Stochastic bandit lower bound

Pierre Ménard

May 16, 2016

Environment and strategy

- K arms bandit problem, $\nu = (\mathcal{B}(\mu_1), .., \mathcal{B}(\mu_K))$ with $\mu_i \in (0, 1)$.

Game, for each round $t \geqslant 1$:

1. Player pulls arm $A_t \in \{1, .., K\}$.
2. He gets a reward $Y_t \sim \mathcal{B}(\mu_{A_t})$.

- Information available at time t :

$$I_t = (Y_1, \dots, Y_t) .$$

Regret

- Optimal arm and gap :

$$\mu^\star = \max_{a=1,\dots,K} \mu_a \qquad \text{and} \qquad \Delta_a = \mu^\star - \mu_a \,.$$

- Number of time arm a is pulled :

$$N_a(T) = \sum_{t=1}^{T} \mathbb{I}_{\{A_t = a\}} \,.$$

- Goal of the player, minimize the expected regret :

$$R_{\nu,T} = T\mu^\star - \mathbb{E}_\nu\left[\sum_{t=1}^{T} Y_t\right] = \sum_{a=1}^{K} \Delta_a \, \mathbb{E}_\nu[N_a(T)] \,.$$

(tower rule)

Tow blocks inequality :

$$\boxed{\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(T)]\,\mathrm{kl}(\mu_a, \mu'_a) = \mathrm{KL}(\mathbb{P}_\nu^{I_T}, \mathbb{P}_{\nu'}^{I_T}) \geqslant \mathrm{kl}(\mathbb{E}_\nu[Z], \mathbb{E}_{\nu'}[Z])\,,} \qquad \text{(F)}$$

where

- $\mathbb{P}_\nu^{I_T}$ and $\mathbb{P}_{\nu'}^{I_T}$ respective distributions of I_T under $\mathbb{P}_\nu$ and $\mathbb{P}_{\nu'}$
- kl the Kullback-Leibler divergence for Bernoulli distributions :

$$\forall p, q \in [0,1]^2, \qquad \mathrm{kl}(p,q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}\,,$$

- Z a $\sigma(I_T)$–measurable random variable with values in $[0,1]$.

Tow blocks inequality :

$$\boxed{\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(T)]\,\mathrm{kl}(\mu_a, \mu'_a) = \mathrm{KL}\left(\mathbb{P}_\nu^{I_T}, \mathbb{P}_{\nu'}^{I_T}\right) \geqslant \mathrm{kl}\left(\mathbb{E}_\nu[Z], \mathbb{E}_{\nu'}[Z]\right),} \tag{F}$$

where

- $\mathbb{P}_\nu^{I_T}$ and $\mathbb{P}_{\nu'}^{I_T}$ respective distributions of I_T under $\mathbb{P}_\nu$ and $\mathbb{P}_{\nu'}$
- kl the Kullback-Leibler divergence for Bernoulli distributions :

$$\forall p, q \in [0,1]^2, \qquad \mathrm{kl}(p,q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q},$$

- Z a $\sigma(I_T)$–measurable random variable with values in $[0,1]$.

Typically $Z = N_a(T)/T$.

Proof.

- Equality in F, an application of chain rule for Kullback-Leibler divergences :

$$\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(T)] \,\mathrm{kl}(\mu_a, \mu'_a) = \mathrm{KL}(\mathbb{P}_\nu^{I_T}, \mathbb{P}_{\nu'}^{I_T})$$

□

Proof.

- Equality in F, an application of chain rule for Kullback-Leibler divergences :

$$\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(T)]\,\mathrm{kl}(\mu_a, \mu'_a) = \mathrm{KL}(\mathbb{P}_\nu^{I_T}, \mathbb{P}_{\nu'}^{I_T})$$

- Inequality in F, use contraction of entropy : Let $V \sim \mathcal{U}[0,1]$ independent of I_T, and the event $E = \{Z \geqslant V\}$ then

$$\mathrm{KL}(\mathbb{P}_\nu^{I_T}, \mathbb{P}_{\nu'}^{I_T}) = \mathrm{KL}(\mathbb{P}_\nu^{I_T} \otimes \mathcal{U}, \mathbb{P}_{\nu'}^{I_T} \otimes \mathcal{U}) \geqslant \mathrm{KL}\Big((\mathbb{P}_\nu^{I_T} \otimes \mathcal{U})^{\mathbb{I}_E}, (\mathbb{P}_{\nu'}^{I_T} \otimes \mathcal{U})^{\mathbb{I}_E}\Big)$$

$$= \mathrm{kl}\big((\mathbb{P}_\nu^{I_T} \otimes \mathcal{U})(E), (\mathbb{P}_{\nu'}^{I_T} \otimes \mathcal{U})(E)\big)\,.$$

The proof is concluded by noting that for all $\alpha = \nu$ or ν',

$$(\mathbb{P}_\alpha^{I_T} \otimes \mathcal{U})(E) = \mathbb{E}_\alpha[Z]\,.$$

□

Definition

A strategy is consistent if for all bandits problems ν, for all suboptimal arms a, i.e. $\Delta_a > 0$, it satisfies $\mathbb{E}_\nu\left[N_a(T)\right] = o(T^\alpha)$ for all $0 < \alpha \leqslant 1$.

Lower bound from Lai & Robbins :

Theorem (asymptotic distribution-dependent lower bounds)

For all consistent strategies, for all bandits problems ν, for all suboptimal arms a,

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\nu[N_a(T)]}{\ln T} \geqslant \frac{1}{\mathrm{kl}(\mu_a, \mu^\star)} .$$

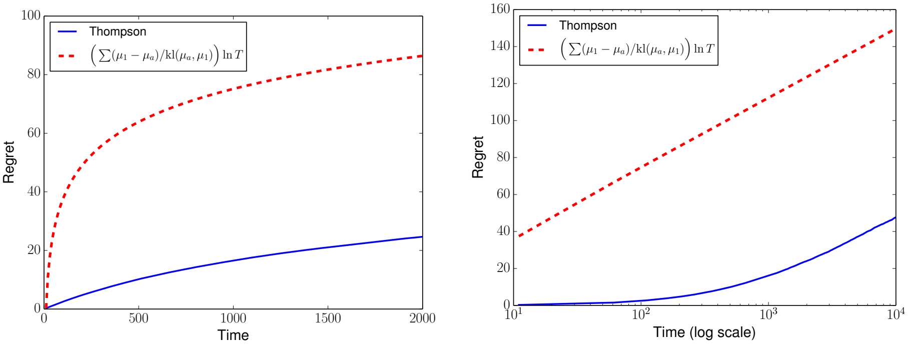

Figure : Bernoulli bandit problem with parameters : $(\mu_a)_{1\leqslant a\leqslant 6} = (0.05, 0.04, 0.02, 0.015, 0.01, 0.005)$

- Linear regret for T small.
- Logarithmic regret for large T (**asymptotic** lower bound).

Absolute lower bound for a suboptimal arm

In what follows $\nu = (\mathcal{B}(\mu_1), .., \mathcal{B}(\mu_K))$ with an unique optimal arm $i^\star$.

Absolute lower bound for a suboptimal arm

In what follows $\nu = (\mathcal{B}(\mu_1), .., \mathcal{B}(\mu_K))$ with an unique optimal arm $i^\star$.

Uniform strategy : pull an arm uniformly at random at each round.

Definition

A strategy is smarter than the uniform strategy if for all bandit problems ν, for all $T \geqslant 1$,

$$\mathbb{E}_\nu\left[N_{i^\star}(T)\right] \geqslant \frac{T}{K}$$
$$\mathbb{E}_\nu\left[N_a(T)\right] \leqslant \frac{T}{K} \quad a \text{ supotimal.}$$

Theorem

For all strategies that are smarter than the uniform strategy, for the bandit problems ν, for all arms a, for all $T \geqslant 1$,

$$\mathbb{E}_\nu\left[N_a(T)\right] \geqslant \frac{T}{K}\left(1 - \sqrt{2T\mathrm{kl}(\mu_a, \mu^\star)}\right).$$

In particular,

$$\forall\, T \leqslant \frac{1}{8\mathrm{kl}(\mu_a, \mu^\star)}, \qquad \mathbb{E}_\nu\left[N_a(T)\right] \geqslant \frac{T}{2K}.$$

Linear regret

a suboptimal arm.
Modified bandit problem $\nu' = (\mathcal{B}(\mu_1), .., \mathcal{B}(\mu'_a), .., \mathcal{B}(\mu_K))$ with $\mu'_a > \mu^\star$.
Tow blocks inequality,

$$\mathbb{E}_\nu[N_a(T)]\,\mathrm{kl}(\mu_a, \mu'_a) \geqslant \mathrm{kl}\Big(\mathbb{E}_\nu[N_a(T)]/T,\ \mathbb{E}_{\nu'}[N_a(T)]/T\Big)$$

smarter than the uniform : $\mathbb{E}_\nu[N_a(T)]/T \leqslant 1/K \leqslant \mathbb{E}'_\nu[N_a(T)]/T$ and $q \mapsto \mathrm{kl}(p, q)$ is increasing on $[p, 1]$,

$$\geqslant \mathrm{kl}\Big(\mathbb{E}_\nu[N_a(T)]/T,\ 1/K\Big)$$

Pinsker inequality,

$$\geqslant \frac{K}{2}\Big(\mathbb{E}_\nu[N_a(T)]/T - 1/K\Big)^2$$

Still with $\mathbb{E}_\nu[N_a(T)]/T \leqslant 1/K$:

$$\mathrm{kl}(\mu_a, \mu'_a)T/K \geqslant \frac{K}{2}\Big(\mathbb{E}_\nu[N_a(T)]/T - 1/K\Big)^2$$