PAC Subset Selection in Stochastic Multi-armed Bandits

Shivaram Kalyanakrishnan
shivaram@yahoo-inc.com

Yahoo Labs Bangalore

January 2014

Today's Talk

Relevant publications

Efficient Selection of Multiple Bandit Arms: Theory and Practice
Shivaram Kalyanakrishnan and Peter Stone, *ICML 2010*.

PAC Subset Selection in Stochastic Multi-armed Bandits
Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone, *ICML 2012*.

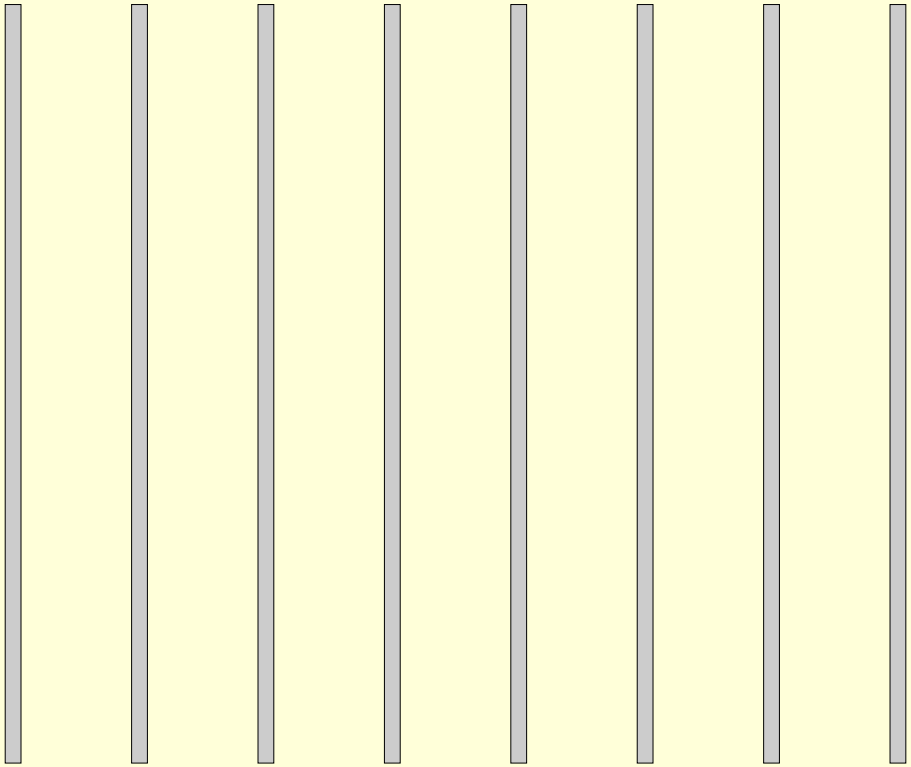
Information Complexity in Bandit Subset Selection
Emilie Kaufmann and Shivaram Kalyanakrishnan, *COLT 2013*.

Outline

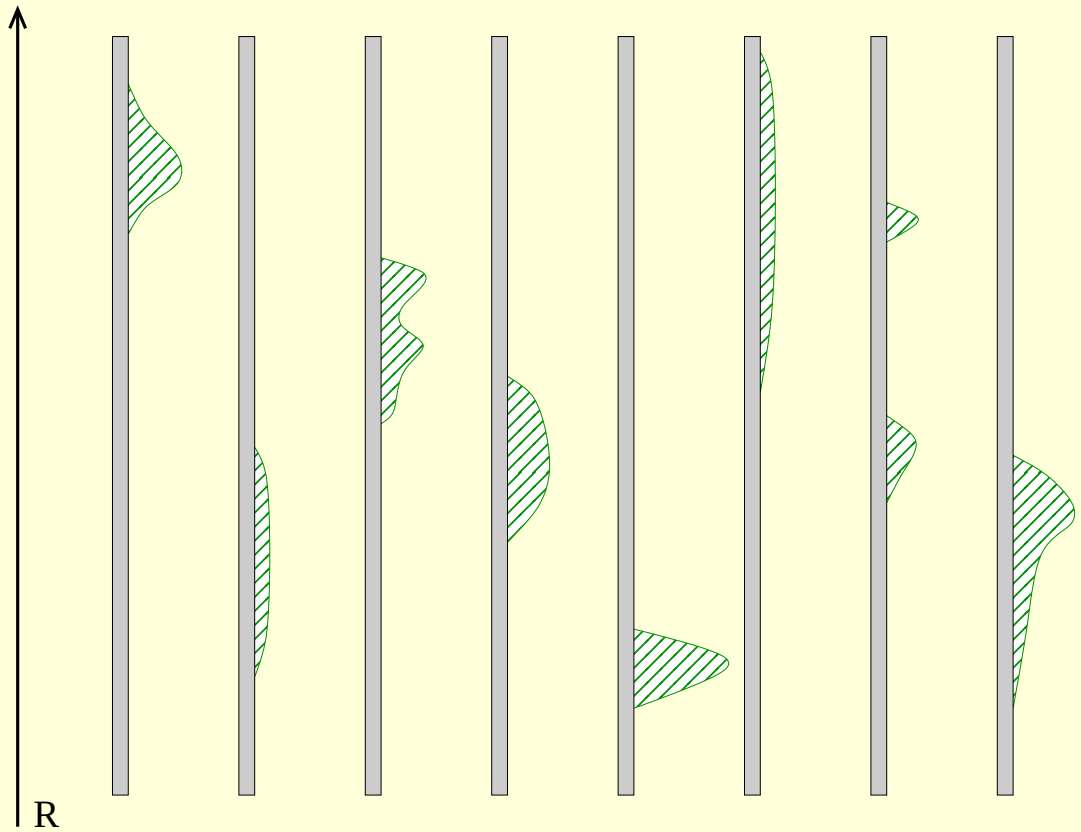
1. Subset selection: PAC formulation
2. Related work
3. Confidence bounds
4. Algorithms and sample-complexity bounds
5. Experiments
6. Future work

2:15

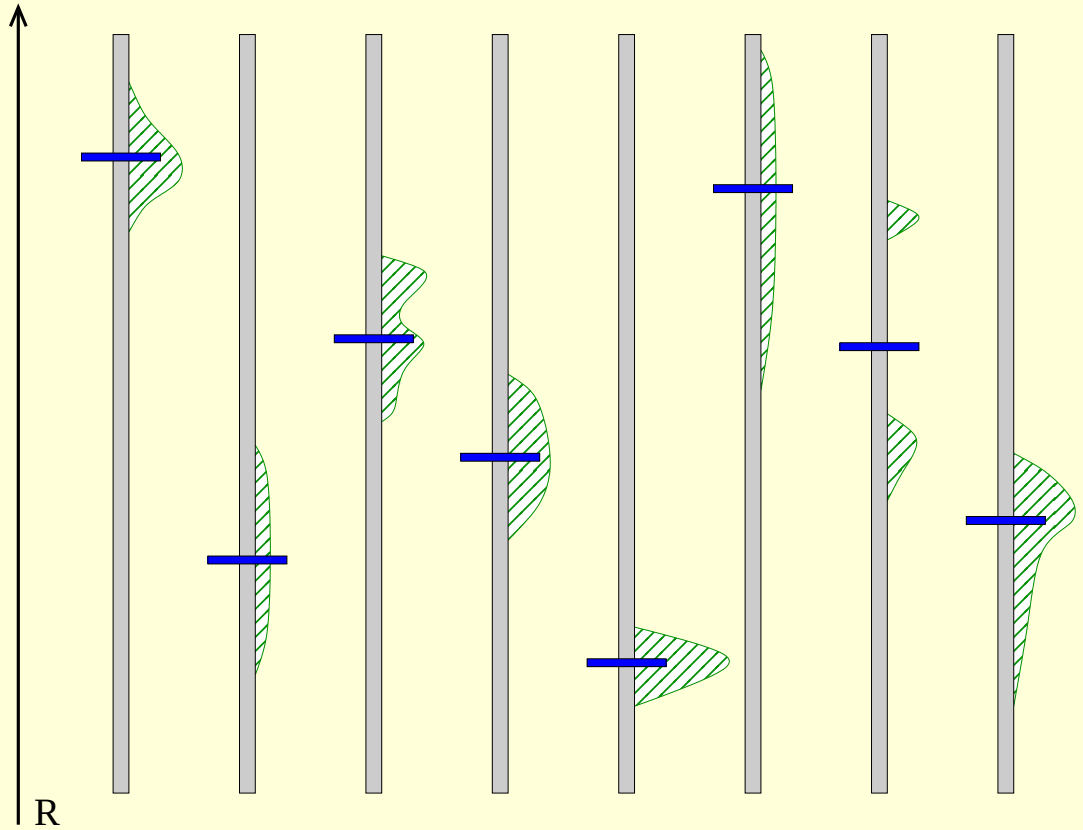
Stochastic Bandits and Subset Selection



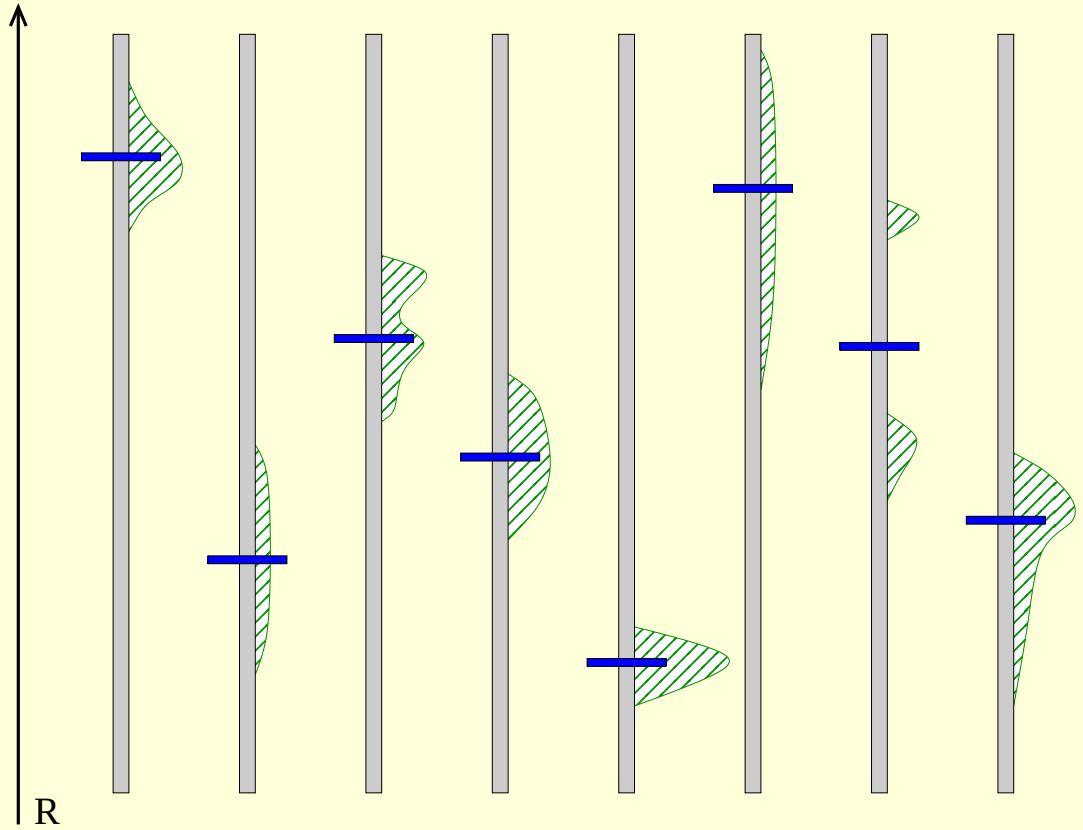
Stochastic Bandits and Subset Selection



Stochastic Bandits and Subset Selection



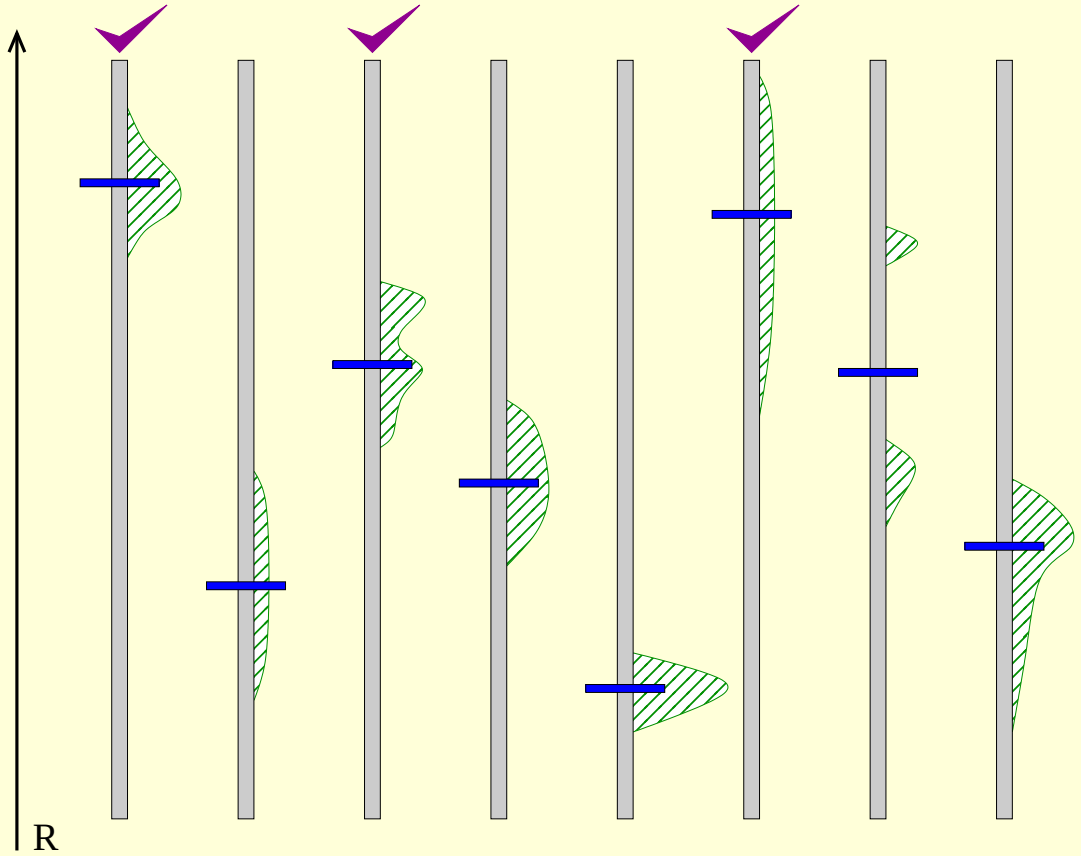
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

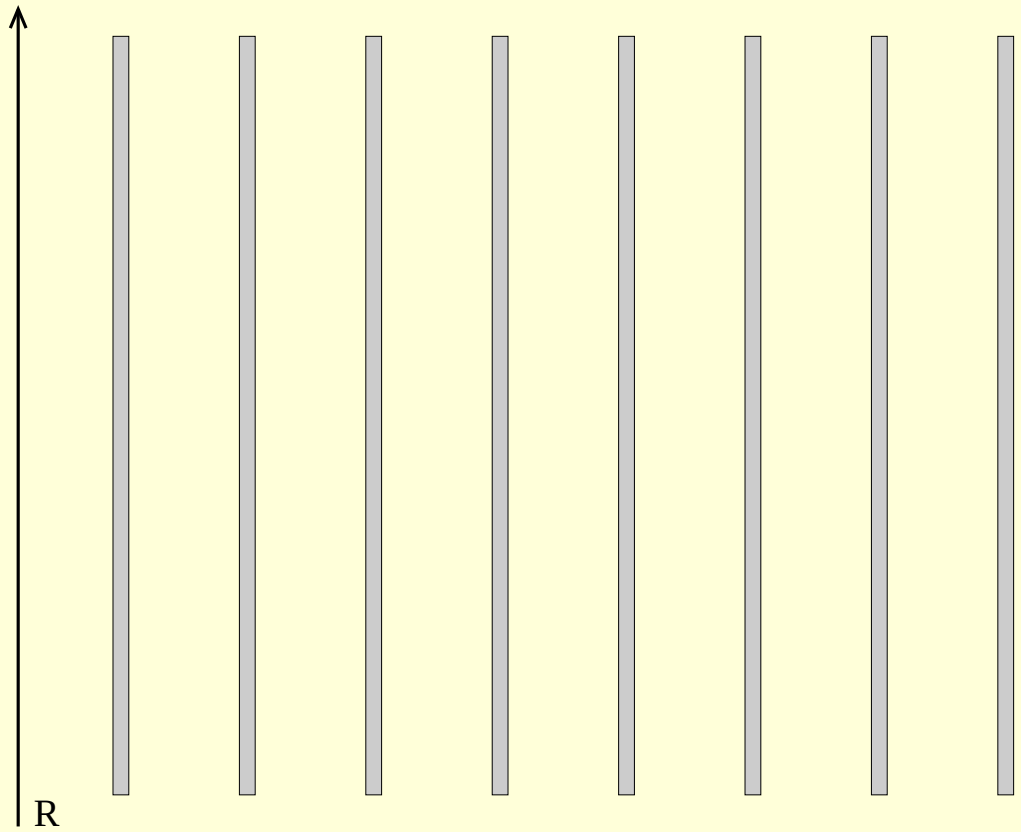
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

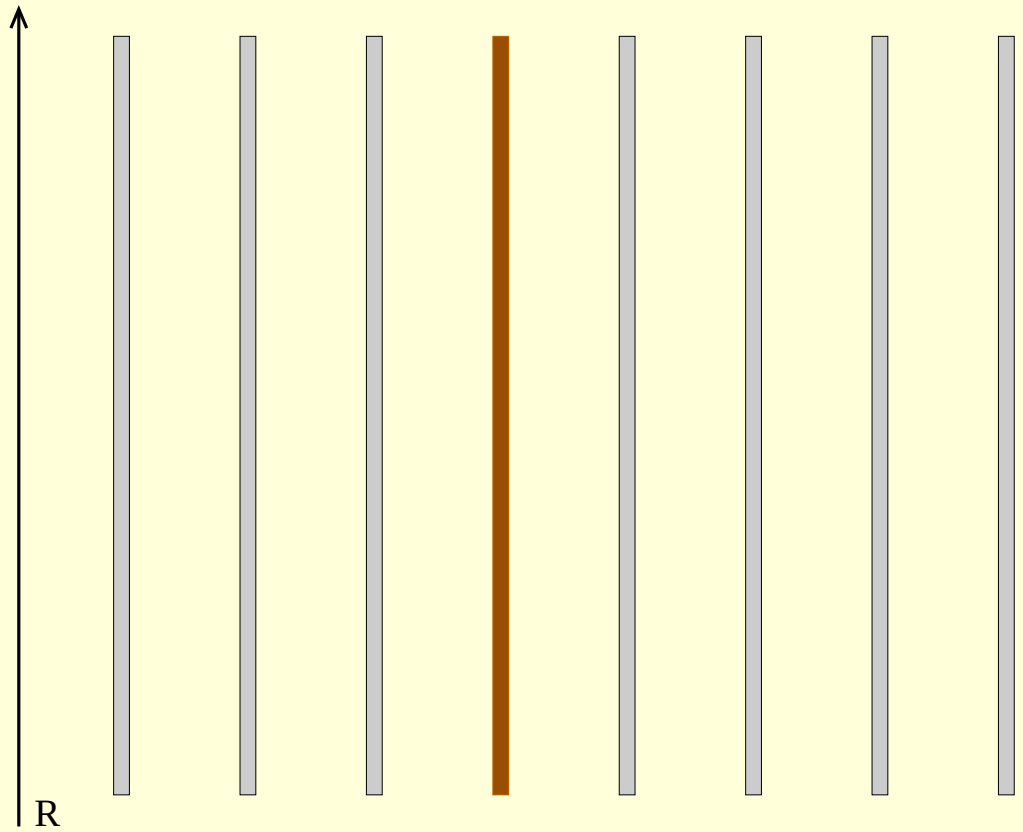
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

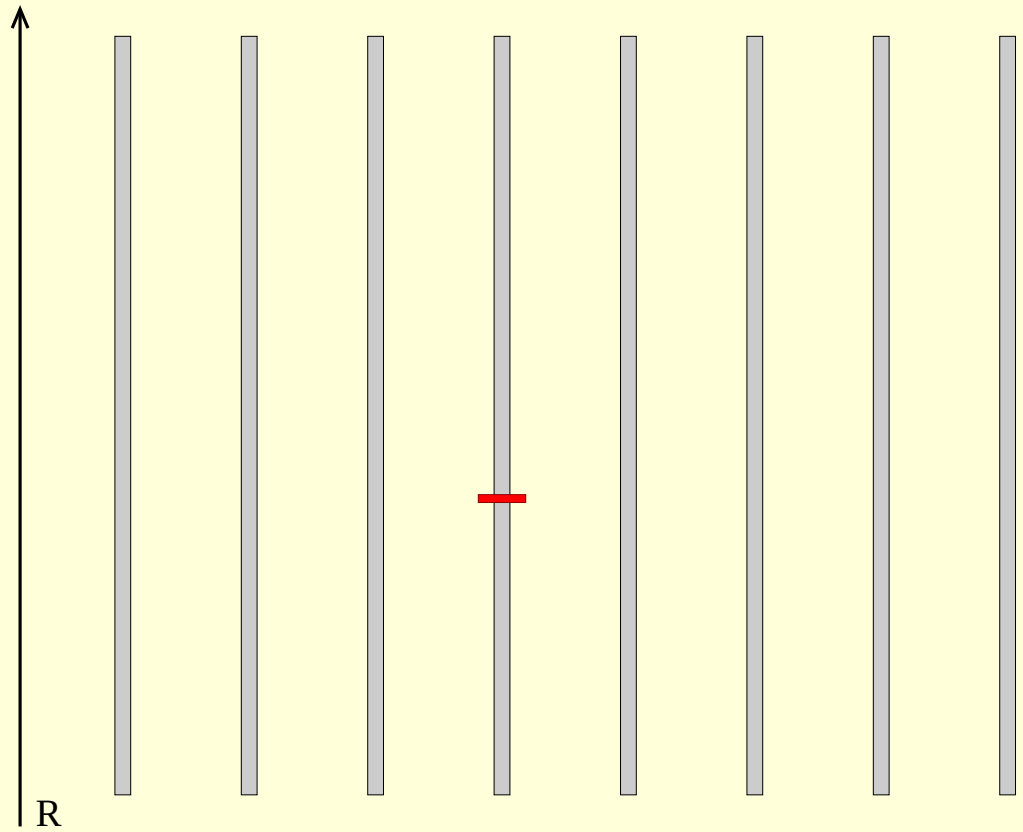
Stochastic Bandits and Subset Selection



In an *n*-armed bandit:

find the *m* arms with the highest means

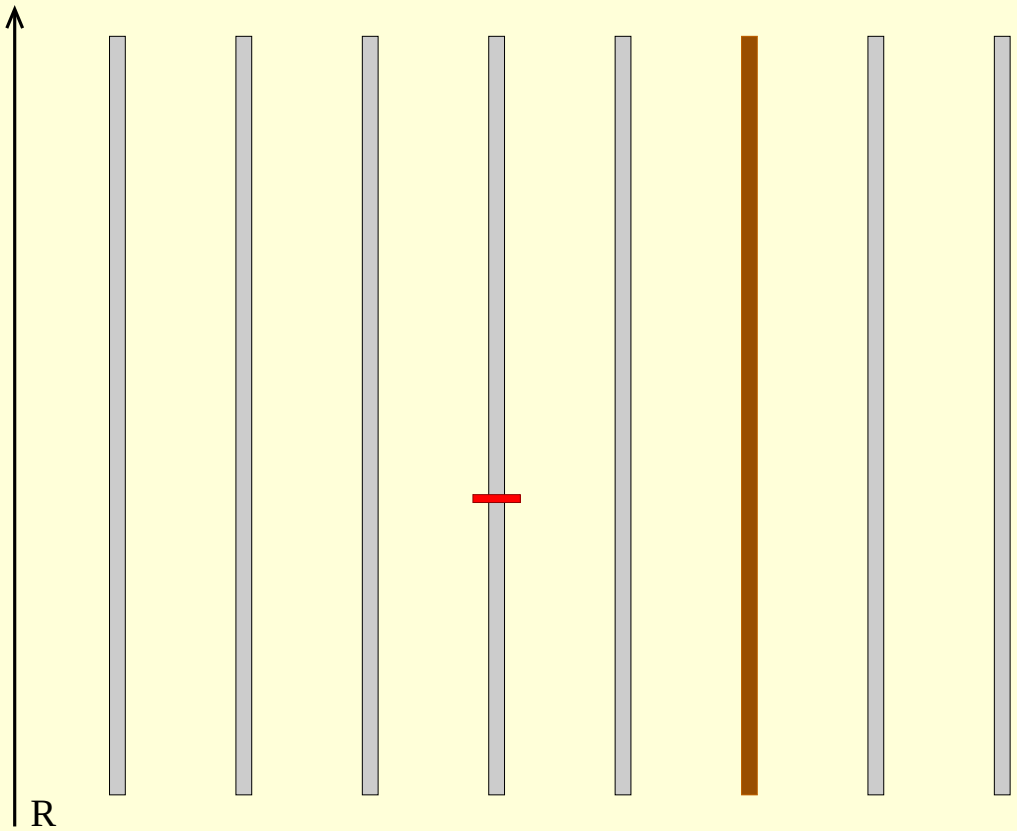
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

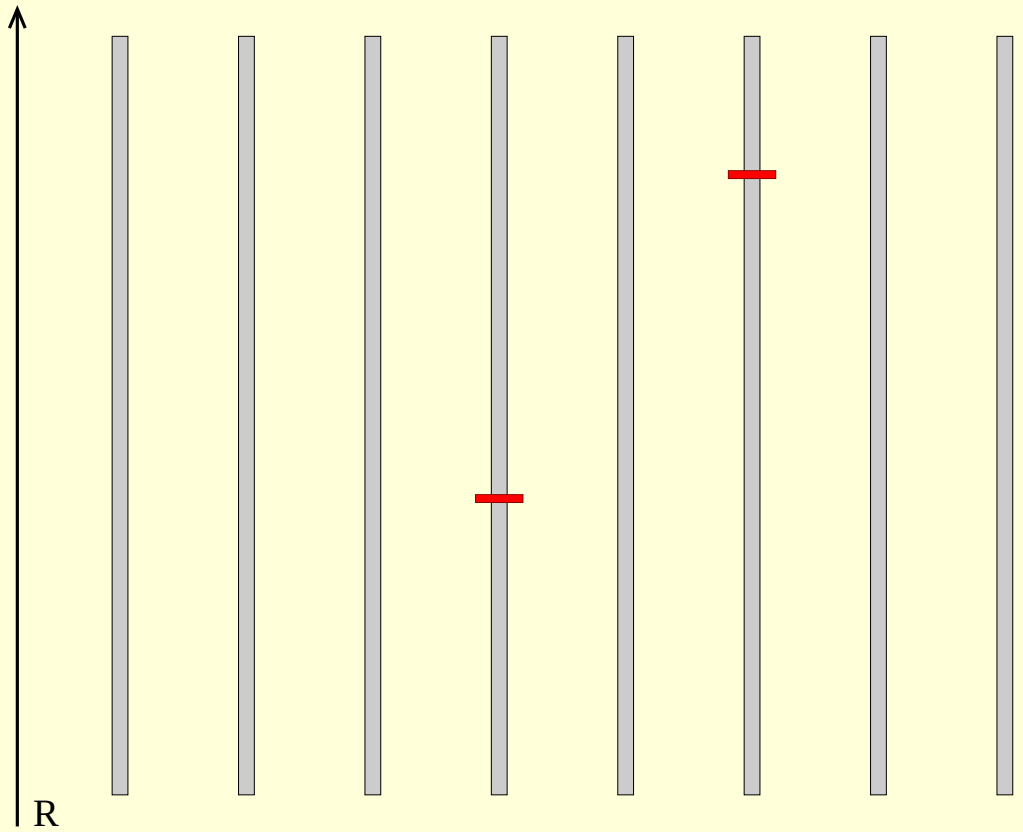
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

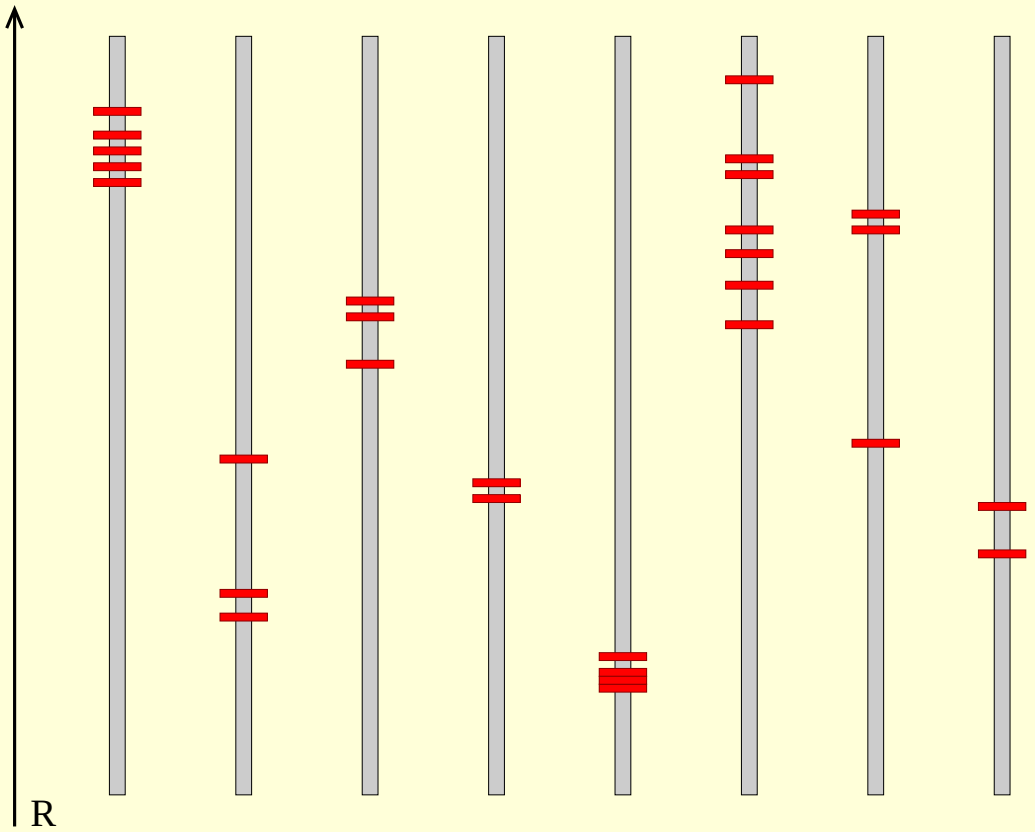
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

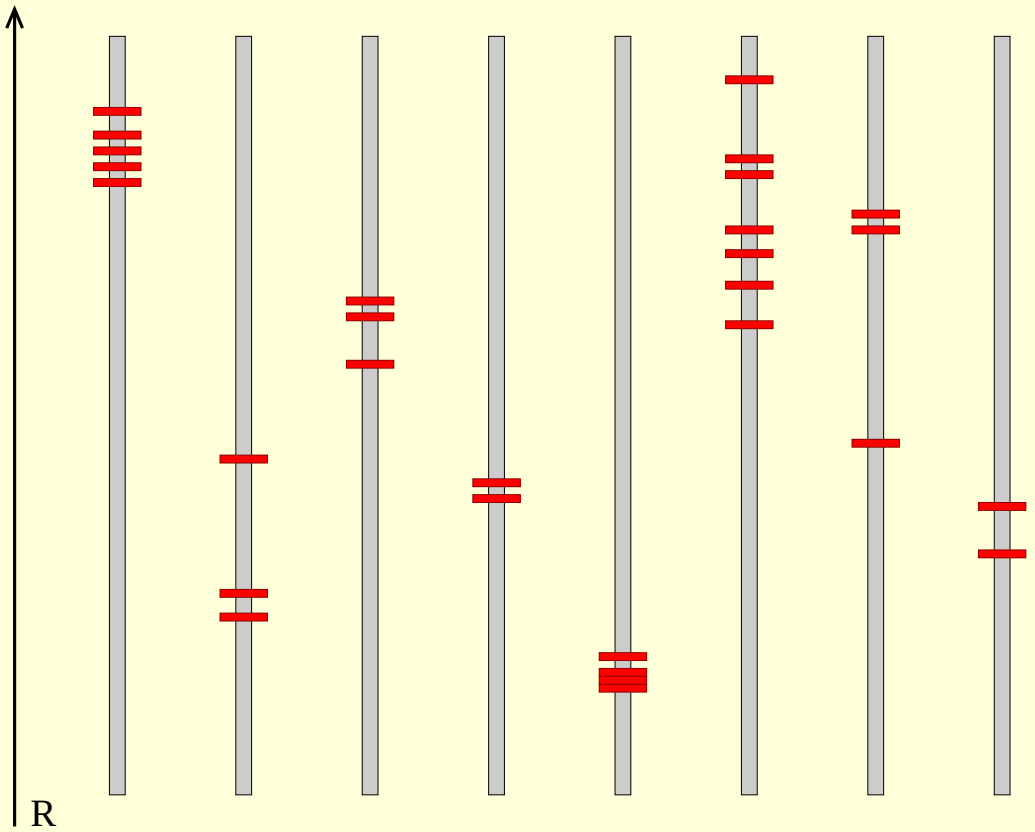
Stochastic Bandits and Subset Selection



In an n-armed bandit:

find the m arms with the highest means

Stochastic Bandits and Subset Selection

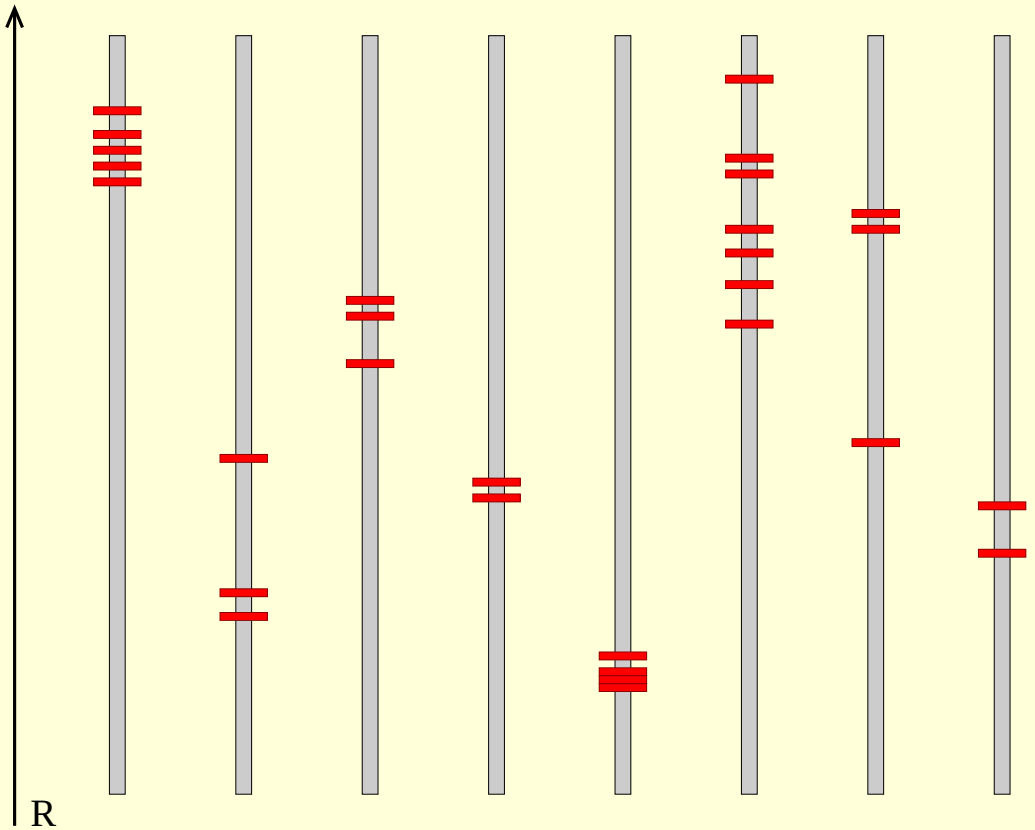


In an n-armed bandit:

find the m arms with the highest means

with high probability

Stochastic Bandits and Subset Selection



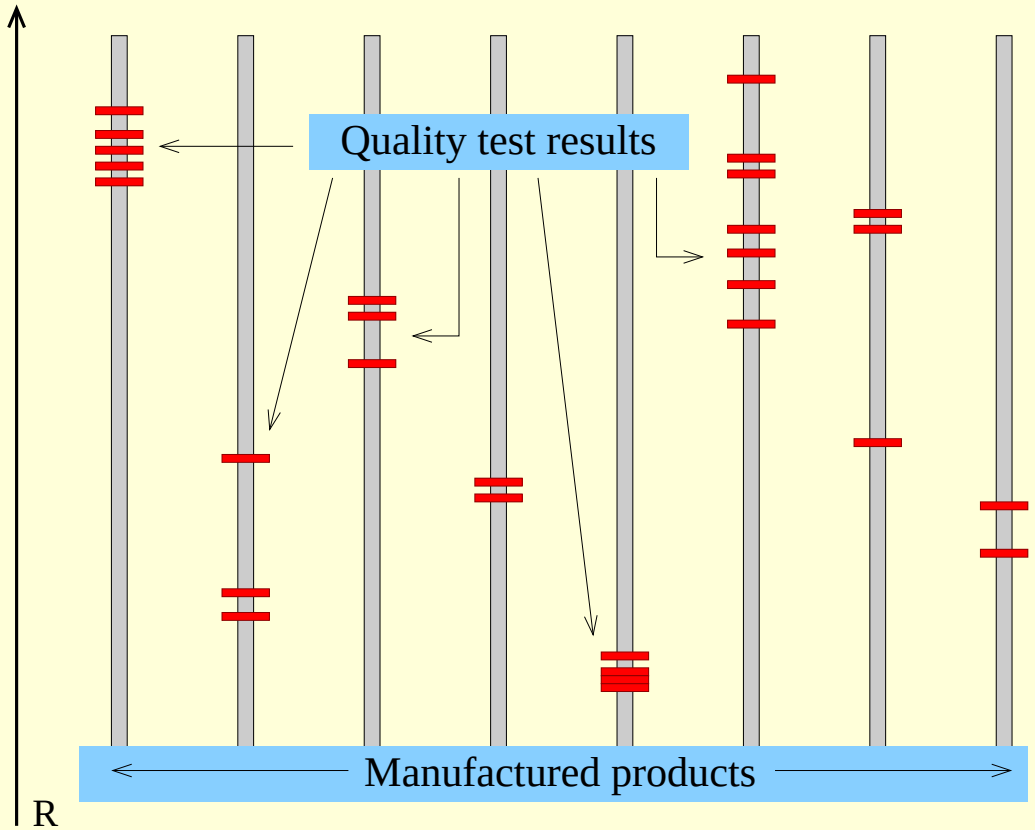
In an n-armed bandit:

find the m arms with the highest means

with high probability

using a *minimal* number of samples.

Stochastic Bandits and Subset Selection



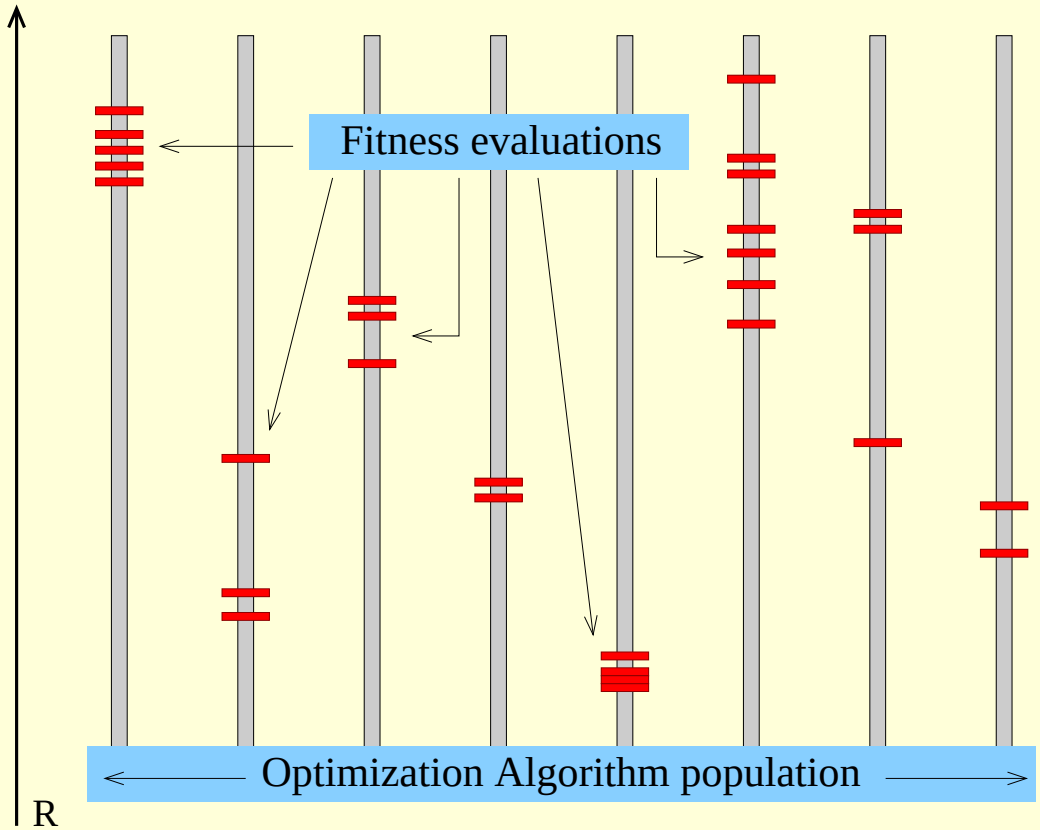
In an *n*-armed bandit:

find the *m* arms with the highest means

with high probability

using a *minimal* number of samples.

Stochastic Bandits and Subset Selection



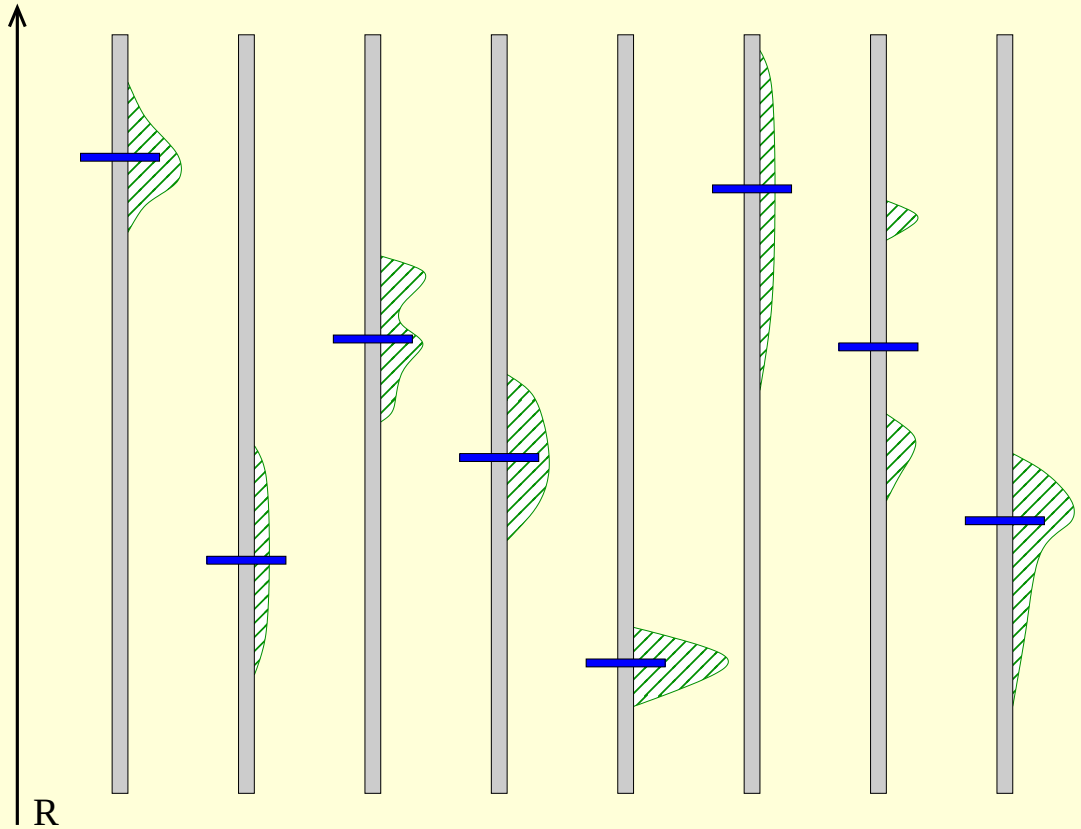
In an *n*-armed bandit:

find the *m* arms with the highest means

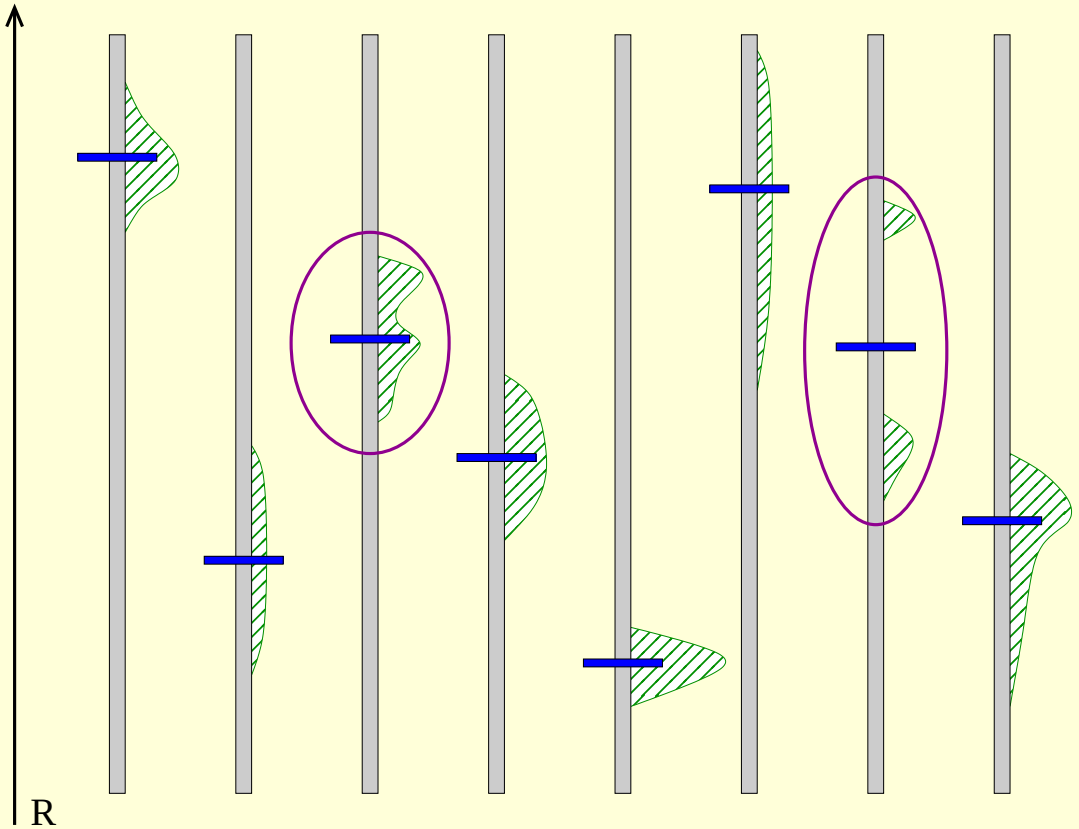
with high probability

using a *minimal* number of samples.

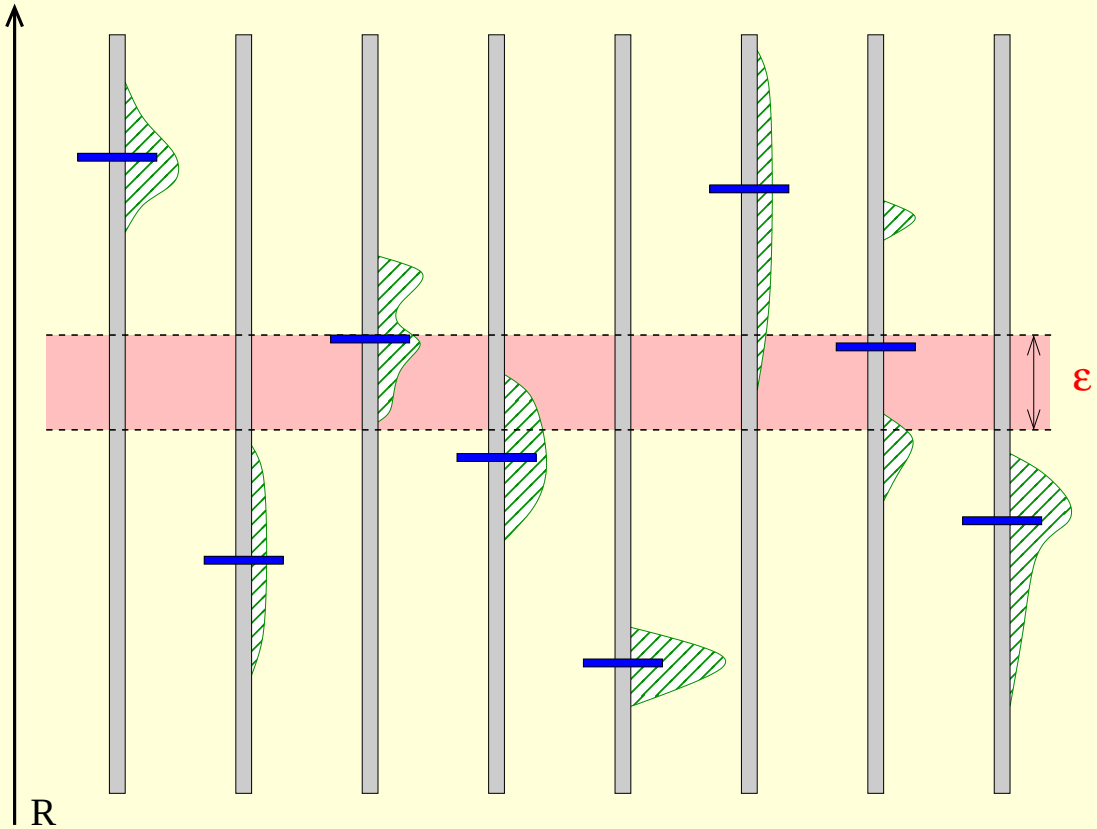
PAC Formulation



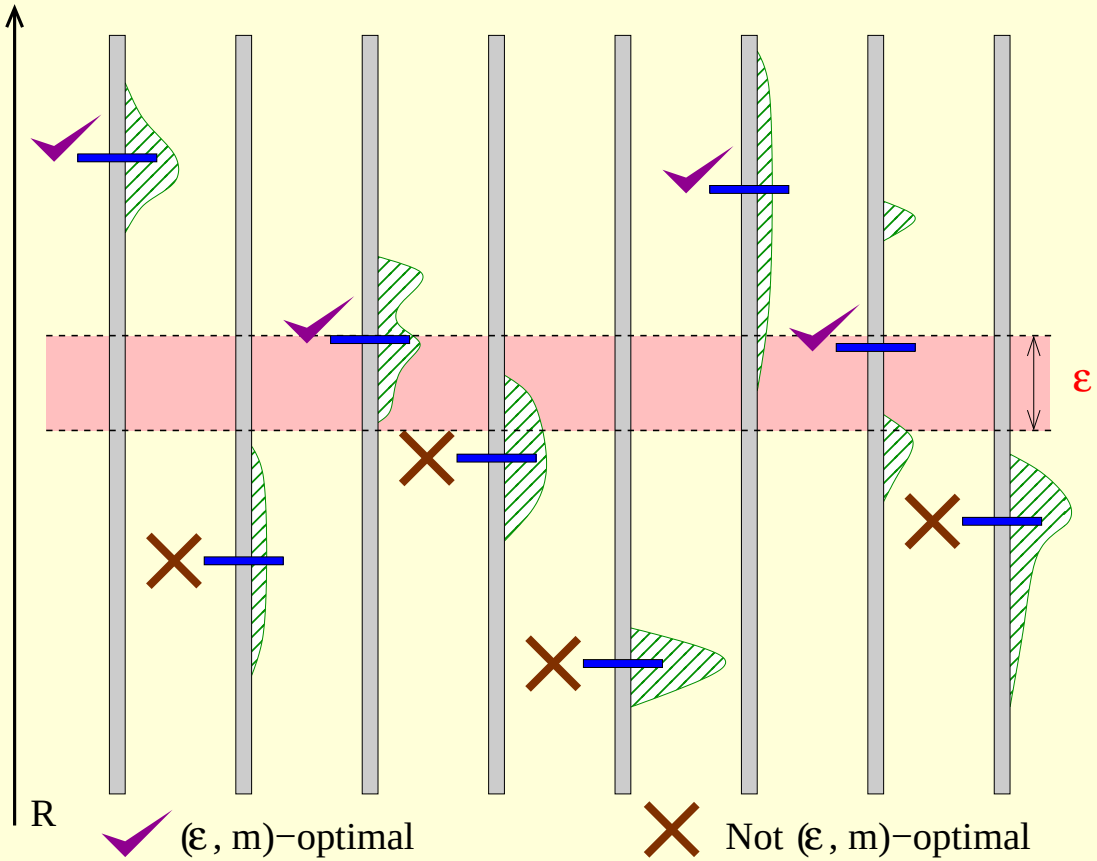
PAC Formulation



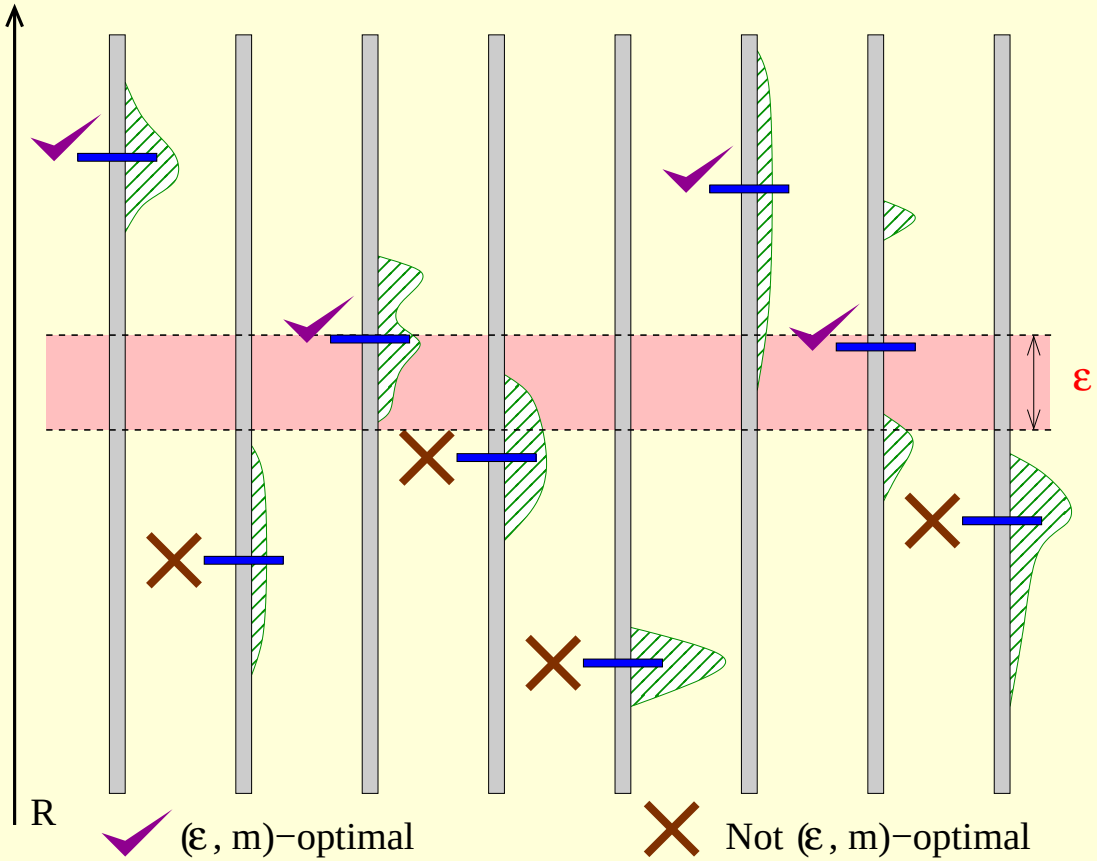
PAC Formulation



PAC Formulation



PAC Formulation



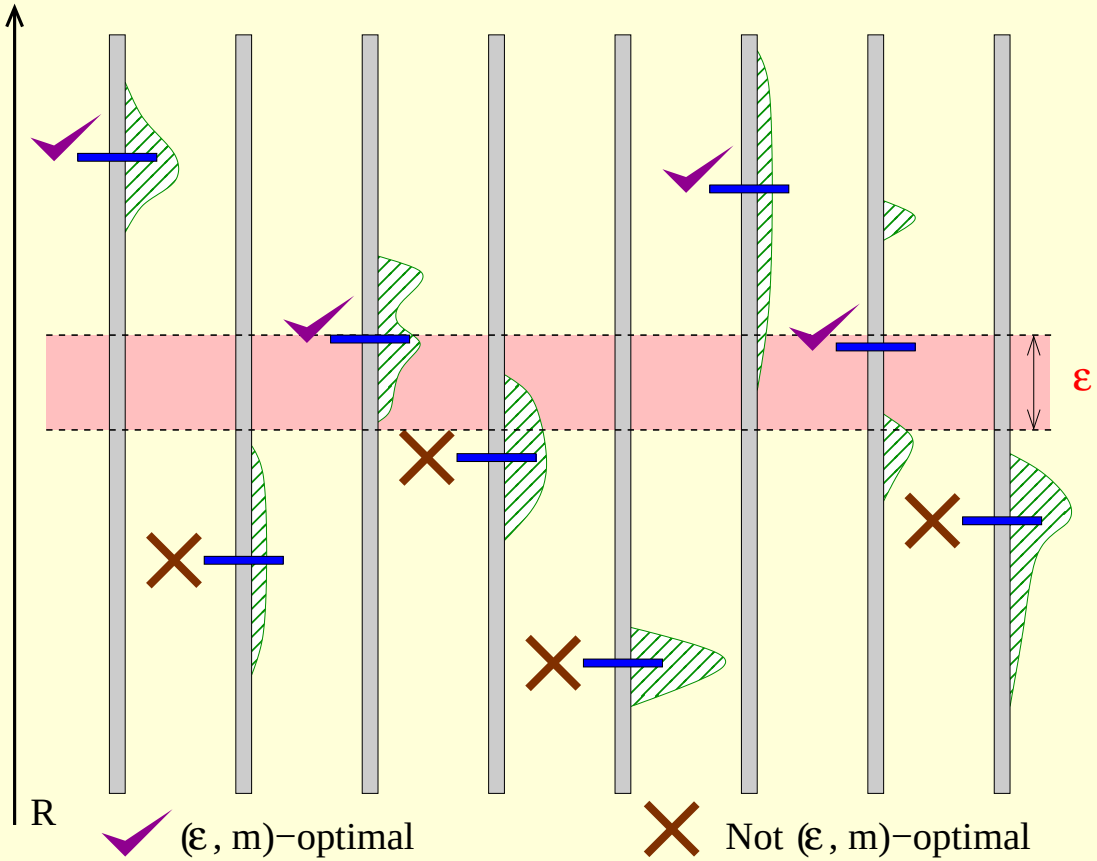
In an n-armed bandit:

find m (ϵ, m)-optimal arms

with probability at least $1 - \delta$

using a minimal number of samples.

PAC Formulation



In an n-armed bandit:

find m (ϵ, m)-optimal arms

with probability at least $1 - \delta$

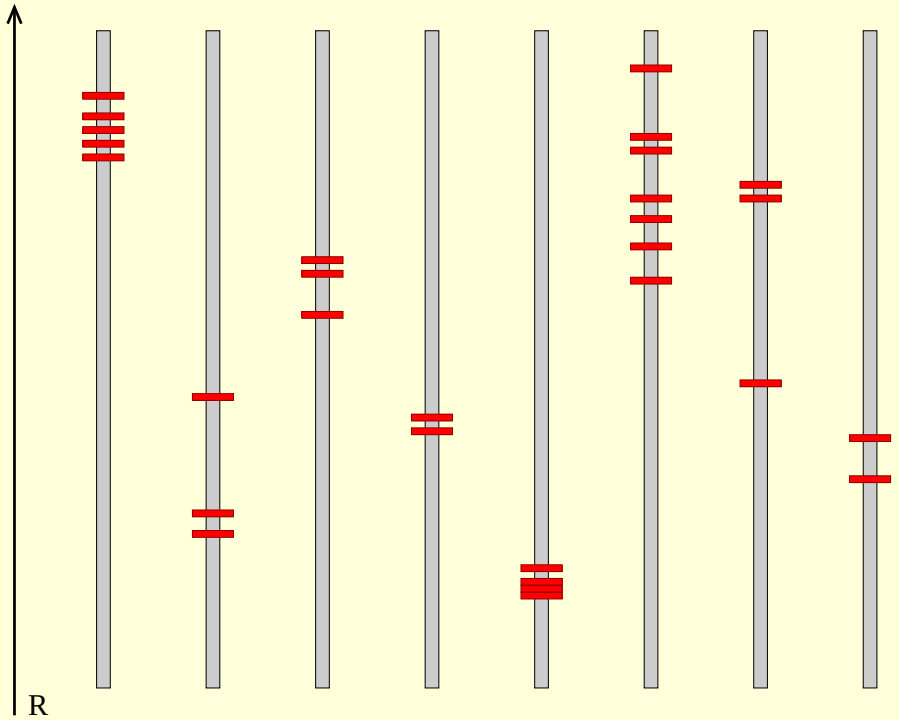
using a minimal number of samples.

$m = 1$: Even-Dar *et al.* (2006)

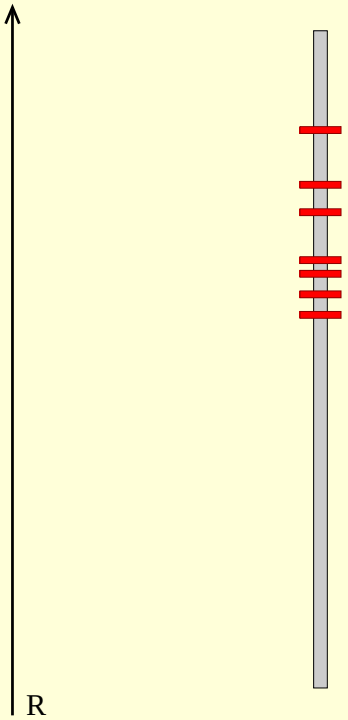
Related Work

Us	Them
m arms	1 arm [Even-Dar, Mannor, and Mansour (2006)]
PAC	Regret [Robbins (1952)] [Auer, Cesa-Bianchi, and Fischer (2002)] Simple regret [Audibert, Bubeck, and Munos (2010)]
Stochastic rewards	Adversarial rewards [Auer, Cesa-Bianchi, Freund, and Schapire (2002)]
Independent arms	Dependent arms [Pandey, Chakrabarti, and Agarwal (2007)] [Kleinberg, Slivkins, and Upfal (2008)]

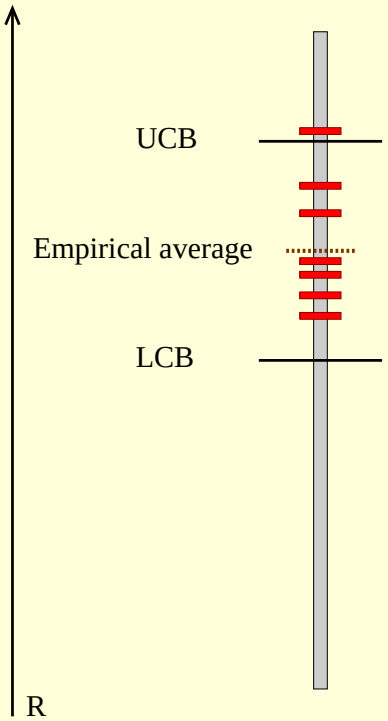
Confidence Bounds on the Mean



Confidence Bounds on the Mean



Confidence Bounds on the Mean

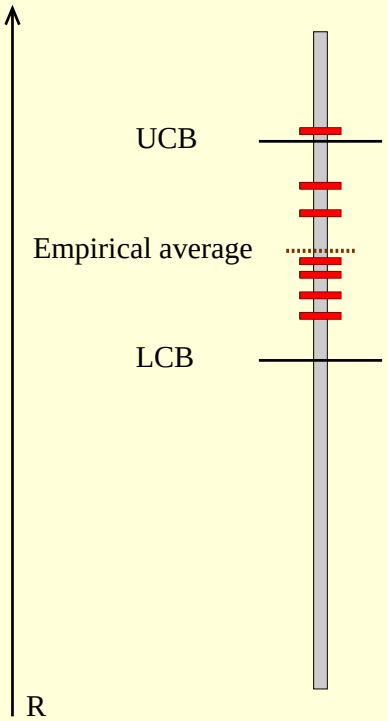


- Hoeffding's inequality: With probability at least $1 - \delta$:

$$\text{True mean} \geq \text{Empirical average} - B\sqrt{\frac{1}{2u}\ln(\frac{1}{\delta})}.$$

$$\text{True mean} \leq \text{Empirical average} + B\sqrt{\frac{1}{2u}\ln(\frac{1}{\delta})}.$$

Confidence Bounds on the Mean



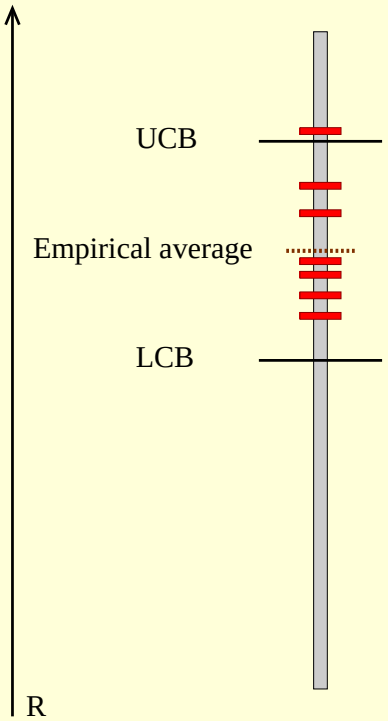
- Hoeffding's inequality: With probability at least $1 - \delta$:

$$\text{True mean} \geq \text{Empirical average} - B\sqrt{\tfrac{1}{2u}\ln(\tfrac{1}{\delta})}.$$

$$\text{True mean} \leq \text{Empirical average} + B\sqrt{\tfrac{1}{2u}\ln(\tfrac{1}{\delta})}.$$

- For simplicity assume $B = 1$; generalizes to distributions with known, finite range.

Confidence Bounds on the Mean



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

$$\text{True mean} \geq \text{Empirical average} - B\sqrt{\tfrac{1}{2u}\ln(\tfrac{1}{\delta})}.$$
$$\text{True mean} \leq \text{Empirical average} + B\sqrt{\tfrac{1}{2u}\ln(\tfrac{1}{\delta})}.$$

- For simplicity assume $B = 1$; generalizes to distributions with known, finite range.
- We employ Hoeffding's inequality and a KL-divergence-based confidence bound.

Algorithms for Subset Selection

- DIRECT Algorithm:

  Sample each arm $\left\lceil \frac{2}{\epsilon^2} ln\left(\frac{n}{\delta}\right) \right\rceil$ times.

  Return m arms with highest *empirical* averages.

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

Algorithms for Subset Selection

- DIRECT Algorithm:

  Sample each arm $\left\lceil \frac{2}{\epsilon^2} ln\left(\frac{n}{\delta}\right) \right\rceil$ times.

  Return m arms with highest *empirical* averages.

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

- HALVING Algorithm:

  Sample each arm $u_1(n, m, \epsilon, \delta)$ times.
  Discard half the arms with lower empirical averages.
  Sample each remaining arm $u_2(n, m, \epsilon, \delta)$ 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 $O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)$.

Algorithms for Subset Selection

- DIRECT Algorithm:

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

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

- HALVING Algorithm:

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

  ⋮

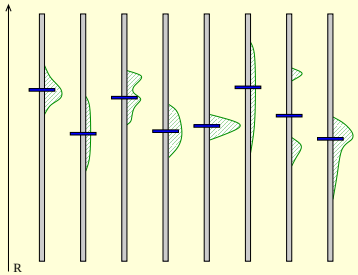
  Until m arms remain.

  - Achieves PAC guarantee.
  - Sequence (u_i) such that total number of samples is $O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)$.

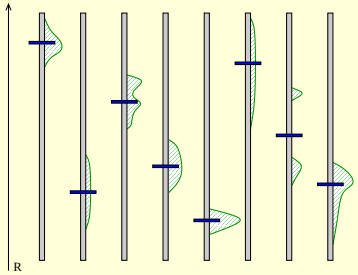
- **Lower bound**: There exist bandit instances (with Bernoulli arms) on which any PAC algorithm needs at least $\Omega\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)$ samples.

Problem Complexity

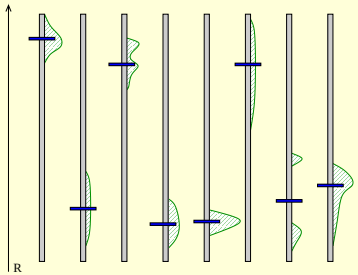
Instance 1



Instance 2

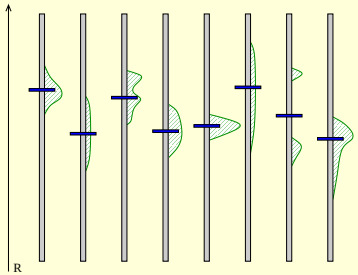


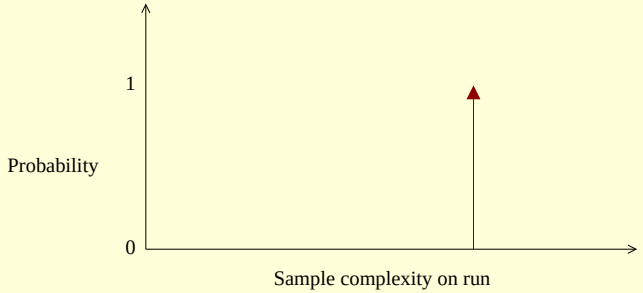
Instance 3



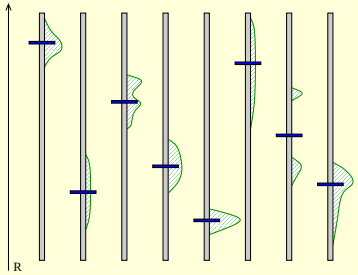
Problem Complexity

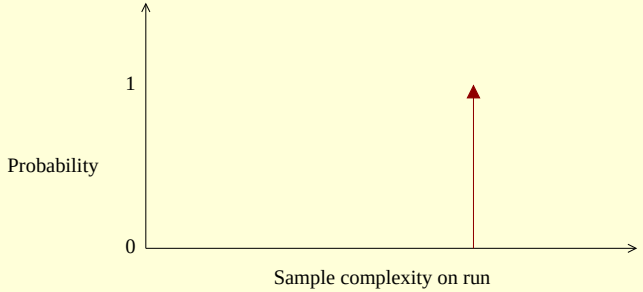
Instance 1



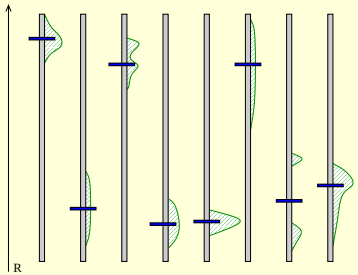


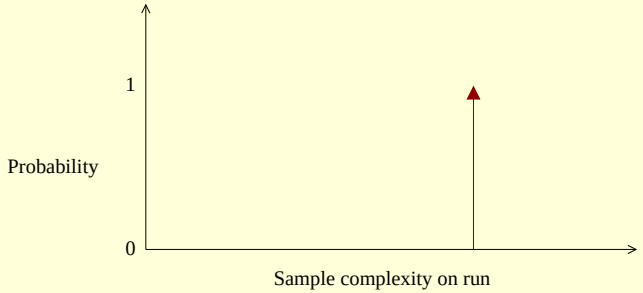
Instance 2





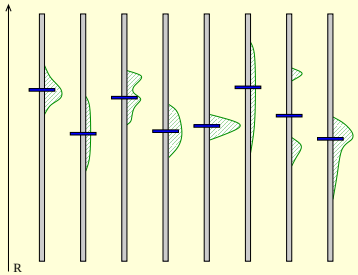
Instance 3

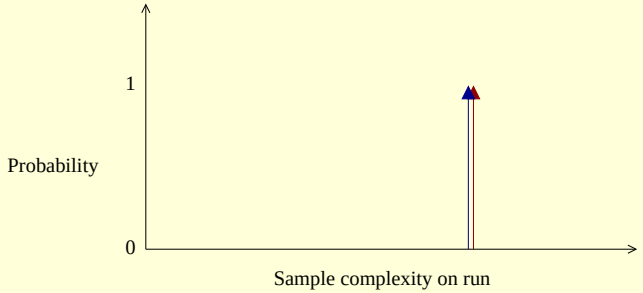




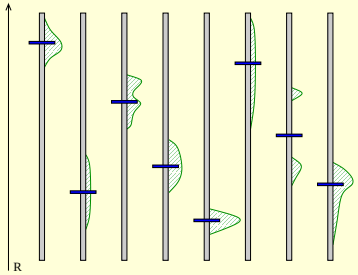
Problem Complexity

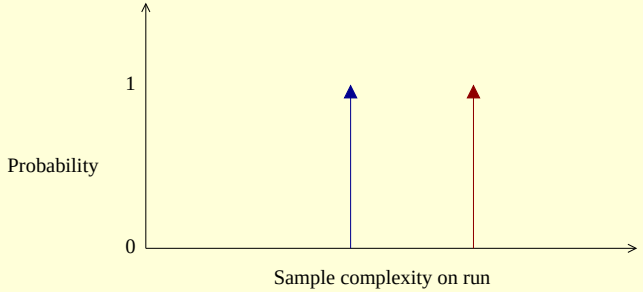
Instance 1



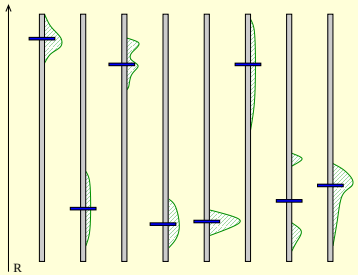


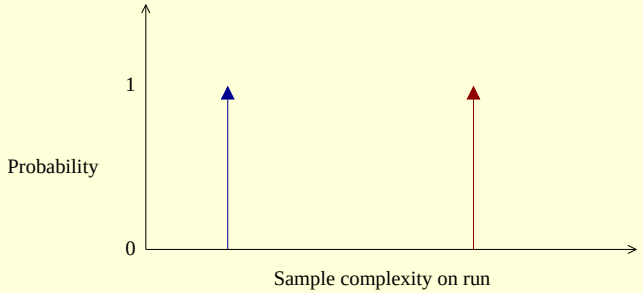
Instance 2





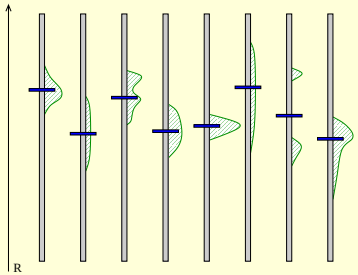
Instance 3

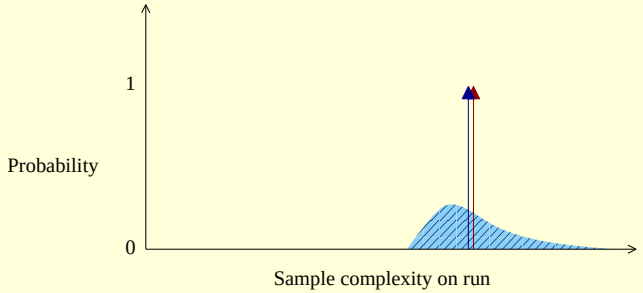




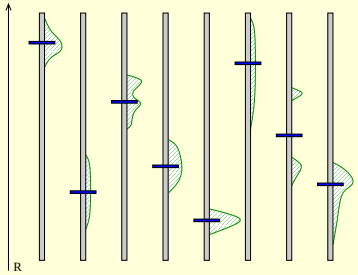
Problem Complexity

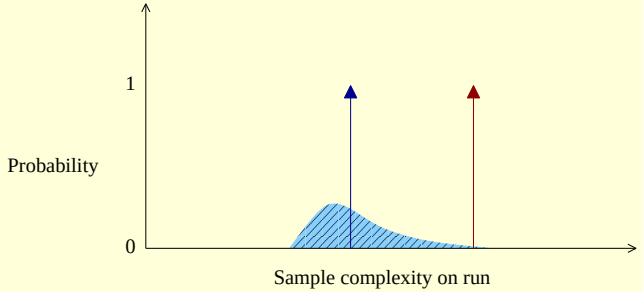
Instance 1



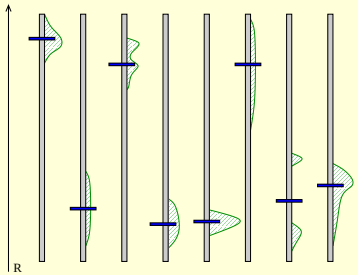


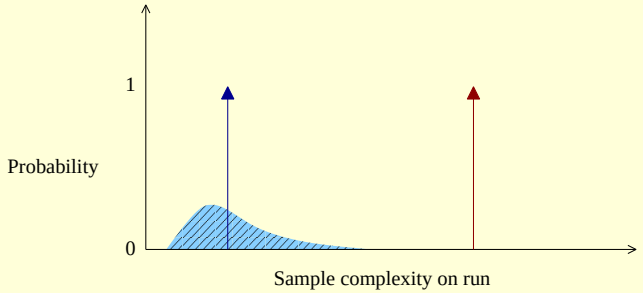
Instance 2





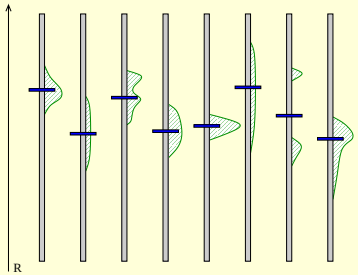
Instance 3

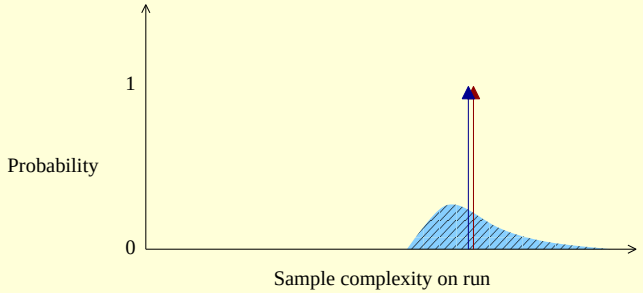




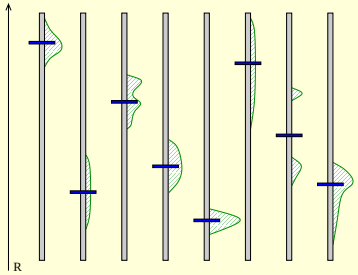
Problem Complexity

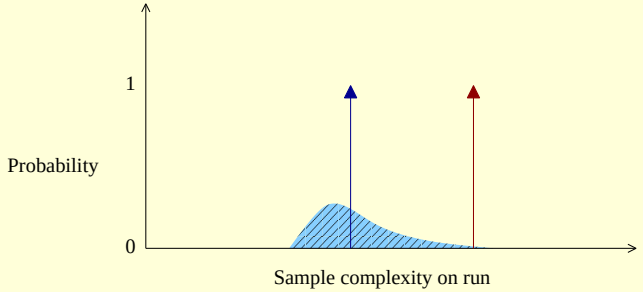
Instance 1



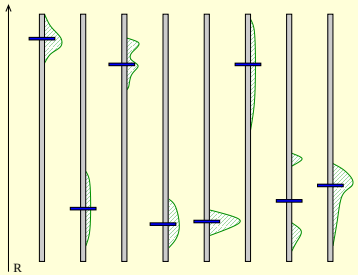


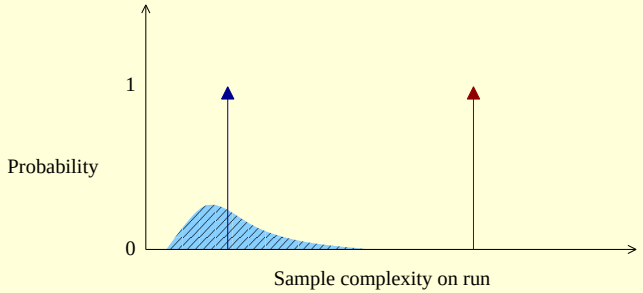
Instance 2





Instance 3



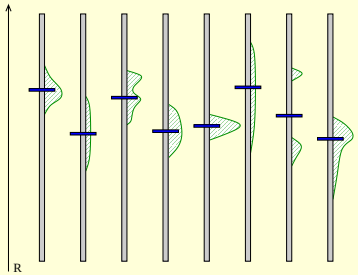


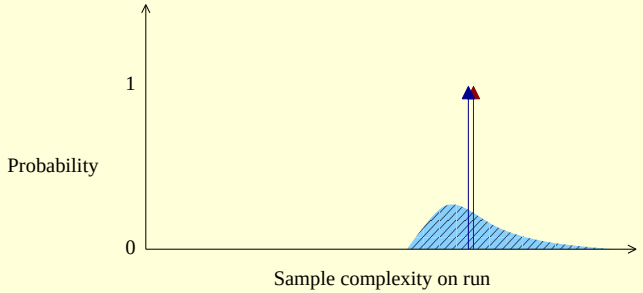
$$\Delta_a \stackrel{\text{def}}{=} \begin{cases} p_a - p_{m+1} & \text{if } 1 \leq a \leq m, \\ p_m - p_a & \text{if } m+1 \leq a \leq n. \end{cases}$$

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

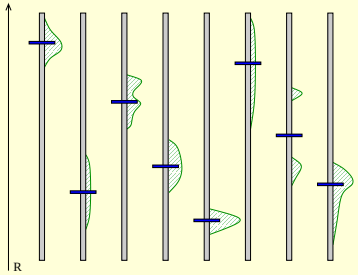
Problem Complexity

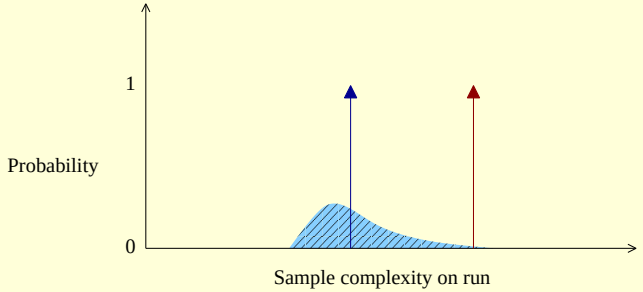
Instance 1



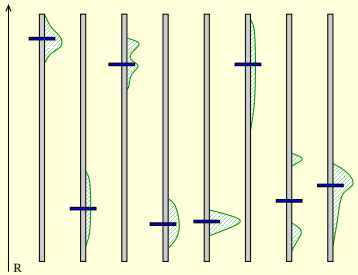


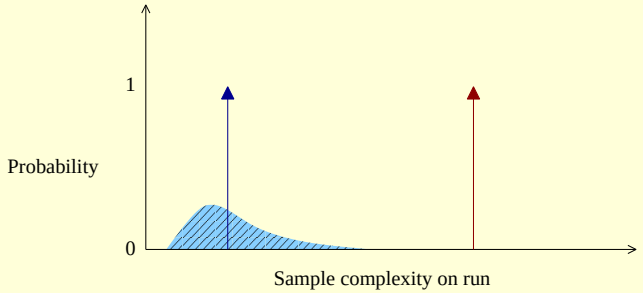
Instance 2





Instance 3





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

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

In practice: $H^{\epsilon} \ll \frac{n}{\epsilon^2}$.

LUCB Algorithm

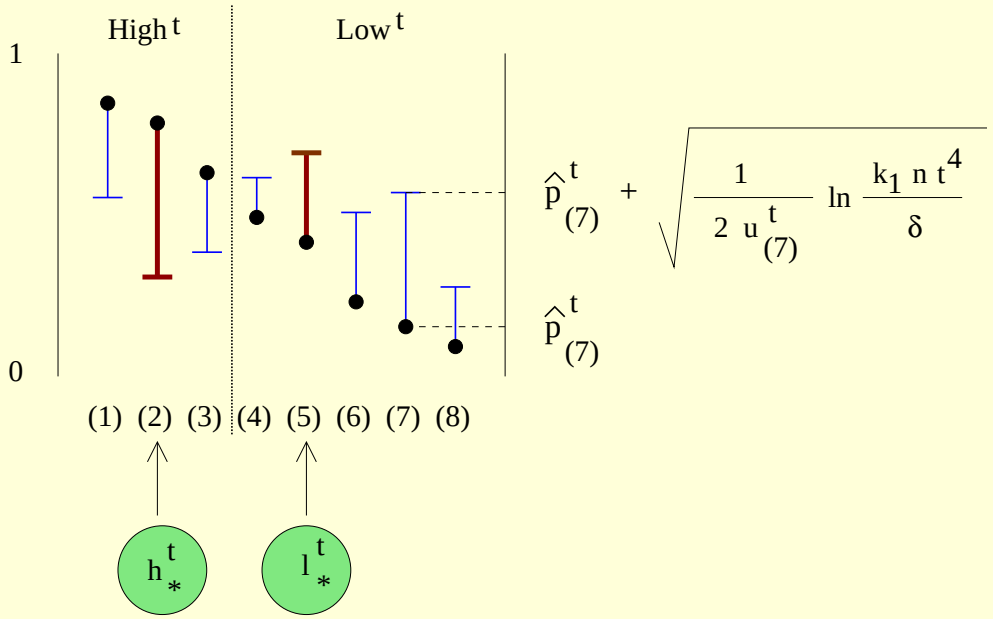
Achieves PAC guarantee.

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

LUCB Algorithm

Achieves PAC guarantee.

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



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:

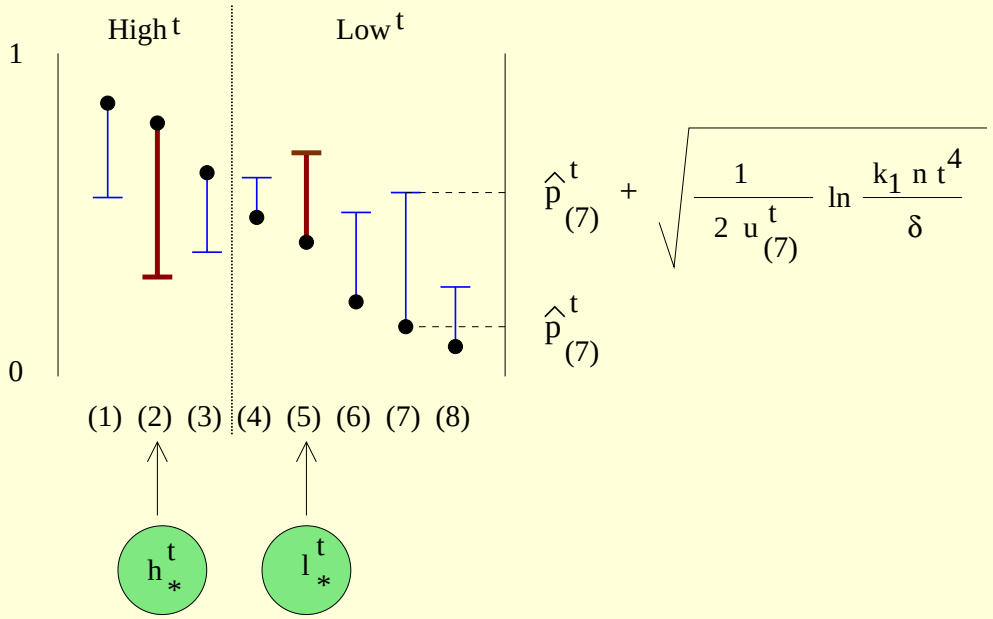
On round t: sample arms h^t_* and l^t_*.

LUCB Algorithm

Achieves PAC guarantee.

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

Bound novel even for $m = 1$.



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_*.

KL-LUCB Algorithm

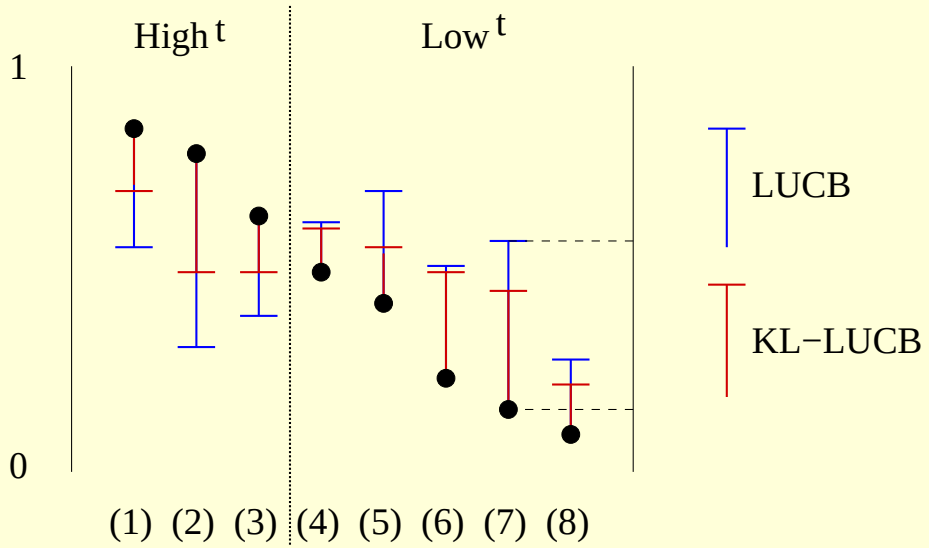
LUCB upper bound = $\hat{p}_a^t + \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{knt^\alpha}{\delta}\right)}$.

LUCB lower bound = $\hat{p}_a^t - \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{knt^\alpha}{\delta}\right)}$.

KL-LUCB upper bound = $\max\left\{q \in [\hat{p}_a^t, 1] : u_a^t KL(\hat{p}_a^t, q) \leq \ln\left(\frac{knt^\alpha}{\delta}\right)\right\}$.

KL-LUCB lower bound = $\min\left\{q \in [0, \hat{p}_a^t] : u_a^t KL(\hat{p}_a^t, q) \leq \ln\left(\frac{knt^\alpha}{\delta}\right)\right\}$.

KL-LUCB confidence bounds provably tighter (Pinsker's Inequality).
Apply same stopping rule and sampling strategy as LUCB.



KL-LUCB Algorithm

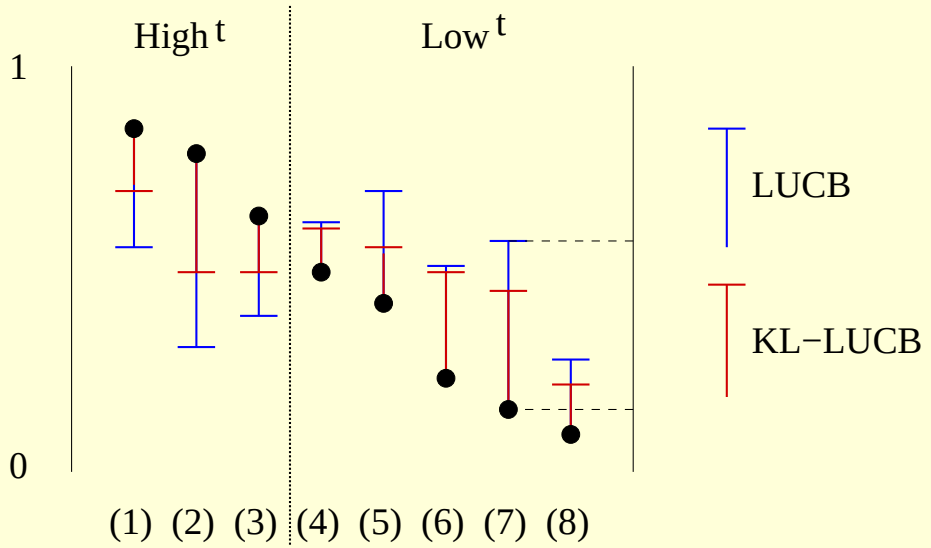
LUCB upper bound = $\hat{p}_a^t + \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{knt^\alpha}{\delta}\right)}$.

LUCB lower bound = $\hat{p}_a^t - \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{knt^\alpha}{\delta}\right)}$.

KL-LUCB upper bound = $\max\left\{q \in [\hat{p}_a^t, 1] : u_a^t KL(\hat{p}_a^t, q) \leq \ln\left(\frac{knt^\alpha}{\delta}\right)\right\}$.

KL-LUCB lower bound = $\min\left\{q \in [0, \hat{p}_a^t] : u_a^t KL(\hat{p}_a^t, q) \leq \ln\left(\frac{knt^\alpha}{\delta}\right)\right\}$.

KL-LUCB confidence bounds provably tighter (Pinsker's Inequality).
Apply same stopping rule and sampling strategy as LUCB.



KL-LUCB Algorithm

Delivers PAC guarantee.

Expected sample complexity =

$$\min\left\{O\left(H'^{\epsilon}\log\left(\frac{H'^{\epsilon}}{\delta}\right)\right), O\left(\frac{n}{\epsilon^2}log\left(\frac{m}{\delta}\right)\right)\right\}, \text{ where}$$

$$H'^{\epsilon} = \min_{c\in[p_{m+1},p_m]}\sum_{a=1}^{n}\frac{1}{\max\left\{d^*(p_a,c),\frac{\epsilon^2}{2}\right\}}.$$

$d^*(x,y)$ is the Chernoff Information between Bernoulli distributions with means x and y, defined as:

$$d^*(x,y) = KL(z^*,x) = KL(z^*,y), \text{ where}$$

z^* is the unique $z \in [\min\{x,y\}, \max\{x,y\}]$ such that $KL(z,x) = KL(z,y)$.

KL-LUCB Algorithm

Delivers PAC guarantee.

Expected sample complexity =

$$\min\left\{O\left(H'^{\epsilon}\log\left(\frac{H'^{\epsilon}}{\delta}\right)\right), O\left(\frac{n}{\epsilon^2}log\left(\frac{m}{\delta}\right)\right)\right\}, \text{ where}$$

$$H'^{\epsilon} = \min_{c\in[p_{m+1},p_m]}\sum_{a=1}^{n}\frac{1}{\max\left\{d^*(p_a,c),\frac{\epsilon^2}{2}\right\}}.$$

$d^*(x,y)$ is the Chernoff Information between Bernoulli distributions with means x and y, defined as:

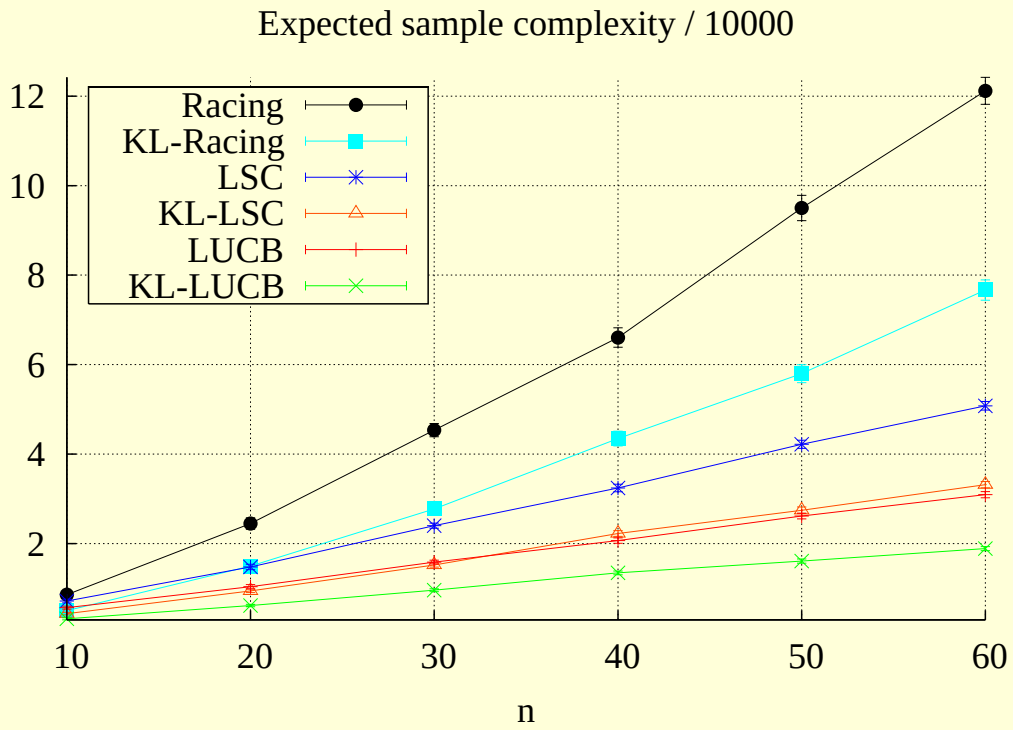
$$d^*(x,y) = KL(z^*,x) = KL(z^*,y), \text{ where}$$

z^* is the unique $z \in [\min\{x,y\}, \max\{x,y\}]$ such that $KL(z,x) = KL(z,y)$.

$H'^{\epsilon} = O(H^{\epsilon})$; typically much smaller.

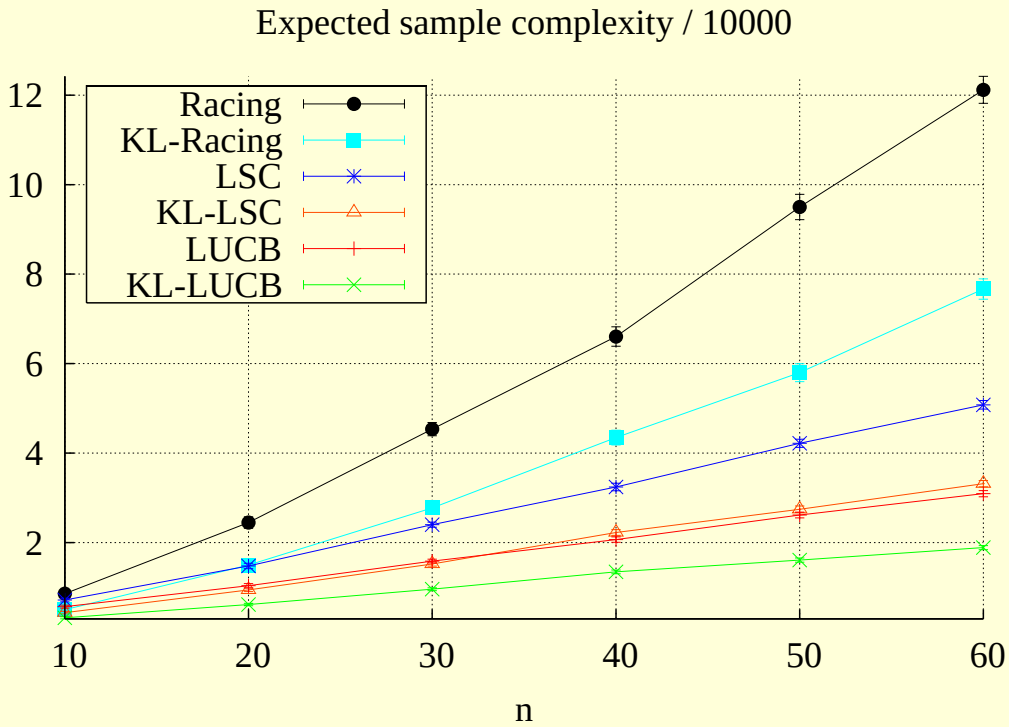
Experiments

- We compare (KL-)LUCB, (KL-)Racing, and (KL-)LSC
- Number of arms n varied.
- 1000 random instances; each arm's mean drawn uniformly at random from [0, 1].
- $m = \frac{n}{5}, \epsilon = 0.1, \delta = 0.1$.



Experiments

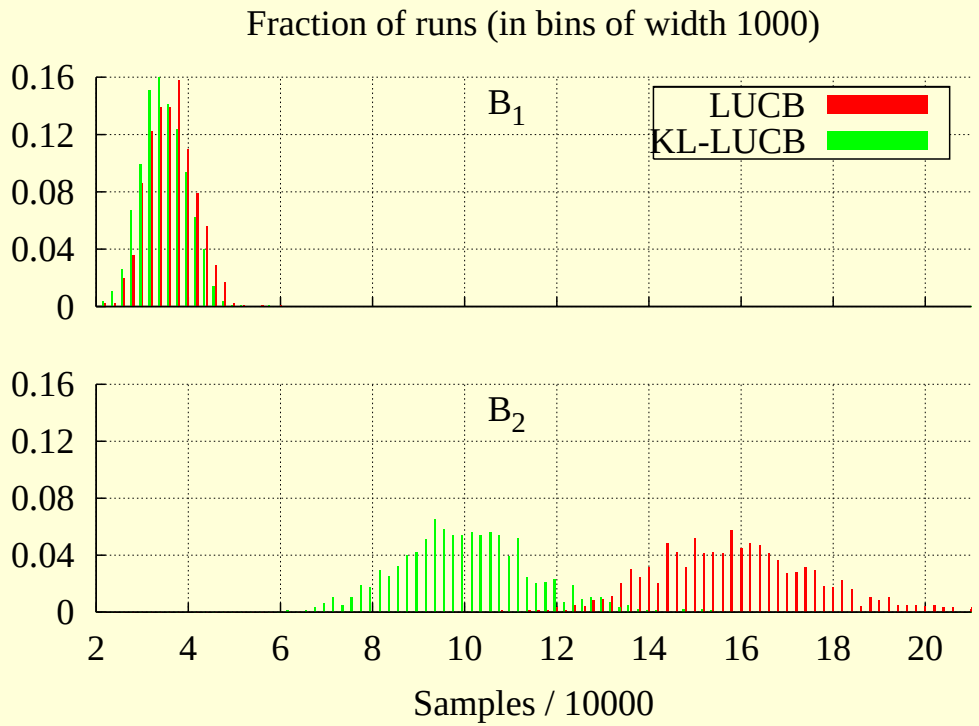
- We compare (KL-)LUCB, (KL-)Racing, and (KL-)LSC
- Number of arms n varied.
- 1000 random instances; each arm's mean drawn uniformly at random from [0, 1].
- $m = \frac{n}{5}, \epsilon = 0.1, \delta = 0.1$.



(KL-)LUCB $>$ (KL-)LSC $>$ (KL-)Racing.
KL-X $>$ X, X $\in$ {LUCB, LSC, Racing}.

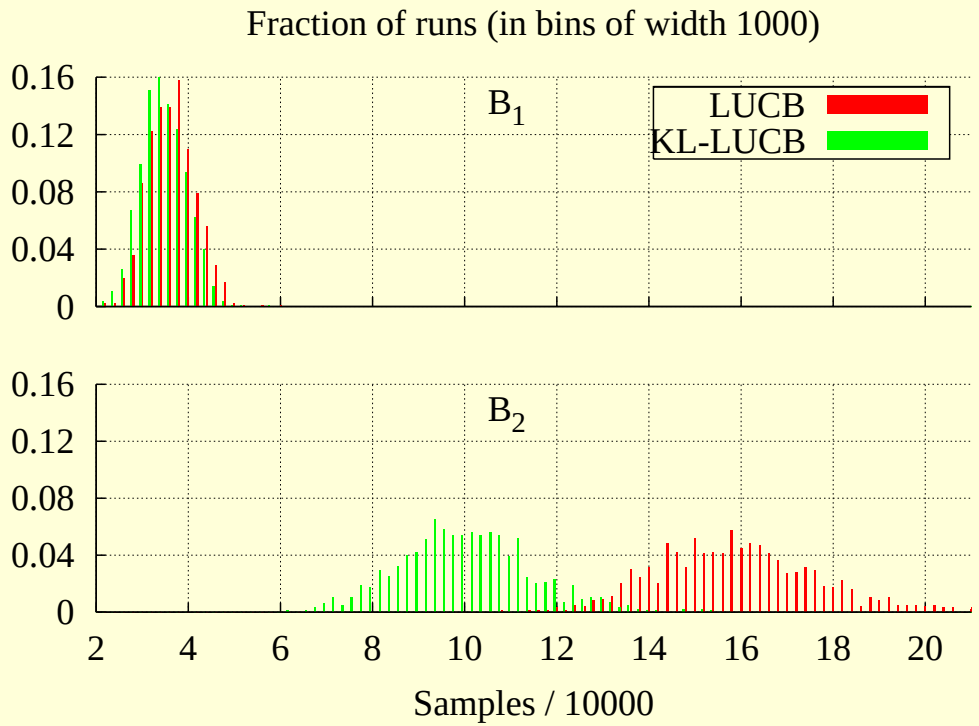
Experiments

- Instance B_1: $n = 15, p_1 = \frac{1}{2}; p_a = \frac{1}{2} - \frac{a}{40}, a = 2, 3, \ldots, n; \epsilon = 0.04$.
- Instance B_2: $n = 15, p_1 = \frac{1}{4}; p_a = \frac{1}{4} - \frac{a}{80}, a = 2, 3, \ldots, n; \epsilon = 0.02$.
- $m = 3, \delta = 0.1$.



Experiments

- Instance B_1: $n = 15, p_1 = \frac{1}{2}; p_a = \frac{1}{2} - \frac{a}{40}, a = 2, 3, \ldots, n; \epsilon = 0.04$.
- Instance B_2: $n = 15, p_1 = \frac{1}{4}; p_a = \frac{1}{4} - \frac{a}{80}, a = 2, 3, \ldots, n; \epsilon = 0.02$.
- $m = 3, \delta = 0.1$.



KL-izing especially economical when means are close to 0 or 1.

Summary

PAC subset selection

n, m, ϵ, δ

Worst case sample complexity upper bound

$O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)$

Worst case sample complexity lower bound

$\Omega\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)$

Expected sample complexity upper bound

LUCB: $\min\left\{O\left(H^{\epsilon} \log\left(\frac{H^{\epsilon}}{\delta}\right)\right), O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)\right\}$

KL-LUCB: $\min\left\{O\left(H'^{\epsilon} \log\left(\frac{H'^{\epsilon}}{\delta}\right)\right), O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)\right\}$

Experiments

(KL-)LUCB $>$ (KL-)LSC $>$ (KL-)Racing

KL-X $>$ X, X $\in$ {LUCB, LSC, Racing}.

Summary

PAC subset selection

n, m, ϵ, δ

Worst case sample complexity upper bound

$O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)$

Worst case sample complexity lower bound

$\Omega\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)$

Expected sample complexity upper bound

LUCB: $\min\left\{O\left(H^{\epsilon} \log\left(\frac{H^{\epsilon}}{\delta}\right)\right), O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)\right\}$

KL-LUCB: $\min\left\{O\left(H'^{\epsilon} \log\left(\frac{H'^{\epsilon}}{\delta}\right)\right), O\left(\frac{n}{\epsilon^2} log\left(\frac{m}{\delta}\right)\right)\right\}$

Experiments

(KL-)LUCB $>$ (KL-)LSC $>$ (KL-)Racing

KL-X $>$ X, X $\in$ {LUCB, LSC, Racing}.

Use KL-LUCB for PAC subset selection!

Future Work

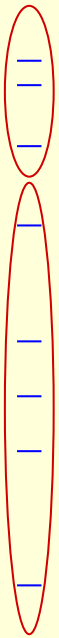
- Expected sample complexity *lower* bound

Future Work

- Expected sample complexity *lower* bound
- Generalized ranking and selection

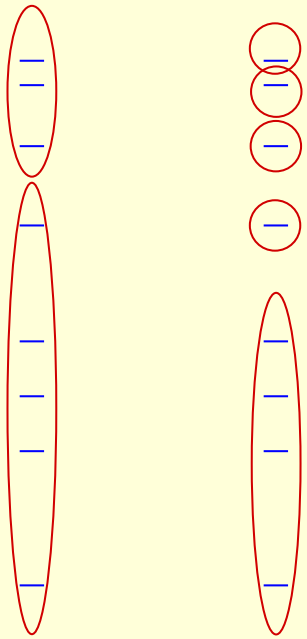
Future Work

- Expected sample complexity *lower* bound
- Generalized ranking and selection



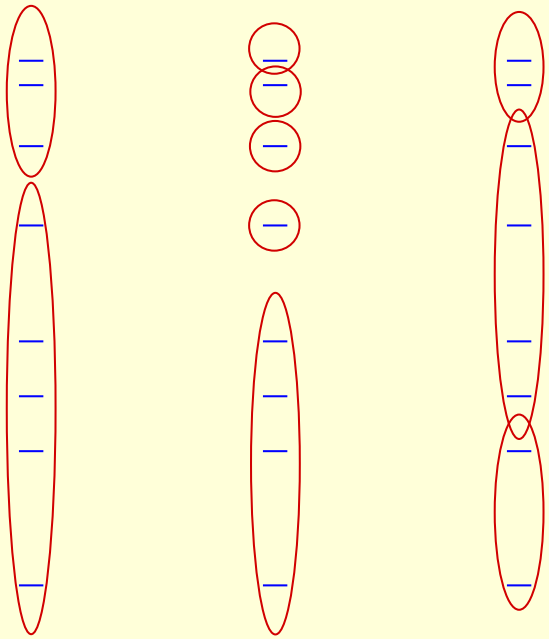
Future Work

- Expected sample complexity *lower* bound
- Generalized ranking and selection



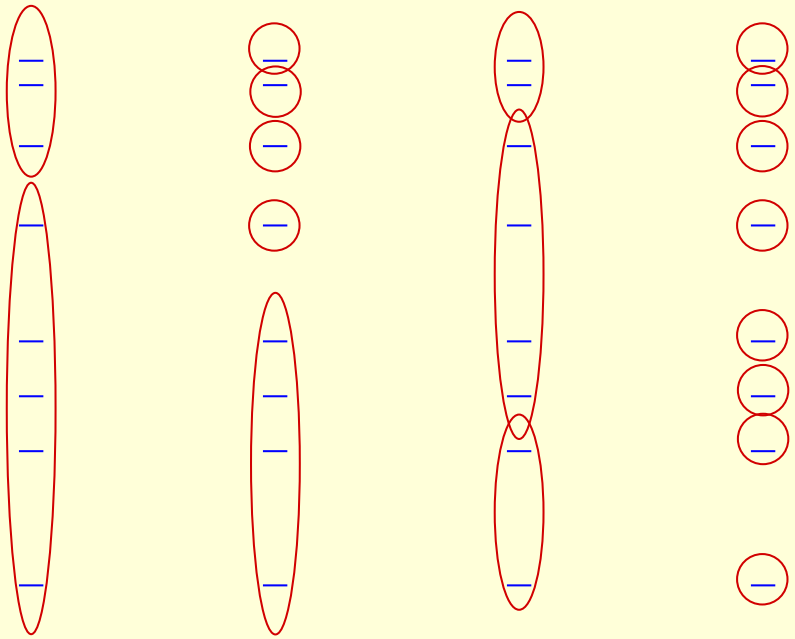
Future Work

- Expected sample complexity *lower* bound
- Generalized ranking and selection



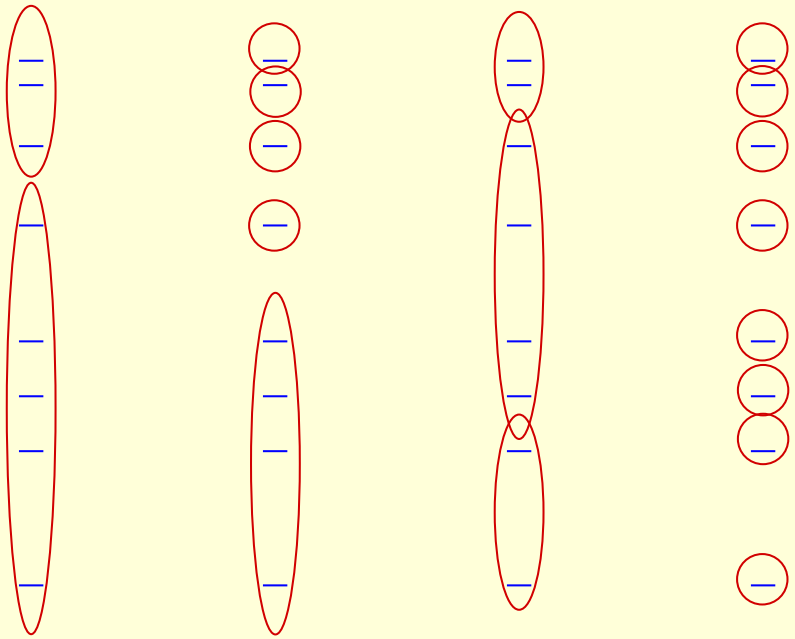
Future Work

- Expected sample complexity *lower* bound
- Generalized ranking and selection



Future Work

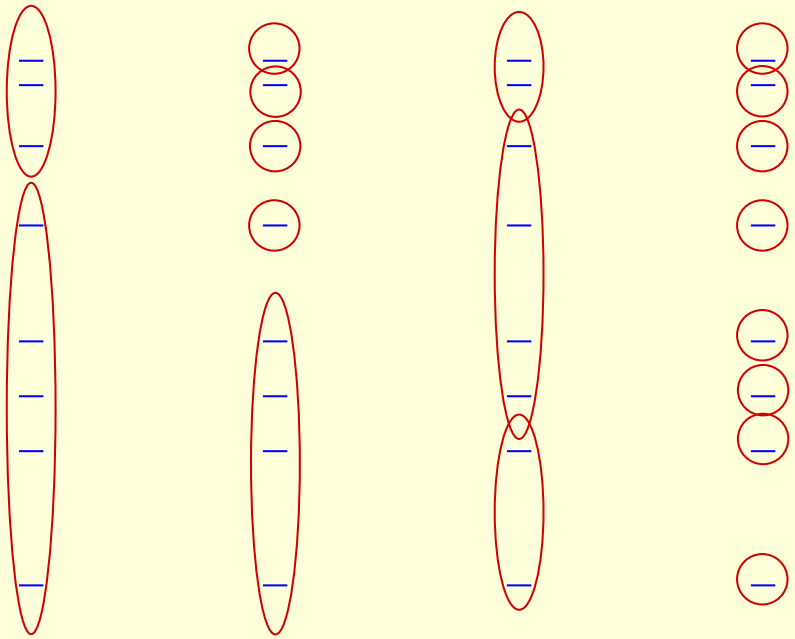
- Expected sample complexity *lower* bound
- Generalized ranking and selection



- Exploration in MDPs with instance-specific sample complexity bounds

Future Work

- Expected sample complexity *lower* bound
- Generalized ranking and selection



- Exploration in MDPs with instance-specific sample complexity bounds

Thank you!