Supplements of "Improving Count-Mean Sketch as the Leading Locally Differentially Private Frequency Estimator for Large Dictionaries"

Mingen Pan
Email: mepan94@gmail.com

S1. Derivation of the Expectation

Let's focus on the term $\hat{y}^{(i)}[h_{j^{(i)}}(x)]$ in the definition of $f(x)$ in Eq. (2), which can be expanded as

$$\begin{aligned}\hat{y}^{(i)}[h_{j^{(i)}}(x)] &= R(h_{j^{(i)}}(X^{(i)}))[h_{j^{(i)}}(x)] \\ &= \sum_{j\in[k]} \mathbf{1}\{j = j^{(i)}\} R(h_j(X^{(i)}))[h_j(x)], \quad \text{(S1)}\end{aligned}$$

where $\mathbf{1}\{*\}$ returns one if $*$ is true. Given that $h_{j^{(i)}}$ is uniformly sampled from $\mathcal{H}$, we have

$$\begin{aligned}E[\hat{y}^{(i)}[h_{j^{(i)}}(x)]] &= \sum_{j\in[k]} E\big[\mathbf{1}\{j = j^{(i)}\} \\ & R(h_j(X^{(i)}))[h_j(x)]\big] = \sum_{j\in[k]} \frac{1}{k} c_j(X^{(i)}, x), \quad \text{(S2)}\end{aligned}$$

where $c_j(X^{(i)}, x)$ indicates whether $X^{(i)}$ and x collide in the hash function h_j, and $E\big[R(h_j(X^{(i)}))[h_j(x)]\big] = c_j(X^{(i)}, x)$ is derived from Property 2.1. Substituting the above equation into Eq. (2), we have

$$\begin{aligned}E[\hat{f}(x)] &= \frac{m}{n(m-1)} \sum_{i\in[n]} E\big[\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big] - \frac{1}{m-1} \\ &= \frac{m}{n(m-1)} \sum_{i\in[n]} \sum_{j\in[k]} \frac{c_j(X^{(i)}, x)}{k} - \frac{1}{m-1} \\ &= \frac{m}{m-1}[f(x) + \sum_{x'\in[d]\setminus x} \sum_{j\in[k]} \frac{c_j(x, x')}{k} f(x')] - \frac{1}{m-1},\end{aligned}$$

and Theorem 3.2 is proved. □

Substituting $\forall x, x'$, where $x \neq x'$ and $\sum_{j\in[k]} \frac{c_j(x,x')}{k} = \frac{1}{m}$ into the above equation, we have

$$\begin{aligned}E[\hat{f}(x)] &= \frac{m}{m-1}[f(x) + \sum_{x'\in[d]\setminus x} \frac{1}{m} f(x')] - \frac{1}{m-1} \\ &= \frac{m}{m-1}[f(x) + (1 - f(x))\frac{1}{m}] - \frac{1}{m-1} = f(x).\end{aligned}$$

Regarding its inverse, whose condition is that there exists an x' such that $\sum_{j\in[k]} \frac{c_j(x,x')}{k} \neq \frac{1}{m}$. If we construct a dataset containing only x' and x, then we have

$$\begin{aligned}E[\hat{f}(x)] = \frac{m}{m-1}[f(x) + (1 - f(x)) \sum_{j\in[k]} \frac{c_j(x, x')}{k}] \\ - \frac{1}{m-1} \neq f(x),\end{aligned}$$

so Corollary 3.1 is proved. □

S2. Derivation of the Variance

Let's start with the variance of an individual response $\hat{y}^{(i)}[h_{j^{(i)}}(x)]$:

$$\begin{aligned}&Var\big(\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big) \\ &\quad = E\big[\big(\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big)^2\big] - \bar{c}(X^{(i)}, x)^2, \quad \text{(S3)}\end{aligned}$$

where $E\big[\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big] = \bar{c}(X^{(i)}, x)$ given Eq. (S2). Then,

$$\begin{aligned}&E\big[\big(\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big)^2\big] \\ &= \sum_{j\in[k]} E\big[\big(\mathbf{1}\{j = j^{(i)}\} R(h_j(X^{(i)}))[h_j(x)]\big)^2\big] \\ &= \sum_{j\in[k]} E[\mathbf{1}\{j = j^{(i)}\}] E\big[\big(R(h_j(X^{(i)}))[h_j(x)]\big)^2\big] \\ &= \frac{1}{k} \sum_{j\in[k]} E\big[\big(R(h_j(X^{(i)}))[h_j(x)]\big)^2\big] \\ &\overset{\text{(a)}}{=} \frac{1}{k} \sum_{j\in[k]} \Big(Var\big(R(h_j(X^{(i)}))[h_j(x)]\big) + c_j(X^{(i)}, x)\Big)\end{aligned}$$

where $\overset{\text{(a)}}{=}$ holds because $E\big[R(h_j(X^{(i)}))[h_j(x)]\big]^2 = c_j(X^{(i)}, x)^2 = c_j(X^{(i)}, x)$. Integrating the above equation into Eq. (S3), we have

$$\begin{aligned}Var\big(\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big) &= \frac{1}{k} \sum_{j\in[k]} Var\big(R(h_j(X^{(i)}))[h_j(x)]\big) \\ &+ \bar{c}(X^{(i)}, x) - \bar{c}(X^{(i)}, x)^2. \quad \text{(S4)}\end{aligned}$$

Now, we will study the variance of $\sum_{i\in[n]} \hat{y}^{(i)}[h_{j^{(i)}}(x)]$:

$$Var\big(\sum_{i\in[n]} \hat{y}^{(i)}[h_{j^{(i)}}(x)]\big) = \sum_{i\in[n]} Var\big(\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big) + \sum_{i_1\neq i_2} Cov\big(\hat{y}^{(i_1)}[h_{j^{(i_1)}}(x)], \hat{y}^{(i_2)}[h_{j^{(i_2)}}(x)]\big).$$

Focusing on the covariance and expanding $\hat{y}^{(i)}[h_{j^{(i)}}(x)]$ as Eq. (S1), we have

$$Cov\big(\hat{y}^{(i_1)}[h_{j^{(i_1)}}(x)], \hat{y}^{(i_2)}[h_{j^{(i_2)}}(x)]\big) = Cov\big(\sum_{j\in[k]} \mathbf{1}\{j = j^{(i_1)}\}R(h_j(X^{(i_1)}))[h_j(x)], \sum_{j\in[k]} \mathbf{1}\{j = j^{(i_2)}\}R(h_j(X^{(i_2)}))[h_j(x)]\big).$$

Given that $Cov(X, Y) = E[XY] - E[X]E[Y]$, we calculate:

$$\begin{aligned}
&E\big[\sum_{j_1\in[k]}\sum_{j_2\in[k]} \mathbf{1}\{j_1 = j^{(i_1)}\}\mathbf{1}\{j_2 = j^{(i_2)}\} R(h_{j_1}(X^{(i_1)}))[h_{j_1}(x)]R(h_{j_2}(X^{(i_2)}))[h_{j_2}(x)]\big]\\
&\overset{\text{(a)}}{=} \sum_{j_1\in[k]}\sum_{j_2\in[k]} E[\mathbf{1}\{j_1 = j^{(i_1)}\}]E[\mathbf{1}\{j_2 = j^{(i_2)}\}]\times E[R(h_{j_1}(X^{(i_1)}))[h_{j_1}(x)]]E[R(h_{j_2}(X^{(i_2)}))[h_{j_2}(x)]]\\
&= \sum_{j_1\in[k]}\sum_{j_2\in[k]} \frac{1}{k^2} c_{j_1}(X^{(i_1)}, x)c_{j_2}(X^{(i_2)}, x)\\
&= \bar{c}(X^{(i_1)}, x)\bar{c}(X^{(i_2)}, x), \qquad \text{(S5)}
\end{aligned}$$

where $\overset{\text{(a)}}{=}$ holds because the assignment of a hash function is simply a uniform sampling, which is independent of everything, and $R(h_{j_1}(X^{(i_1)}))[h_{j_1}(x)]$ is also independent of $R(h_{j_2}(X^{(i_2)}))[h_{j_2}(x)]$ given Property 2.2.

Utilizing Eq. (S2) and (S5), we have

$$\begin{aligned}
&Cov\big(\hat{y}^{(i_1)}[h_{j^{(i_1)}}(x)], \hat{y}^{(i_2)}[h_{j^{(i_2)}}(x)]\big)\\
&= \bar{c}(X^{(i_1)}, x)\bar{c}(X^{(i_2)}, x) - \bar{c}(X^{(i_1)}, x)\bar{c}(X^{(i_2)}, x) = 0.
\end{aligned}$$

As a result, we have

$$\begin{aligned}
Var\big(\sum_{i\in[n]} \hat{y}^{(i)}[h_{j^{(i)}}(x)]\big) &= \sum_{i\in[n]} Var\big(\hat{y}^{(i)}[h_{j^{(i)}}(x)]\big)\\
&\overset{\text{Eq. (S4)}}{=} \sum_{i\in[n]} \Big[\Big(\frac{1}{k}\sum_{j\in[k]} Var\big(R(h_j(X^{(i)}))[h_j(x)]\big)\Big) + \bar{c}(X^{(i)}, x) - \bar{c}(X^{(i)}, x)^2\Big]\\
&= n\sum_{x'\in[d]} \Big[\Big(\frac{1}{k}\sum_{j\in[k]} Var\big(R(h_j(x'))[h_j(x)]\big)\Big) + \bar{c}(x', x) - \bar{c}(x', x)^2\Big] f(x').
\end{aligned}$$

and subsequently, we have

$$Var(\hat{f}(x)) = \frac{m^2}{(m-1)^2 n}\sum_{x'\in[d]} \Big[\Big(\frac{1}{k}\sum_{j\in[k]} Var\big(R(h_j(x'))[h_j(x)]\big)\Big) + \bar{c}(x', x) - \bar{c}(x', x)^2\Big] f(x'), \qquad \text{(S6)}$$

which proves Eq. (4). □

When considering Property 2.3, we have $Var\big(R(h_j(x'))[h_j(x)]\big) = Var(R|=)$ if $h_j(x') = h_j(x)$. Otherwise, it equals $Var(R|\neq)$. When $x' = x$, they will collide in every hash function, so $\bar{c}(x', x) = 1$ and $Var\big(R(h_j(x'))[h_j(x)]\big) = Var(R|=)$ in this case. On the other hand, when $x' \neq x$, we have

$$\begin{aligned}
&\frac{1}{k}\sum_{j\in[k]} Var\big(R(h_j(x'))[h_j(x)]\big)\\
&= \frac{1}{k}\sum_{j\in[k]} [c_j(x', x)Var(R|=) + (1 - c_j(x', x))Var(R|\neq)]\\
&= \bar{c}(x', x)Var(R|=) + (1 - \bar{c}(x', x))Var(R|\neq). \qquad \text{(S7)}
\end{aligned}$$

Therefore, when substituting Eq. (S7) for $x' \neq x$ and the aforementioned parameters with $x = x'$ into Eq. (4), we derived Eq. (5). If $\bar{c}(x', x) = \frac{1}{m}$, it becomes Eq. (6). □

S3. Proof of Theorem 3.5

Recall the definition of $\hat{f}(x)$ in Eq. (2), which is equivalent to the sum of bounded random variables. Define $Y^{(i)} = \frac{m}{m-1}\hat{y}^{(i)}[h_j^{(i)}(x)] - \frac{1}{m-1}$, and we have $n\hat{f}(x) = \sum_i Y^{(i)}$. Note that $\frac{Y^{(i)}-A}{B-A}$ is a Bernoulli random variable where

$$\begin{aligned}
A &= \min \hat{f}(x) = \frac{m}{m-1}\frac{e^{-1}}{e^{\epsilon}-1} - \frac{1}{m-1}\\
B &= \max \hat{f}(x) = \frac{m}{m-1}\frac{e^{\epsilon}+m-2}{e^{\epsilon}-1} - \frac{1}{m-1}.
\end{aligned}$$

Here, we assume that CMS uses RR to perturb the hashed values. Also note that we only consider unbiased CMS, i.e., $E[\hat{f}(x)] = f(x)$. Using the Chernoff (upper) bound, we have $\forall\delta \in [0, 1]$

$$Pr[\sum_i \frac{Y^{(i)} - A}{B - A} \geq (1 + \delta)n\frac{f(x) - A}{B - A}] \leq \exp(-\frac{n}{3}\frac{f(x) - A}{B - A}\delta^2). \qquad \text{(S8)}$$

Define $\delta = \frac{\alpha\sqrt{Var(\hat{f}(x))}}{f(x)-A}$. Given that $\forall\delta \in [0, 1]$, it is equivalent to requiring $\alpha \in [0, \sqrt{\frac{ne^{\epsilon}}{m-1}}]$. Thus, Eq. (S8) can be rewritten as

$$Pr[\hat{f}(x) \geq f(x) + \alpha\sqrt{Var(\hat{f}(x))}] \leq \exp(-\frac{n}{3}\frac{Var(\hat{f}(x))}{(B - A)(f(x) - A)}\alpha^2)$$

$Var(\hat{f}(x))$ is derived in Eq. (19), and $\frac{Var(\hat{f}(x))}{(B-A)(f(x)-A)}$ decreases with $f(x)$ (one can verify it using derivatives). Calculating the minimum at $f(x)=1$, we have

$$\frac{Var(\hat{f}(x))}{(B-A)(f(x)-A)} \geq \frac{m-1}{e^\epsilon+m-1}.$$

As a result,

$$Pr[\hat{f}(x) \geq f(x) + \alpha\sqrt{Var(\hat{f}(x))}] \leq \exp(-\frac{n}{3}\frac{m-1}{e^\epsilon+m-1}\alpha^2).$$

The same proof is also applicable to $\hat{f}(x) \leq f(x) - \alpha\sqrt{Var(\hat{f}(x))}$. Therefore,

$$Pr[|\hat{f}(x) - f(x)| \geq +\alpha\sqrt{Var(\hat{f}(x))}] \leq 2\exp(-\frac{n}{3}\frac{m-1}{e^\epsilon+m-1}\alpha^2). \quad \square$$

S4. Proof regarding Preferring RR to RAPPOR in CMS

The variance of symmetry and asymmetry RAPPOR is listed below:

$$\forall a, b : Var(sRP(a)[b]) = \frac{e^{\epsilon/2}}{(e^{\epsilon/2}-1)^2}. \quad \text{(S9)}$$

$$Var(aRP(a)[b]) = \frac{1}{(e^\epsilon-1)^2}\begin{cases}(e^\epsilon+1)^2 & \text{if } a=b \\ 4e^\epsilon & \text{if } a \neq b,\end{cases} \quad \text{(S10)}$$

Since their $Var(a)[a]$ and $Var(a)[b]$ are independent of m, the corresponding $Var(\hat{f}(x))$ decreases when m increases. Thus, when m is large enough, we have

$$Var(\hat{f}(x)_{RP}) \to \frac{1}{n}[Var(R|\neq)(1-f(x)) + Var(R|=)f(x)]. \quad \text{(S11)}$$

Following Appendix B and given unbiased CMS, the worst-case MSE of symmetry and asymmetry RAPPOR are identical to their variance, which are $\frac{e^{\epsilon/2}}{n(e^{\epsilon/2}-1)^2}$ and $\frac{f^*(e^\epsilon-1)^2+4^\epsilon}{n(e^\epsilon-1)^2}$, respectively.

For asymmetric RAPPOR, we realize $Var(\hat{f}(x)_{aRP}) > Var^*(\hat{f}(x)_{RR})$ when $f^* > 0$, so the CMS with asymmetric RAPPOR cannot serve as the optimized CMS. For symmetric RAPPOR, $Var(\hat{f}(x)_{sRP}) > Var^*(\hat{f}(x)_{RR})$ when $0 \leq f^* < \frac{1}{2}$, and $Var(\hat{f}(x)_{sRP}) = Var^*(\hat{f}(x)_{RR})$ when $\frac{1}{2} \leq f^* \leq 1$. Thus, the CMS with symmetric RAPPOR is also ruled out. As a result, The choice of RR is preferred.

Similar to Appendix B, we can use the variance to derive l_2 loss, which are $\frac{de^{\epsilon/2}}{n(e^{\epsilon/2}-1)^2}$ and $\frac{4de^\epsilon}{n(e^\epsilon-1)^2}$ for symmetric and asymmetric RAPPOR, respectively. we observe that CMS with symmetric RAPPOR will always have a larger l_2 loss than the CMS+RR. For asymmetric RAPPOR, its l_2 loss is the same as that of CMS+RR only when d is large enough. However, its communication cost is too high because RAPPOR requires sending a vector of m bits to the server, and Eq. (S11) requires $m \to \infty$. To determine the necessary value of m, we solve the following equation for asymmetric RAPPOR:

$$l_2(\hat{f}_{aRP}) = dVar(\hat{f}(x)_{aRP}) = (1+\tau)\frac{4de^\epsilon}{(e^\epsilon-1)^2},$$

where τ is a small number like 0.01. m is solved as

$$m = \Omega(\frac{(e^\epsilon-1)^2}{(1+\tau)e^\epsilon}). \quad \text{(S12)}$$

On the other hand, CMS+RR only requires $\log_2(1+e^\epsilon)$ bits to output the perturbed result. Thus, CMS+RR is always preferred when optimizing l_2 loss due to its low communication cost.

S5. Proof regarding Imperfect Hashing

Let's start by proving Theorem 3.9. Observe $\frac{\partial Var(g(x|\mathcal{H}_{api}))}{\partial m'}$, and note that it is linear to $f(x)$. Considering the derivative at $f(x)=1$ and $f(x)=0$ yields:

$$-\max\{\frac{2m'Var(R|=)}{n(m'-1)^3}, \frac{(m'-1)(Var(R|\neq)+1)+(m'+1)Var(R|=)}{n(m'-1)^3}\} \leq \frac{\partial Var(g(x|\mathcal{H}_{api}))}{\partial m'} < 0. \quad \text{(S13)}$$

Observing that $Var(R|=) \geq Var(R|\neq)$ is satisfied by the LDP mechanism such as randomized response and RAPPOR, we have

$$-\frac{2m'Var(R|=)+m'-1}{n(m'-1)^3} \leq \frac{\partial Var(g(x|\mathcal{H}_{api}))}{\partial m'} < 0. \quad \text{(S14)}$$

At the same time, we have $m' = m - \frac{m}{(2q+1)^2}$. Denote $Var(\hat{f}(x))$ as the variance of the CMS using perfect hashing. We want $Var(g(x|\mathcal{H}_{api}))$ to approach $Var(\hat{f}(x))$, which is formulated as $Var(g(x|\mathcal{H}_{api})) \leq Var(\hat{f}(x))(1+\tau)$, where τ is a small number like 0.01. This can be translated to

$$\frac{2mVar(R|=)+m-1}{n(m-1)^3}\frac{m}{(2q+1)^2} \leq \tau Var(\hat{f}(x)),$$

which can be organized as

$$\frac{2mVar(R|=)+m-1}{n(m-1)^3}\frac{m}{\tau Var(\hat{f}(x))} \leq (2q+1)^2, \quad \text{(S15)}$$

thus proving Theorem 3.9. When $m = 1+e^{\epsilon/2}$, $Var(\hat{f}(x))$ becomes $\frac{e^{\epsilon/2}}{(e^{\epsilon/2}-1)^2}$, and $\frac{Var(R|=)}{Var(\hat{f}(x))}$ decreases with ϵ. Thus, we can substitute $\epsilon = 0$ into Eq. (S15), and we have $2q + 1 > \sqrt{\frac{1}{\tau}}$. The MSE part of Corollary 3.5 is proved. $\square$

If $\forall x : Var(g(x)) \leq (1+\tau)Var(\hat{f}(x))$, then $l_2(g) \leq (1+\tau)l_2(\hat{f})$. When $m = 1 + e^\epsilon$, $Var(\hat{f}(x))$ increases with $f(x)$, so we have $Var(\hat{f}(x)|f(x) = 0) \leq Var(\hat{f}(x))$. Thus, if

$$\frac{2mVar(R|=) + m - 1}{n(m-1)^3} \frac{m}{\tau Var(\hat{f}(x)|f(x)=0)} \leq (2q+1)^2 \quad \text{(S16)}$$

is satisfied, Eq. (S15) will also be satisfied. Substituting $\frac{4e^\epsilon}{(e^\epsilon-1)^2}$ into $Var(\hat{f}(x)|f(x) = 0)$ in Eq. (S16) and setting $m = 1 + e^\epsilon$, we also have the left-hand side of Eq. (S16) decreasing with ϵ. Since this inequality is still satisfied when $\epsilon = 0$, it derives $2q+1 > \sqrt{\frac{1}{\tau}}$. Thus, The l_2 part of Corollary 3.5 is proved. $\square$

S6. Proof of Theorem 3.10

Consider each $c_j(x, x')$ as an independent Bernoulli random variable with probability $\frac{1}{m}$ of being one. Given that $E[c_j(x, x')] = \frac{1}{m}$ and $Var(c_j(x, x')) = \frac{m-1}{m}$, we have

$$\begin{aligned} \underset{\forall c_j(x,x')}{E}[E[\hat{f}(x)]] &= \frac{m}{m-1}[f(x) + \sum_{x' \in [d] \setminus x} \frac{k}{km} f(x')] - \frac{1}{m-1} \\ &= \frac{m}{m-1}[f(x) + (1 - f(x))\frac{1}{m}] - \frac{1}{m-1} = f(x), \end{aligned}$$

and

$$\begin{aligned} \underset{\forall c_j(x,x')}{Var}[E[\hat{f}(x)]] &= \frac{m^2}{(m-1)^2}\Big(\sum_{x' \in [d] \setminus x} \frac{m-1}{m^2 k} f(x')^2\Big) \\ &= \frac{1}{(m-1)k} \sum_{x' \in [d] \setminus x} f(x')^2. \end{aligned}$$

Thus, Theorem 3.10 is proved. $\square$

S7. Decoding Hadamard Encoding

Define H to be a scaled Walsh Hadamard matrix, where $H[i, j] = (-1)^{i \cdot j}$ with $\cdot$ representing bitwise multiplication. Hadamard encoding employs a LDP mechanism to perturb $H[X^{(i)} + 1, j^{(i)}]$ of each object, where $X^{(i)}$ and $j^{(i)}$ are as defined in Section 3. [1] has proved that

$$\hat{f}_H(x) = \sum_{i \in [n]} \hat{H}[X^{(i)} + 1, j^{(i)}]H[j^{(i)}, x + 1]$$

unbiasedly estimates the frequency of $f(x)$. Here, $\hat{H}[X^{(i)} + 1, j^{(i)}] = R(H[X^{(i)} + 1, j^{(i)}])$, with R denoting the LDP reconstruction process as detailed in Section 2.2.

Interpreting $H[j^{(i)}, :]$ as a hash function, $\hat{H}[X^{(i)} + 1, j^{(i)}]H[j^{(i)}, x+1]$ yields +1 if both $X^{(i)}$ and x hash to the same value, and -1 otherwise. This behavior is analogous to $\hat{y}^{(i)}[h_{j^i}(x)]$ in Eq. (2). However, $\hat{y}^{(i)}[h_{j^i}(x)]$ returns zero when $X^{(i)}$ and x hash to different values. By considering the constants $\frac{m}{m-1}$ and $\frac{1}{m-1}$ in Eq. (2), and setting $m = 2$, we derive $2\hat{y}^{(i)}[h_{j^i}(x)] - 1$, which produces results identical to $\hat{H}[X^{(i)}+1, j^{(i)}]H[j^{(i)}, x+1]$. Consequently, the decoding process for Hadamard encoding is equivalent to Eq. (2).

S8. Upper Bound of the worst-case MSE estimator

The upper bound of $\overline{MSE}(\hat{f})$ is set to be $[1 + \frac{2}{t}\big(\sqrt{t\log(20|\mathbf{x}|)} + \log(20|\mathbf{x}|)\big)]V$ based on the following theorem:

Theorem S8.1. *If $\forall x \in \mathbf{x} : E[(\hat{f}(x) - f(x))^2] \leq V$, where V is constant, we have*

$$Pr\big(\overline{MSE}(\hat{f}) \geq [1 + \frac{2}{t}\big(\sqrt{t\log(20|\mathbf{x}|)} + \log(20|\mathbf{x}|)\big)]V\big) \leq 0.05,$$

where t is the experiment rounds.

Proof: Similar to Supplement S3, we define $Y^{(i)} = \frac{m}{m-1}\hat{y}^{(i)}[h_j^{(i)}(x)] - \frac{1}{m-1}$, and $\frac{Y^{(i)}-A}{B-A}$ is a Bernoulli random variable (see Supplement S3 for the definition of A and B). All the $X^{(i)}$ can be placed into two groups based on whether $X^{(i)} = x$. If $X^{(i)} = x$, $\frac{Y^{(i)}-A}{B-A} \sim \text{Ber}\left(\frac{e^\epsilon}{e^\epsilon+m-1}\right)$, where Ber$(p)$ denotes the Bernoulli distribution with probability p. On the other hand, if $X^{(i)} \neq x$, $\frac{Y^{(i)}-A}{B-A} \sim \text{Ber}\left(\frac{1}{m}\frac{e^\epsilon}{e^\epsilon+m-1} + \frac{m-1}{m}\frac{1}{e^\epsilon+m-1}\right)$. For convenience, we denote $p_1 = \frac{e^\epsilon}{e^\epsilon+m-1}$ and $p_2 = \frac{1}{m}\frac{e^\epsilon}{e^\epsilon+m-1} + \frac{m-1}{m}\frac{1}{e^\epsilon+m-1}$.

Subsequently, $\frac{n(\hat{f}(x)-A)}{B-A}$ is equivalent to the sum of two binomial random variables, where the first is sampled from BN$(nf(x), p_1)$ and the second is sampled from BN$(n(1 - f(x)), p_2)$, where BN(n, p) denotes the binomial distribution with n trials each having a success probability p.

When n is large enough, both binomial random variables can be approximated as Gaussian random variables, which are sampled from $\mathcal{N}(nf(x)p_1, np_1(1-p_1))$ and $\mathcal{N}(n(1 - f(x))p_2, np_2(1-p_2))$ respectively, where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. Since the sum of these two Gaussian random variables is also a Gaussian random variable, $\hat{f}(x)$ is approximated as

$$\begin{aligned} \frac{n(\hat{f}(x) - A)}{B - A} \sim \mathcal{N}(&nf(x)p_1 + n(1 - f(x))p_2, \\ &nf(x)p_1(1 - p_1) + n(1 - f(x))p_2(1 - p_2)). \end{aligned}$$

One can verify that the above equation is equivalent to

$$\frac{n(\hat{f}(x) - A)}{B - A} \sim \mathcal{N}(\frac{n(f(x) - A)}{B - A}, \frac{n^2 Var(\hat{f}(x))}{(B - A)^2}),$$

which can be refactored as

$$\frac{\hat{f}(x) - f(x)}{\sqrt{Var(\hat{f}(x))}} \sim \mathcal{N}(0, 1).$$

Thus, $\frac{\hat{f}(x)-f(x)}{\sqrt{Var(\hat{f}(x))}}$ of each experiment can be considered a standard Gaussian random variable. There are t rounds of the experiment; using the Laurent-Massart bound [2], we have

$$Pr\Bigg(\sum_t [(\frac{\hat{f}(x)_t - f(x)}{\sqrt{Var(\hat{f}(x))}})^2 - 1] \geq 2(\sqrt{t\alpha} + \alpha)\Bigg) \leq e^{-\alpha}.$$

Define $S(x) = \frac{1}{t}(\hat{f}(x)_t - f(x))^2$, and we have

$$Pr\Bigg(S(x) \geq [1 + \frac{2}{t}(\sqrt{t\alpha} + \alpha)]Var(\hat{f}(x))\Bigg) \leq e^{-\alpha}.$$

Given our assumption that $E[(\hat{f}(x) - f(x))^2] = Var(\hat{f}(x)) \leq V$ is identical for $\forall x \in \mathbf{x}$, we have

$$Pr\Bigg(S(x) \geq [1 + \frac{2}{t}(\sqrt{t\alpha} + \alpha)]V\Bigg) \leq e^{-\alpha}.$$

Note that $\forall x : S(x) < (1 + \frac{2}{t}(\sqrt{t\alpha} + \alpha))V$ is equivalent to $\max_x S(x) < (1+\frac{2}{t}(\sqrt{t\alpha}+\alpha))V$. Given that $\overline{MSE}(\hat{f}) = \max_x S(x)$, we have

$$Pr\big(\overline{MSE}(\hat{f}) < [1 + \frac{2}{t}(\sqrt{t\alpha} + \alpha)]V\big) > (1 - e^{-\alpha})^{|\mathbf{x}|}, \tag{S17}$$

which decreases with $|\mathbf{x}|$. Even when $|\mathbf{x}| \to \infty$, if $e^{-\alpha}|\mathbf{x}| = \frac{1}{20}$, we still have $(1 - e^{-\alpha})^{|\mathbf{x}|} \approx 0.95$. Thus, we have $\alpha = \log(20|\mathbf{x}|)$. Then, we obtain

$$Pr\big(\overline{MSE}(\hat{f}) < [1 + \frac{2}{t}\big(\sqrt{t\log(20|\mathbf{x}|)} + \log(20|\mathbf{x}|)\big)]V\big) > 0.95. \quad \square$$

References

[1] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 127–135.

[2] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Annals of statistics*, pp. 1302–1338, 2000.