CHAPTER 1

DEEP REINFORCEMENT LEARNING

1 Neural Networks

...

2 Distributional Reinforcement Learning

Remember that we defined $\mathbb{P}^{\pi}_{s,a} := \mathbb{P}^{\pi} \otimes \delta_{S_0}(s) \otimes \delta_{A_0}(a)$ as the probability measure of the Markov reward process (S, A, R) started in (s, a). We define the distribution of the return under policy π as

$$Z^{\pi} := \sum_{t=0}^{\infty} \gamma^t R_t, \quad \gamma \in (0, 1).$$

Unlike the methods before, where we were interested in the expected reward $Q^{\pi}(s, a) = \mathbb{E}^{\pi}_{s,a}[Z^{\pi}]$, we are now interested in the distribution of these cumulative rewards. For that define the push forward

$$\eta^{\pi}_{s,a}(B) := \mathbb{P}^{\pi}_{s,a}(Z^{\pi} \in B), \quad B \in \mathcal{B}(\mathbb{R}).$$