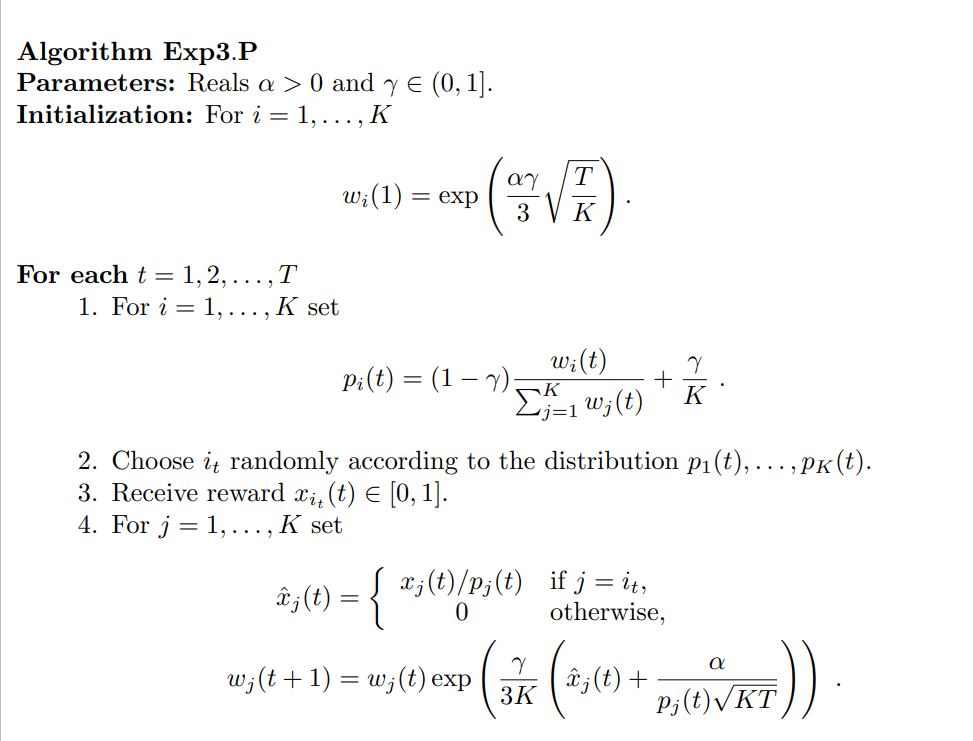
**EXP3.P**

Algorithm Design and Basic Code Structure:

The EXP3.P algorithm is an extension of the EXP3 algorithm, which is a well-known algorithm for multi-armed bandit problems. The EXP3.P algorithm is specifically designed for the setting where the rewards are probabilistic and drawn from an unknown distribution. The high-level description of the EXP3.P algorithm is given as below:



This algorithm combines the concepts of exponential weighting, randomization and probability matching to balance exploration and exploitation. The gamma parameter controls the degree of exploration, where larger values encourage more exploration, while smaller values focus on exploiting arms with higher estimated reward.

\* To control the variance, we modify algorithm Exp3 so that it uses estimates which are based on upper confidence bounds instead of estimates with the correct expectation.

\*The exact code has been attached as a python file with this report. The code contains comments at suitable places to enable the reader to get some insights into the exact implementation.

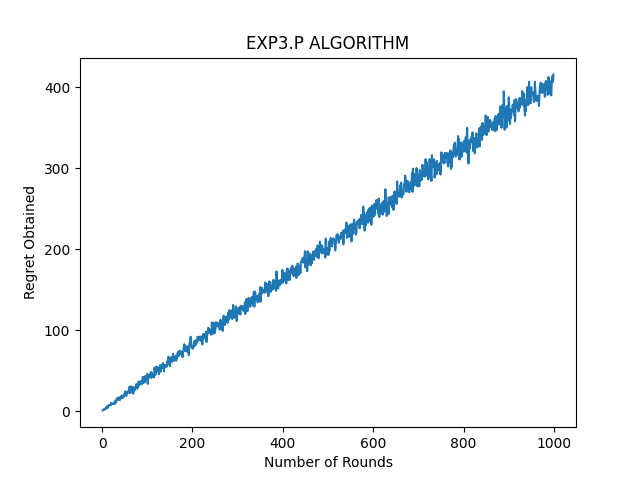
Bandit Instances

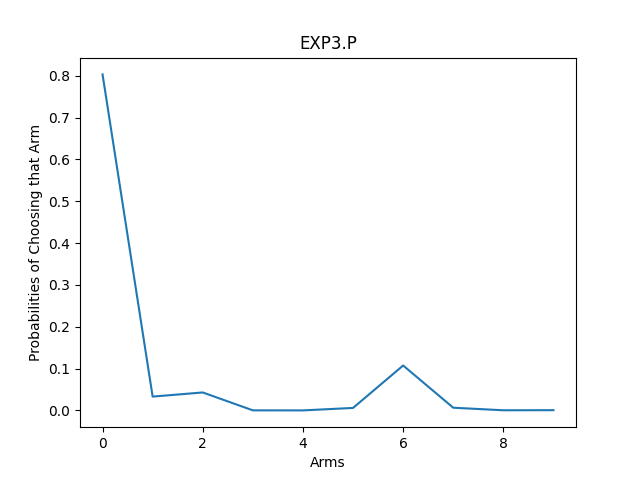
* Number of arms = This can vary depending on the problem domain and its complexity.
* Parameters = The parameters have been adjusted to produce a peak for a suitable arm.
* Reward Distribution = This is generated for each round using a random function.

Assumptions

* Independence of Arms: The algorithm assumes that the rewards from different arms are independent from each other thereby enabling the user to treat each arm as a separate entity.
* Known Number of Arms: The total number of arms are known a prior the algorithm is run.
* Probabilistic and Stationary Rewards: The rewards are drawn from some fixed distribution.

Graph:





References:

* <http://rob.schapire.net/papers/AuerCeFrSc01.pdf>
* chat.openai.com for debugging.

Work completed, compiled and corrected by

Saksham Rathi