Open Problem: Max-of-Means

Wouter Koolen

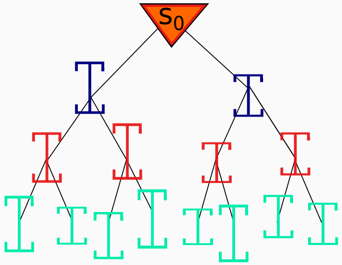
September 23rd, 2020

Motivation

Estimating the value of a state in

- Extensive form games (MCTS)
- MDPs (tabular) (Bellman backup)
- ...

from (noisy) observations



High-level

Simplified Model

Stochastic bandit $\mu_1, \ldots, \mu_K$.

Problem

We want to learn about $\mu^* \coloneqq \max_k \mu_k$.

High-level

Simplified Model

Stochastic bandit $\mu_1, \dots, \mu_K$.

Problem

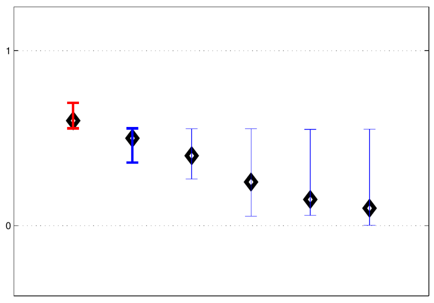
We want to learn about $\mu^* := \max_k \mu_k$.

- Hypothesis test of $\{\mu^* < \gamma\}$ vs $\{\mu^* > \gamma\}$
    - Fixed confidence vs fixed budget
- Make confidence interval $[LCB, UCB]$ for μ^*
    - Uniform sampling vs adaptive sampling
    - Fixed sample size vs any-time valid.

Asymptotic Theory

Asymptotically, **only** the data from arm $i^* \coloneqq \arg\max_k \mu_k$ matters:

$$\mu^* \in \hat{\mu}_{i^*,n} \pm \sqrt{2\sigma^2 \frac{\ln \frac{1}{\delta}}{n/K}}$$



Practice is not asymptotic

Can we get mileage out of data from other arms?

Practice is not asymptotic

Can we get mileage out of data from other arms?

Interpolate adaptively between width $\sim \sqrt{\frac{1}{n/K}}$ and $\sim \sqrt{\frac{1}{n}}$?

Practice is not asymptotic

Can we get mileage out of data from other arms?

Interpolate adaptively between width $\sim \sqrt{\frac{1}{n/K}}$ and $\sim \sqrt{\frac{1}{n}}$?

- Batch data from other arms (those close to maximum)
  - More samples!
  - Bias

  (i.e. can estimate $\frac{1}{K}\sum_i \mu_i$ from all n samples)

Practice is not asymptotic

Can we get mileage out of data from other arms?

Interpolate adaptively between width $\sim \sqrt{\frac{1}{n/K}}$ and $\sim \sqrt{\frac{1}{n}}$?

- Batch data from other arms (those close to maximum)
  - More samples!
  - Bias

  (i.e. can estimate $\frac{1}{K}\sum_i \mu_i$ from all n samples)
- Use multivariate confidence regions
  - Balls/Ellipsoids, KL eggs ($\approx \chi^2_K$)
  - Especially useful for LCB on μ^*

Practice is not asymptotic

Can we get mileage out of data from other arms?

Interpolate adaptively between width $\sim \sqrt{\frac{1}{n/K}}$ and $\sim \sqrt{\frac{1}{n}}$?

- Batch data from other arms (those close to maximum)
  - More samples!
  - Bias

  (i.e. can estimate $\frac{1}{K} \sum_i \mu_i$ from all n samples)
- Use multivariate confidence regions
  - Balls/Ellipsoids, KL eggs ($\approx \chi^2_K$)
  - Especially useful for LCB on μ^*

Incomparable results in practise (Kaufmann, Koolen, and Garivier, 2018).

None of these seem especially principled.

Open Problem

Non-asymptotic instance-dependent

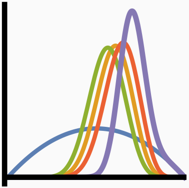
- practical confidence intervals for μ^*
- lower bounds

Open Problem

Non-asymptotic instance-dependent

- practical confidence intervals for μ^*
- lower bounds

Inspiration: Bayesian posterior for μ^* adapts automatically!

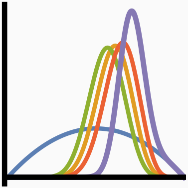


Open Problem

Non-asymptotic instance-dependent

- practical confidence intervals for μ^*
- lower bounds

Inspiration: Bayesian posterior for μ^* adapts automatically!



Applications *everywhere*!

Thanks!

References

- Garivier, A., E. Kaufmann, and W. M. Koolen (June 2016). “Maximin Action Identification: A New Bandit Framework for Games”. In: *Proceedings of the 29th Annual Conference on Learning Theory (COLT)*, pp. 1028–1050.
- Huang, R., M. M. Ajallooeian, C. Szepesvári, and M. Müller (2017). “Structured Best Arm Identification with Fixed Confidence”. In: *International Conference on Algorithmic Learning Theory, ALT 2017, 15-17 October 2017, Kyoto University, Kyoto, Japan*. Vol. 76. Proceedings of Machine Learning Research, pp. 593–616.
- Kaufmann, E. and W. M. Koolen (Oct. 2018). “Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals”. Preprint.

Kaufmann, E. and W. M. Koolen (Dec. 2017). “Monte-Carlo Tree Search by Best Arm Identification”. In: *Advances in Neural Information Processing Systems (NeurIPS) 30*, pp. 4904–4913.

Kaufmann, E., W. M. Koolen, and A. Garivier (Dec. 2018). “Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling”. In: *Advances in Neural Information Processing Systems (NeurIPS) 31*, pp. 6333–6343.