Multi-armed bandit

Marc Lelarge (marc.lelarge@ens.fr)

In probability theory, the multi-armed bandit problem is a problem in which a gambler at a row of slot machines (sometimes known as "one-armed bandits") has to decide which machines to play, how many times to play each machine and in which order to play them. When played, each machine provides a random reward from a distribution specific to that machine. The objective of the gambler is to maximize the sum of rewards earned through a sequence of lever pulls.

Inspired by cognitive radio networks, we will consider a setting where multiple users share several channels modeled as a multi-user multi-armed bandit (MAB) problem [1]. The characteristics of each channel are unknown and are different for each user. Each user can choose between the channels, but her success depends on the particular channel chosen as well as on the selections of other users: if two users select the same channel their messages collide and none of them manages to send any data.

References

[1] Orly Avner, Shie Mannor. Multi-user lax communications: a multi-armed bandit approach.