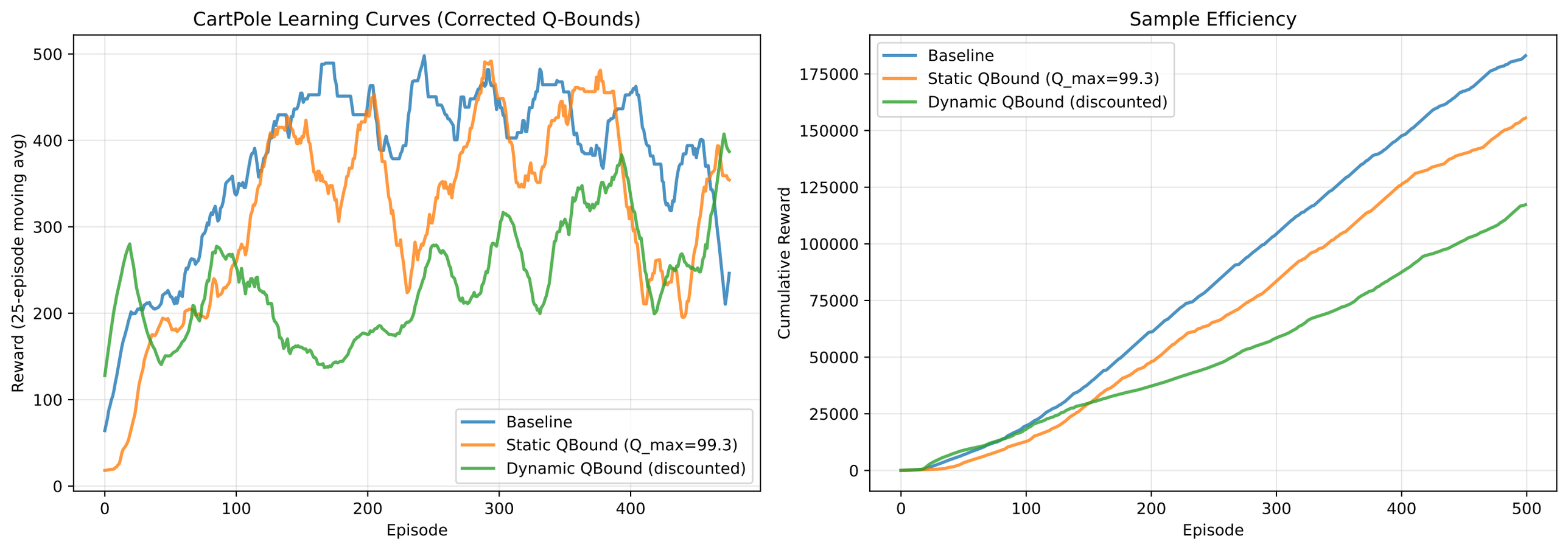
CartPole Learning Curves (Corrected Q-Bounds)
Reward (25-episode moving avg)
Episode
Baseline
Static QBound (Q_max=99.3)
Dynamic QBound (discounted)
Sample Efficiency
Cumulative Reward
Episode
Baseline
Static QBound (Q_max=99.3)
Dynamic QBound (discounted)