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
In an $n$-armed bandit:
find the $m$ arms with the highest means
In an \( n \)-armed bandit:

find the \( m \) arms with the highest means
In an $n$-armed bandit:

find the $m$ arms with the highest means
In an $n$-armed bandit:
find the $m$ arms with the highest means
In an $n$-armed bandit:

- find the $m$ arms with the highest means
In an $n$-armed bandit:

find the $m$ arms with the highest means
In an $n$-armed bandit:
find the $m$ arms with the highest means
In an $n$-armed bandit:
find the $m$ arms with the highest means
In an $n$-armed bandit:

- find the $m$ arms with the highest means
- for a given confidence
In an $n$-armed bandit:

- find the $m$ arms with the highest means
- for a given confidence
- using a minimal number of samples.
In an \( n \)-armed bandit:

- find the \( m \) arms with the highest means
- for a given confidence
- using a minimal number of samples.
In an $n$-armed bandit:

find the $m$ arms with the highest means
for a given confidence
using a minimal number of samples.
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

\[ R \]
PAC Formulation

\[ R \]
PAC Formulation

R

ε
PAC Formulation

\[ \epsilon \text{−optimal} \]

Kalyanakrishnan, Tewari, Auer, and Stone (2012)
In an \( n \)-armed bandit:

- find \( m (\epsilon, m) \)-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 Diagram](image-url)
Confidence Bounds on the Mean
Confidence Bounds on the Mean
Confidence Bounds on the Mean

\[ R \]

\[ \text{Figure: Graph showing confidence bounds on the mean.} \]
Confidence Bounds on the Mean

Empirical average

R
Confidence Bounds on the Mean

Empirical average

LB(u, δ)

R
Confidence Bounds on the Mean

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

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

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

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

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

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

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

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

- **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 $O\left( \frac{n}{\epsilon^2} \log \left( \frac{m}{\delta} \right) \right)$.
- Optimal up to a constant factor.


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

\[ H^{\epsilon/2} = \sum_{a=1}^{n} \frac{1}{\max\{\Delta_a, \frac{\epsilon}{2}\}^2}. \]
Algorithms for Subset Selection (contd.)

- LUCB Algorithm:
  Achieves PAC guarantee.

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

Stopping rule: Terminate iff

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

Sampling strategy:

On round $t$: sample arms $h_{*}^t$ and $l_{*}^t$. 

Kalyanakrishnan, Tewari, Auer, and Stone (2012)
Thank you!