# CMPSCI 711: More Advanced Algorithms

## Section 6-3: Stochastic Streams

Andrew McGregor

Last Compiled: April 29, 2012

# Random-Order Median

- *Goal:* Want to find exact median of a set $S$ of $m$ numbers in $[n]$.
- How much space is required if elements arrive in random order?
- For simplicity assume $m$ odd and elements are distinct.

# Munro-Paterson Algorithm

- Let $S_t$ be first $t$ elements that arrive.
- Algorithm maintains set $A$ of $a$ contiguous (in rank) elements of $S_t$, and $\ell = |\{i \in S_t : i < \min(A)\}|$ and $h = |\{i \in S_t : i > \max(A)\}|$.
- Initially $A \leftarrow S_\ell$ and $\ell = h = 0$.
- For each new item $s$:
  - If $s < \min(A)$ then increment $\ell$.
  - If $s > \max(A)$ then increment $h$.
  - If $\min(A) < s < \max(A)$ then $A \leftarrow A \cup \{s\} \setminus \{s'\}$ where

$$s' = \begin{cases} \min(A) & \text{if } \ell \leq h \\ \max(A) & \text{if } \ell > h \end{cases}$$

    and increment $\ell$ or $h$ based on whether $\min(A)$ or $\max(A)$ removed.
- If at end of stream, $\ell, h < n/2$ we can return the exact median.
- *Thm:* Can find median with high probability if $s = O(\sqrt{m} \log m)$.

# Analysis

- Let $x = \text{median}(S)$.
- For algorithm to fail, there must be some time $t$ when either
  1. $\min(A) \geq x$ and $\ell \leq h$
  2. $\max(A) \leq x$ and $\ell > h$
- Both cases are similar so focus on first case.
- Let $L = S_t \cap \{y < x\}$. Then $|L| \approx t/2$ and specifically,

$$\mathbb{P}\left[||L| - t/2| \leq a/3\right] \geq 1 - 1/m^2$$

  where $a = c\sqrt{m \log m}$ for some $c > 0$
- But $\min(A) \geq x$ and $|L| \geq t/2 - a/3$ implies

$$\ell \geq |L| \geq t/2 - a/3$$

- Contradiction! Since $\ell + h + a = t$, $\ell \leq h$ implies

$$\ell \leq t/2 - a/2$$