Lower bounds on the regret for stochastic bandits

A general inequality to generate them

Gilles Stoltz

CNRS — HEC Paris

Joint work with Aurélien Garivier and Pierre Ménard (UPS Toulouse)

K–armed bandits: framework

K probability distributions $\nu_1, \ldots, \nu_K$
with expectations $\mu_1, \ldots, \mu_K$ $\longrightarrow$ $\mu^\star = \max_{k=1,\ldots,K} \mu_k$

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

1. Statistician picks arm $I_t \in \{1, \ldots, K\}$, possibly using U_{t-1}
2. She gets a reward Y_t with law ν_{I_t} given I_t
3. This is the only feedback she receives

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

Regret:

$$R_T = \sum_{t=1}^{T} \bigl(\mu^\star - \mathbb{E}[Y_t]\bigr) = \sum_{k=1}^{K} \left((\mu^\star - \mu_k)\, \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}_{\{I_t = k\}} \right] \right)$$

Indeed, $Y_t \mid I_t \sim \nu_{I_t}$, thus $\mathbb{E}\bigl[Y_t \mid I_t\bigr] = \mu_{I_t}$

Summary:

At each round, pick I_t (based on U_{t-1} + past) and get $Y_t \mid I_t \sim \nu_{I_t}$

Control the regret $R_T = \sum_{k=1}^{K} (\mu^\star - \mu_k)\, \mathbb{E}\big[N_k(T)\big]$, where $N_k(T) = \sum_{t=1}^{T} \mathbb{I}_{\{I_t = k\}}$

Lower bound $R_T \iff$ Lower bound $\mathbb{E}\big[N_k(T)\big]$ for $\mu_k < \mu^\star$

Randomized strategy $\psi = (\psi_t)_{t \geqslant 0}$: measurable functions

$$\psi_t : H_t = \big(U_0, Y_1, U_1, \ldots, Y_t, U_t\big) \longmapsto \psi_t(H_t) = I_{t+1}$$

Take $U_0, U_1, \ldots$ iid $\sim \mathcal{U}_{[0,1]}$ and denote by $\mathfrak{m}$ the Lebesgue measure

Transition kernel (conditional distributions):

$$\mathbb{P}\big(Y_{t+1} \in B,\ U_{t+1} \in B' \,\big|\, H_t\big) = \nu_{\psi_t(H_t)}(B)\, \mathfrak{m}(B')$$

The fundamental inequality

$$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\,\mathrm{KL}(\nu_k,\nu'_k) \geqslant \mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big],\,\mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big]\Big)$$

Summary: history $H_t = \bigl(U_0, Y_1, U_1, \ldots, Y_t, U_t\bigr)$ and $I_{t+1} = \psi_t(H_t)$

Lower bound $\mathbb{E}\bigl[N_k(T)\bigr]$ for $\mu_k < \mu^\star$, where $N_k(T) = \sum \mathbb{I}_{\{I_t = k\}}$

Transition kernel: $\mathbb{P}\bigl(Y_{t+1} \in B,\ U_{t+1} \in B' \,\big|\, H_t\bigr) = \nu_{\psi_t(H_t)}(B)\,\mathfrak{m}(B')$

Change of measure: $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ vs. $\underline{\nu}' = (\nu'_1, \ldots, \nu'_K)$

Fundamental inequality: performs an implicit change of measure

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

$$\sum_{k=1}^{K} \mathbb{E}_{\underline{\nu}}\bigl[N_k(T)\bigr]\,\mathrm{KL}(\nu_k, \nu'_k) = \mathrm{KL}\bigl(\mathbb{P}^{H_T}_{\underline{\nu}}, \mathbb{P}^{H_T}_{\underline{\nu}'}\bigr)$$

$$\geqslant \mathrm{kl}\bigl(\mathbb{E}_{\underline{\nu}}[Z], \mathbb{E}_{\underline{\nu}'}[Z]\bigr)$$

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

Later use: $\underline{\nu}'$ only differ from $\underline{\nu}$ at k and $Z = N_k(T)/T$

Proof of the equality: chain rule for KL

$H_{t+1} = \big(H_t, \, (Y_{t+1}, U_{t+1})\big)$ and $\mathbb{P}\big(Y_{t+1} \in B, \; U_{t+1} \in B' \,\big|\, H_t\big) = \nu_{\psi_t(H_t)}(B)\,\mathfrak{m}(B')$

$$
\begin{aligned}
&\mathrm{KL}\Big(\mathbb{P}_{\underline{\nu}}^{H_{t+1}}, \, \mathbb{P}_{\underline{\nu}'}^{H_{t+1}}\Big) \\
&= \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_t}, \, \mathbb{P}_{\underline{\nu}'}^{H_t}\big) + \mathrm{KL}\Big(\mathbb{P}_{\underline{\nu}}^{(Y_{t+1},U_{t+1}) \,|\, H_t}, \, \mathbb{P}_{\underline{\nu}'}^{(Y_{t+1},U_{t+1}) \,|\, H_t}\Big) \\
&= \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_t}, \, \mathbb{P}_{\underline{\nu}'}^{H_t}\big) + \mathbb{E}_{\underline{\nu}}\Big[\mathbb{E}_{\underline{\nu}}\Big[\mathrm{KL}\big(\nu_{\psi_t(H_t)} \otimes \mathfrak{m}, \, \nu'_{\psi_t(H_t)} \otimes \mathfrak{m}\big) \,\Big|\, H_t\Big]\Big] \\
&= \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_t}, \, \mathbb{P}_{\underline{\nu}'}^{H_t}\big) + \mathbb{E}_{\underline{\nu}}\Big[\mathbb{E}_{\underline{\nu}}\Big[\mathrm{KL}\big(\nu_{\psi_t(H_t)}, \, \nu'_{\psi_t(H_t)}\big) \,\Big|\, H_t\Big]\Big] \\
&= \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_t}, \, \mathbb{P}_{\underline{\nu}'}^{H_t}\big) + \mathbb{E}_{\underline{\nu}}\left[\sum_{k=1}^{K} \mathrm{KL}(\nu_k, \nu'_k)\, \mathbb{I}_{\{I_{t+1}=k\}}\right]
\end{aligned}
$$

By induction: $\mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_T}, \, \mathbb{P}_{\underline{\nu}'}^{H_T}\big) = \sum_{k=1}^{K} \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\, \mathrm{KL}(\nu_k, \nu'_k)$

References: already present in Auer, Cesa-Bianchi, Freund and Schapire [2002]

Proof of the inequality $\quad \mathrm{KL}\big(\mathbb{P}_{\underline{\nu}}^{H_T}, \mathbb{P}_{\underline{\nu}'}^{H_T}\big) \geqslant \mathrm{kl}\big(\mathbb{E}_{\underline{\nu}}[Z], \mathbb{E}_{\underline{\nu}'}[Z]\big)$

where $\mathrm{kl}(p,q) = \mathrm{KL}\big(\mathrm{Ber}(p), \mathrm{Ber}(q)\big)$ and $Z \in [0,1]$ is $\sigma(H_T)$–measurable

Lemma (Data-processing inequality)

For all random variables $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$,

$$\mathrm{KL}\big(\mathbb{P}^X, \mathbb{Q}^X\big) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$$

Lemma (Data-processing inequality with expectations)

For all random variables $X : (\Omega, \mathcal{F}) \to \big([0,1], \mathcal{B}\big)$,

$$\mathrm{KL}\Big(\mathrm{Ber}\big(\mathbb{E}_{\mathbb{P}}[X]\big), \mathrm{Ber}\big(\mathbb{E}_{\mathbb{Q}}[X]\big)\Big) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$$

Proof of $\mathrm{KL}\big(\mathbb{P}^X, \mathbb{Q}^X\big) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$ — part 1/2

Proof: We may assume that $\mathbb{P} \ll \mathbb{Q}$, otherwise $\mathrm{KL}(\mathbb{P},\mathbb{Q}) = +\infty$ and the inequality is true. We shows that we then have

$$\mathbb{P}^X \ll \mathbb{Q}^X, \quad \text{with} \quad \frac{d\mathbb{P}^X}{d\mathbb{Q}^X} = \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X = \cdot \right] \overset{\text{not.}}{=} \gamma$$

ie, $\gamma(X) = \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X\right]$.

Indeed, for all $B \in \mathcal{F}'$:

$$\mathbb{P}^X(B) = \mathbb{P}\{X \in B\} = \int_\Omega \mathbb{1}_B(X)\, \frac{d\mathbb{P}}{d\mathbb{Q}}\, d\mathbb{Q} \overset{\text{tower rule}}{=} \int_\Omega \mathbb{1}_B(X)\, \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X\right] d\mathbb{Q}$$

$$\overset{\text{not.}}{=} \int_\Omega \mathbb{1}_B(X)\, \gamma(X)\, d\mathbb{Q} \underset{\text{by definition of } \mathbb{Q}^X}{=} \int_{\Omega'} \mathbb{1}_B\, \gamma\, d\mathbb{Q}^X.$$

Proof of $\mathrm{KL}\big(\mathbb{P}^X, \mathbb{Q}^X\big) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$ — part 2/2

Therefore,

$$\mathrm{KL}(\mathbb{P}^X, \mathbb{Q}^X) = \int_{\Omega'} \gamma \ln \gamma \, \mathrm{d}\mathbb{Q}^X = \int_{\Omega} \gamma(X) \ln \gamma(X) \, \mathrm{d}\mathbb{Q}$$

definition of γ

$$= \int_{\Omega} \left(\mathbb{E}_{\mathbb{Q}}\left[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \,\Big|\, X \right] \ln \mathbb{E}_{\mathbb{Q}}\left[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \,\Big|\, X \right] \right) \mathrm{d}\mathbb{Q}$$

conditional version of Jensen's inequality

$$\leqslant \int_{\Omega} \mathbb{E}_{\mathbb{Q}}\left[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \ln \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \,\Big|\, X \right] \mathrm{d}\mathbb{Q}$$

tower rule

$$= \int_{\Omega} \left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \ln \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \right) \mathrm{d}\mathbb{Q} = \mathrm{KL}(\mathbb{P}, \mathbb{Q})$$

Reference: Ali and Silvey [1966]; implies joint convexity of KL

Proof of $\mathrm{KL}\Big(\mathrm{Ber}\big(\mathbb{E}_{\mathbb{P}}[X]\big),\, \mathrm{Ber}\big(\mathbb{E}_{\mathbb{Q}}[X]\big)\Big) \leqslant \mathrm{KL}(\mathbb{P},\mathbb{Q})$

Proof: We denote by $\mathfrak{m}$ the Lebesgue measure over $[0,1]$ and augment the underlying measurable space into $(\Omega\times[0,1],\ \mathcal{F}\otimes\mathcal{B}([0,1]))$, over which we consider the product-distributions $\mathbb{P}\otimes\mathfrak{m}$ and $\mathbb{Q}\otimes\mathfrak{m}$.

$$\mathrm{KL}\Big(\underbrace{(\mathbb{P}\otimes\mathfrak{m})^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{P}\otimes\mathfrak{m}(E))},\ \underbrace{(\mathbb{Q}\otimes\mathfrak{m})^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{Q}\otimes\mathfrak{m}(E))}\Big) \leqslant \mathrm{KL}(\mathbb{P}\otimes\mathfrak{m},\mathbb{Q}\otimes\mathfrak{m}) \underset{\text{cf. product distributions}}{=} \mathrm{KL}(\mathbb{P},\mathbb{Q}) + \mathrm{KL}(\mathfrak{m},\mathfrak{m}) = \mathrm{KL}(\mathbb{P},\mathbb{Q})$$

Thus: $\mathrm{KL}\big(\mathrm{Ber}(\mathbb{P}\otimes\mathfrak{m}(E)),\ \mathrm{Ber}(\mathbb{Q}\otimes\mathfrak{m}(E))\big) \leqslant \mathrm{KL}(\mathbb{P},\mathbb{Q})$

The proof is concluded by picking $E\in\mathcal{F}\otimes\mathcal{B}([0,1])$ such that

$$\mathbb{P}\otimes\mathfrak{m}(E) = \mathbb{E}_{\mathbb{P}}[X] \quad\text{and}\quad \mathbb{Q}\otimes\mathfrak{m}(E) = \mathbb{E}_{\mathbb{Q}}[X]$$

Namely, $E = \big\{(\omega,x)\in\Omega\times[0,1] : \ x\leqslant X(\omega)\big\}$.

By Tonelli's theorem:

$$\mathbb{P}\otimes\mathfrak{m}(E) = \int_\Omega \Big(\int_{[0,1]} \mathbb{1}_{\{x\leqslant X(\omega)\}}\, d\mathfrak{m}(x)\Big)\, d\mathbb{P}(\omega) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \mathbb{E}_{\mathbb{P}}[X]$$

and a similar equality for $\mathbb{Q}\otimes\mathfrak{m}(E)$.

Fundamental inequality: For all $Z \in [0,1]$ and $\sigma(H_T)$–measurable,

$$\sum_{k=1}^{K} \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\,\mathrm{KL}(\nu_k, \nu'_k) \geqslant \mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}[Z],\, \mathbb{E}_{\underline{\nu}'}[Z]\Big)$$

How to use it?

Bandit problem $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ where k is suboptimal: $\mu_k < \mu^\star$

Pick $Z = N_k(T)/T$

Pick $\underline{\nu}'$ that only differs from $\underline{\nu}$ at k:

$$\underline{\nu}' = (\nu_1, \ldots, \nu_{k-1}, \nu'_k, \nu_{k+1}, \ldots, \nu_K)$$

Then $\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\,\mathrm{KL}(\nu_k, \nu'_k) \geqslant \mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big],\, \mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big]\Big)$

Distribution-dependent lower bound for large T

$$\liminf_{T \to \infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}$$

To lower bound $R_T = \sum_{k=1}^{K} (\mu^\star - \mu_k) \, \mathbb{E}\big[N_k(T)\big]$, lower bound $\mathbb{E}\big[N_k(T)\big]$ for $\mu_k < \mu^\star$

Bandit model $\mathcal{D}$: where the $\nu_1, \ldots, \nu_K$ may lie in

Assumption (UFC – uniform fast convergence on $\mathcal{D}$)

The strategy ψ is such that:
For all bandit problems $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ in $\mathcal{D}$, for all $\mu_k < \mu^\star$,

$$\forall \, \alpha \in (0,1], \qquad \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] = o(T^\alpha)$$

Bandit problem $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ where k is suboptimal: $\mu_k < \mu^\star$

Pick $\nu'_k \in \mathcal{D}$ with expectation $\mu'_k > \mu^\star$

Form $\underline{\nu}'$ that only differs from $\underline{\nu}$ at k:

$$\underline{\nu}' = (\nu_1, \ldots, \nu_{k-1}, \nu'_k, \nu_{k+1}, \ldots, \nu_K)$$

Then $\quad \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] = o(T) \quad$ and $\quad T - \mathbb{E}_{\underline{\nu}'}\big[N_k(T)\big] = o(T^\alpha)$

$\mathbb{E}_{\underline{\nu}}[N_k(T)] = o(T)$ and $T - \mathbb{E}_{\underline{\nu}'}[N_k(T)] = o(T^\alpha)$ for a strategy ψ UFC on $\mathcal{D}$

Also, $\mathrm{kl}(p,q) \geqslant (1-p)\ln \dfrac{1}{1-q} - \ln 2$

Fundamental inequality + lower bound on kl:

$$\begin{aligned}
&\mathbb{E}_{\underline{\nu}}[N_k(T)] \\
&\geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu_k')}\, \mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big],\, \mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big]\Big) \\
&\geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu_k')}\left(-\ln 2 + \Big(1-\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big]\Big)\ln \frac{1}{1-\mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big]}\right) \\
&\geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu_k')}\left(-\ln 2 + \big(1-o(1)\big)\ln \frac{1}{T^{\alpha-1}}\right)
\end{aligned}$$

Thus, $\forall\,\alpha \in (0,1]$, $$\liminf_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}[N_k(T)]}{\ln T} \geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu_k')}\,\frac{\ln T^{1-\alpha}}{\ln T}$$

That is, for all models $\mathcal{D}$ (for the first time, no assumption on $\mathcal{D}$)

for all strategies ψ UFC on $\mathcal{D}$ (this is not a real restriction)

for all bandit problems $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ in $\mathcal{D}$

for $\mu_k < \mu^\star$

Lemma

for all ν'_k in $\mathcal{D}$ with $\mu'_k > \mu^\star$,

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathrm{KL}(\nu_k, \nu'_k)}$$

Theorem (see Lai and Robbins [1985], Burnetas and Katehakis [1996])

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}$$

where $\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D}) = \inf\big\{\mathrm{KL}(\nu_k, \nu'_k) : \ \nu'_k \in \mathcal{D} \text{ with } \mu'_k > \mu^\star\big\}$

This distribution-dependent bound is asymptotically optimal:

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}$$

I.e., at least for well-behaved models $\mathcal{D}$, we can exhibit a matching upper bound:

$$\limsup_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \leqslant \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}$$

See Lai and Robbins [1985], Burnetas and Katehakis [1996], Honda and Takemura [2010–2015], Cappé, Garivier, Maillard, Munos and Stoltz [2013], etc.

Replacing the $o(T^\alpha)$ in the definition of UFC by a $O(\ln T)$:

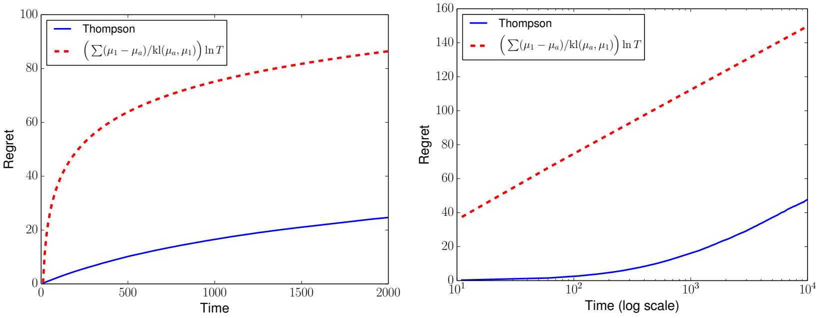
$$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] \geqslant \frac{\ln T}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})} - O\big(\ln(\ln T)\big)$$

Cf. the upper bound of Honda and Takemura [2015]:
This second-order term $-\ln(\ln T)$ is optimal

Distribution-dependent lower bound for small T

We expect them to be linear!

The asymptotic bound is really of an asymptotic nature!



The regret of Thompson Sampling vs. the asymptotic bound

Theorem

For all models $\mathcal{D}$
*for all strategies ψ smarter * than the uniform strategy on $\mathcal{D}$*
for all bandit problems $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ in $\mathcal{D}$
for all arms k, for all $T \geqslant 1$,

$$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] \geqslant \frac{T}{K}\left(1 - \sqrt{2T\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}\right).$$

In particular, for $T \leqslant 1/\big(8\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})\big)$, $\qquad \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] \geqslant \dfrac{T}{2K}$

* A strategy ψ is smarter than the uniform strategy on a model $\mathcal{D}$ if for all bandit problems $\underline{\nu}$ in $\mathcal{D}$, for all optimal arms $a^\star$,

$$\forall\, T \geqslant 1, \qquad \mathbb{E}_{\underline{\nu}}\big[N_{a^\star}(T)\big] \geqslant \frac{T}{K}.$$

Mild requirement; but some requirement needed to get such a universal statement

All previous linear lower bounds were for some (well-chosen) bandit problems in $\mathcal{D}$

Same $\underline{\nu}'$ as before: just replace ν_k by ν'_k with $\mu'_k > \mu^\star$

Thus $\mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big] \geqslant 1/K$ and [wnlog] $\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big] \leqslant 1/K$

Using a local Pinsker's inequality *

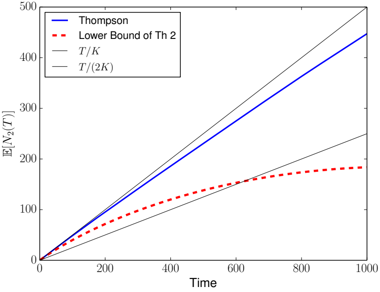
$$\begin{aligned}
\frac{T}{K}\,\mathrm{KL}(\nu_k,\nu'_k) &\geqslant \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\,\mathrm{KL}(\nu_k,\nu'_k) \\
&\geqslant \mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big],\,\mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big]\Big) \\
&\geqslant \mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big],\,1/K\Big) \\
&\geqslant (K/2)\,\Big(\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big] - 1/K\Big)^2
\end{aligned}$$

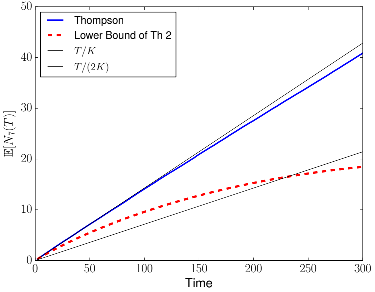
Hence the bound (to be optimized over all relevant ν'_k)

$$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] \geqslant \frac{T}{K}\Big(1 - \sqrt{2T\,\mathrm{KL}(\nu_k,\nu'_k)}\Big)$$

* For $0 \leqslant p < q \leqslant 1$, we have $\quad \mathrm{kl}(p,q) \geqslant \dfrac{1}{2\max\limits_{x\in[p,q]} x(1-x)}(p-q)^2 \geqslant \dfrac{1}{2q}(p-q)^2$

Illustration of our bound





Expected number of times a suboptimal arm is pulled: Thompson Sampling vs. our linear lower bound (look rather at the $T/(2K)$ and T/K lines)

Conclusion: many other bounds!

Many other bandits bounds can be [re-]obtained in a few elementary lines from

$$\sum_{k=1}^{K} \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\, \mathrm{KL}(\nu_k, \nu'_k) \geqslant \mathrm{kl}\big(\mathbb{E}_{\underline{\nu}}[Z],\, \mathbb{E}_{\underline{\nu}'}[Z]\big)$$

For instance,

The $\sqrt{KT}$ distribution-free bound by Auer, Cesa-Bianchi, Freund and Schapire [2002]

The bounds by Bubeck, Perchet and Rigollet [2013] when $\mu^\star$ and/or the gaps $\mu^\star - \mu_k$ are known

And many other new bounds

(our fundamental inequality is already a popular tool!)