Bounded regret in stochastic multi-armed bandits

Sebastien Bubeck, Vianney Perchet, Philippe Rigollet

Bounded regret in stochastic bandits, Bubeck, Perchet and Rigollet

https://blogs.princeton.edu/imabandit/

Unknown parameters: $\nu_1, \ldots, \nu_K$ (subgaussian) probability distributions

Notation: $\mu_i = \mathbb{E}_{X \sim \nu_i} X, \; \mu^* = \max_{i \in [K]} \mu_i, \; \Delta_i = \mu^* - \mu_i$

Game: For $t = 1, \ldots, n$, select $I_t \in \{1, \ldots, K\}$ and receive $Y_t \sim \nu_{I_t}$.

Performance measure: $R_n = n\mu^* - \mathbb{E} \sum_{t=1}^{n} Y_t$

Theorem (Auer, Cesa-Bianchi and Fischer 2002)

$$R_n(\mathrm{UCB}) \leq c \sum_{i : \Delta_i > 0} \frac{\log n}{\Delta_i}$$

Theorem (Lai and Robbins 1985)

Consider a strategy such that if the distributions are Gaussian with variance 1 then $R_n = o(n^a) \; \forall a > 0$

Then for any Gaussian distributions with variance 1 one has

$$\liminf_{n \to +\infty} \frac{R_n}{\log n} \geq c \sum_{i : \Delta_i > 0} \frac{1}{\Delta_i}$$