A PROOF OF THEOREM 5.2

Proof. Let $\mathcal{B} = \{B_{y_1},...,B_{y_t}\}$, and $\overline{\mathcal{B}} = \{B_{y_{t+1}},...,B_{y_{d'}}\}$. We start from the state where all buckets hold non-zero values, i.e., $B_{y_i} \neq 0, i \in \{1,...,d'\}$, and reconstruct the frequency histogram for poison values in $[-C, C]$. The likelihood estimator in Equ. 1 becomes

$$l(F) = \sum_{i=1}^{N} ln(\sum_{k=1}^{d} \hat{x_k} Pr[v_i'|v_i \in B_{x_k}] + \sum_{j=1}^{d'} \hat{y_j} Pr[v_i'|v_i \in B_{y_j}])$$
$$= \sum_{t=1}^{d'} n_t ln(\sum_{k=1}^{d} \hat{x_k} M_{b_t x_k} + \sum_{j=1}^{d'} \hat{y_j} M_{b_t y_j}).$$

Note that $\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{d'} \hat{y_j} = 1$, we employ the Lagrangian multiplier method to derive the extremum. The Lagrangian function can be written as:

$$L(F) = l(F) + \lambda(\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{d'} \hat{y_j} - 1). \tag{12}$$

Let all first-order partial derivatives of L w.r.t. $\hat{x_k}$ and $\hat{y_j}$ equal zero

$$\frac{\partial L(F)}{\partial \hat{x_k}} = \sum_{t=1}^{d'} n_t \frac{M_{b_t x_k}}{\sum_{k=1}^{d} \hat{x_k} M_{b_t x_k} + \hat{y_t}} + \lambda = 0,\ k \in \{1,...,d\}$$
$$\frac{\partial L(F)}{\partial \hat{y_j}} = \sum_{t=1}^{d'} n_t \frac{M_{b_t y_j}}{\sum_{k=1}^{d} \hat{x_k} M_{b_t x_k} + \hat{y_t}} + \lambda = 0,\ j \in \{1,...,d'\},$$

we have:

$$\hat{x_k} = 0,\ k \in \{1,...,d\},\ \ \hat{y_j} = \frac{n_j}{N},\ j \in \{1,...,d'\},\ \lambda = -N.$$

This result shows all collected values converge to poison values if no bucket is suppressed.

When we suppress the bucket $B_{y_{d'}}$ (by setting $\hat{y_{d'}} = 0$) and carry out EMF, the collected values in $B'_{b_{d'}}$ can only converge to the normal values. The collected values in $B'_{b_j} (j \in \{1,...,t\})$ will converge to both the poison values in $B_{y_j} (j \in \{1,...,t\})$ and the normal values, so:

$$\frac{n_{d'}}{N} + \sum_{j=1}^{t} \frac{n_j}{N} \leq (\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0}.$$

Then suppress the bucket $B_{y_{d'-1}}$, similarly, we have:

$$\frac{n_{d'} + n_{d'-1}}{N} + \sum_{j=1}^{t} \frac{n_j}{N} \leq (\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0, y_{d'-1}=0}.$$

Since the collected values in $B'_{b_{d'-1}}$ can only converge to the normal values and the decrement of collected values in $B'_{b_j} (j \in \{1,...,t\})$ will also converge to the normal values, we have:

$$(\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0} \leq (\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0, y_{d'-1}=0}. \tag{13}$$

Suppress all buckets in $\overline{\mathcal{B}}$. Similar to Equ. 13, we have:

$$(\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0} \leq (\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0, y_{d'-1}=0} \leq$$
$$... \leq (\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0,...,y_{t+1}=0},$$

and we infer that:

$$\frac{\sum_{i=t+1}^{d'} n_i}{N} + \sum_{j=1}^{t} \frac{n_j}{N} = (\sum_{k=1}^{d} \hat{x_k} + \sum_{j=1}^{t} \hat{y_j})\Big|_{y_{d'}=0,...,y_{t+1}=0} = 1,$$

which is the optimal case where none of the collected values will converge to buckets in $\overline{\mathcal{B}}$.

When the number of suppressed buckets in $\overline{\mathcal{B}}$ increases, the corresponding interference of $\overline{\mathcal{B}}$ decreases. Therefore, the collected values more accurately converge to the buckets that they should belong to, and thus achieve a better convergence result. □

B PROOF OF THEOREM 5.3

Proof. Let v_{tj} denote the j-th value in group G_t, v'_{tj} denote the perturbed v_{tj}, and M_t denote the mean value of v'_{tj}. The variance of $\tilde{M}$, which is a linear combination of M_t, can be written as:

$$\begin{aligned} Var(\tilde{M}) &= Var(\sum_{t=1}^{h} w_t M_t) = \sum_{t=1}^{h} w_t^2 Var(M_t) \\ &= \sum_{t=1}^{h} w_t^2 Var(\frac{\sum_{j=1}^{n'} v'_{tj}}{n'}) = \sum_{t=1}^{h} \frac{w_t^2}{n'^2} \sum_{j=1}^{n'} Var(v'_{tj}), \end{aligned} \tag{14}$$

where $\sum w_t = 1$, and n' is the number of normal users in each group.

Since $Var(v'_{tj})$ in Equ. 14 relies on the input of each user, we consider the worst-case at the maximum variance, i.e., all inputs v_{tj} are either 1 or -1. The worst-case variance $Var_{worst}(v'_{tj})$ can be expressed as:

$$Var_{worst}(v'_{tj}) = \frac{v_{tj}^2}{e^{\epsilon_t/2} - 1} + \frac{e^{\epsilon_t/2} + 3}{3(e^{\epsilon_t/2} - 1)^2}\Big|_{v_{tj}=\pm 1}$$
$$= \frac{1}{e^{\epsilon_t/2} - 1} + \frac{e^{\epsilon_t/2} + 3}{3(e^{\epsilon_t/2} - 1)^2}.$$

Let $B_t = n' Var_{worst}(v'_{tj})$. Equ. 14 can be rewritten as:

$$Var(\tilde{M}) = \sum_{t=1}^{h} \frac{w_i^2}{n'^2} B_t. \tag{15}$$

We regard the variance as a function of w_t, and the minimal variance is the extreme point of Equ. 15. By the Lagrangian method, we have:

$$\mathcal{L} = \sum_{t=1}^{h} \frac{w_t^2}{n'^2} B_t + C_0(1 - \sum_{t=1}^{h} w_t).$$

The first partial derivatives of $\mathcal{L}$ w.r.t. w_t is:

$$\frac{\partial \mathcal{L}}{\partial w_t} = \frac{2w_t}{n'^2} B_t - C_0. \tag{16}$$

Let $\frac{\partial \mathcal{L}}{\partial w_t} = \frac{2w_t}{n'^2} B_t - C_0 = 0$, then we have $wt = \frac{C_0 n'^2}{2B_t}$. Through the restriction $\sum_{t=1}^{h} w_t = 1$, we figure out

$$C_0 = \frac{2}{n'^2 \sum_{t=1}^{h} \frac{1}{B_t}},\ w_t = \frac{1}{B_t \sum_{i=1}^{h} \frac{1}{B_i}}.$$

And the final minimal variance of $\tilde{M}$ is:

$$Var(\tilde{M})_{min} = [\sum_{t=1}^{h} \frac{n'^2}{B_t}]^{-1}.$$

□