SENTINEL: Taming Uncertainty with Ensemble based Distributional Reinforcement Learning

Hannes Eriksson[1 2], **Debabrota Basu**[3 4], **Mina Alibeigi**[1] & **Christos Dimitrakakis**[2 5]

Zenseact AB, Gothenburg, Sweden[1]

Chalmers University of Technology, Gothenburg, Sweden[2]

Scool, INRIA Lille-Nord Europe, Lille, France[3]

CRIStAL, CNRS, Lille, France[4]

University of Neuchatel, Switzerland and University of Oslo, Norway[5]

Correspondence to: `hannese@chalmers.se`

Abstract

In this work, we consider risk-sensitive sequential decision-making in Reinforcement Learning (RL). Our contributions are two-fold. First, we introduce a novel and coherent quantification of risk, namely composite risk, which quantifies the joint effect of aleatory and epistemic risk during the learning process. We propose an algorithm, SENTINEL-K, based on ensemble bootstrapping and distributional RL for representing epistemic and aleatory uncertainty respectively. The ensemble of K learners uses Follow The Regularised Leader (FTRL) to aggregate the return distributions and obtain the composite risk.

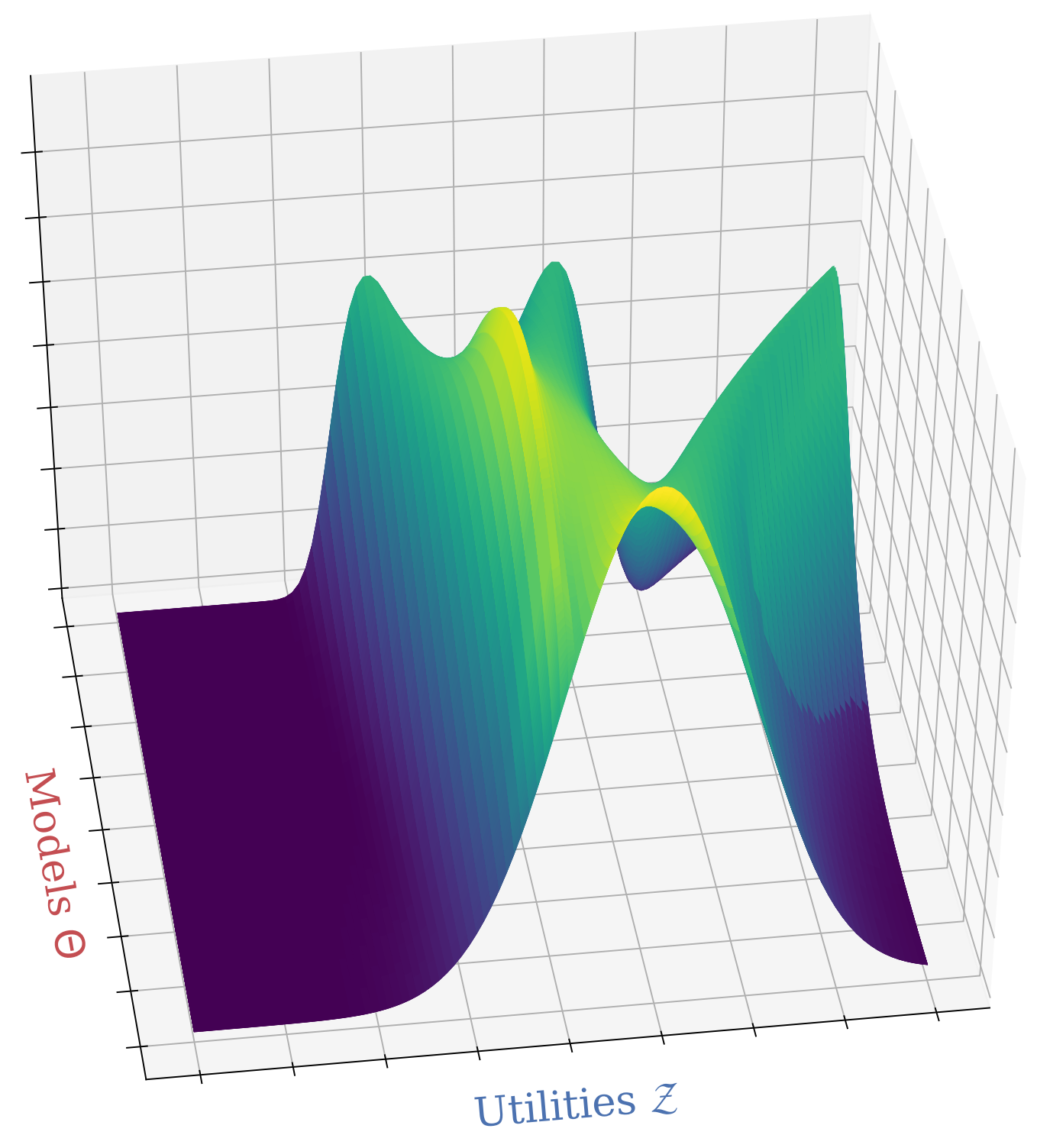

Figure 1: Illustrating the two sources of uncertainty, one related to utility uncertainty ($\mathcal{Z}$) and one related to model uncertainty (Θ).

Contribution. In this work, we propose two main contributions. (1) The *composite risk* formulation, It estimates the total risk more accurately than the previously known additive risk formulation. (2) FTRL as a means of model selection, by weighting each estimator differently instead of model averaging. We empirically demonstrate the superiority of the proposed framework in (i) uncertainty estimation, (ii) performance, and (iii) theoretical properties.

Coherent Composite Risk

A coherent risk measure is *monotonic*, *positive homogenous*, *translation invariant* and *subadditive*.

$$\text{Composite Risk} \triangleq \text{Risk}_{U^E_{\alpha_2}}(\text{Risk}_{U^A_{\alpha_1}}(Z|\theta)|\beta)$$

Theorem 2. Demonstrates the composed risk measure of $U^E_{\alpha_2}$ and $U^A_{\alpha_1}$ is also a coherent risk measure.

Theorem 3. Shows the additive risk formulation is a special case of the composite risk formulation and will in general underestimate the total risk (compared to the composite risk formulation).

SENTINEL-K Algorithm

Our proposed algorithm **SENTINEL-K** is a combination of C51 (Categorical Deep-Q-Network) [1] and Bootstrapped DQN [2], where we have an ensemble of C51 agents.

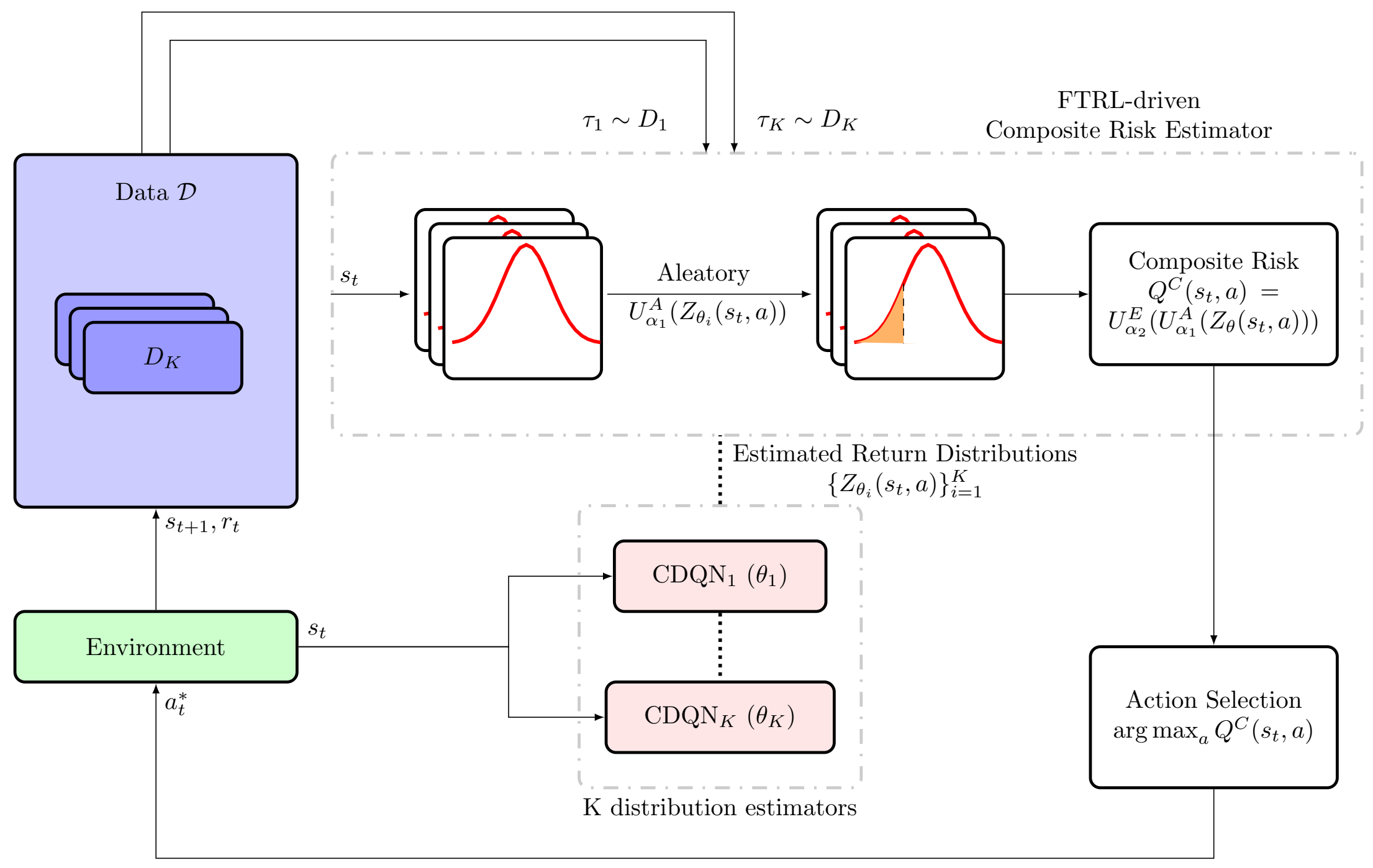

Figure 2: Schema of the proposed algorithm *SENTINEL-K*.

Methods

Experiments

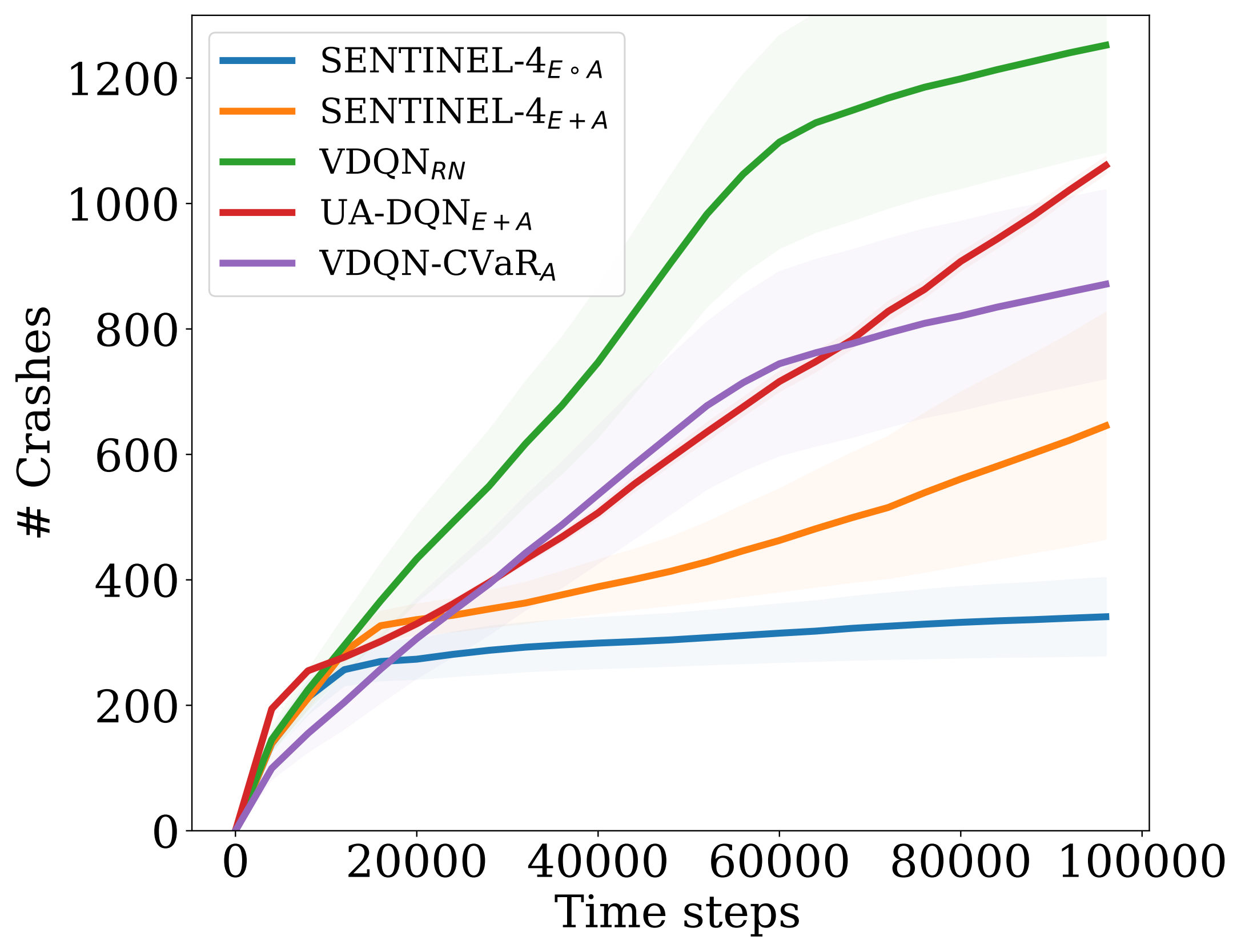

Figure 3: Experiments evaluated for the proposed algorithm (SENTINEL-K) and benchmarked against Variational Deep-Q-Network (VDQN) and Uncertainty-Aware Deep-Q-Network (UA-DQN) for an autonomous driving domain. Fewer # Crashes are better.

Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris facilisis imperdiet nunc, sit amet venenatis eros. Duis ac aliquam arcu. Suspendisse sit amet cursus orci, vel aliquam lacus. Nullam faucibus velit felis, ac vestibulum eros accumsan nec. Praesent sollicitudin venenatis urna quis ornare. Etiam feugiat sagittis iaculis. Maecenas at tellus feugiat metus molestie elementum ut in neque. Donec consequat pulvinar aliquam. Proin ornare rhoncus nisl a eleifend.

Proin vitae urna vel odio tincidunt iaculis. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vivamus at euismod ligula. Nulla maximus diam in arcu malesuada ultricies. Etiam odio est, imperdiet eget justo malesuada, gravida laoreet ante. Nunc placerat sagittis ex sit amet egestas. Proin a euismod magna. Aliquam erat volutpat.

Published at the 38th Conference on Uncertainty in Artificial Intelligence, Eindhoven, Netherlands, 2022. Copyright 2022 by the author(s).

References

[1] Marc G Bellemare, Will Dabney, and Rémi Munos. A distributional perspective on reinforcement learning. In *International Conference on Machine Learning*, pages 449–458. PMLR, 2017.

[2] Ian Osband, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. Deep exploration via bootstrapped dqn. *Advances in neural information processing systems*, 29, 2016.

Acknowledgements

We would like to thank Dapeng Liu for fruitful discussions in the beginning of the project, further, this work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation and the computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE partially funded by the Swedish Research Council through grant agreement no. 2018-05973.