# PAC Subset Selection in Stochastic Multi-armed Bandits

Shivaram Kalyanakrishnan

shivaram@csa.iisc.ernet.in

Department of Computer Science and Automation
Indian Institute of Science

August 2014

# Today's Talk

## Relevant publications

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

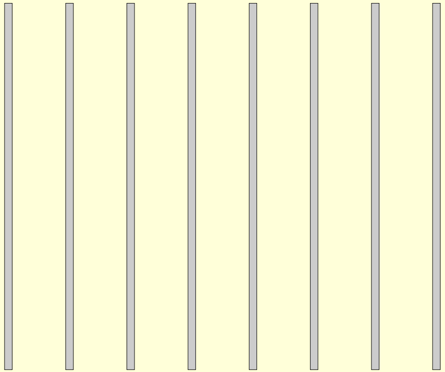
PAC Subset Selection in Stochastic Multi-armed Bandits
Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone, *ICML 2012*.

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

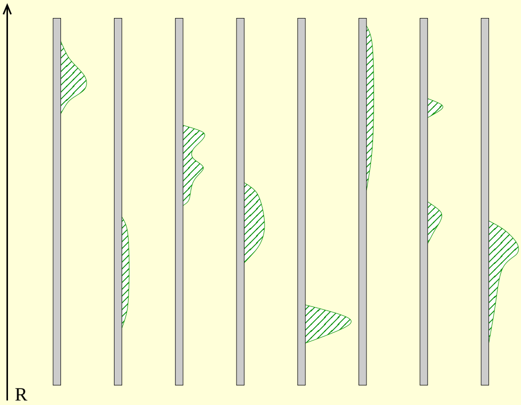
## Outline

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

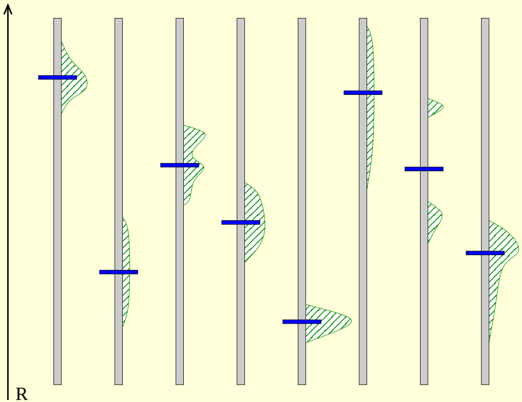
# Stochastic Bandits and Subset Selection



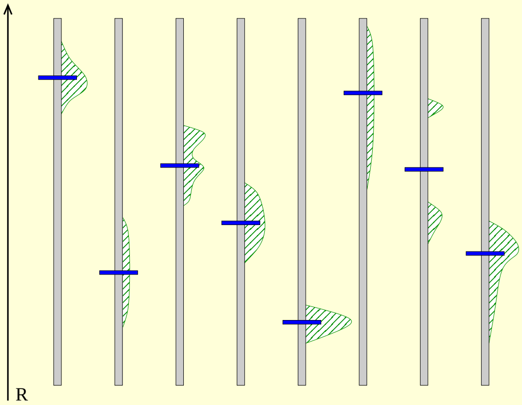
# Stochastic Bandits and Subset Selection



# Stochastic Bandits and Subset Selection



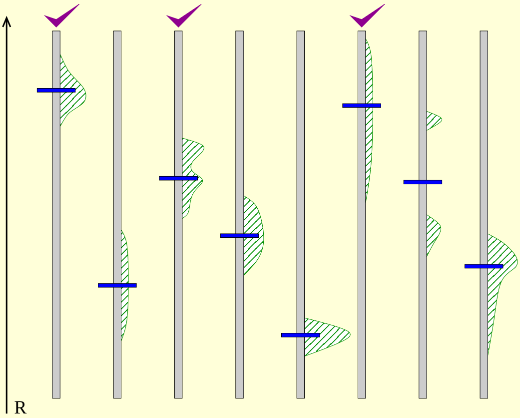
# Stochastic Bandits and Subset Selection




In an $n$-armed bandit:

find the $m$ arms with the highest means

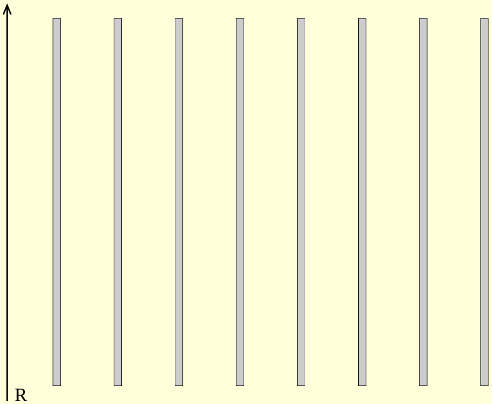
# Stochastic Bandits and Subset Selection




In an $n$-armed bandit:

find the $m$ arms with the highest means

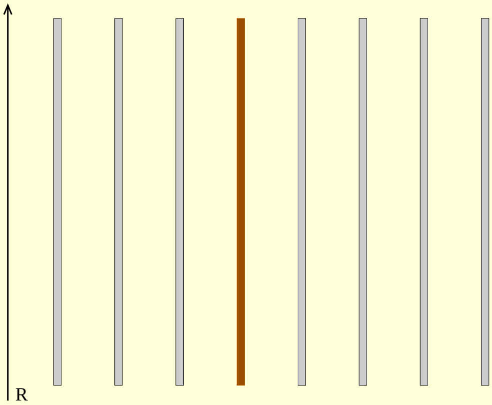
# Stochastic Bandits and Subset Selection




In an *n*-armed bandit:

find the *m* arms with the highest means

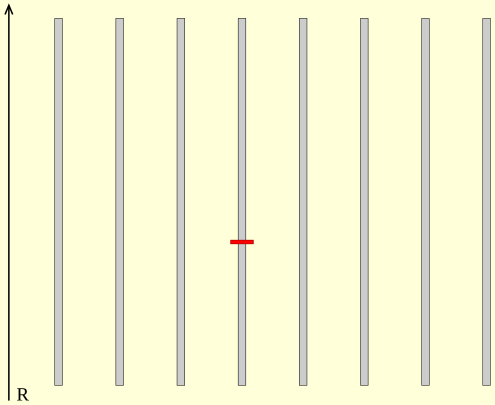
# Stochastic Bandits and Subset Selection




In an $n$-armed bandit:

find the $m$ arms with the highest means

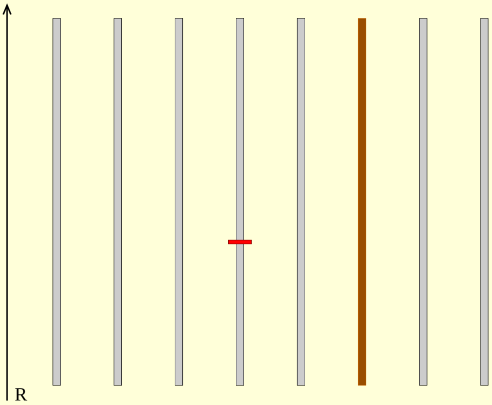
# Stochastic Bandits and Subset Selection




In an $n$-armed bandit:

find the $m$ arms with the highest means

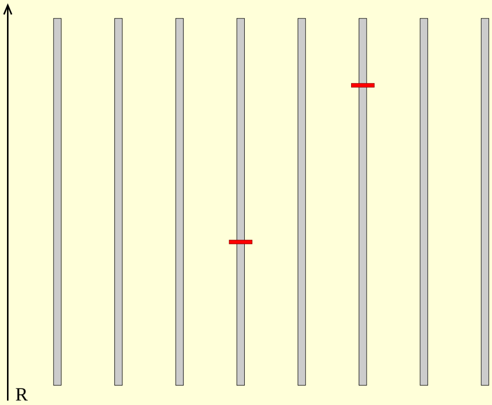
# Stochastic Bandits and Subset Selection




In an *n*-armed bandit:

find the *m* arms with the highest means

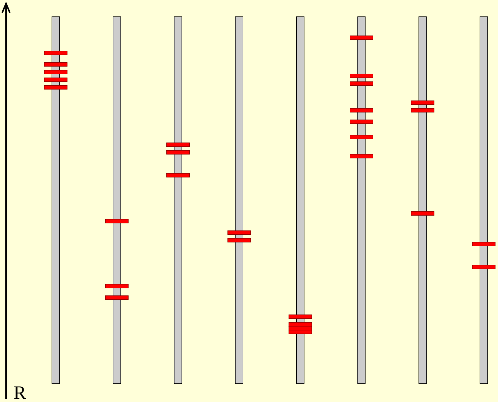
# Stochastic Bandits and Subset Selection




In an $n$-armed bandit:

find the $m$ arms with the highest means

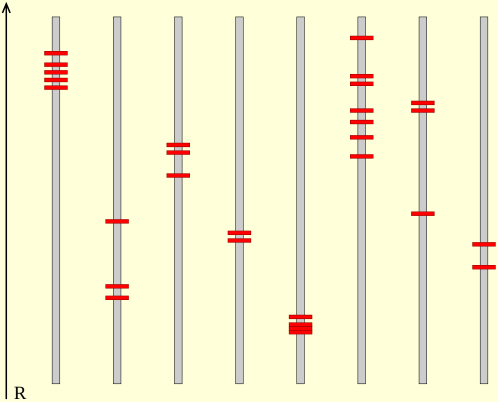
# Stochastic Bandits and Subset Selection




In an $n$-armed bandit:

find the $m$ arms with the highest means

# Stochastic Bandits and Subset Selection

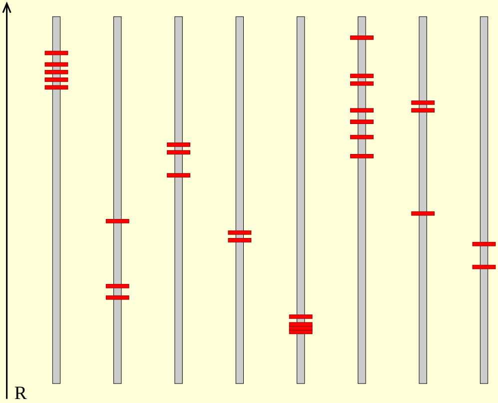



In an $n$-armed bandit:

find the $m$ arms with the highest means

with high probability

# Stochastic Bandits and Subset Selection



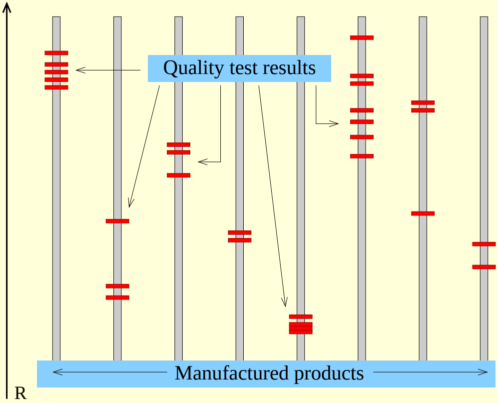

In an *n*-armed bandit:

find the *m* arms with the highest means

with high probability

using a *minimal* number of samples.

# Stochastic Bandits and Subset Selection



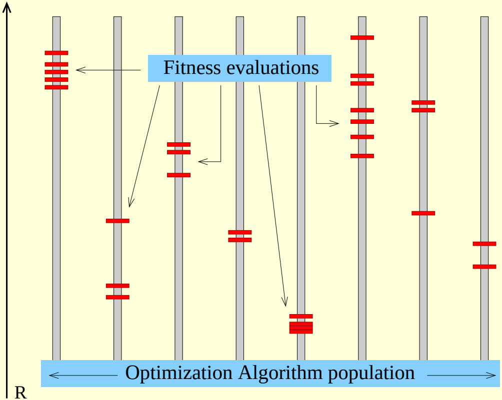

In an $n$-armed bandit:

find the $m$ arms with the highest means

with high probability

using a *minimal* number of samples.

# Stochastic Bandits and Subset Selection



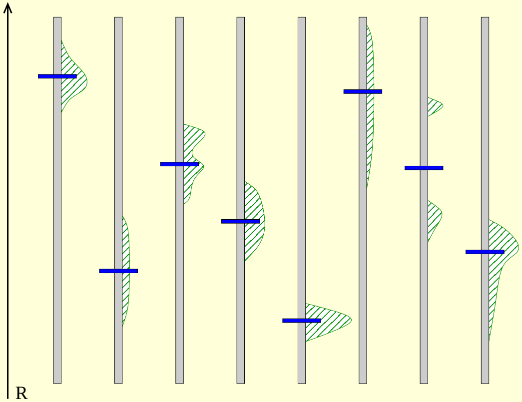

In an $n$-armed bandit:

find the $m$ arms with the highest means

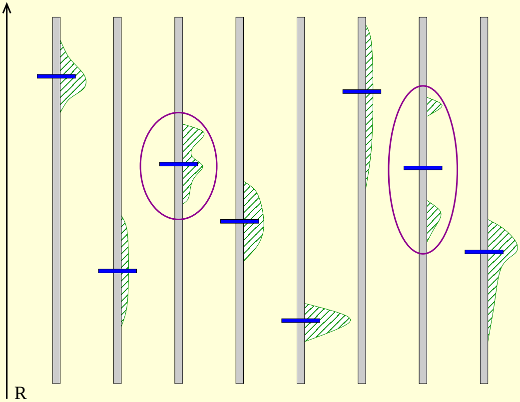
with high probability

using a *minimal* number of samples.

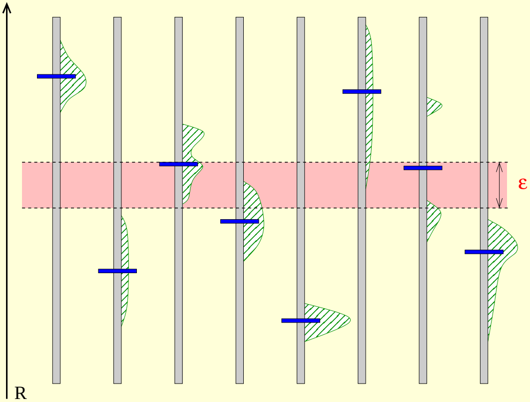
# PAC Formulation



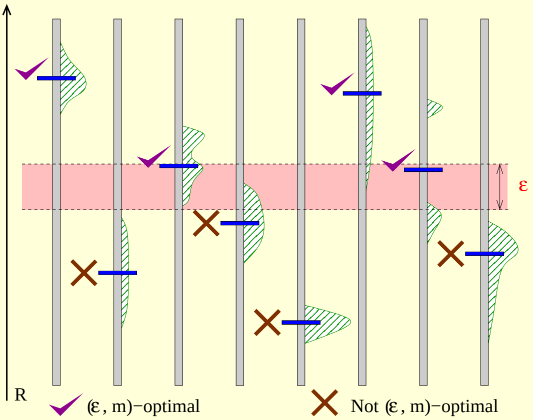
# PAC Formulation



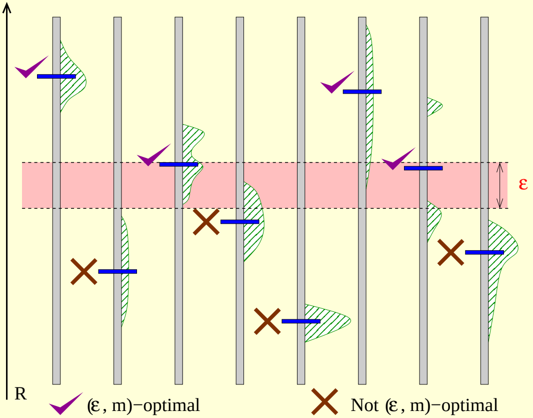
# PAC Formulation



# PAC Formulation



# PAC Formulation



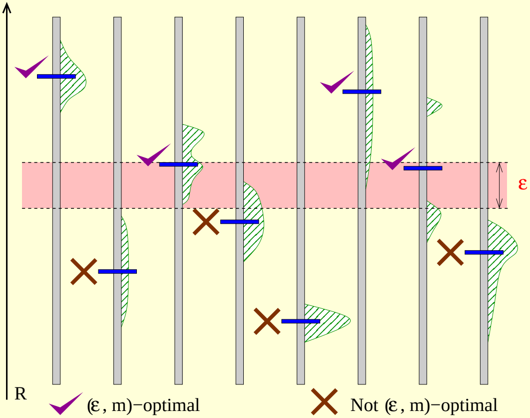

In an *n*-armed bandit:

find *m* $(\epsilon, m)$-optimal arms

with probability at least $1 - \delta$

using a minimal number of samples.

# PAC Formulation




In an $n$-armed bandit:

find $m$ $(\epsilon, m)$-optimal arms

with probability at least $1 - \delta$

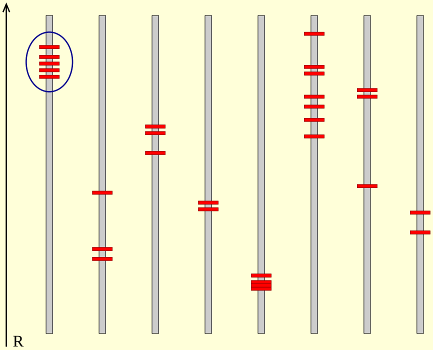
using a minimal number of samples.

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

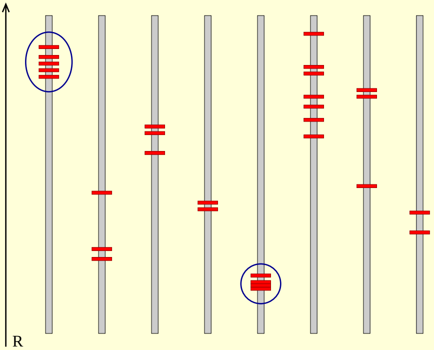
# Related Work

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

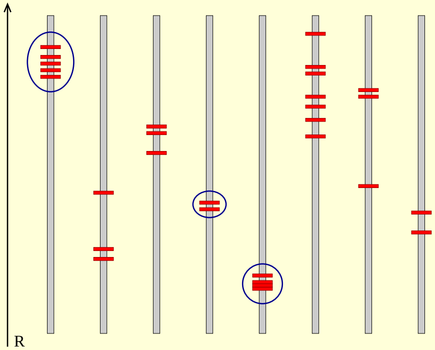
# Confidence Bounds on the Mean



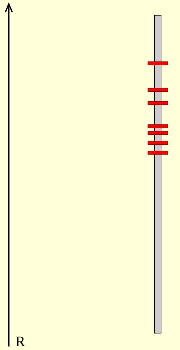
# Confidence Bounds on the Mean



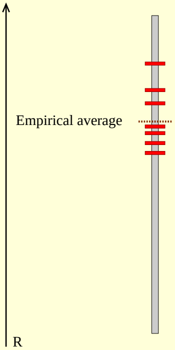
# Confidence Bounds on the Mean



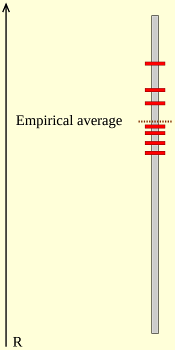
# Confidence Bounds on the Mean



# Confidence Bounds on the Mean



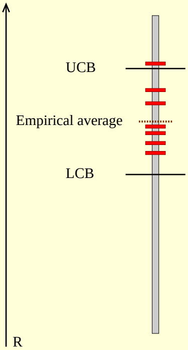
# Confidence Bounds on the Mean




- Hoeffding's inequality (Hoeffding, 1963): With probability at least $1 - \delta$:

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

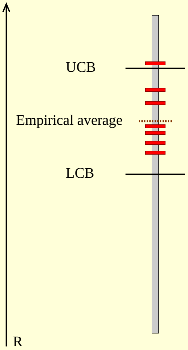
# Confidence Bounds on the Mean




- Hoeffding's inequality (Hoeffding, 1963): 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



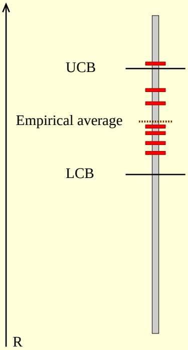

- Hoeffding's inequality (Hoeffding, 1963): With probability at least $1 - \delta$:

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

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

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

# Confidence Bounds on the Mean




- Hoeffding's inequality (Hoeffding, 1963): With probability at least $1 - \delta$:

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

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

- 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.
  $\vdots$
  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:

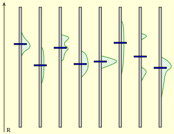
  Sample each arm $u_1(n, m, \epsilon, \delta)$ times.
  Discard half the arms with lower empirical averages.
  Sample each remaining arm $u_2(n, m, \epsilon, \delta)$ times.
  Discard half the remaining arms with lower empirical averages.
  ⋮
  Until $m$ arms remain.

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

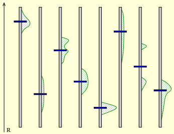
- **Lower bound**: There exist bandit instances (with Bernoulli arms) on which any PAC algorithm needs at least $\Omega\left(\frac{n}{\epsilon^2} log \left(\frac{m}{\delta}\right)\right)$ samples.

# Problem Complexity

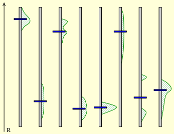
Instance 1




Instance 2

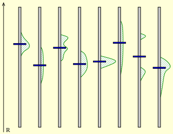


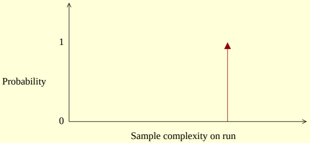

Instance 3



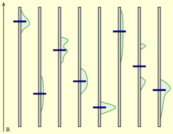
# Problem Complexity

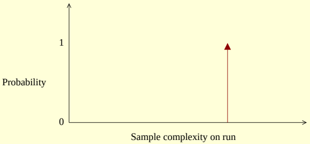
Instance 1



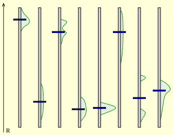



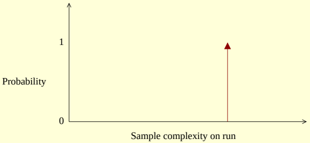

Instance 2



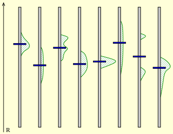

Instance 3

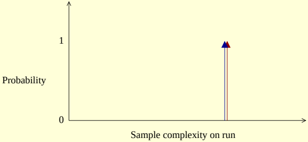


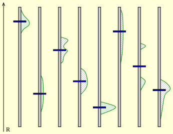
# Problem Complexity

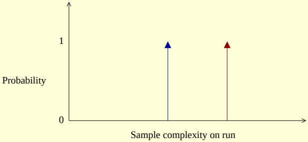
Instance 1



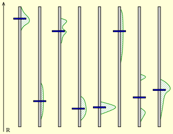



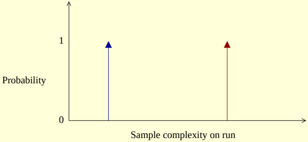

Instance 2



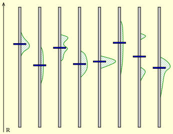

Instance 3

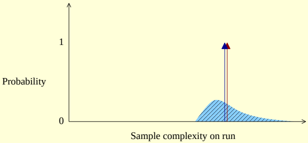


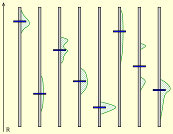
# Problem Complexity

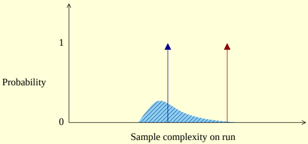
Instance 1



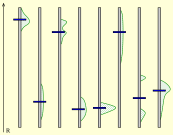



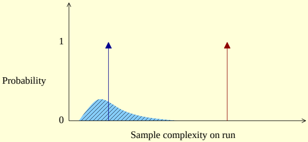

Instance 2




Instance 3



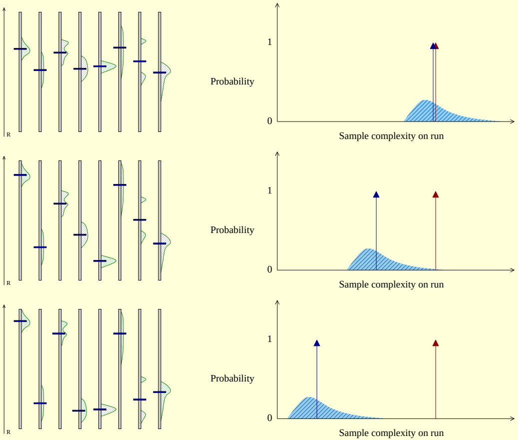



# Problem Complexity

Instance 1

Instance 2

Instance 3




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

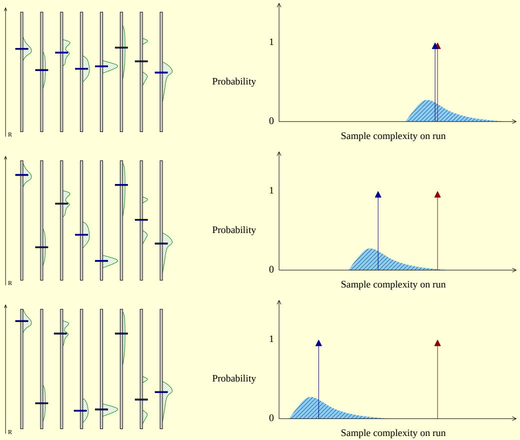
$$\mathbf{H}^{\epsilon} = \sum_{a=1}^{n} \frac{1}{\max\left\{\Delta_a, \frac{\epsilon}{2}\right\}^2}.$$

# 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

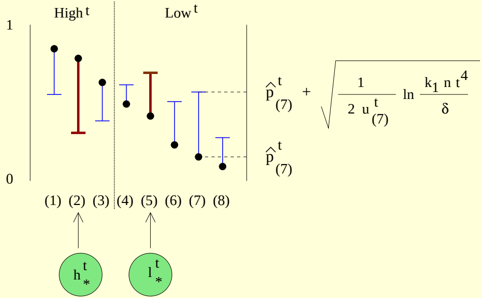
Achieves PAC guarantee.

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:

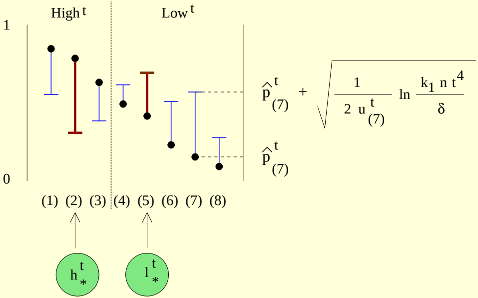
On round $t$: sample arms $h^t_*$ and $l^t_*$.

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

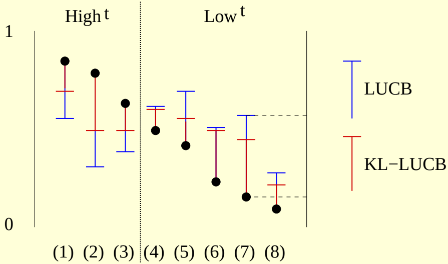
LUCB lower bound = $\hat{p}_a^t - \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{knt^\alpha}{\delta}\right)}$.

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

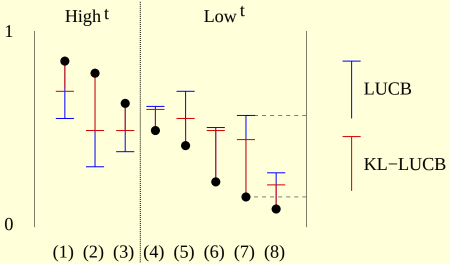
LUCB lower bound = $\hat{p}_a^t - \sqrt{\frac{1}{2u_a^t} \ln\left(\frac{knt^\alpha}{\delta}\right)}$.

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 $KL(z, x) = KL(z, y)$.

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

# 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 $KL(z,x) = KL(z,y)$.

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

Expected-sample-complexity lower bounds fresh off the press!

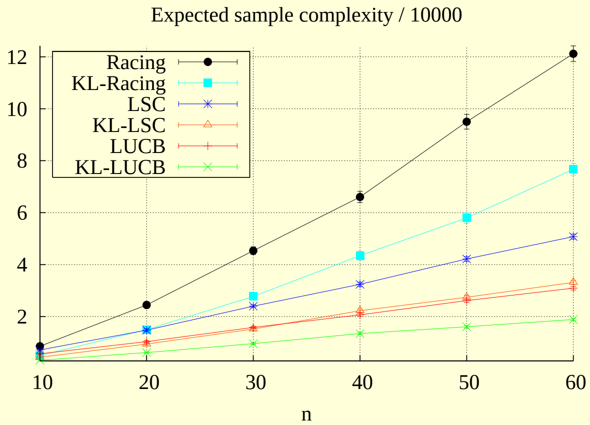
On the Complexity of Best Arm Identification in Multi-Armed Bandit Models
Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier, *2014*.

# Experiments

- We compare (KL-)LUCB, (KL-)Racing, and (KL-)LSC.
- **Racing algorithm** (Heidrich-Meisner and Igel, 2009)
  - Each arm is one of three sets: *Selected*, *Discarded*, *Remaining*.
  - Initially, place all the arms in *Remaining*.
  - In each phase, sample all the arms in *Remaining*. If some arm confidently exceeds $n - m$ others, move it to *Selected*. If some arm confidently is exceeded by $m$ others, move it to *Discarded*.
  - Stop and return *Selected* if it has at least $m$ arms; else go to next phase.
- **LSC algorithm**
  - Akin to LUCB.
  - At each time $t$, among the arms in $High^t$ and $Low^t$ that collide, pick one that has been *sampled the least number of times*.
  - Stop if $High^t$ and $Low^t$ do not collide.
- (KL-)LUCB and (KL-)LSC are "fully sequential", whereas (KL-)Racing is not.

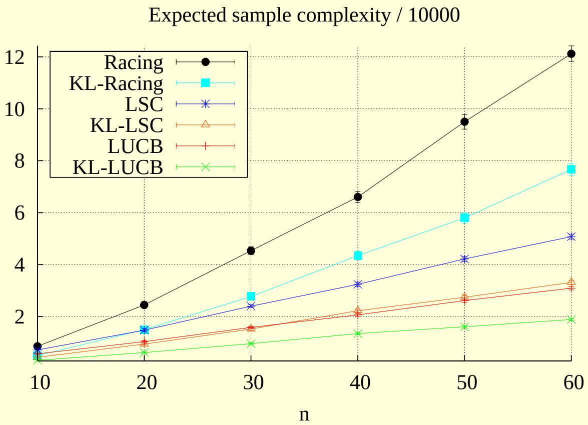
# Experiments

- 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

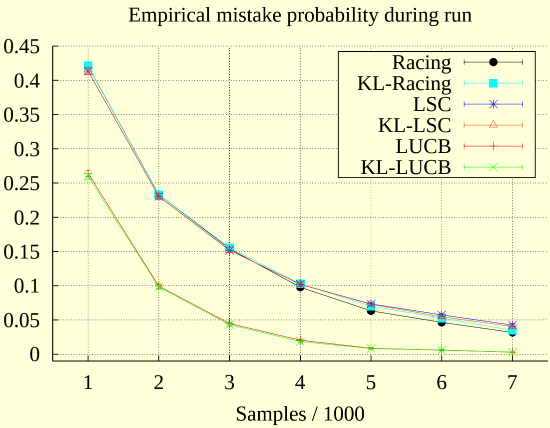
- Number of arms $n$ varied.
- 1000 random instances; each arm's mean drawn uniformly at random from [0, 1].
- $m = \frac{n}{5}, \epsilon = 0.1, \delta = 0.1$.




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

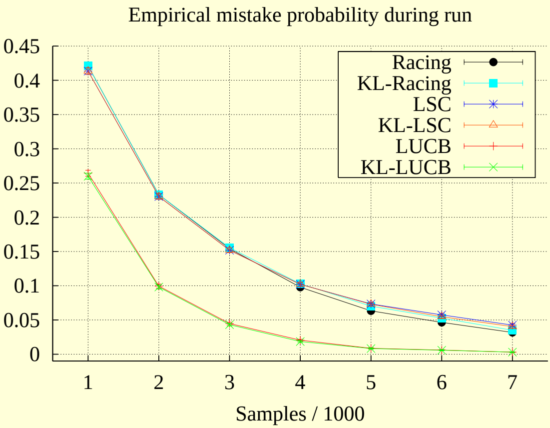
# Experiments

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



# Experiments

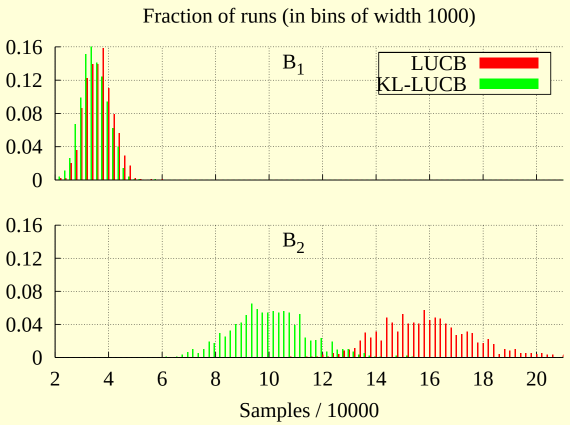
- Instance $B_1$: $n = 15, p_1 = \frac{1}{2}; p_a = \frac{1}{2} - \frac{a}{40}, a = 2, 3, \ldots, n$.
- $m = 3, \epsilon = 0.04, \delta = 0.1$.




(KL-)LUCB separates out arms more quickly.

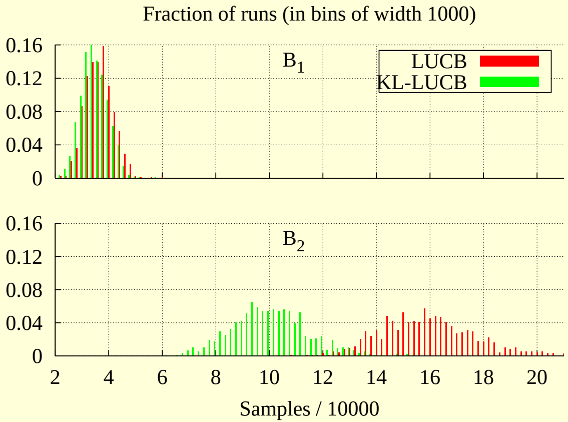
# Experiments

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

# Summary

PAC subset selection

$n, m, \epsilon, \delta$

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

## Summary

PAC subset selection

$n, m, \epsilon, \delta$

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

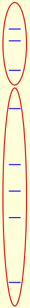
Use KL-LUCB for PAC subset selection!

# Future Work

- Generalized ranking and selection

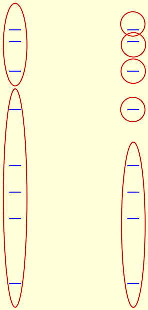
# Future Work

- Generalized ranking and selection



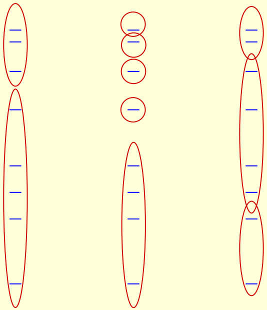
# Future Work

- Generalized ranking and selection



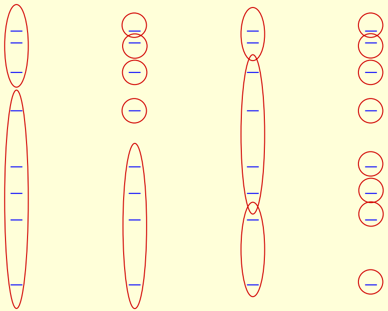
# Future Work

- Generalized ranking and selection



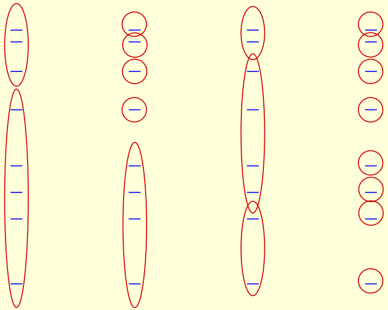
# Future Work

- Generalized ranking and selection



# Future Work

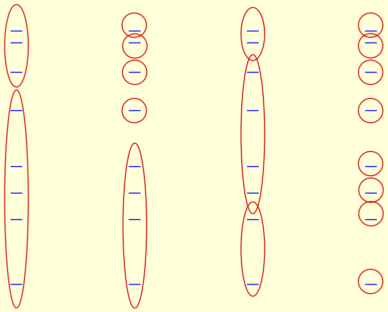
- Generalized ranking and selection



- Exploration in MDPs with instance-specific sample complexity bounds

# Future Work

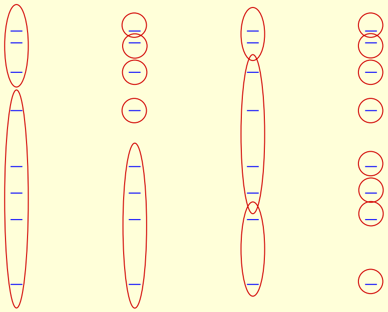
- Generalized ranking and selection



- Exploration in MDPs with instance-specific sample complexity bounds
- Sampling *pairwise* preferences to pick a winner (social choice).

# Future Work

- Generalized ranking and selection



- Exploration in MDPs with instance-specific sample complexity bounds
- Sampling *pairwise* preferences to pick a winner (social choice).

Thank you!

# References

**Herbert Robbins, 1952**. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5): 527–535, 1952.

**Wassily Hoeffding, 1963**. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301): 13–30, 1963.

**Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer, 2002**. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, 47(2–3):235–256, 2002.

**Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire, 2002**. The Nonstochastic Multiarmed Bandit Problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

**Eyal Even-Dar, Shie Mannor, and Yishay Mansour, 2006**. Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems. *Journal of Machine Learning Research*, 7: 1079–1105, 2006.

**Sandeep Pandey, Deepayan Chakrabarti, and Deepak Agarwal, 2007.** Multi-armed bandit problems with dependent arms. In *Proceedings of the Twenty-Fourth International Conference on Machine Learning (ICML 2007)*, pp. 721–728, ACM, 2007.

**Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal, 2008**. Multi-armed bandits in metric spaces. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC 2008)*, pp. 681–690, ACM, 2008.

**Verena Heidrich-Meisner and Christian Igel, 2009**. Hoeffding and Bernstein races for selecting policies in evolutionary direct policy search. In *Proceedings of the Twenty-sixth International Conference on Machine Learning (ICML 2009)*, pp. 401–408, ACM, 2009.

**Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos, 2010**. Best Arm Identification in Multi-Armed Bandits. In *Proceedings of the Twenty-third Conference on Learning Theory (COLT 2010)*, pp. 41–53, Omnipress, 2010.

**Shivaram Kalyanakrishnan and Peter Stone, 2010**. Efficient Selection of Multiple Bandit Arms: Theory and Practice. In *Proceedings of the Twenty-seventh International Conference on Machine Learning (ICML 2010).*, pp. 511–518, Omnipress, 2010.

**Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone, 2012**. PAC Subset Selection in Stochastic Multi-armed Bandits. In *Proceedings of the Twenty-ninth International Conference on Machine Learning (ICML 2012)*, pp. 655–662, Omnipress, 2012.

**Emilie Kaufmann and Shivaram Kalyanakrishnan, 2013**. Information Complexity in Bandit Subset Selection. *JMLR Workshop and Conference Proceedings (Conference on Learning Theory, 2013)*, 30:228–251, 2013.

**Emilie Kaufmann, Aurélien Garivier and Olivier Cappé, 2014**. On the Complexity of Best Arm Identication in Multi-Armed Bandit Models. `http://perso.telecom-paristech.fr/~kaufmann/KCG14.pdf`, 2014.