A MATHEMATICAL ANALYSIS

In this section, we provide detailed proofs of the theorems mentioned in the paper.

A.1 Skewness-Aware Error Bound

THEOREM A.1. ***Skewness-Aware Error Bound*** *Assume the persistence of items follows a Zipf distribution with parameter s, i.e., the persistence of the i-th most frequent item is:*

$$p_i = \frac{1}{i^s H_N^{(s)}}, \quad \text{where } H_N^{(s)} = \sum_{k=1}^{N} \frac{1}{k^s}.$$

The expected error upper bound of the Hypersistent Sketch satisfies:

$$\mathbb{E}[\hat{p}_i - p_i] \leq \underbrace{\frac{H_N^{(s)}}{n}}_{\text{Cold-item error}} + \underbrace{\frac{H_N^{(s-1)}}{m}}_{\text{Medium-hot item error}},$$

where n and m are the number of counters in L_1 and L_2 layers of the Cold Filter, respectively.

PROOF. **Stage 1: Cold Items ($p_i \leq \Delta_1$)**

Collision Probability: For items processed in L_1, the hash collision probability is approximated via Poisson distribution:

$$\mathbb{P}_{\text{coll}}^{(1)} = 1 - \left(1 - \frac{1}{n}\right)^{H_N^{(s)}} \approx \frac{H_N^{(s)}}{n}\left(1 - \frac{H_N^{(s)}}{2n}\right).$$

Expected Error: Summing over all cold items:

$$\mathbb{E}[\epsilon_{\text{cold}}] = \sum_{i=1}^{N} p_i \cdot \mathbb{P}_{\text{coll}}^{(1)} = \frac{1}{H_N^{(s)}} \cdot \frac{H_N^{(s)}}{n} \sum_{i=1}^{N} \frac{1}{i^s} = \frac{H_N^{(s)}}{n}.$$

Stage 2: Medium-hot Items ($\Delta_1 < p_i \leq \Delta_1 + \Delta_2$)

Adjusted Distribution: After filtering by L_1, the remaining items follow a truncated Zipf distribution with parameter $s-1$:

$$p_i' \propto \frac{1}{i^{s-1}}, \quad H_N^{(s-1)} = \sum_{k=1}^{N} \frac{1}{k^{s-1}}.$$

Collision Probability: Collision probability in L_2 becomes:

$$\mathbb{P}_{\text{coll}}^{(2)} \approx \frac{H_N^{(s-1)}}{m}.$$

Expected Error:

$$\mathbb{E}[\epsilon_{\text{mid}}] = \sum_{i=1}^{N} p_i' \cdot \mathbb{P}_{\text{coll}}^{(2)} = \frac{H_N^{(s-1)}}{m}.$$

Stage 3: Extreme-hot Items ($p_i > \Delta_1 + \Delta_2$)

Full ID storage in the Hot Part eliminates hash collisions. Replacement errors decay as:

$$\mathbb{P}_{\text{replace}} \sim \frac{1}{\Delta_2 + 1} \to 0. \quad (s \to \infty).$$

Total Error Bound: Combining all stages:

$$\mathbb{E}[\hat{p}_i - p_i] \leq \frac{H_N^{(s)}}{n} + \frac{H_N^{(s-1)}}{m}.$$

Skewness Sensitivity Analysis The error bound varies with s as follows:

Low Skewness ($s \to 0$):

$$\epsilon(s) \approx \frac{N}{n} + \frac{N^2}{m}. \quad \text{(matches uniform distribution)},$$

moderate Skewness ($1 < s < 2$):

$$\epsilon(s) \approx \frac{\zeta(s)}{n} + \frac{N^{2-s}}{m(2-s)},$$

high Skewness ($s \geq 2$):

$$\epsilon(s) \approx \frac{\zeta(s)}{n} + \frac{\zeta(s-1)}{m}.$$

For $s_1 > s_2 \geq 0$, the improvement ratio satisfies:

$$\frac{\epsilon(s_2)}{\epsilon(s_1)} \sim N^{s_1 - s_2}.$$

□

A.2 Threshold Sensitivity Analysis

Theorem A.2. ***Threshold Sensitivity and Pareto Optimality*** *Let the Cold Filter thresholds be parameterized as:*

$$\Delta_1 = k_1 \cdot \frac{\log n}{\log \log n}, \quad \Delta_2 = k_2 \cdot \Delta_1 = k_1 k_2 \cdot \frac{\log n}{\log \log n}.$$

where k_1, k_2 are tunable constants. The memory-error trade-off satisfies:

$$\textit{Memory Efficiency} \propto \frac{1}{k_1 k_2}.$$

$$\textit{Relative Error} \propto \frac{\sqrt{k_1}}{n^{1/2}} + \frac{\sqrt[3]{k_2}}{m^{1/3}}.$$

Pareto optimality is achieved when:

$$k_1 = \Theta\left(\sqrt{\frac{n}{\log n}}\right), \quad k_2 = \Theta\left(\sqrt[3]{\frac{m}{\log m}}\right).$$

Proof. The Cold Filter memory consumption consists of two layers:

$$M_{\text{cold}} = n \cdot \lceil \log_2 \Delta_1 \rceil + m \cdot \lceil \log_2 \Delta_2 \rceil.$$

Substituting the threshold parameterization:

$$\begin{aligned} M_{\text{cold}} &\approx n \log(k_1 \log n) + m \log(k_1 k_2 \log n) \\ &= n(\log k_1 + \log \log n) + m(\log k_1 + \log k_2 + \log \log n) \\ &\approx n \log k_1 + m(\log k_1 + \log k_2), \end{aligned}$$

under fixed total memory $M_{\text{total}} = M_{\text{cold}} + M_{\text{hot}}$:

$$k_1 k_2 \propto \frac{1}{M_{\text{cold}}}.$$

We analyze the relationship between the error and the thresholds, from Theorem IV.6, the error bound can be expressed as:

$$\epsilon \propto \frac{H_N^{(s)}}{n} + \frac{H_N^{(s-1)}}{m}.$$

For general distributions, using moment bounds:

$$H_N^{(s)} \propto N^{1-s}$$
$$H_N^{(s-1)} \propto N^{2-s},$$

substituting the threshold relationships:

$$\begin{aligned}\epsilon &\propto \frac{N^{1-s}}{n} + \frac{N^{2-s}}{m} \\ &= \frac{N^{1-s}}{\Delta_1^{1/2}} + \frac{N^{2-s}}{\Delta_2^{1/3}} \\ &= \frac{\sqrt{k_1}}{n^{1/2}} + \frac{\sqrt[3]{k_2}}{m^{1/3}}.\end{aligned}$$

Pareto Optimality Condition, Define the optimization problem:

$$\min_{k_1,k_2} \left(\frac{\sqrt{k_1}}{n^{1/2}} + \frac{\sqrt[3]{k_2}}{m^{1/3}} \right) \quad \text{s.t.} \quad k_1 k_2 = C.$$

Using Lagrange multipliers with $\mathcal{L} = \frac{\sqrt{k_1}}{n^{1/2}} + \frac{\sqrt[3]{k_2}}{m^{1/3}} + \lambda(k_1 k_2 - C)$:

$$\frac{\partial \mathcal{L}}{\partial k_1} = \frac{1}{2n^{1/2} k_1^{1/2}} + \lambda k_2 = 0$$
$$\frac{\partial \mathcal{L}}{\partial k_2} = \frac{1}{3m^{1/3} k_2^{2/3}} + \lambda k_1 = 0,$$

dividing the two equations:

$$\frac{3m^{1/3} k_2^{2/3}}{2n^{1/2} k_1^{1/2}} = \frac{k_2}{k_1},$$

solving yields the optimal scaling:

$$k_1 \propto \sqrt{\frac{n}{\log n}}, \quad k_2 \propto \sqrt[3]{\frac{m}{\log m}}.$$

We discuss Pareto optimality of threshold parameters, for real-world deployment with $n = m$, the optimal thresholds satisfy:

$$\frac{\Delta_2}{\Delta_1} = \Theta\left((\log n)^{1/6} \right).$$

□