Qualitative Multi-Armed Bandits: A Quantile-Based Approach

Balázs Szörényi\textsuperscript{1,2}, Róbert Busa-Fekete\textsuperscript{3}, Paul Weng\textsuperscript{4,5}, Eyke Hüllermeier\textsuperscript{3}

\textsuperscript{1}Department of Electrical Engineering, The Technion - Israel Institute of Technology, Haifa, Israel 32000
\textsuperscript{2}MTA-SZTE Research Group on Artificial Intelligence, Tisza Lajos krt. 103., H-6720 Szeged, HUNGARY
\textsuperscript{3}Department of Computer Science, University of Paderborn, Warburger Str. 100, 33098 Paderborn, GERMANY
\textsuperscript{4}SYSU-CMU Joint Institute of Engineering, Guangzhou, China
\textsuperscript{5}SYSU-CMU Shunde International Joint Research Institute, Shunde, China

ICML
July 7, 2015
Outline of the talk

Motivation and problem formulation

Finite horizon case (PAC setting)

Infinite horizon case

Experiments

Concluding remarks
Motivation, problem formulation

Problem: headache

Solutions: Aspirin, Ibuprofen, Beer

Output: mild, severe, no pain
Motivation, problem formulation

Problem: headache

Solutions:

Output: mild, severe, no pain

How to compare the different options?
Motivation, problem formulation

Problem: headache

Solutions:
- ASPIRIN
- Ibuprofen
- Beer

Output: mild | severe | no pain

How to compare the different options?
Consider their quantiles
Formally

- \((L, \preceq)\) is an ordered set \((L \subset \mathbb{R})\)
- \(X_1, \ldots, X_K\): random variables (arms) taking values from \(L\)

EXAMPLE:
\(L = \{\text{no pain, mild, moderate, severe}\}\)
arms = Asprin, Ibuprofen, beer
Formally

- \((L, \prec)\) is an ordered set \((L \subset \mathbb{R})\)
- \(X_1, \ldots, X_K\): random variables (\textit{arms}) taking values from \(L\)

**EXAMPLE:**

\(L = \{\text{no pain, mild, moderate, severe}\}\)
arms = Asprin, Ibuprofen, beer

Arm \(k\) is \(\tau\)-\textit{optimal} if \(Q^{X_k}(\tau) = x^*\), where

- \(F^X\) is the cdf of random variable \(X\),
- \(Q^X(\tau) = \inf\{x \in L : \tau \leq F^X(x)\}\) is the \(\tau\)-\textit{quantile} of \(X\), and
- \(x^* = \max_{1 \leq k' \leq K} Q^{X_{k'}}(\tau)\)
Finite horizon case: PAC setting
PAC model

**GOAL** (informal): find a close-to-optimal arm with high probability
**PAC model**

**GOAL (informal):** find a close-to-optimal arm with high probability

Arm \( k \) is \((\epsilon, \tau)\)-optimal if
\[
Q^{X_k}(\tau + \epsilon) \geq x^*
\]
\(\Leftrightarrow\) slight perturbation of arm \( k \) could render it \( \tau \)-optimal
PAC model

**GOAL (informal):** find a close-to-optimal arm with high probability

Arm $k$ is $(\epsilon, \tau)$-optimal if $Q_{X_k}(\tau + \epsilon) \geq x^*$

(\Leftrightarrow \text{slight perturbation of arm } k \text{ could render it } \tau\text{-optimal})

**GOAL (formal):** with probability at least $(1 - \delta)$ output some $(\epsilon, \tau)$-optimal arm
Algorithm QPAC

Algorithm 1 QPAC($\delta$, $\epsilon$, $\tau$)

1: Set $\mathcal{A} = \{1, \ldots, K\}$ ▷ Active arms
2: $t = 1$
3: while $\mathcal{A} \neq \emptyset$ do
4:   for $k \in \mathcal{A}$ do
5:     Pull arm $k$ and observe $X_{k,t}$
6:     $x_t^+ = \max_{k \in \mathcal{A}} \hat{Q}_t^{X_k} (\tau + c_t (\frac{\delta}{K}))$
7:     $x_t^- = \max_{k \in \mathcal{A}} \hat{Q}_t^{X_k} (\tau - c_t (\frac{\delta}{K}))$
8:   for $k \in \mathcal{A}$ do
9:     if $\hat{Q}_t^{X_k} (\tau + \epsilon + c_t (\frac{\delta}{K})) < x_t^-$ then
10:        $\mathcal{A} = \mathcal{A} \setminus \{k\}$ ▷ Discard $k$
11:     if $x_t^+ \leq \hat{Q}_t^{X_k} (\tau + \epsilon - c_t (\frac{\delta}{K}))$ then
12:        $\hat{k} = k$ ▷ Select $k$
13:   BREAK
14: $t = t + 1$
15: return $\hat{k}$

Based on the elimination strategy by Even-Dar et al., 2002
Sample complexity

\[ \mathcal{K}_{\epsilon, \tau} \]: set of \((\epsilon, \tau)\)-optimal arms
\[ \Delta^\epsilon_k = \sup \left\{ \Delta \in [0, 1] \mid Q^{X_k}(\tau + \epsilon + \Delta) < \max_{h \in \mathcal{K}_{\epsilon, \tau}} Q^{X_h}(\tau - \Delta) \right\} \]

QPAC SAMPLE COMPLEXITY BOUND:
\[ O \left( \sum_{k=1}^{K} \frac{1}{(\epsilon \vee \Delta^\epsilon_k)^2} \log \frac{K}{(\epsilon \vee \Delta^\epsilon_k)\delta} \right) \]

GENERAL SAMPLE COMPLEXITY LOWER BOUND:
\[ \Omega \left( \sum_{k=1}^{K} \frac{1}{(\epsilon \vee \Delta^\epsilon_k)^2} \log \frac{1}{\delta} \right) \]
Infinite horizon case: regret minimization
Regret minimization

**GOAL:** minimize the expected cumulative regret $R_t$, where...

- $G \subset L$: set of "good" outcomes
- $\text{regret}(x, y) = \mathbb{I}\{x \in G\} - \mathbb{I}\{y \in G\}$
- In round $t$, choosing arm $k_t$ incurs immediate regret $r_t = \rho_{k_t}$
- $\mathbb{E}[\text{regret}(X'_{k_t}, X_{k_t})] = \max_{k' = 1, \ldots, K} \mathbb{P}[X_{k'} \in G] - \mathbb{P}[X_k \in G]$
- $R_t = \mathbb{E}[\sum_{t'=1}^{t} r_t]$: expected cumulative regret

How to set $G$?
Regret minimization

**GOAL:** minimize the expected cumulative regret $R_t$, where...

- $G \subset L$: set of “good” outcomes
- regret$(x, y) = \mathbb{I}\{x \in G\} - \mathbb{I}\{y \in G\}$
- In round $t$, choosing arm $k_t$ incurs immediate regret $r_t = \rho_{k_t}$ where

\[
\rho_k = \max_{k' = 1, \ldots, K} \mathbb{E}[\text{regret}(X'_k, X_k)] \\
= \max_{k' = 1, \ldots, K} \mathbb{P}[X_{k'} \in G] - \mathbb{P}[X_k \in G]
\]

- $R_t = \mathbb{E}\left[\sum_{t'=1}^{t} r_t\right]$: expected cumulative regret
Regret minimization

GOAL: minimize the expected cumulative regret $R_t$, where...

- $G \subset L$: set of “good” outcomes
- $\text{regret}(x, y) = \mathbb{I}\{x \in G\} - \mathbb{I}\{y \in G\}$
- In round $t$, choosing arm $k_t$ incurs immediate regret $r_t = \rho_{k_t}$
  where

$$
\rho_k = \max_{k' = 1,\ldots,K} \mathbb{E}[\text{regret}(X'_{k'}, X_k)]
= \max_{k' = 1,\ldots,K} \mathbb{P}[X_{k'} \in G] - \mathbb{P}[X_k \in G]
$$

- $R_t = \mathbb{E}\left[\sum_{t'=1}^{t} r_t\right]$ : expected cumulative regret

How to set $G$?
How to set $G$?

- Natural candidate: $G := \{x \in L : x \succeq x^*\}$,
  Problematic regret bounds
How to set $G$?

- Natural candidate: $G := \{ x \in L : x \succeq x^* \}$.
- Problematic regret bounds

- Alternatively:

\[
x^*(\tau') = \max_{k=1,\ldots,K} Q^X_k(\tau') \text{ for } \tau' \in [0, 1]
\]

\[
G = L_\tau = \{ x \in L : x \succeq x^*(\tau') \text{ for some } \tau' > \tau \}.
\]

Accordingly, the best arm is

\[
k^* = \arg\max_{1 \leq k \leq K} \mathbb{P}[X_k \in L_\tau].
\]
Based on UCB1 (Auer et al., 2002)

Doubled optimism in the face of uncertainty

\[ p^X(x) = \mathbb{P}[X < x] \]
Expected regret bounds

- $\Delta_0 = \min_{k'} \mathbb{P}[X_{k'} \leq \inf L_\tau] - \tau$
- Suboptimal arm $k$: $0 < \mathbb{P}[X_k \notin L_\tau] - \tau =: \Delta_k$
- Optimal arm $k$: $\Delta_k := \min(\rho_k, \Delta_0)$

**QUCB Regret Bound:**

$$R_t = O \left( \sum_{k: \Delta_k > 0} \frac{\rho_k}{(\Delta_k)^2} \log t \right).$$

**General Regret Lower Bound:**

$$R_t \geq \min \left( c_1 t, c_2 \sum_{k: \rho_k > 0} \frac{\rho_k}{(\Delta_k)^2} \log t \right)$$
for some $c_1, c_2 > 0$
Experiments
Finite horizon case

Samp. complexity

- QPAC ($\epsilon = 0.01$)
- SE ($\epsilon = 0.01$)
- QPAC ($\epsilon = 0.03$)
- SE ($\epsilon = 0.03$)
- QPAC ($\epsilon = 0.05$)
- SE ($\epsilon = 0.05$)
Infinite horizon case

\[ \tau = 0.9 \]

\[ \tau = 0.5 \]
Concluding remarks

- Formulation of a qualitative MAB problem for both the infinite and finite horizon case
- Algorithmic solutions
- (Almost) matching lower bounds

TO DO: thorough investigation of the possible regret notions
Thank you for the attention