Statistics and Samples in Distributional Reinforcement Learning

Mark Rowland, Robert Dadashi, Saurabh Kumar, Rémi Munos, Marc G. Bellemare, Will Dabney

ICML 2019
Distributional Reinforcement Learning

Distributional RL aims to learn full return distributions.

Return distribution:

\[ Z^\pi(x, a) = \sum_{t=0}^{\infty} \gamma^t R_t \mid X_0 = x, A_0 = a \]

Distributional Bellman equation:

\[ Z^\pi \overset{D}{=} T^\pi Z^\pi \]

[Bellemare et al., 2017]
In practice, we often work with parametric approximate distributions.

Non-parametric

\[ T^\pi \]
In practice, we often work with parametric approximate distributions.

Non-parametric

Categorical [Bellemare et al., 2017]
In practice, we often work with **parametric approximate distributions**.

- **Non-parametric**
- **Categorical** [Bellemare et al., 2017]
- **Dirac deltas** [Dabney et al., 2018]
Main Contribution: An Alternative Perspective

Distributional RL algorithms learn **statistical functionals** of the return distribution.

- Moments, tail probabilities, expectations, etc.
Main Contribution: An Alternative Perspective

Distributional RL algorithms learn **statistical functionals** of the return distribution.

- Moments, tail probabilities, expectations, etc.

**Theory:** What properties of return distributions can be learnt through dynamic programming?

**Algorithmic:** A general framework for approximate learning of statistics.
Main Contribution: An Alternative Perspective

Distributional RL algorithms learn **statistical functionals** of the return distribution.

- Moments, tail probabilities, expectations, etc.

**Theory:** What properties of return distributions can be learnt through dynamic programming?

**Algorithmic:** A general framework for approximate learning of statistics.
A General Framework for Distributional RL Algorithms

Current statistics

Imputation strategy

Imputed samples

Bellman-updated statistics

$s_1$, $s_2$, $s_K$

$T^\pi$

Bellman-updated distribution
A General Framework for Distributional RL Algorithms

Current statistics

Imputation strategy

Imputed samples

$T^\pi$

Bellman-updated statistics

$s_1$, $s_2$, $s_K$

Bellman-updated distribution
A General Framework for Distributional RL Algorithms

Current statistics

Imputation strategy

Imputed samples

Bellman-updated statistics

\( T^\pi \)

Bellman-updated distribution
A General Framework for Distributional RL Algorithms

Current statistics

Imputation strategy

Imputed samples

Bellman-updated statistics

$s_1$, $s_2$, $s_K$

$T^\pi$

Bellman-updated distribution
A General Framework for Distributional RL Algorithms

Current statistics

Bellman-updated statistics

Imputation strategy

Imputed samples

Bellman-updated distribution
A General Framework for Distributional RL Algorithms

Currently, statistics are used to represent the state of the system. Imputation strategies are employed to fill in missing data. Bellman-updated statistics incorporate new information into the existing statistics. Imputed samples are used to represent the possible outcomes. Bellman-updated distributions reflect the updated state of the system after applying the Bellman operator.
Application: Expectiles

We apply this framework to learn **expectiles** of return distributions.

New deep RL agent: **Expectile Regression DQN (ER-DQN)**, with **improved mean performance** on Atari-57 relative to QR-DQN.
Summary

A new perspective on distributional RL

Theoretical progress on what it is possible to learn

A general framework for distributional RL algorithms
THANK YOU

Poster #113