A calculus for the optimal quantification of uncertainties

Houman Owhadi

- Bayesian Brittleness.
- Brittleness of Bayesian inference and new Selberg formulas.

Kavli Royal Society 2014
Reduction calculus

\[ A = \left\{ (f, \mu) \mid \begin{array}{l} f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \to \mathbb{R}, \\ \mu = \mu_1 \otimes \cdots \otimes \mu_m, \\ \mathcal{G}(f, \mu) \leq 0 \end{array} \right\} \]

\[ \mathcal{G}(f, \mu) \leq 0 \iff \begin{cases} n' \text{ generalized moment constraints on } \mu, & \mathbb{E}_\mu[\varphi^f_j] \leq 0 \\ n_k \text{ generalized moment constraints on } \mu_k, & \mathbb{E}_{\mu_k}[\psi^f_{k,j}] \leq 0 \end{cases} \]

**Theorem**

\[ \sup_{(f, \mu) \in A} \mathbb{E}_\mu[qf] = \sup_{(f, \mu) \in A_\Delta} \mathbb{E}_\mu[qf] \]

\[ A_\Delta = \left\{ (f, \mu) \in A \mid \begin{array}{l} \mu_k \text{ is a sum of at most } n' + n_k + 1 \text{ weighted Dirac measures on } \chi_k \end{array} \right\} \]
Non-convex and infinite dimensional optimization problems can be considered as a generalization of classical Chebyshev inequalities. Our proof rely on a form of Linear programming in infinite dimensional spaces. 

\[ \sup_{(f, \mu) \in A} \mathbb{E}_\mu [qf] \]

Connection between Chebyshev inequalities and optimization theory:
- Classical Markov-Krein theorem (Karlin, Studden, 1958)
- Karlin, Studden (1966, Tchebycheff systems with applications in analysis and statistics)
- Dynkin (1978, Sufficient statistics & extreme points)
- Karr (1983, Extreme points of probability measures with applications)
- Artzner et al (1997, risk measures, value at risk, etc…)
- Betsimas & Popescu (2008, convex optimization approach to inequalities in prob. theo.)

Our proof rely on a form of Linear programming in infinite dimensional spaces:
- Winkler (1988, Extreme points of moment sets)
- Follows from an extension of Choquet theory (Phelps 2001, lectures on Choquet’s theorem) by Von Weizsacker & Winkler (1979, Integral representation in the set of solutions of a generalized moment problem)
- Kendall (1962, Simplexes & Vector lattices)
Reduction of optimization variables

\[ \{ f : \mathcal{X} \rightarrow \mathbb{R}, \, \mu \in \mathcal{P}(\mathcal{X}) \} \]

\[ \{ f : \mathcal{X} \rightarrow \mathbb{R}, \, \mu \in \mathcal{P}(\mathcal{X}) \mid \mu = \sum_{i=1}^{k} \alpha_k \delta_{x_k} \} \]

\[ \{ f : \{1, 2, \ldots, n\} \rightarrow \mathbb{R}, \, \mu \in \mathcal{P}(\{1, 2, \ldots, n\}) \} \]

\[ \{ \{1, 2, \ldots, q\}, \, \mu \in \mathcal{P}(\{1, 2, \ldots, n\}) \} \]
Example: Optimal concentration inequality

\[ \mathcal{A}_{MD} := \left\{ (f, \mu) \mid f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \to \mathbb{R}, \mu \in \mathcal{M}(\mathcal{X}_1) \otimes \cdots \otimes \mathcal{M}(\mathcal{X}_m), \mathbb{E}_\mu[f] \leq 0, \text{Osc}_i(f) \leq D_i \right\} \]

\[ \text{Osc}_i(f) := \sup_{(x_1, \ldots, x_m) \in \mathcal{X}} \sup_{x'_i \in \mathcal{X}_i} (f(\ldots, x_i, \ldots) - f(\ldots, x'_i, \ldots)) \]

\[ \mathcal{U}(\mathcal{A}_{MD}) := \sup_{(f, \mu) \in \mathcal{A}_{MD}} \mu[f(X) \geq a] \]

McDiarmid inequality
\[ \mathcal{U}(\mathcal{A}_{MD}) \leq \exp \left( -2 \frac{a^2}{\sum_{i=1}^{m} D_i^2} \right) \]
Reduction of optimization variables

\[ \mathcal{A}_C := \left\{ (C, \alpha) \ \bigg| \ \alpha \in \bigotimes_{i=1}^{m} \mathcal{M}(\{0, 1\}), \ \mathbb{E}_{\alpha}[h^C] \leq 0 \right\} \]

\[ h^C : \{0, 1\}^m \rightarrow \mathbb{R} \]

\[ t \rightarrow a - \min_{s \in C} \sum_{i : s_i \neq t_i} D_i \]

\[ \mathcal{U}(\mathcal{A}_C) := \sup_{(C, \alpha) \in \mathcal{A}_C} \alpha[h^C \geq a] \]

Theorem

\[ \mathcal{U}(\mathcal{A}_{MD}) = \mathcal{U}(\mathcal{A}_C) \]
Explicit Solution $m=2$

**Theorem**

$$m = 2$$

$$\mathcal{U}(A_{MD}) = \begin{cases} 
0 & \text{if } D_1 + D_2 \leq a \\
\frac{(D_1+D_2-a)^2}{4D_1D_2} & \text{if } |D_1 - D_2| \leq a \leq D_1 + D_2 \\
1 - \frac{a}{\max(D_1,D_2)} & \text{if } 0 \leq a \leq |D_1 - D_2|
\end{cases}$$

**OUQ bound** $a=1$

$$C = \{(1, 1)\}$$

$$h^C(s) = a - (1 - s_1)D_1 - (1 - s_2)D_2$$

**Corollary**

If $D_1 \geq a + D_2$, then

$$\mathcal{U}(A_{MD})(a, D_1, D_2) = \mathcal{U}(A_{MD})(a, D_1, 0)$$
Reduction calculus with measures over measures

\[ M(\mathcal{X}) \supset A \xrightarrow{\Psi} Q \]

\[ M(A) \supset \Pi \xleftarrow{\Psi^{-1}} Q \subset M(Q) \]

**Theorem**

\[
\sup_{\pi \in \Psi^{-1} Q} \mathbb{E}_{\mu \sim \pi} \left[ \Phi(\mu) \right] = \sup_{Q \in \Omega} \left[ \mathbb{E}_{q \sim Q} \left[ \sup_{\mu \in \Psi^{-1}(q)} \Phi(\mu) \right] \right]
\]
A simple example

10,000 children are given one pound of play-doh. On average, how much mass can they put above $a$ while, on average, keeping the seesaw balanced around $m$?

Paul is given one pound of play-doh. What can you say about how much mass he is putting above $a$ if all you have is the belief that he is keeping the seesaw balanced around $m$?
What is the least upper bound on
\[ \mathbb{E}_{\mu \sim \pi} \left[ \mu \left[ X \geq a \right] \right] \]

If all you know is
\[ \mathbb{E}_{\mu \sim \pi} \left[ \mathbb{E}_\mu \left[ X \right] \right] = m \]

\[ \mu \in \mathcal{A} := \mathcal{M}([0, 1]) \]

Answer
\[ \sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu \left[ X \geq a \right] \right] \]

\[ \Pi := \left\{ \pi \in \mathcal{M}(\mathcal{A}) : \mathbb{E}_{\mu \sim \pi} \left[ \mathbb{E}_\mu \left[ X \right] \right] = m \right\} \]
Theorem

\[
\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu [X \geq a] \right] = \sup_{Q \in \mathcal{M}([0,1]) : \mathbb{E}_Q[q] = m} \mathbb{E}_{q \sim Q} \left[ \sup_{\mu \in \mathcal{M}([0,1]) : \mathbb{E}_\mu[X] = q} \mu [X \geq a] \right]
\]
\[
\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu[X \geq a] \right] = \sup_{Q \in \mathcal{M}([0,1]) : \mathbb{E}_Q[q] = m} \mathbb{E}_{q \sim Q} \left[ \min\left(\frac{q}{a}, 1\right) \right]
\]
$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu[X \geq a] \right]$$

$$\Pi := \left\{ \pi \in M(M([0,1])) : \mathbb{E}_{\mu \sim \pi} \left[ \mathbb{E}_{\mu}[X] \right] = m \right\}$$

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu[X \geq a] \right] = \frac{m}{a}$$
Can this form of calculus in infinite dimensional spaces facilitate the process of scientific discovery?

New Reproducing Kernel Hilbert Spaces and Selberg Integral formulas

Forrester and Warnaar 2008

The importance of the Selberg integral

“Used to prove outstanding conjectures in Random matrix theory and cases of the Macdonald conjectures”

“Central role in random matrix theory, Calogero-Sutherland quantum many-body systems, Knizhnik-Zamolodchikov equations, and multivariable orthogonal polynomial theory”
The truncated moment problem

\[ \mathcal{M}[0, 1] \xrightarrow{\Psi} \mathbb{R}^k \]

\[ \mu \xrightarrow{\Psi} (\mathbb{E}_{X \sim \mu}[X], \mathbb{E}_{X \sim \mu}[X^2], \ldots, \mathbb{E}_{X \sim \mu}[X^k]) \]

Study of the geometry of \( M_k := \Psi(\mathcal{M}([0, 1])) \)

P. L. Chebyshev
1821-1894

A. A. Markov
1856-1922

M. G. Krein
1907-1989
\[ M[0, 1] \xrightarrow{\Psi} \mathbb{R}^k \]

\[ \mu \xrightarrow{\Psi} (E_{X \sim \mu}[X], E_{X \sim \mu}[X^2], \ldots, E_{X \sim \mu}[X^k]) \]

\[ M_k := \Psi(M([0, 1])) \]

Origin of these new Selberg integral formulas and new RKHS

Compute Vol(\(M_k\)) using different (finite-dimensional) representations in \(M([0, 1])\)
Let us compute $\text{Vol}(M_k)$ using different extreme points representations.
\[
\mu = \sum_{j=1}^{N} \lambda_j \delta_{t_j} \quad \xrightarrow{\Psi} \quad (q_1, \ldots, q_k)
\]
\[
q_i = \sum_{j=1}^{N} \lambda_j t^i_j
\]
\[ \mu = \sum_{j=1}^{N} \lambda_j \delta_{t_j} \]

**Index** \( i(\mu) \): Number of support points of \( \mu \)

Counting interior points with weight 1 and boundary points with weight \( \frac{1}{2} \)

\( \mu \) is called
- principal if \( i(\mu) = \frac{k+1}{2} \)
- canonical if \( i(\mu) = \frac{k+2}{2} \)
- upper if support points include 1
- lower if support points do not include 1

**Theorem**

Every point \( q \in \text{Int}(M_k) \) has a unique upper and lower principal representation.

![Graph showing upper and lower representations with intervals from 0 to 1, including points t1, t2, tj, tN.](attachment:graph.png)
\[
\frac{1}{(m-1)!} S_{m-1}(3, 3, 2) = \frac{1}{m!} S_m(1, 1, 2)
\]

\[
S_m(1, 3, 2) = S_m(3, 1, 2)
\]

Selberg Identities

\[
S_n(\alpha, \beta, \gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}
\]

\[
S_n(\alpha, \beta, \gamma) := \int_{[0,1]^n} \prod_{j=1}^{n} t_j^{\alpha-1}(1-t_j)^{\beta-1}|\Delta(t)|^{2\gamma} dt.
\]

\[
\Delta(t) := \prod_{j<k} (t_k - t_j)
\]

\[ \mu = \sum_{j=1}^{N} \lambda_j \delta_{t_j} \]

**Index**  \( i(\mu) \): Number of support points of \( \mu \)

Counting interior points with weight 1 and boundary points with weight \( \frac{1}{2} \)

\( \mu \) is called
- principal if \( i(\mu) = \frac{k+1}{2} \)
- canonical if \( i(\mu) = \frac{k+2}{2} \)
- upper if support points include 1
- lower if support points do not include 1

**Theorem**

For \( t_* \in (0, 1) \), every point \( q \in \text{Int}(M_k) \) has a unique canonical representation whose support contains \( t_* \).

When \( t_* = 0 \) or 1, there exists a unique principal representation whose support contains \( t_* \).
New Reproducing Kernel Hilbert Spaces and Selberg Integral formulas related to the Markov-Krein representations of moment spaces.

\[ M[0, 1] \xrightarrow{\Psi} [0, 1]^k \]

\[ \mu \xrightarrow{} (\mathbb{E}_{X \sim \mu}[X], \mathbb{E}_{X \sim \mu}[X^2], \ldots, \mathbb{E}_{X \sim \mu}[X^k]) \]

\[ \int_{I^m} \sum t^{-1} \cdot \prod_{j=1}^{m} t_j^2 (1 - t_j)^2 \Delta_4^m(t) dt = \frac{S_m(5,1,2) - S_m(3,3,2)}{2} \]

\[ \int_{I^m} \sum t^{-1} \cdot \prod_{j=1}^{m} t_j^2 \cdot \Delta_4^m(t) dt = \frac{m}{2} S_{m-1}(5, 3, 2) \]

\[ \Delta_m(t) := \prod_{j<k} (t_k - t_j) \quad I := [0, 1] \]

\[ (\Sigma \phi)(t) := \sum_{j=1}^{m} \phi(t_j), \quad t \in I^m \]

\[ S_n(\alpha, \beta, \gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)} \]
\[ e_j(t) := \sum_{i_1 < \cdots < i_j} t_{i_1} \cdots t_{i_j} \]

\( \Pi^n_0 \): \( n \)-th degree polynomials which vanish on the boundary of \([0, 1]\)

\( M_n \subset \mathbb{R}^n \): set of \( q = (q_1, \ldots, q_n) \in \mathbb{R}^n \) such that there exists a probability measure \( \mu \) on \([0, 1]\) with \( \mathbb{E}_\mu[X^i] = q_i \) with \( i \in \{1, \ldots, n\} \).

**Theorem**

**Bi-orthogonal systems of Selberg Integral formulas**

Consider the basis of \( \Pi_0^{2m-1} \) consisting of the associated Legendre polynomials \( Q_j, j = 2, \ldots, 2m - 1 \) of order 2 translated to the unit interval \( I \). For \( k = 2, \ldots, 2m - 1 \) define

\[ a_{jk} := \frac{(j + k + k^2)\Gamma(j + 2)\Gamma(j)}{\Gamma(j + k + 2)\Gamma(j - k + 1)}, \quad k \leq j \leq 2m - 1 \]

\[ \tilde{h}_k(t) := \sum_{j=k}^{2m-1} (-1)^{j+1} a_{jk} e_{2m-1-j}(t, t). \]

Then for \( j = k \mod 2, j, k = 2, \ldots, 2m - 1 \), we have

\[
\int_{I^{m-1}} \tilde{h}_k(t) \sum_{j=1}^{m-1} t_j^2 \prod_{j' = 1}^{m-1} \Delta_4^{m-1}(t) dt = Vol(M_{2m-1})(2m-1)!(m-1)! \frac{(k + 2)!}{(8k + 4)(k - 2)!} \delta_{jk}.
\]