Stochastic differential equations for limiting description of UCB rule for Gaussian multi-armed bandits

Sergey Garbar*

August 19, 2022

We consider the upper confidence bound strategy for Gaussian multi-armed bandits with known control horizon sizes N and build its limiting description with a system of stochastic differential equations and ordinary differential equations. Rewards for the arms are assumed to have unknown expected values and known variances. A set of Monte-Carlo simulations was performed for the case of close distributions of rewards, when mean rewards differ by the magnitude of order $N^{-1/2}$, as it yields the highest normalized regret, to verify the validity of the obtained description. The minimal size of the control horizon when the normalized regret is not noticeably larger than maximum possible was estimated.

*Yaroslav-the-Wise Novgorod State University, Sergey.Garbar@novsu.ru. The reported study was funded by RFBR, project number 20-01-00062.