Universal and Tight Bounds on Counting Errors of Count-Min Sketch with Conservative Updates

Younes Ben Mazziane (Inria and University of Avignon LIA)
Othmane Marfoq (Meta)

12 June 2025

ACM SIGMETRICS 2025

Stony Brook, New York, USA.

Outline

1) Streaming Algorithms

2) Classical Streaming Algorithms

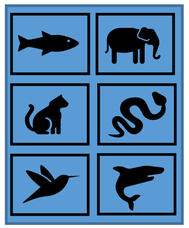
3) Count-Min Sketch with Conservative Updates

4) Research Gap and Motivation

5) Contribution

Streaming Algorithms

Catalog of items $\mathcal{I} = \{1, \ldots, N\}$



Data stream $r_1, \ldots, r_T$

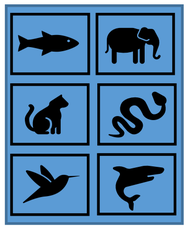
Objective: Answer queries

1. The count of item i?
2. Top k items?
3. $\phi -$ heavy hitters?
4. Distinct items?

$$n_i(t) = \sum_{s=1}^{t-1} I(r_s = i)$$

Streaming Algorithms

Catalog of items $\mathcal{I} = \{1, \ldots, N\}$



Data stream $r_1, \ldots, r_T$

Objective: Answer queries

1. The count of item i?
2. Top k items?
3. $\phi -$ heavy hitters?
4. Distinct items?

$$n_i(t) = \sum_{s=1}^{t-1} I(r_s = i)$$

Constraints

1. High arrival rate, e.g., 1M item/s
2. Limited memory

$\hat{\mathbf{n}}(t)$ using $o(N)$ memory

And one pass algorithm

Network monitoring + clickstream analytics

Classical Streaming Algorithms

Counter-based

Id + count for a subset of items, e.g., Space Saving [4], Frequent [1,2]

[1]Misra, S., & Gries, D. (1982). *Finding repeated elements*. **Science of Computer Programming**, 2(2), 143–152.
[2] Demaine, E. D., López-Ortiz, A., & Munro, J. I. (2002). *Frequency estimation of internet packet streams with limited space*. In **European Symposium on Algorithms (ESA)** (pp. 348–360).
[3] Charikar, M., Chen, K., & Farach-Colton, M. (2002). *Finding frequent items in data streams*. In **International Colloquium on Automata, Languages and Programming (ICALP)** (pp. 693–703).
[4] Metwally, A., Agrawal, D., & El Abbadi, A. (2005). *Efficient computation of frequent and top-k elements in data streams*. In **International Conference on Database Theory (ICDT)** (pp. 398–412).
[5] Cormode, G., & Muthukrishnan, S. (2005). *An improved data stream summary: The Count–Min sketch and its applications*. **Journal of Algorithms**, 55(1), 58–75.

Classical Streaming Algorithms

Counter-based

Id + count for a subset of items, e.g., Space Saving [4], Frequent [1,2]

Sketch-based

Random projection of counts vector, e.g., Count-Sketch [3], Count-Min-Sketch [5].

$$(Y_1(t), \ldots, Y_M(t)) = \mathbf{n}(t)\mathbf{\Pi} : \ \mathbf{\Pi} \in \mathcal{M}_{N,M}(\mathbb{R})$$

Sparse random matrix

[1]Misra, S., & Gries, D. (1982). *Finding repeated elements*. **Science of Computer Programming**, 2(2), 143–152.
[2] Demaine, E. D., López-Ortiz, A., & Munro, J. I. (2002). *Frequency estimation of internet packet streams with limited space*. In **European Symposium on Algorithms (ESA)** (pp. 348–360).
[3] Charikar, M., Chen, K., & Farach-Colton, M. (2002). *Finding frequent items in data streams*. In **International Colloquium on Automata, Languages and Programming (ICALP)** (pp. 693–703).
[4] Metwally, A., Agrawal, D., & El Abbadi, A. (2005). *Efficient computation of frequent and top-k elements in data streams*. In **International Conference on Database Theory (ICDT)** (pp. 398–412).
[5] Cormode, G., & Muthukrishnan, S. (2005). *An improved data stream summary: The Count–Min sketch and its applications*. **Journal of Algorithms**, 55(1), 58–75.

Count-Min Sketch (CMS)

$$\mathbf{Y}(t) = \mathbf{n}(t) \cdot \mathbf{\Pi}$$

$$\mathbf{\Pi} = (\pi_{i,j}) \in \mathcal{M}_{N,M}(\{0,1\}) : \sum_{j=1}^{M} \pi_{i,j} = d \ll M$$

Popularity estimation

$$\hat{n}_i(t) = \min_{j \in [M]:\, \pi_{i,j}=1} Y_j(t)$$

Count-Min Sketch (CMS)

$$\mathbf{Y}(t) = \mathbf{n}(t) \cdot \mathbf{\Pi}$$

$$\mathbf{\Pi} = (\pi_{i,j}) \in \mathcal{M}_{N,M}(\{0,1\}) : \sum_{j=1}^{M} \pi_{i,j} = d \ll M$$

Popularity estimation

$$\hat{n}_i(t) = \min_{j \in [M]:\, \pi_{i,j}=1} Y_j(t)$$

Example:

$N = 5,\ M = 4,\ d = 2$

$$\mathbf{n} = [10 \quad 10 \quad 1 \quad 1 \quad 1] \qquad \mathbf{\Pi} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$$\sum_{i \in \mathcal{N}(4)} n_i = n_2 + n_4$$

$$\mathbf{Y} = [21 \quad 2 \quad 12 \quad 11]$$

$$\hat{n}_1 = \min(21, 12) = 12$$

Count-Min Sketch (CMS)

Implicit storage via d hash functions: $\forall l \in [d],\ h_l : \mathcal{I} \mapsto [M]$

$$\pi_{i,j} = 1 \iff \exists l \in [d] :\ h_l(i) = j$$

Example

$$\mathbf{Y}(t) = [21 \quad 2 \quad 12 \quad 11] \longrightarrow \mathbf{Y}(t+1) = [22 \quad 2 \quad 12 \quad 12]$$

$$h_1(r_{t+1}) = 1 \qquad h_2(r_{t+1}) = 4$$

Count-Min Sketch (CMS)

Implicit storage via d hash functions: $\forall l \in [d],\ h_l : \mathcal{I} \mapsto [M]$

$$\pi_{i,j} = 1 \iff \exists l \in [d] :\ h_l(i) = j$$

Example

$$\mathbf{Y}(t) = [21 \quad 2 \quad 12 \quad 11] \longrightarrow \mathbf{Y}(t+1) = [22 \quad 2 \quad 12 \quad 12]$$

$$h_1(r_{t+1}) = 1 \qquad h_2(r_{t+1}) = 4$$

Update Rule

$$Y_c(t+1) = \begin{cases} Y_c(t+1) + 1, & \text{if } \exists l \in [d] :\ h_l(r_{t+1}) = c, \\ Y_c(t), & \text{otherwise.} \end{cases}$$

Count Estimation

$$\hat{n}_i(t) = \min_{l \in [d]} Y_{h_l(i)}(t)$$

CMS with Conservative Updates (CMS-CU)

Example

$$\mathbf{Y}(t) = [21 \quad 2 \quad 12 \quad 11] \quad \Rightarrow \quad Y(t+1) = [21 \quad 2 \quad 12 \quad 12]$$

$$h_1(r_{t+1}) = 1 \qquad\qquad h_2(r_{t+1}) = 4$$

CMS with Conservative Updates (CMS-CU)

Example

$$\mathbf{Y}(t) = [21 \quad 2 \quad 12 \quad 11] \longrightarrow Y(t+1) = [21 \quad 2 \quad 12 \quad 12]$$

$$h_1(r_{t+1}) = 1 \qquad h_2(r_{t+1}) = 4$$

Update Rule

$$Y_c(t+1) = \begin{cases} Y_c(t+1) + 1, & \text{if } c \in \operatorname{argmin}_{l \in [d]} Y_{h_l(r_{t+1})}(t+1), \\ Y_c(t), & \text{otherwise.} \end{cases}$$

Count Estimation

$$\hat{n}_i(t) = \min_{l \in [d]} Y_{h_l(i)}(t)$$

Counting Error (CE)

$$e_i(t) = \hat{n}_i(t) - n_i(t)$$

Research Gap and Motivation

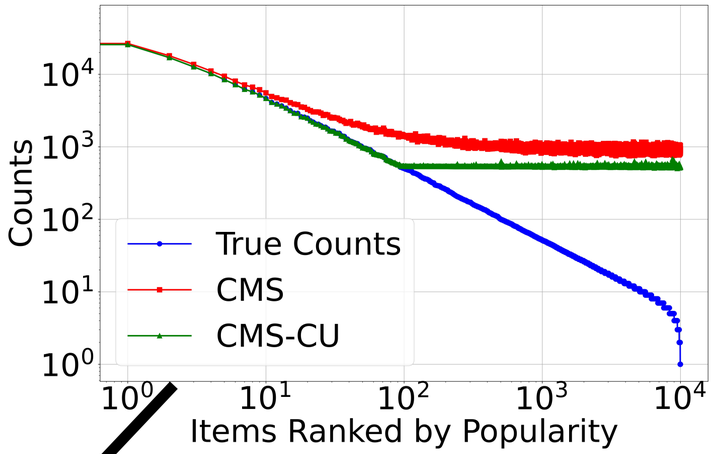
Quantifying CE of infrequent Items is crucial: Risk of misclassification as Heavy Hitters

Research Gap and Motivation

Quantifying CE of infrequent Items is crucial: Risk of misclassification as Heavy Hitters

Stream: $N = 10^4, T = 5 * 10^5, p_i \sim 1/i$

Sketch: $(M, d) = (1000,5)$



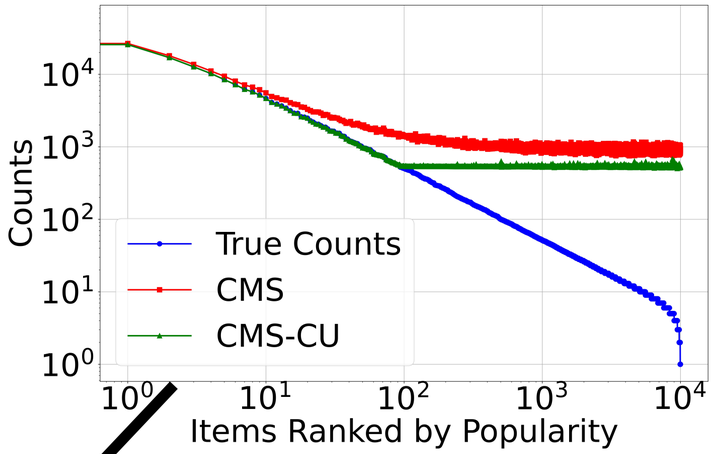
Inflated Count ≈ CE of infrequent items

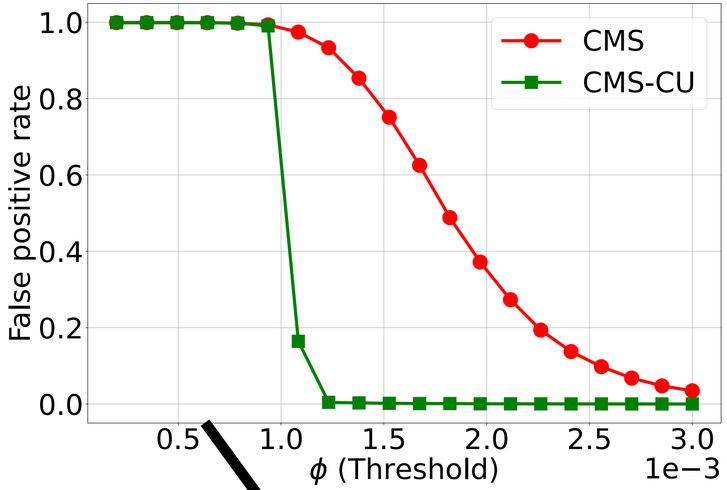
Research Gap and Motivation

Quantifying CE of infrequent Items is crucial: Risk of misclassification as Heavy Hitters

Stream: $N = 10^4, T = 5 * 10^5, p_i \sim 1/i$

Sketch: $(M, d) = (1000,5)$





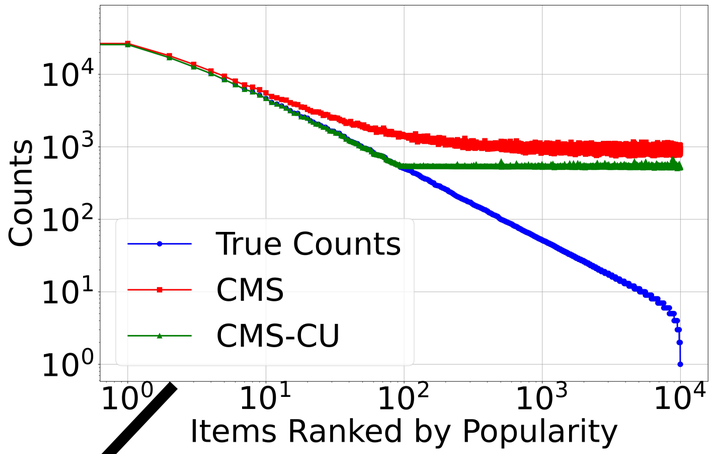
Inflated Count $\approx$ CE of infrequent items $\sim$ Critical Threshold ϕ_c

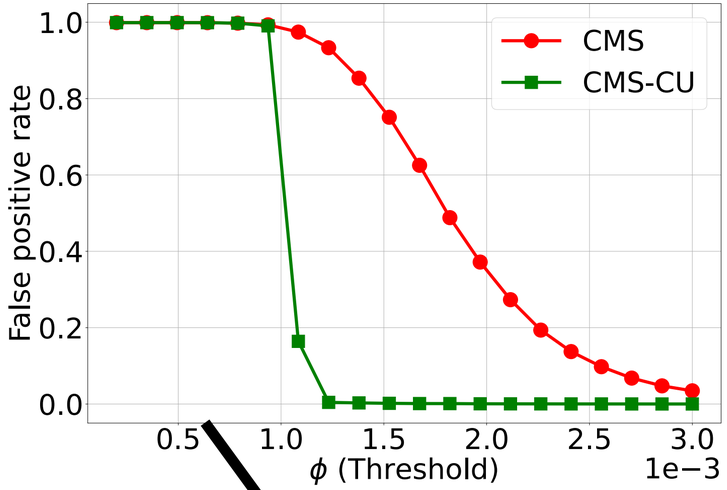
Research Gap and Motivation

Quantifying CE of infrequent Items is crucial: Risk of misclassification as Heavy Hitters

Stream: $N = 10^4, T = 5 * 10^5, p_i \sim 1/i$

Sketch: $(M, d) = (1000,5)$





Inflated Count ≈ CE of infrequent items $\sim$ Critical Threshold ϕ_c

Quantifying CE of infrequent items → Near-optimal Configuration of CMS-CU

Contribution: Definitions

Definition 1 [Counting Error (CE)]

$$e_i(t) \triangleq \hat{n}_i(t) - n_i(t)\,.$$

Definition 2 [Unseen Items' Error (UE)]

$$e^0(t) \triangleq \min_{c \in \mathbf{X}} Y_c(t) : \ \mathbf{X} \sim \text{Unif}(\binom{[M]}{d})\,.$$

Contribution: Definitions

Definition 1 [Counting Error (CE)]

$$e_i(t) \triangleq \hat{n}_i(t) - n_i(t)\,.$$

Definition 2 [Unseen Items' Error (UE)]

$$e^0(t) \triangleq \min_{c \in \mathbf{X}} Y_c(t) : \ \mathbf{X} \sim \text{Unif}\left(\begin{pmatrix} [M] \\ d \end{pmatrix}\right).$$

Definition 3 [Reference Error (RE)]

$$e^*(t) \triangleq \min_{c \in \mathbf{X}} \tilde{Y}_c(t),$$

$$\tilde{Y}_c(t+1) = \begin{cases} \tilde{Y}_c(t) + 1, & \text{if } c \in \text{argmin}_{u \in \mathbf{X}_{t+1}} \tilde{Y}_u(t), \\ \tilde{Y}_c(t), & \text{otherwise.} \end{cases}$$

, (X_t) and $\mathbf{X}$ are i.i.d.

Contribution: RE is important

Theorem 1 [From CE to UE]:

$$\Pr\left(e_i(t) \geq x\right) \leq \Pr\left(e^0(t) \geq x\right)$$

1) UE is a proxy for the errors of infrequent items

2) The false positive rate in ϕ-Heavy Hitters can be bounded in terms of UE

Contribution: RE is important

Theorem 1 [From CE to UE]: $\Pr\left(e_i(t) \geq x\right) \leq \Pr\left(e^0(t) \geq x\right)$

1) UE is a proxy for the errors of infrequent items

2) The false positive rate in ϕ-Heavy Hitters can be bounded in terms of UE

Theorem 2 [From UE to RE]: For a class of skewed streams: $\mathrm{E}\left[e^0(t)\right] \leq C_k + \alpha_k \mathrm{E}\left[\hat{e}^{*,k}(t)\right]$

	Counters	Stream
$e^*(T)$	$[M]$	$\mathbf{r} = (r_t)_{t \in [T]}$
$\hat{e}^{*,k}(T)$	$[M] \setminus \cup_{i=1}^{k} \mathbf{h}(i)$	$\mathbf{r} \setminus$ Top k

Contribution: RE is important

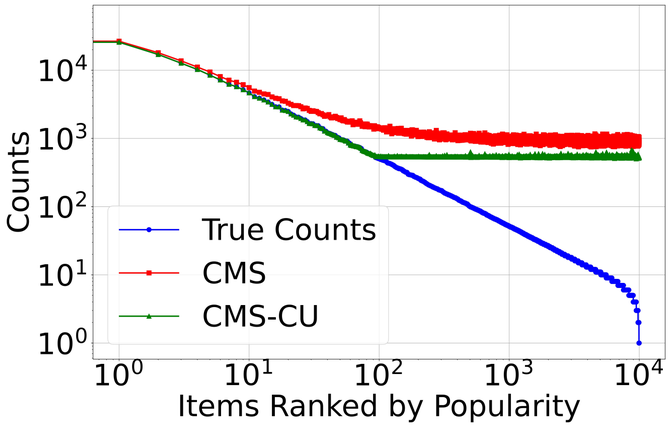
Theorem 1 [From CE to UE]: $\Pr\left(e_i(t) \geq x\right) \leq \Pr\left(e^0(t) \geq x\right)$

1) UE is a proxy for the errors of infrequent items

2) The false positive rate in ϕ-Heavy Hitters can be bounded in terms of UE

Theorem 2 [From UE to RE]: For a class of skewed streams: $\mathrm{E}\left[e^0(t)\right] \leq C_k + \alpha_k \mathrm{E}\left[\hat{e}^{*,k}(t)\right]$

	Counters	Stream
$e^*(T)$	$[M]$	$\mathbf{r} = (r_t)_{t \in [T]}$
$\hat{e}^{*,k}(T)$	$[M] \setminus \cup_{i=1}^{k} \mathbf{h}(i)$	$\mathbf{r} \setminus$ Top k



1) These bounds are provably tight for some streams

2) These bounds are valid for any stream if:

Conjecture 1: $\mathrm{E}\left[\hat{e}^{0,k}(t)\right] \leq \mathrm{E}\left[\hat{e}^{*,k}(t)\right]$

Computing RE is important

Contribution: Computing RE

Analyzing $\tilde{Y}$ is difficult

$$\tilde{Y}_c(t+1) = \begin{cases} \tilde{Y}_c(t) + 1, & \text{if } c \in \text{argmin}_{u \in \mathbf{X}_{t+1}} \tilde{Y}_u(t), \\ \tilde{Y}_c(t), & \text{otherwise.} \end{cases}$$

1) Complex dependency on $\mathbf{X}_{t+1}$, vs CMS: $Y_c(t) = \sum_{s=1}^{t} I\left(c \in \mathbf{X}_s\right)$

2) $\tilde{Y}$ is a Markov process but its state space grows fast with time

Contribution: Computing RE

Analyzing $\tilde{Y}$ is difficult

$$\tilde{Y}_c(t+1) = \begin{cases} \tilde{Y}_c(t) + 1, & \text{if } c \in \text{argmin}_{u \in \mathbf{X}_{t+1}} \tilde{Y}_u(t), \\ \tilde{Y}_c(t), & \text{otherwise.} \end{cases}$$

1) Complex dependency on $\mathbf{X}_{t+1}$, vs CMS: $Y_c(t) = \sum_{s=1}^{t} I\left(c \in \mathbf{X}_s\right)$

2) $\tilde{Y}$ is a Markov process but its state space grows fast with time

3) Similar to a Balls and Bins with Power of d: $\tilde{B}$

$$\tilde{B}_c(t+1) = \begin{cases} \tilde{B}_c(t) + 1, & \text{if } c \in \text{min argmin}_{u \in \mathbf{X}_{t+1}} \tilde{B}_u(t), \\ \tilde{B}_c(t), & \text{otherwise.} \end{cases},$$

Number of increments is d in $\tilde{B}$ but random in $\tilde{Y}$

Contribution: Computing RE

Analyzing $\tilde{Y}$ is difficult

$$\tilde{Y}_c(t+1) = \begin{cases} \tilde{Y}_c(t) + 1, & \text{if } c \in \text{argmin}_{u \in \mathbf{X}_{t+1}} \tilde{Y}_u(t), \\ \tilde{Y}_c(t), & \text{otherwise.} \end{cases}$$

1) Complex dependency on $\mathbf{X}_{t+1}$, vs CMS: $Y_c(t) = \sum_{s=1}^{t} I\left(c \in \mathbf{X}_s\right)$

2) $\tilde{Y}$ is a Markov process but its state space grows fast with time

3) Similar to a Balls and Bins with Power of d: $\tilde{B}$

$$\tilde{B}_c(t+1) = \begin{cases} \tilde{B}_c(t) + 1, & \text{if } c \in \text{min argmin}_{u \in \mathbf{X}_{t+1}} \tilde{B}_u(t), \\ \tilde{B}_c(t), & \text{otherwise.} \end{cases},$$

Number of increments is d in $\tilde{B}$ but random in $\tilde{Y}$

Key Ideas

1) Alternate representation:

$$\mathbf{\Delta}(t): \ \Delta_l(t) \triangleq \sum_{c=1}^{m} I\left(\tilde{Y}_c(t) - \tilde{Y}_{\min}(t) = l\right).$$

Contribution: Computing RE

Analyzing $\tilde{Y}$ is difficult

$$\tilde{Y}_c(t+1) = \begin{cases} \tilde{Y}_c(t) + 1, & \text{if } c \in \text{argmin}_{u \in \mathbf{X}_{t+1}} \tilde{Y}_u(t), \\ \tilde{Y}_c(t), & \text{otherwise.} \end{cases}$$

1) Complex dependency on $\mathbf{X}_{t+1}$, vs CMS: $Y_c(t) = \sum_{s=1}^{t} I\left(c \in \mathbf{X}_s\right)$

2) $\tilde{Y}$ is a Markov process but its state space grows fast with time

3) Similar to a Balls and Bins with Power of d: $\tilde{B}$

$$\tilde{B}_c(t+1) = \begin{cases} \tilde{B}_c(t) + 1, & \text{if } c \in \text{min argmin}_{u \in \mathbf{X}_{t+1}} \tilde{B}_u(t), \\ \tilde{B}_c(t), & \text{otherwise.} \end{cases},$$

Number of increments is d in $\tilde{B}$ but random in $\tilde{Y}$

Key Ideas

1) Alternate representation:

$$\mathbf{\Delta}(t): \ \Delta_l(t) \triangleq \sum_{c=1}^{m} I\left(\tilde{Y}_c(t) - \tilde{Y}_{\min}(t) = l\right).$$

2) Variants of $\tilde{Y}$:

CU-LB ($\tilde{\mathbf{Y}}^{\text{lb},g}$) and CU-UB ($\tilde{\mathbf{Y}}^{\text{ub},g}$)

2.1) $(\mathbf{\Delta}^g(t))_t$ is a finite state space Markov chain of size $\binom{m+g-d}{g}$

2.2) $\tilde{\mathbf{Y}}^{\text{lb},g} \leq \tilde{\mathbf{Y}} \leq \tilde{\mathbf{Y}}^{\text{ub},g}$

Contribution: Computing RE

If $\mathbf{X}_{t+1} \subset \text{argmax}_{c \in [m]} Y_c^{\text{lb,g}}(t)$, and $G^{\text{lb},g}(t) = g$

$\longrightarrow$ $\tilde{\mathbf{Y}}^{\text{lb},g}(t+1) = \tilde{\mathbf{Y}}^{\text{lb},g}(t)$

If $\mathbf{X}_{t+1} \subset \text{argmax}_{c \in [m]} Y_c^{\text{ub,g}}(t)$, and $G^{\text{ub},g}(t) = g$

$\longrightarrow$ Increment all min value counters and those in $\mathbf{X}_{t+1}$

Contribution: Computing RE

If $\mathbf{X}_{t+1} \subset \text{argmax}_{c\in[m]} Y_c^{\text{lb,g}}(t)$, and $G^{\text{lb},g}(t) = g$

$\rightarrow$ $\tilde{\mathbf{Y}}^{\text{lb},g}(t+1) = \tilde{\mathbf{Y}}^{\text{lb},g}(t)$

Otherwise:

$\rightarrow$ $\tilde{\mathbf{Y}}^{\text{lb},g}(t+1) = \text{CU}\left(\tilde{\mathbf{Y}}^{\text{lb},g}(t)\right)$

If $\mathbf{X}_{t+1} \subset \text{argmax}_{c\in[m]} Y_c^{\text{ub,g}}(t)$, and $G^{\text{ub},g}(t) = g$

$\rightarrow$ Increment all min value counters and those in $\mathbf{X}_{t+1}$

Otherwise:

$\rightarrow$ $\tilde{\mathbf{Y}}^{\text{ub},g}(t+1) = \text{CU}\left(\tilde{\mathbf{Y}}^{\text{ub},g}(t)\right)$

$$G^{\text{lb},g}(t) \leq g,\ G^{\text{ub},g}(t) \leq g$$

Contribution: Computing RE

If $\mathbf{X}_{t+1} \subset \text{argmax}_{c\in[m]} Y_c^{\text{lb,g}}(t)$, and $G^{\text{lb},g}(t) = g$

$\longrightarrow$ $\tilde{\mathbf{Y}}^{\text{lb},g}(t+1) = \tilde{\mathbf{Y}}^{\text{lb},g}(t)$

Otherwise:

$\longrightarrow$ $\tilde{\mathbf{Y}}^{\text{lb},g}(t+1) = \text{CU}\left(\tilde{\mathbf{Y}}^{\text{lb},g}(t)\right)$

If $\mathbf{X}_{t+1} \subset \text{argmax}_{c\in[m]} Y_c^{\text{ub,g}}(t)$, and $G^{\text{ub},g}(t) = g$

$\longrightarrow$ Increment all min value counters and those in $\mathbf{X}_{t+1}$

Otherwise:

$\longrightarrow$ $\tilde{\mathbf{Y}}^{\text{ub},g}(t+1) = \text{CU}\left(\tilde{\mathbf{Y}}^{\text{ub},g}(t)\right)$

$$G^{\text{lb},g}(t) \leq g,\ G^{\text{ub},g}(t) \leq g$$

Theorem 3:

1) $\ell_g < \ell_{g+1} \leq \mathrm{E}\left[e^*(T)/T\right] \leq U_{g+1} < U_g$

2) ℓ_g and U_g can be computed in $\mathcal{O}\left((T+g)m\binom{m+g-d}{g}\right)$

3) $\lim_{t\to\infty} \Pr\left(\ell_g(\infty) \leq e^*(t)/t \leq U_g(\infty)\right) = 1$

Table 1: Results for different values of g with $m = 50$, $d = 4$, and $T = 250$.

g	1	2	3	4	5
$\ell_g(T)$	0.01860	0.02956	0.03420	0.03540	0.03559
$U_g(T)$	0.07654	0.04090	0.03637	0.03572	0.03562
Execution time in (s)	0.05	0.68	17	257	3241

Thank you.

CMS-CU: Previous studies

[6] Uniform Counters selection via mean-field analysis: Solving a large system of ODES

[7,8] Assumes stationary Stream: Analysis captures errors for popular items only

[9,10] Uniform and phase transition

[6] Bianchi, G., Duffy, K., Leith, D., & Shneer, V. (**2012**). Modeling conservative updates in multi-hash approximate count sketches. In *24th International Teletraffic Congress (**ITC**)* (pp. 1–8).

[7] Einziger, G., & Friedman, R. (**2015**). A formal analysis of conservative update based approximate counting. In *2015 International Conference on Computing, Networking and Communications (**ICNC**)* (pp. 255–259)

[8] Ben Mazziane, Y., Alouf, S., & Neglia, G. (**2022**). Analyzing count–min sketch with conservative updates. ***Computer Networks***, 217, 109315.

[9] Fusy, É., & Kucherov, G. (**2023**). Phase transition in count approximation by count–min sketch with conservative updates. In *International Conference on Algorithms and Complexity (**CIAC**)* (pp. 232–246).

[10] Fusy, É., & Kucherov, G. (**2023**). Count–min sketch with variable number of hash functions: an experimental study. In *String Processing and Information Retrieval (**SPIRE**)* (pp. 218–232).