IEEE Conference on Decision and Control 2003

Arbitrary Side Observations in Bandit Problems

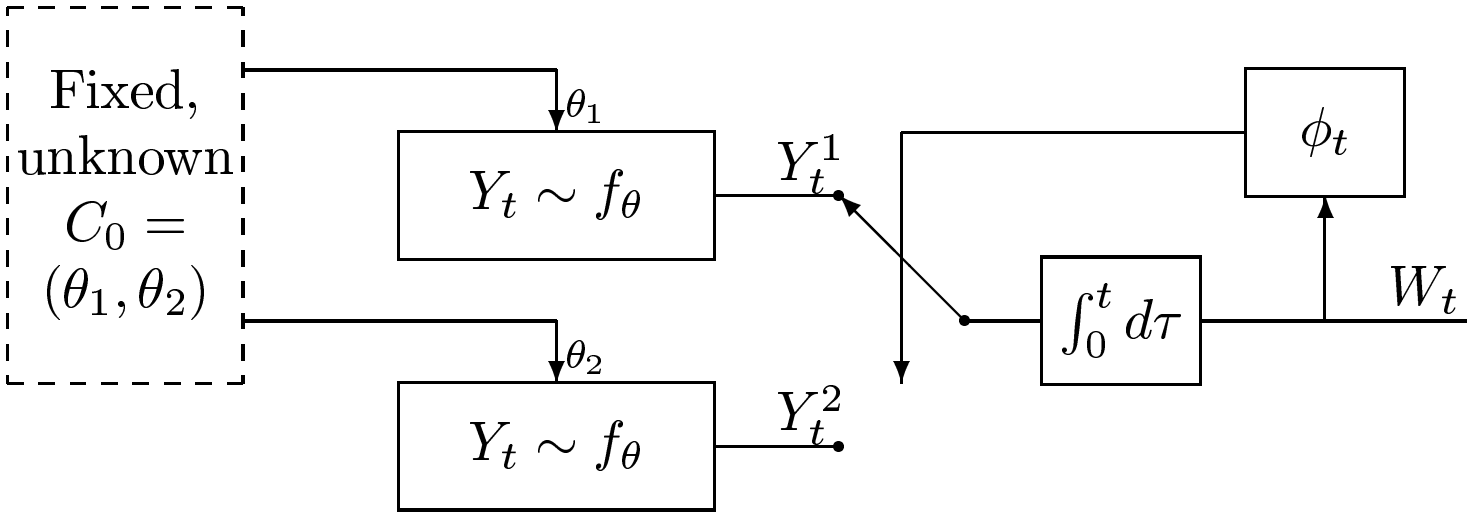
Chih-Chun Wang, Sanjeev. R. Kulkarni, H. Vincent Poor

Princeton University

December 10th, 2003

Contents

Asymptotic setting for bandit problems



- Non-Bayesian, infinite horizon setting [Robbins 52].
- Y_t^i is i.i.d., the family $\{f_\theta\}_{\theta \in \mathbf{\Theta}}$ is known.
- Only the fixed configuration pair $C_0 = (\theta_1, \theta_2)$ is unknown.

Goal:

$$\max \mathsf{E}\{W_t\}.$$

Asymptotic setting for bandit problems

Dilemma:

$$\boxed{\textbf{learning } C_0 := (\theta_1, \theta_2) \text{ vs. } \textbf{control } \phi_t = M_{C_0} := \text{argmax}\{\mu_1, \mu_2\}.}$$

Known results: $\lim_{t\to\infty} \frac{\mathsf{E}\{W_t\}}{t} = \mu^* := \max\{\mu_1, \mu_2\}.$

Question: How fast can it approach μ^*?

Asymptotic setting: uniformly good rules

For the asymptotic analysis, maximizing its convergence speed is equivalent to minimizing the growth rate of the inferior sampling time:

$$\mathsf{E}\{T_{inf}(t)\} = \mathsf{E}\left\{\sum_{\tau=1}^{t} 1_{\{\phi_\tau \neq M_{C_0}\}}\right\}.$$

Definition 1 (Uniformly good rules) [Lai and Robbins 85] *The decision rule ϕ_t is uniformly good iff for any pair of $C_0 = (\theta_1, \theta_2)$, $\mathsf{E}\{T_{inf}(t)\} = o(t^\alpha)$, $\forall \alpha > 0$.*

Question 1: Whether uniformly good rules exist?

Question 2: If they exist, how slow can the growth rate of $\mathsf{E}\{T_{inf}(t)\}$ be? (At least sub-polynomial)

$\log t$ lower bound

Theorem 1 ($\log t$ lower bound) [Lai and Robbins 85] *For any uniformly good rule, suppose $\theta_1 < \theta_2$, and μ_θ is strictly increasing with respect to θ, with some other mild conditions, we have*

$$\lim_{t\to\infty} \mathsf{P}\left(T_1(t) \geq \frac{\log t}{\inf_{\epsilon>0} I(\theta_1, \theta_2+\epsilon)}\right) = 1,$$

and by Markov inequality.

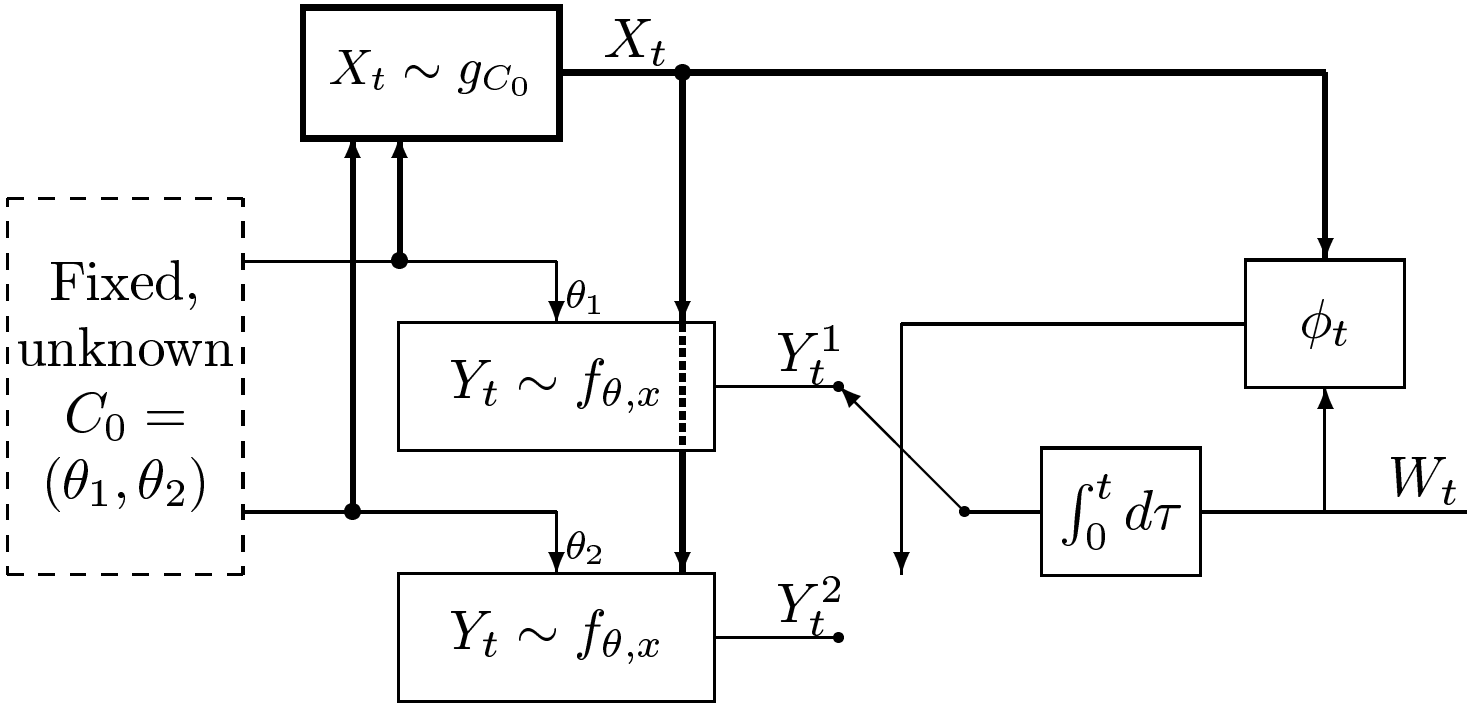
$$\liminf_{t\to\infty} \frac{\mathsf{E}\{T_1(t)\}}{\log t} \geq \frac{1}{\inf_{\epsilon>0} I(\theta_1, \theta_2+\epsilon)},$$

where $C_0 = (\theta_1, \theta_2)$, $I(\theta_1, \theta_2) = \mathsf{E}_{\theta_1}\left\{\log(\frac{dP_{\theta_1}}{dP_{\theta_2}})\right\}$.

Theorem 2 (Tightness) [Lai and Robbins 85] *$\exists \phi_t$, such that $\forall C_0$*

$$\limsup_{t\to\infty} \frac{\mathsf{E}\{T_1(t)\}}{\log t} \leq \frac{1}{\inf_{\epsilon>0} I(\theta_1, \theta_2+\epsilon)}.$$

I.I.D. Side Observations



- Y_t^i is independently distributed according to $f_{\theta,x}$.
- X_t is independently distributed according to g_{C_0}.

Example: X_t is the GPS, ϕ_t chooses between two modulation schemes, and Y_t^i is whether the current transmission is successful.

I.I.D. Side Observations

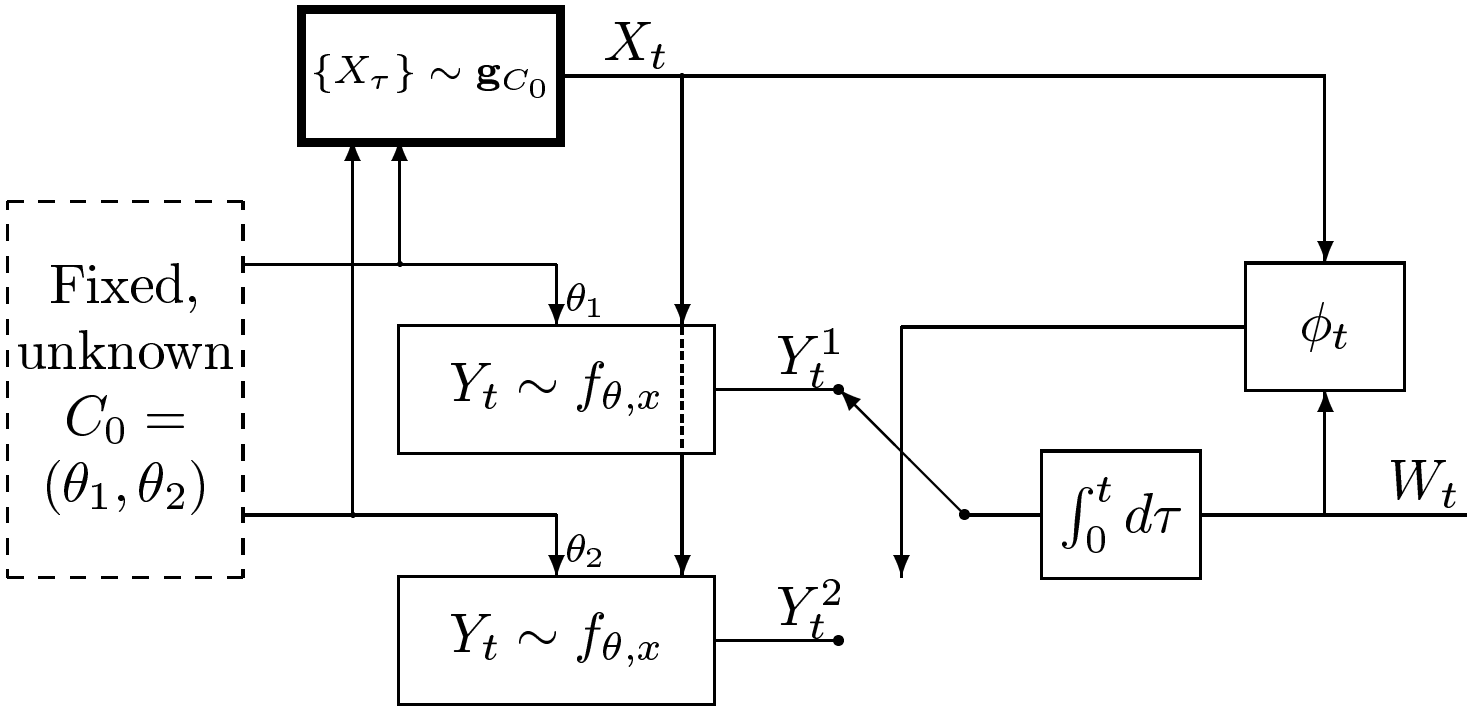
Goal: ϕ_t chooses $M_{C_0}(x) := \text{argmax}\{\mu_1(X_t), \mu_2(X_t)\}$ as *fast* as possible.

For any finite $\mathbf{X}$, it can be viewed as we are having several two-armed bandit machines, with common pair (θ_1, θ_2). However, the access opportunities are determined by a random sequence $\{X_\tau\}$.

Previous work:

- [Woodroofe 79]: $Y = \mathcal{N} + \theta + X$. Conclusion: if the cdf $G(x) < 1$, $\forall x$, then a myopic approach is asymptotically optimal.
- [Sarkar 91]: exponential families. [Kulkarni 93]: learnable distribution class. [Zoubeidi 94]: the concept of "indices".
- [Wang02]: Both θ and X_t take values in **finite sets**. Four different cases are discussed.

Arbitrary Side Observations

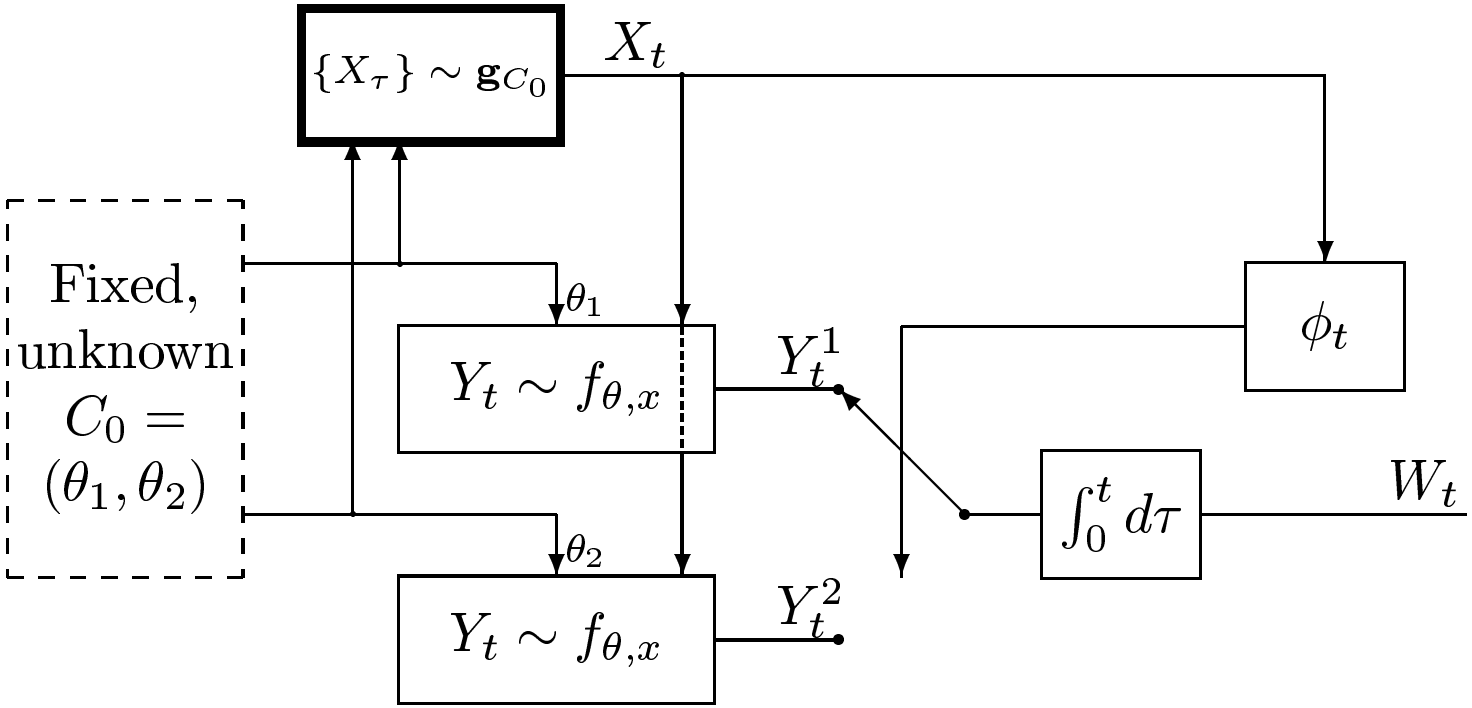


In this paper, we generalized the results of four different cases in [Wang02] to arbitrarily side information sequence taking values in **finite sets**, including

- I.i.d. sequences,
- Markov chains,
- Deterministic periodic sequences.

Case 1: Direct Information

(i) $g_{C'} = g_{C''}$ iff $C' = (\theta_1', \theta_2') = C''(\theta_1'', \theta_2'')$.
(ii) if $\hat{C}$ is close to C_0, $\forall x, M_{\hat{C}}(x) = M_{C_0}(x)$.



Case 1: Direct Information

Theorem 3 (Previous Result) *Consider i.i.d.* $\{X_\tau\}$ *in Case 1. There exists* ϕ_t*, such that* $\mathsf{E}\{T_{inf}(t)\} < \infty, \forall C_0$.

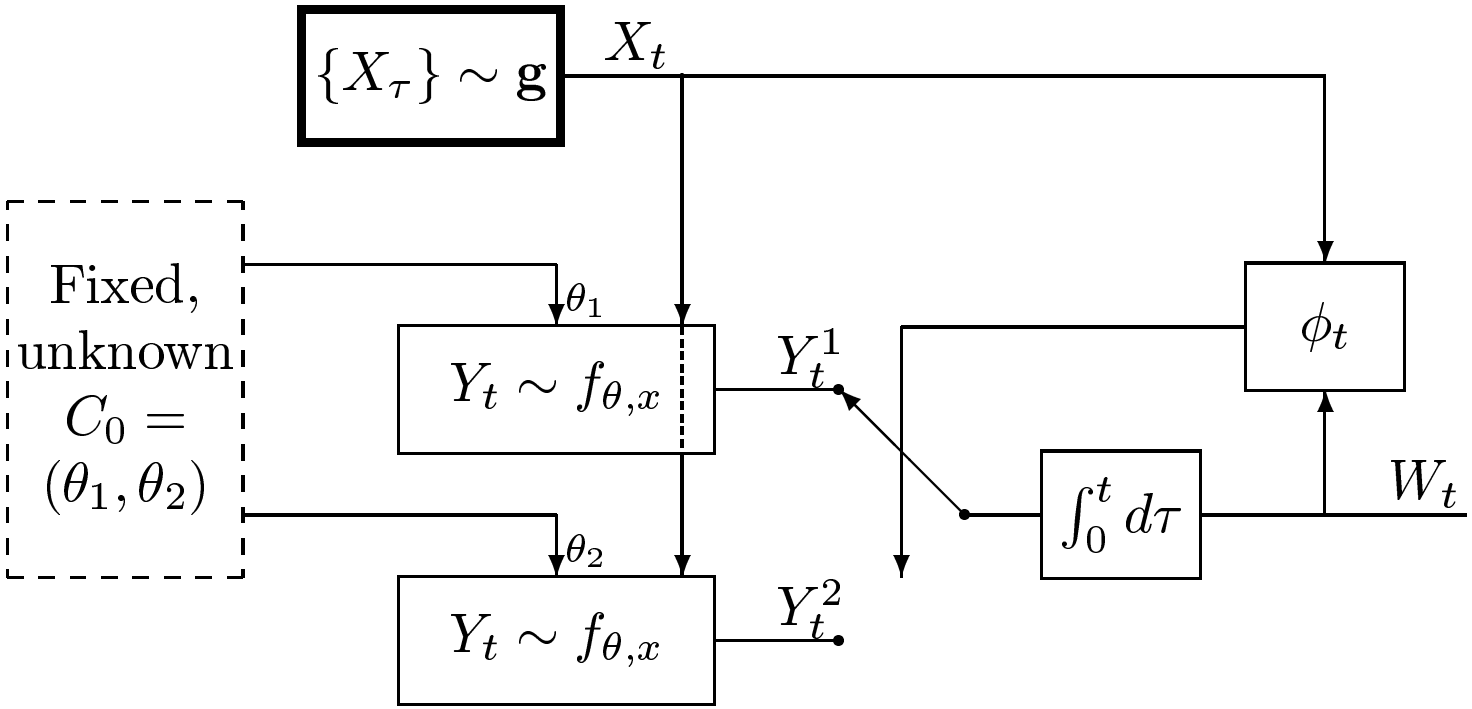
Theorem 4 (New Result) *Suppose* $M_{\hat{C}_t}(x) = M_{C_0}(x)$ *for all* x*. We have a decision rule* ϕ_t *and* $\epsilon > 0$ *such that*

$$\lim_{t\to\infty} \frac{\mathsf{E}\{T_{inf}(t)\}}{1 + \sum_{\tau=1}^{t} \mathsf{P}(|\hat{C}_\tau - C_0| > \epsilon)} \leq 1.$$

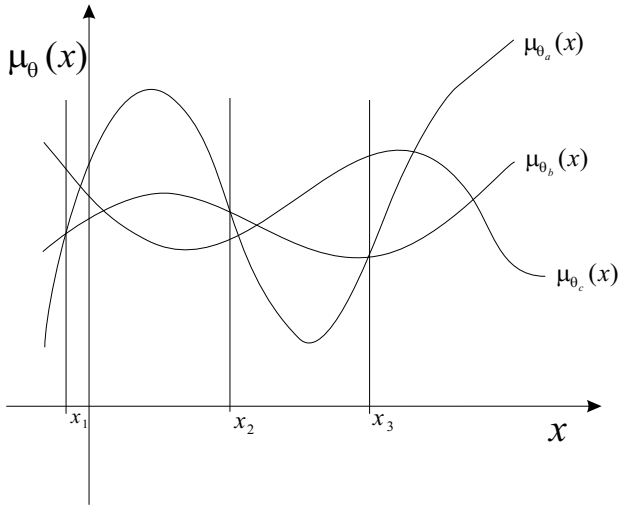
Special cases:

- i.i.d. sequences.
- Markov chains.
- deterministic periodic sequences.

$\{X_\tau\}$ Does Not Reveal $C_0 = (\theta_1, \theta_2)$



Case 2: Best Arm Depends on X_t



(i) $g_{C_0} = g$ does not depend on the unknown C_0.
(ii) For any C_0, there exist x_1 and x_2 such that $M_{C_0}(x_1) = 1$ and $M_{C_0}(x_2) = 2$.

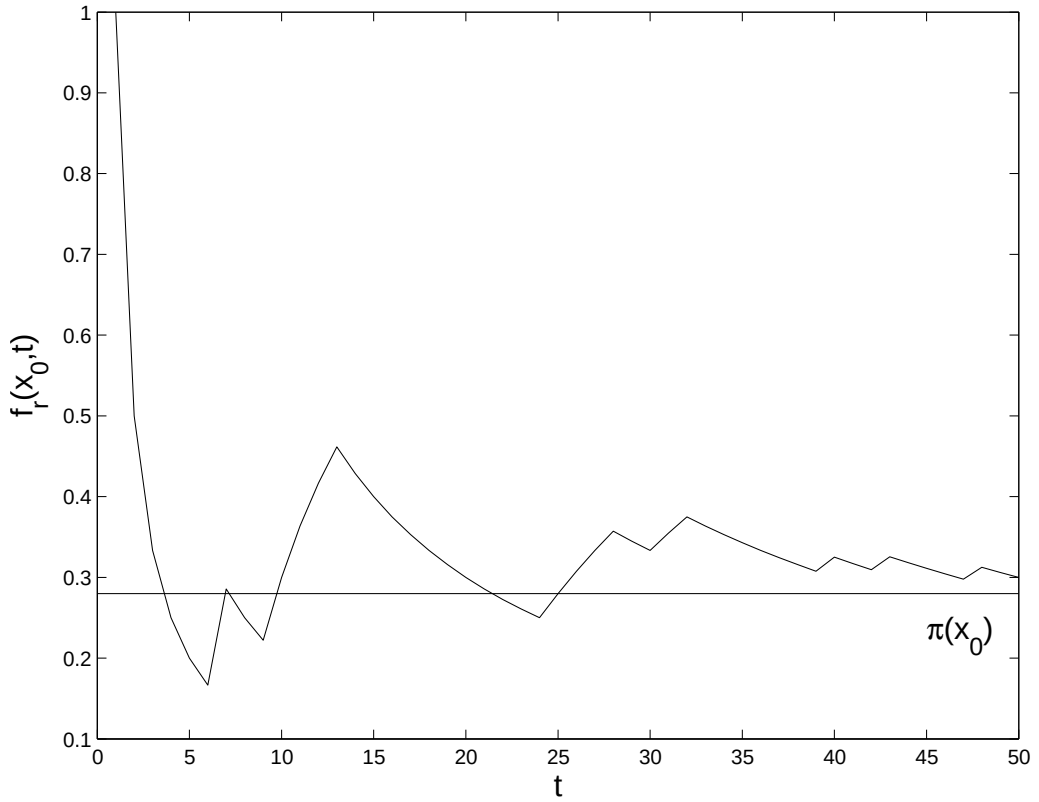
Theorem 5 (Previous Result) *Consider i.i.d.* $\{X_\tau\}$ *in Case 2. There exists* ϕ*, such that* $\mathsf{E}\{T_{\inf}(t)\} < \infty, \forall C_0$.

Intuition: For any $\hat{C}$, the even appearances of x_1 and x_2 result that the myopic decision $\phi_t = M_{\hat{C}}$ evenly samples both arms often enough. The dilemma does not exist and then we have $\mathsf{E}\{T_{inf}(t)\} < \infty$.

Evenly Distributed in Probability Series

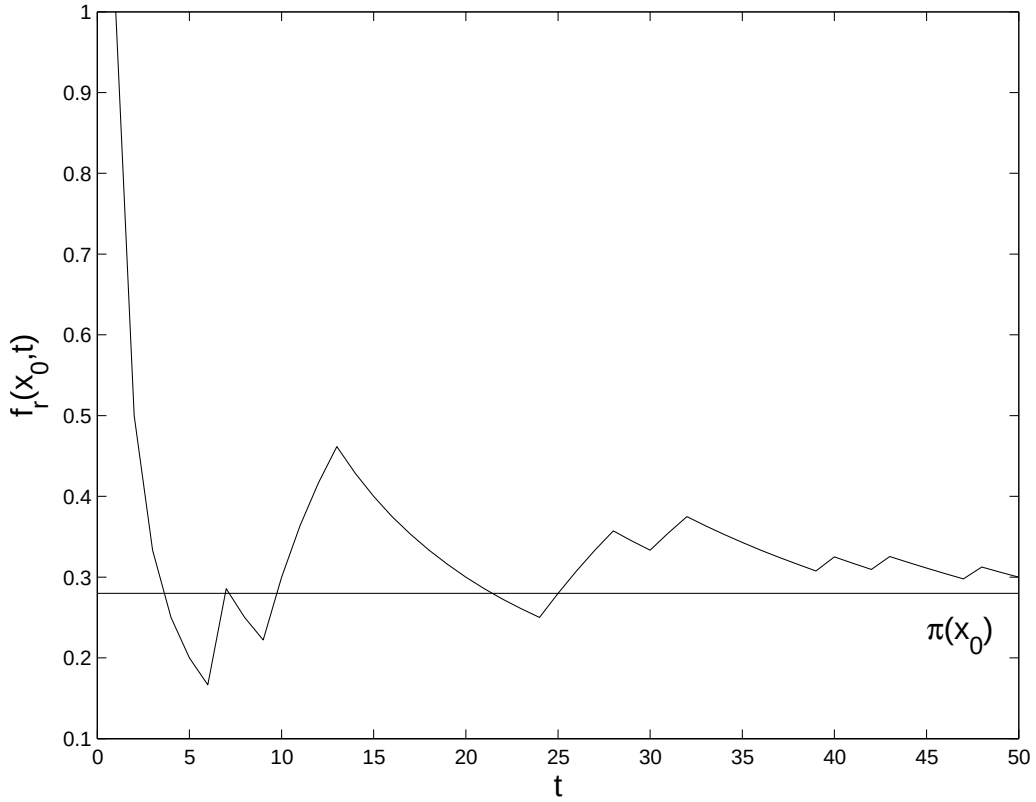
Definition 2 *$\{X_\tau\}$ is evenly distributed "in probability series" if there exists a strictly positive mapping $\pi(x) > 0$, such that for any x_0,*

$$E\left\{\sum_{t=1}^{\infty} 1\{f_r(x_0,t) < \pi(x_0)\}\right\} < \infty.$$

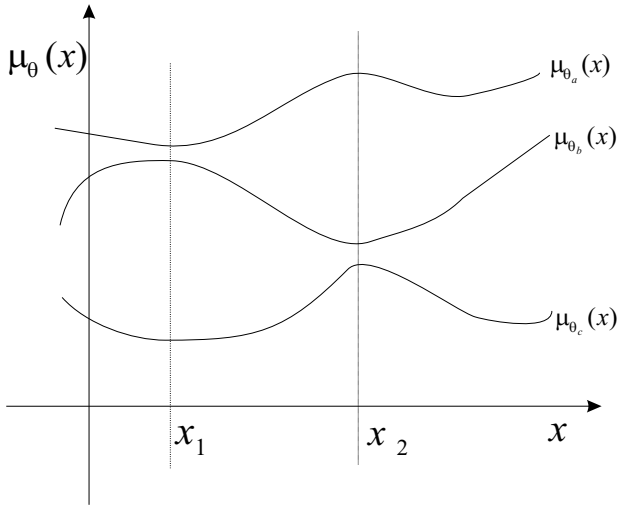


Arbitrary $\{X_\tau\}$ in Case 2

Theorem 6 (New Result) *Theorem 5 holds for all $\{X_\tau\}$ being evenly distributed in probability series. I.e. we have finite $\mathsf{E}\{T_{inf}(t)\}$ if $\{X_\tau\}$ is evenly distributed.*



Case 3: Best Arm Doesn't Depend on X_t



(i) $g_{C_0} = g$, which is independent of C_0.
(ii) It is either $M_{C_0}(x) = 1, \forall x$ or $M_{C_0}(x) = 2, \forall x$, which case is active depends on the unknown C_0.

For i.i.d. X_t, the $\log t$ lower bound and the tightness theorem still hold, by replacing

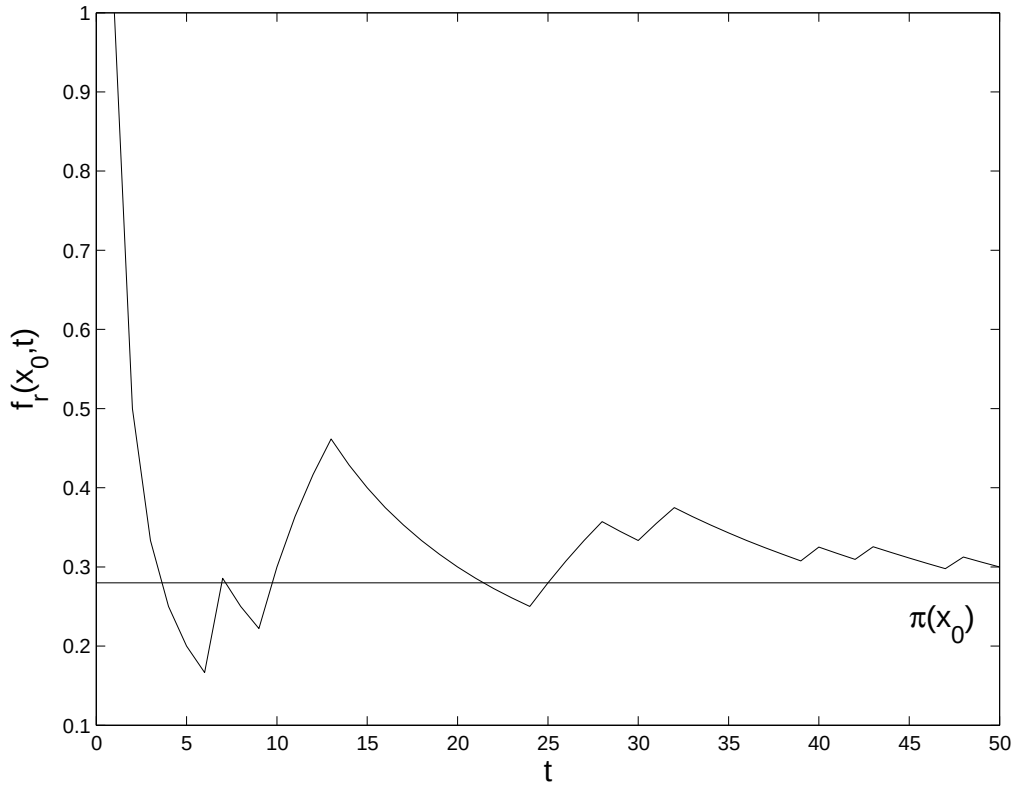
$$\frac{\log t}{I(\theta_1, \theta_2 + \epsilon)} \longrightarrow \frac{\log t}{\max_x I(\theta_1, \theta_2 + \epsilon | x)}.$$

In essence, it is like a **two-player-zero-sum game**. The decision maker postpones the forced sampling (learning) to the most informative sub-bandit.

Arbitrary $\{X_\tau\}$ in Case 3

The $\log t$ lower bound:

For all $\{X_\tau\}$ being *evenly distributed in probability series*, the $\log t$ lower bound on the inferior sampling time still holds.

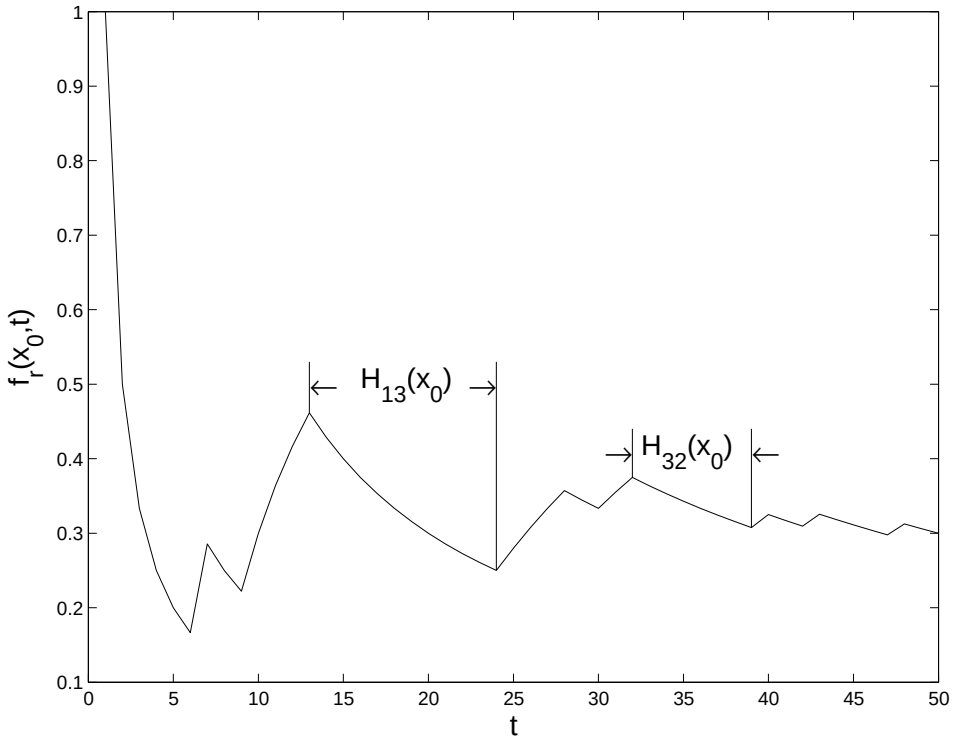


Uniformly Strongly Evenly (u.s.e.) Distributed in L^1

Definition 3 *$\{X_\tau\}$ is u.s.e. distributed in L^1, if for any stopping time T, the conditional expectation of the first hitting time of x after T, has a global upper bound, i.e., $\exists d > 0$ such that*

$$E\{H_T(x)|T\} \leq d < \infty, \forall T, x,$$

where $H_T(x) := \inf\{l > 0 | X_{T+l} = x\}$.



Arbitrary $\{X_\tau\}$ in Case 3

Definition 4 (Value of the Game Condition)

$$\inf_\theta \sup_x \{I(\theta_1, \theta|x)\} = \sup_x \inf_\theta \{I(\theta_1, \theta|x)\}.$$

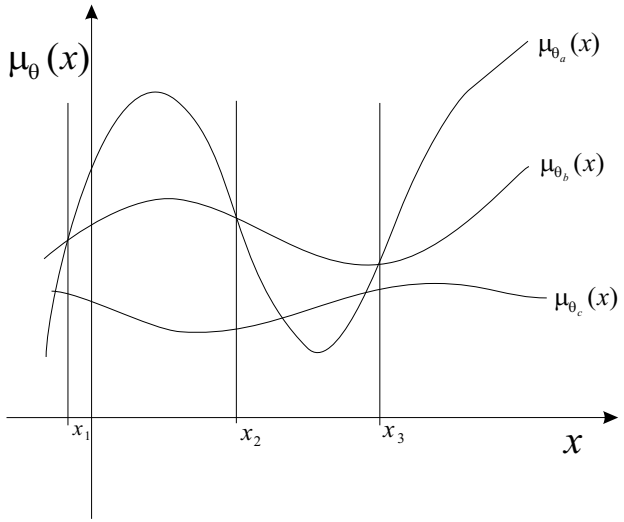
Tightness of the $\log t$ lower bound

If

- $\{X_\tau\}$ is uniformly strongly distributed in L^1,
- The value of the game exists,

then we can find a scheme achieving the $\log t$ lower bound. Thus the $\log t$ lower bound is tight.

Case 4: Mixed Case



(i) $g_{C_0} = g$, which is independent of C_0.
(ii) For some C_0, $M_{C_0}(x)$ is *constant* for all x, for other C_0, $M_{C_0}(x)$ *varies* with respect to x.

For i.i.d. $\{X_\tau\}$, if $M_{C_0}(x)$ is constant for all x, the $\log t$ lower bound still holds, by replacing

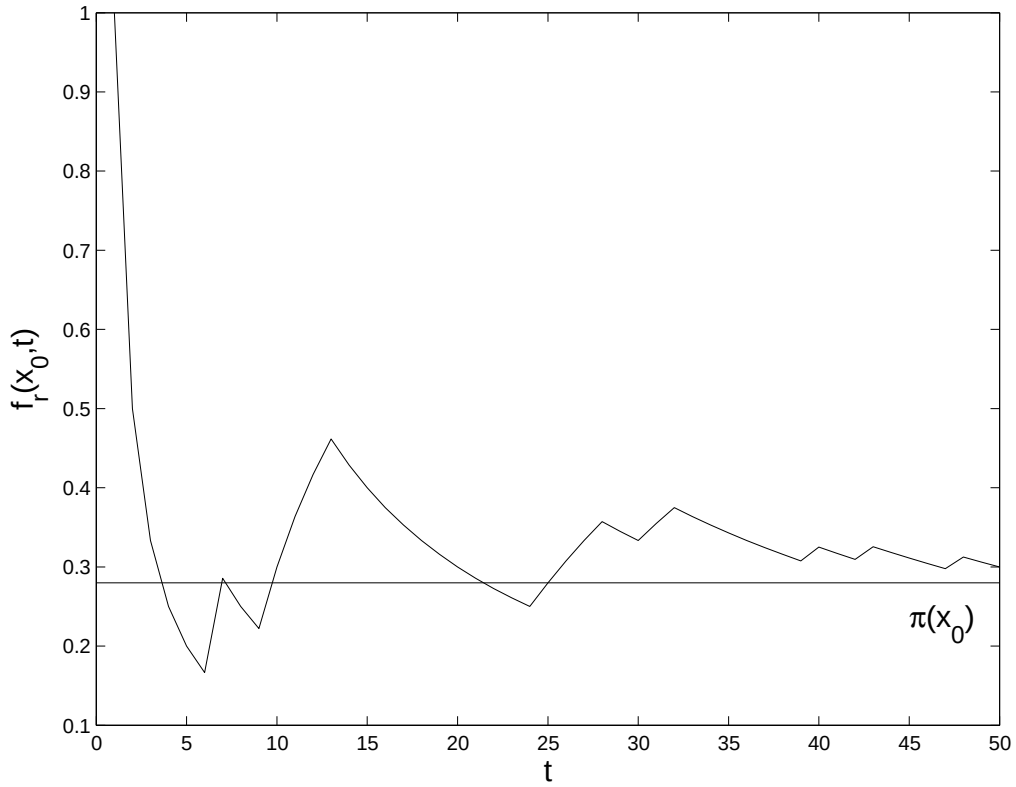
$$\frac{\log t}{I(\theta_1, \theta_2 + \epsilon)} \longrightarrow \frac{\log t}{\inf_{\{\theta:\ \exists x, \mu_\theta(x) > \mu_{\theta_2}(x)\}} \sup_x \{I(\theta_1, \theta | x)\}}.$$

We also found an adaptive scheme which achieves the modified $\log t$ lower bound if $M_{C_0}(x)$ is constant, and has bounded $\mathsf{E}\{T_{inf}(t)\}$ if $M_{C_0}(x)$ varies with respect to x.

Arbitrary $\{X_\tau\}$ in Case 4

The $\log t$ lower bound:

For all $\{X_\tau\}$ being *evenly distributed in probability series*, the $\log t$ lower bound on the inferior sampling time still holds.



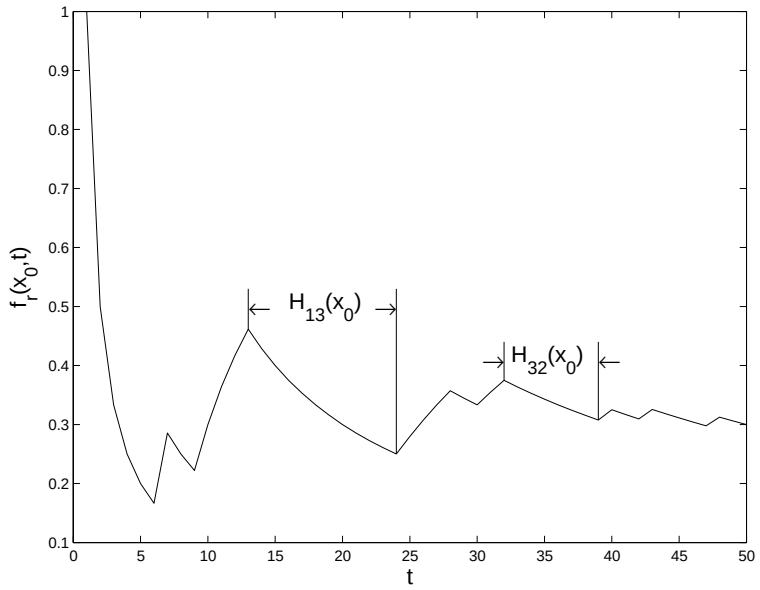
Arbitrary $\{X_\tau\}$ in Case 4

Tightness of the $\log t$ lower bound

If

- $\{X_\tau\}$ is uniformly strongly distributed in L^1,
- The value of the game exists,

then we can find a scheme achieving the $\log t$ lower bound in one case and bounded expected inferior sampling time in the other case.



Conclusions

- Essence of bandit problems: conflicts of learning and control.
- Four cases:
  - Direct Information: Bounded $\mathsf{E}\{T_{inf}(t)\}$.
  - Best Arm Depends on X_t: Bounded $\mathsf{E}\{T_{inf}(t)\}$.
  - Best Arm Doesn't Depends on X_t: $\mathsf{E}\{T_{inf}(t)\}$ is still lower bounded by $\log t$, but with smaller constant.
  - Mixed Case: Bounded $\mathsf{E}\{T_{inf}(t)\}$ or $\mathsf{E}\{T_{inf}(t)\}$ is lower bounded by $\log t$, depending on $C_0 = (\theta_1, \theta_2)$.
- Using the side information as indices of different sub-bandit machines.
- **Evenly distributed appearances** are the key property for beneficial side information, not the **randomness** of i.i.d. sequence.
  - $f_r(x,t)$: the bounded expected duration of small relative frequency,
  - $H_T(x)$: the bounded hitting time of the next x.