Finite-time Analysis of Frequentist Strategies for Multi-armed Bandits

Subhojyoti Mukherjee
CS15S300
Guide: Dr. Balaraman Ravindran
Co-Guide: Dr. Nandan Sudarsanam
Collaborator: Dr. K.P. Naveen

IIT Madras

January 8, 2018

Overview

SMAB Setting (Part 1)	TBP Setting (Part 2)
• Problem Definition • Contributions • EUCBV algorithm • Theory • Experiments	• Problem Definition • Contributions • AugUCB algorithm • Theory • Experiments

Stochastic Multi-Armed Bandit Problem (SMAB) (Chapter 2)

- A finite set of actions or arms belonging to set $\mathbb{A}$ such that $|\mathbb{A}| = K$.

Stochastic Multi-Armed Bandit Problem (SMAB) (Chapter 2)

- A finite set of actions or arms belonging to set $\mathbb{A}$ such that $|\mathbb{A}| = K$.
- The rewards for each of the arms are i.i.d random variables drawn from distribution specific to the arm which are fixed throughout the time horizon denoted by T.

Stochastic Multi-Armed Bandit Problem (SMAB) (Chapter 2)

- A finite set of actions or arms belonging to set $\mathbb{A}$ such that $|\mathbb{A}| = K$.
- The rewards for each of the arms are i.i.d random variables drawn from distribution specific to the arm which are fixed throughout the time horizon denoted by T.
- The learner does not know the mean $r_i, \forall i \in \mathbb{A}$ of the distribution or the variance σ_i^2.

Problem Definition of SMAB (Chapter 2)

- **Primary aim:** Minimize the cumulative regret by quickly identifying the arm whose expected mean is r^* such that $r^* > r_i, \forall i \in \mathbb{A}$.

Problem Definition of SMAB (Chapter 2)

- **Primary aim:** Minimize the cumulative regret by quickly identifying the arm whose expected mean is r^* such that $r^* > r_i, \forall i \in \mathbb{A}$.
- **Condition:** This has to be achieved within a finite T timesteps.

Problem Definition of SMAB (Chapter 2)

- **Primary aim:** Minimize the cumulative regret by quickly identifying the arm whose expected mean is r^* such that $r^* > r_i, \forall i \in \mathbb{A}$.
- **Condition:** This has to be achieved within a finite T timesteps.
- The expected regret of an algorithm after T timesteps is give by,

$$\mathbb{E}[R_T] = \sum_{i=1}^{K} \mathbb{E}[z_i(T)]\Delta_i,$$

where $\Delta_i = r^* - r_i$ is the gap.

Contributions in SMAB (Chapter 2)

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.

Contributions in SMAB (Chapter 2)

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.

Contributions in SMAB (Chapter 2)

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.

Contributions in SMAB (Chapter 2)

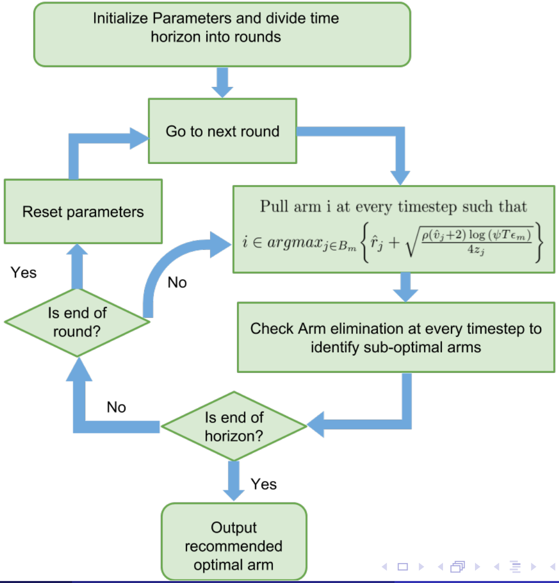
- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.
- It addresses an open problem discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.

Contributions in SMAB (Chapter 2)

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.
- It addresses an open problem discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.
- Theoretically it achieves an order-optimal regret bound, the first for an arm elimination algorithm in SMAB setting.

Contributions in SMAB (Chapter 2)

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.
- It addresses an open problem discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.
- Theoretically it achieves an order-optimal regret bound, the first for an arm elimination algorithm in SMAB setting.
- Empirically, it outperforms all the state-of-the-art algorithms for the considered environments.


Initialize Parameters and divide time horizon into rounds
Go to next round
Pull arm i at every timestep such that
$i \in argmax_{j \in B_m} \left\{ \hat{r}_j + \sqrt{\frac{\rho(\hat{v}_j+2)\log(\psi T \epsilon_m)}{4z_j}} \right\}$
Check Arm elimination at every timestep to identify sub-optimal arms
Is end of horizon?
Yes
Output recommended optimal arm
No
Is end of round?
No
Yes
Reset parameters

Expected Regret of EUCBV (Chapter 3)

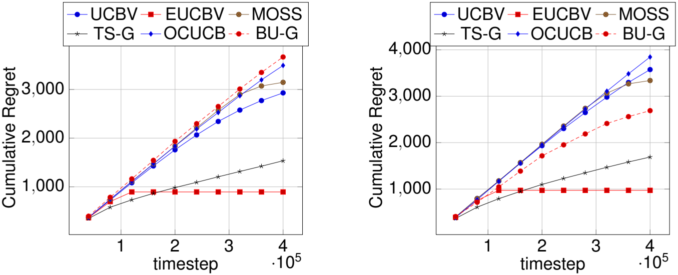
Corollary (***Gap-Independent Bound***)

The regret of EUCBV is upper bounded by the following gap-independent expression:

$$\mathbb{E}[R_T] \leq \frac{C_3 K^5}{T^{\frac{1}{4}}} + 80\sqrt{KT}.$$

Algorithm	GD Bound	GI Bound	Var
EUCBV	$O\left(\frac{K\sigma_{\max}^2 \log(\frac{T\Delta^2}{K})}{\Delta}\right)$	$O\left(\sqrt{KT}\right)$	Yes
UCBV	$O\left(\frac{K\sigma_{\max}^2 \log T}{\Delta}\right)$	$O\left(\sqrt{KT \log T}\right)$	Yes
MOSS	$O\left(\frac{K^2 \log(T\Delta^2/K)}{\Delta}\right)$	$O\left(\sqrt{KT}\right)$	No
OCUCB	$O\left(\frac{K \log(T/H_i)}{\Delta}\right)$	$O\left(\sqrt{KT}\right)$	No

Experiments in SMAB (Chapter 3)



(a) Expt-3: Failure of TS (b) Expt-4: 3 Group Variance

- **Setting:** The Thresholding Bandit Problem (TBP) is a special case of SMAB setting.
- **Primary aim:** Identify all the arms whose mean of the reward distribution (r_i) is above a particular threshold τ given as input.

Problem Definition of TBP (Chapter 4)

- **Setting:** The Thresholding Bandit Problem (TBP) is a special case of SMAB setting.
- **Primary aim:** Identify all the arms whose mean of the reward distribution (r_i) is above a particular threshold τ given as input.
- **Condition:** This has to be achieved within T timesteps of exploration and this is termed as a fixed-budget problem.

Problem Definition of TBP (Chapter 4)

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.

Problem Definition of TBP (Chapter 4)

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.
- S_τ^c denote the complement of S_τ, i.e., $S_\tau^c = \{i \in \mathcal{A} : r_i < \tau\}$.

Problem Definition of TBP (Chapter 4)

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.
- S_τ^c denote the complement of S_τ, i.e., $S_\tau^c = \{i \in \mathcal{A} : r_i < \tau\}$.
- Let $\hat{S}_\tau$ denote the recommendation of a learning algorithm after T time units of exploration, while $\hat{S}_\tau^c$ denotes its complement.

Problem Definition of TBP (Chapter 4)

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.
- S_τ^c denote the complement of S_τ, i.e., $S_\tau^c = \{i \in \mathcal{A} : r_i < \tau\}$.
- Let $\hat{S}_\tau$ denote the recommendation of a learning algorithm after T time units of exploration, while $\hat{S}_\tau^c$ denotes its complement.
- The goal of the learning agent is to minimize the expected loss:

$$\mathbb{E}[\mathcal{L}(T)] = \mathbb{P}\Big(\underbrace{\{S_\tau \cap \hat{S}_\tau^c \neq \emptyset\}}_{\textbf{Rejected good arms}} \cup \underbrace{\{\hat{S}_\tau \cap S_\tau^c \neq \emptyset\}}_{\textbf{Accepted bad arms}} \Big)$$

Contributions in TBP (Chapter 4)

- We propose the Augmented UCB (AugUCB) algorithm for the fixed-budget TBP setting.

Contributions in TBP (Chapter 4)

- We propose the Augmented UCB (AugUCB) algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the empirical variances of the arms along with mean estimates.

Contributions in TBP (Chapter 4)

- We propose the Augmented UCB (AugUCB) algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the empirical variances of the arms along with mean estimates.
- It is the first variance-based arm elimination algorithm for the considered TBP settings.

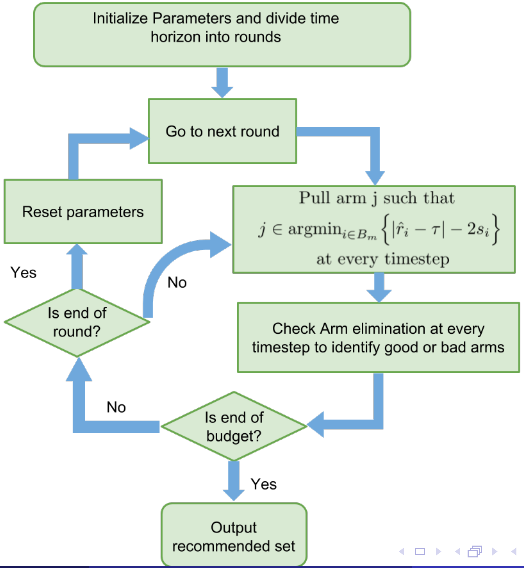
Contributions in TBP (Chapter 4)

- We propose the Augmented UCB (AugUCB) algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the empirical variances of the arms along with mean estimates.
- It is the first variance-based arm elimination algorithm for the considered TBP settings.
- We also define a new problem complexity which uses empirical variance estimates along with arm's mean for giving the theoretical bound.

Contributions in TBP (Chapter 4)

- We propose the Augmented UCB (AugUCB) algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the empirical variances of the arms along with mean estimates.
- It is the first variance-based arm elimination algorithm for the considered TBP settings.
- We also define a new problem complexity which uses empirical variance estimates along with arm's mean for giving the theoretical bound.
- Empirically, in the considered environments AugUCB outperforms all the algorithms that rely only on mean estimation to conduct exploration.

AugUCB Algorithm for TBP (Chapter 5)



Problem Complexity in TBP (Chapter 5)

- We define $\Delta_i = |r_i - \tau|$ as in Locatelli et al. (2016).

Problem Complexity in TBP (Chapter 5)

- We define $\Delta_i = |r_i - \tau|$ as in Locatelli et al. (2016).
- We define $H_1 = \sum_{i=1}^{K} \frac{1}{\Delta_i^2}$ and $H_2 = \min_{i \in \mathcal{A}} \frac{i}{\Delta_{(i)}^2}$ as in Audibert and Bubeck (2010).

Problem Complexity in TBP (Chapter 5)

- We define $\Delta_i = |r_i - \tau|$ as in Locatelli et al. (2016).
- We define $H_1 = \sum_{i=1}^{K} \frac{1}{\Delta_i^2}$ and $H_2 = \min_{i \in \mathcal{A}} \frac{i}{\Delta_{(i)}^2}$ as in Audibert and Bubeck (2010).
- We define $H_{\sigma,1}$ (as in Gabillon et al. (2011)) that incorporates variances as:

$$H_{\sigma,1} = \sum_{i=1}^{K} \frac{\sigma_i + \sqrt{\sigma_i^2 + (16/3)\Delta_i}}{\Delta_i^2}.$$

Problem Complexity in TBP (Chapter 5)

- We define $\Delta_i = |r_i - \tau|$ as in Locatelli et al. (2016).
- We define $H_1 = \sum_{i=1}^{K} \frac{1}{\Delta_i^2}$ and $H_2 = \min_{i \in \mathcal{A}} \frac{i}{\Delta_{(i)}^2}$ as in Audibert and Bubeck (2010).
- We define $H_{\sigma,1}$ (as in Gabillon et al. (2011)) that incorporates variances as:

$$H_{\sigma,1} = \sum_{i=1}^{K} \frac{\sigma_i + \sqrt{\sigma_i^2 + (16/3)\Delta_i}}{\Delta_i^2}.$$

- Finally, analogous to H_2, we introduce $H_{\sigma,2}$, such that $H_{\sigma,2} = \max_{i \in \mathcal{A}} \frac{i}{\tilde{\Delta}_{(i)}^2}$, where $\tilde{\Delta}_i^2 = \frac{\Delta_i^2}{\sigma_i + \sqrt{\sigma_i^2 + (16/3)\Delta_i}}$, $(\tilde{\Delta}_{(i)})$ is an increasing ordering of $(\tilde{\Delta}_i)$.

Problem Complexity in TBP (Chapter 5)

- We define $\Delta_i = |r_i - \tau|$ as in Locatelli et al. (2016).
- We define $H_1 = \sum_{i=1}^{K} \dfrac{1}{\Delta_i^2}$ and $H_2 = \min_{i \in \mathcal{A}} \dfrac{i}{\Delta_{(i)}^2}$ as in Audibert and Bubeck (2010).
- We define $H_{\sigma,1}$ (as in Gabillon et al. (2011)) that incorporates variances as:

$$H_{\sigma,1} = \sum_{i=1}^{K} \frac{\sigma_i + \sqrt{\sigma_i^2 + (16/3)\Delta_i}}{\Delta_i^2}.$$

- Finally, analogous to H_2, we introduce $H_{\sigma,2}$, such that $H_{\sigma,2} = \max_{i \in \mathcal{A}} \frac{i}{\tilde{\Delta}_{(i)}^2}$, where $\tilde{\Delta}_i^2 = \frac{\Delta_i^2}{\sigma_i + \sqrt{\sigma_i^2 + (16/3)\Delta_i}}$, $(\tilde{\Delta}_{(i)})$ is an increasing ordering of $(\tilde{\Delta}_i)$.
- From Audibert and Bubeck (2010), we can show that

$$H_{\sigma,2} \leq H_{\sigma,1} \leq \log(2K) H_{\sigma,2}.$$

Expected Loss of AugUCB (Chapter 5)

Theorem

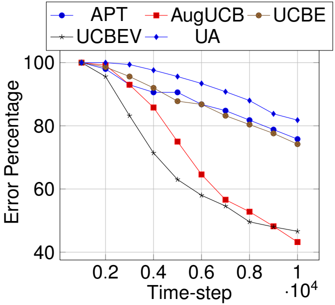
For $K \geq 4$ and $\rho = 1/3$, the expected loss of the AugUCB algorithm is given by,

$$\mathbb{E}[\mathcal{L}(T)] \leq 2KT \exp\left(- \frac{T}{4096 \log(K \log K) H_{\sigma,2}} \right).$$

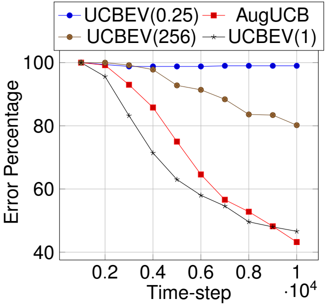
Table: AugUCB vs. State of the art

Algorithm	Upper Bound on Expected Loss	Oracle
AugUCB	$\exp\left(- \frac{T}{4096 \log(K \log K) H_{\sigma,2}} + \log(2KT) \right)$	No
UCBEV	$\exp\left(- \frac{1}{512} \frac{T-2K}{H_{\sigma,1}} + \log(6KT) \right)$	Yes
APT	$\exp\left(- \frac{T}{64H_1} + 2\log((\log(T) + 1)K) \right)$	No
UCBE	$\exp\left(- \frac{T-K}{18H_1} - 2\log(\log(T)K \right)$	Yes

Experiments in TBP (Chapter 5)



(c) Expt-1: Two Group Setting (Advance)



(d) Expt-2: Two Group Setting (Advance)

Conclusion (Chapter 6)

- We proposed the EUCBV algorithm for the SMAB setting which uses variance and mean estimation along with arm elimination to give an order-optimal theoretical guarantee.

Conclusion (Chapter 6)

- We proposed the EUCBV algorithm for the SMAB setting which uses variance and mean estimation along with arm elimination to give an order-optimal theoretical guarantee.
- We proposed the AugUCB algorithm for the fixed budget TBP which uses variance estimation and arm elimination to give improved theoretical and experimental guarantees than mean estimation based algorithms.

Conclusion (Chapter 6)

- We proposed the EUCBV algorithm for the SMAB setting which uses variance and mean estimation along with arm elimination to give an order-optimal theoretical guarantee.
- We proposed the AugUCB algorithm for the fixed budget TBP which uses variance estimation and arm elimination to give improved theoretical and experimental guarantees than mean estimation based algorithms.
- Further studies are required to establish a lower bound on the expected loss of AugUCB.

Conclusion (Chapter 6)

- We proposed the EUCBV algorithm for the SMAB setting which uses variance and mean estimation along with arm elimination to give an order-optimal theoretical guarantee.
- We proposed the AugUCB algorithm for the fixed budget TBP which uses variance estimation and arm elimination to give improved theoretical and experimental guarantees than mean estimation based algorithms.
- Further studies are required to establish a lower bound on the expected loss of AugUCB.
- A more detailed analysis of the non-uniform arm selection and parameter selection is also required for both AugUCB and EUCBV.

Papers based on Thesis

- Subhojyoti Mukherjee, K.P. Naveen, Nandan Sudarsanam, and Balaraman Ravindran, "*Thresholding Bandit with Augmented UCB*", *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017,2515-2521*.
- Subhojyoti Mukherjee, K.P. Naveen, Nandan Sudarsanam, and Balaraman Ravindran, "*Efficient UCBV: An Almost Optimal Algorithm using Variance Estimates*", *To appear in Proceedings of the Thirty-Second Association for the Advancement of Artificial Intelligence, AAAI 2018, New Orleans, Louisiana, USA, February 2-7*.

Thank You