Tic-Tac-Toe
Trilokinath Modi

1 Results

For this task, the learning rate was kept constant at 0.1 and the probability of picking optimal move was kept constant at 1 for first 2500 games. For further games, it was decreased by factor 0.9 after every 500 games. This allowed random actions for at least 2500 games thus leading to good exploration. Since player 1 takes the first move in every game, the number of games player 1 won is higher than number of games player 2 won. As the players played enough games, they started to learn and the number of draws kept on increasing. The final game where a player has won appeared at around 48000 games after which only draw appeared. However, the steady state reached at some point after 50000 games. The total number of games played by both players = $1e6$. The figure 1 portrays the fraction of number of wins by each player and also the number of draws. The reward of win, draw and lose was kept constant as $\{+1, 0, -1\}$ respectively.

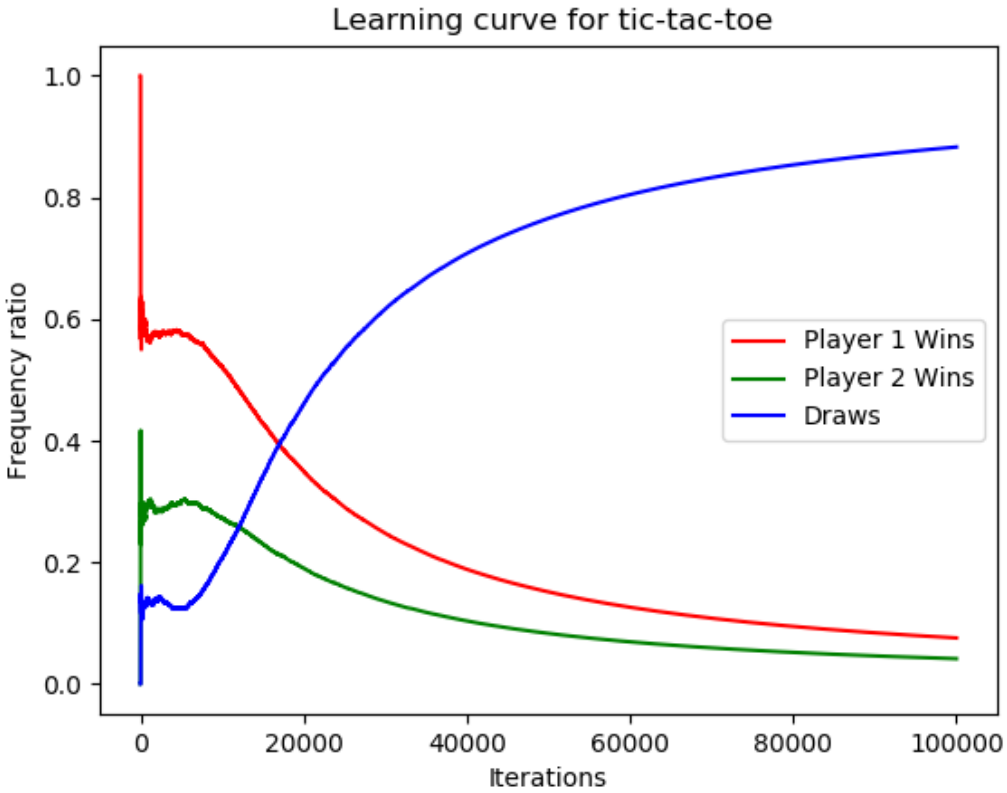

Figure 1: Ratio of games drawn or won by each player against total number of games