Best Arm Identification in Multi-Armed Bandits

Jean-Yves Audibert
Imagine, Université Paris Est
&
Willow, CNRS/ENS/INRIA, Paris, France
audibert@imagine.enpc.fr

Sébastien Bubeck, Rémi Munos
SequeL Project, INRIA Lille
40 avenue Halley,
59650 Villeneuve d'Ascq, France
{sebastien.bubeck, remi.munos}@inria.fr

Abstract

This is the supplemental material of the COLT'10 paper untitled "Best Arm Identification in Multi-Armed Bandits".

A Lower bound for UCB-E

Theorem 1 *If $\nu_2, \dots, \nu_K$ are Dirac distributions concentrated at $\frac{1}{2}$ and if ν_1 is the Bernoulli distribution of parameter $3/4$, the UCB-E algorithm satisfies $4\mathbb{E}r_n = e_n \geq 4^{-(4a+1)}$.*

Proof: Consider the event $\mathcal{E}$ on which the reward obtained from the first $m = \lceil 4a \rceil$ draws of arm 1 are equal to zero. On this event of probability 4^{-m}, UCB-E will not draw arm 1 more than m times. Indeed, if it is drawn m times, it will not be drawn another time since $B_{1,m} \leq \frac{1}{2} < B_{2,s}$ for any s. On the event $\mathcal{E}$, we have $J_n \neq 1$. ■

B Application of Hoeffding's maximal inequality in the proof of Theorem 4

Let $i \in \{2, \dots, L\}$ and $j \in \{1, \dots, L\}$. First note that, by definition of ν' and since $i \neq 1$,

$$\mathbb{E}_{\nu'} \widehat{\mathrm{KL}}_{i,t}(\nu_i, \nu_j) = t\, \mathrm{KL}(\nu_i, \nu_j).$$

Since $\nu_i = Ber(\mu_i)$ and $\nu_j = Ber(\mu_j)$, with $\mu_i, \mu_j \in [p, 1-p]$, we have

$$\left| \log \left(\frac{d\nu_i(X_{i,t})}{d\nu_j(X_{i,t})} \right) \right| \leq \log(p^{-1}).$$

From Hoeffding's maximal inequality, see e.g. (Cesa-Bianchi and Lugosi, 2006, Section A.1.3), we have to bound almost surely the quantity, with $\mathbb{P}_{\nu'}$-probability at least $1 - \frac{1}{2L^2}$, we have for all $t \in \{1, \dots, n\}$,

$$\widehat{\mathrm{KL}}_{i,t}(\nu_i, \nu_j) - t\, \mathrm{KL}(\nu_i, \nu_j) \leq 2 \log(p^{-1}) \sqrt{\frac{\log(L^2) n}{2}}.$$

Similarly, with $\mathbb{P}_{\nu'}$-probability at least $1 - \frac{1}{2L^2}$, we have for all $t \in \{1, \dots, n\}$,

$$\widehat{\mathrm{KL}}_{1,t}(\nu_L, \nu_j) - t\, \mathrm{KL}(\nu_L, \nu_j) \leq 2 \log(p^{-1}) \sqrt{\frac{\log(L^2) n}{2}}.$$

A simple union bound argument then gives $\mathbb{P}_{\nu'}(C_n) \geq 1/2$.

References

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.