p(x, y; \theta)} p(x, y; \theta + \varepsilon v) R(dy) \pi_q(dx) \\ &= \frac{|\varepsilon|}{2} \langle v, \mathcal{I}(p(\theta)) v \rangle + O(\varepsilon^2), \end{aligned}$$

where

$$\mathcal{I}(p(\theta)) = \int_{\mathcal{X}^2} \nabla_\theta \log p(x, y; \theta) [\nabla_\theta \log p(x, y; \theta)]^T p(x, y; \theta) R(dy) \pi_p(dx)$$

Sensitivity bounds for ergodic problem

$$\mathcal{I}(p(\theta)) = \int_{\mathcal{X}^2} \nabla_\theta \log p(x,y;\theta)[\nabla_\theta \log p(x,y;\theta)]^T p(x,y;\theta) R(dy) \pi_p(dx)$$

is called a *path Fisher Information Matrix*.

Sensitivity bounds for ergodic problem

$$\mathcal{I}(p(\theta)) = \int_{\mathcal{X}^2} \nabla_\theta \log p(x,y;\theta)[\nabla_\theta \log p(x,y;\theta)]^T p(x,y;\theta) R(dy) \pi_p(dx)$$

is called a *path Fisher Information Matrix*. From

$$\left| E_{\pi_q}[g] - E_{\pi_p}[g] \right| \leq \sqrt{v_p[g] 2 r(q \,\|\, p)} + O(r(q \,\|\, p)),$$

and expansion of RER, for any QoI g one has bounds

$$\frac{1}{|\varepsilon|} \left| E_{\pi_{p(\theta+\varepsilon v)}}[g] - E_{\pi_{p(\theta)}}[g] \right| \leq \sqrt{v_{p(\theta)}[g]} \sqrt{\langle v, \mathcal{I}(p(\theta)) v \rangle} + O(\varepsilon)$$

Sensitivity bounds for ergodic problem

The sensitivity bound

$$\frac{1}{|\varepsilon|}\left|E_{\pi_{p(\theta+\varepsilon v)}}[g]-E_{\pi_{p(\theta)}}[g]\right|\leq\sqrt{v_{p(\theta)}[g]}\sqrt{\langle v,\mathcal{I}(p(\theta))v\rangle}+O(\varepsilon)$$

allows for sensitivity screening: which parameters are less important.

Sensitivity bounds for ergodic problem

The sensitivity bound

$$\frac{1}{|\varepsilon|}\left|E_{\pi_{p(\theta+\varepsilon v)}}[g]-E_{\pi_{p(\theta)}}[g]\right|\leq\sqrt{v_{p(\theta)}[g]}\sqrt{\langle v,\mathcal{I}(p(\theta))v\rangle}+O(\varepsilon)$$

allows for sensitivity screening: which parameters are less important. The path FIM

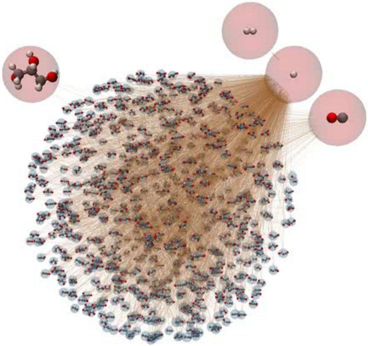
$$\mathcal{I}(p(\theta))=\int_{\mathcal{X}^2}\nabla_\theta\log p(x,y;\theta)[\nabla_\theta\log p(x,y;\theta)]^T p(x,y;\theta)R(dy)\pi_p(dx)$$

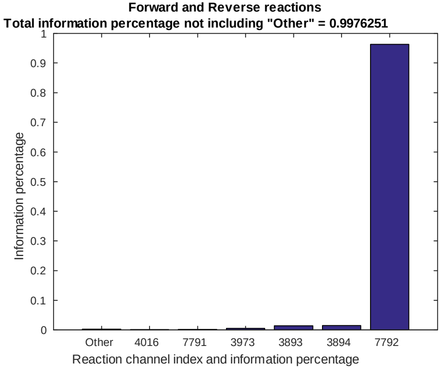
can be obtained via Monte Carlo since by ergodic theorem, with $\{X_t\}\sim p(x,dy;\theta)$

$$\frac{1}{T}\sum_{t=0}^{T-1}\delta_{(X_t,X_{t+1})}(dx\,dy)\Rightarrow\pi_p(dx)p(x,y;\theta)R(dy)$$

Sensitivity bounds for ergodic problem

Example: glycerol reforming. With $\theta =$ vector of rate parameters, and "information percentage" is fraction of trace of $\mathcal{I}(p(\theta))$ associated with given parameter[†]





[†] Computations courtesy Pedro Vilanova

Other uses

Optimally robust for ergodic cost. Suppose $g = g_\alpha$ with $\alpha \in A$. Now to solve

$$\min_{\alpha \in A} \max_{q : r(q \| p) \le r} E_{\pi_q}[g_\alpha].$$

Other uses

Optimally robust for ergodic cost. Suppose $g = g_\alpha$ with $\alpha \in A$. Now to solve

$$\min_{\alpha \in A} \max_{q: r(q \| p) \le r} E_{\pi_q}[g_\alpha].$$

Then choose α according to

$$\begin{aligned}
&\min_{\alpha \in A} \left[\max_{q} \min_{c>0} \left(E_{\pi_q}[g_\alpha] + \frac{1}{c}[r - r(q \| p)] \right) \right] \\
&\quad = \min_{\alpha \in A} \left[\min_{c>0} \max_{q} \left(E_{\pi_q}[g_\alpha] - \frac{1}{c} r(q \| p) \right) + \frac{1}{c} r \right] \\
&\quad = \min_{\alpha \in A} \min_{c>0} \frac{1}{c} \left(r + \Lambda_\alpha(c) \right).
\end{aligned}$$

Situations that require new ideas/open problems

- **Issue:** Want alternative measures that are not absolutely continuous w.r.t. design. **Possible solution:** Distances not limited to divergences, e.g., Wasserstein. However chain rule only for product measures, and in general there is dependence on probability space.

Situations that require new ideas/open problems

- **Issue:** Want alternative measures that are not absolutely continuous w.r.t. design. **Possible solution:** Distances not limited to divergences, e.g., Wasserstein. However chain rule only for product measures, and in general there is dependence on probability space.
- Multi-scale problems, e.g., models of form

$$
\begin{aligned}
dX^{\varepsilon} &= b(X^{\varepsilon}, Y^{\varepsilon})dt + \sigma(X^{\varepsilon}, Y^{\varepsilon})dW_t \\
dY^{\varepsilon} &= \frac{1}{\varepsilon} g(X^{\varepsilon}, Y^{\varepsilon})dt + \frac{1}{\varepsilon^{1/2}} \gamma(X^{\varepsilon}, Y^{\varepsilon})d\bar{W}_t
\end{aligned}
$$

  Issue: relative entropy counts fast process more heavily. **Possible solution:** "multi-level" exponential integral and corresponding variational inequalities.

Situations that require new ideas/open problems

- **Issue:** Want alternative measures that are not absolutely continuous w.r.t. design. **Possible solution:** Distances not limited to divergences, e.g., Wasserstein. However chain rule only for product measures, and in general there is dependence on probability space.
- Multi-scale problems, e.g., models of form

$$
\begin{aligned}
dX^{\varepsilon} &= b(X^{\varepsilon}, Y^{\varepsilon})dt + \sigma(X^{\varepsilon}, Y^{\varepsilon})dW_t \\
dY^{\varepsilon} &= \frac{1}{\varepsilon} g(X^{\varepsilon}, Y^{\varepsilon})dt + \frac{1}{\varepsilon^{1/2}} \gamma(X^{\varepsilon}, Y^{\varepsilon})d\bar{W}_t
\end{aligned}
$$

  Issue: relative entropy counts fast process more heavily. **Possible solution:** "multi-level" exponential integral and corresponding variational inequalities.
- **Issue:** QoI is itself risk-sensitive (i.e., depends of rare events). KL divergence is not suitable for robust bounds.

Rare event performance measures and Rényi divergence

Recently derived variational formula relates risk-sensitive QoI and Rényi divergence.

Rare event performance measures and Rényi divergence

Recently derived variational formula relates risk-sensitive QoI and Rényi divergence. Let $0 < \beta < \gamma$. Then

$$\frac{1}{\gamma} \log E_P\left[e^{\gamma f}\right] = \sup_{Q \ll P} \left[\frac{1}{\beta} \log E_Q\left[e^{\beta f}\right] - \frac{1}{\gamma - \beta} R_{\frac{\gamma}{\gamma-\beta}}(Q \,\|\, P)\right],$$

where for mutually absolutely continuous P, Q and $\alpha > 1$

$$R_\alpha(Q \,\|\, P) = \frac{1}{\alpha(\alpha-1)} \log \int_S \left(\frac{dQ}{dP}\right)^{\alpha-1} dQ.$$

Rare event performance measures and Rényi divergence

Recently derived variational formula relates risk-sensitive QoI and Rényi divergence. Let $0 < \beta < \gamma$. Then

$$\frac{1}{\gamma}\log E_P\left[e^{\gamma f}\right] = \sup_{Q \ll P}\left[\frac{1}{\beta}\log E_Q\left[e^{\beta f}\right] - \frac{1}{\gamma-\beta} R_{\frac{\gamma}{\gamma-\beta}}(Q\,\|\,P)\right],$$

where for mutually absolutely continuous P, Q and $\alpha > 1$

$$R_\alpha(Q\,\|\,P) = \frac{1}{\alpha(\alpha-1)}\log\int_S \left(\frac{dQ}{dP}\right)^{\alpha-1} dQ.$$

As $\beta \downarrow 0$ recover Kullback-Leibler formula. Bounds on risk-sensitive QoI for various Q at level β in terms of one at level γ in terms of design:

$$\frac{1}{\beta}\log E_Q\left[e^{\beta f}\right] \le \frac{1}{\gamma}\log E_P\left[e^{\gamma f}\right] + \frac{1}{\gamma-\beta} R_{\frac{\gamma}{\gamma-\beta}}(Q\,\|\,P).$$

Example use: sensitivity bounds

Rare event performance measures and Rényi divergence

Let $P = P^{\theta}, Q = P^{\theta+\varepsilon v}$, where $P^{\theta}(dx) = p(x;\theta)R(dx)$.

Rare event performance measures and Rényi divergence

Let $P = P^{\theta}, Q = P^{\theta+\varepsilon v}$, where $P^{\theta}(dx) = p(x;\theta)R(dx)$. Then

$$\frac{d}{d\varepsilon}\frac{1}{\beta}\log E_{P^{\theta}}\left[e^{\beta f}\right] \leq \frac{1}{\beta^2}\inf_{c>0}\left[\frac{R(P^{\theta}_{\beta}\,\|P^{\theta}) + \log\int e^{c\beta v^T\nabla_{\theta}p(\cdot;\theta)}dP^{\theta}}{c}\right],$$

where

$$dP^{\theta}_{\beta} = \frac{e^{\beta f}\,dP^{\theta}}{\int e^{\beta f}\,dP^{\theta}},$$

and corresponding lower bound.

Rare event performance measures and Rényi divergence

Let $P = P^{\theta}$, $Q = P^{\theta+\varepsilon v}$, where $P^{\theta}(dx) = p(x;\theta)R(dx)$. Then

$$\frac{d}{d\varepsilon}\frac{1}{\beta}\log E_{P^{\theta}}\left[e^{\beta f}\right] \leq \frac{1}{\beta^2}\inf_{c>0}\left[\frac{R(P^{\theta}_{\beta}\,\|P^{\theta}) + \log\int e^{c\beta v^T\nabla_{\theta}p(\cdot;\theta)}dP^{\theta}}{c}\right],$$

where

$$dP^{\theta}_{\beta} = \frac{e^{\beta f}\,dP^{\theta}}{\int e^{\beta f}\,dP^{\theta}},$$

and corresponding lower bound. Again decomposes into term depending on g and term plays role of Fisher information matrix. With $f = -\infty 1_{A^c}$,

$$\frac{d}{d\varepsilon}\frac{1}{\beta}\log P^{\theta}(A) \leq \frac{1}{\beta^2}\inf_{c>0}\left[\frac{-\log P^{\theta}(A) + \log\int e^{c\beta v^T\nabla_{\theta}p(\cdot;\theta)}dP^{\theta}}{c}\right],$$

with bounds uniform w.r.t. events A with uniformly bounded $-\log P^{\theta}(A)$.

Rare event performance measures and Rényi divergence

Negatives of Rényi:

- A chain rule for product measures, but not for Markov measures
- Quantity one would optimize over in robust design (here γ) appears also in $R_{\frac{\gamma}{\gamma-\beta}}(Q \,\|\, P)$. Complicates formulation of robust optimization

Rare event performance measures and Rényi divergence

Negatives of Rényi:

- A chain rule for product measures, but not for Markov measures
- Quantity one would optimize over in robust design (here γ) appears also in $R_{\frac{\gamma}{\gamma-\beta}}(Q \| P)$. Complicates formulation of robust optimization

Positives of Rényi:

- Bounds still scale with meaningful limits large time/system size, even for Markov measures
- Bounds independent of underlying probability space

Summary

- Variational formulas relating QoI and information divergences/distances provide a natural way to quantitatively deal with model uncertainty and robust design
- Can treat various issues such as sensitivity bounds, optimal robust design, etc.
- Important properties are (i) preservation of bounds under scalings, (ii) computationally tractable, (iii) bounds should not depend on underlying probabilistic formulation
- Variational formula with nicest properties seems to be that of Kullback-Leibler divergence and ordinary QoI
- However for some problems this framework is not appropriate, and for these alternatives are being (or need to be) investigated

References

- Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games, D.H. Jacobson, *IEEE Trans. on Auto. Control*, **18**, (1973), pp. 124–131.
- Robust properties of risk–sensitive control, D, M.R. James and I. R. Petersen, *Math. of Control, Signals and Systems*. **13** (2000) pp. 318–332.
- Minimax optimal control of stochastic uncertain systems with relative entropy constraints, I.R. Petersen, M.R. James and D, *IEEE Trans. on Auto. Control.* **45** (2000) pp. 398–412.
- Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification, K. Chowdhary and D, *ESAIM: Mathematical Modelling and Numerical Analysis*, **47**, (2013), 635–662.

References

- Robust bounds on risk-sensitive functionals via Rényi divergence, R. Atar, K. Chowdhary and D, *SIAM/ASA J. Uncertainty Quantification*, **3**, (2015), 18–33.
- Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics, D, M.A. Katsoulakis, Y. Pantazis and P. Plecháč, *SIAM/ASA J. Uncertainty Quantification*, **4**, (2016), 80-111.
- Sensitivity analysis for rare events based on Rényi divergence, D, M.A. Katsoulakis, Y. Pantazis and L. Rey-Bellet, *preprint*.