Multi-armed bandit problems: a statistical view, focused on lower bounds

Gilles Stoltz

Laboratoire de mathématiques d'Orsay

Joint works with Hédi Hadiji, now at CentraleSupélec,
and with ENS Lyon colleagues: Aurélien Garivier, Pierre Ménard, Antoine Barrier

K–armed stochastic bandits

Framework, possible objectives, index strategies

K probability distributions $\nu_1, \ldots, \nu_K$
with expectations $\mu_1, \ldots, \mu_K$ $\longrightarrow$ $\mu^\star = \max_{a \in [K]} \mu_a$

At each round $t = 1, 2, \ldots,$

1. Statistician picks arm $A_t \in [K]$
2. She gets a reward Y_t drawn according to ν_{A_t}
3. This is the only feedback she receives

$\longrightarrow$ Exploration–exploitation dilemma
estimate the ν_a vs. get high rewards Y_t

Link with UQ? Emmanuel Vazquez told me:

Conceptually	arms $\leftrightarrow$ parameters of numerical experiments
Technically	leverage bandit techniques to study EI strategy

Setting: at round each round $t \geqslant 1$, pick arm $A_t \in [K]$, get and observe $Y_t \sim \nu_{A_t}$

Objective #1: Maximize cumulative rewards $\leftrightarrow$
Minimize pseudo-regret

$$R_T = \sum_{t=1}^{T} (\mu^\star - \mathbb{E}[Y_t]) = \sum_{t=1}^{T} (\mu^\star - \mathbb{E}[\mu_{A_t}])$$

$$= \sum_{a \in [K]} \left((\mu^\star - \mu_a) \, \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}_{\{A_t = a\}} \right] \right) = \sum_{a \in [K]} (\mu^\star - \mu_a) \, \mathbb{E}\big[N_a(T) \big]$$

$\leftrightarrow$ Control the $\mathbb{E}\big[N_a(T)\big]$

Objective #2: Identify best arm $\leftrightarrow$ Minimize $\mathbb{P}\Big(I_T \notin \arg\max_{a \in [K]} \mu_a \Big)$

Model: $\nu_1, \ldots, \nu_K$ are distributions over $[0, 1]$

A classical strategy: UCB [upper confidence bound]

Auer, Cesa-Bianchi and Fisher [2002]

For $t \geqslant K$, pick
$$A_{t+1} \in \operatorname*{arg\,max}_{a \in [K]} \left\{ \widehat{\mu}_a(t) + \sqrt{\frac{2 \ln t}{N_a(t)}} \right\}$$

Exploitation: cf. empirical mean $\widehat{\mu}_a(t)$

Exploration: cf. $\sqrt{2 \ln t / N_a(t)}$ favors arms a not pulled often

Suboptimal regret bounds of two types

– Distribution-dependent bound:
$$R_T \lesssim \sum_{a : \mu_a < \mu^\star} \frac{8 \ln T}{\mu^\star - \mu_a}$$

– Distribution-free bound:
$$\sup_{\nu_1, \ldots, \nu_K} R_T \lesssim \sqrt{8 K T \ln T}$$

Model: $\nu_1, \ldots, \nu_K$ are distributions over $[0,1]$

Another index-based strategy:

MOSS [minimax optimal strategy in the stochastic setting]

Audibert and Bubeck [2009]

For $t \geqslant K$, pick $A_{t+1} \in \operatorname*{arg\,max}_{a \in [K]} \left\{ \widehat{\mu}_a(t) + \sqrt{\dfrac{1}{N_a(t)} \ln_+ \dfrac{T}{K N_a(t)}} \right\}$

$\ln_+ = \max\{\ln, 0\}$; there exist anytime versions

Distribution-free regret bounds $\sup_{\nu_1,\ldots,\nu_K} R_T$ of optimal order $\sqrt{KT}$

– Upper bound: $49\sqrt{KT}$ for MOSS

– Lower bound: $(1/20)\sqrt{KT}$

Auer, Cesa-Bianchi, Freund and Schapire [2002]

Model: $\nu_1, \ldots, \nu_K$ are distributions over $[0,1]$

KL-UCB strategy

Honda and Takemura [2015]; Cappé, Garivier, Maillard, Munos, Stoltz [2013]; Garivier, Hadiji, Ménard, Stoltz [2022]

Key quantity $\qquad \mathcal{K}_{\inf}(\nu_a, \mu^\star) = \inf\big\{ \mathrm{KL}(\nu_a, \nu'_a) : E(\nu'_a) > \mu^\star \big\}$

Indices $\quad U_a(t) = \sup\left\{ \mu \in [0,1] : \quad \mathcal{K}_{\inf}\big(\hat{\nu}_a(t), \mu\big) \leqslant \dfrac{\varphi\big(t, N_a(t)\big)}{N_a(t)} \right\}$

Typically, $\varphi\big(t, N_a(t)\big)$ of order $\ln t$; $\quad$ for $t \geqslant K$, pick $\quad A_{t+1} \in \operatorname*{arg\,max}_{a \in [K]} U_a(t)$

Optimal distribution-dependent regret bounds:

$$\sum_{a : \mu_a < \mu^\star} \frac{\mu^\star - \mu_a}{\mathcal{K}_{\inf}(\nu_a, \mu^\star)} \ln T - \Theta(\ln \ln T)$$

For lower bounds: Lai and Robbins [1985]; Burnetas and Katehakis [1996]; Garivier, Ménard and Stoltz [2019]

Model: $\nu_1, \ldots, \nu_K$ are distributions over $[0, 1]$

KL-UCB-Switch strategy

Garivier, Hadiji, Ménard, Stoltz [2022]

Index strategy of the form: for each arm $a \in [K]$, use

- KL-UCB index if $N_a(t) \leqslant (t/K)^5$
- MOSS index if $N_a(t) \geqslant (t/K)^5$

Optimal bounds of the two types:

- Distribution-dependent bound, with the $-\Theta(\ln \ln T)$ term
- Distribution-free bound: $\displaystyle \sup_{\nu_1,\ldots,\nu_K} R_T \lesssim K + 23\sqrt{KT}$

Model: $\nu_1, \ldots, \nu_K$ are distributions over $[0, 1]$

Summary / Reviewed index strategies all of the form:

For $t \geqslant K$, pick $\quad A_{t+1} \in \operatorname*{arg\,max}_{a \in [K]} \Big\{ \widehat{\mu}_a(t) + expl\big(t, N_a(t)\big) \Big\}$

Possibly with fancy, or null, exploration bonuses $expl\big(t, N_a(t)\big)$

Exploitation: cf. empirical mean $\widehat{\mu}_a(t)$

Various bounds achieved, depending on how $expl\big(t, N_a(t)\big)$ is set

– Optimal distribution-dependent bounds:

$$\sum_{a:\mu_a<\mu^\star} \frac{\mu^\star - \mu_a}{\mathcal{K}_{\inf}\big(\nu_a, \mu^\star\big)} \ln T - \Theta(\ln \ln T)$$

– Optimal distribution-free bounds: $\quad \sup_{\nu_1,\ldots,\nu_K} R_T = \Theta\big(\sqrt{KT}\big)$

Proofs for upper bounds: control $\mathbb{E}\big[N_a(T)\big]$

Proofs of the regret lower bounds on $[0, 1]$

(At least, high-level ideas...)

Proof ideas for the lower bounds

Strategy ψ: maps $H_t = (Y_1, \ldots, Y_t) \mapsto A_{t+1} = \psi_t(H_t)$

Change of measure: compare distributions of H_T
under $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ vs. $\underline{\nu}' = (\nu_1', \ldots, \nu_K')$

Fundamental inequality: performs an implicit change of measure

Reference: Lai and Robbins [1985], Auer et al. [2002], Garivier et al. [2019]

For all Z taking values in $[0,1]$ and $\sigma(H_T)$–measurable

(chain rule) $$\sum_{a \in [K]} \mathbb{E}_{\underline{\nu}}\big[N_a(T)\big] \, \mathrm{KL}(\nu_a, \nu_a') = \mathrm{KL}\Big(\mathbb{P}_{\underline{\nu}}^{H_T}, \mathbb{P}_{\underline{\nu}'}^{H_T}\Big)$$

(data-proc. ineq.) $$\geqslant \mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}[Z], \, \mathbb{E}_{\underline{\nu}'}[Z]\Big)$$

where $\mathrm{kl}(p,q) = \mathrm{KL}\big(\mathrm{Ber}(p), \mathrm{Ber}(q)\big)$

Later use: $\underline{\nu}'$ only differs from $\underline{\nu}$ at some a, with $Z = N_a(T)/T$

Distribution-free lower bound, for distributions over $[0,1]$

Problem $\underline{\nu}_0 = \big(\mathrm{Ber}(1/2)\big)_{a\in[K]}$ vs. $\underline{\nu}_k = \Big(\mathrm{Ber}\big(1/2 + \varepsilon\,\mathbb{I}_{\{a=k\}}\big)\Big)_{a\in[K]}$

$$R_T \stackrel{\text{def}}{=} \sum_{a\neq k} \varepsilon\,\mathbb{E}_{\underline{\nu}_k}\big[N_a(T)\big] = T\varepsilon\Big(1 - \mathbb{E}_{\underline{\nu}_k}\big[N_k(T)/T\big]\Big)$$

Thus, $\sup_{\underline{\nu}} R_T \geqslant \sup_{\varepsilon\in(0,1)} \max_{k\in[K]} T\varepsilon\Big(1 - \mathbb{E}_{\underline{\nu}_k}\big[N_k(T)/T\big]\Big)$

Fundamental inequality, with $Z = N_k(T)/T$

+ Pinsker's inequality and $k \in [K]$ such that $\mathbb{E}_{\underline{\nu}_0}\big[N_k(T)/T\big] \leqslant 1/K$

$$\overbrace{\mathbb{E}_{\underline{\nu}_0}\big[N_k(T)\big]}^{\leqslant T/K}\ \overbrace{\mathrm{KL}\big(\mathrm{Ber}(1/2),\ \mathrm{Ber}(1/2+\varepsilon)\big)}^{=-\ln(1-4\varepsilon^2)/2\ \leqslant\ 2.5\varepsilon^2}$$

$$\geqslant \mathrm{kl}\big(\mathbb{E}_{\underline{\nu}_0}[Z],\ \mathbb{E}_{\underline{\nu}_k}[Z]\big) \geqslant 2\Big(\mathbb{E}_{\underline{\nu}_k}\big[N_0(T)/T\big] - \mathbb{E}_{\underline{\nu}_k}\big[N_k(T)/T\big]\Big)^2$$

Thus, $\sup_{\underline{\nu}} R_T \geqslant \sup_{\varepsilon\in(0,1/4)} T\varepsilon\big(1 - 1/K - \varepsilon\sqrt{1.25\,T/K}\big) \geqslant \Theta\big(\sqrt{KT}\big)$

Distribution-dependent bound: $$R_T = \sum_{a \in [K]} (\mu^\star - \mu_a)\, \mathbb{E}_{\underline{\nu}}\big[N_a(T)\big]$$

We lower bound each $\mathbb{E}_{\underline{\nu}}\big[N_a(T)\big]$ for a fixed a with $\mu_a < \mu^\star$; let ν'_a with $\mu_a > \mu^\star$

Problems $\underline{\nu} = (\nu_a)_{a \in [K]}$ vs. $\underline{\nu}' = \big(\nu_1, \ldots, \nu_{a-1}, \nu'_a, \nu_{a+1}, \ldots, \nu_K\big)$

Fundamental inequality
on "good" strategies $\forall \alpha \in (0,1],\quad \mathbb{E}[N_k(T)] = o(T^\alpha)$ for subopt. k
& lower bound on kl $\mathrm{kl}(p,q) \geqslant (1-p)\ln\big(1/(1-q)\big) - \ln 2$

$$\mathbb{E}_{\underline{\nu}}\big[N_a(T)\big]\,\mathrm{KL}(\nu_a,\nu'_a) \geqslant \mathrm{kl}\Big(\overbrace{\mathbb{E}_{\underline{\nu}}\big[N_a(T)/T\big]}^{=o(1)},\, \mathbb{E}_{\underline{\nu}'}\big[N_a(T)/T\big]\Big)$$
$$\gtrsim \ln\Big(1/\big(1-\mathbb{E}_{\underline{\nu}'}\big[N_a(T)/T\big]\big)\Big)$$

Since $\mathbb{E}_{\underline{\nu}'}\big[N_a(T)/T\big] = 1 - \sum_{k \neq a} \mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big] \gtrsim 1 - T^{\alpha-1}$, we get:

$$\mathbb{E}_{\underline{\nu}}\big[N_a(T)\big]\,\mathrm{KL}(\nu_a,\nu'_a) \gtrsim \ln T^{1-\alpha}$$

Distribution-dependent bound: $\displaystyle R_T = \sum_{a \in [K]} \bigl(\mu^\star - \mu_a\bigr)\, \mathbb{E}_{\underline{\nu}}\bigl[N_a(T)\bigr]$

We lower bound each $\mathbb{E}_{\underline{\nu}}\bigl[N_a(T)\bigr]$ for a fixed a with $\mu_a < \mu^\star$; let ν'_a with $\mu_a > \mu^\star$

$$\mathbb{E}_{\underline{\nu}}\bigl[N_a(T)\bigr]\, \mathrm{KL}(\nu_a, \nu'_a) \gtrsim \ln T^{1-\alpha}, \qquad \text{that is,} \qquad \frac{\mathbb{E}_{\underline{\nu}}\bigl[N_a(T)\bigr]\, \mathrm{KL}(\nu_a, \nu'_a)}{\ln T} \gtrsim 1 - \alpha \to 1$$

Therefore, “good” strategies can ensure, at best:

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}\bigl[N_a(T)\bigr]}{\ln T} \geqslant \sup_{\nu'_a : \mu'_a > \mu^\star} \frac{1}{\mathrm{KL}(\nu_a, \nu'_a)} \stackrel{\text{def}}{=} \frac{1}{\mathcal{K}_{\inf}(\nu_a, \mu^\star)}$$

By summing over suboptimal arms:

$$\liminf_{T\to\infty} \frac{R_T}{\ln T} \geqslant \sum_{a \in [K]} \frac{\mu^\star - \mu_a}{\mathcal{K}_{\inf}(\nu_a, \mu^\star)}$$

Note: general proof, valid for any model $\mathcal{D}$

Adaptation to the range

Bounded but unknown range

Reference for this part of the talk: Hadiji and Stoltz [2020]

That is, model: $$\mathcal{D} = \bigcup_{m,M:m<M} \mathcal{D}_{m,M}$$

where $\mathcal{D}_{m,M}$ set of distributions ν over a given interval $[m, M]$
Before, we were only dealing with $\mathcal{D}_{0,1}$

What changes?

Same distribution-free lower bound:

$\Theta\big((M-m)\sqrt{KT}\big)$ by rescaling

No worsening due to ignorance of the range

Different distribution-dependent lower bound:

$R_T / \ln T \to +\infty$ as $\mathcal{K}_{\inf}(\nu_a, \mu^\star, \mathcal{D}) = 0$

But any rate $\gg \ln T$ may be achieved

Focus on the UCB strategy

With a known range $[m, M]$, reads (knowledge of the range is key!)

$$A_{t+1} \in \mathop{\arg\max}_{a \in [K]} \left\{ \widehat{\mu}_a(t) + (M-m)\sqrt{\frac{2 \ln t}{N_a(t)}} \right\}$$

Extension to an unknown range:

$$A_{t+1} \in \mathop{\arg\max}_{a \in [K]} \left\{ \widehat{\mu}_a(t) + \sqrt{\frac{\varphi(t)}{N_a(t)}} \right\}$$

where $\ln t \ll \varphi(t) \ll t$; eventually, $\sqrt{\varphi(t)} \geqslant (M-m)\sqrt{2 \ln t}$

Guarantee: for all bandit problems $\nu_1, \ldots, \nu_K$ in $\mathcal{D}$,

$$\limsup \frac{R_T}{\varphi(T)} < +\infty$$

$\Phi_{\text{dep}} = \varphi$ is the corresponding distribution-dependent rate for adaptation to the range

Distribution-free rate for adaptation to the range

$\Phi_{\mathrm{free}} : \mathbb{N} \to (0, +\infty)$ such that

$\forall m < M,$
$\forall \nu_1, \ldots, \nu_K$ in $\mathcal{D}_{m,M},$
$\forall T \geqslant 1,$ $\qquad R_T \leqslant (M - m)\Phi_{\mathrm{free}}(T)$

By the lower bound proved for $[m, M] = [0, 1]$:

$$\Phi_{\mathrm{free}}(T) \geqslant \Theta\big(\sqrt{KT}\big)$$

AdaHedge on estimated payoffs + mixing achieves

$$\Phi_{\mathrm{free}}(T) \approx 7(M - m)\sqrt{TK \ln K}$$

Reference for AdaHedge: Cesa-Bianchi, Mansour, Stoltz [2005, 2007] and De Rooij, van Erven, Grünwald, Koolen [2014]

Note: $\sqrt{\ln K}$ shaved off (with different strategy) when M is known

AdaHedge on estimated payoffs + mixing

Randomized strategy: $A_t \sim \mathbf{p}_t$ for $t \geqslant K+1$

Unbiased estimated payoffs: $\widehat{X}_{t,a} = \dfrac{Y_t - C}{p_{t,a}} + C$

where C is the average of the payoffs in the first K rounds

AdaHedge: $q_{t+1,a} \propto \exp\left(-\eta_t \sum_{s=K+1}^{t} \widehat{X}_{t,a}\right)$

Mixing: $\mathbf{p}_t = (1-\gamma_t)\mathbf{q}_t + \gamma_t \, \mathbf{1}/K$

Strategy actually built for adversarial payoffs

(= arbitrary sequences)

What about simultaneous bounds?

Reminder for known range $[0,1]$: ln T and $\sqrt{T}$ rates for regret upper bounds

Theorem: If $\Phi_{\text{free}}(T) \ll T$ then $\Phi_{\text{dep}}(T) \times \Phi_{\text{free}}(T) \geqslant \Theta(T)$

Example: $\Phi_{\text{free}}(T) = \Theta(\sqrt{T})$ now forces $\Phi_{\text{dep}}(T) \geqslant \Theta(\sqrt{T})$

$\rightarrow$ We finally exhibit some heavy price for adaptation!

Proof: by the fundamental inequality
$+$ lack of upper end on payoffs in $\mathcal{D}$

AdaHedge on estimated payoffs + mixing simultaneously achieves

$$\Phi_{\text{free}}(T) = \Theta\big(\sqrt{T}\big) \quad \text{and} \quad \Phi_{\text{dep}}(T) = \Theta\big(\sqrt{T}\big)$$

Analysis heavily based on Seldin and Lugosi [2017]

Actually, all pairs $\Phi_{\text{free}}(T) = \Theta(T^{\alpha})$ and $\Phi_{\text{dep}}(T) = \Theta(T^{1-\alpha})$
with $\alpha \in [1/2,\ 1)$ may be achieved, by setting the mixing factor properly

FYI—Proof of "If $\Phi_{\mathrm{free}}(T) \ll T$ then $\Phi_{\mathrm{dep}}(T) \times \Phi_{\mathrm{free}}(T) \geqslant \Theta(T)$"

Based on fundamental inequality + lack of upper end on payoffs in $\mathcal{D}$

We lower bound each $\mathbb{E}_{\underline{\nu}}\big[N_a(T)\big]$ for a fixed a with $\mu_a < \mu^\star$

Problems $\underline{\nu}$, $\underline{\nu}'$ only differing at $\nu'_a = (1-\varepsilon)\nu_a + \varepsilon\,\delta_{\mu_a + c/\varepsilon}$

such that $\nu_a \perp \delta_{\mu_a + c/\varepsilon}$ and $\mu'_a > \mu^\star$

$f = \dfrac{\mathrm{d}\nu_a}{\mathrm{d}\nu'_a} = \dfrac{1}{1-\varepsilon}$ so that $\mathrm{KL}(\nu_a, \nu'_a) = \mathbb{E}_{\nu_a}[\ln f] \approx \varepsilon$

Fundamental inequality and $\mathrm{kl}(p,q) \gtrsim (1-p)\ln\big(1/(1-q)\big)$

$$\mathbb{E}_{\underline{\nu}}\big[N_a(T)\big]\,\overbrace{\mathrm{KL}(\nu_a,\nu'_a)}^{\approx\varepsilon} \geqslant \mathrm{kl}\Big(\overbrace{\mathbb{E}_{\underline{\nu}}\big[N_a(T)/T\big]}^{=o(1)}, \mathbb{E}_{\underline{\nu}'}\big[N_a(T)/T\big]\Big)$$
$$\gtrsim \ln\Big(1/\big(1-\mathbb{E}_{\underline{\nu}'}\big[N_a(T)/T\big]\big)\Big)$$

Indeed: $(\mu^\star - \mu_a)\,\mathbb{E}_{\underline{\nu}}\big[N_a(T)\big] \leqslant R_T(\underline{\nu}) \leqslant (M-m)\,\Phi_{\mathrm{free}}(T) \ll T$

Similarly: $\ln\Big(1/\big(1-\mathbb{E}_{\underline{\nu}'}\big[N_a(T)/T\big]\big)\Big) \gtrsim \ln\big(c'\,\Phi_{\mathrm{free}}(T)\,/(T\varepsilon)\big)$

As: $(\mu'_a - \mu^\star)\,\Big(T - \mathbb{E}_{\underline{\nu}'}\big[N_a(T)\big]\Big) \leqslant R_T(\underline{\nu}') \leqslant (M + c/\varepsilon - m)\,\Phi_{\mathrm{free}}(T)$

Picking $\varepsilon \sim \Phi_{\mathrm{free}}(T)/T$: $\big(\Phi_{\mathrm{free}}(T)/T\big)\,\mathbb{E}_{\underline{\nu}}\big[N_a(T)\big] \gtrsim \mathrm{cst}$

Best-arm identification

With fixed budget and for possibly non-parametric models

Objective #2: BAI with fixed budget T

Reference for this final part of the talk: Barrier, Garivier, and Stoltz [2022]

Bandit problem $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ with unique optimal arm $a^\star(\underline{\nu})$

where optimality is in expectation: $\mu_{a^\star} = \max_{a \in [K]} \mu_a$

T rounds, where arms A_t are pulled, rewards Y_t are obtained

Then: issue a recommendation $I_T \in [K]$

Goal: upper and lower bound $\mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big)$

Note—BAI with fixed confidence δ well understood

Track and Stop strategies, by Aurélien Garivier, Emilie Kaufmann, and co-authors

Typical strategy: Successive rejects

By Audibert and Bubeck [2010], with analysis based on Hoeffding's inequality

$K-1$ regimes, and in each regime $r=1,\ldots,K-1$:

- Denote by S_r the set of arms not dropped so far; $S_1=[K]$
- Play each arm in S_r an equal number of times
- Drop arm with smallest average payoff since the beginning
  (not smallest average payoff in regime r)

By carefully setting regimes (based on T and K): when $\mathcal{D} \subseteq \mathcal{P}_{0,1}$,

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \leqslant -\frac{1}{\overline{\ln} K} \min_{2\leqslant k\leqslant K} \frac{\big(\mu_{a^\star}-\mu_{(k)}\big)^2}{k}$$

where $\mu_{a^\star}=\mu_{(1)}>\mu_{(2)}\geqslant \ldots \geqslant \mu_{(K)}$

$\rightarrow$ Called a gap-based bound

Lower bounds?

Gap-based approach by Audibert and Bubeck [2010]

Studied $\mathcal{D}_p = \big\{ \mathrm{Ber}(x) : \; x \in [p, 1-p] \big\}$ for $p > 0$

Methodology actually extends to models $\mathcal{D}$ such that

$$\forall \nu, \nu' \text{ in } \mathcal{D}, \qquad \mathrm{KL}(\nu, \nu') \leqslant C_{\mathcal{D}} \big(\mathrm{E}(\nu) - \mathrm{E}(\nu') \big)^2$$

For instance, $C_{\mathcal{D}_p} = 1/\big(2p(1-p)\big)$

Careful and explicit analysis leading to: for all strategies,

$$\liminf_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu}) \big) \geqslant -5\, C_{\mathcal{D}} \min_{2 \leqslant k \leqslant K} \frac{\big(\mu_{a^\star} - \mu_{(k)}\big)^2}{k}$$

Difference: $-5\, C_{\mathcal{D}}$ vs. $-1/\overline{\ln} K$ in front of the min

New non-parametric approach

Key quantities: note the reverse order in the KL compared to $\mathcal{K}_{\inf}$

$$\mathcal{L}_{\inf}^{<}(x,\nu) = \inf\bigl\{\mathrm{KL}(\zeta,\nu) : \zeta \in \mathcal{D} \text{ s.t. } \mathrm{E}(\zeta) < x\bigr\}$$

and

$$\mathcal{L}_{\inf}^{>}(x,\nu) = \inf\bigl\{\mathrm{KL}(\zeta,\nu) : \zeta \in \mathcal{D} \text{ s.t. } \mathrm{E}(\zeta) > x\bigr\}$$

Analysis of Successive rejects based on Cramér-Chernoff bounds:

$$\limsup_{T\to+\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\bigl(I_T \neq a^\star(\underline{\nu})\bigr) \leqslant -\frac{1}{\overline{\ln} K} \min_{2\leqslant k\leqslant K} \frac{\mathcal{L}\bigl(\nu_{(k)},\nu_{a^\star}\bigr)}{k}$$

where $\mathcal{L}\bigl(\nu_{(k)},\nu_{a^\star}\bigr) = \inf_{x\in[\mu_{(k)},\mu_{a^\star}]} \Bigl\{\mathcal{L}_{\inf}^{\geqslant}\bigl(x,\nu_{(k)}\bigr) + \mathcal{L}_{\inf}^{\leqslant}\bigl(x,\nu_{a^\star}\bigr)\Bigr\}$

By Pinsker's inequality: yields the gap-based upper bound for $\mathcal{D} = \mathcal{P}_{0,1}$

Special case $x = \mu_{(k)}$ for the lower bounds

New non-parametric approach: simple lower bound

Alternative problem $\underline{\nu}'$ differing from $\underline{\nu}$ only at $a^\star(\underline{\nu})$,
with distribution ζ s.t. $\mathrm{E}(\zeta) < \mu_{(K)}$

$$q_T \stackrel{\text{def}}{=} \mathbb{P}_{\underline{\nu}'}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant \mathbb{P}_{\underline{\nu}'}\big(I_T = a^\star(\underline{\nu}')\big) \xrightarrow[T \to +\infty]{} 1\,,$$

$$\text{while} \qquad p_T \stackrel{\text{def}}{=} \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \xrightarrow[T \to +\infty]{} 0$$

Fundamental inequality:

$$-\frac{1}{T} \ln p_T \sim \frac{\mathrm{KL}\big(\mathrm{Ber}(p_T), \mathrm{Ber}(q_T)\big)}{T} \leqslant \frac{\overbrace{\mathbb{E}_{\underline{\nu}'}\big[N_{a^\star(\underline{\nu})}\big]}^{\leqslant 1/K}}{T} \, \mathrm{KL}\big(\zeta, \nu_{a^\star}\big)$$

Hence,

$$\liminf_{T \to +\infty} \frac{1}{T} \ln \mathbb{P}_{\underline{\nu}}\big(I_T \neq a^\star(\underline{\nu})\big) \geqslant -\frac{1}{K} \overbrace{\inf_{\mathrm{E}(\zeta) < \mu_{(K)}} \mathrm{KL}\big(\zeta, \nu_{a^\star}\big)}^{\mathcal{L}_{\inf}^{<}(\mu_{(K)}, \nu_{a^\star})}$$

Pruning argument to get the min over $k \in \{2, \ldots, K\}$