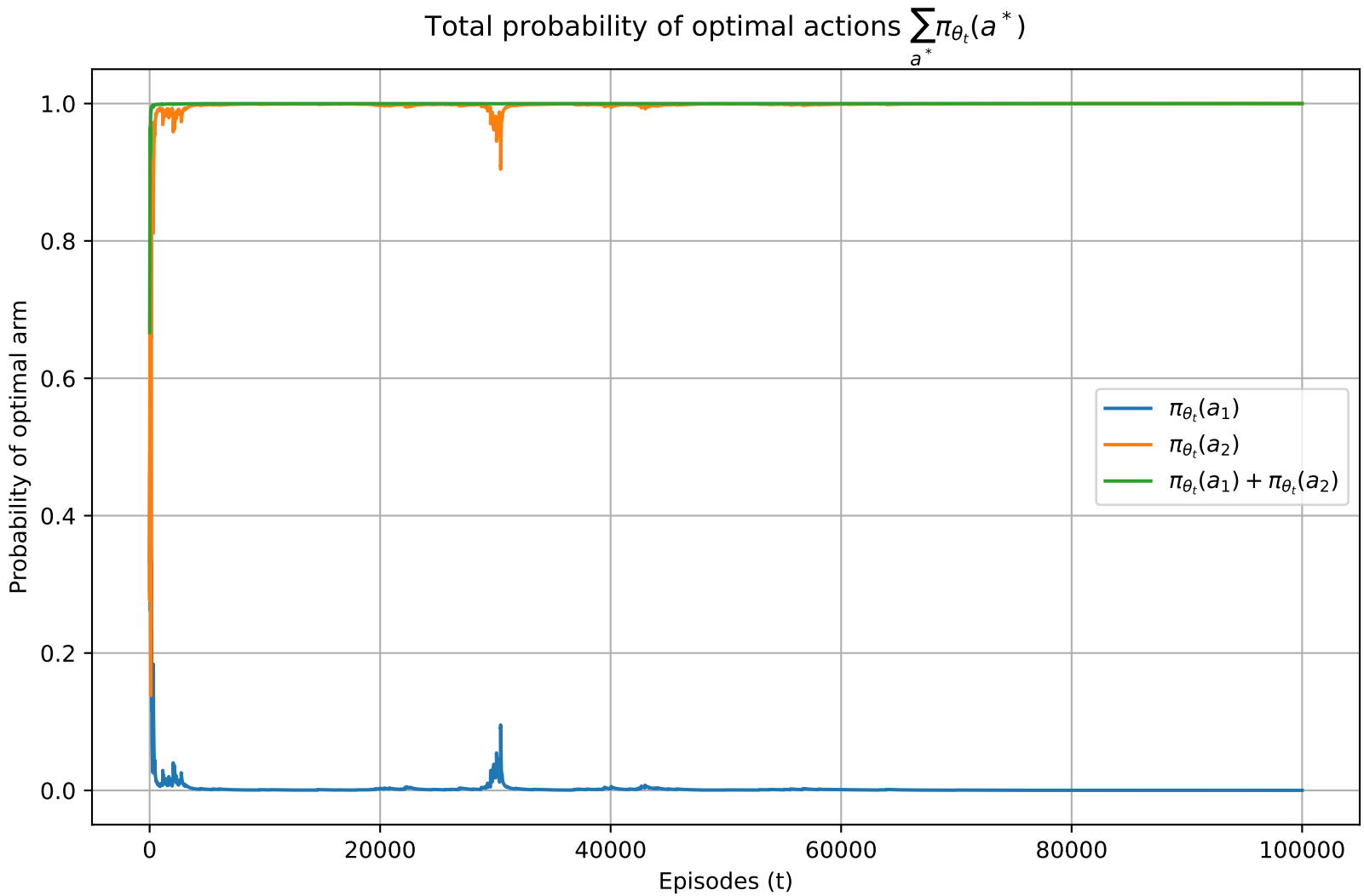
Total probability of optimal actions $\sum_{a^*} \pi_{\theta_t}(a^*)$
Probability of optimal arm
Episodes (t)
1.0
0.8
0.6
0.4
0.2
0.0
0
20000
40000
60000
80000
100000
$\pi_{\theta_t}(a_1)$
$\pi_{\theta_t}(a_2)$
$\pi_{\theta_t}(a_1) + \pi_{\theta_t}(a_2)$