# Safe Interactive Motion Planning with Learning-based Distributionally Robust Optimal Control

Erik Börve, Nikolce Murgovski, Morteza Haghir Chehreghani, and Leo Laine

## I. Introduction

Motion planning is a crucial component of any robotic application, such as manipulators, UAVs, and Autonomous Vehicles (AVs). Although much work has been dedicated to this topic, arguably, the most difficult challenge still remains, safely dealing with humans. The difficulties stem from the fact that human movement is notoriously challenging to model, can depend on the robots' own motion, and can be inherently uncertain. On a high level, one may decompose this uncertainty into two components; 1.) *Motion uncertainty*, e.g., "Is the humans' velocity $1.0\,\mathrm{m\,s^{-1}}$ or $1.1\,\mathrm{m\,s^{-1}}$?" 2.) *Decision uncertainty*, e.g., "Will the human stop or go through the intersection?". In this work we explore Learning-based Distributionally Robust Optimization (LB-DRO) as a solution for deriving safety guarantees for motion planning with uncertain human decisions, by modeling humans as Markov Decision Processes (MDP) with a discrete set of a priori know policies. Our work is suitable for treating safety when the decision uncertainty is the primary cause of concern, e.g., when an AV needs to interact with a human driver at an intersection. Note that the work is not final; we are currently finalizing the proofs and a simulation study that we look forward to presented at the conference.

## II. Learning-based Distributionally Robust Optimal Control Problem

Consider a decision vector $\mathbf{u} \in \mathbf{U} \subseteq \mathbb{R}^{n_u}$ and a random vector $\mathbf{y} : \mathbf{Y} \mapsto \Omega \subseteq \mathbb{R}^{n_y}$ with probability measure $\mathbf{p}^\star$ on the measurable space $(\mathbf{Y}, \Omega)$. In addition, consider a random cost function $J(\mathbf{u}, \mathbf{y}) : \mathbf{U} \times \mathbf{Y} \mapsto \mathbb{R}$ and a vector of random constraints $g(\mathbf{u}, \mathbf{y}) : \mathbf{U} \times \mathbf{Y} \mapsto \mathbb{R}^{n_g}$. LB-DRO concerns itself with solving problems of the following structure,

$$\inf_{\mathbf{u}} \sup_{\mathbf{p} \in \mathcal{P}_\theta} \left\{ \mathcal{R}_{\mathbf{p}} \left[ J(\mathbf{u}, \mathbf{y}) \right] \;\middle|\; \sup_{\mathbf{p} \in \mathcal{P}_\theta} \mathcal{R}_{\mathbf{p}} \left[ g(\mathbf{u}, \mathbf{y}) \right] \leq 0 \right\} \tag{1}$$

where $\mathcal{R}_{\mathbf{p}}$ is some risk measure and $\mathcal{P}_\theta$ is an "*ambiguity set*" of probability measures that depends on some learnable parameters $\theta$. The goal of the learning problem is to find a $\mathcal{P}_\theta$ that contains $\mathbf{p}^\star$ with high probability, i.e., $\mathbb{P}\left[\mathbf{p}^\star \in \mathcal{P}_\theta\right] \geq 1 - \alpha$ for some $\alpha \in [0, 1]$. [1].

### A. Prior Work

Existing work [2], [3] have extended the LB-DRO framework to a predictive multistage setting over a discrete time horizon $k = 0, 1, \ldots, N$. Crucially, they consider $\mathbf{Y} = \{0, \ldots, n_y\}$ as a discrete set of possible decisions for a human at each time $k$. The random decisions $\mathbf{y}$ are modeled as a discrete-time Markov chain by treating a *transition kernel* $\mathbf{p}^\star = (\mathbf{p}^\star_{ij})_{i,j \in \mathbf{Y}}$ where $\mathbf{p}^\star_{ij} = \mathbb{P}\left[\mathbf{y} = y_j | \mathbf{y} = y_i\right]$. To learn the transition kernel they rely on a Sample Average Approximation (SAA) by collecting $t$ measurements of $\mathbf{y}$ in an online-fashion,

$$\hat{\mathbf{p}}_{i,j}(t) \coloneqq \theta_{i,j}(t) = \frac{1}{t} \sum_{i=1}^{t} \mathbb{1}_{[\mathbf{y}(t) = y_i, \mathbf{y}(t-1) = y_j]} \tag{2}$$

where $\mathbb{1}$ is an indicator function. Assuming that the Markov Chain is ergodic and that we may obtain i.i.d. measurements of $\mathbf{y}$, one may utilize uniform convergence properties of $\mathbf{p}^\star$ to construct an ambiguity set as follows,

$$\mathcal{P}_\theta = \left\{ \mathbf{p} \;\middle|\; ||\mathbf{p} - \hat{\mathbf{p}}(t)||_1 \leq \sqrt{\frac{n_y^2 \log 2 - \alpha}{t}} \right\} \tag{3}$$

where $\mathbb{P}[\mathbf{p}^\star \in \mathcal{P}_\theta] \geq 1 - \alpha$. In this setting, [2], [3] considers the Conditional Value-at-Risk (CVaR), i.e. $\mathcal{R}_{\mathbf{p}}[\cdot] = \mathrm{CVaR}_{\mathbf{p}}[\cdot]$, and performs extensive derivations to transform (1) into a tractable multistage LB-DRO problem. All details cannot be covered here, but a particularly crucial step is to enumerate all possible combinations of $\mathbf{y}$ over the prediction horizon to construct a scenario tree. The problem is finally solved in a receding horizon fashion where the solution $\mathbf{u}^\star$ yields the AV control actions e.g., acceleration and steering.

### B. Our Contribution

Indeed, the above SAA approach is restrictive and cannot be used to treat interactive motion planning, as the dependence of $\mathbf{p}^\star$ is limited to the previous actions of the human driver via $\mathbf{y}(t-1)$. In this work, we develop theoretical results for conditional distributions $\mathbf{p}^\star_i = \mathbb{P}[\mathbf{y} = y_i | \mathbf{x}]$ where $\mathbf{x} = [\mathbf{x}_{\mathrm{AV}}, \mathbf{x}_{\mathrm{h}}]$ contains continuous state variables for the AV and human, e.g. position and velocity. To maintain the i.i.d. assumption, we consider an offline learning setting where we obtain measurements for humans interacting with another vehicle in a specific driving scenario, e.g., an intersection.

## III. Machine Learning Problem

### A. Basic Definitions

We consider a classification problem with random variables $\mathbf{y} \in \mathbf{Y} = \{0, 1, \ldots, n_y\}$, $\mathbf{x} \in \mathbf{X} \subseteq \mathbb{R}^{n_x}$ with joint distribution $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}$. In this setting we may describe the joint distribution as,

$$\mathcal{D} \coloneqq F^\star_\theta(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}) f^\star_\theta(\mathbf{y} | \mathbf{x}) \tag{4}$$

where $\theta \in \mathbb{R}^{n_x}$ are parameters and $P(\mathbf{x})$ is the state distribution. We assume to obtain $n$ i.i.d. samples from $\mathcal{D}$ and construct a training set $\mathbb{D}_{\mathrm{train}} = \{x_i, y_i\}_{i=1}^n$ with the aim of

learning the conditional distribution using an empirical risk minimizer $\hat{f}_\theta(\mathbf{y}|\mathbf{x})$ (ERM). To this end, we define a risk measure utilizing the cross-entropy, i.e,

$$\mathcal{R}_{\mathcal{D}}(\hat{f}_\theta) = \mathbb{E}_{F_\theta^\star}\big[-\log \hat{F}_\theta\big] = H(F_\theta^\star, \hat{F}_\theta) \tag{5}$$

where $H(p,q)$ notes the cross-entropy between two distributions $p$, $q$. To minimize (5) we consider an empirical estimate of the risk,

$$\mathcal{R}_n(\hat{f}_\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\mathbf{y}=y_i}\big[-\log\big(\hat{f}_\theta(\mathbf{y}=y_i, \mathbf{x}=x_i)\big)\big] \tag{6}$$

where $\mathbb{1}_{\mathbf{y}=y_i}$ is an indicator function for $\mathbf{y}=y_i$.

### B. Excess Risk Bounds

Much work has been dedicated to deriving bounds on the excess risk for many different machine learning algorithms. Such bounds commonly take the following form,

$$\mathbb{P}\Big[\mathcal{R}_{\mathcal{D}}(\hat{f}_\theta) - \mathcal{R}_{\mathcal{D}}(f_\theta^\star) \leq r(n,\alpha)\Big] \geq 1-\alpha \tag{7}$$

for some $r(n,\alpha) \geq 0$, depending on the number of samples $n$, and some $\alpha \in [0,1]$. Similarly to the vast majority of learning-based DRO approaches, most such risk bounds rely on uniform convergence, which we may state more formally with the following assumption.

*Assumption 1 (Uniform Convergence):* Consider some risk measure $\mathcal{R}_{\mathcal{D}}$ with empirical estimator $\mathcal{R}_n$, and some estimator $\hat{f}_\theta$ of the distribution $f_\theta^\star$. Uniform convergence then implies that,

$$\lim_{n\to\infty} \mathbb{P}\Big[\big|\mathcal{R}_n(\hat{f}_\theta) - \mathcal{R}_{\mathcal{D}}(f_\theta^\star)\big| > \epsilon\Big] = 0. \tag{8}$$

for some $\epsilon \geq 0$.

Bounds of the type (7) are additionally challenging in the setting of Section III-A, since the risk is unbounded, but may be obtained under uniform convergence by imposing additional constraints on the random variable $\mathbf{x}$ and the learnable weights $\theta$. Different approaches, e.g., VC-dimensions [4], Rademarcher complexity [5] or PAC-Bayes bounds [6], yield different constraints with varying restrictiveness in different applications. Note that our approach is flexible with regards to the choice of bound, and may be adapted based on the application.

### C. Problem Formulation

For completeness we present a case where $\mathbf{y} \in \{-1, 1\}$ is conditionally Bernoulli distributed as,

$$f_\theta^\star(\mathbf{y}|\mathbf{x}) = \mathrm{Ber}\big(\sigma(\langle \mathbf{x}, \theta\rangle)\big) \tag{9}$$

where $\sigma(t) = \frac{1}{1+\exp(t)}$ and the parameters $\theta$ are unknown. In this setting, we use a Rademarcher complexity bound by additionally considering $\mathbf{X} = \{\mathbf{x}\,|\,||\mathbf{x}|| \leq B\}$ and the parameters $\theta \in \Theta = \{\theta\,|\,||\theta|| \leq R\}$. The ERM is now obtained from the following constrained optimization problem,

$$\hat{f}_\theta = \arg\min_{f_\theta} \mathcal{R}_n(f_\theta), \text{ s.t. } \theta \in \Theta \tag{10}$$

The Rademarcher complexity bound then gives the following result.

*Proposition 1 (Excess Risk bound):* Consider the learning problem in Section III-A with $\mathbf{X} = \{\mathbf{x}\,|\,||\mathbf{x}|| \leq B\}$, $\Theta = \{\theta\,|\,||\theta|| \leq R\}$. Under Assumption 1, we may bound the excess risk as,

$$\mathcal{R}_{\mathcal{D}}(\hat{f}_\theta) - \mathcal{R}_{\mathcal{D}}(f_\theta^\star) \leq r(\alpha,n) = \frac{4BR}{\sqrt{n}} + 6\sqrt{\frac{\log(2/\alpha)}{2n}} \tag{11}$$

which holds with a probability larger or equal to $1-\alpha$.
*Proof:* Application of [5], details cannot fit in this format.

## IV. LB-DRO with Ambiguity Sets from Excess Risk

We now present our main result, deriving a valid ambiguity set for a conditional distribution using excess risk bounds.

*Proposition 2 (Conditional Ambiguity Set from Excess Risk):* Consider the setting of Section III-A with a well defined bound on the excess risk, such as in (10). For a given realization of the random variable $\mathbf{x} = x_i$ and an estimated conditional distribution $\hat{\mathbf{p}}_\theta(x_i) = \hat{f}_\theta(\mathbf{y}|x_i) \in (0,1)^2$, we propose the following ambiguity set.

$$\mathcal{P}_\theta(x_i) = \big\{\mathbf{p}\,\big|\,\mathrm{D}_{\mathrm{KL}}\big(\mathbf{p}||\hat{\mathbf{p}}_\theta(x_i)\big) \leq \eta\big(r(\alpha,n)\big)\big\}$$

where $\mathbf{p} \in (0,1)^2$, $\sum_i \mathbf{p}_i = 1$ and the following probabilistic guarantee holds,

$$\mathbb{P}\left[f_\theta^\star(\mathbf{y}|x_i) \in \mathcal{P}_\theta(x_i)\right] \geq \left(1 - \frac{r(\alpha,n)}{\eta\big(r(\alpha,n)\big)}\right)(1-\alpha)$$

where $\eta(r)$ is a function such that,

$$\lim_{r\to 0}\eta(r) = 0, \quad \lim_{r\to 0} 1 - \frac{r}{\eta(r)} = 1$$

*Proof:* Details cannot fit in this format.

With derivations similar to [2], [3] we are then able to obtain a tractable multistage LB-DRO problem that can treat a much more extensive family of distributions $\mathbf{p}^\star$.

## V. Current and Future Work

Our current efforts are focused on finalizing a demonstration for an intersection scenario in which a human driver and an AV need to negotiate crossing priority. There is also plenty of exciting future work including formulating recursive feasibility, treating online learning, and treating more complex learning methods, e.g., small neural networks.

## References


[1] F. Lin, X. Fang, and Z. Gao, "Distributionally robust optimization: A review on theory and applications," *Numerical Algebra, Control and Optimization*, vol. 12, no. 1, pp. 159–212, 2022.

[2] P. Sopasakis, M. Schuurmans, and P. Patrinos, "Risk-averse risk-constrained optimal control," in *2019 18th European Control Conference (ECC)*. IEEE, 2019, pp. 375–380.

[3] M. Schuurmans and P. Patrinos, "A general framework for learning-based distributionally robust mpc of markov jump systems," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 2950–2965, 2023.

[4] L. Devroye, L. Györfi, and G. Lugosi, *A probabilistic theory of pattern recognition*. Springer Science & Business Media, 2013, vol. 31.

[5] P. L. Bartlett and S. Mendelson, "Rademacher and gaussian complexities: Risk bounds and structural results," *Journal of Machine Learning Research*, vol. 3, no. Nov, pp. 463–482, 2002.

[6] S. Nakakita, "Dimension-free uniform concentration bound for logistic regression," *arXiv preprint arXiv:2405.18055*, 2024.