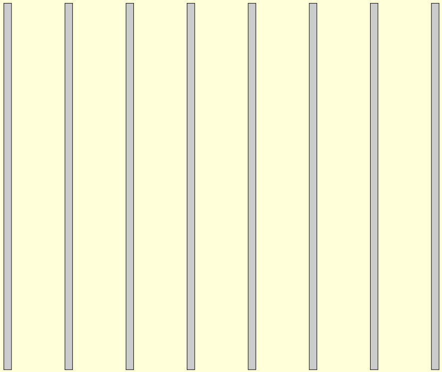
PAC Subset Selection in Stochastic Multi-armed Bandits

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone

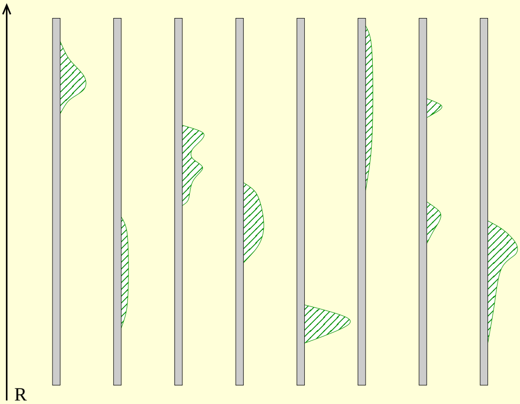
Yahoo! Labs Bangalore
Department of Computer Science, The University of Texas at Austin
Chair for Information Technology, University of Leoben

July 2012

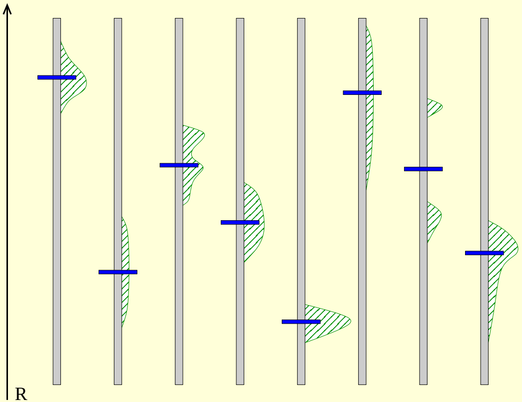
Stochastic Bandits and Subset Selection



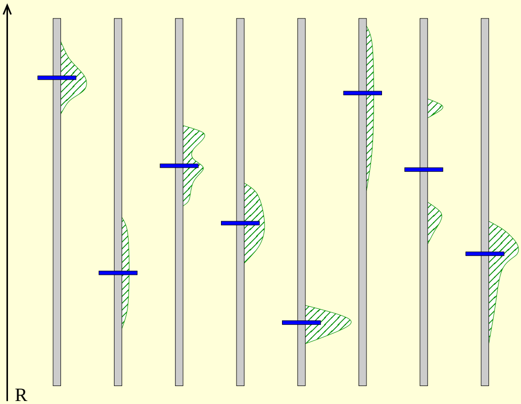
Stochastic Bandits and Subset Selection



Stochastic Bandits and Subset Selection



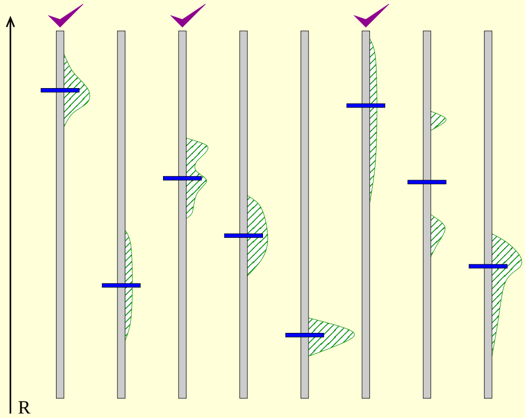
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

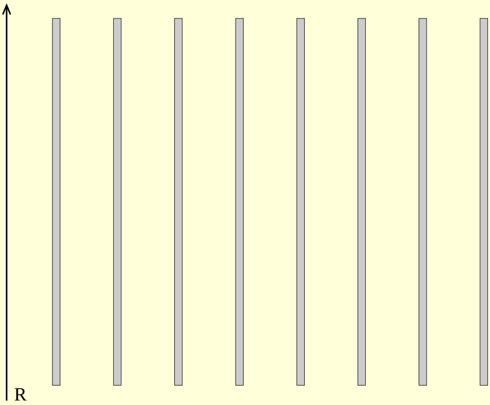
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

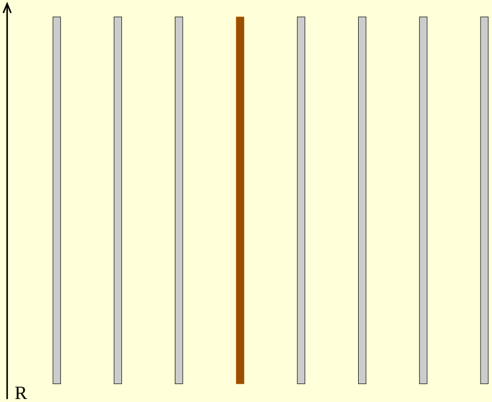
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

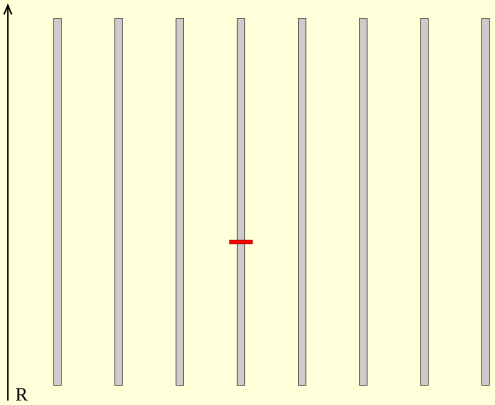
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

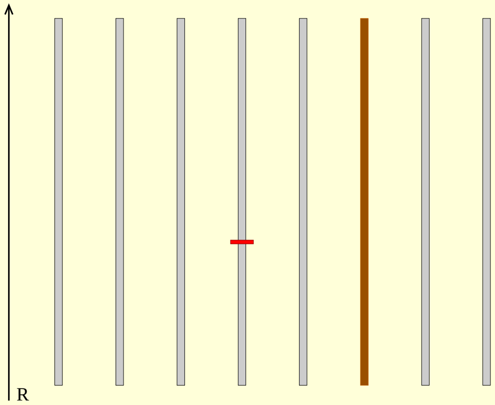
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

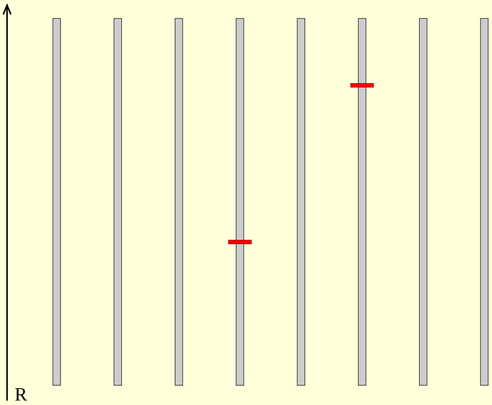
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

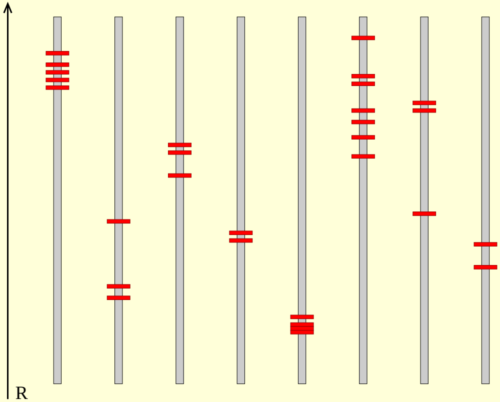
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

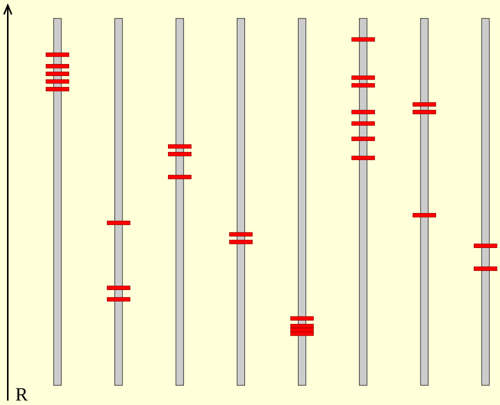
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

Stochastic Bandits and Subset Selection

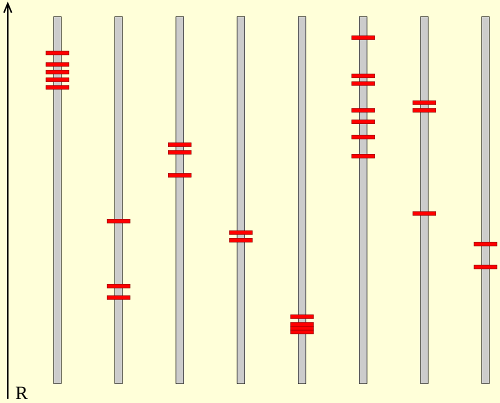


In an *n*-armed bandit:

find the *m* arms with the highest means

for a given confidence

Stochastic Bandits and Subset Selection



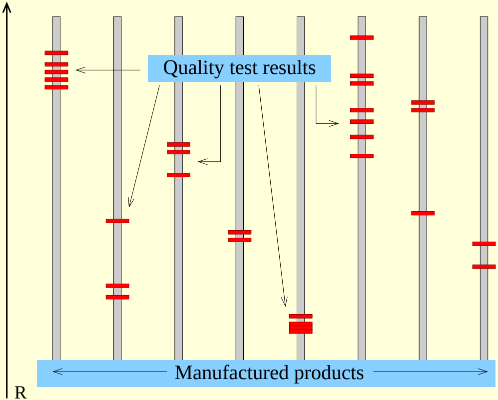
In an *n*-armed bandit:

find the *m* arms with the highest means

for a given confidence

using a minimal number of samples.

Stochastic Bandits and Subset Selection



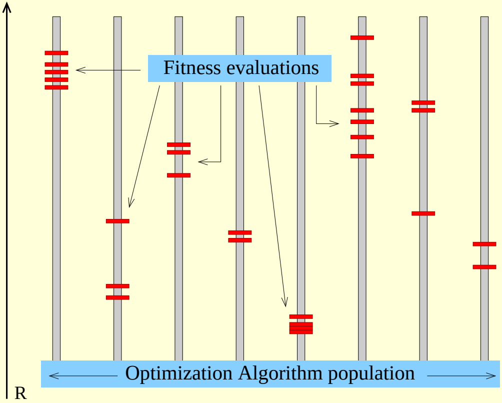
In an *n*-armed bandit:

find the *m* arms with the highest means

for a given confidence

using a minimal number of samples.

Stochastic Bandits and Subset Selection



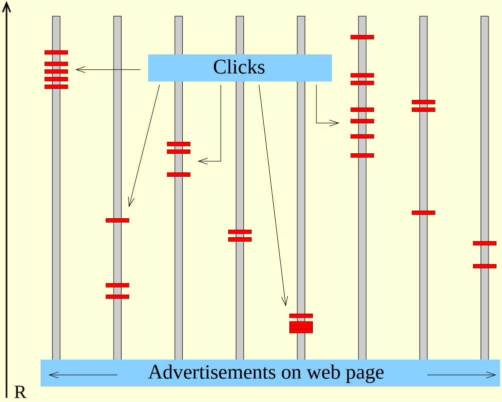
In an *n*-armed bandit:

find the *m* arms with the highest means

for a given confidence

using a minimal number of samples.

Stochastic Bandits and Subset Selection



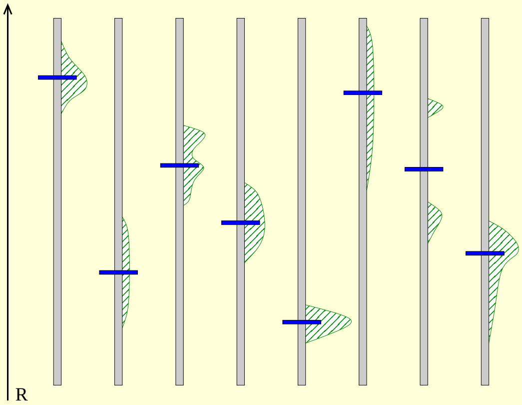
In an *n*-armed bandit:

find the *m* arms with the highest means

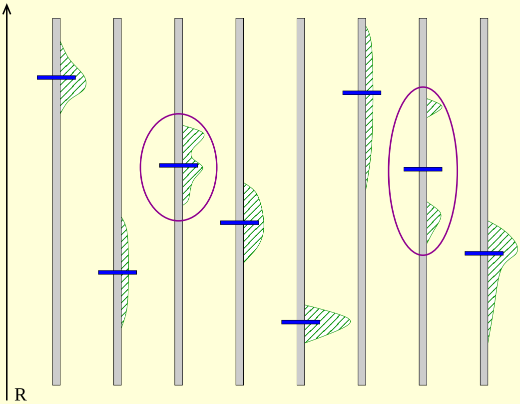
for a given confidence

using a minimal number of samples.

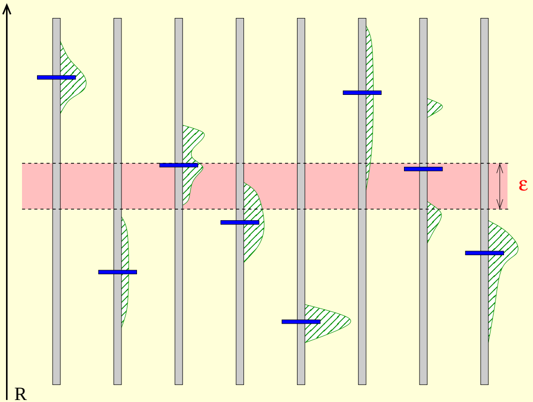
PAC Formulation



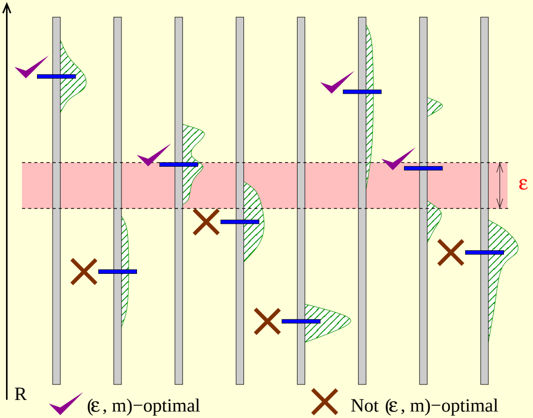
PAC Formulation



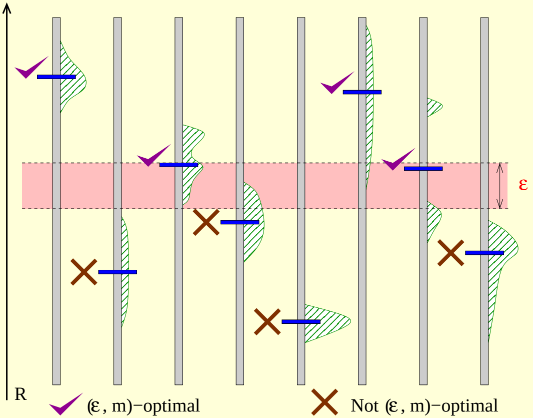
PAC Formulation



PAC Formulation



PAC Formulation



In an *n*-armed bandit:

find m (ϵ, m)-optimal arms

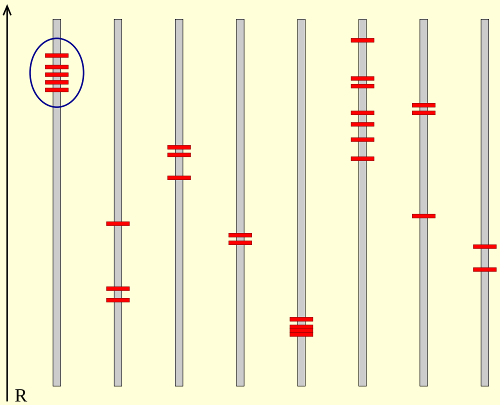
with probability at least $1 - \delta$

using a minimal number of samples.

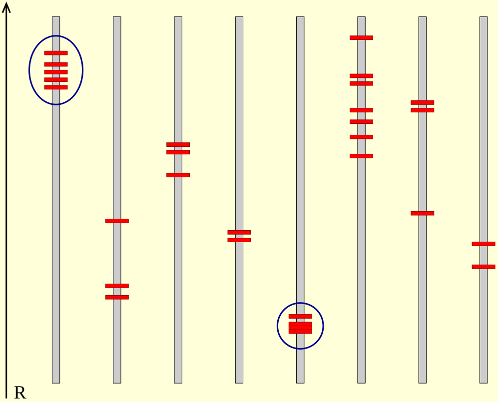
Bandit Variations

- PAC vs. *Regret* setting.
- Independent vs. *Dependent* arms.
- Stochastic vs. *Adversarial* rewards.

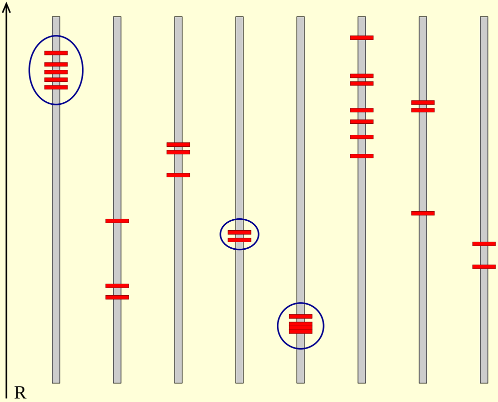
Confidence Bounds on the Mean



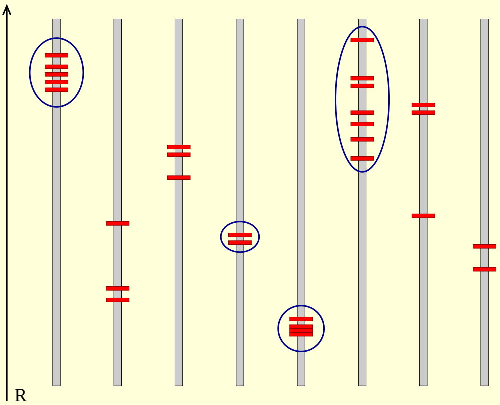
Confidence Bounds on the Mean



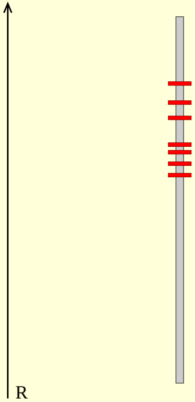
Confidence Bounds on the Mean



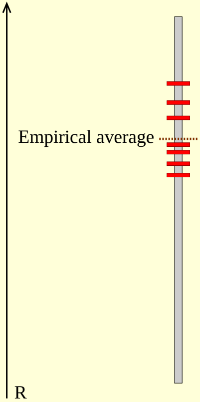
Confidence Bounds on the Mean



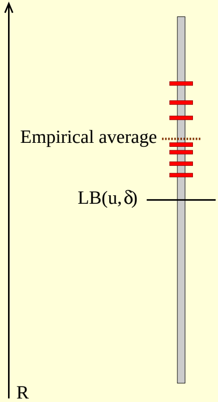
Confidence Bounds on the Mean



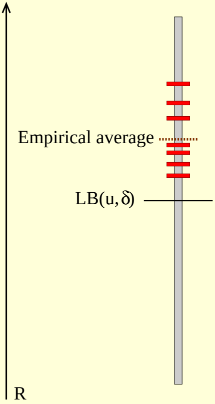
Confidence Bounds on the Mean



Confidence Bounds on the Mean



Confidence Bounds on the Mean



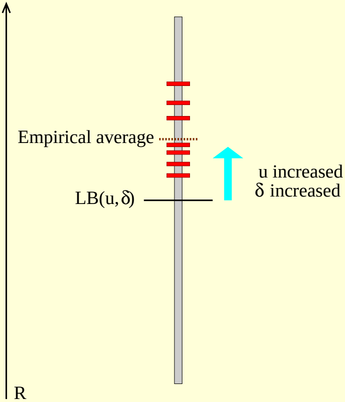
Hoeffding's inequality: With probability at least $1 - \delta$,

$$\text{True mean} \geq \text{Empirical average} - R\sqrt{\frac{1}{2u} log(\frac{1}{\delta})} = \text{LB}(u, \delta).$$

Empirical Bernstein bound: With probability at least $1 - \delta$,

$$\text{True mean} \geq \text{Empirical average} - (\sqrt{\frac{\sigma^2 log(\frac{3}{\delta})}{2u}} + \frac{3R log(\frac{3}{\delta})}{2u}) = \text{LB}(u, \sigma^2, \delta).$$

Confidence Bounds on the Mean



Hoeffding's inequality: With probability at least $1 - \delta$,

$$\text{True mean} \geq \text{Empirical average} - R\sqrt{\frac{1}{2u} log(\frac{1}{\delta})} = \text{LB}(u, \delta).$$

Empirical Bernstein bound: With probability at least $1 - \delta$,

$$\text{True mean} \geq \text{Empirical average} - (\sqrt{\frac{\sigma^2 log(\frac{3}{\delta})}{2u}} + \frac{3Rlog(\frac{3}{\delta})}{2u}) = \text{LB}(u, \sigma^2, \delta).$$

Algorithms for Subset Selection

- DIRECT Algorithm:

  Sample each arm $\lceil \frac{2}{\epsilon^2} log \frac{n}{\delta} \rceil$ times.
  Return *m* arms with highest *empirical* averages.

  - Achieves PAC guarantee.
  - Sample complexity: $O(\frac{n}{\epsilon^2} log(\frac{n}{\delta}))$.

Algorithms for Subset Selection

- DIRECT Algorithm:

  Sample each arm $\lceil \frac{2}{\epsilon^2} log \frac{n}{\delta} \rceil$ times.
  Return m arms with highest *empirical* averages.

  - Achieves PAC guarantee.
  - Sample complexity: $O(\frac{n}{\epsilon^2} log(\frac{n}{\delta}))$.

- HALVING Algorithm:

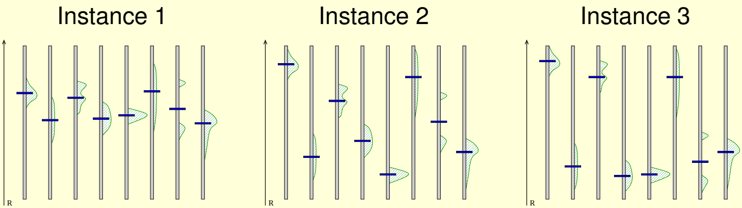
  Sample each arm $u_1(m, \epsilon, \delta)$ times.
  Discard half the arms with lower empirical averages.
  Sample each remaining arm $u_2(m, \epsilon, \delta)$ times.
  Discard half the remaining arms with lower empirical averages.

  ⋮

  Until m arms remain.

  - Achieves PAC guarantee.
  - Sequence (u_i) such that total number of samples is $O(\frac{n}{\epsilon^2} log(\frac{m}{\delta}))$.
  - Optimal up to a constant factor.

Problem Complexity



$$\Delta_a \stackrel{\text{def}}{=} \begin{cases} p_a - p_{m+1} & \text{if } 1 \leq a \leq m, \\ p_m - p_a & \text{if } m+1 \leq a \leq n. \end{cases}$$

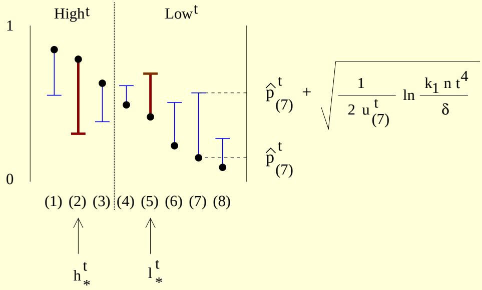
$$\mathbf{H}^{\epsilon/2} = \sum_{a=1}^{n} \frac{1}{\max\{\Delta_a, \frac{\epsilon}{2}\}^2}.$$

Algorithms for Subset Selection (contd.)

- LUCB Algorithm:

  Achieves PAC guarantee.

  Expected sample complexity of $O\left(H^{\epsilon/2} \log\left(\frac{H^{\epsilon/2}}{\delta}\right)\right)$.



Stopping rule: Terminate iff

$$\left(\hat{p}^t_{l^t_*} + \beta(u^t_{l^t_*}, t)\right) - \left(\hat{p}^t_{h^t_*} - \beta(u^t_{h^t_*}, t)\right) < \epsilon.$$

Sampling strategy:

On round t: sample arms h^t_* and l^t_*.

Thank you!