Order-Optimal Sequential 1-Bit Mean Estimation

Ivan Lau & Jonathan Scarlett

National University of Singapore

AISTATS Extended

1. Problem Setup

- **Mean:** $\mathbb{E}[X] \in [-\lambda, \lambda]$
- **Moment bound:** $\mathbb{E}[|X-\mu|^k] \leq \sigma^k$
  - $k = 2$: finite variance
  - $k > 2$: light-tailed (e.g., sub-Gaussian)
  - $k \in (1,2)$: heavy-tailed

Feedback constraint: 1 bit per sample

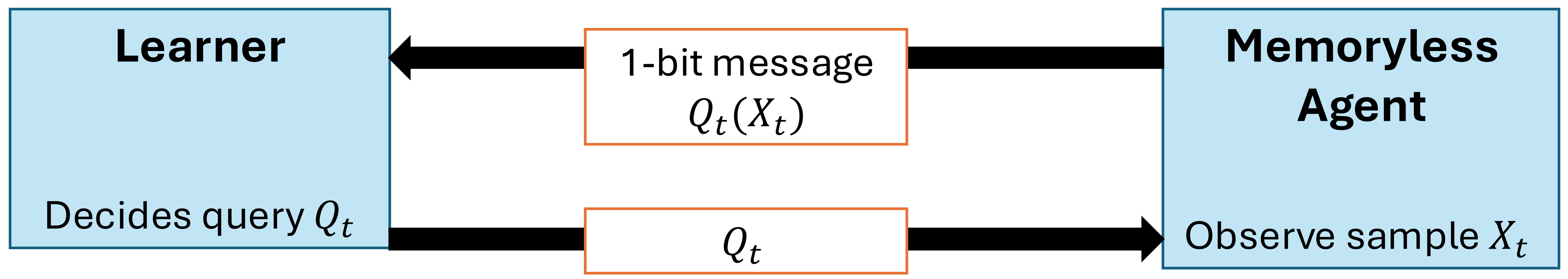
Goal: (ϵ, δ)-PAC estimate of $\mathbb{E}[X]$

2. Order-Optimal Sample Complexity

$$\underbrace{\log\left(\frac{\lambda}{\sigma}\right)}_{\text{Localization}} + \underbrace{\begin{cases} \frac{\sigma^2}{\epsilon^2} \cdot \log\left(\frac{1}{\delta}\right) & k > 2 \\ \frac{\sigma^2}{\epsilon^2} \cdot \log\left(\frac{\sigma}{\epsilon}\right) \cdot \log\left(\frac{1}{\delta}\right) & k = 2 \\ \left(\frac{\sigma}{\epsilon}\right)^{\frac{k}{k-1}} \cdot \log\left(\frac{1}{\delta}\right) & k \in (1,2) \end{cases}}_{\text{“Refinement”}}$$

3. Cost of 1-Bit Quantization

- **Localization:** Incurs additive $\log(\lambda/\sigma)$ penalty absent in *unquantized* estimation
- **Refinement:**
  - Matches *unquantized* rates for $k \neq 2$
  - Suffers $\log(\sigma/\epsilon)$ penalty for $k = 2$



Our estimator

- Uses only **threshold queries**
- **Distribution free** (only moment bound)
- **Order-optimal** sample complexity across all tail regimes

Prior work

- **Parametric family** only (e.g., Gaussian, Laplace)
- Requires CDF inversions and/or complicated queries

4. Sketch of Estimator $(k = 2)$

1. Localize mean to $O(\sigma)$

- Noisy binary search using $O(\log(\lambda/\sigma))$ queries
- WLOG, center of interval is 0

2. Truncate at $t = \Theta(\sigma^2/\epsilon)$

- $\mathbb{E}[X] \approx \mathbb{E}[X \cdot \mathbf{1}(|X| \leq t)]$ with bias $= O(\epsilon)$

3. Geometric Partition on $[-t, t]$

- Symmetric doubling grids:
  - $R_1 = [0, 2\sigma)$, $R_2 = [2\sigma, 6\sigma), \ldots$; $R_i = -R_{-i}$
  - No. of grids $O(\log(\sigma/\epsilon))$
- Decompose into local contributions

$$\mathbb{E}[X \cdot \mathbf{1}(|X| \leq t)] = \sum_i \underbrace{\mathbb{E}[X \cdot \mathbf{1}(X \in R_i)]}_{\text{Form local estimates } \hat{\mu}_i}$$

4. Threshold Query for each $R_i = [a, b)$

- Reduced to estimating boundary probabilities:

$$\mathbb{E}[X \cdot \mathbf{1}(X \in [a,b))] = a \cdot \Pr(a \leq X \leq U) + b \cdot \Pr(U < X < b)$$

where $U \sim \mathrm{Unif}(a, b)$

- Estimated through threshold queries:

$$\Pr(a \leq X \leq U) = \mathbb{E}[\mathbf{1}(X \leq U)] - \mathbb{E}[\mathbf{1}(a \leq X)]$$

5. Allocation & Concentration

- Allocate queries per grid $n_i = O(\sigma^2/\epsilon^2)$
  - Summing over all grids gives $O(\sigma^2/\epsilon^2 \cdot \log(\sigma/\epsilon))$
  - Overall variance $O(\epsilon^2)$
- Median-of-means to achieve (ϵ, δ)-PAC
  - Repeat $O(\log(1/\delta))$ times