Communicating in Multi-Armed Multiplayer Bandits

Raymond Zhang

Supervisors: Richard Combes & Sheng Yang

Multi-Armed Multiplayer Bandits (MMAB)

J players that do not know how many they are.

At each time $t \in [T]$:

- Each player $j \in [J]$ selects an arm $A_j(t) \in [K]$.
- Observes a reward: $Y_j(t) = X_{A_j(t)}(t)\left(1 - C_{A_j(t)}(t)\right)$.

Where :

- $X_k(t)$ is the reward of arm k at time t.
- $X(t)$ i.i.d. following $\bigotimes_{k\in[K]} Ber(\mu_k)$ distribution.
- $\mu_k \in [0,1]$ the average reward of arm k.
- $C_k(t) = \{|\{j \in [J], A_j(t) = k\}| \geqslant 2\} \in \{0,1\}$ is the collision indicator.
- The players do not observe $C_k(t)$ nor know the μ_k's.

The goal is to minimize the regret:

$$R(T) = T\sum_{j=1}^{J}\mu_{(j)} - \mathbb{E}\left[\sum_{t=1}^{T}\sum_{j=1}^{J} X_{A_j(t)}(t)\left(1 - C_{A_j(t)}(t)\right)\right] \qquad (1)$$

Synchronisation and Communication with Collisions

The channel model from player 1 perspective is defined as follows :

$$Y_1(t) = 0 \text{ if } A_1(t) = A_2(t)$$

Otherwise,

$$Y_1(t) = \begin{cases} 0 & \text{w.p. } 1 - \mu_{A_1(t)} \\ 1 & \text{w.p. } \mu_{A_1(t)} \end{cases}$$

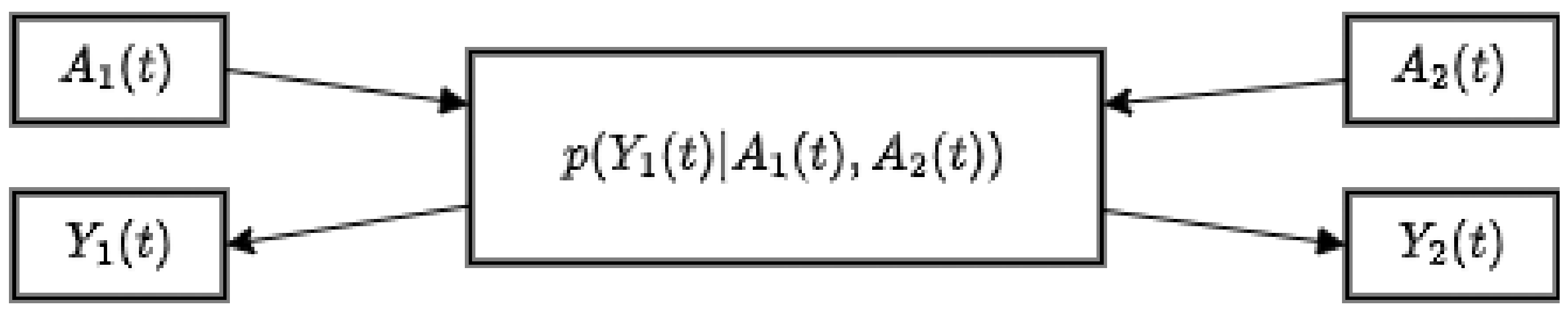

Figure 1. MMAB channel model for 2 players

There are known protocols for synchronizing the players and communicating the μ's estimates [1], [2]. However, we believe that they are suboptimal. They need the knowledge of $\mu_{(K)}$ or require an excessively long time to synchronize.

Optimality Gap

An asymptotic upper and lower bound of the centralized regret problem is given by:

$$R_{\text{centralized}}(T) = O\left(\sum_{k>J}\frac{1}{\mu_{(J)} - \mu_{(k)}}\log T\right) \qquad (2)$$

One can notice that if the number of players is larger or equal to the number of arms, the regret of the centralized algorithm is exactly 0.

Regarding the regret of the decentralized algorithm, the regret due to synchronization and communication is dominated by :

$$O\left(K^2 J \ln(T)\right) \text{ and } O\left(KJ^2\ln\left(\frac{1}{\mu_{(J)} - \mu_{(J+1)}}\right)\ln(T)\right) \qquad (3)$$

The goal is to find a lower bound on the regret due to communication and synchronization using information-theoretic tools.

Study of a two players case $J = 2$

Assuming that player 1 and player 2 use the same random code book with distribution $(p_k)_{k\in[K]} \in \Delta^K$.

Using a random coding argument, see chapter 5 of [3], the probability that player 1 makes an error decoding the message of player 2 is upper bounded by:

- $P_{e_1} < (M-1)^\rho\left[\sum_{y\in\{0,1\}}\sum_{x_1\in[K]} p_{x_1}\left[\sum_{x_2\in[K]} p_{x_2} p(y|x_1,x_2)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right]^n$

The term inside the outer brackets can be decomposed and written as a sum of two terms:

- For $y = 0$,
$$\sum_{x_1\in[K]} p_{x_1}(1 - p_{x_1})^{1+\rho}\mu_{x_1} \qquad (4)$$
- for $y = 1$,
$$\sum_{x_1\in[K]} p_{x_1}\left[p_{x_1} + (1 - p_{x_1})(1 - \mu_{x_1})^{\frac{1}{1+\rho}}\right]^{1+\rho} \qquad (5)$$

Note that taking minus the logarithm of (4)+(5) we get the random coding error exponent $E_0(p,\rho,\mu)$ for the MMAB problem.

The objective is to find the optimal $p^\star_{\rho,\mu}$ that minimizes the error exponent $E_0(p,\rho,\mu)$.

Special cases

For $\rho = 1$, the optimal $p^\star$ is such that:

$$\forall \mu, ||p^\star(1,\mu)||_0 = 2 \text{ and } p^\star_k(1,\mu) = 0 \text{ if } \mu_k < \mu_{(2)} \qquad (6)$$

Using second order KKT conditions on a minimum and upper bounding the second derivative of (4)+(5). We can show that :

$$\forall \mu, \rho, ||p^\star(\rho,\mu)||_0 \leqslant 3 \qquad (7)$$

$p^\star_{\rho,\mu}$ has at most 3 non zero components.

Work in Progress

- Finish to show that the optimal $p^\star$ has only 2 non-zero components on the best arms.
- What happens with more than 2 players?
- What happens when $J > K$?
- Is the error exponent convex in ρ? Do we get a notion of mutual information if we differentiate the error exponent at 0? Can we get a notion of capacity?
- What happens in the mismatched case?
- Find effective codebooks and messages to be transmitted.

References

[1] E. Boursier and V. Perchet, *SIC-MMAB: Synchronisation Involves Communication in Multiplayer Multi-Armed Bandits*, 2019.

[2] W. Huang, R. Combes, and C. Trinh, *Towards Optimal Algorithms for Multi-Player Bandits without Collision Sensing Information*, 2022.

[3] R. Gallager, *Information Theory and Reliable Communication*, en. Vienna: Springer Vienna, 1972.