

# Bandits : optimality in exponential families

Odalric-Ambrym Maillard

IHES, January 2016



- ① Stochastic multi-armed bandits
- ② Boundary crossing probabilities
- ③ Analysis for exponential families

# **Stochastic multi-armed bandits**

## Clinical trials

- ▶  $K$  possible treatments (unknown effects)



- ▶ What treatment allocate to each patient, depending on observed effects on previous patients?

## Online advertising

- ▶  $K$  ads that can be displayed



- ▶ What ad show to each user, depending on observed clicks from past users?

## Setting

Set of choices  $\mathcal{A}$ . Each  $a \in \mathcal{A}$  is associated with an unknown probability distribution  $\nu_a \in \mathcal{D}$  with mean  $\mu_a$ .

## Game

The game is *sequential*: At each round  $t \geq 1$ ,

- ▶ the player first picks an arm  $A_t \in \mathcal{A}$  based on previous observations.
- ▶ then receives (and sees) a stochastic payoff  $Y_t \sim \nu_{A_t}$ .

## Goal

Maximize the cumulative payoff  $\mathbb{E} \left[ \sum_{t=1}^T Y_t \right]$  or equivalently minimize the **regret** at round  $T$ :

$$R_T \stackrel{\text{def}}{=} T\mu^* - \mathbb{E} \left[ \sum_{t=1}^T Y_t \right] \quad \text{where}$$

$$\mu^* = \max \{ \mu_a ; a \in \mathcal{A} \}, \quad a^* \in \operatorname{argmax} \{ \mu_a ; a \in \mathcal{A} \}.$$

# Applications of Bandits

amazon.com

Recommended for You

Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own.



What book to display?

(+Few interactions)



What drug is the most efficient?

(+Economic/Human cost)



What power to inject in the electrical network?

(+Fast decisions)



What news to display?

(+Short lifespan)



What action to choose?

(+Risk Aversion)



How to organize future cities?

(+Long-term decisions)

# DECISION MAKING UNDER UNCERTAINTY

## EXPLORATION VERSUS EXPLOITATION

- ▶ Sample more information or trust current estimates ?

## OPTIMISM IN FACE OF (MODELED) UNCERTAINTY

- ▶ Build empirical confidence bounds on value. Choose point with highest value.

- ▶ 1933: Introduction by Thompson ("Bayesian" way)
- ▶ 1952: Frequentist formalization by Robbins.
- ▶ 1985: Lai and Robbins' first lower-performance bounds.
- ▶ 1996: Burnetas and Katehakis get general **lower-performance** bounds.
  
- ▶ 2002: Upper-Confidence Bound (UCB) algorithm gets first **non-asymptotic** guarantee.
- ▶ 2010+: Provably optimal strategies KL-UCB, D-MED, Thompson Sampling for **specific** distributions.
- ▶ This talk: Extension to **general exponential families**.

## Lower bound (Burnetas and Katehakis 96)

For all "consistent" algorithm, for arbitrary  $\mathcal{D} \subset \mathcal{P}([0, 1])$

$$\liminf_{T \rightarrow \infty} \frac{R_T}{\log(T)} \geq \sum_{a \in \mathcal{A}} \frac{\mu_\star - \mu_a}{K_{inf}(\nu_a, \mu_\star)},$$

where  $K_{inf}(\nu_a, \mu_\star) = \inf\{KL(\nu_a, \nu) : \nu \in \mathcal{D}, \mathbb{E}_\nu[Y] > \mu_\star\}$ .

## Upper bounds (Cappe, Garivier, M., Munos, Stoltz 2013)

KL-UCB achieves for specific  $\mathcal{D}$  (e.g. 1-dim exponential families):

$$R_T \leq \sum_{a \in \mathcal{A}} \frac{\mu_\star - \mu_a}{K_{inf}(\nu_a, \mu_\star)} \log(T) + \diamond \log(T)^{2/3}.$$

Pull arm  $A_{t+1}$  that maximizes the upper bound

$$\sup \left\{ \mathbb{E}_\nu[Y] : \nu \in \mathcal{D}, KL\left(\Pi_{\mathcal{D}}(\widehat{\nu}_{a,N_a(t)}), \nu\right) \leq \frac{f(t)}{N_a(t)} \right\}$$

where

- ▶  $\widehat{\nu}_{a,n} = \frac{1}{n} \sum_{m=1}^n \delta_{X_{a,m}}$ ,  $N_a(t) = \sum_{s=1}^t \mathbb{I}\{A_s = a\}$ .
- ▶  $\Pi_{\mathcal{D}} : \mathfrak{M}_1(\mathbb{R}) \rightarrow \mathcal{D}$  projection on  $\mathcal{D}$ .
- ▶  $f: \mathbb{N} \rightarrow \mathbb{R}$  non-decreasing.

Build empirical distribution for arm  $a$ .

Pull arm  $A_{t+1}$  that maximizes the upper bound

$$\sup \left\{ \mathbb{E}_\nu[Y] : \nu \in \mathcal{D}, KL\left(\Pi_{\mathcal{D}}(\widehat{\nu}_{a,N_a(t)}), \nu\right) \leq \frac{f(t)}{N_a(t)} \right\}$$

where

- ▶  $\widehat{\nu}_{a,n} = \frac{1}{n} \sum_{m=1}^n \delta_{X_{a,m}}$ ,  $N_a(t) = \sum_{s=1}^t \mathbb{I}\{A_s = a\}$ .
- ▶  $\Pi_{\mathcal{D}} : \mathfrak{M}_1(\mathbb{R}) \rightarrow \mathcal{D}$  projection on  $\mathcal{D}$ .
- ▶  $f: \mathbb{N} \rightarrow \mathbb{R}$  non-decreasing.

Project it on  $\mathcal{D}$ .

Pull arm  $A_{t+1}$  that maximizes the upper bound

$$\sup \left\{ \mathbb{E}_\nu[Y] : \nu \in \mathcal{D}, KL\left(\Pi_{\mathcal{D}}(\widehat{\nu}_{a,N_a(t)}), \nu\right) \leq \frac{f(t)}{N_a(t)} \right\}$$

where

- ▶  $\widehat{\nu}_{a,n} = \frac{1}{n} \sum_{m=1}^n \delta_{X_{a,m}}$ ,  $N_a(t) = \sum_{s=1}^t \mathbb{I}\{A_s = a\}$ .
- ▶  $\Pi_{\mathcal{D}} : \mathfrak{M}_1(\mathbb{R}) \rightarrow \mathcal{D}$  projection on  $\mathcal{D}$ .
- ▶  $f : \mathbb{N} \rightarrow \mathbb{R}$  non-decreasing. **Threshold**.

Consider all  $\nu \in \mathcal{D}$  close to this distribution.

Pull arm  $A_{t+1}$  that maximizes the upper bound

$$\sup \left\{ \mathbb{E}_\nu[Y] : \nu \in \mathcal{D}, KL\left(\Pi_{\mathcal{D}}(\widehat{\nu}_{a,N_a(t)}), \nu\right) \leq \frac{f(t)}{N_a(t)} \right\}$$

where

- ▶  $\widehat{\nu}_{a,n} = \frac{1}{n} \sum_{m=1}^n \delta_{X_{a,m}}$ ,  $N_a(t) = \sum_{s=1}^t \mathbb{I}\{A_s = a\}$ .
- ▶  $\Pi_{\mathcal{D}} : \mathfrak{M}_1(\mathbb{R}) \rightarrow \mathcal{D}$  projection on  $\mathcal{D}$ .
- ▶  $f: \mathbb{N} \rightarrow \mathbb{R}$  non-decreasing.

Compute the maximal mean of these candidate distributions.

Pull arm  $A_{t+1}$  that maximizes the upper bound

$$\sup \left\{ \mathbb{E}_\nu[Y] : \nu \in \mathcal{D}, KL\left(\Pi_{\mathcal{D}}(\widehat{\nu}_{a,N_a(t)}), \nu\right) \leq \frac{f\left(t/N_a(t)\right)}{N_a(t)} \right\}$$

where

- ▶  $\widehat{\nu}_{a,n} = \frac{1}{n} \sum_{m=1}^n \delta_{X_{a,m}}$ ,  $N_a(t) = \sum_{s=1}^t \mathbb{I}\{A_s = a\}$ .
- ▶  $\Pi_{\mathcal{D}} : \mathfrak{M}_1(\mathbb{R}) \rightarrow \mathcal{D}$  projection on  $\mathcal{D}$ .
- ▶  $f : \mathbb{N} \rightarrow \mathbb{R}$  non-decreasing.

The regret is given by

$$R_T = \sum_{a \in \mathcal{A}} (\mu^* - \mu_a) \mathbb{E}[N_a(T)] \leq \sum_{a \in \mathcal{A}} (\mu^* - \mu_a) \underbrace{\frac{\log(T)}{K_{inf}(\nu_a, \mu^*)}_{(a)}} + \underbrace{\diamond \log(T)^{2/3}}_{(b)}.$$

## Theorem

Let  $0 < \varepsilon < \min\{\mu^* - \mu_a, a \in \mathcal{A}\}$ . Then, the number of pulls of a sub-optimal arm  $a \in \mathcal{A}$  by Algorithm KL-UCB satisfies

$$\mathbb{E}[N_T(a)] \leq 2 + \underbrace{\sum_{n=1}^T \mathbb{P}\left\{ \mathcal{K}_{inf}(\Pi_D(\hat{\nu}_{a,n}), \mu^* - \varepsilon) \leq f(T)/n \right\}}_{(a) \text{ Easy term, via Sanov}} + \underbrace{\sum_{t=|\mathcal{A}|}^{T-1} \mathbb{P}\left\{ \mathcal{K}_{inf}(\Pi_D(\hat{\nu}_{a^*, N_{a^*}(t)}), \mu^* - \varepsilon) \geq f(t)/N_{a^*}(t) \right\}}_{(b) \text{ Difficult: Boundary Crossing Probability}}.$$

The regret is given by

$$R_T = \sum_{a \in \mathcal{A}} (\mu^* - \mu_a) \mathbb{E}[N_a(T)] \leq \sum_{a \in \mathcal{A}} (\mu^* - \mu_a) \underbrace{\frac{\log(T)}{K_{inf}(\nu_a, \mu^*)}_{(a)}} + \underbrace{\diamond \log(T)^{2/3}}_{(b)}.$$

## Theorem

Let  $0 < \varepsilon < \min\{\mu^* - \mu_a, a \in \mathcal{A}\}$ . Likewise, the number of pulls of a sub-optimal arm  $a \in \mathcal{A}$  by Algorithm **KL-UCB<sup>+</sup>** satisfies

$$\mathbb{E}[N_T(a)] \leq 2 + \underbrace{\sum_{n=1}^T \mathbb{P}\left\{ \mathcal{K}_{inf}(\Pi_D(\hat{\nu}_{a,n}), \mu^* - \varepsilon) \leq f(T/n)/n \right\}}_{(a) \text{ Easy term, via Sanov}} + \underbrace{\sum_{t=|\mathcal{A}|}^{T-1} \mathbb{P}\left\{ \mathcal{K}_{inf}(\Pi_D(\hat{\nu}_{a^*, N_{a^*}(t)}), \mu^* - \varepsilon) \geq f(t/N_{a^*}(t))/N_{a^*}(t) \right\}}_{(b) \text{ Difficult: Boundary Crossing Probability}}.$$

$\mathcal{E}$  parametric family of distributions of the form

$$\text{where } \nu_\theta(x) = \exp(\langle \theta, F(x) \rangle - \psi(\theta)) \nu_0(x),$$

- ▶  $F : \mathcal{X} \rightarrow \mathbb{R}^K$ ,  $\nu_0 \in \mathcal{P}(\mathcal{X})$ : reference function/measure.
- ▶  $\theta \in \Theta_{\mathcal{E}} \stackrel{\text{def}}{=} \left\{ \theta \in \mathbb{R}^K ; \psi(\theta) < \infty \right\}$  natural parameter.
- ▶  $\psi(\theta) \stackrel{\text{def}}{=} \log \int_{\mathcal{X}} \exp(\langle \theta, F(x) \rangle) \nu_0(dx)$ : log-partition function.

## Example

- ▶ Bernoulli:  $\mathcal{X} = \{0, 1\}$ ,  $F : x \rightarrow x$  ( $K = 1$ ),  $\Theta_{\mathcal{D}} = \mathbb{R}$ ,  
 $\psi(\theta) = \log(1 + e^\theta)$ .  $\mathcal{B}(\mu)$ : parameter  $\theta = \log(\mu/(1 - \mu))$ .
- ▶ Gaussian:  $\mathcal{X} = \mathbb{R}$ ,  $F : x \rightarrow (x, x^2)$  ( $K = 2$ ),  $\Theta_{\mathcal{D}} = \mathbb{R} \times \mathbb{R}_*^-$ ,  
 $\psi(\theta) = -\frac{\theta_1^2}{4\theta_2} + \frac{1}{2} \log\left(\frac{-\pi}{\theta_2}\right)$ .  $\mathcal{N}(\mu, \sigma^2)$ : parameter  $\theta = (\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2})$ .

- ▶  $\mathbb{E}_{\nu_\theta}(F(X)) = \nabla\psi(\theta)$

- ▶  $\mathbb{E}_{\nu_\theta}(F(X)) = \nabla\psi(\theta)$
- ▶  $KL(\nu_\theta, \nu_{\theta'}) = \underbrace{\mathcal{B}^\psi(\theta, \theta')}_\text{Bregman div.} = \langle \underbrace{\theta - \theta'}_\text{natural}, \underbrace{\nabla\psi(\theta)}_\text{dual} \rangle - \psi(\theta) + \psi(\theta').$

- ▶  $\mathbb{E}_{\nu_\theta}(F(X)) = \nabla\psi(\theta)$
- ▶  $KL(\nu_\theta, \nu_{\theta'}) = \underbrace{\mathcal{B}^\psi(\theta, \theta')}_\text{Bregman div.} = \langle \underbrace{\theta - \theta'}_\text{natural}, \underbrace{\nabla\psi(\theta)}_\text{dual} \rangle - \psi(\theta) + \psi(\theta').$
- ▶ Estimation for regular family:

- ▶  $\mathbb{E}_{\nu_\theta}(F(X)) = \nabla\psi(\theta)$
- ▶  $KL(\nu_\theta, \nu_{\theta'}) = \underbrace{\mathcal{B}^\psi(\theta, \theta')}_\text{Bregman div.} = \langle \underbrace{\theta - \theta'}_\text{natural}, \underbrace{\nabla\psi(\theta)}_\text{dual} \rangle - \psi(\theta) + \psi(\theta').$
- ▶ Estimation for regular family:
  - ▶ Let  $\{X_i\}_{i \leq n} \sim \nu_{\theta^*}$ ,  $\widehat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ .

- ▶  $\mathbb{E}_{\nu_\theta}(F(X)) = \nabla\psi(\theta)$
- ▶  $KL(\nu_\theta, \nu_{\theta'}) = \underbrace{\mathcal{B}^\psi(\theta, \theta')}_\text{Bregman div.} = \langle \underbrace{\theta - \theta'}_\text{natural}, \underbrace{\nabla\psi(\theta)}_\text{dual} \rangle - \psi(\theta) + \psi(\theta').$
- ▶ Estimation for regular family:
  - ▶ Let  $\{X_i\}_{i \leq n} \sim \nu_{\theta^*}$ ,  $\widehat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ .
  - ▶  $\widehat{F}_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n F(X_i) = \mathbb{E}_{\widehat{\nu}_n}(F(X)) \in \overline{\nabla\psi(\Theta_{\mathcal{D}})}$

In the sequel, we would like to have  
 $\Pi_{\mathcal{E}}(\widehat{\nu}_{\theta^*, n}) = \nu_{\widehat{\theta}_n}$  for some  $\widehat{\theta}_n \in \Theta_{\mathcal{E}}$ .

## Technical assumption: Well-defined empirical parameter $\hat{\theta}_n$

We assume that  $\theta^* \in \Theta$  and  $\hat{F}_n \in \nabla \psi(\Theta_\rho)$  where

- ▶  $\Theta$  is convex
- ▶ Neighborhood  $\Theta_{2\rho} \subset \mathring{\Theta}_I$  where
  - ▶  $\Theta_{2\rho} = \{\theta : \inf_{\theta' \in \Theta} \|\theta - \theta'\| < 2\rho\}$
  - ▶  $\Theta_I \stackrel{\text{def}}{=} \left\{ \theta \in \Theta_{\mathcal{E}} ; 0 < \lambda_{\text{MIN}}(\nabla^2 \psi(\theta)) \leq \lambda_{\text{MAX}}(\nabla^2 \psi(\theta)) < \infty \right\},$   
with  $\lambda_{\text{MIN/MAX}}(M)$ : min/max eigenvalues of a sdp matrix  $M$ .
- ▶  $2\rho < \rho^* = d(\theta^*, \Theta_I^c)$

$$\theta^* \in \Theta \subset \Theta_{2\rho} \subset \mathring{\Theta}_I \subset \Theta_{\mathcal{E}}$$

Then we can form  $\Pi_{\mathcal{E}}(\widehat{\nu}_{a^*, n}) = \nu_{\widehat{\theta}_n}$  with  $\widehat{\theta}_n \in \Theta_\rho$  and

$$\mathcal{K}_{\inf}(\Pi_{\mathcal{D}}(\widehat{\nu}_{a^*, n}), \mu^* - \varepsilon) = \inf \{ \mathcal{B}^\psi(\widehat{\theta}_n, \theta) : \mathbb{E}_{\nu_\theta}[X] \geq \mu^* - \varepsilon \}$$

# **Boundary crossing probabilities**

for general exponential families

$$\mathbb{P}_{\theta^*} \left\{ \mathcal{K}_{inf}(\Pi_{\mathcal{E}}(\widehat{\nu}_{a^*}, N_{a^*}(t)), \mu^* - \varepsilon) \geq f(t/N_{a^*}(t))/N_{a^*}(t) \right\} \leq \\ \mathbb{P}_{\theta^*} \left\{ \bigcup_{n=1}^t \inf \{ \mathcal{B}^\psi(\widehat{\theta}_n, \theta) : \mathbb{E}_{\nu_\theta}[X] \geq \mu^* - \varepsilon \} \geq f(t/n)/n \right\}$$

E.g. for canonical exp. family of dimension 1:

Unique  $\theta = \theta_\varepsilon^* \in \mathbb{R}$  such that  $\mathbb{E}_{\nu_\theta}[X] = \mu^* - \varepsilon$ , thus:

$$\mathbb{P}_{\theta^*} \left\{ \bigcup_{n=1}^t \mathcal{B}^\psi(\widehat{\theta}_n, \theta_\varepsilon^*) \geq f(t/n)/n \right\}$$

We want this to be  $o(1/t)$  for well-chosen  $f$ .

## Theorem (Cappé, Garivier, M., Munos, Stoltz 2013)

For canonical ( $F(x) = x \in \mathcal{X}$ ) exp. families of dimension  $K = 1$ :

$$\mathbb{P}_{\theta^*} \left\{ \bigcup_{n=1}^t \mathcal{B}^\psi(\hat{\theta}_n, \theta^*) \geq f(t)/n \right\} \leq e^{\lceil f(t) \log(t) \rceil} e^{-f(t)}.$$

For  $f(x) = \log(x) + \xi \log \log(x)$ , the bound is  $O(\frac{\log^{2-\xi}(t)}{t})$ .

A similar result is shown for discrete distributions.

## Limitations

- ▶ Requires  $\xi > 2$ , while experiments shows  $\xi \leq 0$  still works (and even better).
- ▶ Only for threshold  $f(t)/n$ , does not work for  $f(t/n)/n$ .
- ▶ Consider  $\theta^*$ : does not take advantage of  $\varepsilon > 0$ .

Theorem (Lai, 1988; exponential family of dimension  $K$ )

For  $\gamma \in (0, 1)$  let  $\mathcal{C}_\gamma(\theta) = \{\theta' \in \mathbb{R}^K : \langle \theta', \theta \rangle \geq \gamma |\theta| |\theta'| \}$ . Then, for  $f(x) = \log(x) + \xi \log \log(x)$  it holds for all  $\theta^\dagger \in \Theta_\rho$  such that  $|\theta^\dagger - \theta^*|^2 \geq \delta_t$ , where  $\delta_t \rightarrow 0$ ,  $t\delta_t \rightarrow \infty$  as  $t \rightarrow \infty$ ,

$$\mathbb{P}_{\theta^*} \left\{ \bigcup_{n=1}^t \widehat{\theta}_n \in \Theta_\rho \cap \mathcal{B}^\psi(\widehat{\theta}_n, \theta^\dagger) \geq f(t/n)/n \cap \nabla \psi(\widehat{\theta}_n) - \nabla \psi(\theta^\dagger) \in \mathcal{C}_\gamma(\theta^\dagger - \theta^*) \right\} \\ \stackrel{t \rightarrow \infty}{=} O\left( \frac{|\theta^\dagger - \theta^*|^{-2}}{t} \log^{-\xi-1+K/2}(t|\theta^\dagger - \theta^*|^2) \right).$$

### Pros

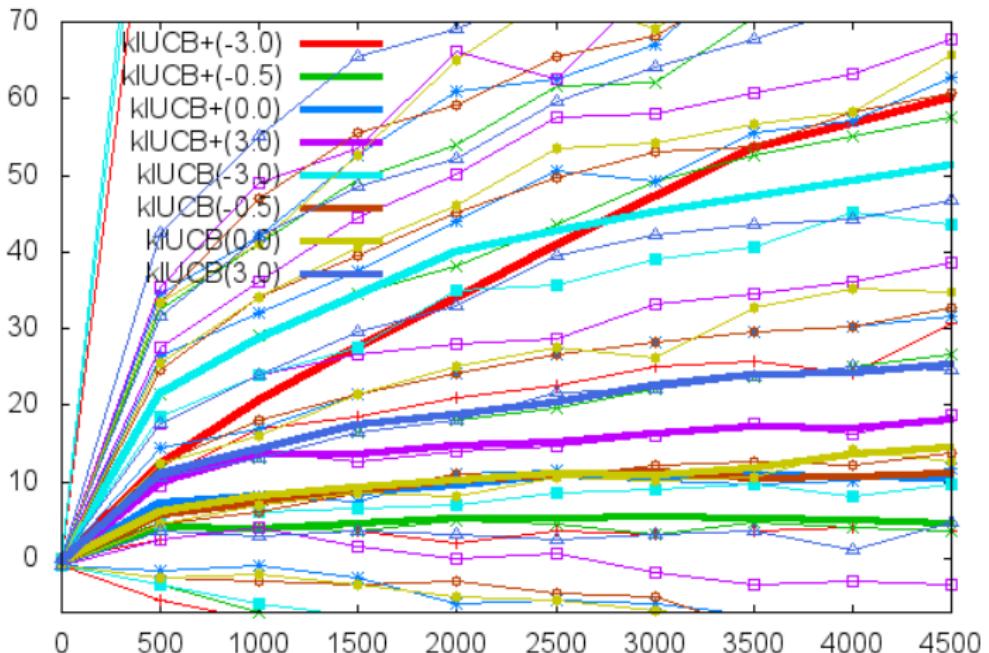
- ▶  $\xi > K/2 - 1$  is enough.
- ▶ For  $f(t/n)$

### Cons

- ▶ Blue term
- ▶ Asymptotic

"Boundary crossing problems for sample means", 1988.

# Illustrative experiment on Bernoulli distributions



Regret as function of time for KL-UCB and KL-UCB+  
(mean and quantiles).

- ▶ Explain the proof technique for this result:
  - ➊ Peeling (standard) and covering
  - ➋ Change of measure argument
  - ➌ Localization + change of measure
  - ➍ Concentration (standard)
  - ➎ Fine tuning (gain  $\log^2$  factor)
- ▶ Show that Lai's theorem can be made non-asymptotic, and used to control the general case.

# **Analysis and proof technique.**

(up to some details for presentation purpose)

$$\mathbb{P}_{\theta^*} \left\{ \bigcup_{1 \leq n \leq t} \underbrace{\widehat{\theta}_n \in \Theta_\rho \cap \mathcal{K}_{inf}(\Pi_{\mathcal{E}}(\widehat{\nu}_{a^*,n}), \mu^* - \varepsilon)}_{E_n} \geq f(t/n)/n \right\}$$

## I. Peeling

Let  $n_0 = 1 < n_1 < \dots < n_{I_t} = t + 1$  for some  $I_t \in \mathbb{N}_\star$ .

$$\mathbb{P}_{\theta^*} \left\{ \bigcup_{1 \leq n \leq t} E_n \right\} \leq \sum_{i=0}^{I_t-1} \mathbb{P}_{\theta^*} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_n \right\}.$$

For some  $b > 1$  and  $\beta \in (0, 1)$ , use:

$$n_i = \begin{cases} b^i & \text{if } i < I_t \stackrel{\text{def}}{=} \lceil \log_b(\beta t + \beta) \rceil \\ t + 1 & \text{if } i = I_t, \end{cases}$$

(valid sequence since  $n_{I_t-1} \leq b^{\log_b(\beta t + \beta)} < t + 1 = n_{I_t}$ )

## II. Cone covering

① First:

$$\begin{aligned}\mathcal{K}_{inf}(\Pi_{\mathcal{E}}(\widehat{\nu}_{\theta^*, n}), \mu^* - \varepsilon) \\ = \inf \{ \mathcal{B}^\psi(\widehat{\theta}_n, \theta) : \theta \in \Theta_{\mathcal{E}}, \mu_\theta \geq \mu^* - \varepsilon \} \\ = \inf \{ \mathcal{B}^\psi(\widehat{\theta}_n, \theta^* - \Delta) : \theta^* - \Delta \in \Theta_{\mathcal{E}}, \mu_{\theta^* - \Delta} \geq \mu^* - \varepsilon \}\end{aligned}$$

② Then

$$\begin{aligned}\mathcal{B}^\psi(\widehat{\theta}_n, \theta^* - \Delta) \\ = \mathcal{B}^\psi(\widehat{\theta}_n, \theta^*) - \underbrace{\mathcal{B}^\psi(\theta^* - \Delta, \theta^*) - \langle \Delta, \nabla \psi(\theta^* - \Delta) - \widehat{\nabla} \psi(\widehat{\theta}_n) \rangle}_{\text{shift}}.\end{aligned}$$

## II. Cone covering

Consider half-cone  $\mathcal{C}_\gamma(\Delta) = \left\{ \theta' \in \mathbb{R}^K : \langle \theta', \Delta \rangle \geq \gamma \|\theta'\| \|\Delta\| \right\}$ ,

then set  $\{\Delta_c\}_{1 \leq c \leq C_{\gamma,K}}$  of directions s.t.  $\mathbb{R}^K = \bigcup_{c=1}^{C_{\gamma,K}} \mathcal{C}_\gamma(\Delta_c)$ .

- ▶  $\mathcal{C}_\gamma(\Delta)$  invariant by rescaling of  $\Delta$ : We can choose  $\Delta_c$  small enough so that  $\theta^\star - \Delta_c \in \Theta_\rho$ .
- ▶  $C_{\gamma,1} = 2$ ,  $C_{\gamma,K} \rightarrow \infty$  when  $\gamma \rightarrow 1$ .

## II. Cone covering

- $\mathcal{K}_{inf}(\Pi_{\mathcal{E}}(\widehat{\nu}_{\theta^*, n}), \mu^* - \varepsilon)$   
 $= \inf \{ \mathcal{B}^\psi(\widehat{\theta}_n, \theta^* - \Delta) : \theta^* - \Delta \in \Theta_{\mathcal{E}}, \mu_{\theta^* - \Delta} \geq \mu^* - \varepsilon \}$   
 $\leq \mathcal{B}^\psi(\widehat{\theta}_n, \theta^* - \Delta_c) \quad \forall \theta^* - \Delta_c \in \Theta_\rho, \mu_{\theta^* - \Delta_c} \geq \mu^* - \varepsilon, c \in [C_{\gamma, K}]$

- Thus  $\mathbb{P}_{\theta^*} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_n \right\} \leq \min_c \sum_{c'=1}^{C_{\gamma, K}} \mathbb{P}_{\theta^*} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n, c, c'} \right\}$

where  $E_{n, c, c'} = \left\{ \widehat{\theta}_n \in \Theta_\rho \cap \nabla \psi(\theta^* - \Delta_c) - \widehat{F}_n \in \mathcal{C}_\gamma(\Delta'_c) \right.$   
 $\left. \cap \mathcal{B}^\psi(\widehat{\theta}_n, \theta^* - \Delta_c) \geq \frac{f(t/n)}{n} \right\}$

$E_{n, c, c'}$  is almost the event appearing in Lai's theorem.

$$\mathbb{P}_{\theta^*} \left\{ \bigcup_{1 \leq n \leq t} E_n \right\} \leq \sum_{i=0}^{I_t-1} \min_c \sum_{c'=1}^{C_{\gamma,K}} \mathbb{P}_{\theta^*} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n,c,c'} \right\}.$$

## Lemma (Change of measure)

If  $n \rightarrow nf(t/n)$  is non-decreasing and  $\theta^* - \Delta_c \in \Theta_\rho$ , then

$$\begin{aligned} \mathbb{P}_{\theta^*} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n,c,c'} \right\} &\leq C \exp \left( -\frac{n_i}{2} v_\rho \delta_c^2 - \gamma \delta_c v_\rho \sqrt{\frac{2n_i f(t/n_i)}{V_\rho}} \right) \\ &\quad \times \mathbb{P}_{\theta^* - \Delta_c} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n,c,c'} \right\}, \end{aligned}$$

where  $v_\rho = \inf_{\theta \in \Theta_\rho} \lambda_{MIN}(\nabla^2 \psi(\theta))$ ,  $V_\rho = \sup_{\theta \in \Theta_\rho} \lambda_{MAX}(\nabla^2 \psi(\theta))$ ,  $\delta_c = \|\Delta_c\|$ .

Concentration of  $\mathcal{B}^\psi(\hat{\theta}_n, \theta^* - \Delta_c)$  for distribution  $\nu_{\theta^*}$  vs  
 concentration of  $\mathcal{B}^\psi(\hat{\theta}_n, \theta^* - \Delta_c)$  for distribution  $\nu_{\theta^* - \Delta_c}$ .

### III. Change of measure: proof details

- ▶ for  $n$  i.i.d. variables look at the ratio

$$\begin{aligned}\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}_{\theta^* - \Delta_c}} &= \frac{\prod_{k=1}^n \nu_{\theta^*}(X_k)}{\prod_{k=1}^n \nu_{\theta^* - \Delta_c}(X_k)} \\ &= \exp \left( n \langle \Delta_c, \hat{F}_{a^*, n} \rangle - n (\psi(\theta^*) - \psi(\theta^* - \Delta_c)) \right) \\ &= \exp \left( - n \underbrace{\langle \Delta_c, \nabla \psi(\theta^* - \Delta_c) - \hat{F}_{a^*, n} \rangle}_{(a)} - n \underbrace{\mathcal{B}^\psi(\theta^* - \Delta_c, \theta^*)}_{(b)} \right).\end{aligned}$$

### III. Change of measure: proof details

- ▶ for  $n$  i.i.d. variables look at the ratio

$$\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}_{\theta^* - \Delta_c}} = \exp \left( -n \underbrace{\langle \Delta_c, \nabla \psi(\theta^* - \Delta_c) - \hat{F}_{a^*, n} \rangle}_{(a)} - n \underbrace{\mathcal{B}^\psi(\theta^* - \Delta_c, \theta^*)}_{(b)} \right).$$

- ▶ For (b):

$$\mathcal{B}^\psi(\theta^* - \Delta_c, \theta^*) \geq \frac{1}{2} v_\rho \delta_c^2.$$

### III. Change of measure: proof details

- for  $n$  i.i.d. variables look at the ratio

$$\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}_{\theta^* - \Delta_c}} = \exp \left( -n \underbrace{\langle \Delta_c, \nabla \psi(\theta^* - \Delta_c) - \hat{F}_{a^*, n} \rangle}_{(a)} - n \underbrace{\mathcal{B}^\psi(\theta^* - \Delta_c, \theta^*)}_{(b)} \right).$$

- For (b):

$$\mathcal{B}^\psi(\theta^* - \Delta_c, \theta^*) \geq \frac{1}{2} v_\rho \delta_c^2.$$

- For (a):

(up to going from  $\Delta_c$  to  $\Delta'_c$ )

$$\begin{aligned} \langle \Delta_{c'}, \nabla \psi(\theta^* - \Delta_c) - \hat{F}_n \rangle &\geq \gamma ||\Delta_{c'}|| ||\nabla \psi(\theta^* - \Delta_c) - \hat{F}_n|| \\ &\geq \gamma \delta_{c'} v_\rho ||\hat{\theta}_n^* - \theta^* + \Delta_c|| \\ &\geq \gamma \delta_{c'} v_\rho \sqrt{\frac{2}{V_\rho} \mathcal{B}^\psi(\hat{\theta}_n, \theta^* - \Delta_c)} \\ &\geq \gamma \delta_{c'} v_\rho \sqrt{\frac{2f(t/n)}{V_\rho n}}. \end{aligned}$$

- ▶ By construction:

$$\nabla \psi(\theta^* - \Delta_c) - \hat{F}_n \in \mathcal{C}_p(\Delta_{c'}) \Leftrightarrow$$

$$\langle \Delta_{c'}, \nabla \psi(\theta^* - \Delta_c) - \hat{F}_n \rangle \geq \gamma ||\Delta_{c'}|| ||\nabla \psi(\theta^* - \Delta_c) - \hat{F}_n||,$$

Thus, we look at  $||\nabla \psi(\theta^* - \Delta_c) - \hat{F}_n||$ .

- ▶ Decomposition, for some constant  $\varepsilon_{t,i,c} > 0$

$$\begin{aligned} & \mathbb{P}_{\theta^* - \Delta_c} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n,c} \right\} \\ & \leq \underbrace{\mathbb{P}_{\theta^* - \Delta_c} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n,c} \cap ||\nabla \psi(\theta^* - \Delta_c) - \hat{F}_n|| < \varepsilon_{t,i,c} \right\}}_{(a) \text{ Requires some care}} \\ & \quad + \underbrace{\mathbb{P}_{\theta^* - \Delta_c} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n,c} \cap ||\nabla \psi(\theta^* - \Delta_c) - \hat{F}_n|| \geq \varepsilon_{t,i,c} \right\}}_{(b) \text{ "Standard" concentration}}. \end{aligned}$$

## Lemma (Concentration (admitted))

*Under some technical assumption,*

$$(b) \leq 2K \exp\left(-\frac{n_i^2 \varepsilon_{t,i,c}^2}{2KV_{2\rho} n_{i+1}}\right)$$

For  $\varepsilon_{t,i,c} = \sqrt{\frac{2KV_{2\rho} n_{i+1} f(t/n_{i+1})}{n_i^2}}$ , then (b)  $\leq 2K \exp(-f(t/n_{i+1}))$ .

## Lemma (Localized change of measure)

For the  $\varepsilon_{t,i,c}$  considered, we have

$$(a) \leq 2C_{b,\rho} \exp\left(-f(t/n_{i+1})\right) f(t/n_{i+1})^{K/2},$$

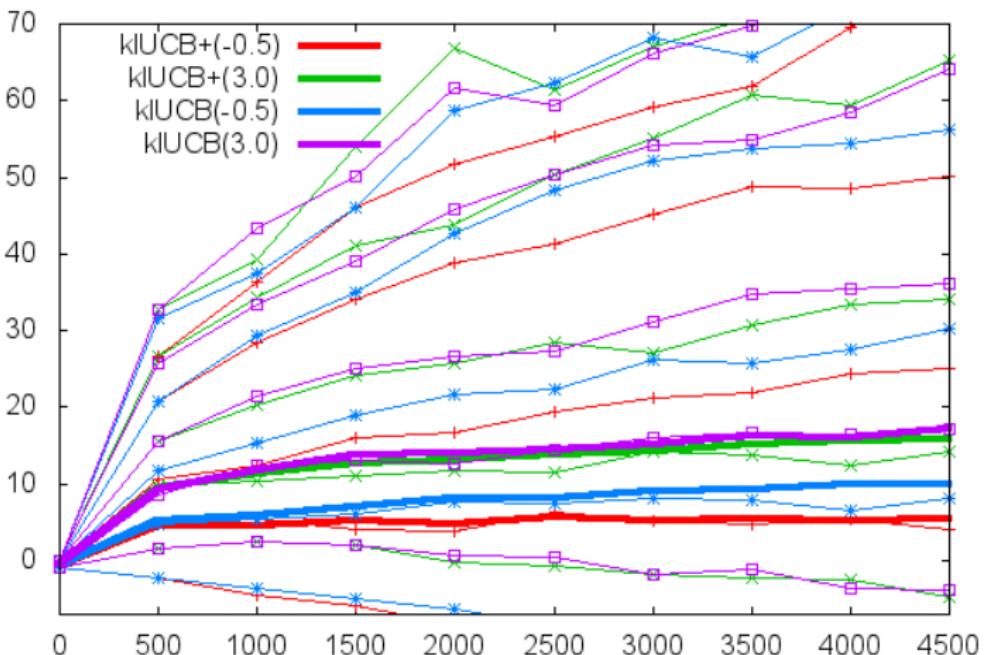
for an explicit constant  $C_{b,\rho} = \max\left\{\frac{2KbV_{2\rho}}{\rho^2 v_\rho^2}, 4, \frac{8K^2 b^2 V_{2\rho}^2}{(K+2)v_\rho^2}\right\}^{K/2}$ .

$$\begin{aligned}
 & \mathbb{P}_{\theta^*} \left\{ \bigcup_{1 \leq n \leq t} \widehat{\theta}_n \in \Theta_\rho \cap \mathcal{K}_{inf}(\Pi_{\mathcal{E}}(\widehat{\nu}_{a^*, n}), \mu^* - \varepsilon) \geq f(t/n)/n \right\} \\
 & \leq \underbrace{\sum_{c=1}^{C_{\gamma, K}} \sum_{i=0}^{I_t-1}}_{\text{Peeling+Covering}} \underbrace{Ce^{-\frac{n_i}{2} v_\rho \delta_c^2 - \gamma \delta_c v_\rho \sqrt{\frac{2n_i f(t/n_i)}{v_\rho}}}}_{\text{Change of measure}} \\
 & \quad \times \underbrace{e^{-f(t/n_{i+1})} \left[ 2CC_{b,\rho} f(t/n_{i+1})^{K/2} + 2K \right]}_{\text{Localized change of measure + Concentration}}
 \end{aligned}$$

End of the proof:

- ▶ "Change of measure" term kills  $n_i \gtrsim \delta_c^{-2}/\log(t\delta_c^2)$  terms
- ▶ Remains  $e^{-f(t\delta_c^2)} f(t\delta_c^2)^{K/2}/\log(t\delta_c^2) \simeq \frac{1}{t\delta_c^2} \log(t\delta_c^2)^{-\xi+K/2-1}$

## Illustrative experiment on Bernoulli distributions



Regret as function of time for KL-UCB and KL-UCB+ (mean and quantiles).

$$\mathbb{P}_{\theta^* - \Delta_c} \left\{ \bigcup_{n_i \leq n < n_{i+1}} E_{n,c} \cap \|\nabla \psi(\theta^* - \Delta_c) - \widehat{F}_n\| < \varepsilon_{t,i,c} \right\}$$

- ▶ Let  $\theta_c^* \stackrel{\text{def}}{=} \theta^* - \Delta_c \in \Theta_\rho$
- ▶  $E_{n,c} \subset \{\widehat{\theta}_n \in \Theta_\rho \cap \mathcal{B}^\psi(\widehat{\theta}_n, \theta_c^*) \geq f(t/n)/n\}.$

## Back to V.(a) Localized change of measure: idea

Change of measure from  $\mathbb{P}_{\theta_c^*}$  to the mixture  $Q_B(\cdot) = \int_{\theta' \in B} \mathbb{P}_{\theta'}(\cdot)$  where

$$B = \Theta_{2\rho} \cap \text{Ball}(\theta_c^*, \frac{2\varepsilon_{t,i,c}}{v_\rho}).$$

- We have  $\text{Ball}(\widehat{\theta}_n, \min\{\frac{\varepsilon_{t,i,c}}{v_\rho}, \rho\}) \subset B \subset \Theta_{2\rho}$ .
- Study for  $n$  iid points the ratio

$$\begin{aligned}\frac{d\mathbb{P}_\theta}{dQ_B} &= \left[ \int_{\theta' \in B} \frac{\prod_{k=1}^n \nu_{\theta'}(X_k)}{\prod_{k=1}^n \nu_\theta(X_k)} d\theta' \right]^{-1} \\ &= \left[ \int_{\theta' \in B} \exp \left( \underbrace{n\langle \theta' - \theta, \widehat{F}_{a^*, n} \rangle - n(\psi(\theta') - \psi(\theta))}_{h(\theta')} \right) d\theta' \right]^{-1}\end{aligned}$$

## Back to V.(a) Localized change of measure: idea

Studying  $h$ , it comes for  $\theta' \in \Theta_{2\rho}$

$$\begin{aligned} h(\theta') &\geq n\mathcal{B}^\psi(\hat{\theta}_n, \theta) - \frac{n}{2}(\theta' - \hat{\theta}_n)' \nabla^2 \psi(\tilde{\theta})(\theta' - \hat{\theta}_n) \\ &\geq f(t/n) - \frac{n}{2} \|\theta' - \hat{\theta}_n\|^2 V_{2\rho} \quad (\text{convexity of } \Theta_{2\rho}) \end{aligned}$$

Thus using that  $\text{Ball}(\hat{\theta}_n, \min\{\frac{\varepsilon_{t,i,c}}{v_\rho}, \rho\}) \subset B$

$$\begin{aligned} \frac{d\mathbb{P}_\theta}{dQ_B} &\leq \left[ \int_{\theta' \in B} \exp\left(f(t/n) - \frac{n}{2} \|\theta' - \hat{\theta}_n\|^2 V_{2\rho}\right) d\theta' \right]^{-1} \\ &\leq e^{-f(t/n_{i+1})} \left[ \int_{x \in \text{Ball}(0, \min\{\rho, \frac{\varepsilon_{t,i,c}}{v_\rho}\})} \exp\left(-\frac{n_{i+1}}{2} \|x\|^2 V_{2\rho}\right) dx \right]^{-1}, \end{aligned}$$

Thus for our event of interest  $\Omega$ ,

$$\mathbb{P}_{\theta_c^*}\{\Omega\} \leq e^{-f(t/n_{i+1})} \underbrace{\left[ \int_{\text{Ball}(0, \min\{\rho, \frac{\varepsilon_{t,i,c}}{v_\rho}\})} e^{-\frac{n_{i+1}}{2}|x|^2 V_{2\rho}} dx \right]}_{\text{Volume of a Gaussian ball}}^{-1} \underbrace{\int_B \mathbb{P}_{\theta'}\{\Omega\} d\theta'}_{\leq |B|},$$

And after some easy computations

$$\mathbb{P}_{\theta_c^*}\{\Omega\} \leq 2 \exp\left(-f(t/n_{i+1})\right) \min\left\{\rho^2 v_\rho^2, \varepsilon_{t,i,c}^2, \frac{(K+2)v_\rho^2}{Kn_{i+1}V_{2\rho}}\right\}^{-K/2} 2^K \varepsilon_{t,i,c}^K.$$

We conclude using  $\varepsilon_{t,i,c} = \sqrt{\frac{2KV_{2\rho}n_{i+1}f(t/n_{i+1})}{n_i^2}}$ .  $\square$

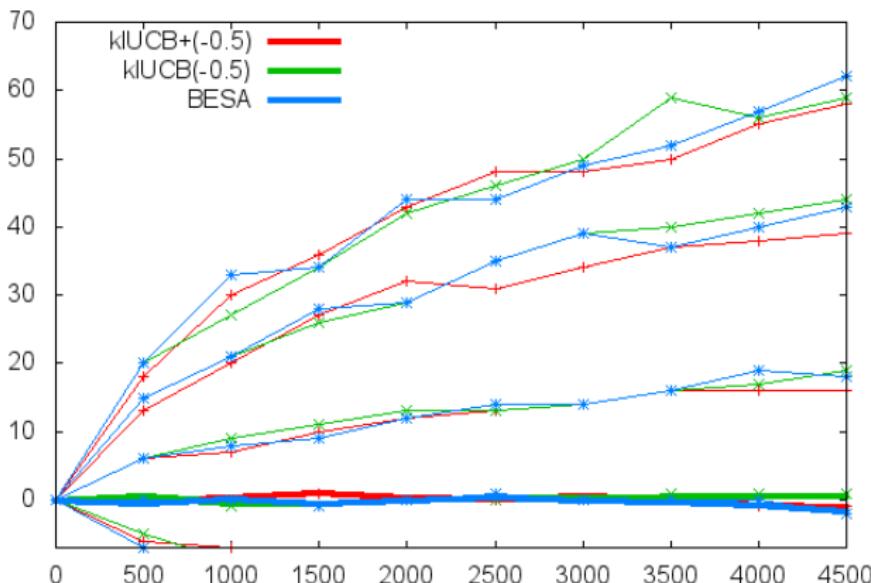
## Conclusion

For a general exponential family  $\mathcal{E}$ , known, we can\*

- ▶ Exhibit a **concentration of measure** for the Bregman divergence induced by the family:
  - ▶ in **finite time**.
  - ▶ for all **dimension  $K$** .
  - ▶ **powerful** proof technique.
- ▶ Get fine understanding of the threshold  
 $f(x) = \log(x) + \xi \log \log(x)$ 
  - ▶ improve over existing work:  $\xi > K/2 - 1$  is enough + hold for all  $K$  + KL-UCB+ as well.
  - ▶ closely follows what is observed in experiments
- ▶ ... show this was **known 30 years ago** by Lai.

... but we still need to know the family  $\mathcal{E}$ .

\* we skipped some regularity restrictions on  $\mathcal{E}$ .



Regret as function of time for KL-UCB, KL-UCB+ and BESA  
(mean and quantiles)

BESA **does not** know the family  $\mathcal{E}$ .

"Sub-sampling for the multi-armed bandit", A. Baransi et al. 2014

- ▶ Looking for Interns, PhD candidates, collaborations...

Thank you