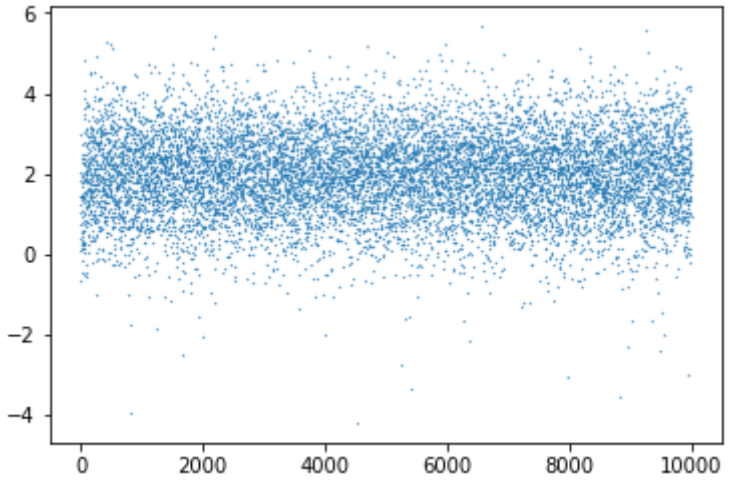
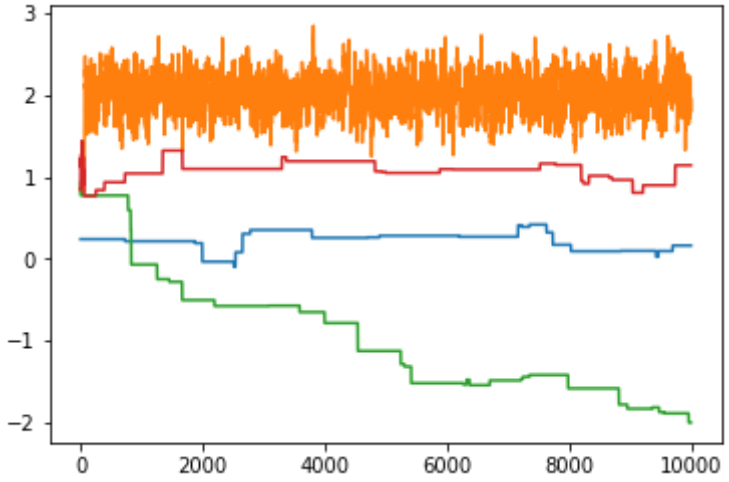
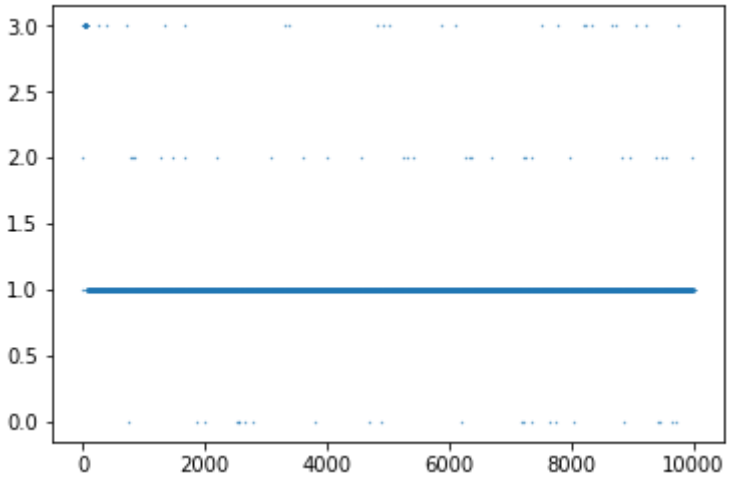
```
import numpy as np
import matplotlib.pyplot as plt

K= 4
action_values = np.array([0, 2, -2, 1])

# experiment
T = int(1e4)
eps = 1e-2
alpha = 1e-1
G = []
estimates = np.zeros([T + 1, K])
estimates[0] = np.random.random(K)
actions = []
for t in range(T):
    if np.random.random() < eps:
        action = np.random.randint(len(action_values))
    else:
        action = np.argmax(estimates[t])
    reward = np.random.normal(action_values[action], 1.0)
    actions.append(action)
#     temp = np.zeros(
    estimates[t + 1] = estimates[t]
    estimates[t + 1][action] += alpha*(reward - estimates[t + 1][action])
    G.append(reward)
# print(np.cumsum(G))

plt.scatter(x = np.arange(T), y = G, s = 0.1)
plt.figure()
plt.plot(estimates)
plt.figure()
plt.scatter(x = np.arange(T), y = actions, s = 0.1)
plt.show()

```


6
4
2
0
−2
−4
0
2000
4000
6000
8000
10000


3
2
1
0
−1
−2
0
2000
4000
6000
8000
10000


3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
2000
4000
6000
8000
10000

```
1
```