Extreme Risk Mitigation in Reinforcement Learning Using Extreme Value Theory

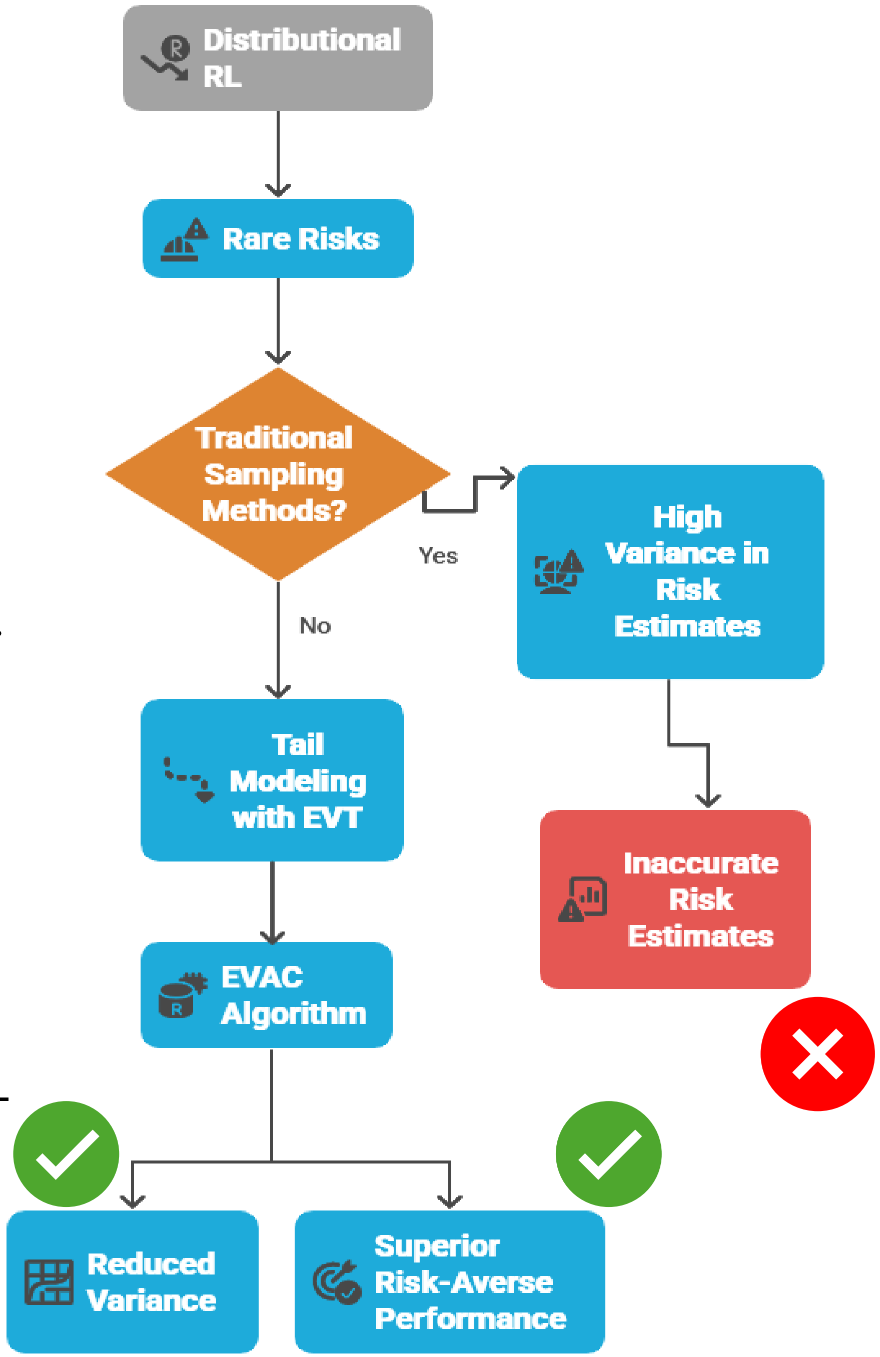
Karthik Somayaji NS   Yu Wang   Malachi Schram,
Jan Drgona   Mahantesh Halappanavar
Frank Liu   Peng Li

Background

- **Distributional RL**
  - Instead of expected return - models the **entire distribution** of possible returns.
  - Captures the full uncertainty and variability in outcomes - essential for **safety-critical decisions**.

- **Rare Risks**
  - Catastrophic events that **occur infrequently** but can have severe consequences.
  - Represent infrequent but catastrophic events lying in the **tail of the return distribution.**
  - Traditional sampling methods struggle to accurately model these low-probability events, leading to **high variance** in risk estimates.

- **Our Contributions**
  - **Tail Modeling with EVT:** Leverage Extreme Value Theory to model the tail of the return distribution using a Generalized Pareto Distribution (GPD).
  - **EVAC Algorithm:** Develop a novel Extreme Valued Actor-Critic that integrates EVT-based tail modeling into distributional RL for improved risk aversion .
  - **Empirical Validation:** Demonstrate through experiments on benchmark environments that our method theoretically and practically reduces variance in risk metrics and achieves superior risk-averse performance.



The Challenge of Modeling Extreme Risks

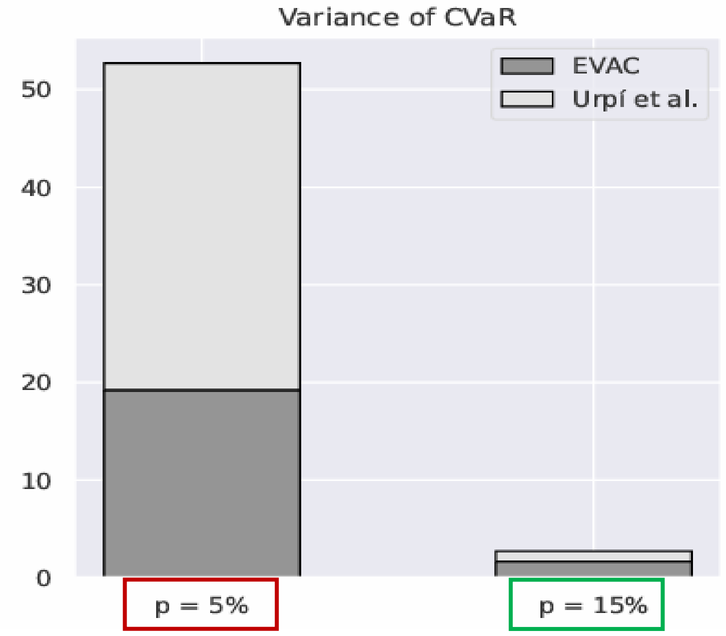
- **Setup**
  - For example, in Mujoco - HalfCheetah environments, we simulate rare risk by applying a penalty via a Bernoulli variable
  - Two risk levels when velocity of agent exceeds threshold:
    – **15% Penalty**: Moderate rare risk
    – **5% Penalty**: Extreme rare risk
  - Evaluation via :

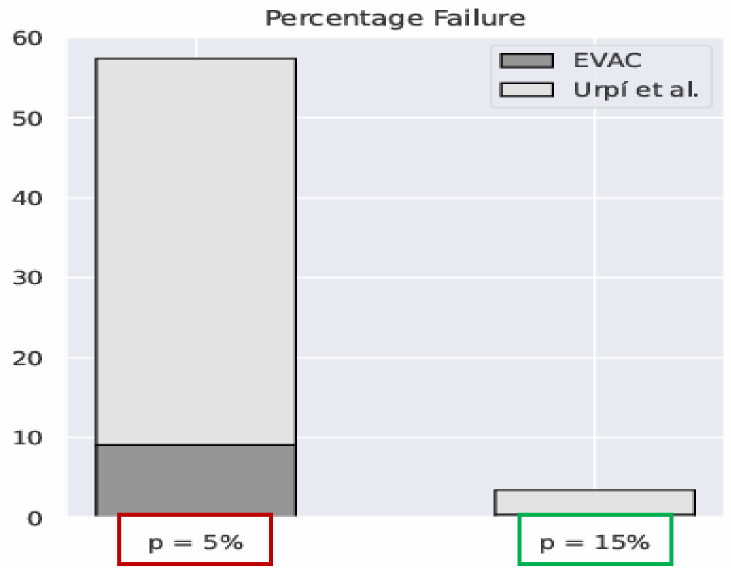
    - **Empirical risk aversion** (Percentage Failure – how many times does agent go to unsafe states)

    - Emprical **CVaR (conditional value at risk) variance**

- **Conclusion**

  – Accurate tail modeling is critical; EVAC reduces variance and enhances risk-sensitive policy convergence





* Núria Armengol Urpí, Sebastian Curi, and Andreas Krause. Risk-averse offline reinforcement learning. In Proc. International Conference on Learning Representations (ICLR), May 2021.

Modeling the Tail Distribution

- **Pickands-Balkema-de Haan Theorem:**
  - Approximates the conditional excess distribution above a high threshold using the Generalized Pareto Distribution ($H_{\xi,\sigma}(x)$)

(Pickands-Balkema-de Haan Theorem) *Pickands III (1975) Let $X_1 \cdots X_n$ be a sequence of IID random variables with distribution function (CDF) given by F whose limiting behavior approaches the GEV distribution. Let $F_u(x) = P(X - u \leq x | X > u)$ be the conditional excess distribution. Then,*

$$\lim_{u \to \infty} F_u(x) \xrightarrow{D} H_{\xi,\sigma}(x),$$

- **Key Idea:**

  - For values above a high threshold u, the conditional excess (i.e., the amount by which values exceed u) converges in distribution to a Generalized Pareto Distribution (GPD).

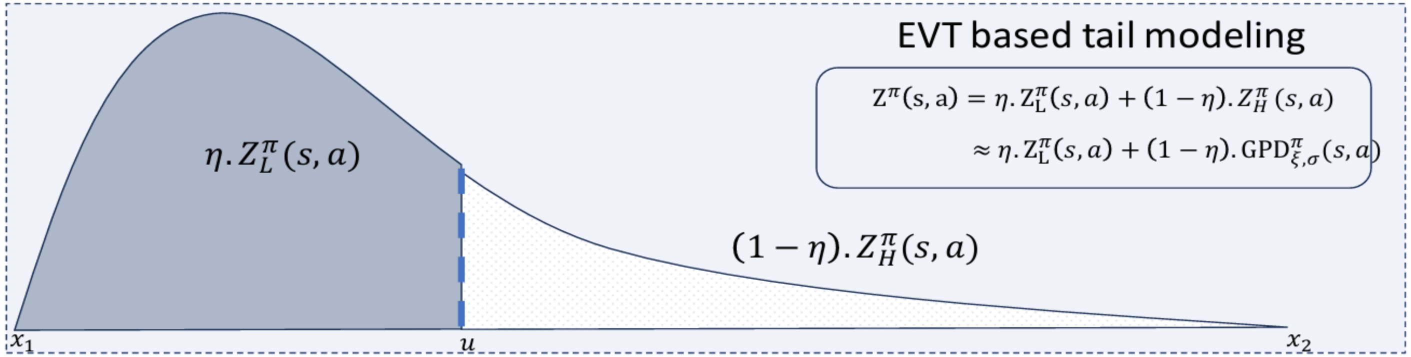
- **Tail Construction Process:**
  – ***Threshold Selection*:** Choose u (or a quantile level η) that partitions the distribution into a non-tail region and a tail region.

  – ***Parameter Estimation*:** Represent excess values $(X - u \mid X > u)$ using the GPD with parameters ξ (shape) and σ (scale). Fit the GPD to the observed excesses (via maximum likelihood or other methods) to obtain robust estimates of the tail behavior.

- **Benefits:**
  – Provides an efficient, parameterized model of extreme events that reduces variance in risk measures (e.g., CVaR) and supports more robust risk-averse policy learning.

Extreme Valued Actor Critic

- **Decomposed Critic Architecture**
  - **Non-Tail Region:**
    - Quantile-based modeling (e.g., standard quantile regression) for values below a chosen threshold (η).
  - **Tail Region:**
    - Parameterized by a Generalized Pareto Distribution (GPD) for excess values above the threshold.
    - GPD parameters (ξ, σ) updated via maximum likelihood estimation to capture rare, extreme events accurately.

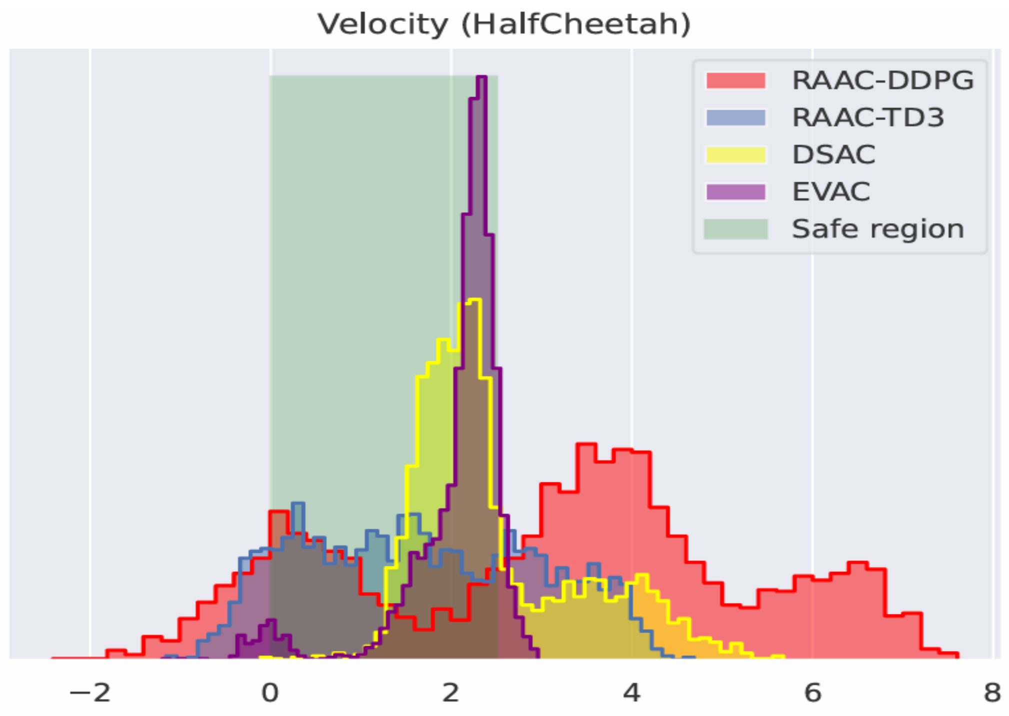


- **Convergence Guarantee**
  We prove that the critic modeling above under a fixed policy is a contraction

- **Integration with Actor-Critic**
  We optimize the actor using a risk measure (e.g., CVaR) derived from the refined (tail + non-tail) distribution.

- **Variance Reduction in quantile regression estimate of the tail region**

Experiments



Algorithm	Percentage Failure	Cumulative Reward	CVaR
RAAC-DDPG	16.55± 4.43	637.81± 319.78	135.18 ± 13.02
RAAC-TD3	41.3± 16.6	836.8± 195.85	129.81 ± 38.75
D-SAC	30.04 ± 22.26	1356.05 ± 269.63	149.76 ± 27.38
EVAC	**2.87 ± 1.3**	**1502.46 ± 94.25**	**156.71 ± 11.07**

Rare Risk Setup (HalfCheetah):

- Penalize the agent whenever its velocity exceeds a threshold (e.g., 2.5).
- Penalty is triggered with a small probability (*p=5%*), simulating extreme rare events.
- Safe region is velocity < 2.5

Rare Risk Setup (Navigation with rare obstacles):

- Penalize the agent whenever it enters the red circles.
- Penalty is triggered with a small probability (*p=5%*), simulating extreme rare events