Estimating the Maximum Expected Value through Gaussian Approximation

C. D'Eramo, M. Restelli and A. Nuara

{carlo.deramo, marcello.restelli}@polimi.it
{alessandro.nuara}@mail.polimi.it

POLITECNICO MILANO 1863

Problem

- Compute the **Maximum Expected Value (MEV)** of a set of two or more independent random variables $X = \{X_1, ..., X_M\}$ given samples $S = \{S_1, ..., S_N\}$
- Most RL algorithms need to **approximate MEV**
- A good estimation of MEV is **critical** in many real-world applications

Contributions

1. We propose the **Weighted Estimator (WE)** to approximate MEV
2. The approximation is done by a **weighted average** of sample means
3. We provide theoretical and empirical **comparisons** of the **performance** of WE and other MEV estimators

Maximum Expected Value Estimation

Idea

- Compute the *Maximum Expected Value*

$$\mu_*(X) = \max_i \mu_i = \max_i \int_{-\infty}^{+\infty} x f_i(x)\,\mathrm{d}x$$

- of independent random variables ($X = \{X_1, X_2, \dots, X_M\}$) whose PDFs f_i are unknown

- **Issue:** μ_* cannot be computed analytically
- Given a set of noisy samples $S = \{S_1, \dots, S_N\}$ retrieved by the unknown distributions of each X_i

GOAL

$$\boxed{\mu_*(X) \approx \hat{\mu}_*(S)}$$

Naïve approach

- *Maximum Estimator (ME)*
- i.e., take the maximum of the sample means

$$\hat{\mu}^{ME}(S) = \max_i \hat{\mu}_i(S) \approx \mu_*(X)$$

- **Positive** bias can cause problems in some applications (e.g., Q-Learning)

Double Estimator (DE) [Van Hasselt, 2010]

1. Split dataset S in two disjoint sets

$$S^A = \{S_1^A, ..., S_N^A\} \quad \text{and} \quad S^B = \{S_1^B, ..., S_N^B\}$$

2. Estimate the maximum index in each set

$$a^* = \arg\max_i \hat{\mu}_i^{ME}(S^A) \quad \text{and} \quad b^* = \arg\max_i \hat{\mu}_i^{ME}(S^B)$$

3. Take the average maximum value

$$\hat{\mu}^{DE}(S) = \frac{\hat{\mu}_{b^*}^{ME}(S^A) + \hat{\mu}_{a^*}^{ME}(S^B)}{2} \approx \mu_*(X)$$

Negative bias may solve ME issues in many applications

Weighted Estimator (WE)

WE

$$\boxed{\hat{\mu}^{WE}(S) = \sum_{i=1}^{M} \hat{\mu}_i(S) w_i^S}$$

Weights the sample means by the probability of being the maximum

$$w_i^S = P\left(\hat{\mu}_i(S) = \max_j \hat{\mu}_j(S)\right) = \int_{-\infty}^{+\infty} \overbrace{\hat{f}_i^S(x)}^{\text{PDF unknown}} \prod_{j \neq i} \underbrace{\hat{F}_j^S(x)}_{\text{CDF unknown}} \,\mathrm{d}x$$

- As the number of samples increases, by the **central limit theorem**:

$$\hat{\mu}_i(S) \sim \mathcal{N}\Bigg(\underbrace{\mu_i}_{\text{sample mean}}, \frac{\overbrace{\sigma_i^2}^{\text{sample variance}}}{|S_i|}\Bigg) \quad \text{i.e.,} \quad \hat{f}_i^S = \textit{normal distribution}$$

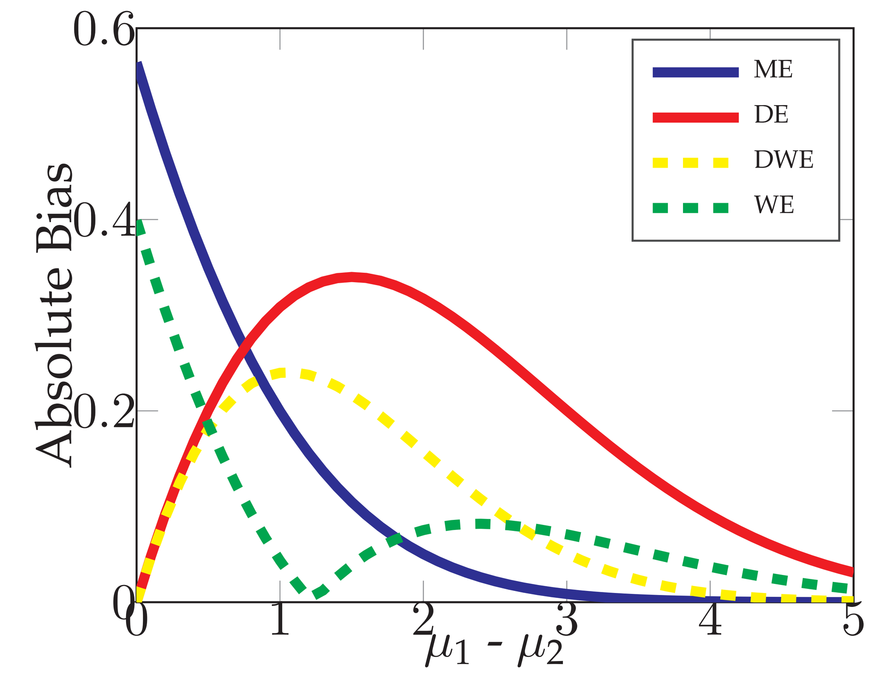
Theoretical Results

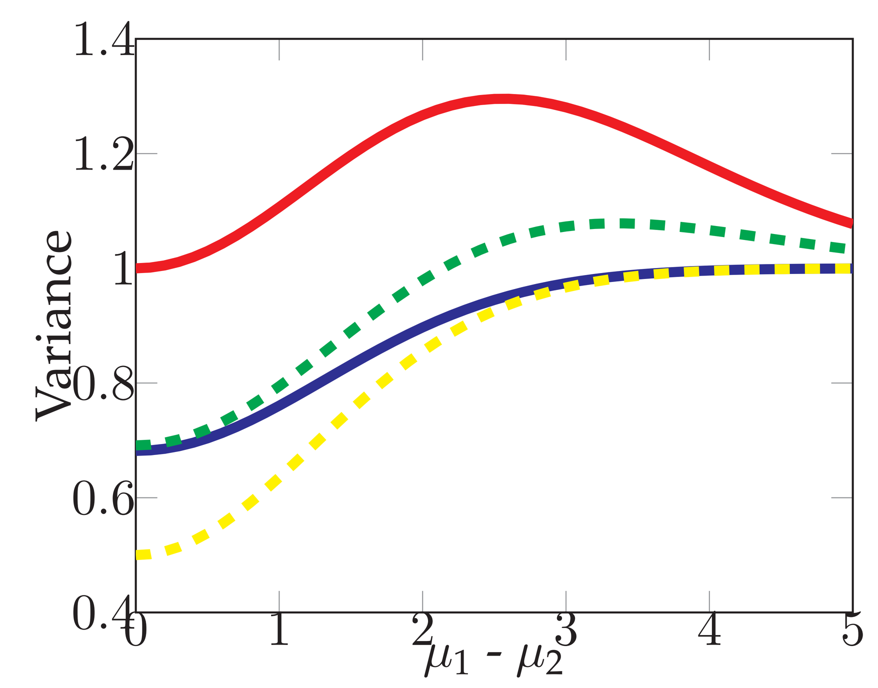
Bias and variance of the estimators

$$\text{Bias}\left(\hat{\mu}^{DE}\right) \geq -\frac{1}{2}\left(\sqrt{\sum_{i=1}^{M} \frac{\sigma_i^2}{|S_i^A|}} + \sqrt{\sum_{i=1}^{M} \frac{\sigma_i^2}{|S_i^B|}}\right)$$

$$\text{Bias}(\hat{\mu}^{WE}) \leq \text{Bias}(\hat{\mu}^{ME}) \leq \sqrt{\frac{M-1}{M} \sum_{i=1}^{M} \frac{\sigma_i^2}{|S_i|}}$$

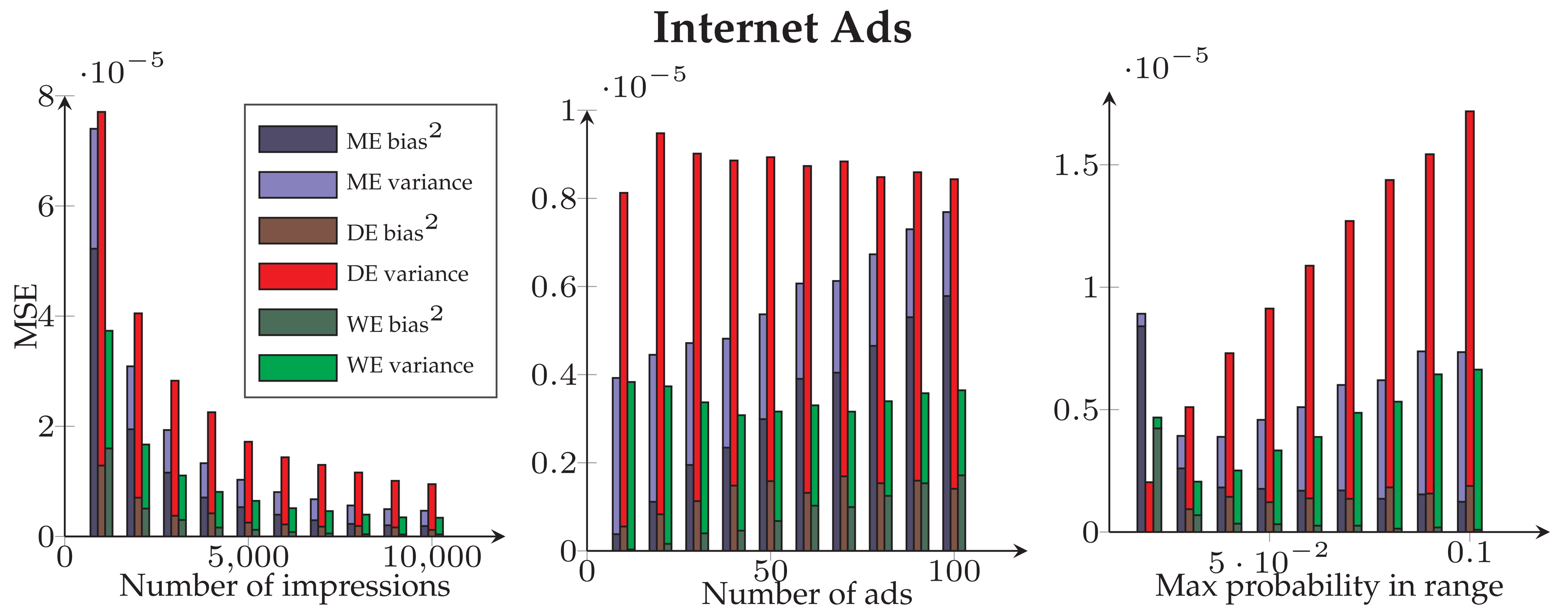
$$\text{Var}\left(\hat{\mu}^{ME,DE,WE}\right) \leq \sum_{i=1}^{M} \frac{\sigma_i^2}{|S_i|}.$$

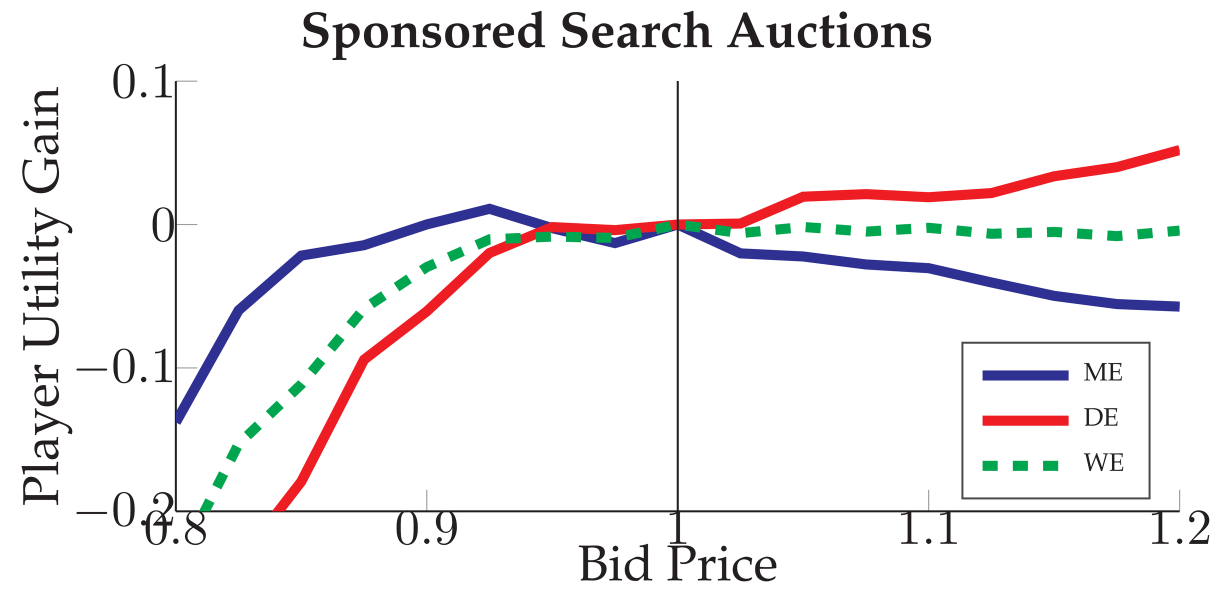




Empirical Results

Multi-Armed Bandits





Markov Decision Processes