Lower Bound for Multi-Player Bandits: Erratum for the paper *Multi-player bandits revisited*

Lilian Besson and Emilie Kaufmann

September 4, 2019

This note is aimed at correcting the statement of Theorem 6 in [Besson and Kaufmann, 2018], which provides a lower bound on the number of arms selections and on the regret in multi-player bandits. Theorem 6 was claimed to improve a result established by [Liu and Zhao, 2010], also based on a change-of-distribution argument. However, [Boursier and Perchet, 2019] pointed out that these two lower bounds are wrong by exhibiting an algorithm, called SIC-MMAB, achieving a regret which is smaller to the lower bound in Theorem 6. Here we clarify what is wrong or right about the lower bound in Theorem 6 of [Besson and Kaufmann, 2018]. It turns out that this result still applies to *some algorithms*, but indeed not to algorithms exploiting collision information in the way SIC-MMAB does. More precisley, we provide a sufficient condition on algorithms, that says in spirit "the collisions do not bring too much information on the arm means", under which the existing lower bound is valid.

Consider Bernoulli arms with mean utilities $\mu_1, \dots, \mu_K$ (shared by all players) and assume, to simplify the presentation, that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$, with $\mu_M > \mu_{M+1}$. Letting $N_k^m(T)$ be the number of selections of arm k by player m, Theorem 6 claims that, for each player m and each arm $k > M$, any uniformly efficient algorithm $\mathcal{A}$ satisfies

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_k^m(T)]}{\log(T)} \geq \frac{1}{\mathrm{kl}(\mu_k, \mu_M)}. \tag{1}$$

Fix a player m and consider the observations $\mathcal{O}_t$ gathered by this player after t round of an algorithm:

$$\mathcal{O}_t = (U_1, Y_1, C_1, \dots, U_t, Y_t, C_t),$$

where $Y_t := Y_{A^m(t),t}$ denotes the sensing information, $C_t := \mathbb{1}_{C_{A^m(t),t}}$ denotes the collision information and U_t denotes some external source of randomness useful to select $A^m(t+1)$. By definition of a sequential strategy, $A^m(t+1)$ is $\sigma(\mathcal{O}_t)$-measurable. The (wrong) lower bound (1) is obtained by combining two steps, among which we highlight that the first is correct.

The first step is to introduce, for $\epsilon > 0$, the alternative model parameterized by $\boldsymbol{\lambda}$ such that

$$\begin{cases} \lambda_\ell &= \mu_\ell \quad \text{for all } \ell \neq k, \\ \lambda_k &= \mu_{M^*} + \varepsilon. \end{cases}$$

This model is such that $M\text{-best}_\mu = \{1, \dots, M\}$ and $M\text{-best}_\lambda = \{1, \dots, (M-1), k\}$. Information theoretic arguments given in [Garivier et al., 2019] show that for any event $\mathcal{E}_T$ that is $\sigma(\mathcal{O}_T)$ measurable,

$$\mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_T}, \mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_T}\right) \geq \mathrm{kl}\left(\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{E}_T), \mathbb{P}_{\boldsymbol{\lambda}}(\mathcal{E}_T)\right),$$

where $\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_T}$ (resp. $\mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_T}$) is the distribution of the vector $\mathcal{O}_T$ under the model $\boldsymbol{\mu}$ (resp. $\boldsymbol{\lambda}$) when algorithm $\mathcal{A}$ is applied and KL denotes the Kullback-Leibler divergence. Now the event

$$\mathcal{E}_T = \left(N_k^m(T) > \frac{T}{2M} \right)$$

is supposed to have a small probability under $\boldsymbol{\mu}$ (under which k is sub-optimal) and a large probability under $\boldsymbol{\lambda}$ (under which k is one of the optimal arms, and is likely to be drawn a lot). More precise arguments detailed in [Besson and Kaufmann, 2018] permit to show that

$$\liminf_{T\to\infty} \frac{\mathrm{kl}\left(\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{E}_T), \mathbb{P}_{\boldsymbol{\lambda}}(\mathcal{E}_T)\right)}{\log(T)} \geq 1,$$

which yields

$$\liminf_{T\to\infty} \frac{\mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_T}, \mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_T}\right)}{\log(T)} \geq 1. \tag{2}$$

We emphasize here that the statement (2) may still be useful to derive a lower bound. The wrong conclusion in [Besson and Kaufmann, 2018] came from the second step, which is the computation of the Kullback-Leibler divergence $\mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_t}, \mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_t}\right)$. We detail (correct) computation of this quantity now, that rely on the chain rule for KL-divergence, computing some terms and using induction (more details can be found in [Besson and Kaufmann, 2018]):

$$\begin{aligned}
\mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_T}, \mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_T}\right) &= \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_{T-1}}\right) + \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{Y_T, C_T, U_T|\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{Y_T, C_T, U_T|\mathcal{O}_{T-1}}\right) \\
&= \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_{T-1}}\right) + \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{Y_T|\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{Y_T|\mathcal{O}_{T-1}}\right) + \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{C_T|\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{C_T|\mathcal{O}_{T-1}}\right) \\
&= \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_{T-1}}\right) + \mathbb{E}_{\boldsymbol{\mu}}\left[\sum_{\ell=1}^{K} \mathbb{1}_{(A^m(T)=\ell)} \mathrm{kl}\left(\mu_\ell, \lambda_\ell\right)\right] + \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{C_T|\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{C_T|\mathcal{O}_{T-1}}\right) \\
&= \sum_{\ell=1}^{K} \mathbb{E}[N_\ell^m(T)] \mathrm{kl}\left(\mu_\ell, \lambda_\ell\right) + \sum_{t=1}^{T} \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{C_T|\mathcal{O}_{T-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{C_T|\mathcal{O}_{T-1}}\right) \\
&= \mathbb{E}[N_k^m(T)] \mathrm{kl}\left(\mu_k, \mu_M + \epsilon\right) + \sum_{t=1}^{T} \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{C_t|\mathcal{O}_{t-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{C_t|\mathcal{O}_{t-1}}\right),
\end{aligned}$$

where the last step uses that in the alternative model $\boldsymbol{\lambda}$ there is only a single arm that is modified, hence $\mathrm{kl}\left(\mu_\ell, \lambda_\ell\right) = 0$ except for $\ell = k$.

We introduce the collision information term for algorithm $\mathcal{A}$ as

$$\mathcal{I}_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathcal{A}, T) := \sum_{t=1}^{T} \mathrm{KL}\left(\mathbb{P}_{\boldsymbol{\mu}}^{C_t|\mathcal{O}_{t-1}}, \mathbb{P}_{\boldsymbol{\lambda}}^{C_t|\mathcal{O}_{t-1}}\right).$$

The wrong claim in [Besson and Kaufmann, 2018] is that the collision information term is zero for any algorithm. However, Lemma 1 below provides a sufficient condition on the algorithm $\mathcal{A}$ for the lower bound (1) to still be correct. This condition is that the collision information term is negligible with respect to $\log(T)$. As the lower bound (1) is violated by the SIC-MMAB algorithm of [Boursier and Perchet, 2019], this algorithm doesn't have a negligible information term. This is expected as under SIC-MMAB the collisions depend on the empirical means of other arms, which vary between two models $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$. However, the condition in Lemma 1 may be true for other algorithms, like the Rand-TopM algorithm of [Besson and Kaufmann, 2018].

Lemma 1. *Any uniformly efficient algorithm that satisfies* $\mathcal{I}_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathcal{A},T) = o(\log(T))$ *satisfies that for any player* m *and sub-optimal arm* k

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_k^m(T)]}{\log(T)} \geq \frac{1}{\mathrm{kl}(\mu_k,\mu_M)}.$$

Proof. From the above, for all $\epsilon > 0$ we have

$$\begin{aligned}\liminf_{T\to\infty} \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_k^m(T)]\mathrm{kl}(\mu_k,\mu_M+\epsilon)}{\log(T)} &= \liminf_{T\to\infty} \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_k^m(T)]\mathrm{kl}(\mu_k,\mu_M+\epsilon) + \mathcal{I}_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathcal{A},T)}{\log(T)} \\ &= \liminf_{T\to\infty} \frac{\mathrm{KL}(\mathbb{P}_{\boldsymbol{\mu}}^{\mathcal{O}_T},\mathbb{P}_{\boldsymbol{\lambda}}^{\mathcal{O}_T})}{\log(T)} \geq 1\end{aligned}$$

Hence, for all $\epsilon > 0$,

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_k^m(T)]}{\log(T)} \geq \frac{1}{\mathrm{kl}(\mu_k,\mu_M+\epsilon)}$$

and the conclusion follows by letting ϵ go to zero. □

References

[Besson and Kaufmann, 2018] Besson, L. and Kaufmann, E. (2018). Multi-player Bandits Revisited. In *Algorithmic Learning Theory (ALT)*.

[Boursier and Perchet, 2019] Boursier, E. and Perchet, V. (2019). SIC-MMAB: synchronisation involves communication in multiplayer multi-armed bandits.

[Garivier et al., 2019] Garivier, A., Ménard, P., and Stoltz, G. (2019). Explore first, exploit next: The true shape of regret in bandit problems. *Math. Oper. Res.*, 44(2):377–399.

[Liu and Zhao, 2010] Liu, K. and Zhao, Q. (2010). Distributed learning in Multi-Armed Bandit with multiple players. *IEEE Transaction on Signal Processing*, 58(11):5667–5681.