Streaming Algs

see elts going by, very little space to store anything.

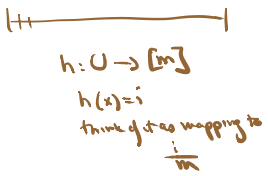
$a_1, a_2, \ldots, a_t \ldots$ each $a_i \in U$

Distinct Elements

	a_1	a_2	a_3	....								
	32	5	17	32	14	5	17	5	17	5	32	17
$Y(t)$	0.43	0.43	0.19	0.19	—	—	—	—	—			0.19

$h: U \to (0,1)$

$h(32) = 0.43$
$h(5) = 0.61$
$h(17) = 0.19$
$h(14) = 0.85$



Define $Y(t) = \min_{1 \le j \le t} h(a_j)$

storing one hash value only

$E[Y(t)] = \frac{1}{n_t + 1}$

n_t = # of distinct elts in $a_1, a_2, \ldots, a_t$

$Y(t) = \frac{1}{n_t+1} \Rightarrow n_t + 1 = \frac{1}{Y(t)}$

$$\hat{n}_t = \frac{1}{Y(t)} - 1$$

$Var(Y(t)) \approx \frac{1}{(n_t+1)^2}$

Chebychev:

$$\Pr\left(|X - E(X)| \geq c\sigma\right) \leq \frac{1}{c^2}$$

$$\Pr\left(|Y(t) - \frac{1}{n_{t+1}}| \geq c\,\frac{1}{n_{t+1}}\right) \leq \frac{1}{c^2}$$