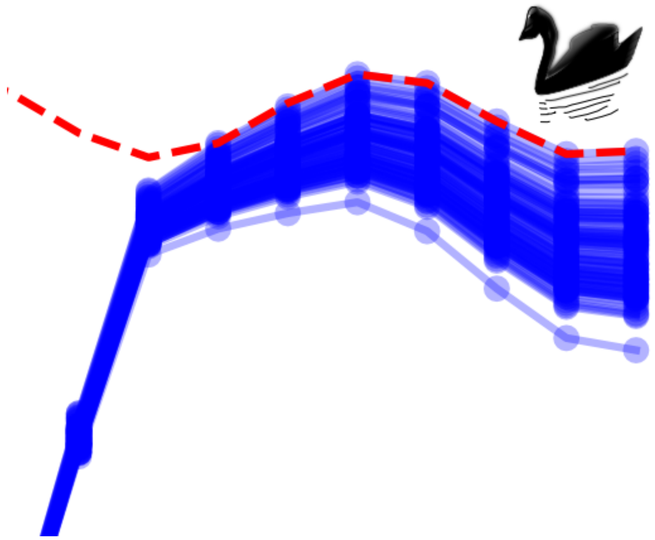
Worst-Case Risk Quantification under Distributional Ambiguity using Kernel Mean Embedding in Moment Problem

Jia-Jie Zhu[*], Wittawat Jitkrittum[*],
Moritz Diehl[**], Bernhard Schölkopf[*]

[*]Empirical Inference Department, Max Planck Institute for Intelligent Systems
Tübingen, Germany
[**]Department of Microsystems Engineering & Department of Mathematics University of Freiburg
Freiburg, Germany

CDC, December 2020

Distributional shift in data-driven control



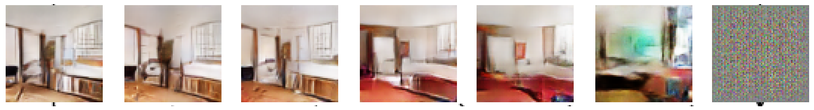


Image credit: Arjovsky'17

- For example, consider the Monte Carlo estimate for probability of constraint satisfaction:

$$\mathbb{P}(c(x) \leq 0) \approx \sum_{i=1}^{N} \frac{1}{N} 1\{c(x_i) \leq 0\} = 1$$

- If I have never seen a black swan, they must not exist…
- Stochastic and robust control with uncertainty-aware models robustifies against ***known*** **uncertainty**
  - but not the (inevitable) data distribution shift.
  - Watch out for the black swan!
    - Implication: sim2real failure, off-policy in RL

Mathematical problem of moments

Empirical risk: $\mathbb{E}_{\xi\sim\hat{P}} \; l(\xi)$

Empirical data dist. $\hat{P} = \sum_{i=1}^{N} \frac{1}{N} \delta_{\xi_i}$

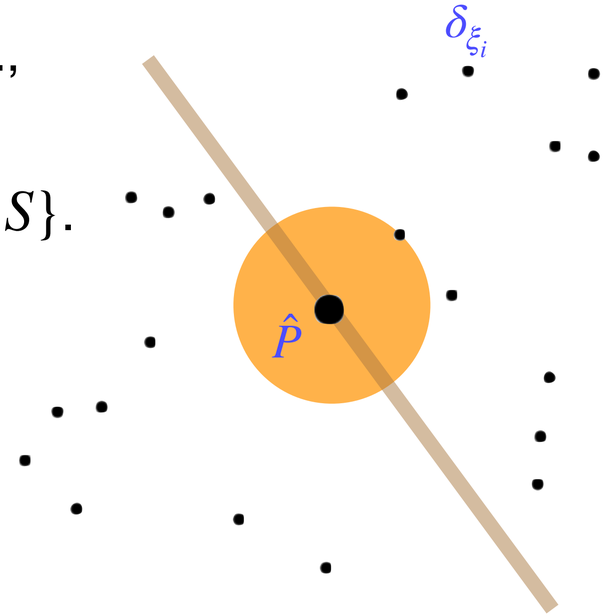
Robust risk $\sup_{\xi\in\mathcal{X}} \; l(\xi)$

Uncertainty set

$$\sup_{P\in\mathcal{K}} \; \mathbb{E}_P l(\xi)$$

Generalized Moment Problem [Stieltjes, Hausdorff, Hamburger 100+ yrs]
Find the worst-case distribution (adversary)!

- The worst case risk in a set of probability measures $\mathcal{K}$ (***ambiguity set***), e.g.,
  - Modern approaches to the problem of moments use, e.g.,
    - moment ambiguity set, e.g.,$\mathcal{K} := \{P : \mathbb{E}_P 1 = 1, \mathbb{E}_P \xi = \mu, \mathbb{E}_P \xi\xi^T = S\}$.
      This is a hyperplane in $\mathcal{P}$.
    - $\mathcal{K}$ can be a metric-ball centered at $\hat{P}$, e.g., Wasserstein metric , ϕ -divergence.
- This paper proposes a generalized approach using a functional perspective rooted in the theory of RKHS.
- Our approach can be viewed as generalizations of the above ambiguity sets.

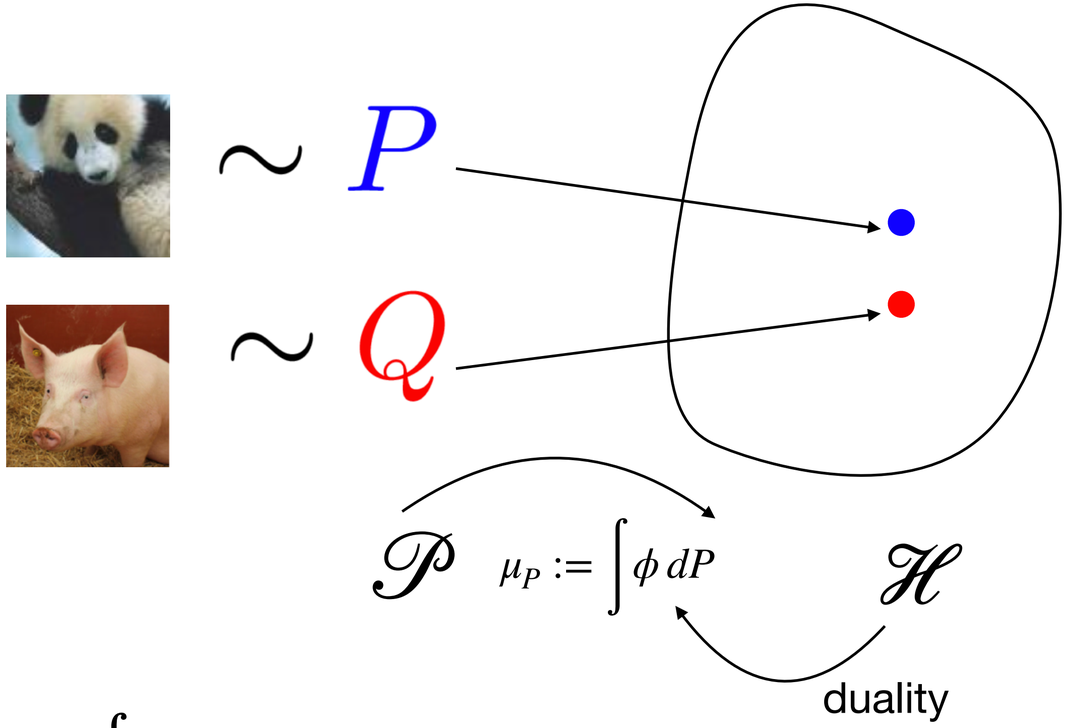


Elements of learning with kernels

- A kernel is a symmetric function $k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, e.g., Gaussian kernel $k(x, x') = \exp\left(-\|x - x'\|_2^2 \,/\, 2\sigma^2\right)$.

- $\phi(x) := k(x, \cdot\,)$ is called the feature  .

- A positive semi-definite k defines a Hilbert space $\mathcal{H}$, which satisfies the *reproducing property* $f(x) = \langle f, \phi(x) \rangle_{\mathcal{H}}, \forall f \in \mathcal{H}, x \in \mathcal{X}$.



$\mu := \int \phi \, dP$ is the *kernel mean embedding* of P in $\mathcal{H}$.

μ can be viewed as a generalized moment vector
e.g., let $\phi(x) = [x, x^2]^\top$

Illustration design inspired by Gretton, Sutherland, Jitkrittum NeurIPS 2019 tutorial

Problem of moments with kernel mean embedding

$$\begin{aligned} \underset{P,\mu}{\text{maximize}} \quad & \int l(x)\ dP(x) \\ \text{subject to} \quad & \|\mu - \mu_{\hat{P}}\|_{\mathcal{H}} \leq \epsilon \\ & \int \phi(x)\ dP(x) = \mu \\ & P \in \mathcal{P}, \mu \in \mathcal{H}, \end{aligned}$$

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{N} \alpha_i l(z_i) \\ \text{subject to} \quad & \alpha^\top K_z \alpha - 2\frac{1}{M}\alpha^\top K_{zx}\mathbf{1} + \frac{1}{M^2}\mathbf{1}^\top K_x \mathbf{1} \leq \epsilon^2 \\ & \sum_{i=1}^{N} \alpha_i = 1, \alpha_i \geq 0, i = 1 \dots N. \end{aligned}$$

- where $\phi(x) := k(x, \cdot)$ is the RKHS feature.
- If we choose the kernel to be the p-th order polynomial kernel (non-characteristic), we recover the moment problem with p-th order moment bound in the literature.
- This can be generalized to the class of integral probability metric, which includes the Wasserstein metrics.

We prove the solution convergence guarantee

$$\sum_{i=1}^{N} \alpha_i^* l(z_i) \xrightarrow{N\to\infty} \int l(x)\ dP^*(x).$$

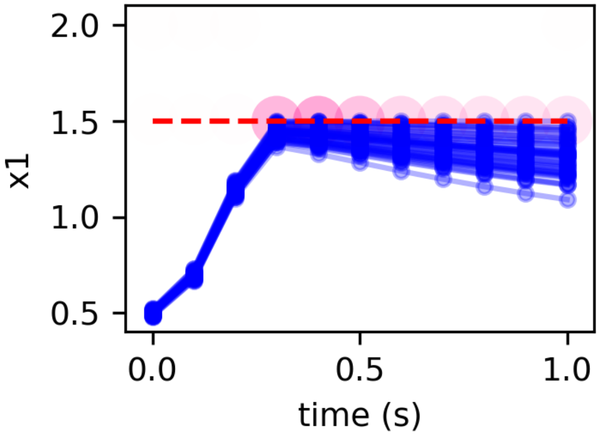
Note: We do not assume the loss to be quadratic or have a known Lipschitz constant as in many approaches.

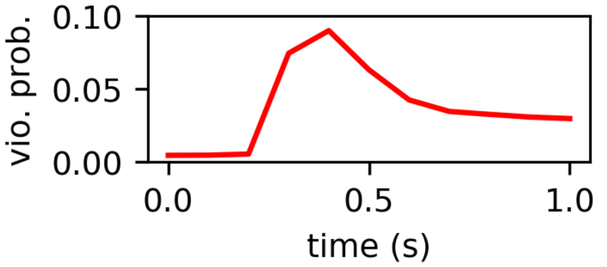
Uncertainty quantification in stochastic MPC

- Constrained stochastic control system

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -0.1\left(1-x_1^2\right)x_2 - x_1 + u \end{bmatrix}.$$

- We solve the stochastic OCP with scenario SMPC. All simulated trajectories satisfy constraints.
- Assuming an ambiguity set $\|\mu_P - \mu_{\hat{P}}\|_{\mathscr{H}} \leq \epsilon$, we solve the kernel moment problem to quantify the worst-case constraint violation probability in SMPC.





Conclusions

- This paper introduced the *kernel mean embedding* framework as a tool for solving generalized moment problems to quantify the worst-case risk.
- We propose a practical solution method and proved the consistency of the solution.
- Our theory unifies the existing ambiguity description such as the Wasserstein distance, through *IPM,* and moment ambiguity sets.
- Subsequent works:
  - Solve DRO with a generalized duality results with RKHS functions (*Kernel DRO*).
  - Principled constraint tightening for distributionally robust MPC with *Kernel DRO.*
  - *MMD* has attractive convergence rates for control applications in terms of dimensionality.

Thank you! This talk is based on

- Z, Jitkrittum, Diehl, Schölkopf, 2020. **Worst-Case Risk Quantification under Distributional Ambiguity using Kernel Mean Embedding in Moment Problem**. CDC.
- Z, Jitkrittum, Diehl, Schölkopf, 2020. Kernel Distributionally Robust Optimization. Preprint.
- Nemmour, Schölkopf, Z. 2020. Constraint Tightening Techniques for Distributionally Robust Model Predictive Control: A Functional Approach. Preprint.

For more information, contact me at jzhu@tuebingen.mpg.de