

PH.D. SEMINAR ON BANDIT MODELS

TENTATIVE SYLLABUS

INSTRUCTOR

Alessandro Arlotto
The Fuqua School of Business
Duke University
Email: aa249@duke.edu

OVERVIEW

You enter a famous casino in Las Vegas and you are facing several slot machines (bandits with one arm). The probability of winning for each slot machine is unknown, and you need to choose which slot machine to play to maximize the expected sum of the rewards earned through a sequence of arm pulls. How would you do that?

The theory of bandit models studies the sequential allocation of resources (or efforts) between a number of competing projects. Examples include job scheduling, random search problems, project selection, allocation of treatments in clinical trials, employment decisions and, of course, gambling.

Since their first appearance in the 1950s, bandit models have received incredible attention from researchers in statistics, management sciences, economics and computer sciences. This Ph.D. seminar covers the basics of the bandit framework, several classical results, as well as some more recent approaches.

LIST OF TOPICS

1. Markovian bandits I: examples, formulation, and preliminary results

Further readings:

- Berry, D. A. and Fristedt, B. (1985). *Bandit problems: sequential allocation of experiments*. Monographs on Statistics and Applied Probability. Chapman & Hall, London

- Gittins, J. C., Glazebrook, K. D., and Weber, R. (2011). *Multi-armed bandit allocation indices*. John Wiley & Sons, Ltd., Chichester

2. Markovian bandits II: optimality of the Gittins index

Further readings:

- Berry, D. A. and Fristedt, B. (1985). *Bandit problems: sequential allocation of experiments*. Monographs on Statistics and Applied Probability. Chapman & Hall, London
- Gittins, J. C., Glazebrook, K. D., and Weber, R. (2011). *Multi-armed bandit allocation indices*. John Wiley & Sons, Ltd., Chichester

3. Stochastic bandits: regret analysis

Further readings:

- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58:527–535
- Lai, T. L. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.*, 6(1):4–22
- Agrawal, R. (1995). Sample mean based index policies with $O(\log n)$ regret for the multi-armed bandit problem. *Adv. in Appl. Probab.*, 27(4):1054–1078
- Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122

4. Stochastic bandits: probably-approximately-correct analysis

Further readings:

- Even-Dar, E., Mannor, S., and Mansour, Y. (2002). PAC bounds for multi-armed bandit and Markov decision processes. In *Computational learning theory (Sydney, 2002)*, volume 2375 of *Lecture Notes in Comput. Sci.*, pages 255–270. Springer, Berlin
- Mannor, S. and Tsitsiklis, J. N. (2004). The sample complexity of exploration in the multi-armed bandit problem. *J. Mach. Learn. Res.*, 5:623–648