1153-37-204 **Erol Pekoz**, **Sheldon Ross** and **Zhengyu Zhang*** (zhan892@usc.edu). *Dueling bandit problems.* Preliminary report.

The problem of using pairwise comparisons to determine a consumer's favorite item is a timely active research area with many applications in online content and product recommendation systems. In the formulation as a bandit problem, there is a set of n bandits and at every stage 2 of the bandits are chosen to play a game, with the result of a game being learned. In the "weak regret problem", we suppose there is a "best" bandit that wins each game it plays with probability at least $p > 1/2$, with the value of p being unknown. The objective is to choose bandits to maximize the number of times that one of the bandits in the game is the best bandit overall. In the "strong regret problem" we suppose that bandit i has unknown value v_i, $i = 1, ..., n$, and that i beats j with probability $v_i/(v_i + v_j)$. We are maximizing the number of stages in which the arm with largest value is declared to be best. In the weak regret problem we propose a policy and obtain an analytic bound on the expected number of stages over an infinite time frame that the best arm is not one of the competitors when this policy is employed. In the strong regret problem we propose a Thompson sampling type algorithm and empirically show that its performance matches the state-of-art algorithm (Received August 27, 2019)