Tic Tac Toe

For the task, I started with an $\epsilon = 1$ such that during the first 20,000 games, the agents favor exploration of different states by moving to a random free spot on the board, rather than choosing the best action. After 20,000 games, I start to decay ϵ with a factor of 0.95 every 100 games. This is to gradually favor exploitation of the best actions. Finally, when ϵ has reached 0, we only choose the best actions possible in order to reach a draw, i.e for the players to maximize the future expected reward. The procedure was repeated for 100,000 games, just to make sure that no player ever wins.

Every time a game ended, i incremented a counter for player 1 wins, player 2 wins or draw, based on what the outcome was. Then after 250 games, i calculated the frequencies of wins and draw based on the round interval of 250 games. Finally i plotted these points in the same figure, see Figure 1. Studying Figure 1 further we can see that the frequencies of wins and draw are relatively stable, they oscillate around a certain point. However, when we begin with the ϵ-greedy policy, we see that the probability of draws increase drastically, and the probability of a player winning decreases. We also see that the probability of the outcome being a draw converges to 1.

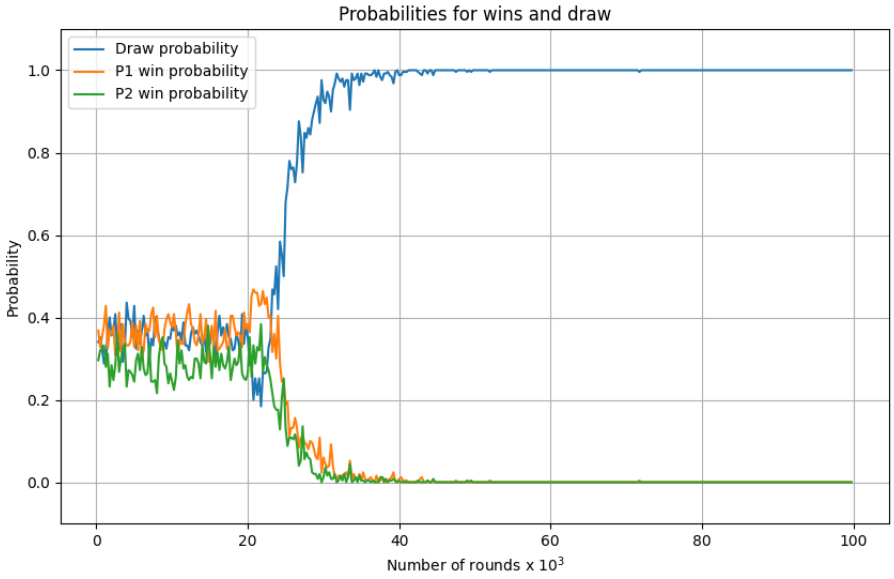

Figure 1: The figure represents the convergence of whether player 1 wins, players 2 wins, or if they end up in a draw.