Are sample means in multi-armed bandits positively or negatively biased?

Jaehyeok Shin¹, Aaditya Ramdas¹,² and Alessandro Rinaldo¹
Dept. of Statistics and Data Science¹, Machine Learning Dept.², CMU

Poster #12 @ Hall B + C
Stochastic multi-armed bandit

$\mu_1$  $\mu_2$  $\ldots$  $\mu_K$

$Y \sim \text{"Random reward"}$
Adaptive sampling scheme to maximize rewards / to identify the best arm.

Time

$\mu_1$  $\mu_2$  . . .  $\mu_K$
Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time

\[ t = 1 \]
Adaptive sampling scheme to maximize rewards / to identify the best arm

Time

\[ t = 1 \]
Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time

\[ t = 1 \]

\[ Y_1 \]
Adaptive sampling scheme
to maximize rewards / to identify the best arm

Time

\[ t = 1 \]

\[ t = 2 \]
Adaptive sampling scheme to maximize rewards / to identify the best arm

Time

\[ t = 1 \]

\[ Y_1 \]

\[ t = 2 \]
Adaptive sampling scheme

to maximize rewards / to identify the best arm

Time

\begin{align*}
t &= 1 \\
t &= 2
\end{align*}

\begin{align*}
Y_1 \\
Y_2
\end{align*}
Adaptive sampling scheme
to maximize rewards / to identify the best arm

\[
\begin{align*}
\text{Time} & \quad \mu_1 & \quad \mu_2 & \quad \ldots & \quad \mu_K \\
\text{ } & \quad \vdots & \quad \vdots & \quad \vdots & \quad \vdots \\
t = 1 & \quad Y_1 \\
t = 2 & \quad Y_2 \\
\vdots & \quad \vdots
\end{align*}
\]
Adaptive sampling scheme
to maximize rewards / to identify the best arm

\begin{align*}
\text{Time} & \quad \mu_1 \quad \mu_2 \quad \ldots \quad \mu_K \\
t = 1 & \quad Y_1 \\
t = 2 & \quad Y_2 \\
\vdots & \\
\mathcal{T} & \\
\text{Stopping time} & \end{align*}
Collected data can be used to identify an interesting arm...

Time

\[ t = 1 \]

\[ t = 2 \]

\[ \vdots \]

\[ t = T \]

\[ Y_1 \]

\[ Y_2 \]

Collected data can be used to identify an interesting arm...

"Interesting!"
...and data can be used to estimate the mean.

Time

\[ t = 1 \]

\[ Y_1 \]

\[ t = 2 \]

\[ Y_2 \]

\[ \vdots \]

\[ \mathcal{T} \]

\[ \hat{\mu}_\kappa(\mathcal{T}) \]

Sample mean of chosen arm \( \kappa \)
Q. Bias of sample mean?

\[ \mathbb{E} \left[ \hat{\mu}_\kappa(\mathcal{T}) - \mu_\kappa \right] \leq \textbf{or} \geq 0 ? \]
Nie et al. 2018

: Sample mean is **negatively** biased.

\[
\mathbb{E} \left[ \hat{\mu}_k(t) - \mu_k \right] \leq 0
\]
Nie et al. 2018: Sample mean is negatively biased.

\[ \mathbb{E} \left[ \hat{\mu}_k(t) - \mu_k \right] \leq 0 \]

Fixed Arm \quad \text{Fixed Time}
Nie et al. 2018
: Sample mean is **negatively** biased.

\[ \mathbb{E} \left[ \hat{\mu}_k(t) - \mu_k \right] \leq 0 \]

- **Fixed Arm**
- **Fixed Time**

This work
: Sample mean of chosen arm at stopping time

\[ \mathbb{E} \left[ \hat{\mu}_\kappa(\mathcal{T}) - \mu_\kappa \right] \]

- **Chosen Arm**
- **Stopping Time**
This work

Sample mean of chosen arm at stopping time is ...

\[ \mathbb{E} \left[ \hat{\mu}_\kappa(T) - \mu_\kappa \right] \]
This work

: Sample mean of chosen arm at stopping time is ...

\[
\mathbb{E} \left[ \hat{\mu}_k(T) - \mu_k \right]
\]

(a) negatively biased under ‘optimistic sampling'.
This work

: Sample mean of chosen arm at stopping time is ...

\[ \mathbb{E} \left[ \hat{\mu}_\kappa(\mathcal{T}) - \mu_\kappa \right] \]

(a) **negatively** biased under ‘optimistic sampling'.

(b) **positively** biased under ‘optimistic stopping’.
This work
: Sample mean of chosen arm at stopping time is ...

\[ \mathbb{E} \left[ \hat{\mu}_k(\mathcal{T}) - \mu_k \right] \]

(a) **negatively** biased under ‘optimistic sampling’.

(b) **positively** biased under ‘optimistic stopping’.

(c) **positively** biased under ‘optimistic choosing’.
Monotone effect of a sample

**Theorem [Informal]**

Sample from arm $k$ \[ \frac{1(\kappa = k)}{N_k(\mathcal{T})} \]
Monotone effect of a sample

**Theorem** [Informal]

Sample from arm $k$ \[ \xrightarrow{\text{Increasing}} \] \[ \frac{1(\kappa = k)}{N_k(\mathcal{T})} \] \[ \text{Positive bias} \]
Monotone effect of a sample

**Theorem** [Informal]

Sample from arm $k$ \[ \frac{1(\kappa = k)}{N_k(\mathcal{T})} \]

Positive bias, Negative bias
Monotone effect of a sample

**Theorem** [Informal]

Sample from arm $k$ → \( \frac{1(\kappa = k)}{N_k(\mathcal{T})} \)

Positive bias

Negative bias

Increasing

Decreasing

Agnostic to algorithm
Monotone effect of a sample

**Theorem** [Informal]

Sample from arm $k$ \[ \frac{1(\kappa = k)}{N_k(\mathcal{T})} \] Positive bias  
Increasing  
Negative bias  
Decreasing  

Agnostic to algorithm  
Includes Nie et al. 2018 as a special case
Monotone effect of a sample

**Theorem** [Informal]

Sample from arm $k$  \( \xrightarrow{\text{Increasing}} \) \( \frac{1(\kappa = k)}{N_k(\mathcal{T})} \)  \( \xrightarrow{\text{Decreasing}} \) Positive bias

Increasing

Decreasing

- Agnostic to algorithm
- Includes Nie et al. 2018 as a special case
- Positive bias under best arm identification, sequential testing
Are sample means in multi-armed bandits positively or negatively biased?

Jaehyeok Shin, Aaditya Ramdas and Alessandro Rinaldo