Safe Decision-Making for Autonomous Driving

Hannes Eriksson
hannese@chalmers.se, hannes.eriksson@zenseact.com
Zenseact AB, Chalmers University of Technology

Abstract

For sequential decision-making problems such as autonomous driving it is imperative to consider the full range of outcomes as they might range from arriving to a location faster than expected, to being part of a catastrophic crash. In particular, we are mostly concerned with the left-tail properties of the distribution of outcomes.
By devising risk-aware agents that focus on performance in the worst outcomes, we can arrive at *safer* decision-makers.

Paper I: Epistemic Risk-sensitive Reinforcement Learning

In this work we studied the concept of *epistemic risk*, that is the risk that arises due to uncertainty about the model parameters μ. Typically this situation occurs when we have a belief over MDPs $\xi(\mu)$ and we want to optimize for a risk-sensitive objective w.r.t. the uncertainty due to ξ.
The main contributions were defining an entropic risk measure for epistemic risk and the delivery of two algorithms, one based on **Approximate Dynamic Programming** and one based on **Bayesian policy gradient**.

$$\pi^E(U) = \arg\max_{\pi} \frac{1}{\beta} \log \mathbb{E}\Big(\exp(\beta R)\Big). \tag{1}$$

Eq. 1. defines the objective over the utility function mirroring an entropic risk measure with some nice properties. By also considering the uncertainty induced by ξ we can arrive at the full objective in Eq. 2., by replacing U with the considered utility function in Eq. 1.

$$\pi^E(U, \xi) \triangleq \arg\max_{\pi} \int_{\mathcal{M}} U(\mathbb{E}^{\pi}_{\mu}[R]) \, \mathrm{d}\xi(\mu). \tag{2}$$

SENTINEL Schematic

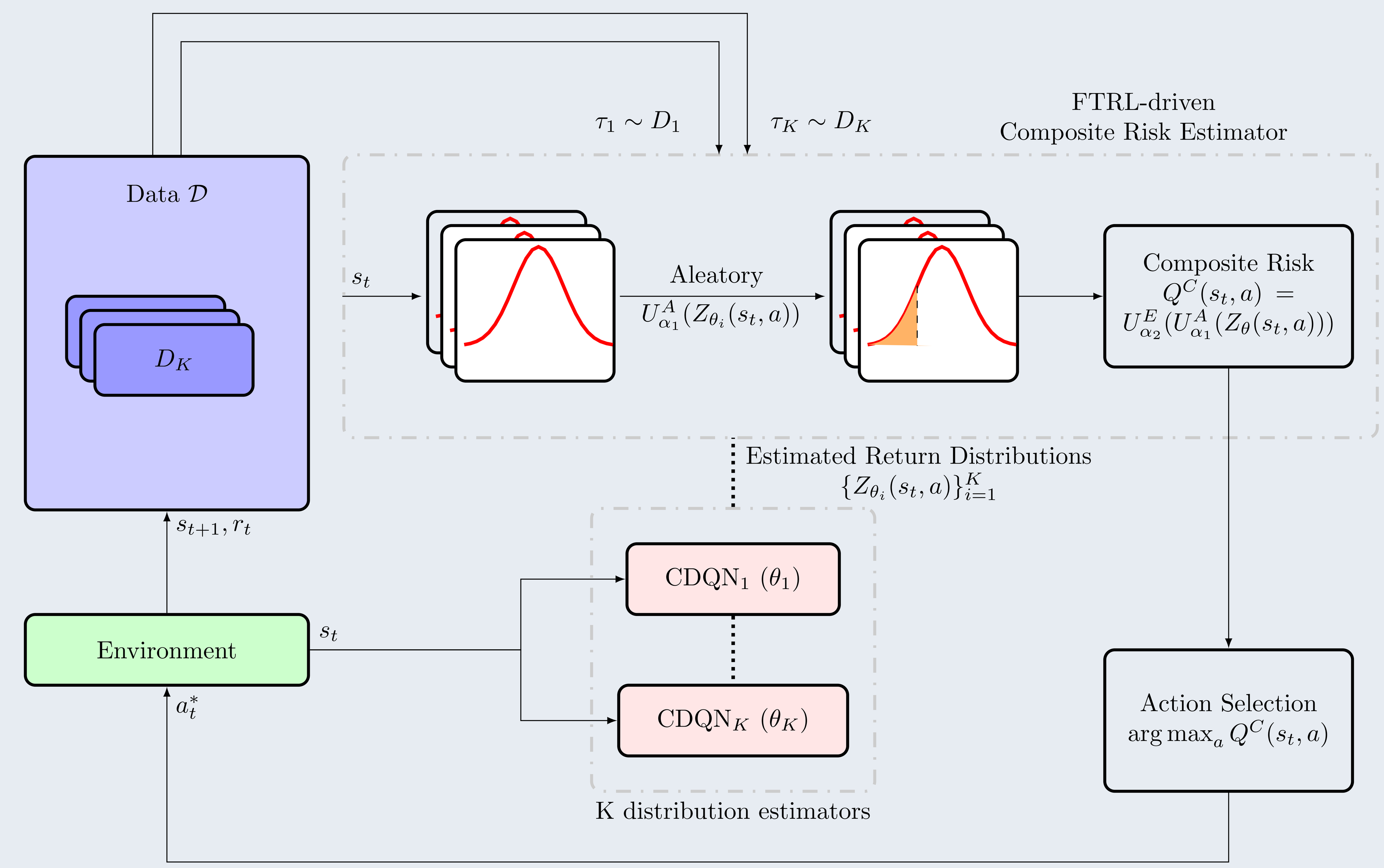

Figure: Overview of the model structure of the SENTINEL work.

Paper II: Inferential Induction: A Novel Framework for Bayesian Reinforcement Learning

In this work we introduced a novel Bayesian Reinforcement Learning framework that correctly infers value function distributions from data. From this framework, depending on what you marginalize over, gives arise to a whole new class of BRL algorithms. In particular, we develop and demonstrate comparable to state-of-the-art performance of **Bayesian Backwards Induction**.

$$\mathbb{P}^{\pi}_{\beta}(V_i \mid D) = \int_{\mathcal{V}} \mathbb{P}^{\pi}_{\beta}(V_i \mid V_{i+1}, D) \, \mathrm{d}\mathbb{P}^{\pi}_{\beta}(V_{i+1} \mid D). \tag{3}$$

Paper III: SENTINEL: Taming Uncertainty with Ensemble-based Distributional Reinforcement Learning

In *SENTINEL*, we study a novel kind of risk measure, in this work termed *composite risk*, which combines both the risk due to aleatory uncertainty and the risk due to epistemic uncertainty into one risk measure. We prove that this new risk measure better estimates the total risk than one that considers both risks separately.

$$\text{Comp. Risk} \triangleq \int_{\Theta} \int_{\mathcal{Z}} Z \, \mathrm{d}(U^A_{\alpha_1} \circ \mathrm{Pr})(Z|\theta) \, \mathrm{d}(U^E_{\alpha_2} \circ \beta)(\theta) \tag{4}$$

Further, we demonstrate how to design an agent that optimizes for this risk, using distribution estimators, as seen in the above schematic.

Paper IV: *In progress* Transfer Reinforcement Learning with Risk

In an ongoing work we are considering techniques that leverage knowledge transfer from a set of *source* domains to a *target* domain. This setting is interesting when you for instance have a task that you know how to solve and you now want to solve a different task but with similar structure. For instance, knowing how to drive in **Europe** should inform you to some extent how to drive in the **US**, but there are some differences, namely traffic rules, road signs and road behavior.

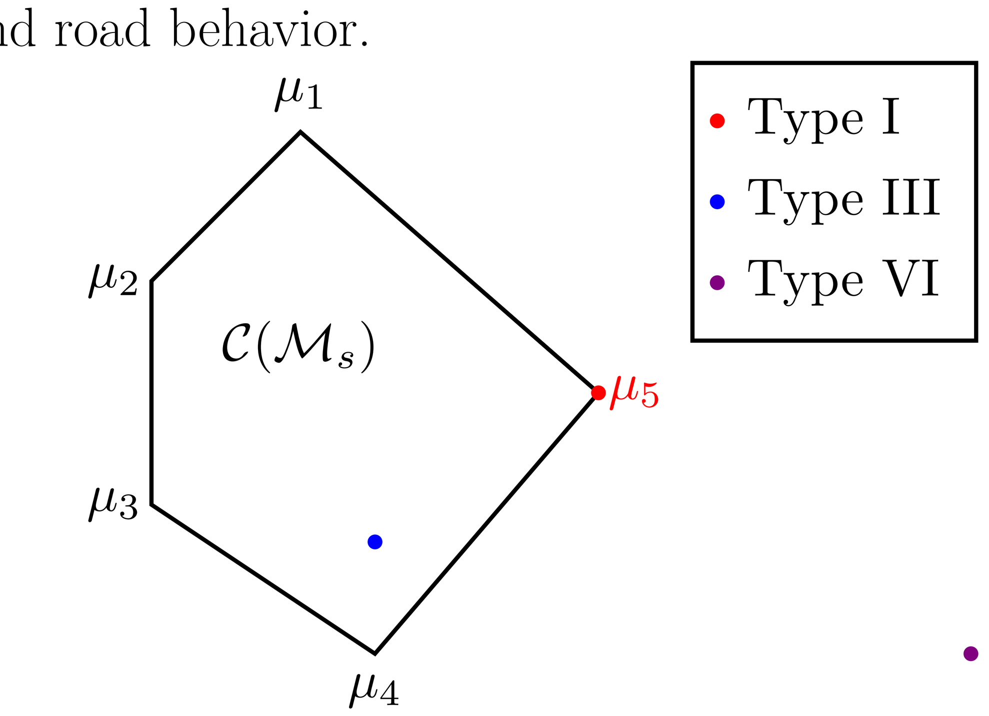

Figure: Overview some of the types of settings that may arise in Transfer Reinforcement Learning.

In particular, we consider three kinds of structures over model space. The first, which gives arise to *Type I*, assumes the target MDP μ_t is part of the finite set of source MDPs $\mathcal{M}_s$. The second, *Type III*, which considers the convex set of source MDPs $\mathcal{C}(\mathcal{M}_s)$, searches for the best representative $\hat{\mu}_t \in \mathcal{C}(\mathcal{M}_s)$. Finally, the last type of structure is the general case where the target MDP can be arbitrarily different from the source MDPs. In that case, you are studying problems of *Type V*.

Acknowledgements

I want to thank all my co-authors and the people I have worked with, some of these are, Emilio Jorge, Divya Grover, Debabrota Basu, Tommy Tram, Christos Dimitrakakis, Mina Ali-Beigi Nabi, Aristide Tossou.