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 \( \mu_1, \ldots, \mu_K \)

\[ \mu^* = \max_{k=1,\ldots,K} \mu_k \]

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 \( \nu_{I_t} \) given \( I_t \)
3. This is the only feedback she receives

\[ \text{Exploration–exploitation dilemma} \]

estimate the \( \nu_k \) vs. get high rewards \( Y_t \)

Regret:

\[ R_T = \sum_{t=1}^{T} (\mu^* - \mathbb{E}[Y_t]) = \sum_{k=1}^{K} \left( \mu^* - \mu_k \right) \mathbb{E}\left[ \sum_{t=1}^{T} \mathbb{I}\{I_t = k\} \right] \]

Indeed, \( Y_t \mid I_t \sim \nu_{I_t} \), thus \( \mathbb{E}[Y_t \mid I_t] = \mu_{I_t} \)
Summary:
At each round, pick $I_t$ (based on $U_{t-1}$ + past) and get $Y_t | I_t \sim \nu_{I_t}$

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

Lower bound $R_T \iff$ Lower bound $E[N_k(T)]$ for $\mu_k < \mu^*$

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

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

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

Transition kernel (conditional distributions):

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

$$\mathbb{E}_\nu[N_k(T)] \KL(\nu_k, \nu'_k) \geq \kl\left(\mathbb{E}_\nu[N_k(T)/T], \mathbb{E}_{\nu'}[N_k(T)/T]\right)$$
Summary: history $H_t = (U_0, Y_1, U_1, \ldots, Y_t, U_t)$ and $I_{t+1} = \psi_t(H_t)$

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

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

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

**Fundamental inequality:** performs an implicit change of measure

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

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

$$\geq \text{kl}(\mathbb{E}_\nu[Z], \mathbb{E}_{\nu'}[Z])$$

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

**Later use:** $\nu'$ only differ from $\nu$ at $k$ and $Z = N_k(T)/T$
Proof of the equality: chain rule for KL

\[ H_{t+1} = (H_t, (Y_{t+1}, U_{t+1})) \text{ and } P(Y_{t+1} \in B, U_{t+1} \in B' \mid H_t) = \nu_{\psi_t(H_t)}(B) m(B') \]

\[
\text{KL}\left( \mathbb{P}^{H_{t+1}}_{\nu}, \mathbb{P}^{H_{t+1}}_{\nu'} \right) \\
= \text{KL}\left( \mathbb{P}^{H_t}_{\nu}, \mathbb{P}^{H_t}_{\nu'} \right) + \text{KL}\left( \mathbb{P}(Y_{t+1}, U_{t+1}) \mid H_t, \mathbb{P}(Y_{t+1}, U_{t+1}) \mid H_t \right) \\
= \text{KL}\left( \mathbb{P}^{H_t}_{\nu}, \mathbb{P}^{H_t}_{\nu'} \right) + \mathbb{E}_\nu \left[ \mathbb{E}_\nu \left[ \text{KL}\left( \nu_{\psi_t(H_t)} \otimes m, \nu'_{\psi_t(H_t)} \otimes m \right) \mid H_t \right] \right] \\
= \text{KL}\left( \mathbb{P}^{H_t}_{\nu}, \mathbb{P}^{H_t}_{\nu'} \right) + \mathbb{E}_\nu \left[ \mathbb{E}_\nu \left[ \text{KL}\left( \nu_{\psi_t(H_t)}, \nu'_{\psi_t(H_t)} \right) \mid H_t \right] \right] \\
= \text{KL}\left( \mathbb{P}^{H_t}_{\nu}, \mathbb{P}^{H_t}_{\nu'} \right) + \mathbb{E}_\nu \left[ \sum_{k=1}^{K} \text{KL}(\nu_k, \nu'_{k}) I\{l_{t+1}=k\} \right] \\
\]

By induction: \[ \text{KL}\left( \mathbb{P}^{H_T}_{\nu}, \mathbb{P}^{H_T}_{\nu'} \right) = \sum_{k=1}^{K} \mathbb{E}_\nu \left[ N_k(T) \right] \text{KL}(\nu_k, \nu'_{k}) \]

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

\[ \text{KL}(\mathbb{P}_{\mathcal{H}^T}, \mathbb{P}_{\mathcal{H}^T}') \geq \text{kl}(\mathbb{E}_\mathcal{H}[Z], \mathbb{E}_{\mathcal{H}'}[Z]) \]

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

**Lemma (Data-processing inequality)**

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

\[ \text{KL}(\mathbb{P}^X, \mathbb{Q}^X) \leq \text{KL}(\mathbb{P}, \mathbb{Q}) \]

**Lemma (Data-processing inequality with expectations)**

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

\[ \text{KL}\left(\text{Ber}(\mathbb{E}_\mathbb{P}[X]), \text{Ber}(\mathbb{E}_\mathbb{Q}[X])\right) \leq \text{KL}(\mathbb{P}, \mathbb{Q}) \]
Proof of $\text{KL}(\mathbb{P}^X, \mathbb{Q}^X) \leq \text{KL}(\mathbb{P}, \mathbb{Q})$ — part 1/2

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

$$\mathbb{P}^X \ll \mathbb{Q}^X,$$

with

$$\frac{d\mathbb{P}^X}{d\mathbb{Q}^X} = \mathbb{E}_{\mathbb{Q}} \left[ \frac{d\mathbb{P}}{d\mathbb{Q}} | X = \cdot \right] = \gamma,$$

i.e.,

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

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

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

by definition of $\mathbb{Q}^X$. 

\[\text{by definition of } \mathbb{Q}^X\]
Proof of $\text{KL}(\mathbb{P}^X, \mathbb{Q}^X) \leq \text{KL}(\mathbb{P}, \mathbb{Q})$ — part 2/2

Therefore,

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

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

\[ \leq \int_{\Omega} \text{E}_\mathbb{Q} \left[ \frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} | X \right] \ d\mathbb{Q} \]

(conditional version of Jensen's inequality)

Lower bound

\[ = \int_{\Omega} \left( \frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} \right) \ d\mathbb{Q} = \text{KL}(\mathbb{P}, \mathbb{Q}) \]

Reference: Ali and Silvey [1966]; implies joint convexity of KL
Proof of \( \text{KL}(\text{Ber}(E_{\mathbb{P}}[X]), \text{Ber}(E_{\mathbb{Q}}[X])) \leq \text{KL}(\mathbb{P}, \mathbb{Q}) \)

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

\[
\text{KL}( (\mathbb{P} \times m)_{\mathbb{E}} \circ \mathbb{E}_{m}(\mathbb{E}), (\mathbb{Q} \times m)_{\mathbb{E}} \circ \mathbb{E}_{m}(\mathbb{E})) \leq \text{KL}( (\mathbb{P} \times m, \mathbb{Q} \times m) \circ \mathbb{E}_{m}(\mathbb{E}))
\]

Thus:

\[
\text{KL}( \text{Ber}(\mathbb{P} \times m(E)), \text{Ber}(\mathbb{Q} \times m(E))) \leq \text{KL}(\mathbb{P}, \mathbb{Q})
\]

The proof is concluded by picking \( E \in \mathcal{F} \otimes \mathcal{B}(\mathbb{R}[0,1]) \) such that \( \mathbb{P} \times m(E) = E_{\mathbb{P}}[x] \) and \( \mathbb{Q} \times m(E) = E_{\mathbb{Q}}[x] \).

Namely, \( E = \{(w, x) \in \Omega \times [0,1] : x \leq x(w)\} \).

By Tonelli’s theorem:

\[
(\mathbb{P} \times m)(E) = \int_{\Omega} \left( \int_{[0,1]} \mathbb{1}_{\{x \leq x(w)\}} \text{d}m(x) \right) \text{d}P(w)
= \int_{\Omega} X(w) \text{d}P(w) = E_{\mathbb{P}}[x]
\]

and a similar equality for \( \mathbb{Q} \times m(E) \).
Fundamental inequality: For all $Z \in [0, 1]$ and $\sigma(H_T)$–measurable,

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

How to use it?

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

Pick $Z = N_k(T)/T$

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

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

Then

$$
\mathbb{E}_\nu[N_k(T)] \ KL(\nu_k, \nu'_k) \geq \text{kl}(\mathbb{E}_\nu[N_k(T)/T], \mathbb{E}_{\nu'}[N_k(T)/T])
$$
Distribution-dependent lower bound for large $T$

\[ \liminf_{T \to \infty} \frac{\mathbb{E}_\nu \left[ N_k(T) \right]}{\ln T} \geq \frac{1}{\mathcal{K}_{\text{inf}}(\nu_k, \mu^*, \mathcal{D})} \]
To lower bound $R_T = \sum_{k=1}^{K} (\mu^* - \mu_k) \mathbb{E}[N_k(T)]$, lower bound $\mathbb{E}[N_k(T)]$ for $\mu_k < \mu^*$

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

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

*The strategy $\psi$ is such that:*

*For all bandit problems $\nu = (\nu_1, \ldots, \nu_K)$ in $\mathcal{D}$, for all $\mu_k < \mu^*$,*

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

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

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

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

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

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

Also, \( kl(p, q) \geq (1 - p) \ln \frac{1}{1 - q} - \ln 2 \)

**Fundamental inequality + lower bound on kl:**

\[
\begin{align*}
\mathbb{E}_\nu[N_k(T)] & \geq \frac{1}{\text{KL}(\nu_k, \nu'_{k})} \text{kl}\left( \mathbb{E}_\nu[N_k(T)/T], \mathbb{E}_{\nu'}[N_k(T)/T] \right) \\
& \geq \frac{1}{\text{KL}(\nu_k, \nu'_{k})} \left( -\ln 2 + \left( 1 - \mathbb{E}_\nu[N_k(T)/T] \right) \ln \frac{1}{1 - \mathbb{E}_{\nu'}[N_k(T)/T]} \right) \\
& \geq \frac{1}{\text{KL}(\nu_k, \nu'_{k})} \left( -\ln 2 + (1 - o(1)) \ln \frac{1}{T^{\alpha-1}} \right)
\end{align*}
\]

Thus, \( \forall \alpha \in (0, 1], \quad \liminf_{T \to \infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\ln T} \geq \frac{1}{\text{KL}(\nu_k, \nu'_{k})} \frac{\ln T^{1-\alpha}}{\ln T} \)
That is, for all models $\mathcal{D}$ (for the first time, no assumption on $\mathcal{D}$)
for all strategies $\psi$ UFC on $\mathcal{D}$ (this is not a real restriction)
for all bandit problems $\nu = (\nu_1, \ldots, \nu_K)$ in $\mathcal{D}$
for $\mu_k < \mu^*$

**Lemma**

for all $\nu_k'$ in $\mathcal{D}$ with $\mu_k' > \mu^*$,

$$
\liminf_{T \to \infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\ln T} \geq \frac{1}{\text{KL}(\nu_k, \nu_k')} \tag{1}
$$

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

$$
\liminf_{T \to \infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\ln T} \geq \frac{1}{\mathcal{K}_\inf(\nu_k, \mu^*, \mathcal{D})} \tag{2}
$$

where $\mathcal{K}_\inf(\nu_k, \mu^*, \mathcal{D}) = \inf\{\text{KL}(\nu_k, \nu_k') : \nu_k' \in \mathcal{D} \text{ with } \mu_k' > \mu^*\}$
This distribution-dependent bound is asymptotically optimal:

$$\lim \inf_{T \to \infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\ln T} \geq \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^*, D)}$$

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

$$\lim \sup_{T \to \infty} \frac{\mathbb{E}_\nu[N_k(T)]}{\ln T} \leq \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^*, D)}$$

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 $O(\ln T)$:

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

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 $\psi$ 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 $T \geq 1$,

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

In particular, for $T \leq 1/(8K_{\inf}(\nu_k, \mu^*, \mathcal{D}))$,

$$\mathbb{E}_{\underline{\nu}}[N_k(T)] \geq \frac{T}{2K}.$$

* A strategy $\psi$ 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 $a^*$,

$$\forall T \geq 1, \quad \mathbb{E}_{\underline{\nu}}[N_{a^*}(T)] \geq \frac{T}{K}.$$  

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 $\nu'$ as before: just replace $\nu_k$ by $\nu'_k$ with $\mu'_k > \mu^*$

Thus $\mathbb{E}_{\nu'}[N_k(T)/T] \geq 1/K$ and $[\text{wnlog}] \mathbb{E}_\nu[N_k(T)/T] \leq 1/K$

Using a local Pinsker's inequality *

$$
\frac{T}{K} \KL(\nu_k, \nu'_k) \geq \mathbb{E}_\nu[N_k(T)] \KL(\nu_k, \nu'_k)
\geq \kl\left(\mathbb{E}_\nu[N_k(T)/T], \mathbb{E}_{\nu'}[N_k(T)/T]\right)
\geq \kl\left(\mathbb{E}_\nu[N_k(T)/T], 1/K\right)
\geq \left(\frac{K}{2}\right)\left(\mathbb{E}_\nu[N_k(T)/T] - 1/K\right)^2
$$

Hence the bound (to be optimized over all relevant $\nu'_k$)

$$
\mathbb{E}_\nu[N_k(T)] \geq \frac{T}{K} \left(1 - \sqrt{2T \KL(\nu_k, \nu'_k)}\right)
$$

* For $0 \leq p < q \leq 1$, we have $\kl(p, q) \geq \frac{1}{2 \max_{x \in [p, q]} x(1-x) (p-q)^2} \geq \frac{1}{2q} (p-q)^2$
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

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

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 $\mu^*$ and/or the gaps $\mu^* - \mu_k$ are known

And many other new bounds
(our fundamental inequality is already a popular tool!)