

A PROOF OF THEOREM 5.2

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

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

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:

$$L(F) = l(F) + \lambda \left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^{d'} \hat{y}_j - 1 \right). \quad (12)$$

Let all first-order partial derivatives of L w.r.t. \hat{x}_k and \hat{y}_j equal zero

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

we have:

$$\hat{x}_k = 0, k \in \{1, \dots, d\}, \hat{y}_j = \frac{n_j}{N}, j \in \{1, \dots, d'\}, \lambda = -N.$$

This result shows all collected values converge to poison values if no bucket is suppressed, and we can obtain:

$$\left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^{d'} \hat{y}_j \right) \Big|_{y_i \neq 0, i \in \{1, \dots, d'\}} = \sum_{j=1}^{d'} \frac{n_j}{N}.$$

When we suppress the bucket $B_{y_{d'}}$ (by setting $\hat{y}_{d'} = 0$) and carry out EMF. The collected values in $B'_{b_{d'}}$ will converge to all buckets for normal users, because normal values from each B_{x_k} ($k \in \{1, \dots, d\}$) can be perturbed into $B'_{b_{d'}}$. Hence, every \hat{x}_k will increase. However, normal values from buckets B_{x_k} can be perturbed into other buckets B'_{b_i} ($i \in \{1, \dots, d' - 1\}$), so some collected data in B'_{b_i} ($i \in \{1, \dots, d' - 1\}$) will also converge to normal values and those collected values converge to poison values decrease. Therefore, suppressing $B_{y_{d'}}$ leads to the increase of all \hat{x}_k , which in turn results in the decrease of all \hat{y}_j . However, since the decrease of \hat{y}_j ($j \in \{1, \dots, t\}$) is a part of increment of \hat{x} , we can figure out:

$$\left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_i \neq 0, i \in \{1, \dots, d'\}} \leq \left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_{d'}=0}.$$

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

$$\left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_{d'}=0} \leq \left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_{d'}=0, y_{d'-1}=0}. \quad (13)$$

Suppress all buckets in $\bar{\mathcal{B}}$, we have:

$$\left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_i \neq 0, i \in \{1, \dots, d'\}} \leq \left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_{d'}=0} \leq \\ \left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_{d'}=0, y_{d'-1}=0} \leq \dots \leq \left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_{d'}=0, \dots, y_{t+1}=0}.$$

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

After suppressing all buckets in $\bar{\mathcal{B}}$, all collected data will convergence to normal values and poison values in B_{y_j} ($j \in \{1, \dots, t\}$) and we can infer that:

$$\left(\sum_{k=1}^d \hat{x}_k + \sum_{j=1}^t \hat{y}_j \right) \Big|_{y_{d'}=0, \dots, y_{t+1}=0} = 1,$$

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

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:

$$\text{Var}(\tilde{M}) = \text{Var} \left(\sum_{t=1}^h w_t M_t \right) = \sum_{t=1}^h w_t^2 \text{Var}(M_t) \\ = \sum_{t=1}^h w_t^2 \text{Var} \left(\frac{\sum_{j=1}^{n'} v'_{tj}}{n'} \right) = \sum_{t=1}^h \frac{w_t^2}{n'^2} \sum_{j=1}^{n'} \text{Var}(v'_{tj}), \quad (14)$$

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

Since $\text{Var}(v'_{tj})$ 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 $\text{Var}_{\text{worst}}(v'_{tj})$ can be expressed as:

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

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

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

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:

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

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

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

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

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

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

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

□