What if we generalize from "frequency" to "point weight"?

↳ "general turnstile model", $(x_1, v_1), \ldots, (x_t, v_t)$ where $v_i \in \mathbb{Z}$ (so $v_i < 0$ possible)

$$n_x = \sum_{\substack{(x_i, v_i) \in S \\ \text{s.t. } x_i = x}} v_i \in \mathbb{Z}$$

in particular, $n_x < 0$ possible

then can't use $\hat{n}_x = \min_{i \in [k]} \{ \hat{n}_x^i \}$ as a reliable estimator

<u>INSTEAD</u>: use $\hat{n}_x = \operatorname{median}_{i \in [k]} \{ \hat{n}_x^i \}$

↳ $\mathbb{E}[\,|\hat{n}_x^i - n_x|\,] = \mathbb{E}[\,|Z_i|\,] \leq \frac{N}{m}$ $\quad N = \sum_{(x_i, v_i) \in S} |v_i|$

(to see this, recall that if all $v_i \geq 0$, then $\mathbb{E}[\,|Z_i|\,] = \mathbb{E}[Z_i] \leq \frac{N}{m}$; and similarly if all $v_i < 0$ then $\mathbb{E}[\,|Z_i|\,] = -\mathbb{E}[Z_i] \leq \frac{N}{m}$. So if some $v_i \geq 0$ and some $v_i < 0$, the expected value must still satisfy $\mathbb{E}[\,|Z_i|\,] \leq N/m$.)

$\forall i \in [k]$:

↳ $$\Pr(|\hat{n}_x^i - n_x| \geq 3\varepsilon N) \underset{\text{by Markov}}{\leq} \frac{N}{m} \frac{1}{3\varepsilon N} = \frac{1}{3\varepsilon N}$$

taking $m = \frac{e}{\varepsilon}$ as before

$$= \frac{1}{3e}$$

$$\leq \frac{1}{8}$$

↳ Let $B_i = 1 \Leftrightarrow |\hat{n}_x^i - n_x| \geq 3\varepsilon N$. Then: $\mathbb{E}[B_i] \leq \frac{1}{8}$.

↳ Let $B = \sum_{i=1}^{k} B_i$, then $\mathbb{E}[B] \leq \frac{k}{8}$.

Now, recall that $\hat{n}_x = \text{median}\{\hat{n}_x^1, \dots, \hat{n}_x^k\}$.

Then $\Pr(|\hat{n}_x - n_x| \geq 3\varepsilon N)$

$= \Pr(B \geq \frac{k}{2})$ by the "Median trick"

Note that the $B_i$ are i.i.d., so Chernoff bounds apply to $B = \sum_i B_i$.

$$\Pr(B \geq \tfrac{k}{2}) = \Pr(B \geq 4(\tfrac{k}{8}))$$
$$\leq \Pr(\tilde{B} \geq 4(\tfrac{k}{8}))$$

where $\tilde{B} = \sum_{i=1}^{k} \tilde{B}_i$ and the $\tilde{B}_i$ are such that

$\Pr(\tilde{B}_i = 1) = \frac{1}{8}$, rather than $\Pr(B_i = 1) \leq \frac{1}{8}$

So $\mu = \mathbb{E}[\tilde{B}] = \frac{k}{8}$. (Making the $\tilde{B}_i$ / $\tilde{B}$ easier to work with)

$$\Pr(B \geq \tfrac{k}{2}) \leq \Pr(\tilde{B} \geq 4\mu) \leq \left(\frac{e^3}{4^4}\right)^{k/8}$$

Now, some "trickery": $4^4 = e^{\ln 4^4} = e^{4\ln 4}$,

so $\frac{e^3}{4^4} = e^{(3 - 4\ln 4)} \leq e^{-2}$

↳ $\Pr(B \geq \frac{k}{2}) \leq e^{-k/4}$ so with $k = c\ln(\frac{1}{\delta})$

$\leq \delta^{c/4}$  $\left(\frac{1}{1000}\right)^{1/4} = 0.178$, $\left(\frac{1}{1000}\right)^{1/2} = 0.03$

$k = 7$  $k = 14$