Karlsruhe Institute of Technology

Paper https://doi.org/10.1145/3637528.3671833

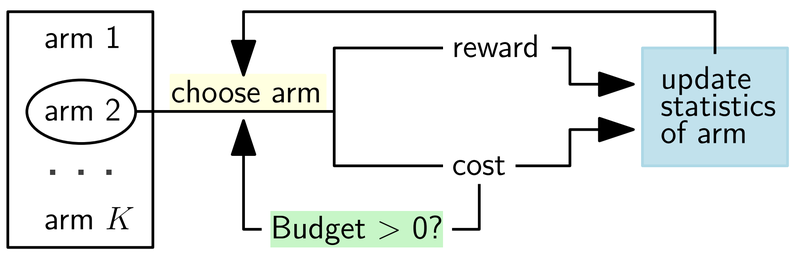
GitHub https://github.com/heymarco/OmegaUCB

Budgeted Multi-Armed Bandits with Asymmetric Confidence Intervals

Marco Heyden (marco.heyden@kit.edu), Vadim Arzamasov, Edouard Fouché, Klemens Böhm

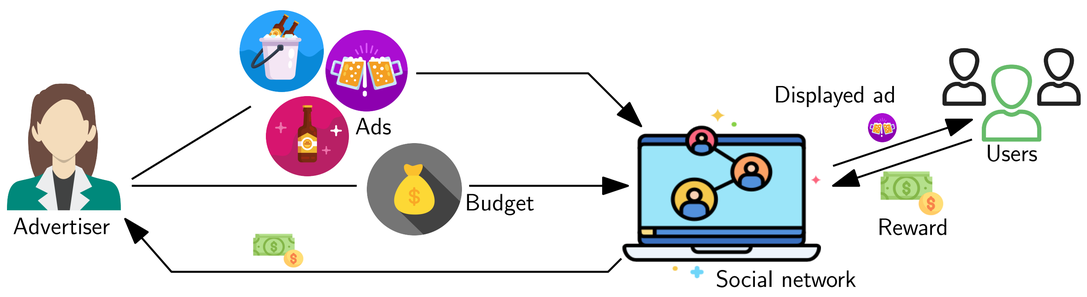
Budgeted Multi-Armed Bandits

- While budget B not empty:
  - Play one of K arms
  - Observe reward, pays cost (both are random)
  - Update strategy



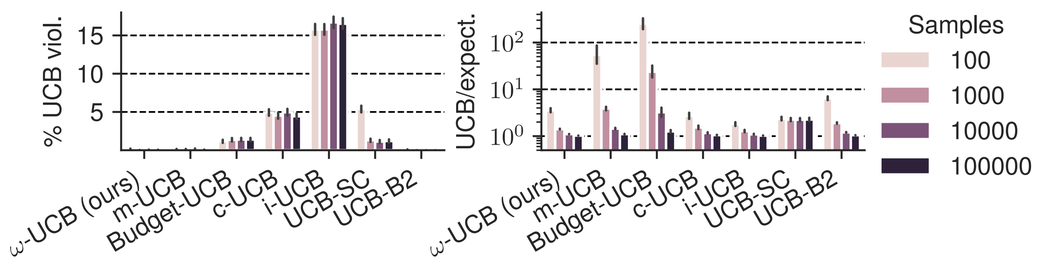
Applications

- Social media advertising (there are more applications in the paper)



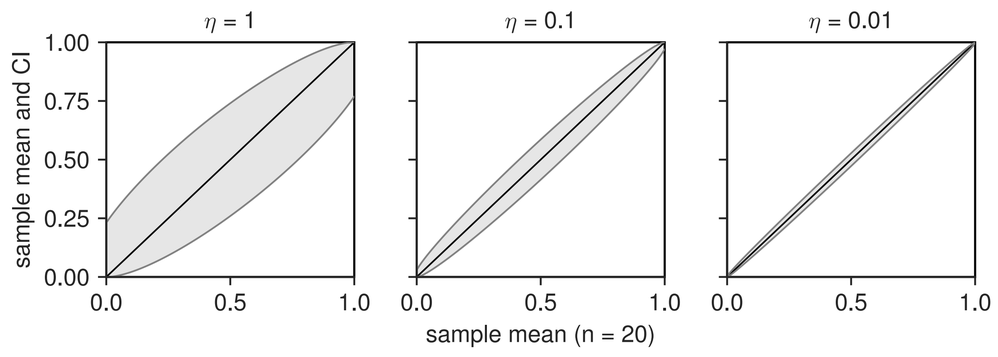
Related work

- Upper confidence bound (UCB) sampling: m-UCB, i-UCB, c-UCB [1], Budget-UCB [2], UCB-SC [3], UCB-B2 [4]
- UCB is often either **too tight** (left plot, higher is worse)
- or **too loose** (right plot, higher is worse)



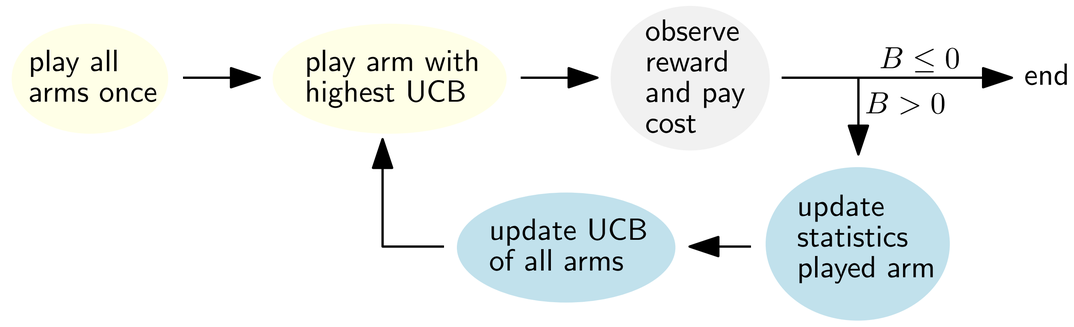
Asymmetric confidence intervals

- Generalization of the Wilson Score Interval for binomial proportions [5]
  - (refer to the paper for the math)
- Figure illustrates asymmetry of confidence interval
- $\eta \in [0, 1]$: variance relative to Bernoulli variable with same mean
- Confidence interval stays within bounds of random variable



Our Algorithm – ω-UCB

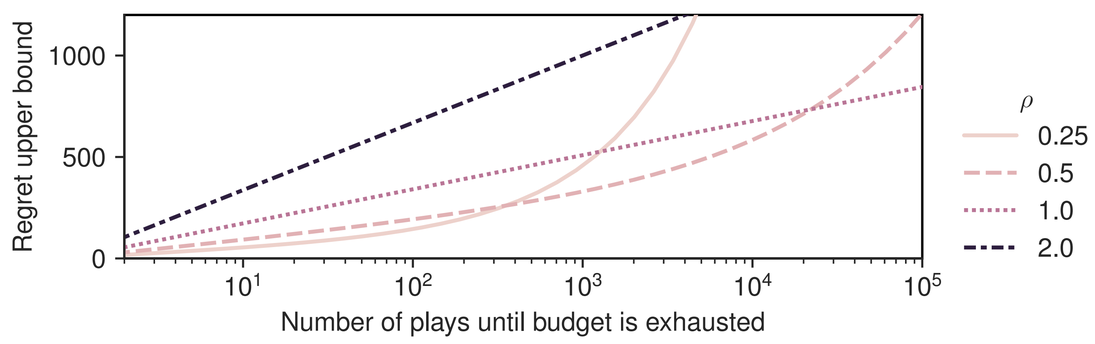
- Upper confidence bound (UCB) sampling
  - **Choose arm with highest UCB of reward-cost ratio**
  - "optimism under uncertainty"
- Compute UCB using **asymmetric confidence interval**
- **Increase confidence level over time** according to $\sqrt{1 - t^{-\rho}}$
  - ρ: scaling parameter of confidence interval



Regret analysis

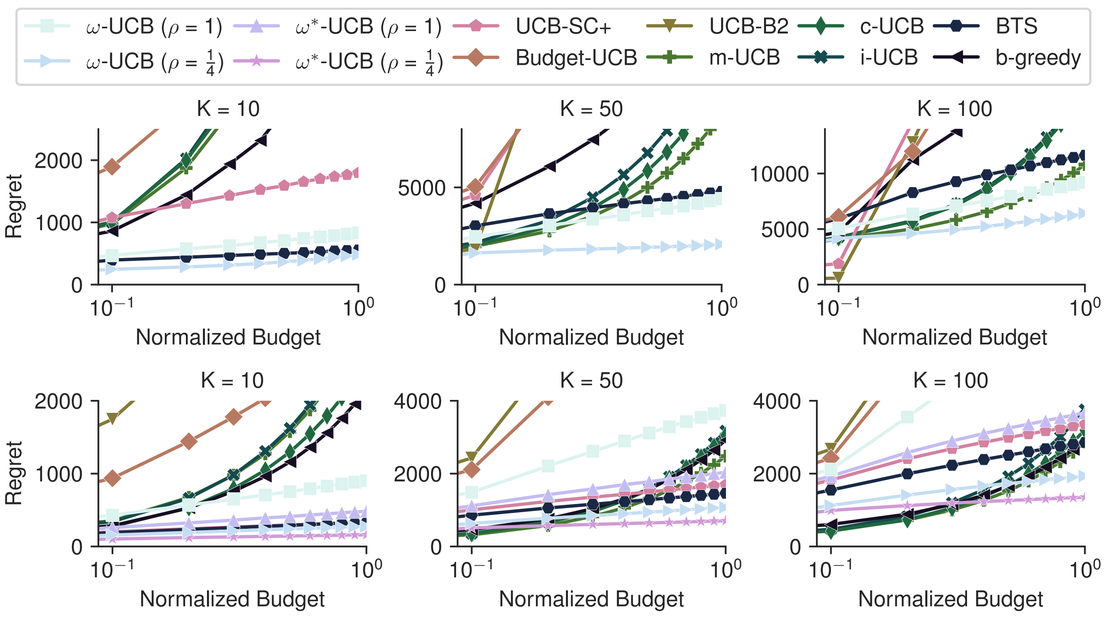
$$\text{Regret} \in \underbrace{\mathcal{O}\left(B^{1-\rho}\right) \text{ for } 0 < \rho < 1}_{\text{favors exploitation}}, \text{ and in } \underbrace{\mathcal{O}\left(\log B\right) \text{ for } \rho \geq 1}_{\text{favors exploration}}$$

- The figure below illustrates upper regret bound for a 2-armed bandit
- For (relatively) small budgets, choose small ρ to favor exploitation



Experiments on synthetic data

- First row: Bernoulli distributed rewards and costs
- Second row: rewards and costs sampled from $\{0, 0.25, 0.5, 0.75, 1\}$
- Approach ω^*-UCB approximates η-parameter



[1] Y. Xia, T. Qin, W. Ding, *et al.*, "Finite budget analysis of multi-armed bandit problems," *Neurocomputing*, vol. 258, pp. 13–29, 2017, ISSN: 0925-2312.
[2] Y. Xia, W. Ding, X.-D. Zhang, N. Yu, and T. Qin, "Budgeted Bandit Problems with Continuous Random Costs," in *ACML*, ser. JMLR Workshop and Conference Proceedings, vol. 45, JMLR.org, 2015, pp. 317–332.
[3] R. Watanabe, J. Komiyama, A. Nakamura, and M. Kudo, "KL-UCB-Based Policy for Budgeted Multi-Armed Bandits with Stochastic Action Costs," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. 100-A, no. 11, pp. 2470–2486, 2017.
[4] S. Cayci, A. Eryilmaz, and R. Srikant, "Budget-constrained bandits over general cost and reward distributions," in *AISTATS*, S. Chiappa and R. Calandra, Eds., ser. PMLR, vol. 108, PMLR, 2020, pp. 4388–4398.
[5] E. B. Wilson, "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209–212, 1927.