Optimal Compression of Locally Differentially Private Mechanisms

Abhin Shah
Massachusetts Institute of Technology
abhin@mit.edu

Wei-Ning Chen
Stanford University
wnchen@stanford.edu

Johannes Balle
Google Research
jballe@google.com

Peter Kairouz
Google Research
kairouz@google.com

Lucas Theis
Google Research
theis@google.com

Abstract

Compressing the output of ε-locally differentially private (LDP) randomizers naively leads to suboptimal utility. In this work, we demonstrate the benefits of using schemes that jointly compress and privatize the data using shared randomness. In particular, we investigate a family of schemes based on Minimal Random Coding (Havasi et al., 2019) and prove that they offer optimal privacy-accuracy-communication tradeoffs. Our theoretical and empirical findings show that our approach can compress `PrivUnit`$_2$ (Bhowmick et al., 2018) and `Subset Selection` (Ye and Barg, 2018), the best known LDP algorithms for mean and frequency estimation, to the order of ε bits of communication while preserving their privacy and accuracy guarantees.

1 INTRODUCTION

Machine learning and data analytics are critical tools for designing better products and services. So far, these tools have been predominantly applied in datacenters on data that was curated from millions of users. However, centralized data collection and processing can expose individuals to privacy risks and organizations to legal risks if data is not properly managed. Indeed, increasing privacy concerns are fueling the demand for distributed learning and analytics systems that ensure that the underlying data remains private and secure. This is evident from the recent surge of interest in federated learning and analytics (e.g., Ramage and Mazzocchi, 2020; Kairouz et al., 2021).

Designing private and efficient distributed learning and analytics systems involves addressing three main challenges: (a) preserving the privacy of the user's local data, (b) communicating the privatized data efficiently to a central server, and (c) achieving high accuracy on a task (e.g., mean or frequency estimation). Privacy is often achieved by enforcing ε-local differential privacy (ε-LDP) (Warner, 1965; Evfimievski et al., 2003; Dwork et al., 2006; Kasiviswanathan et al., 2011), which guarantees that the outcome from a privatization mechanism will not release too much individual information statistically. Efficient communication, on the other hand, is achieved via compression and dimensionality reduction techniques (Suresh et al., 2017; Alistarh et al., 2017; Wen et al., 2017; Wang et al., 2018; Han et al., 2018a,b; Agarwal et al., 2018; Gandikota et al., 2019; Barnes et al., 2020; Chen et al., 2021).

Most existing works focus on addressing two of the three above-mentioned challenges, such as achieving good privacy-accuracy or good communication-accuracy tradeoffs separately. However, doing so can lead to suboptimal performance where all three desiderata are concerned. It is thus important to investigate the joint privacy-communication-accuracy tradeoffs when designing communication-efficient and private distributed algorithms. Under ε-LDP constraints,

Work done while A.S and W.C were interns at Google. J.B. provided the initial idea. P.K. and L.T. gave the conceptual and theoretical framework. A.S. and L.T. designed the algorithm. A.S., W.C., and L.T., devised the proofs. A.S. designed and performed the simulations with support from J.B.. A.S., W.C., P.K., and L.T. wrote the manuscript. J.B., P.K., and L.T., are listed alphabetically.

Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

Chen et al. (2020) presents minimax order-optimal mechanisms for frequency and mean estimation that require only ε bits (independent of the underlying dimensionality of the problem) by using shared randomness[1]. However, as noted by Feldman and Talwar (2021), the algorithms of Chen et al. (2020) are not competitive in terms of accuracy with the best known schemes – `Subset Selection` for frequency estimation (Ye and Barg, 2018) and $\texttt{PrivUnit}_2$ for mean estimation (Bhowmick et al., 2018). Motivated by this fact, the present work addresses the following fundamental question: *Can we attain the best known accuracy under ε-LDP while only using on the order of ε bits of communication*? We answer this question affirmatively by leveraging a technique based on importance sampling called Minimal Random Coding (Havasi et al., 2019; Cuff, 2008; Song et al., 2016).

1.1 Our Contributions

We first demonstrate that Minimal Random Coding (`MRC`) can compress any ε-LDP mechanism in a near-lossless fashion using only on the order of ε bits of communication (see Theorem 3.1). We also prove that the resulting compressed mechanism is 2ε-LDP (see Theorem 3.2). Thus, to achieve ε-LDP, one has to simulate an $\varepsilon/2$ mechanism and pay the corresponding penalty in accuracy. Similar to Chen et al. (2020), this approach can achieve the order optimal privacy-accuracy tradeoffs with about ε bits of communication but is not competitive with the best known LDP schemes. However, we show that this approach is optimal if one is willing to accept approximate LDP with a small δ (see Theorem 3.3).

To overcome the limitations of `MRC` in the pure LDP case, we present a modified version (`MMRC`) such that the resulting compressed mechanism is ε-LDP (see Theorem 3.4). We show that `MMRC` can simulate a large class of LDP mechanisms in a near-lossless fashion using only on the order of ε bits of communication (see Theorem 3.5 in conjunction with Theorem 3.1).

While the class of LDP mechanisms `MMRC` can simulate includes the best-known schemes for mean and frequency estimation, `MMRC` (similar to `MRC`) is biased for a fixed number of bits of communication. We show that `MMRC` simulating $\texttt{PrivUnit}_2$ and `Subset Selection` can be debiased (see Lemma 4.1 and Lemma 5.1), while preserving the corresponding accuracy guarantees (see Theorem 4.1 and Theorem 5.1).

Finally, we empirically demonstrate that `MMRC` achieves an accuracy comparable to $\texttt{PrivUnit}_2$ and `Subset Selection` (see Section 4.2 and Section 5.2)[2] while only using about ε bits.

We discuss interesting open problems in Section 6 and defer all additional results and experiments to the accompanying supplementary material.

1.2 Related Work

Recent works have examined approaches for compressing LDP schemes in the presence of shared randomness. When $\varepsilon \leq 1$, for frequency estimation, Bassily and Smith (2015) showed that a single bit is enough to simulate any LDP randomizer with (almost) no impact on its utility although with a large amount of shared randomness. Their result was improved upon, in terms of the amount of shared randomness required, by Bassily et al. (2017), Bun et al. (2019), and Acharya and Sun (2019).

Chen et al. (2020) generalized these methods to arbitrary ε's, and provided order-optimal schemes for both frequency and mean estimation that only use on the order of ε bits. However, their method is only order-optimal and cannot achieve the accuracy of the best known schemes: $\texttt{PrivUnit}_2$ (Bhowmick et al., 2018) for mean estimation and `Subset Selection` (Ye and Barg, 2018)[3] for frequency estimation. We show how one can achieve the accuracy of these schemes with on the order of ε bits of communication (when $\varepsilon \geq 1$). While we don't advocate large ε (our methods work for $\varepsilon = 1$ as well), we note that larger ε are both of theoretical and practical interest since *amplification via shuffling* can convert a local $\varepsilon > 1$ to a small central ε (Erlingsson et al., 2019; Balle et al., 2019; Erlingsson et al., 2020).

In the absence of shared randomness, Girgis et al. (2021b), Girgis et al. (2021a), Chen et al. (2020) provided order-optimal mechanisms for frequency and mean estimation but their mechanisms do not achieve the best known accuracy. Feldman and Talwar (2021) presented an approach for compressing ε-LDP schemes in a lossless fashion using a pseudorandom generator (PRG). Their approach, which relies on cryptographic hardness of the PRG, can compress `Subset Selection` to $O(\ln d)$ bits and $\texttt{PrivUnit}_2$ to $O(\varepsilon + \ln d)$ bits, where d is the dimension of the underlying problem. Their approach, similar to ours, can achieve the privacy vs accuracy tradeoffs of the best known schemes, i.e., `Subset Selection` and

[1] We assume that the encoder and the decoder can depend on a random quantity that both the server and the user have access to. See Section 2.2 for details.

[2] The source code of our implementation is available at `https://tinyurl.com/rcc-dp`.

[3] `Subset Selection` is similar to asymmetric RAPPOR (Erlingsson et al., 2014) in the sense that both have the same marginal distribution. Here, we focus on simulating `Subset Selection`.

$\texttt{PrivUnit}_2$. Nevertheless, their approach is designed to work without shared randomness, therefore requiring more bits than necessary if shared randomness is available, as in our work.

Unlike previous work, our technique of compressing generic LDP schemes relies on Minimal Random Coding (MRC), which was designed to simulate noisy channels. Several papers in information theory and related fields have studied the problem of efficiently simulating noisy channels over digital channels (e.g., Bennett and Shor, 2002; Harsha et al., 2007; Li and El Gamal, 2018) and proposed general solutions. In particular, these papers showed that any noisy channel can be simulated at a bit-rate which is close to the mutual information between the information available to the sender and the receiver. However, this result only holds if a shared source of randomness is available. Without such a source, the achievable rate has been shown to be close to Wyner's common information (Wyner, 1975; Cuff, 2008), which can be significantly larger than the mutual information (Xu et al., 2011). While promising as a recipe for simulating arbitrary differentially private mechanisms, the general coding schemes discussed in these papers have not been analyzed for their effect on differential privacy guarantees. MRC (Havasi et al., 2019), which we analyze and build upon here, is one of these schemes and is also known as likelihood encoder in information theory (Cuff, 2008; Song et al., 2016).

Finally, mean and frequency estimation under LDP constraints, two canonical problems in distributed learning and analytics, have been widely studied (Duchi et al., 2013; Nguyên et al., 2016; Bhowmick et al., 2018; Wang et al., 2019; Gandikota et al., 2019; Erlingsson et al., 2014; Bassily and Smith, 2015; Kairouz et al., 2016; Ye and Barg, 2018; Acharya et al., 2019).

2 PRELIMINARIES

2.1 Locally Differentially Private (LDP)

Suppose $\boldsymbol{x} \in \mathcal{X}$ is some user's data that must remain private. A privatization mechanism q is a randomized mapping that maps $\boldsymbol{x} \in \mathcal{X}$ to $\boldsymbol{z} \in \mathcal{Z}$ with probability $q(\boldsymbol{z}|\boldsymbol{x})$ where $\mathcal{Z}$ can be arbitrary. The user transmits $\boldsymbol{z} \sim q(\cdot|\boldsymbol{x})$, i.e., a privatized version of $\boldsymbol{x}$ to the server. Further, q is ε-LDP if

$$\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z},\ q(\boldsymbol{z}|\boldsymbol{x}) \leq e^{\varepsilon} q(\boldsymbol{z}|\boldsymbol{x}') \qquad (1)$$

and q is (ε, δ)-LDP if $\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, Z \subseteq \mathcal{Z}$,

$$\sum_{\boldsymbol{z} \in Z} q(\boldsymbol{z}|\boldsymbol{x}) \leq e^{\varepsilon} \sum_{\boldsymbol{z} \in Z} q(\boldsymbol{z}|\boldsymbol{x}') + \delta.$$

Here, we focus on ε-LDP mechanisms where $\varepsilon \geq 1$.

2.2 Shared Randomness

Here, we allow ε-LDP mechanisms to use *shared randomness*. That is, q can depend on a random variable $\boldsymbol{u} \in \mathcal{U}$ that is known to both the user and the server (but $\boldsymbol{u}$ is independent of $\boldsymbol{x}$). The corresponding ε-LDP constraint is

$$\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z}, \boldsymbol{u} \in \mathcal{U},\ q(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{u}) \leq \exp(\varepsilon) q(\boldsymbol{z}|\boldsymbol{x}', \boldsymbol{u}).$$

The server wishes to reconstruct $\boldsymbol{x}$ from $\boldsymbol{z}$ and the corresponding estimator is allowed to implicitly depend on $\boldsymbol{u}$. However, for simplicity, we suppress the dependence on $\boldsymbol{u}$ in our notation. In practice, shared randomness can be achieved via downlink communication, that is, the server generates $\boldsymbol{u}$ (e.g., a random seed) and communicates it to the user. Further, we note that such shared randomness can be established well before the advent of any private data[4].

2.3 $\texttt{PrivUnit}_2$

The $\texttt{PrivUnit}_2$ mechanism $q^{\texttt{pu}}$, proposed by Bhowmick et al. (2018), is an ε-LDP sampling scheme when the input alphabet $\mathcal{X}$ is the $d-$dimensional unit ℓ_2 sphere $\mathbb{S}^{d-1}$. Formally, given a vector $\boldsymbol{x} \in \mathbb{S}^{d-1}$, $\texttt{PrivUnit}_2$ draws a random vector $\boldsymbol{z}$ from a spherical cap $\{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x} \rangle \geq \gamma\}$ with probability p_0 or from its complement $\{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x} \rangle < \gamma\}$ with probability $1 - p_0$, where $\gamma \in [0, 1]$ and $p_0 \geq 1/2$ are parameters (depending on ε and d) that trade accuracy and privacy (see Appendix D). In other words, $q^{\texttt{pu}}$ is as follows:

$$q^{\texttt{pu}}(\boldsymbol{z}|\boldsymbol{x}) = \begin{cases} \dfrac{2p_0}{A(1,d) I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2})} & \text{if } \langle \mathbf{x}, \mathbf{z} \rangle \geq \gamma \\ \dfrac{2(1-p_0)}{2A(1,d) - A(1,d) I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2})} & \text{otherwise} \end{cases}$$

where $A(1,d)$ denotes the area of $\mathbb{S}^{d-1}$ and $I_x(a,b)$ denotes the regularized incomplete beta function. The estimator of the $\texttt{PrivUnit}_2$ mechanism (denoted by $\hat{\boldsymbol{x}}^{\texttt{pu}}$) is obtained by dividing every coordinate of $\boldsymbol{z}$ by $m_{\texttt{pu}}$ i.e., $\hat{\boldsymbol{x}}^{\texttt{pu}} := \boldsymbol{z}/m_{\texttt{pu}}$ where

$$m_{\texttt{pu}} := \tfrac{(1-\gamma^2)^{\alpha}}{2^{d-2}(d-1)} \left[\tfrac{p_0}{B(1;\alpha,\alpha) - B(\tau;\alpha,\alpha)} - \tfrac{1-p_0}{B(\tau;\alpha,\alpha)} \right] \quad (2)$$

with $\alpha = (d-1)/2$, $\tau = (1+\gamma)/2$, and $B(x;\alpha,\beta)$ denoting the incomplete beta function. The estimator $\hat{\boldsymbol{x}}^{\texttt{pu}}$ is (a) unbiased i.e., $\mathbb{E}[\hat{\boldsymbol{x}}^{\texttt{pu}}|\boldsymbol{x}] = \boldsymbol{x}$, (b) has order-optimal utility i.e., $\mathbb{E}[\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\|_2^2] = \Theta\left(\frac{d}{\min(\varepsilon, (e^{\varepsilon}-1)^2, d)}\right)$, and (c) achieves the best known constants for mean estimation. See Appendix D for more details on $\texttt{PrivUnit}_2$.

[4] Quantifying the amount of such shared randomness required remains an open question. See Section 6.

2.4 Subset Selection

The `Subset Selection` mechanism $q^{\texttt{ss}}$, proposed by Ye and Barg (2018), is an ε-LDP sampling scheme when the input alphabet $\mathcal{X}$ can take d different values. Without loss of generality, let $\mathcal{X} := \{e_1, e_2, ..., e_d\}$, where $e_j \in \{0,1\}^d$ is the j^{th} standard unit vector, i.e., the one-hot encoding of j. The output alphabet $\mathcal{Z}$ is the set of all d-bit binary strings with Hamming weight $s := \lceil \frac{d}{1+e^\varepsilon} \rceil$, i.e.,

$$\mathcal{Z} = \left\{ \boldsymbol{z} = (z^{(1)}, \cdots, z^{(d)}) \in \{0,1\}^d : \textstyle\sum_{i=1}^d z^{(i)} = s \right\}.$$

Given $\boldsymbol{x} \in \mathcal{X}$, `Subset Selection` maps it to $\boldsymbol{z} \in \mathcal{Z}$ with the following conditional probability:

$$q^{\texttt{ss}}(\boldsymbol{z}|\boldsymbol{x}) := \begin{cases} \frac{e^\varepsilon}{\binom{d-1}{s-1}e^\varepsilon + \binom{d-1}{s}} & \text{if } \boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{x}} \\ \frac{1}{\binom{d-1}{s-1}e^\varepsilon + \binom{d-1}{s}} & \text{if } \boldsymbol{z} \in \mathcal{Z} \setminus \mathcal{Z}_{\boldsymbol{x}} \end{cases} \tag{3}$$

where $\mathcal{Z}_{\boldsymbol{x}} = \left\{ \boldsymbol{z} = (z^{(1)}, \cdots, z^{(d)}) \in \mathcal{Z} : z^{(x)} = 1 \right\}$ is the set of elements in $\mathcal{Z}$ with 1 in the x^{th} location. The estimator of the `Subset Selection` mechanism (denoted by $\hat{\boldsymbol{x}}^{\texttt{ss}}$) is obtained by subtracting $b_{\texttt{ss}}$ from every component of $\boldsymbol{z}$ and dividing every component of the result by $m_{\texttt{ss}}$ i.e., $\hat{\boldsymbol{x}}^{\texttt{ss}} := (\boldsymbol{z} - b_{\texttt{ss}})/m_{\texttt{ss}}$ where

$$m_{\texttt{ss}} := \tfrac{s(d-s)(e^\varepsilon-1)}{(d-1)(s(e^\varepsilon-1)+d)}, \; b_{\texttt{ss}} := \tfrac{s((s-1)e^\varepsilon+(d-s))}{(d-1)(s(e^\varepsilon-1)+d)}. \tag{4}$$

The estimator $\hat{\boldsymbol{x}}^{\texttt{ss}}$ is (a) unbiased i.e., $\mathbb{E}[\hat{\boldsymbol{x}}^{\texttt{ss}}|\boldsymbol{x}] = \boldsymbol{x}$, (b) has optimal utility i.e., $\mathbb{E}[\|\hat{\boldsymbol{x}}^{\texttt{ss}} - \boldsymbol{x}\|_2^2] = \Theta\left(\frac{d}{\min\left(e^\varepsilon, (e^\varepsilon-1)^2, d\right)}\right)$ and (c) achieves the best known constants for frequency estimation. See Appendix G for more details on `Subset Selection`.

3 MAIN RESULTS

In this section, first, we describe the Minimal Random Coding algorithm for compressing any ε-LDP mechanism and prove its order-optimal privacy-accuracy-communication tradeoffs. Then, we propose the Modified Minimal Random Coding algorithm for compressing any ε-LDP *cap-based mechanism*[5] and prove that it achieves optimal privacy-accuracy-communication tradeoffs.

3.1 Minimal Random Coding (`MRC`)

Consider an ε-LDP mechanism $q(\cdot|\boldsymbol{x})$ that we wish to compress. Under `MRC`, first, a number of candidates $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ are drawn from a fixed reference distribution $p(\cdot)$ (known to both the user and the server). This can be achieved via a pseudorandom number generator with a known seed. Next, the user transmits an index $K \in [N]$ to the server where K is drawn according to some distribution $\pi^{\texttt{mrc}}(\cdot)$ such that $\boldsymbol{z}_K \sim q(\cdot|\boldsymbol{x})$ approximately. The distribution $\pi^{\texttt{mrc}}$ is such that, $\forall k \in [N]$, $\pi^{\texttt{mrc}}(k) \propto w(k)$ where $w(k) := q(\boldsymbol{z}_k|\boldsymbol{x})/p(\boldsymbol{z}_k)$ are the importance weights[6] (see Algorithm 1). To communicate the index K of `MRC`, $\log N$ bits are required.

Algorithm 1: `MRC`

```
Input: ε-LDP mechanism $q(\cdot|\boldsymbol{x})$, reference
        distribution $p(\cdot)$, number of candidates $N$
Draw samples $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ from $p(\boldsymbol{z})$ // Using the
  shared source of randomness
for $k \in \{1, \cdots, N\}$ do
    $w(k) \leftarrow q(\boldsymbol{z}_k|\boldsymbol{x})/p(\boldsymbol{z}_k)$
$\pi^{\texttt{mrc}}(\cdot) \leftarrow w(\cdot)/\sum_k w(k)$
Output: $\pi^{\texttt{mrc}}(\cdot), \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N\}$
```

Let $q^{\texttt{mrc}}$ denote the distribution of $\boldsymbol{z}_K$ where $K \sim \pi^{\texttt{mrc}}(\cdot)$. The following theorem shows that when the number of candidates is exponential in ε, samples drawn from $q^{\texttt{mrc}}$ will be similar to samples drawn from $q(\cdot|\boldsymbol{x})$ in terms of ℓ_2 error. In other words, $q^{\texttt{mrc}}$ can compress $q(\cdot|\boldsymbol{x})$ to the order of ε bits of communication as well as simulate it in a near-lossless fashion. A proof can be found in Appendix B.1.

Theorem 3.1 (Utility of `MRC`)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP mechanism $q(\cdot|\boldsymbol{x})$. Consider any reference distribution $p(\cdot)$ such that $|\ln(q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z}))| \leq \varepsilon \; \forall \; \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z}$.*[7] *Let the number of candidates be $N = 2^{(\log e + 4c)\varepsilon}$ for some constant $c \geq 0$. Then, for $\alpha \in [0, 1/2]$, q^{mrc} is such that*

$$\left| \mathbb{E}_{q^{mrc}}\left[\|\boldsymbol{z} - \boldsymbol{x}\|^2\right] - \mathbb{E}_q\left[\|\boldsymbol{z} - \boldsymbol{x}\|^2\right] \right| \leq \frac{2\alpha\sqrt{\mathbb{E}_q[\|\boldsymbol{z}-\boldsymbol{x}\|^4]}}{1-\alpha} \tag{5}$$

holds with probability at least $1 - 2\alpha$, with c and α related by the following: $\alpha = \sqrt{2^{-c\varepsilon} + 2^{-c^2/\log e+1}}$.

In general, $\mathbb{E}_q\left[\|\boldsymbol{z} - \boldsymbol{x}\|^4\right]$ in (5) can be well-controlled. See Remark B.1 in Appendix B.1 for more details.

In the next Theorem, we show that $\pi^{\texttt{mrc}}$ is 2ε-LDP. Hence, the compressed mechanism $q^{\texttt{mrc}}$ is 2ε-LDP.

Theorem 3.2 (Pure DP guarantee of `MRC`)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, and data $\boldsymbol{x} \in \mathcal{X}$. Consider any ε-LDP mechanism $q(\cdot|\boldsymbol{x})$, reference distribution $p(\cdot)$, and number of candidates $N \geq 1$. Then, $\pi^{\texttt{mrc}}(\cdot)$ obtained from Algorithm 1 is a 2ε-LDP mechanism.*

[5] The family of cap-based mechanisms includes $\texttt{PrivUnit}_2$ and `Subset Selection`. See Definition 3.1.

[6] We suppress dependence of $\pi^{\texttt{mrc}}$ & w on $\boldsymbol{x}$ for simplicity.

[7] Note that this condition holds for many reference distributions $p(\cdot)$. For example, one can simply choose $p(\cdot) = q(\cdot|\boldsymbol{x}^*)$ for some $\boldsymbol{x}^* \in \mathcal{X}$.

A proof is provided in Appendix B.2.1 and it relies on fact that the following ratio can be bounded by $e^{2\varepsilon}$:

$$\frac{\pi^{\mathtt{mrc}}_{\boldsymbol{x}}(k)}{\pi^{\mathtt{mrc}}_{\boldsymbol{x}'}(k)} = \frac{q(\boldsymbol{z}_k|\boldsymbol{x})}{q(\boldsymbol{z}_k|\boldsymbol{x}')} \cdot \frac{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x}')/p(\boldsymbol{z}_{k'})}{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})}.$$

In the following Theorem, we show that $\pi^{\mathtt{mrc}}$ is $(\varepsilon + \varepsilon_0, \delta)$-LDP implying that the compressed mechanism $q^{\mathtt{mrc}}$ is $(\varepsilon + \varepsilon_0, \delta)$-LDP where $\varepsilon_0 > 0$ and $\delta \leq 1$ are free parameters. This Theorem can be viewed complementary to Theorem 3.2 where a stronger privacy parameter can be achieved (i.e., $\varepsilon + \varepsilon_0$ which can get arbitrarily close to ε as opposed to 2ε) albeit at the cost of trading pure privacy for approximate privacy. A proof is provided in Appendix B.2.2.

Theorem 3.3 (Approximate DP guarantee of MRC)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP mechanism $q(\cdot|\boldsymbol{x})$. Consider any reference distribution $p(\cdot)$ such that $|\ln(q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z}))| \leq \varepsilon\ \forall\ \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z}$.[5] Let $c_0 \geq 0$ be some constant and let the number of candidates $N = \exp(2\varepsilon + 2c_0)$. Then, for any $\delta \leq 1$, $\pi^{\mathtt{mrc}}(\cdot)$ obtained from Algorithm 1 is $(\varepsilon + \varepsilon_0, \delta)$-LDP mechanism where*

$$\varepsilon_0 := \ln \frac{1 + a_0}{1 - a_0} \qquad \text{and} \qquad a_0 := \exp(-c_0)\sqrt{\tfrac{1}{2} \ln \tfrac{2}{\delta}}.$$

3.2 Modified Minimal Random Coding (MMRC)

While the results regarding MRC in Section 3.1 are general and offer order optimal privacy-accuracy tradeoffs with about ε bits of communication, the resulting compressed mechanism is not exactly ε-LDP. More specifically, Theorem 3.2 introduces an additional factor of 2 in the LDP guarantee and Theorem 3.3 provides an approximate privacy guarantee instead of a pure privacy guarantee. To address these limitations, we focus on a class of ε-LDP mechanisms which we call *cap-based* mechanisms and propose a modification to MRC such that the resulting compressed mechanism is ε-LDP. Further, like MRC, MMRC can simulate the underlying ε-LDP mechanism in a near-lossless fashion while using only on the order of ε bits.

We start with the definition of cap-based mechanism which is inspired from the structure of $\mathtt{PrivUnit}_2$ and Subset Selection.

Definition 3.1 (Cap-based Mechanisms)**.** *An ε-LDP mechanism $q(\boldsymbol{z}|\boldsymbol{x})$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Z}$ is a cap-based mechanism if it can be written in the following way:*

$$q(\boldsymbol{z}|\boldsymbol{x}) = \begin{cases} c_1(\varepsilon, d) & \text{if } \boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}} \\ c_2(\varepsilon, d) & \text{if } \boldsymbol{z} \notin \mathsf{Cap}_{\boldsymbol{x}} \end{cases} \tag{6}$$

where (a) $c_1(\varepsilon, d)$ and $c_2(\varepsilon, d)$ are constants with respect to $\boldsymbol{x}$ and $\boldsymbol{z}$ such that $c_1(\varepsilon, d) \geq c_2(\varepsilon, d)$, and (b) $\mathsf{Cap}_{\boldsymbol{x}} \subseteq \mathcal{Z}$ such that $\mathbb{P}_{\boldsymbol{z}\sim\mathrm{Unif}(\mathcal{Z})}(\boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}})$ is independent of $\boldsymbol{x}$ and is at least $c_2(\varepsilon, d)/2c_1(\varepsilon, d)$.

In words, a cap-based ε-LDP mechanism samples uniformly either from $\mathsf{Cap}_{\boldsymbol{x}}$ or from $\mathcal{Z} \setminus \mathsf{Cap}_{\boldsymbol{x}}$ where $\mathsf{Cap}_{\boldsymbol{x}} \subseteq \mathcal{Z}$ is such that if $\boldsymbol{z}$ is sampled uniformly from $\mathcal{Z}$, it will belong to $\mathsf{Cap}_{\boldsymbol{x}}$ with probability at least $c_2(\varepsilon, d)/2c_1(\varepsilon, d)$. It is easy to see that $q^{\mathtt{ss}}$ defined in (3) is a cap-based mechanism with $\mathsf{Cap}_{\boldsymbol{x}} = \mathcal{Z}_{\boldsymbol{x}}$, $c_1(\varepsilon, d) = \frac{e^{\varepsilon}}{\binom{d-1}{s-1}e^{\varepsilon} + \binom{d-1}{s}}$, and $c_2(\varepsilon, d) = \frac{1}{\binom{d-1}{s-1}e^{\varepsilon} + \binom{d-1}{s}}$. See Appendix G where we evaluate $\mathbb{P}_{\boldsymbol{z}\sim\mathrm{Unif}(\mathcal{Z})}(\boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{x}})$ and show that it is at least $1/2e^{\varepsilon}$. In Appendix D, we show that $q^{\mathtt{pu}}$ is a cap-based mechanism.

For a cap-based ε-LDP mechanism $q(\boldsymbol{z}|\boldsymbol{x})$ and a uniform reference distribution $p(\cdot)$, the distribution $\pi^{\mathtt{mrc}}$ obtained from Algorithm 1 takes a special form:

$$\pi^{\mathtt{mrc}}(k) = \begin{cases} \frac{1}{N} \times \frac{c_1(\varepsilon,d)}{\theta c_1(\varepsilon,d) + (1-\theta)c_2(\varepsilon,d)} & \text{if } \boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}} \\ \frac{1}{N} \times \frac{c_2(\varepsilon,d)}{\theta c_1(\varepsilon,d) + (1-\theta)c_2(\varepsilon,d)} & \text{if } \boldsymbol{z}_k \notin \mathsf{Cap}_{\boldsymbol{x}} \end{cases} \tag{7}$$

where θ is the fraction of candidates inside the $\mathsf{Cap}_{\boldsymbol{x}}$, i.e., $\theta = \frac{1}{N}\sum_k \mathbb{1}(\boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}})$. As is, $\pi^{\mathtt{mrc}}$ in (7) is not necessarily ε-LDP because θ can be different for $\boldsymbol{x}$ and $\boldsymbol{x}'$. However, as $N \to \infty$, $\theta \to \mathbb{E}[\theta] = \mathbb{P}_{\boldsymbol{z}\sim\mathrm{Unif}(\mathcal{Z})}(\boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}})$, which is not a function of $\boldsymbol{x}$, implying that $\pi^{\mathtt{mrc}}_{\boldsymbol{x}}(k)/\pi^{\mathtt{mrc}}_{\boldsymbol{x}'}(k) \leq c_1(\varepsilon, d)/c_2(\varepsilon, d) \leq \exp(\varepsilon)$[8]. This shows that $\pi^{\mathtt{mrc}}$ is ε-LDP when $N \to \infty$. This motivates us to modify $\pi^{\mathtt{mrc}}$ to $\pi^{\mathtt{mmrc}}$ such that $\pi^{\mathtt{mmrc}}$ is ε-LDP irrespective of N. Further, when N is large enough, the modification is not by much, i.e., a sample from $\pi^{\mathtt{mrc}}$ is similar to a sample from $\pi^{\mathtt{mmrc}}$.

To that end, define an upper threshold $t_u = \frac{1}{N} \times \frac{c_1(\varepsilon,d)}{\mathbb{E}[\theta]c_1(\varepsilon,d) + (1-\mathbb{E}[\theta])c_2(\varepsilon,d)}$ and a lower threshold $t_l = \frac{1}{N} \times \frac{c_2(\varepsilon,d)}{\mathbb{E}[\theta]c_1(\varepsilon,d) + (1-\mathbb{E}[\theta])c_2(\varepsilon,d)}$, and initialize $\pi^{\mathtt{mmrc}}$ to be equal to $\pi^{\mathtt{mrc}}$. We want to modify $\pi^{\mathtt{mmrc}}$ so as to ensure:

$$t_l \leq \pi^{\mathtt{mmrc}}(k) \leq t_u\ \forall k \in [N], \tag{8}$$

which, as argued above, guarantees ε-LDP irrespective of the choice of N. First, it is easy to see that $\theta c_1(\varepsilon, d) + (1 - \theta)c_2(\varepsilon, d)$ is an increasing function of θ. Next, we will look at 3 cases depending on the relationship between θ and $\mathbb{E}[\theta]$: (A) If $\theta = \mathbb{E}[\theta]$, then $\pi^{\mathtt{mmrc}}$ already satisfies (8); (B) If $\theta < \mathbb{E}[\theta]$, then only the upper threshold is violated and we set $\pi^{\mathtt{mmrc}}(k) = t_u\ \forall k : \boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}}$ and re-normalize the remaining $\pi^{\mathtt{mmrc}}(k)$; (C) If $\theta > \mathbb{E}[\theta]$, then only the lower threshold is violated, we set $\pi^{\mathtt{mmrc}}(k) = t_l\ \forall k : \boldsymbol{z}_k \notin \mathsf{Cap}_{\boldsymbol{x}}$ and re-normalize the remaining $\pi^{\mathtt{mmrc}}(k)$. The re-normalization step does not violate (8). We provide pseudo-code to calculate $\pi^{\mathtt{mmrc}}$ in Algorithm 2.

[8] This follows from (1) and (6) because $q(\cdot|\boldsymbol{x})$ is ε-LDP.

Algorithm 2: `MMRC`

```
Input: $\varepsilon$-LDP cap-based mechanism $q(\cdot|\boldsymbol{x})$, the
       associated $\mathsf{Cap}_{\boldsymbol{x}}$, reference distribution
       $p(\cdot)$, number of candidates $N$, lower
       threshold $t_l$, upper threshold $t_u$
$\pi^{\mathtt{mrc}}(\cdot), \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N\} \leftarrow \mathtt{MRC}(p(\cdot), q(\cdot|\boldsymbol{x}), N)$
$\theta \leftarrow \frac{1}{N}\sum_k \mathbb{1}(\boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}})$ // Compute the fraction
  of candidates inside the cap
Initialization: $\pi^{\mathtt{mmrc}}(\cdot) \leftarrow \pi^{\mathtt{mmrc}}(\cdot)$
if $\max_k \pi^{mmrc}(k) > t_u$ then
    // Upper threshold is violated
    $\pi^{\mathtt{mmrc}}(k) \leftarrow t_u, \forall\, k : \boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}}$
    $\pi^{\mathtt{mmrc}}(k) \leftarrow \frac{1-N\theta t_u}{N(1-\theta)}, \forall\, k : \boldsymbol{z}_k \notin \mathsf{Cap}_{\boldsymbol{x}}$
else if $\min_k \pi^{mmrc}(k) < t_l$ then
    // Lower threshold is violated
    $\pi^{\mathtt{mmrc}}(k) \leftarrow t_l, \forall\, k : \boldsymbol{z}_k \notin \mathsf{Cap}_{\boldsymbol{x}}$
    $\pi^{\mathtt{mmrc}}(k) \leftarrow \frac{1-N(1-\theta)t_l}{N\theta}, \forall\, k : \boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}}$
Output: $\pi^{\mathtt{mmrc}}(\cdot), \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N\}$
```

Let $q^{\mathtt{mmrc}}$ denote the distribution of $\boldsymbol{z}_K$ where $K \sim \pi^{\mathtt{mmrc}}(\cdot)$. In the following Theorem, we show that $\pi^{\mathtt{mmrc}}$ is ε-LDP implying that the compressed mechanism $q^{\mathtt{mmrc}}$ is ε-LDP. The proof follows from (8) and can be found in Appendix C.1.

Theorem 3.4 (DP guarantee of `MMRC`)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP cap-based mechanism $q(\cdot|\boldsymbol{x})$. Let the reference distribution $p(\cdot)$ be the uniform distribution on $\mathcal{Z}$. Consider any number of candidates $N \geq 1$. Then, $\pi^{mmrc}(\cdot)$ obtained from Algorithm 2 is an ε-LDP mechanism.*

The following Theorem shows that, with number of candidates exponential in ε, samples drawn from $q^{\mathtt{mmrc}}$ will be similar to the samples drawn from $q^{\mathtt{mrc}}$ in terms of ℓ_2 error. A proof can be found in Appendix C.3.

Theorem 3.5 (Utility of `MMRC`)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP cap-based mechanism $q(\cdot|\boldsymbol{x})$. Let the reference distribution $p(\cdot)$ be the uniform distribution on $\mathcal{Z}$. Let N denote the number of candidates. Then, q^{mmrc} is such that*

$$\mathbb{E}_{q^{mmrc}}\big[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\big] \leq \mathbb{E}_{q^{mrc}}\big[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\big] + \sqrt{\frac{\rho(1+\varepsilon)}{2}}\max_{\boldsymbol{x},\boldsymbol{z}}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2 \qquad (9)$$

where $\rho \in (0,1)$ is such that $N = \frac{2(\exp(\varepsilon)-1)^2}{\rho^2}\ln\frac{2}{\rho}$.

For bounded mechanisms, $\max_{\boldsymbol{x},\boldsymbol{z}}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2$ in (9) can be well-controlled. See Remark C.1 in Appendix C.3 for a discussion.

In conjunction with Theorem 3.1, Theorem 3.5 implies that $q^{\mathtt{mmrc}}$ can compress $q(\cdot|\boldsymbol{x})$ to the order of ε bits of communication and simulate it in a near-lossless fashion. This is stated formally and proved in Appendix C.3.

4 MEAN ESTIMATION

In this section, we focus on the mean estimation problem, which is a canonical statistical task in distributed estimation with applications in distributed stochastic gradient descent, federated learning, etc. Let the input space $\mathcal{X}$ be the d-dimensional unit ℓ_2 sphere, i.e., $\mathcal{X} = \mathbb{S}^{d-1}$. Consider n users where user i has some data $\boldsymbol{x}_i \in \mathcal{X}$. For every $i \in [n]$, let $\boldsymbol{x}_i$ be privatized using an ε-LDP mechanism $q(\cdot|\boldsymbol{x}_i)$ and potentially post-processed to obtain an estimate $\hat{\boldsymbol{x}}_i$ of $\boldsymbol{x}_i$. We are interested in estimating the *empirical mean* $\boldsymbol{\mu} \triangleq \frac{1}{n}\sum_i \boldsymbol{x}_i$ using $\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_n$ such that the mean estimation error defined below is minimized

$$r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}, q\right) \triangleq \max_{\boldsymbol{x}^n \in \mathcal{X}^n} \mathbb{E}\left[\|\hat{\boldsymbol{\mu}}\left(\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_n\right) - \boldsymbol{\mu}\|_2^2\right], \quad (10)$$

where $\hat{\boldsymbol{\mu}}$ is an estimate of $\boldsymbol{\mu}$ and the expectation is with respect to $q(\cdot|\boldsymbol{x}_i)$ as well as all (possibly shared) randomness used by $q(\cdot|\boldsymbol{x}_i)$ $\forall i \in [n]$.

Bhowmick et al. (2018) show that $\mathtt{PrivUnit}_2$ achieves the order-optimal privacy-accuracy trade-off for mean estimation, i.e., $r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{\mathtt{pu}}, q^{\mathtt{pu}}\right) = \Theta\left(\frac{d}{\min(\varepsilon,(e^{\varepsilon}-1)^2,d)}\right)$ where $\hat{\boldsymbol{\mu}}^{\mathtt{pu}} := \frac{1}{n}\sum_i \hat{\boldsymbol{x}}_i^{\mathtt{pu}}$. Moreover, compared to other (order-optimal) ε-LDP mean estimation mechanisms, $\mathtt{PrivUnit}_2$ admits the best constants and gives the smallest ℓ_2 error in practice (see Feldman and Talwar (2021)). However, $\mathtt{PrivUnit}_2$ requires each user to send a d-dimensional real vector, so without any compression, the communication needed is $\Theta(d)$ bits, which can be an issue in many practical scenarios.

To compress and simulate $\mathtt{PrivUnit}_2$, one can directly apply the generic `MMRC` mechanism defined in Section 3.2. However, for a fixed number of candidates N, `MMRC` yields a biased estimate of $\boldsymbol{x}$ and hence cannot get the correct (optimal) order of estimation error in (10), i.e., the error would not decay with n[9]. Fortunately, we show (in Section 4.1) that the bias can be corrected by appropriately scaling the privatized version of $\boldsymbol{x}$, i.e., by using an estimator which is slightly different compared to the original estimator of $\mathtt{PrivUnit}_2$. Further, we also show (in Section 4.2) that the resulting unbiased estimator for mean estimation ($\hat{\boldsymbol{\mu}}^{\mathtt{mmrc}}$) can simulate $\mathtt{PrivUnit}_2$ closely while only using on the order of ε bits of communication.

[9] We note that this does not undermine the significance of Theorem 3.1 and Theorem 3.5. These are useful in single-user settings (i.e., $n = 1$) and are generic as they can compress (near-losslessly) any ε-LDP and ε-LDP cap-based mechanism, respectively.

4.1 Debiasing `MMRC` to simulate `PrivUnit`$_2$

Let us focus on a single user and consider some data $\boldsymbol{x} \in \mathcal{X}$. Recall the `PrivUnit`$_2$ ε-LDP mechanism $q^{\mathtt{pu}}$ described in Section 2 with parameters p_0 and γ. `PrivUnit`$_2$ is a cap-based mechanism with $\mathsf{Cap}_{\boldsymbol{x}} = \{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x} \rangle \geq \gamma\}$ (see Appendix D for details). Let $\pi^{\mathtt{mmrc}}$ be the distribution and $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_N$ be the candidates obtained from Algorithm 2 when the reference distribution is $\mathrm{Unif}(\mathbb{S}^{d-1})$. Let $K \sim \pi^{\mathtt{mmrc}}(\cdot)$. Therefore, $\boldsymbol{z}_K$ is the privatized version of $\boldsymbol{x}$ using `MMRC`.

Define $p_{\mathtt{mmrc}} := \mathbb{P}(\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}})$ to be the probability with which the sampled candidate $\boldsymbol{z}_K$ belongs to the spherical cap associated with `PrivUnit`$_2$. Define $m_{\mathtt{mmrc}}$ as the scaling factor in (2) when p_0 in (2) is replaced by $p_{\mathtt{mmrc}}$. Define $\hat{\boldsymbol{x}}^{\mathtt{mmrc}} := \boldsymbol{z}_K / m_{\mathtt{mmrc}}$ as the estimator of the `MMRC` mechanism simulating `PrivUnit`$_2$. The following Lemma shows that $\hat{\boldsymbol{x}}^{\mathtt{mmrc}}$ is an unbiased estimator. See Appendix F.1 for a proof.

Lemma 4.1. *Let $\hat{\boldsymbol{x}}^{mmrc}$ be the estimator of the* `MMRC` *mechanism simulating* `PrivUnit`$_2$ *as defined above. Then, $\mathbb{E}_{q^{mmrc}}[\hat{\boldsymbol{x}}^{mmrc}] = \boldsymbol{x}$.*

4.2 Simulating `PrivUnit`$_2$ using `MMRC`

Finally, we consider estimating the empirical mean $\boldsymbol{\mu}$ defined earlier using the `MMRC` scheme simulating `PrivUnit`$_2$. To that end, consider n users and let $\hat{\boldsymbol{x}}_i^{\mathtt{mmrc}}$ be the unbiased estimator of $\boldsymbol{x}_i$ at the i^{th} user. Let the (unbiased) estimate of $\boldsymbol{\mu}$ be $\hat{\boldsymbol{\mu}}^{\mathtt{mmrc}} := \frac{1}{n} \sum_i \hat{\boldsymbol{x}}_i^{\mathtt{mmrc}}$.

The following Theorem shows that, for mean estimation, `MMRC` can simulate `PrivUnit`$_2$ in a near-lossless manner (when n is large and λ is small) while only using on the order of ε bits of communication. A proof can be found in Appendix F.2. The key idea in the proof is to show that when the number of candidates N is exponential in ε, the scaling factor $m_{\mathtt{mmrc}}$ is close to the scaling parameter associated with `PrivUnit`$_2$ (i.e., $m_{\mathtt{pu}}$ defined in (2)).

Theorem 4.1. *Let $r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right)$ and $r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{mmrc}, q^{mmrc}\right)$ be the empirical mean estimation error for* `PrivUnit`$_2$ *with parameter p_0 and* `MMRC` *simulating* `PrivUnit`$_2$ *with N candidates respectively. Consider any $\lambda > 0$. Then,*

$$r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{mmrc}, q^{mmrc}\right) \leq (1+\lambda)^2\, r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right) + 2(1+\lambda)(2+\lambda)\sqrt{\frac{r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right)}{n}} + \frac{(2+\lambda)^2}{n}.$$

as long as

$$N \geq \frac{e^{2\varepsilon}}{2}\left(\frac{2(1+\lambda)}{\lambda\,(p_0 - 1/2)}\right)^2 \ln\left(\frac{4(1+\lambda)}{\lambda\,(p_0 - 1/2)}\right). \quad (11)$$

We note that while a specific value of λ can be chosen (say 0.1 or smaller) in (11), in practice, the number of bits could be fixed (see Section 4.3), determining the value of λ.

4.3 Empirical Comparisons

Next, we empirically demonstrate the privacy-accuracy-communication tradeoffs of `MMRC` simulating `PrivUnit`$_2$. Along with `PrivUnit`$_2$, we compare against the SQKR algorithm of Chen et al. (2020) which offers order-optimal privacy-accuracy tradeoffs while requiring only ε bits. Following Chen et al. (2020), we generate data independently but non-identically to capture the distribution-free setting as well as ensure that the data non-central, i.e. $\boldsymbol{\mu} \neq 0$. More specifically, we set $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n/2} \overset{\text{i.i.d.}}{\sim} N(1,1)^{\otimes d}$ and $\boldsymbol{x}_{n/2+1}, \cdots, \boldsymbol{x}_n \overset{\text{i.i.d.}}{\sim} N(10,1)^{\otimes d}$. Further, to ensure that each data lies on $\mathbb{S}^{d-1}$, we normalize each $\boldsymbol{x}_i$ by setting $\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i / \left\|\boldsymbol{x}_i\right\|_2$. We report the average ℓ_2 estimation error over 10 runs. See more variations in Appendix F.3.

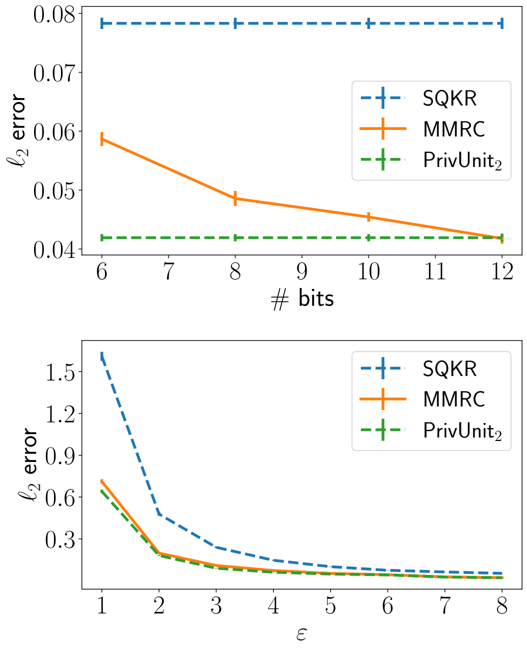

Figure 1: Comparing `PrivUnit`$_2$, `MMRC` simulating `PrivUnit`$_2$ and SQKR for mean estimation with $d = 500$ and $n = 5000$. **Top:** ℓ_2 error vs #bits for $\varepsilon = 6$. **Bottom:** ℓ_2 error vs ε for #bits $= \max\{\lceil(\varepsilon/\ln 2)\rceil + 2, 8\}$. SQKR uses #-bits $= \varepsilon$ for both as it leads to a poor performance if #-bits $> \varepsilon$.

In Figure 1 (top), we show the communication-accuracy tradeoffs. We see that with correct order of bits, the accuracy of `MMRC` simulating `PrivUnit`$_2$ converges to the accuracy of the uncompressed `PrivUnit`$_2$. In Figure 1 (bottom), we show the privacy-accuracy tradeoffs. We see that `MMRC` simulating `PrivUnit`$_2$ can attain accuracy of the uncompressed $\texttt{PrivUnit}_2$ for the range of ε's typically considered by LDP mechanisms while only using $\max\{\lceil(\varepsilon/\ln 2)\rceil + 2, 8\}$ bits.

5 FREQUENCY ESTIMATION

In this section, we study the frequency estimation problem, which is another canonical statistical task in distributed distribution estimation, with application to federated analytics (Ramage and Mazzocchi, 2020).

Let $\mathcal{X}$ be a set of d distinct symbols and without loss of generality $\mathcal{X} := \{e_1, e_2, ..., e_d\}$, where $e_j \in \{0,1\}^d$ is the j^{th} standard unit vector i.e., e_j is the one-hot encoding of j. Consider n users where user i has some data $\boldsymbol{x}_i \in \mathcal{X}$. For every $i \in [n]$, let $\boldsymbol{x}_i$ be privatized using an ε-LDP mechanism $q(\cdot|\boldsymbol{x}_i)$ and potentially post-processed to obtain an estimate $\hat{\boldsymbol{x}}_i$ of $\boldsymbol{x}_i$. We are interested in estimating the *empirical distribution* of $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ defined as $\Pi \triangleq \frac{1}{n}\sum_i \boldsymbol{x}_i$ using $\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_n$ such that the estimation error defined below is minimized:

$$r_{\textsf{FE}}\left(\hat{\Pi}, q, \ell\right) \triangleq \max_{\boldsymbol{x}^n \in \mathcal{X}^n} \mathbb{E}\left[\ell\left(\hat{\Pi}(\hat{\boldsymbol{x}}_1, ..., \hat{\boldsymbol{x}}_n), \Pi\right)\right], \quad (12)$$

where $\ell = \|\cdot\|_1$ or $\|\cdot\|_2^2$, $\hat{\Pi}$ is an estimate of Π and the expectation is with respect to $q(\cdot|\boldsymbol{x}_i)$ as well as all (possibly shared) randomness used by $q(\cdot|\boldsymbol{x}_i)$ $\forall i \in [n]$. For simplicity, we only focus on ℓ_2 error i.e., $\ell = \|\cdot\|_2^2$.

Ye and Barg (2018) show that the $\texttt{Subset Selection}$ achieves the order-optimal privacy-accuracy trade-off for frequency estimation i.e., $r_{\textsf{FE}}(\hat{\Pi}^{\texttt{ss}}, q^{\texttt{ss}}) = \Theta\left(\frac{d}{\min(e^\varepsilon, (e^\varepsilon - 1)^2, d)}\right)$ (where $\hat{\Pi}^{\texttt{ss}} := \frac{1}{n}\sum_i \hat{\boldsymbol{x}}_i^{\texttt{ss}}$). Like $\texttt{PrivUnit}_2$, compared to other (order-optimal) ε-LDP frequency estimation mechanisms, $\texttt{Subset Selection}$ admits the best constants and gives the smallest ℓ_2 error in practice (see Chen et al. (2020)). However, the communication cost associated with $\texttt{Subset Selection}$ is $O\left(\frac{d}{e^\varepsilon + 1}\right)$ bits per user, which which can be an issue for small and moderate ε.

Similar to $\texttt{PrivUnit}_2$, one could apply the generic $\texttt{MMRC}$ scheme defined in Section 3 to compress and simulate $\texttt{Subset Selection}$. However, for a fixed number of candidates N, it yields a biased estimate of $\boldsymbol{x}$ and hence cannot get the correct (optimal) order of estimation error in (12) i.e., the error would not decay with n. Fortunately, similar to $\texttt{PrivUnit}_2$, we show (in Section 5.1) that the bias can be corrected by appropriately translating and scaling the privatized version of $\boldsymbol{x}$ i.e., by using an estimator which is slightly different compared to the original estimator of $\texttt{Subset Selection}$. Further, we also show (in Section 5.2) that the resulting unbiased estimator for frequency estimation ($\hat{\Pi}^{\texttt{mmrc}}$) can simulate $\texttt{Subset Selection}$ closely while only using on the order of ε-bits communication.

5.1 Debiasing $\texttt{MMRC}$ to simulate $\texttt{Subset Selection}$

Let us focus on a single user and consider some data $\boldsymbol{x} \in \mathcal{X}$. Recall the $\texttt{Subset Selection}$ ε-LDP mechanism $q^{\texttt{ss}}$ described in Section 2 with $s := \lceil\frac{d}{1+e^\varepsilon}\rceil$. $\texttt{Subset Selection}$ is cap-based mechanism as discussed in Section 3 and Appendix G with $\mathsf{Cap}_{\boldsymbol{x}} = \mathcal{Z}_{\boldsymbol{x}}$ and $\mathbb{P}_{\boldsymbol{z} \sim \text{Unif}(\mathcal{Z})}(\boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}}) = s/d$. Similar to Section 4.1, let $\boldsymbol{z}_K$ be the privatized version of $\boldsymbol{x}$ using $\texttt{MMRC}$. We define $\hat{\boldsymbol{x}}^{\texttt{mmrc}} := (\boldsymbol{z}_K - b_{\texttt{mmrc}})/m_{\texttt{mmrc}}$ as the estimator of the $\texttt{MMRC}$ mechanism simulating $\texttt{Subset Selection}$ where $m_{\texttt{mmrc}}$ and $b_{\texttt{mmrc}}$ (defined in Appendix I.1) are translation and scaling factor analogous to $m_{\texttt{ss}}$ and $b_{\texttt{ss}}$ in (4). The following Lemma shows that $\hat{\boldsymbol{x}}^{\texttt{mmrc}}$ is an unbiased estimator. See Appendix I.1 for a proof.

Lemma 5.1. *Let $\hat{\boldsymbol{x}}^{mmrc}$ be the estimator of the* $\texttt{MMRC}$ *mechanism simulating* $\texttt{Subset Selection}$ *as defined above. Then, $\mathbb{E}[\hat{\boldsymbol{x}}^{mmrc}] = \boldsymbol{x}$.*

5.2 Simulating $\texttt{Subset Selection}$ using $\texttt{MMRC}$

Finally, we consider estimating the empirical frequency Π defined earlier using the $\texttt{MMRC}$ scheme simulating $\texttt{Subset Selection}$. To that end, consider n users and let $\hat{\boldsymbol{x}}_i^{\texttt{mmrc}}$ be the unbiased estimator of $\boldsymbol{x}_i$ at the i^{th} user. Let the (unbiased) estimate of Π be $\hat{\Pi}^{\texttt{mmrc}} := \frac{1}{n}\sum_i \hat{\boldsymbol{x}}_i^{\texttt{mmrc}}$. The following Theorem shows that, for frequency estimation, $\texttt{MMRC}$ can simulate $\texttt{Subset Selection}$ in a near-lossless manner (when λ is small) while only using on the order of ε bits of communication. A proof can be found in Appendix I.2. Similar to $\texttt{PrivUnit}_2$, the key idea in the proof is to show that when the number of candidates N is exponential in ε, the scaling factor $m_{\texttt{mmrc}}$ is close to the scaling parameter associated with $q^{\texttt{ss}}$ (i.e., $m_{\texttt{ss}}$ defined in (4)).

Theorem 5.1. *Let $r_{\textsf{FE}}\left(\hat{\Pi}^{ss}, q^{ss}\right)$ and $r_{\textsf{FE}}\left(\hat{\Pi}^{mmrc}, q^{mmrc}\right)$ be the empirical mean estimation error for* $\texttt{Subset Selection}$ *and* $\texttt{MMRC}$ *simulating* $\texttt{Subset Selection}$ *with N candidates respectively. Consider any $\lambda > 0$. Then*

$$r_{\textsf{FE}}\left(\hat{\Pi}^{mmrc}, q^{mmrc}\right) \leq \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right) r_{\textsf{FE}}\left(\hat{\Pi}^{ss}, q^{ss}\right),$$

as long as

$$N \geq \frac{2(e^\varepsilon + 1)^2(1+\lambda)^2}{0.24^2\lambda^2} \ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right). \quad (13)$$

Similar to mean estimation, while a specific value of λ can be chosen in (13), in practice, the number of bits could be fixed (see Section 5.3), determining the value of λ.

5.3 Empirical Comparisons.

Next, we empirically demonstrate the privacy-accuracy-communication tradeoffs of `MMRC` simulating `Subset Selection`. Along with `Subset Selection`, we compare against the RHR algorithm of Chen et al. (2020) which offers order-optimal privacy-accuracy tradeoffs while requiring only ε bits. Following Acharya et al. (2019), we generate $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ from the Zipf distribution with degree 1. We report the average ℓ_2 estimation error over 10 runs. See more variations in Appendix I.3.

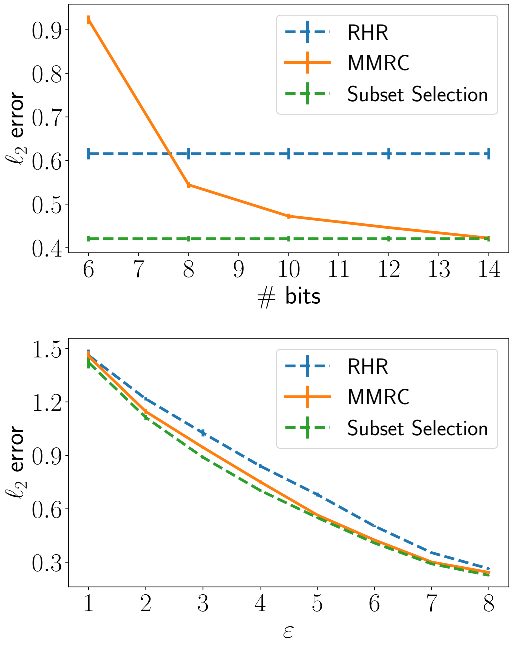

Figure 2: Comparing `Subset Selection`, `MMRC` simulating `Subset Selection` and RHR for frequency estimation with $d = 500$ and $n = 5000$. **Top:** ℓ_2 error vs #bits for $\varepsilon = 6$. **Bottom:** ℓ_2 error vs ε for #bits $= \max\{\lceil(\varepsilon/\ln 2)\rceil + 3, 8\}$. RHR uses #-bits $= \varepsilon$ for both as it leads to a poor performance if #-bits $> \varepsilon$.

In Figure 2 (top), we show the communication-accuracy tradeoffs. We see that with correct order of bits, the accuracy of `MMRC` simulating `Subset Selection` converges to the accuracy of the uncompressed `Subset Selection`. In Figure 2 (bottom), we show the privacy-accuracy tradeoffs. More specifically, `MMRC` simulating `Subset Selection` can attain the accuracy of the uncompressed `Subset Selection` for the range of ε's typically considered by LDP mechanisms while only using $\max\{\lceil(\varepsilon/\ln 2)\rceil + 3, 8\}$ bits.

6 CONCLUSION AND OPEN PROBLEMS

We demonstrated how Minimal Random Coding can be used to simulate a class of ε-LDP mechanisms in a manner which is communication efficient while preserving accuracy and differential privacy guarantees. Further, for mean and frequency estimation, we proposed unbiased versions of our schemes (relying only on translation and scaling) that attain the privacy-accuracy tradeoffs of the best known schemes i.e., $\texttt{PrivUnit}_2$ and `Subset Selection`, while requiring on the order of ε bits of communication.

We now briefly discuss a few non-trivial and interesting open questions.

Computational Cost. The computational cost of our approach, similar to Feldman and Talwar (2021) grows linearly in d and exponentially in ε (as we need $N = \exp(O(\varepsilon))$ candidates to properly simulate the optimal mechanisms). An important question for future research is therefore how to increase the computational efficiency of `MRC` and `MMRC` with respect to ε.

Privacy Amplification via Shuffling. As mentioned in Section 1.2, privacy amplification via shuffling techniques ensure a central $\varepsilon \approx 1$ even when the local $\varepsilon > 1$. While our method could be combined with these amplification techniques in principle, we leave the analysis of the privacy, utility, and communication guarantees of the resulting scheme as a question for future research.

Other schemes to simulate noisy channels. `MRC` is only one of several channel simulation schemes studied in information theory which could be considered for compression of ε-LDP mechanisms. Similar to `MRC`, other channel simulation schemes, e.g., rejection sampling (Harsha et al., 2007) or schemes based on the Poisson functional representation (Li and El Gamal, 2018), can also compress noisy signals to a number of bits which is close to the information contained in the signal (which decreases as noise increases). Analyzing these schemes for their effect on differential privacy guarantees is an interesting open question.

Shared Randomness. Finally, here we assumed the existence of a shared source of randomness. We further assumed that each user is using a different source of shared randomness. While shared randomnesss only adds to the cost of downlink and not uplink communication (which is usually the bottleneck in settings like federated learning), a question left for future research is how much communication is required to establish and select these sources of randomness.

Acknowledgements

We thank the anonymous reviewers of AISTATS 2022 for their comments and suggestions. We sincerely thank Jakub Konečný and Wennan Zhu for helpful discussions. We also thank Zachary Charles for support with the software infrastructure.

References

J. Acharya and Z. Sun. Communication complexity in locally private distribution estimation and heavy hitters. In *International Conference on Machine Learning*, pages 51–60, 2019.

J. Acharya, Z. Sun, and H. Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1120–1129. PMLR, 2019.

N. Agarwal, A. T. Suresh, F. X. X. Yu, S. Kumar, and B. McMahan. cpsgd: Communication-efficient and differentially-private distributed sgd. In *Advances in Neural Information Processing Systems*, pages 7564–7575, 2018.

D. Alistarh, D. Grubic, J. Li, R. Tomioka, and M. Vojnovic. Qsgd: Communication-efficient sgd via gradient quantization and encoding. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 1709–1720. Curran Associates, Inc., 2017.

B. Balle, J. Bell, A. Gascón, and K. Nissim. The privacy blanket of the shuffle model. In *Annual International Cryptology Conference*, pages 638–667. Springer, 2019.

L. P. Barnes, H. A. Inan, B. Isik, and A. Ozgur. rtop-k: A statistical estimation approach to distributed sgd, 2020.

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*, pages 127–135, 2015.

R. Bassily, K. Nissim, U. Stemmer, and A. Thakurta. Practical locally private heavy hitters. *arXiv preprint arXiv:1707.04982*, 2017.

C. H. Bennett and P. W. Shor. Entanglement-Assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem. *IEEE Trans. Info. Theory*, 48 (10), 2002.

A. Bhowmick, J. Duchi, J. Freudiger, G. Kapoor, and R. Rogers. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984*, 2018.

M. Bun, J. Nelson, and U. Stemmer. Heavy hitters and the structure of local privacy. *ACM Transactions on Algorithms (TALG)*, 15(4):1–40, 2019.

W.-N. Chen, P. Kairouz, and A. Özgür. Breaking the communication-privacy-accuracy trilemma. *arXiv preprint arXiv:2007.11707*, 2020.

W.-N. Chen, P. Kairouz, and A. Özgür. Breaking the dimension dependence in sparse distribution estimation under communication constraints. *arXiv preprint arXiv:2106.08597*, 2021.

P. Cuff. Communication requirements for generating correlated random variables. In *2008 IEEE International Symposium on Information Theory*, pages 1393–1397, 2008.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

U. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 21st ACM Conference on Computer and Communications Security*, Scottsdale, Arizona, 2014. URL https://arxiv.org/abs/1407.6981.

Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2468–2479. SIAM, 2019.

Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, S. Song, K. Talwar, and A. Thakurta. Encode, shuffle, analyze privacy revisited: Formalizations and empirical evaluation. *arXiv preprint arXiv:2001.03618*, 2020.

A. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 211–222, 2003.

V. Feldman and K. Talwar. Lossless compression of efficient private local randomizers. *arXiv preprint arXiv:2102.12099*, 2021.

V. Gandikota, D. Kane, R. K. Maity, and A. Mazumdar. vqsgd: Vector quantized stochastic gradient descent, 2019.

A. Girgis, D. Data, S. Diggavi, P. Kairouz, and A. T. Suresh. Shuffled model of differential privacy in federated learning. In *International Conference on Artificial Intelligence and Statistics*, pages 2521–2529. PMLR, 2021a.

A. M. Girgis, D. Data, S. Diggavi, P. Kairouz, and A. T. Suresh. Shuffled model of federated learning: Privacy, accuracy and communication trade-offs. *IEEE Journal on Selected Areas in Information Theory*, 2(1):464–478, 2021b.

Y. Han, P. Mukherjee, A. Ozgur, and T. Weissman. Distributed statistical estimation of high-dimensional and nonparametric distributions. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 506–510. IEEE, 2018a.

Y. Han, A. Özgür, and T. Weissman. Geometric lower bounds for distributed parameter estimation under communication constraints. *arXiv preprint arXiv:1802.08417*, 2018b.

P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan. The communication complexity of correlation. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 10–23. IEEE, 2007.

M. Havasi, R. Peharz, and J. M. Hernández-Lobato. Minimal Random Code Learning: Getting Bits Back from Compressed Model Parameters. In *International Conference on Learning Representations*, 2019.

P. Kairouz, K. Bonawitz, and D. Ramage. Discrete distribution estimation under local privacy. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pages 2436–2444, New York, New York, USA, 20–22 Jun 2016.

P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, R. G. L. D'Oliveira, H. Eichner, S. E. Rouayheb, D. Evans, J. Gardner, Z. Garrett, A. Gascón, B. Ghazi, P. B. Gibbons, M. Gruteser, Z. Harchaoui, C. He, L. He, Z. Huo, B. Hutchinson, J. Hsu, M. Jaggi, T. Javidi, G. Joshi, M. Khodak, J. Konecný, A. Korolova, F. Koushanfar, S. Koyejo, T. Lepoint, Y. Liu, P. Mittal, M. Mohri, R. Nock, A. Özgür, R. Pagh, H. Qi, D. Ramage, R. Raskar, M. Raykova, D. Song, W. Song, S. U. Stich, Z. Sun, A. T. Suresh, F. Tramèr, P. Vepakomma, J. Wang, L. Xiong, Z. Xu, Q. Yang, F. X. Yu, H. Yu, and S. Zhao. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021. ISSN 1935-8237. doi: 10.1561/2200000083. URL http://dx.doi.org/10.1561/2200000083.

S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40 (3):793–826, 2011.

C. T. Li and A. El Gamal. Strong functional representation lemma and applications to coding theorems. *IEEE Transactions on Information Theory*, 64(11):6967–6978, 2018.

T. T. Nguyên, X. Xiao, Y. Yang, S. C. Hui, H. Shin, and J. Shin. Collecting and analyzing data from smart device users with local differential privacy, 2016.

D. Ramage and S. Mazzocchi. Federated analytics: Collaborative data science without data collection. https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.html, 2020.

E. C. Song, P. Cuff, and H. V. Poor. The likelihood encoder for lossy compression. *IEEE Transactions on Information Theory*, 62(4):1836–1849, 2016. doi: 10.1109/TIT.2016.2529657.

A. T. Suresh, F. X. Yu, S. Kumar, and H. B. McMahan. Distributed mean estimation with limited communication. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 3329–3337. JMLR.org, 2017.

H. Wang, S. Sievert, S. Liu, Z. Charles, D. Papailiopoulos, and S. Wright. Atomo: Communication-efficient learning via atomic sparsification. In *Advances in Neural Information Processing Systems*, pages 9850–9861, 2018.

T. Wang, J. Zhao, X. Yang, and X. Ren. Locally differentially private data collection and analysis. *arXiv preprint arXiv:1906.01777*, 2019.

S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60 (309):63–69, 1965.

W. Wen, C. Xu, F. Yan, C. Wu, Y. Wang, Y. Chen, and H. Li. Terngrad: Ternary gradients to reduce communication in distributed deep learning. In *Advances in neural information processing systems*, pages 1509–1519, 2017.

A. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975. doi: 10.1109/TIT.1975.1055346.

G. Xu, W. Liu, and B. Chen. Wyners common information for continuous random variables - a lossy source coding interpretation. In *45th Annual Conference on Information Sciences and Systems*, pages 1–6, 2011. doi: 10.1109/CISS.2011.5766249.

M. Ye and A. Barg. Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Transactions on Information Theory*, 64(8):5662–5676, 2018.

Supplementary Material:
Optimal Compression of Locally Differentially Private Mechanisms

Organization. The Appendix is organized as follows. In Appendix A, we discuss the societal impact associated with our work. In Appendix B, we focus on `MRC` and provide the proofs of Theorem 3.1, Theorem 3.2, and Theorem 3.3. In Appendix C, we focus on `MMRC` and provide the proofs of Theorem 3.4 and Theorem 3.5. Further, we also provide Theorem C.1 where we show that `MMRC` can simulate any ε-LDP cap-based mechanism in a nearly lossless fashion with about ε bits of communication. In Appendix D, we provide additional preliminary on $\texttt{PrivUnit}_2$ and also show that $\texttt{PrivUnit}_2$ is a cap-based mechanism (Definition 3.1). In Appendix E, we show how $\texttt{PrivUnit}_2$ can be simulated using `MRC` analogous to how we simulated $\texttt{PrivUnit}_2$ using `MMRC` in Section 4. Along with the theoretical guarantees, we also provide some empirical comparisons between `MRC` simulating $\texttt{PrivUnit}_2$ and $\texttt{PrivUnit}_2$. In Appendix F, we provide the proofs of Lemma 4.1 and Theorem 4.1 as well as some additional empirical comparisons between `MMRC` simulating $\texttt{PrivUnit}_2$ and $\texttt{PrivUnit}_2$. In Appendix G, we provide additional preliminary on `Subset Selection` and also show that `Subset Selection` is a cap-based mechanism (Definition 3.1). In Appendix H, we show how `Subset Selection` can be simulated using `MRC` analogous to how we simulated `Subset Selection` using `MMRC` in Section 5. Along with the theoretical guarantees, we also provide some empirical comparisons between `MRC` simulating `Subset Selection` and `Subset Selection`. In Appendix I, we provide the proofs of Lemma 5.1 and Theorem 5.1 as well as some additional empirical comparisons between `MMRC` simulating `Subset Selection` and `Subset Selection`.

A SOCIETAL IMPACT

Collecting large datasets allows us to build better machine learning models which can facilitate our lives in many different ways. However, harnessing data from devices can expose their users to privacy risks. Research into differential privacy can help to minimize these risks. At present, our work is mostly theoretical in nature as there are a few unsolved questions. In particular, for large ε the computational complexity of our approach may be too expensive to be practical.

B MINIMAL RANDOM CODING

Let $q(\boldsymbol{z}|\boldsymbol{x})$ be an ε-LDP mechanism for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{Z}$. Let $p(\boldsymbol{z})$ be the fixed reference distribution over $\mathcal{Z}$ and let $\{\boldsymbol{z}_k\}_{k=1}^N$ be N candidates drawn from $p(\boldsymbol{z})$. From Algorithm 1, the distribution over the indices $k \in [N]$ under minimal random coding (`MRC`) is as follows:

$$\pi_{\boldsymbol{x}}^{\texttt{mrc}}(k) := \frac{q(\boldsymbol{z}_k|\boldsymbol{x})/p(\boldsymbol{z}_k)}{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})}$$

$\pi^{\texttt{mrc}}$ can be viewed as a function that maps $\boldsymbol{x}$ and $(\boldsymbol{z}_1, ..., \boldsymbol{z}_N)$ to a distribution in $[N]$. However for notational convenience, when the context is clear, we will omit the dependence on $\boldsymbol{x}$ and $(\boldsymbol{z}_1, ..., \boldsymbol{z}_N)$.

Let $q^{\texttt{mrc}}$ denote the distribution of $\boldsymbol{z}_K$ where $K \sim \pi^{\texttt{mrc}}(\cdot)$ i.e., with $\delta(\cdot)$ denoting the Dirac delta function:

$$q^{\texttt{mrc}}(\boldsymbol{z}|\boldsymbol{x}) := \sum_k \pi^{\texttt{mrc}}(k)\delta(\boldsymbol{z} - \boldsymbol{z}_k).$$

B.1 Utility of `MRC`

In this section, we prove Theorem 3.1 i.e., we show that `MRC` can simulate any ε-LDP mechanism in a nearly lossless fashion with about ε bits of communication

Theorem 3.1 (Utility of MRC)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP mechanism $q(\cdot|\boldsymbol{x})$. Consider any reference distribution $p(\cdot)$ such that $|\ln(q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z}))| \le \varepsilon \;\forall\; \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z}$.*[10] *Let the number of candidates be $N = 2^{(\log e + 4c)\varepsilon}$ for some constant $c \ge 0$. Then, for $\alpha \in [0, 1/2]$, $q^{\texttt{mrc}}$ is such that*

$$\left|\mathbb{E}_{q^{mrc}}\left[\|\boldsymbol{z}-\boldsymbol{x}\|^2\right]-\mathbb{E}_q\left[\|\boldsymbol{z}-\boldsymbol{x}\|^2\right]\right| \le \frac{2\alpha\sqrt{\mathbb{E}_q[\|\boldsymbol{z}-\boldsymbol{x}\|^4]}}{1-\alpha} \tag{5}$$

holds with probability at least $1-2\alpha$, with c and α related by the following: $\alpha = \sqrt{2^{-c\varepsilon} + 2^{-c^2/\log e + 1}}$.

Proof. In order to prove this theorem, we invoke Theorem 3.2 of Havasi et al. (2019).

Recall Theorem 3.2 (Havasi et al., 2019): Let q' and p be distributions over $\mathcal{Z}$. Let $t \ge 0$ be some constant and let $N' = 2^{\left(D_{\mathsf{KL}}\left(q'(\boldsymbol{z})\,\|p(\boldsymbol{z})\right)+t\right)}$. Let $\tilde{q}$ be a discrete distribution constructed by drawing N' samples $\{\boldsymbol{z}_k\}_{k=1}^{N'}$ from p and defining

$$\tilde{q}(\boldsymbol{z}) \coloneqq \sum_{k=1}^{N'} \frac{q(\boldsymbol{z}_k)/p(\boldsymbol{z}_k)}{\sum_{k'} q(\boldsymbol{z}_{k'})/p(\boldsymbol{z}_{k'})}\delta(\boldsymbol{z}-\boldsymbol{z}_k).$$

Furthermore, let f be a measurable function and $\|f\|_{q'} = \sqrt{\mathbb{E}_{q'(\boldsymbol{z})}[f^2(\boldsymbol{z})]}$ be its 2-norm under q'. Then it holds that

$$\mathbb{P}\left(\left|\mathbb{E}_{\tilde{q}(\boldsymbol{z})}[f(\boldsymbol{z})] - \mathbb{E}_{q'(\boldsymbol{z})}[f(\boldsymbol{z})]\right| \ge \frac{2\|f\|_{q'}\alpha'}{1-\alpha'}\right) \le 2\alpha'$$

where

$$\alpha' = \sqrt{2^{-t/4} + 2\sqrt{\mathbb{P}(\log(q'(\boldsymbol{z})/p(\boldsymbol{z})) > D_{\mathsf{KL}}\left(q'(\boldsymbol{z})\,\|p(\boldsymbol{z})\right) + t/2)}}.$$

We apply Theorem 3.2 (Havasi et al., 2019) to $q'(\boldsymbol{z}) \coloneqq q(\boldsymbol{z}|\boldsymbol{x})$ and $f(\boldsymbol{z}) \coloneqq \|\boldsymbol{z}-\boldsymbol{x}\|^2$. We identify $\tilde{q}(\boldsymbol{z}) = q^{\texttt{mrc}}(\boldsymbol{z}|\boldsymbol{x})$ and $N' = N$. To prove Theorem 3.1, it suffices to show that $\alpha \ge \alpha'$. Note that

$$D_{\mathsf{KL}}\left(q(\boldsymbol{z}|\boldsymbol{x})\,\|p(\boldsymbol{z})\right) \overset{(a)}{=} \mathbb{E}_{q(\boldsymbol{z}|\boldsymbol{x})}\left[\log\left(\frac{q(\boldsymbol{z}|\boldsymbol{x})}{p(\boldsymbol{z})}\right)\right] \overset{(b)}{\le} \varepsilon\log e,$$

where (a) follows the definition of KL-divergence and (b) follows since $|\log(q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z}))| \le \varepsilon\log e$ by the assumption on p. We therefore have

$$t = (\log e + 4c)\varepsilon - D_{\mathsf{KL}}\left(q(\boldsymbol{z}|\boldsymbol{x})\,\|p(\boldsymbol{z})\right) \ge 4c\varepsilon.$$

It follows that

$$\begin{aligned}\mathbb{P}\big(\log(q(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z})) > D_{\mathsf{KL}}\left(q(\boldsymbol{z}|\boldsymbol{x})\,\|p(\boldsymbol{z})\right) + t/2\big) &\le \mathbb{P}\big(\log(q(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z})) > \mathbb{E}\big[\log(q(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z}))\big] + 2c\varepsilon\big)\\ &\overset{(b)}{\le} \exp(-2c^2/(\log e)^2) = 2^{-2c^2/\log e}.\end{aligned}$$

where (b) follows from Hoeffding's inequality since $|\log(q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z}))| \le \varepsilon\log e$ by the assumption on p. Therefore,

$$\alpha' \le \sqrt{2^{-c\varepsilon} + 2\sqrt{2^{-2c^2/\log e}}} = \sqrt{2^{-c\varepsilon} + 2^{-c^2/\log e + 1}} = \alpha.$$

□

Remark B.1. *For most ε-LDP mechanisms $q(\cdot|\boldsymbol{x})$, the term $\mathbb{E}_q\left[\|\boldsymbol{z}-\boldsymbol{x}\|^4\right]$ in (5) can be well-controlled. For instance, for* Subset Selection *and* PrivUnit$_2$*, the output spaces are bounded, and therefore, $\sqrt{\mathbb{E}_q[\|\boldsymbol{z}-\boldsymbol{x}\|^4]}$ is of the same order as $\mathbb{E}_q\left[\|\boldsymbol{z}-\boldsymbol{x}\|^2\right]$. Therefore, by making α small enough (in Theorem 3.1) i.e. by increasing c, the estimation error of* MRC *can be arbitrarily close to the estimation error of the underlying scheme it is simulating.*

[10] Note that this condition holds for many reference distributions $p(\cdot)$. For example, one can simply choose $p(\cdot) = q(\cdot|\boldsymbol{x}^*)$ for some $\boldsymbol{x}^* \in \mathcal{X}$.

B.2 Privacy of MRC

B.2.1 Pure Privacy of MRC

In this section, we prove Theorem 3.2 i.e., we show that $\pi^{\texttt{mrc}}$ is a 2ε-LDP mechanism.

Theorem 3.2 (Pure DP guarantee of MRC)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, and data $\boldsymbol{x} \in \mathcal{X}$. Consider any ε-LDP mechanism $q(\cdot|\boldsymbol{x})$, reference distribution $p(\cdot)$, and number of candidates $N \geq 1$. Then, $\pi^{\texttt{mrc}}(\cdot)$ obtained from Algorithm 1 is a 2ε-LDP mechanism.*

Proof. For any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z}$, using the definition of an ε-LDP mechanism, we have

$$q(\boldsymbol{z}|\boldsymbol{x}) \leq \exp(\varepsilon) q(\boldsymbol{z}|\boldsymbol{x}'). \tag{14}$$

For any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \{\boldsymbol{z}_k\}_{k=1}^N \in \mathcal{Z}^N$ and $k \in [N]$, we have

$$\begin{aligned}
\frac{\pi_{\boldsymbol{x}}^{\texttt{mrc}}(k)}{\pi_{\boldsymbol{x}'}^{\texttt{mrc}}(k)} &\overset{(a)}{=} \frac{q(\boldsymbol{z}_k|\boldsymbol{x})}{q(\boldsymbol{z}_k|\boldsymbol{x}')} \times \frac{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x}')/p(\boldsymbol{z}_{k'})}{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})} \\
&\overset{(b)}{\leq} \exp(\varepsilon) \times \frac{\sum_{k'} \exp(\varepsilon) q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})}{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})} \\
&= \exp(\varepsilon) \times \frac{\exp(\varepsilon) \sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})}{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})} \\
&= \exp(2\varepsilon).
\end{aligned}$$

where (a) follows from the definition of $\pi^{\texttt{mrc}}$ and (b) follows from (14). □

B.2.2 Approximate Privacy of MRC

In this section, we prove Theorem 3.3 i.e., we provide the approximate DP guarantee of $\pi^{\texttt{mrc}}$.

Theorem 3.3 (Approximate DP guarantee of MRC)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP mechanism $q(\cdot|\boldsymbol{x})$. Consider any reference distribution $p(\cdot)$ such that $|\ln(q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z}))| \leq \varepsilon$ $\forall\ \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \mathcal{Z}$.[5] Let $c_0 \geq 0$ be some constant and let the number of candidates $N = \exp(2\varepsilon + 2c_0)$. Then, for any $\delta \leq 1$, $\pi^{\texttt{mrc}}(\cdot)$ obtained from Algorithm 1 is $(\varepsilon + \varepsilon_0, \delta)$-LDP mechanism where*

$$\varepsilon_0 \coloneqq \ln \frac{1 + a_0}{1 - a_0} \qquad \textit{and} \qquad a_0 \coloneqq \exp(-c_0) \sqrt{\tfrac{1}{2} \ln \tfrac{2}{\delta}}.$$

Proof. Fix any $\boldsymbol{x} \in \mathcal{X}$. Let us define the following random variable:

$$w(\boldsymbol{z}|\boldsymbol{x}) = q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z}). \tag{15}$$

Assuming $\boldsymbol{z} \sim p(\cdot)$, the expected value of the random variable $w(\boldsymbol{z}|\boldsymbol{x})$ is

$$\mathbb{E}_p[w(\boldsymbol{z}|\boldsymbol{x})] = \mathbb{E}_p[q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z})] = \int_{\boldsymbol{z} \in \mathcal{Z}} q(\boldsymbol{z}|\boldsymbol{x}) = 1.$$

Further, the random variable $w(\boldsymbol{z}|\boldsymbol{x})$ can be bounded as follows:

$$|w(\boldsymbol{z}|\boldsymbol{x})| = |q(\boldsymbol{z}|\boldsymbol{x})/p(\boldsymbol{z})| \overset{(a)}{\leq} \exp(\varepsilon).$$

where (a) follows from the assumption on $p(\cdot)$. Therefore, we have

$$\mathbb{P}\left(\left|\frac{1}{N} \sum_{k=1}^N w(\boldsymbol{z}_k|\boldsymbol{x}) - 1\right| \geq a_0\right) \overset{(a)}{\leq} 2 \exp\left(\frac{-2Na_0^2}{(\exp(\varepsilon) - \exp(-\varepsilon))^2}\right) \leq 2 \exp\left(\frac{-2Na_0^2}{\exp(2\varepsilon)}\right) \overset{(b)}{=} \delta \tag{16}$$

where (a) follows from Hoeffding's inequality and (b) follows from the definition of a_0 and N. Now, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \{\boldsymbol{z}_k\}_{k=1}^N \in \mathcal{Z}^N$ and $k \in [N]$, we have

$$\frac{\pi_{\boldsymbol{x}}^{\texttt{mrc}}(k)}{\pi_{\boldsymbol{x}'}^{\texttt{mrc}}(k)} \overset{(a)}{=} \frac{q(\boldsymbol{z}_k|\boldsymbol{x})}{q(\boldsymbol{z}_k|\boldsymbol{x}')} \times \frac{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x}')/p(\boldsymbol{z}_{k'})}{\sum_{k'} q(\boldsymbol{z}_{k'}|\boldsymbol{x})/p(\boldsymbol{z}_{k'})}$$

$$\begin{aligned}&\overset{(b)}{=} \frac{q(\boldsymbol{z}_k|\boldsymbol{x})}{q(\boldsymbol{z}_k|\boldsymbol{x}')} \times \frac{\sum_{k'} w(\boldsymbol{z}_{k'}|\boldsymbol{x}')}{\sum_{k'} w(\boldsymbol{z}_{k'}|\boldsymbol{x})}\\ &\overset{(c)}{\leq} \exp(\varepsilon) \times \frac{\frac{1}{N}\sum_{k'} w(\boldsymbol{z}_{k'}|\boldsymbol{x}')}{\frac{1}{N}\sum_{k'} w(\boldsymbol{z}_{k'}|\boldsymbol{x})}\end{aligned} \tag{17}$$

where (a) follows from the definition of $\pi^{\texttt{mrc}}$, (b) follows from (15) and (c) follows from (14). Now, using (16) in (17), we have with probability at least $1-\delta$:

$$\frac{\pi^{\texttt{mrc}}_{\boldsymbol{x}}(k)}{\pi^{\texttt{mrc}}_{\boldsymbol{x}'}(k)} \leq \exp(\varepsilon) \times \frac{1+a_0}{1-a_0} \overset{(a)}{=} \exp(\varepsilon+\varepsilon_0)$$

where (a) follows from the definition of ε_0. □

C MODIFIED MINIMAL RANDOM CODING

Let $q(\boldsymbol{z}|\boldsymbol{x})$ be an ε-LDP cap-based mechanism (see definition 3.1) for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{Z}$. Let $p(\boldsymbol{z})$ be the uniform distribution over $\mathcal{Z}$ and let $\{\boldsymbol{z}_k\}_{k=1}^N$ be N candidates drawn from $p(\boldsymbol{z})$. Let θ denote the fraction of candidates inside the $\texttt{Cap}_{\boldsymbol{x}}$ associated with $q(\boldsymbol{z}|\boldsymbol{x})$. Let $\pi^{\texttt{mmrc}}$ be the distribution over the indices $k \in [N]$ under modified minimal random coding (MMRC) obtained from Algorithm 2. Recall that $\pi^{\texttt{mmrc}}(k)$ is bounded by an upper threshold t_u and a lower threshold t_l (Section 3.2),

$$t_u = \frac{1}{N} \times \frac{c_1(\varepsilon,d)}{\mathbb{E}[\theta]c_1(\varepsilon,d)+(1-\mathbb{E}[\theta])c_2(\varepsilon,d)}, \qquad t_l = \frac{1}{N} \times \frac{c_2(\varepsilon,d)}{\mathbb{E}[\theta]c_1(\varepsilon,d)+(1-\mathbb{E}[\theta])c_2(\varepsilon,d)}.$$

Similar to $\pi^{\texttt{mrc}}$, $\pi^{\texttt{mmrc}}$ can be be viewed as a function that maps $\boldsymbol{x}$ and $(\boldsymbol{z}_1,...,\boldsymbol{z}_N)$ to a distribution in $[N]$. However, to reduce clutter, we will generally omit the dependence on $\boldsymbol{x}$ and $(\boldsymbol{z}_1,...,\boldsymbol{z}_N)$. Further, since $\pi^{\texttt{mmrc}}$ depends on $(\boldsymbol{z}_1,...,\boldsymbol{z}_N)$ only through θ, we will sometimes show this dependence as $\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}$.

Finally, let $q^{\texttt{mmrc}}$ denote the distribution of $\boldsymbol{z}_K$ where $K \sim \pi^{\texttt{mmrc}}$. That is, with δ denoting the Dirac delta function:

$$q^{\texttt{mmrc}}(\boldsymbol{z}|\boldsymbol{x}) := \sum_k \pi^{\texttt{mmrc}}(k)\delta(\boldsymbol{z}-\boldsymbol{z}_k).$$

C.1 Privacy of MMRC

In this section, we prove Theorem 3.4 i.e., we show that $\pi^{\texttt{mmrc}}$ is a ε-LDP mechanism.

Theorem 3.4 (DP guarantee of MMRC)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP cap-based mechanism $q(\cdot|\boldsymbol{x})$. Let the reference distribution $p(\cdot)$ be the uniform distribution on $\mathcal{Z}$. Consider any number of candidates $N \geq 1$. Then, $\pi^{\texttt{mmrc}}(\cdot)$ obtained from Algorithm 2 is an ε-LDP mechanism.*

Proof. For any ε-LDP cap-based $q(\cdot|\boldsymbol{x})$, we have the following from (1) and (6):

$$\frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)} \leq \exp(\varepsilon). \tag{18}$$

Further, the modification of $\pi^{\texttt{mrc}}$ to $\pi^{\texttt{mmrc}}$ ensures that (8) is true, that is,

$$t_l \leq \pi^{\texttt{mmrc}}(k) \leq t_u \ \forall k \in [N].$$

Therefore, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ and $k \in [N]$, we have

$$\frac{\pi^{\texttt{mmrc}}_{\boldsymbol{x}}(k)}{\pi^{\texttt{mmrc}}_{\boldsymbol{x}'}(k)} \leq \frac{t_u}{t_l} \overset{(a)}{=} \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)} \overset{(b)}{\leq} \exp(\varepsilon),$$

where (a) follows from the definitions of t_u and t_l and (b) follows from (18). □

C.2 Supporting Lemmas to prove the utility of MMRC

To prove Theorem 3.5 (Section C.3), we prove that the expected KL divergence between $\pi^{\texttt{mrc}}$ and $\pi^{\texttt{mmrc}}$ can be controlled arbitrarily when the number of candidates is of the right order (Lemma C.2). To prove Lemma C.2, we first show that the KL divergence between $\pi^{\texttt{mrc}}$ and $\pi^{\texttt{mmrc}}$, for a given fraction of candidates inside the $\mathsf{Cap}_{\boldsymbol{x}}$, can be bounded in terms of ε (Lemma C.1).

C.2.1 The KL divergence between $\pi^{\texttt{mrc}}$ and $\pi^{\texttt{mmrc}}$ is small

Lemma C.1. *Let $q(\boldsymbol{z}|\boldsymbol{x})$ be an ε-LDP cap-based mechanism. Let $p(\boldsymbol{z})$ be the uniform distribution over $\mathcal{Z}$ and let $\{\boldsymbol{z}_k\}_{k=1}^N$ be N candidates drawn from $p(\boldsymbol{z})$. Let θ denote the fraction of candidates inside the $\mathsf{Cap}_{\boldsymbol{x}}$ associated with $q(\boldsymbol{z}|\boldsymbol{x})$. Let π^{mrc} be the distribution over the indices $k \in [N]$ under MRC obtained from Algorithm 1 and π^{mmrc} be the distribution over the indices $k \in [N]$ under MMRC obtained from Algorithm 2. Then,*

$$D_{\mathsf{KL}}\left(\pi^{mrc}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{mmrc}_{\boldsymbol{x},\theta}(\cdot)\right) \leq \varepsilon \log e$$

Proof. We consider three different cases depending on whether $\theta = \mathbb{E}[\theta]$, $\theta < \mathbb{E}[\theta]$ or $\theta > \mathbb{E}[\theta]$.

1. For $\theta = \mathbb{E}[\theta]$, we have $\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot) = \pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)$. Therefore,

$$D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) = D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\right) = 0 \leq \varepsilon \log e. \tag{19}$$

2. If $\theta < \mathbb{E}[\theta]$, then $\pi^{\texttt{mrc}}$ violates the upper threshold t_u so that $\pi^{\texttt{mmrc}}(k) = t_u$ for all $k \in \mathsf{Cap}_{\boldsymbol{x}}$ and we have

$$\begin{aligned}
&D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) = \sum_k \pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k) \log \frac{\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k)}{\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)} \\
&\overset{(a)}{=} \sum_{k\in\mathsf{Cap}_{\boldsymbol{x}}} \frac{1}{N} \times \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)+\theta(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \log \frac{c_2(\varepsilon,d)+\mathbb{E}[\theta]\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{c_2(\varepsilon,d)+\theta\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))} + \sum_{k\notin\mathsf{Cap}_{\boldsymbol{x}}} \frac{1}{N}\times \\
&\qquad \frac{c_2(\varepsilon,d)}{c_2(\varepsilon,d)+\theta(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \Bigg[\log \frac{c_2(\varepsilon,d)+\mathbb{E}[\theta]\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{c_2(\varepsilon,d)+\theta\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \\
&\qquad\qquad + \log \frac{(1-\theta)\times c_2(\varepsilon,d)}{(1-\mathbb{E}[\theta])\times c_2(\varepsilon,d)+(\mathbb{E}[\theta]-\theta)c_1(\varepsilon,d)}\Bigg] \\
&\overset{(b)}{=} \frac{\theta c_1(\varepsilon,d)}{c_2(\varepsilon,d)+\theta(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \log \frac{c_2(\varepsilon,d)+\mathbb{E}[\theta]\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{c_2(\varepsilon,d)+\theta\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))} + \\
&\qquad \frac{(1-\theta)c_2(\varepsilon,d)}{c_2(\varepsilon,d)+\theta(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \Bigg[\log \frac{c_2(\varepsilon,d)+\mathbb{E}[\theta]\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{c_2(\varepsilon,d)+\theta\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \\
&\qquad\qquad + \log \frac{(1-\theta)\times c_2(\varepsilon,d)}{(1-\mathbb{E}[\theta])\times c_2(\varepsilon,d)+(\mathbb{E}[\theta]-\theta)c_1(\varepsilon,d)}\Bigg] \\
&= \log \frac{c_2(\varepsilon,d)+\mathbb{E}[\theta]\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{c_2(\varepsilon,d)+\theta\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \\
&\quad + \frac{(1-\theta)c_2(\varepsilon,d)}{c_2(\varepsilon,d)+\theta(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \left[\log \frac{(1-\theta)\times c_2(\varepsilon,d)}{(1-\mathbb{E}[\theta])\times c_2(\varepsilon,d)+(\mathbb{E}[\theta]-\theta)c_1(\varepsilon,d)}\right] \\
&\overset{(c)}{\leq} \log \frac{c_2(\varepsilon,d)+\mathbb{E}[\theta]\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{c_2(\varepsilon,d)+\theta\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))} \overset{(d)}{\leq} \log\left(\frac{c_2(\varepsilon,d)+\mathbb{E}[\theta]\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{c_2(\varepsilon,d)}\right) \qquad (20) \\
&\overset{(e)}{\leq} \log \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)} \overset{(f)}{\leq} \varepsilon \log e
\end{aligned}$$

where (a) follows from the definition of $\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k)$ and $\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)$, (b) follows because $|\{k : k \in \mathsf{Cap}_{\boldsymbol{x}}\}| = \theta N$ and $|\{k : k \notin \mathsf{Cap}_{\boldsymbol{x}}\}| = (1-\theta)N$, (c) follows because $\log \frac{(1-\theta)\times c_2(\varepsilon,d)}{(1-\mathbb{E}[\theta])\times c_2(\varepsilon,d)+(\mathbb{E}[\theta]-\theta)c_1(\varepsilon,d)} \leq 0$, (d) follows because $\theta \geq 0$, (e) follows because $\mathbb{E}[\theta] \leq 1$, and (f) follows because $c_1(\varepsilon,d)/c_2(\varepsilon,d) \leq \exp(\varepsilon)$.

3. For $\theta > \mathbb{E}[\theta]$, we have

$$
\begin{aligned}
&D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) = \sum_{\boldsymbol{z}_i} \pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k) \log \frac{\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k)}{\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)} \\
&\overset{(a)}{=} \sum_{k \notin \mathsf{Cap}_{\boldsymbol{x}}} \frac{1}{N} \times \frac{c_2(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \log \frac{c_2(\varepsilon,d) + \mathbb{E}[\theta] \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))} + \sum_{k \in \mathsf{Cap}_{\boldsymbol{x}}} \frac{1}{N} \times \\
&\quad \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \left[\log \frac{c_2(\varepsilon,d) + \mathbb{E}[\theta] \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))} + \right. \\
&\qquad \left. \log \frac{\theta c_1(\varepsilon,d)}{\mathbb{E}[\theta] c_1(\varepsilon,d) + (\theta - \mathbb{E}[\theta]) \times c_2(\varepsilon,d)} \right] \\
&\overset{(b)}{=} \frac{(1-\theta) \times c_2(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \log \frac{c_2(\varepsilon,d) + \mathbb{E}[\theta] \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))} + \\
&\quad \frac{\theta c_1(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \left[\log \frac{c_2(\varepsilon,d) + \mathbb{E}[\theta] \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))} + \right. \\
&\qquad \left. \log \frac{\theta c_1(\varepsilon,d)}{\mathbb{E}[\theta] c_1(\varepsilon,d) + (\theta - \mathbb{E}[\theta]) \times c_2(\varepsilon,d)} \right] \\
&\overset{(c)}{\leq} \frac{\theta c_1(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \log \left(\frac{\theta c_1(\varepsilon,d)}{\mathbb{E}[\theta] c_1(\varepsilon,d) + (\theta - \mathbb{E}[\theta]) \times c_2(\varepsilon,d)} \right) \qquad (21) \\
&\overset{(d)}{\leq} \log \left(\frac{c_1(\varepsilon,d)}{\mathbb{E}[\theta] c_1(\varepsilon,d) + (1 - \mathbb{E}[\theta]) \times c_2(\varepsilon,d)} \right) \overset{(e)}{\leq} \log \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)} \overset{(f)}{\leq} \varepsilon \log e
\end{aligned}
$$

where (a) follows from the definition of $\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k)$ and $\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)$, (b) follows because $|\{k : k \in \mathsf{Cap}_{\boldsymbol{x}}\}| = \theta N$ and $|\{k : k \notin \mathsf{Cap}_{\boldsymbol{x}}\}| = (1-\theta)N$, (c) follows because $\log \frac{c_2(\varepsilon,d) + \mathbb{E}[\theta] \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))} \leq 0$, (d) follows because $\theta \leq 1$, (e) follows because $\mathbb{E}[\theta] \geq 0$, and (f) follows because $c_1(\varepsilon,d)/c_2(\varepsilon,d) \leq \exp(\varepsilon)$.

□

C.2.2 The expected KL divergence between the distribution of indices in MRC and MMRC can be controlled arbitrarily when N is in the right order

Lemma C.2. *Let $q(\boldsymbol{z}|\boldsymbol{x})$ be an ε-LDP cap-based mechanism. Let $p(\boldsymbol{z})$ be the uniform distribution over $\mathcal{Z}$ and let $\{\boldsymbol{z}_k\}_{k=1}^N$ be N candidates drawn from $p(\boldsymbol{z})$. Let θ denote the fraction of candidates inside the $\mathsf{Cap}_{\boldsymbol{x}}$ associated with $q(\boldsymbol{z}|\boldsymbol{x})$. Let $\pi^{\texttt{mrc}}$ be the distribution over the indices $k \in [N]$ under `MRC` obtained from Algorithm 1 and $\pi^{\texttt{mmrc}}$ be the distribution over the indices $k \in [N]$ under `MMRC` obtained from Algorithm 2. Then,*

$$\mathbb{E}_\theta \left[D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) \right] \leq \rho \times \log e \times (1 + \varepsilon)$$

where $\rho \in (0,1)$ is a free variable that is related to N as follows:

$$N = \frac{2\left(\exp(\varepsilon) - 1\right)^2}{\rho^2} \ln \frac{2}{\rho}.$$

Proof. Let θ denote the fraction of candidates inside the cap, i.e.,

$$\theta = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}(\boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}}).$$

Therefore, we have

$$\mathbb{E}[\theta] = \mathbb{P}_{\boldsymbol{z}_k \sim \mathrm{Unif}(\mathcal{Z})}\left(\boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}}\right) = \mathbb{P}_{\boldsymbol{z} \sim \mathrm{Unif}(\mathcal{Z})}\left(\boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}}\right). \qquad (22)$$

Now, using the Hoeffding's inequality, we have $\mathbb{P}\left\{|\theta - \mathbb{E}[\theta]| \geq \sqrt{\frac{\ln(2/\rho)}{2N}}\right\} \leq \rho$. Letting $\hat{\rho} = \sqrt{\frac{\ln(2/\rho)}{2N}}$, we have

$$\begin{aligned}\mathbb{E}_\theta\left[D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right)\right] &= \sum_{\theta}\mathbb{P}(\theta) \times D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right)\\ &= \sum_{\theta:|\theta-\mathbb{E}[\theta]|\leq\hat{\rho}}\mathbb{P}(\theta) D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) + \sum_{\theta:|\theta-\mathbb{E}[\theta]|>\hat{\rho}}\mathbb{P}(\theta) D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right)\end{aligned} \tag{23}$$

Now, we will upper bound $D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right)$ whenever θ is such that $|\theta - \mathbb{E}[\theta]| \leq \hat{\rho}$. As in the proof of Lemma C.1, we have three different cases depending on whether $\theta = \mathbb{E}[\theta]$, $\theta < \mathbb{E}[\theta]$ or $\theta > \mathbb{E}[\theta]$.

1. For $\theta = \mathbb{E}[\theta]$, using (19), we have $D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) = 0$.

2. For $\theta < \mathbb{E}[\theta]$, using (20), we have

$$\begin{aligned}D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) &\leq \log\frac{c_2(\varepsilon,d) + \mathbb{E}[\theta] \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}\\ &\overset{(a)}{=} \log\frac{c_2(\varepsilon,d) + \mathbb{E}[\theta] \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + (\mathbb{E}[\theta] - t) \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}\\ &= \log\left(1 + \frac{t \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + (\mathbb{E}[\theta] - t) \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}\right)\\ &\overset{(b)}{\leq} \frac{\log e \times t \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + (\mathbb{E}[\theta] - t) \times (c_1(\varepsilon,d) - c_2(\varepsilon,d))}\\ &\overset{(c)}{\leq} \log e \times t \times \left(\frac{c_1(\varepsilon,d) - c_2(\varepsilon,d)}{c_2(\varepsilon,d)}\right) \overset{(d)}{\leq} \log e \times \hat{\rho} \times \left(\frac{c_1(\varepsilon,d) - c_2(\varepsilon,d)}{c_2(\varepsilon,d)}\right)\end{aligned} \tag{24}$$

   where (a) follows by letting $\theta = \mathbb{E}[\theta] - t$ with $t > 0$, (b) follows by using $\log(1+x) \leq x\log e$ for $x = \frac{t\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{1+(\mathbb{E}[\theta]-t)\times(c_1(\varepsilon,d)-c_2(\varepsilon,d))} > 0$, (c) follows because $\mathbb{E}[\theta] - t = \theta \geq 0$, and (d) follows because $t = \mathbb{E}[\theta] - \theta \leq \hat{\rho}$.

3. For $\theta > \mathbb{E}[\theta]$, using (21), we have

$$\begin{aligned}D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) &\leq \frac{\theta c_1(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))}\log\left(\frac{\theta c_1(\varepsilon,d)}{\mathbb{E}[\theta]c_1(\varepsilon,d) + (\theta - \mathbb{E}[\theta]) \times c_2(\varepsilon,d)}\right)\\ &\overset{(a)}{\leq} \log\left(\frac{\theta c_1(\varepsilon,d)}{\mathbb{E}[\theta]c_1(\varepsilon,d) + (\theta - \mathbb{E}[\theta]) \times c_2(\varepsilon,d)}\right)\\ &\overset{(b)}{=} \log\left(\frac{(\mathbb{E}[\theta] + t)c_1(\varepsilon,d)}{\mathbb{E}[\theta]c_1(\varepsilon,d) + tc_2(\varepsilon,d)}\right)\\ &= \log\left(1 + \frac{t\,(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{\mathbb{E}[\theta]c_1(\varepsilon,d) + tc_2(\varepsilon,d)}\right)\\ &\overset{(c)}{\leq} \frac{\log e \times t\,(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{\mathbb{E}[\theta]c_1(\varepsilon,d) + tc_2(\varepsilon,d)}\\ &\overset{(d)}{\leq} \frac{\log e \times t\,(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{\mathbb{E}[\theta]c_1(\varepsilon,d)} \overset{(e)}{\leq} \frac{\log e \times \hat{\rho}\,(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{\mathbb{E}[\theta]c_1(\varepsilon,d)}\end{aligned} \tag{25}$$

   where (a) follows because $\theta \leq 1$, (b) follows by letting $\theta = \mathbb{E}[\theta] + t$ with $t > 0$, (c) follows by using $\log(1+x) \leq x\log e$ for $x = \frac{t(c_1(\varepsilon,d)-c_2(\varepsilon,d))}{\mathbb{E}[\theta]c_1(\varepsilon,d)+tc_2(\varepsilon,d)} > 0$, (d) follows because $t > 0$, and (e) follows because $t = \theta - \mathbb{E}[\theta] \leq \hat{\rho}$.

Therefore, for θ such that $|\theta - \mathbb{E}[\theta]| \leq \hat{\rho}$, we have the following from (24) and (25):

$$D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right) \leq \frac{\log e \times \hat{\rho}\,(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{\min\left\{c_2(\varepsilon,d), \mathbb{E}[\theta]c_1(\varepsilon,d)\right\}} \overset{(a)}{=} \frac{\log e \times \hat{\rho}\,(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{\min\left\{c_2(\varepsilon,d), c_1(\varepsilon,d)\mathbb{P}_{\boldsymbol{z}\sim\mathrm{Unif}(\mathcal{Z})}\left(\boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}}\right)\right\}}$$

$$\overset{(b)}{\leq} \frac{2\log e \times \hat{\rho}\left(c_1(\varepsilon,d) - c_2(\varepsilon,d)\right)}{c_2(\varepsilon,d)}$$
$$\overset{(c)}{\leq} 2\log e \times \hat{\rho}\left(\exp(\varepsilon) - 1\right) \tag{26}$$

where (a) follows from (22), (b) follows because $\mathbb{P}_{\boldsymbol{z}\sim\mathrm{Unif}(\mathcal{Z})}\left(\boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}}\right) \geq c_2(\varepsilon,d)/2c_1(\varepsilon,d)$ from the definition of cap-based mechanisms, and (c) follows because $c_1(\varepsilon,d)/c_2(\varepsilon,d) \leq \exp(\varepsilon)$.

Using (26) and Lemma C.1 in (23), we have

$$\begin{aligned}\mathbb{E}_\theta\left[D_{\mathsf{KL}}\left(\pi^{\mathtt{mrc}}_{\boldsymbol{x},\theta}(\cdot)\,\middle\|\,\pi^{\mathtt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right)\right] &\leq \sum_{\theta:|\theta-\mathbb{E}[\theta]|\leq\hat{\rho}} \mathbb{P}(\theta) \times 2\log e \times \hat{\rho}\left(\exp(\varepsilon)-1\right) + \sum_{\theta:|\theta-\mathbb{E}[\theta]|>\hat{\rho}} \mathbb{P}(\theta) \times \varepsilon\log e \\ &\overset{(a)}{\leq} 2\log e \times \hat{\rho}\left(\exp(\varepsilon)-1\right) + \rho\varepsilon\log e \\ &\overset{(b)}{\leq} 2\log e \times \sqrt{\frac{\ln(2/\rho)}{2N}}\left(\exp(\varepsilon)-1\right) + \rho\varepsilon\log e \\ &\overset{(c)}{\leq} \log e \times \rho(1+\varepsilon)\end{aligned}$$

where (a) follows because $\mathbb{P}\left(|\theta - \mathbb{E}[\theta]| \leq \hat{\rho}\right) \leq 1$ and $\mathbb{P}\left(|\theta - \mathbb{E}[\theta]| \geq \hat{\rho}\right) \leq \rho$, (b) follows by plugging in $\hat{\rho} = \sqrt{\frac{\ln(2/\rho)}{2N}}$, and (c) follows by plugging in N. □

C.3 Utility of MMRC

In this section, we first prove Theorem 3.5 i.e., we show that, with number of candidates exponential in ε, samples drawn from $q^{\mathtt{mmrc}}$ will be similar to the samples drawn from $q^{\mathtt{mrc}}$ in terms of ℓ_2 error.

Then, in Theorem C.1, we show that MMRC can simulate any ε-LDP cap-based mechanism in a nearly lossless fashion with about ε bits of communication.

C.3.1 Utility of MMRC with respect to $q^{\mathtt{mrc}}$

Theorem 3.5 (Utility of MMRC)**.** *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP cap-based mechanism $q(\cdot|\boldsymbol{x})$. Let the reference distribution $p(\cdot)$ be the uniform distribution on $\mathcal{Z}$. Let N denote the number of candidates. Then, q^{mmrc} is such that*

$$\begin{aligned}\mathbb{E}_{q^{mmrc}}\left[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\right] &\leq \mathbb{E}_{q^{mrc}}\left[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\right] \\ &+\sqrt{\frac{\rho(1+\varepsilon)}{2}}\max_{\boldsymbol{x},\boldsymbol{z}}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\end{aligned} \tag{9}$$

where $\rho \in (0,1)$ is such that $N = \frac{2(\exp(\varepsilon)-1)^2}{\rho^2}\ln\frac{2}{\rho}$.

Proof. We will first upper bound the difference between $\mathbb{E}_{q^{\mathtt{mmrc}}}\left[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\right]$ and $\mathbb{E}_{q^{\mathtt{mrc}}}\left[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\right]$ in terms of the total variation distance between $q^{\mathtt{mrc}}$ and $q^{\mathtt{mmrc}}$. Due to a property of the total variation distance (e.g., Song et al., 2016), we have

$$\mathbb{E}_{q^{\mathtt{mmrc}}}\left[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\right] - \mathbb{E}_{q^{\mathtt{mrc}}}\left[\,\|\boldsymbol{z}-\boldsymbol{x}\|_2^2\,\right] \leq \max_{\boldsymbol{x},\boldsymbol{z}}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2 \times \|q^{\mathtt{mrc}}(\boldsymbol{z}|\boldsymbol{x}) - q^{\mathtt{mmrc}}(\boldsymbol{z}|\boldsymbol{x})\|_{\mathsf{TV}}. \tag{27}$$

Next, we will upper bound the total variation distance between $q^{\mathtt{mrc}}$ and $q^{\mathtt{mmrc}}$ using Pinsker's inequality as follows:

$$\|q^{\mathtt{mrc}}(\boldsymbol{z}|\boldsymbol{x}) - q^{\mathtt{mmrc}}(\boldsymbol{z}|\boldsymbol{x})\|_{\mathsf{TV}} \leq \sqrt{\frac{1}{2\log e}D_{\mathsf{KL}}\left(q^{\mathtt{mrc}}(\boldsymbol{z}|\boldsymbol{x})\,\|\,q^{\mathtt{mmrc}}(\boldsymbol{z}|\boldsymbol{x})\right)}. \tag{28}$$

Next, we will upper bound the KL divergence between $q^{\mathtt{mrc}}(\boldsymbol{z}|\boldsymbol{x})$ and $q^{\mathtt{mmrc}}(\boldsymbol{z}|\boldsymbol{x})$. To that end, for every $\boldsymbol{x} \in \mathcal{X}$, let $p^{\mathtt{mrc}}(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N,K,\boldsymbol{z}_K|\boldsymbol{x})$ denote the joint distribution of the candidates $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N$ drawn from $p(\mathbf{z})$, the transmitted index K under MRC, and the sample $\boldsymbol{z}_K$ corresponding to K. We have

$$p^{\mathtt{mrc}}(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N,K,\boldsymbol{z}_K|\boldsymbol{x}) = p(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N|\boldsymbol{x}) \times p^{\mathtt{mrc}}(K|\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N,\boldsymbol{x}) \times p^{\mathtt{mrc}}(\boldsymbol{z}_K|\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N,K,\boldsymbol{x})$$

$$
\begin{aligned}
&\overset{(a)}{=} p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times p^{\texttt{mrc}}(K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, \boldsymbol{x}) \times p^{\texttt{mrc}}(\boldsymbol{z}_K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{x}) \\
&\overset{(b)}{=} p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times \pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k) \times p^{\texttt{mrc}}(\boldsymbol{z}_K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{x}) \\
&\overset{(c)}{=} p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times \pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k)
\end{aligned} \tag{29}
$$

where (a) follows because $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ are independent of $\boldsymbol{x}$, (b) follows because $p^{\texttt{mrc}}(K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, \boldsymbol{x}) = \pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k)$, and (c) follows because $p^{\texttt{mrc}}(\boldsymbol{z}_K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{x}) = 1$ (note that $\boldsymbol{z}_K$ can be viewed as a function of $(\boldsymbol{z}_1, ..., \boldsymbol{z}_N, K)$).

Similarly, for every $\boldsymbol{x} \in \mathcal{X}$, let $p^{\texttt{mmrc}}(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{z}_K | \boldsymbol{x})$ denote the joint distribution of the candidates $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ drawn from $p(\mathbf{z})$, the transmitted index K under MMRC, and the sample $\boldsymbol{z}_K$ corresponding to K. We have

$$
\begin{aligned}
p^{\texttt{mmrc}}(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{z}_K | \boldsymbol{x}) &= p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N | \boldsymbol{x}) \times p^{\texttt{mmrc}}(K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, \boldsymbol{x}) \times p^{\texttt{mmrc}}(\boldsymbol{z}_K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{x}) \\
&\overset{(a)}{=} p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times p^{\texttt{mmrc}}(K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, \boldsymbol{x}) \times p^{\texttt{mmrc}}(\boldsymbol{z}_K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{x}) \\
&\overset{(b)}{=} p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times \pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k) \times p^{\texttt{mmrc}}(\boldsymbol{z}_K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{x}) \\
&\overset{(c)}{=} p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times \pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)
\end{aligned} \tag{30}
$$

where (a) follows because $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ are independent of $\boldsymbol{x}$, (b) follows because $p^{\texttt{mmrc}}(K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, \boldsymbol{x}) = \pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)$, and (c) follows because $p^{\texttt{mmrc}}(\boldsymbol{z}_K | \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{x}) = 1$.

We are now in a position to upper bound the KL divergence between $q^{\texttt{mrc}}(\boldsymbol{z}_K | \boldsymbol{x})$ and $q^{\texttt{mmrc}}(\boldsymbol{z}_K | \boldsymbol{x})$:

$$
\begin{aligned}
D_{\mathsf{KL}}\left(q^{\texttt{mrc}}(\boldsymbol{z}|\boldsymbol{x}) \,\|q^{\texttt{mmrc}}(\boldsymbol{z}|\boldsymbol{x})\right) &\overset{(a)}{\leq} D_{\mathsf{KL}}\left(p^{\texttt{mrc}}(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{z}_K | \boldsymbol{x}) \,\|p^{\texttt{mmrc}}(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N, K, \boldsymbol{z}_K | \boldsymbol{x})\right) \\
&\overset{(b)}{=} D_{\mathsf{KL}}\left(p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times \pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k) \,\middle\|\, p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \times \pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)\right) \\
&\overset{(c)}{=} \mathbb{E}_{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N}\left[D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(k) \,\middle\|\, \pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(k)\right)\right] \\
&\overset{(d)}{=} \mathbb{E}_{\theta}\left[D_{\mathsf{KL}}\left(\pi^{\texttt{mrc}}_{\boldsymbol{x},\theta}(\cdot) \,\middle\|\, \pi^{\texttt{mmrc}}_{\boldsymbol{x},\theta}(\cdot)\right)\right] \overset{(e)}{\leq} \rho \times \log e \times (1 + \varepsilon)
\end{aligned} \tag{31}
$$

where (a) follows because by the chain rule for KL-divergence, (b) follows from (29) and (30), (c) follows by the definition of KL-divergence, (d) follows because $\pi^{\texttt{mrc}}$ and $\pi^{\texttt{mmrc}}$ depend on $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ only via θ for cap-based mechanisms, and (e) follows from Lemma C.2 because $N = \frac{2(\exp(\varepsilon)-1)^2}{\rho^2} \ln \frac{2}{\rho}$. Combining (27), (28),and (31), we have

$$
\mathbb{E}_{q^{\texttt{mmrc}}}\left[\, \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 \,\right] \leq \mathbb{E}_{q^{\texttt{mrc}}}\left[\, \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 \,\right] + \sqrt{\frac{\rho(1+\varepsilon)}{2}} \times \max_{\boldsymbol{x},\boldsymbol{z}} \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 .
$$

□

Remark C.1. *For bounded ε-LDP mechanisms such as* PrivUnit$_2$ *and* Subset Selection*, the term $\max_{\boldsymbol{x},\boldsymbol{z}} \|\boldsymbol{z} - \boldsymbol{x}\|_2^2$ in (9) is of the same order as $\mathbb{E}_q\left[\|\boldsymbol{z} - \boldsymbol{x}\|^2\right]$. Therefore, by picking a large N in Theorem 3.5 (i.e. $\log N \geq C\varepsilon$ for a sufficiently large C), ρ can be made arbitrarily small and the estimation error of* MMRC *can be arbitrarily close to the estimation error of* MRC*.*

C.3.2 Utility of MMRC with respect to q

Theorem C.1. *Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $\boldsymbol{x} \in \mathcal{X}$, and ε-LDP cap-based mechanism $q(\cdot|\boldsymbol{x})$. Let the reference distribution $p(\cdot)$ be the uniform distribution on $\mathcal{Z}$. Let N denote the number of candidates. Then, q^{mmrc} is such that*

$$
\mathbb{E}_{q^{mmrc}}\left[\, \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 \,\right] \leq \mathbb{E}_{q}\left[\, \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 \,\right] + \sqrt{\frac{\rho(1+\varepsilon)}{2}} \times \max_{\boldsymbol{x},\boldsymbol{z}} \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 + \frac{2\alpha}{1-\alpha} \times \sqrt{\mathbb{E}_q[\|\boldsymbol{z} - \boldsymbol{x}\|^4]}
$$

holds with probability at least $1 - 2\alpha$ where

$$
\alpha = \sqrt{2^{-c\varepsilon} + 2^{-c^2/\log e + 1}}.
$$

and c and $\rho \in (0,1)$ are free variables such that

$$N = \max\left\{ 2^{(\log e + 4c)\varepsilon}, \frac{2\left(\exp(\varepsilon)-1\right)^2}{\rho^2} \ln\frac{2}{\rho} \right\}$$

Proof. The proof follows from Theorem 3.1 and Theorem 3.5. □

D PRELIMINARY ON $\texttt{PrivUnit}_2$

First, we briefly recap the $\texttt{PrivUnit}_2$ mechanism ($q^{\texttt{pu}}$) proposed in Bhowmick et al. (2018). $\texttt{PrivUnit}_2$ is a private sampling scheme when the input alphabet $\mathcal{X}$ is the $d-$dimensional unit ℓ_2 sphere $\mathbb{S}^{d-1}$. More formally, given a vector $\boldsymbol{x} \in \mathbb{S}^{d-1}$, $\texttt{PrivUnit}_2$ (see Algorithm 3) draws a vector $\boldsymbol{z}$ from a spherical cap $\{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x}\rangle \geq \gamma\}$ with probability $p_0 \geq 1/2$ or from its complement $\{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x}\rangle < \gamma\}$ with probability $1-p_0$, where $\gamma \in [0,1]$ and p_0 are constants that trade accuracy and privacy. In other words, the conditional density $q^{\texttt{pu}}(\boldsymbol{z}|\boldsymbol{x})$ is:

$$q^{\texttt{pu}}(\mathbf{z}|\mathbf{x}) = \begin{cases} p_0 \times \dfrac{2}{A(1,d) I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2})} & \text{if } \langle \mathbf{x}, \mathbf{z}\rangle \geq \gamma \\ (1-p_0) \times \dfrac{2}{2A(1,d) - A(1,d) I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2})} & \text{otherwise} \end{cases} \tag{32}$$

where $A(1,d)$ denotes the area of $\mathbb{S}^{d-1}$ and $I_x(a,b)$ denotes the regularized incomplete beta function.

```
Algorithm 3: Privatized Unit Vector: $\texttt{PrivUnit}_2$
Require: $\boldsymbol{x} \in \mathbb{S}^{d-1}$, $\gamma \in [0,1]$, $p_0 \geq 1/2$.
  Draw random vector $\boldsymbol{z}$ according to the distribution
      $\boldsymbol{z} = \begin{cases} \text{uniform on } \{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x}\rangle \geq \gamma\} & \text{with probability } p_0 \\ \text{uniform on } \{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x}\rangle < \gamma\} & \text{otherwise.} \end{cases}$
  Set $\alpha = \frac{d-1}{2}$, $\tau = \frac{1+\gamma}{2}$, and
      $m_{\text{pu}} = \frac{(1-\gamma^2)^\alpha}{2^{d-2}(d-1)} \left[ \frac{p_0}{B(\alpha,\alpha) - B(\tau;\alpha,\alpha)} - \frac{1-p_0}{B(\tau;\alpha,\alpha)} \right]$      (33)
  return $\hat{\boldsymbol{x}}^{\texttt{pu}} = \frac{\boldsymbol{z}}{m_{\text{pu}}}$
```

Given its inputs $\boldsymbol{x}, \gamma$, and p_0, Algorithm 3 returns an estimator $\hat{\boldsymbol{x}}^{\texttt{pu}} := \boldsymbol{z}/m_{\texttt{pu}}$ which is differentially private and unbiased where $m_{\texttt{pu}}$ is a scaling factor. The choice of γ described in Theorem D.1 ensures differential privacy and the choice of the scaling factor m described in (33) ensures unbiasedness where

$$B(x;\alpha,\beta) := \int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt \quad \text{where} \quad B(\alpha,\beta) := B(1;\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

denotes the incomplete beta function.

D.1 $\texttt{PrivUnit}_2$ is a differentially private mechanism

The following theorem borrowed from Bhowmick et al. (2018) describes the choice of γ and provides the precise associated differential privacy guarantee of the $\texttt{PrivUnit}_2$ mechanism.

Theorem D.1 (Bhowmick et al. (2018))**.** *Let $\gamma \in [0,1]$ and $p_0 = \frac{e^{\varepsilon_0}}{1+e^{\varepsilon_0}}$. Then algorithm $\texttt{PrivUnit}_2(\cdot, \gamma, p_0)$ is*

$\varepsilon = (\bar{\varepsilon} + \varepsilon_0)$-differentially private whenever $\gamma \geq 0$ is such that

$$\begin{aligned} &\bar{\varepsilon} \geq \log \frac{1 + \gamma \cdot \sqrt{2(d-1)/\pi}}{\left(1 - \gamma \cdot \sqrt{2(d-1)/\pi}\right)_+}, \quad \textit{i.e.} \quad \gamma \leq \frac{e^{\bar{\varepsilon}} - 1}{e^{\bar{\varepsilon}} + 1}\sqrt{\frac{\pi}{2(d-1)}}, \\ &\textit{or} \\ &\bar{\varepsilon} \geq 1/2 \log(d) + \log 6 - \frac{d-1}{2}\log(1-\gamma^2) + \log \gamma \quad \textit{and} \quad \gamma \geq \sqrt{\frac{2}{d}}. \end{aligned} \tag{34}$$

Here, ε can be viewed as the total privacy budget. Typically, μ fraction of this budget is allocated for the spherical cap threshold γ and $1 - \mu$ fraction is allocated to the probability parameter p_0 with which a particular spherical cap is chosen i.e., $\bar{\varepsilon} = \mu\varepsilon$ and $\varepsilon_0 = (1 - \mu)\varepsilon$ for some $\mu \in [0, 1]$. While the parameter μ can be optimized over as described in Feldman and Talwar (2021), we will view it as a constant for convenience. Our results on MRC and MMRC simulating $\texttt{PrivUnit}_2$ can be easily extended to the setup where μ needs to be optimized over.

D.2 $\texttt{PrivUnit}_2$ is unbiased and order-optimal

The following lemma borrowed from Bhowmick et al. (2018) shows that the output of the $\texttt{PrivUnit}_2$ mechanism (a) is unbiased, (b) has a bounded norm, and (c) has order-optimal utility.

Proposition D.1 (Bhowmick et al. (2018))**.** *Let $\hat{\boldsymbol{x}}^{pu} = \texttt{PrivUnit}_2(\boldsymbol{x}, \gamma, p_0)$ for some $\boldsymbol{x} \in \mathbb{S}^{d-1}$, $\gamma \in [0, 1]$, and $p_0 \in [1/2, 1]$. Then, $\mathbb{E}[\hat{\boldsymbol{x}}^{pu}] = \boldsymbol{x}$. Further, assume that $0 \leq \varepsilon \leq d$. Then, there exists a numerical constant $c < \infty$ such that if γ saturates either of the two inequalities (34), then $\gamma \gtrsim \min\{\varepsilon/\sqrt{d}, \sqrt{\varepsilon/d}\}$, and*

$$\|\hat{\boldsymbol{x}}^{pu}\|_2 \leq c \cdot \sqrt{\frac{d}{\varepsilon} \vee \frac{d}{(e^{\varepsilon} - 1)^2}}.$$

Additionally, $\mathbb{E}[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|_2^2] \lesssim \frac{d}{\varepsilon} \vee \frac{d}{(e^{\varepsilon}-1)^2}$.

D.3 $\texttt{PrivUnit}_2$ is a cap-based mechanism

The randomness in the estimator $\hat{\boldsymbol{x}}^{\text{pu}}$ obtained from the $\texttt{PrivUnit}_2(\boldsymbol{x}, \gamma, p_0)$ mechanism comes from $\boldsymbol{z}$. Therefore, we obtain a convenient expression for the conditional distribution of $\boldsymbol{z}$ conditioned on $\boldsymbol{x}$ i.e., $q^{\text{pu}}(\boldsymbol{z}|\boldsymbol{x})$. Define $\mathsf{Cap}_{\boldsymbol{x}} \coloneqq \{\boldsymbol{z} | \langle \boldsymbol{x}, \boldsymbol{z} \rangle \geq \gamma\}$. Recall from (34) that γ is a function of ε and d. Further, as described in Section D.1, when the budget split parameter μ is known, p_0 can viewed as a function of ε. Then, the conditional distribution $q^{\text{pu}}(\boldsymbol{z}|\boldsymbol{x})$ in (32) can be written as follows:

$$q^{\text{pu}}(\boldsymbol{z}|\boldsymbol{x}) = \begin{cases} c_1(\varepsilon, d) & \text{if } \boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}} \\ \\ c_2(\varepsilon, d) & \text{if } \boldsymbol{z} \notin \mathsf{Cap}_{\boldsymbol{x}} \end{cases} \tag{35}$$

where $c_1(\varepsilon, d) = p_0 \times \dfrac{2}{A(1,d) I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2})}$ and $c_2(\varepsilon, d) = (1 - p_0) \times \dfrac{2}{2A(1,d) - A(1,d) I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2})}$ are functions of ε and d.

Further, $\mathbb{P}_{\boldsymbol{z} \sim \text{Unif}(\mathcal{Z})}\left(\boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{x}}\right) = \frac{I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2})}{2}$. Therefore,

$$\frac{c_1(\varepsilon, d)}{c_2(\varepsilon, d)} \times \mathbb{P}_{\boldsymbol{z} \sim \text{Unif}(\mathcal{Z})}\left(\boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{x}}\right) = \frac{p_0 \times (2 - I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2}))}{2(1-p_0)} \overset{(a)}{\geq} \frac{p_0}{2(1-p_0)} \overset{(b)}{\geq} \frac{1}{2}$$

where (a) follows because $I_{1-\gamma^2}(\frac{d-1}{2}, \frac{1}{2}) \leq 1$ and (b) follows because $p_0 \geq 1/2$.

E SIMULATING $\texttt{PrivUnit}_2$ USING MINIMAL RANDOM CODING

In this section, we simulate $\texttt{PrivUnit}_2$ using MRC analogous to how we simulate $\texttt{PrivUnit}_2$ using MMRC in Section 4. First, in Appendix E.1, we provide an unbiased estimator for MRC simulating $\texttt{PrivUnit}_2$. Next, in Appendix

E.2 we provide the utility guarantee associated with MRC simulating $\texttt{PrivUnit}_2$. To do that, first, in Appendix E.2.1, we show that when the number of candidates N is exponential in ε, the scaling factor $m_{\texttt{mrc}}$ is close to the scaling parameter associated with $\texttt{PrivUnit}_2$ (i.e., $m_{\texttt{pu}}$). Next, in Appendix E.2.2, we provide the relationship between the mean squared error associated with MRC simulating $\texttt{PrivUnit}_2$ and the mean squared error associated with $\texttt{PrivUnit}_2$. In Appendix E.2.3, we combine everything and show that, for mean estimation, MRC can simulate $\texttt{PrivUnit}_2$ in a near-lossless manner while only using on the order of ε-bits of communication. Finally, in Appendix E.3, we provide some empirical comparisons.

E.1 Unbiased Minimal Random Coding simulating $\texttt{PrivUnit}_2$

Consider the $\texttt{PrivUnit}_2$ ε-LDP mechanism $q^{\texttt{pu}}$ described in Section 2 with parameters p_0 and γ. $\texttt{PrivUnit}_2$ is a cap-based mechanism with $\mathsf{Cap}_{\boldsymbol{x}} = \{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x}\rangle \geq \gamma\}$ as discussed in Appendix D. Let $\pi^{\texttt{mrc}}$ be the distribution and $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_N$ be the candidates obtained from Algorithm 1 when the reference distribution is $\text{Unif}(\mathbb{S}^{d-1})$. Let $K \sim \pi^{\texttt{mrc}}(\cdot)$. Define $p_{\texttt{mrc}} \coloneqq \mathbb{P}(\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}})$ to be the probability with which the sampled candidate $\boldsymbol{z}_K$ belongs to the spherical cap associated with $\texttt{PrivUnit}_2$. Define $m_{\texttt{mrc}}$ as the scaling factor in (2) when p_0 in (2) is replaced by $p_{\texttt{mrc}}$. Define $\hat{\boldsymbol{x}}^{\texttt{mrc}} \coloneqq \boldsymbol{z}_K / m_{\texttt{mrc}}$ as the estimator of the MRC mechanism simulating $\texttt{PrivUnit}_2$. The following Lemma shows that $\hat{\boldsymbol{x}}^{\texttt{mrc}}$ is an unbiased estimator.

Lemma E.1. *Let $\hat{\boldsymbol{x}}^{mrc}$ be the estimator of the* MRC *mechanism simulating* $\texttt{PrivUnit}_2$ *as defined above. Then,* $\mathbb{E}_{q^{mrc}}[\hat{\boldsymbol{x}}^{mrc}] = \boldsymbol{x}$.

Proof. For $k \in [N]$, let $A_k \coloneqq \mathbb{1}(\boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}})$. Then, $p_{\texttt{mrc}} = \mathbb{P}(A_K = 1)$. Using the definition of $\hat{\boldsymbol{x}}^{\texttt{mrc}}$, we have

$$\mathbb{E}_{q^{\texttt{mrc}}}[\hat{\boldsymbol{x}}^{\texttt{mrc}}] = \frac{1}{m_{\texttt{mrc}}} \mathbb{E}_{q^{\texttt{mrc}}}[\boldsymbol{z}_K].$$

Let us evaluate $\mathbb{E}_{q^{\texttt{mrc}}}[\boldsymbol{z}_K]$. We have

$$\begin{aligned}
\mathbb{E}_{q^{\texttt{mrc}}}[\boldsymbol{z}_K] &\overset{(a)}{=} \mathbb{E}_{K, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N}[\boldsymbol{z}_K] \\
&\overset{(b)}{=} \mathbb{E}_{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N}\left[\sum_{k=1}^{N} \pi^{\texttt{mrc}}_{\boldsymbol{x}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N}(k) \times \boldsymbol{z}_k\right] \\
&\overset{(c)}{=} \mathbb{E}_{A_1, \cdots, A_N}\left[\mathbb{E}_{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N}\left[\sum_{k=1}^{N} \pi^{\texttt{mrc}}_{\boldsymbol{x}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N}(k) \times \boldsymbol{z}_k \middle| A_1, \cdots, A_N\right]\right] \\
&\overset{(d)}{=} \sum_{k=1}^{N} \mathbb{E}_{A_1, \cdots, A_N}\left[\mathbb{E}_{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N}(k) \times \boldsymbol{z}_k \middle| A_1, \cdots, A_N\right]\right] \\
&\overset{(e)}{=} \sum_{k=1}^{N} \mathbb{E}_{A_1, \cdots, A_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \ldots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N}\left[\boldsymbol{z}_k \middle| A_1, \cdots, A_N\right]\right] \\
&\overset{(f)}{=} \sum_{k=1}^{N} \mathbb{E}_{A_1, \cdots, A_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \ldots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N}\left[\boldsymbol{z}_k \middle| A_k\right]\right] \\
&\overset{(g)}{=} \sum_{k=1}^{N} \mathbb{E}_{A_1, \cdots, A_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \ldots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_k}\left[\boldsymbol{z}_k \middle| A_k\right]\right] \\
&\overset{(h)}{=} \sum_{k=1}^{N} \mathbb{E}_{A_k}\left[\mathbb{E}_{A_1, \cdots, A_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \ldots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_k}\left[\boldsymbol{z}_k \middle| A_k\right] \middle| A_k\right]\right] \\
&\overset{(i)}{=} \sum_{k=1}^{N} \mathbb{P}(A_k = 1)\left[\mathbb{E}_{A_1, \cdots, A_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \ldots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_k}\left[\boldsymbol{z}_k \middle| A_k\right] \middle| A_k = 1\right]\right] \\
&\quad + \sum_{k=1}^{N} \mathbb{P}(A_k = 0)\left[\mathbb{E}_{A_1, \cdots, A_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \ldots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_k}\left[\boldsymbol{z}_k \middle| A_k\right] \middle| A_k = 0\right]\right] \\
&= \sum_{k=1}^{N} \mathbb{P}(A_k = 1)\left[\mathbb{E}_{A_1, \cdots, A_N}\left[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \cdots, A_k = 1, \cdots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_k}\left[\boldsymbol{z}_k \middle| A_k = 1\right]\right]\right]
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{k=1}^{N} \mathbb{P}(A_k = 0) \Big[\mathbb{E}_{A_1, \cdots, A_N} \Big[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \cdots, A_k = 0, \cdots, A_N}(k) \mathbb{E}_{\boldsymbol{z}_k} \big[\boldsymbol{z}_k \big| A_k = 0 \big] \Big] \Big] \\
&\overset{(j)}{=} \mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 1 \big] \sum_{k=1}^{N} \mathbb{P}(A_k = 1) \Big[\mathbb{E}_{A_1, \cdots, A_N} \Big[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \cdots, A_k = 1, \cdots, A_N}(k) \Big] \Big] \\
&+ \mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 0 \big] \sum_{k=1}^{N} \mathbb{P}(A_k = 0) \Big[\mathbb{E}_{A_1, \cdots, A_N} \Big[\pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \cdots, A_k = 0, \cdots, A_N}(k) \Big] \Big] \\
&\overset{(k)}{=} \mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 1 \big] \sum_{k=1}^{N} \mathbb{P}(A_k = 1) \pi^{\texttt{mrc}}_{\boldsymbol{x}, A_k = 1}(k) \\
&+ \mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 0 \big] \sum_{k=1}^{N} \mathbb{P}(A_k = 0) \pi^{\texttt{mrc}}_{\boldsymbol{x}, A_k = 0}(k) \\
&\overset{(l)}{=} \mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 1 \big] \mathbb{P}(A_K = 1) + \mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 0 \big] \mathbb{P}(A_K = 0) \\
&\overset{(m)}{=} m_{\texttt{mrc}} \boldsymbol{x} \qquad (36)
\end{aligned}
$$

where (a) follows because the randomness in $q^{\texttt{mrc}}$ comes from the randomness in $K, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$, (b) follows by calculating the expectation over K and showing the dependence of $\pi^{\texttt{mrc}}$ on $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ explicitly, (c) follows by the tower property of expectation, (d) follows by linearity of expectation, (e) follows because $\pi^{\texttt{mrc}}_{\boldsymbol{x}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N}(k) = \pi^{\texttt{mrc}}_{\boldsymbol{x}, A_1, \ldots, A_N}(k)$ since $\pi^{\texttt{mrc}}$ depends on $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ via $A_1, \cdots, A_N$, (f) follows because $\boldsymbol{z}_k$ is independent of $A_1, \cdots, A_{k-1}, A_{k+1}, \cdots, A_N$ given A_k, (g) follows by marginalizing $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_{k-1}, \boldsymbol{z}_{k+1}, \cdots, \boldsymbol{z}_N$, (h) follows by the tower property of expectation, (i) follows by evaluating the expectation over A_k, (j) follows because $\mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 1 \big] := \mathbb{E}_{\boldsymbol{z}_k} \big[\boldsymbol{z}_k \big| A_k = 1 \big]$ and $\mathbb{E}_{\boldsymbol{z}} \big[\boldsymbol{z} \big| A = 0 \big] := \mathbb{E}_{\boldsymbol{z}_k} \big[\boldsymbol{z}_k \big| A_k = 0 \big]$ are constants for every $k \in [N]$, (k) follows by marginalizing $A_1, \cdots, A_N$, (l) follows from the definitions of $\mathbb{P}(A_K = 1)$ and $\mathbb{P}(A_K = 0)$, and (m) follows from rotational symmetry (see the proof of Lemma 4.1 in Bhowmick et al. (2018) for details). Therefore, we can write

$$
\mathbb{E}_{q^{\texttt{mrc}}} [\hat{\boldsymbol{x}}^{\texttt{mrc}}] = \frac{1}{m_{\texttt{mrc}}} \mathbb{E}_{q^{\texttt{mrc}}} [\boldsymbol{z}_K] \overset{(a)}{=} \boldsymbol{x}
$$

where (a) follows from (36). □

E.2 Utility of Minimal Random Coding simulating $\texttt{PrivUnit}_2$

E.2.1 The scaling factors of $\texttt{PrivUnit}_2$ and $\texttt{MRC}$ are close when N is of the right order

In the following Lemma, we show that when the number of candidates N is exponential in ε, then the scaling parameters associated with $\texttt{PrivUnit}_2$ and the $\texttt{MRC}$ scheme simulating $\texttt{PrivUnit}_2$ are close.

Lemma E.2. *Let N denote the number of candidates used in the* $\texttt{MRC}$ *scheme. Let $K \sim \pi^{mrc}$ where π^{mrc} is the distribution over the indices $[N]$ associated the* $\texttt{MRC}$ *scheme simulating* $\texttt{PrivUnit}_2(\boldsymbol{x}, \gamma, p_0)$*. Consider any $\lambda > 0$. Then, the scaling factor m_{pu} associated with* $\texttt{PrivUnit}_2$ *and the scaling factor m_{mrc} associated with the* $\texttt{MRC}$ *scheme simulating* $\texttt{PrivUnit}_2$ *are such that*

$$
m_{pu} - m_{mrc} \leq \lambda \cdot m_{mrc}
$$

as long as

$$
N \geq 2e^{2\varepsilon} \left(\frac{2(1+\lambda)}{\lambda \left(p_0 - 1/2 \right)} \right)^2 \ln \left(\frac{4(1+\lambda)}{\lambda \left(p_0 - 1/2 \right)} \right).
$$

Proof. Following the proofs of Lemma 4.1 and Proposition 4 in Bhowmick et al. (2018), we can write $m_{\texttt{pu}} = \gamma_+ p_0 + \gamma_- (1 - p_0)$ and $m_{\texttt{mrc}} = \gamma_+ p_{\texttt{mrc}} + \gamma_- (1 - p_{\texttt{mrc}})$ where

$$
\gamma_+ \triangleq \frac{(1-\gamma^2)^\alpha}{2^{d-2}(d-1) \left(B(\alpha, \alpha) - B(\tau; \alpha, \alpha) \right)}, \qquad \text{and} \qquad \gamma_- \triangleq \frac{(1-\gamma^2)^\alpha}{2^{d-2}(d-1) \left(B(\tau; \alpha, \alpha) \right)}.
$$

Therefore, we have

$$\frac{1}{m_{\texttt{mrc}}} - \frac{1}{m_{\texttt{pu}}} = \frac{m_{\texttt{pu}} - m_{\texttt{mrc}}}{m_{\texttt{pu}} \cdot m_{\texttt{mrc}}} = \frac{1}{m_{\texttt{pu}}}\left(\frac{(\gamma_+ - \gamma_-)\cdot(p_0 - p_{\texttt{mrc}})}{((\gamma_+ - \gamma_-)p_{\texttt{mrc}} + \gamma_-)}\right) = \frac{1}{m_{\texttt{pu}}}\left(\frac{p_0 - p_{\texttt{mrc}}}{p_{\texttt{mrc}} + \dfrac{\gamma_-}{\gamma_+ - \gamma_-}}\right) \tag{37}$$

From Bhowmick et al. (2018), we have $\gamma_- \leq 0 \leq \gamma_+$ and $|\gamma_+| \geq |\gamma_-|$. These inequalities imply $\frac{\gamma_-}{\gamma_+ - \gamma_-} \geq -\frac{1}{2}$. Plugging this in (37), we have

$$\frac{1}{m_{\texttt{mrc}}} - \frac{1}{m_{\texttt{pu}}} \leq \frac{1}{m_{\texttt{pu}}}\left(\frac{p_0 - p_{\texttt{mrc}}}{p_{\texttt{mrc}} - 1/2}\right) = \frac{1}{m_{\texttt{pu}}}\left(\frac{1}{\dfrac{p_0 - 1/2}{p_0 - p_{\texttt{mrc}}} - 1}\right) \tag{38}$$

We will now upper bound $\dfrac{p_0 - p_{\texttt{mrc}}}{p_0 - 1/2}$. We start by obtaining convenient expressions for $p_{\texttt{mrc}}$ and p_0. To compute $p_{\texttt{mrc}} = \mathbb{P}(\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}})$, recall that θ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$. Let $c_1(\varepsilon, d)$ and $c_2(\varepsilon, d)$ be as defined in (35). Let $\bar{c}_1(\varepsilon, d) = c_1(\varepsilon, d) \times A(1, d)$ and $\bar{c}_2(\varepsilon, d) = c_2(\varepsilon, d) \times A(1, d)$. It is easy to see from Algorithm 3 and (35) that $\mathbb{P}(\boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}}) = \bar{c}_1(\varepsilon, d)/p_0$. Further, since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\theta \sim \frac{1}{N}\mathsf{Binom}\left(N, \frac{\bar{c}_1(\varepsilon, d)}{p_0}\right),$$

so we have

$$\begin{aligned}
p_{\texttt{mrc}} = \mathbb{P}\left\{\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}}\right\} &= \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}} | \theta\right\}\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\frac{\bar{c}_1(\varepsilon, d)\theta}{\bar{c}_1(\varepsilon, d)\theta + \bar{c}_2(\varepsilon, d)(1-\theta)}\right] \\
&= \frac{\bar{c}_1(\varepsilon, d)}{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}\mathbb{E}\left[\frac{(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d))\theta}{(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d))\theta + \bar{c}_2(\varepsilon, d)}\right] \\
&\overset{(b)}{=} \frac{\bar{c}_1(\varepsilon, d)\bar{c}_2(\varepsilon, d)}{(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d))^2}\mathbb{E}\left[\frac{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}{\bar{c}_2(\varepsilon, d)} - \frac{1}{\theta + \dfrac{\bar{c}_2(\varepsilon, d)}{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}}\right]
\end{aligned} \tag{39}$$

where (a) follows from (7) because q^{pu} is a cap-based mechanism and (b) follows by simple manipulations.

To compute p_0, observe that we have the following relationship between $\bar{c}_1(\varepsilon, d)$, $\bar{c}_2(\varepsilon, d)$, and p_0 from (35):

$$\frac{p_0}{\bar{c}_1(\varepsilon, d)} + \frac{1 - p_0}{\bar{c}_2(\varepsilon, d)} = 1$$

Using this and with some simple manipulations, we have

$$p_0 = \frac{\bar{c}_1(\varepsilon, d)\bar{c}_2(\varepsilon, d)}{(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d))^2}\left(\frac{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}{\bar{c}_2(\varepsilon, d)} - \frac{1}{\mathbb{E}\left[\theta\right] + \dfrac{\bar{c}_2(\varepsilon, d)}{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}}\right) \tag{40}$$

From (39) and (40), we have

$$\begin{aligned}
p_0 - p_{\texttt{mrc}} &= \frac{\bar{c}_1(\varepsilon, d)\bar{c}_2(\varepsilon, d)}{(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d))^2}\left(\mathbb{E}\left[\frac{1}{\theta + \dfrac{\bar{c}_2(\varepsilon, d)}{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}} - \frac{1}{\mathbb{E}[\theta] + \dfrac{\bar{c}_2(\varepsilon, d)}{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}}\right]\right) \\
&= \frac{\bar{c}_1(\varepsilon, d)\bar{c}_2(\varepsilon, d)}{(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d))^2}\left(\mathbb{E}\left[\frac{\mathbb{E}[\theta] - \theta}{\left(\theta + \dfrac{\bar{c}_2(\varepsilon, d)}{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}\right)\left(\mathbb{E}[\theta] + \dfrac{\bar{c}_2(\varepsilon, d)}{\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}\right)}\right]\right).
\end{aligned}$$

Now, using the Hoeffding's inequality, we have $\mathbb{P}\left\{|\theta - \mathbb{E}[\theta]| \geq \sqrt{\frac{\ln(2/\beta)}{2N}}\right\} \leq \beta$. Conditioned on the event $\left\{|\theta - \mathbb{E}[\theta]| \leq \sqrt{\frac{\ln(2/\beta)}{2N}}\right\}$ and using the fact that $|p_0 - p_{\texttt{mrc}}| \leq 1$, we have

$$p_0 - p_{\texttt{mrc}} \leq \frac{\bar{c}_1(\varepsilon,d)\bar{c}_2(\varepsilon,d)}{(\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d))^2}\left(\frac{\sqrt{\frac{\ln(2/\beta)}{2N}}}{\left(\frac{p_0}{\bar{c}_1(\varepsilon,d)} - \sqrt{\frac{\ln(2/\beta)}{2N}} + \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)}\right)\left(\frac{p_0}{\bar{c}_1(\varepsilon,d)} + \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)}\right)}\right) + \beta \tag{41}$$

where we have also plugged in $\mathbb{E}[\theta] = \dfrac{p_0}{\bar{c}_1(\varepsilon,d)}$. Now, we can lower bound $\left(\dfrac{p_0}{\bar{c}_1(\varepsilon,d)} + \dfrac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)}\right)$ as follows:

$$\left(\frac{p_0}{\bar{c}_1(\varepsilon,d)} + \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)}\right) \overset{(a)}{\geq} \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)} \overset{(b)}{\geq} \frac{1}{\exp(\varepsilon)-1}$$

where (a) follows by lower bounding $p_0/\bar{c}_1(\varepsilon,d)$ by 0 and (b) follows because we have $\bar{c}_1(\varepsilon,d)/\bar{c}_2(\varepsilon,d) \leq \exp(\varepsilon)$. Further, if we pick $N \geq 2\ln(2/\beta)(\exp(\varepsilon)-1)^2$, then

$$\sqrt{\frac{\ln(2/\beta)}{2N}} \leq \frac{1}{2} \times \frac{1}{\exp(\varepsilon)-1} \leq \frac{1}{2}\left(\frac{p_0}{\bar{c}_1(\varepsilon,d)} + \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)}\right). \tag{42}$$

Using (42) in (41), we have

$$\begin{aligned} p_0 - p_{\texttt{mrc}} &\leq \frac{\bar{c}_1(\varepsilon,d)\bar{c}_2(\varepsilon,d)}{(\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d))^2}\left(\frac{2\sqrt{\frac{\ln(2/\beta)}{2N}}}{\left(\frac{p_0}{\bar{c}_1(\varepsilon,d)} + \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)}\right)\left(\frac{p_0}{\bar{c}_1(\varepsilon,d)} + \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)-\bar{c}_2(\varepsilon,d)}\right)}\right) + \beta \\ &= \left(\frac{2\bar{c}_1(\varepsilon,d)\bar{c}_2(\varepsilon,d)\sqrt{\frac{\ln(2/\beta)}{2N}}}{\left(p_0\left(1-\frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)}\right)+\bar{c}_2(\varepsilon,d)\right)^2}\right) + \beta \\ &\overset{(a)}{\leq} \left(\frac{2\bar{c}_1(\varepsilon,d)}{\bar{c}_2(\varepsilon,d)}\sqrt{\frac{\ln(2/\beta)}{2N}}\right) + \beta \overset{(b)}{\leq} \left(2\exp(\varepsilon)\sqrt{\frac{\ln(2/\beta)}{2N}}\right) + \beta \overset{(c)}{\leq} \frac{\lambda(p_0-1/2)}{1+\lambda}. \end{aligned} \tag{43}$$

where (a) follows because $p_0\left(1-\frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d)}\right) \geq 0$, (b) follows because we have $\bar{c}_1(\varepsilon,d)/\bar{c}_2(\varepsilon,d) \leq \exp(\varepsilon)$ and (c) follows if we pick

$$\beta \leq \frac{\lambda(p_0-1/2)}{2(1+\lambda)} \qquad \text{and} \qquad N \geq \frac{2\exp(2\varepsilon)\ln(2/\beta)}{\left(\frac{\lambda(p_0-1/2)}{1+\lambda}-\beta\right)^2} = 2\exp(2\varepsilon)\left(\frac{2(1+\lambda)}{\lambda(p_0-1/2)}\right)^2\ln\left(\frac{4(1+\lambda)}{\lambda(p_0-1/2)}\right). \tag{44}$$

Further, it is easy to verify that (42) holds since the choice of N in (44) is such that $N \geq \frac{1}{2}\ln(2/\beta)(\exp(\varepsilon)-1)^2$. Now, rearranging (43) gives us an upper bound on $\dfrac{p_0 - p_{\texttt{mrc}}}{p_0 - 1/2}$, i.e.,

$$\frac{p_0 - p_{\texttt{mrc}}}{p_0 - 1/2} \leq \frac{\lambda}{1+\lambda}. \tag{45}$$

Using (45) in (38), we have

$$\frac{1}{m_{\texttt{mrc}}} - \frac{1}{m_{\texttt{pu}}} \leq \frac{\lambda}{m_{\texttt{pu}}}. \tag{46}$$

Rearranging (46) completes the proof. □

E.2.2 Relationship between mean squared errors associated with $\texttt{PrivUnit}_2$ and $\texttt{MRC}$ simulating $\texttt{PrivUnit}_2$

In the following Proposition, we show that if the scaling factor $m_{\texttt{mrc}}$ is close to the scaling parameter $m_{\texttt{pu}}$, then the mean squared error associated with $\texttt{MRC}$ simulating $\texttt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\texttt{mrc}}}\big[\,\|\hat{\boldsymbol{x}}^{\texttt{mrc}} - \boldsymbol{x}\|_2^2\,\big]$) is close to the mean squared error associated with $\texttt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\texttt{pu}}}\big[\,\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\|_2^2\,\big]$).

Proposition E.1. *Let $q^{pu}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP $\texttt{PrivUnit}_2$ mechanism with parameters p_0 and γ and estimator $\hat{\boldsymbol{x}}^{pu}$. Let $q^{mrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the $\texttt{MRC}$ privatization mechanism simulating $\texttt{PrivUnit}_2$ with N candidates and estimator $\hat{\boldsymbol{x}}^{mrc}$. Let m_{pu} denote the scaling factor associated with $\texttt{PrivUnit}_2$ and m_{mrc} denote the scaling factor associated with the $\texttt{MRC}$ scheme simulating $\texttt{PrivUnit}_2$. Consider any $\lambda > 0$. If $m_{pu} - m_{mrc} \leq \lambda \cdot m_{mrc}$, then*

$$\mathbb{E}_{q^{mrc}}\big[\,\|\hat{\boldsymbol{x}}^{mrc} - \boldsymbol{x}\|_2^2\,\big] \leq (1+\lambda)^2\,\mathbb{E}_{q^{pu}}\big[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\big] + 2(1+\lambda)(2+\lambda)\sqrt{\mathbb{E}_{q^{pu}}\big[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\big]} + (2+\lambda)^2.$$

Proof. We will start by upper bounding $1/m_{\texttt{pu}}$ in terms of $\mathbb{E}_{q^{\texttt{pu}}}\big[\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\|^2\big]$. First, observe that

$$\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\| \overset{(a)}{\geq} \|\hat{\boldsymbol{x}}^{\texttt{pu}}\| - \|\boldsymbol{x}\| \overset{(b)}{\geq} \frac{1}{m_{\texttt{pu}}} - 1 \tag{47}$$

where (a) follows from the triangle inequality and (b) follows because $\|\hat{\boldsymbol{x}}^{\texttt{pu}}\| = 1/m_{\texttt{pu}}$ and $\|\boldsymbol{x}\| \leq 1$. Next, we have

$$\frac{1}{m_{\texttt{pu}}} = \frac{1}{m_{\texttt{pu}}} - 1 + 1 \overset{(a)}{\leq} \sqrt{\mathbb{E}_{q^{\texttt{pu}}}\big[\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\|^2\big]} + 1 \tag{48}$$

where (a) follows from (47). We will now upper bound $\mathbb{E}_{q^{\texttt{mrc}}}[\|\hat{\boldsymbol{x}}^{\texttt{mrc}} - \boldsymbol{x}\|^2]$. We have

$$\begin{aligned}
\mathbb{E}_{q^{\texttt{mrc}}}[\|\hat{\boldsymbol{x}}^{\texttt{mrc}} - \boldsymbol{x}\|^2] &= \mathbb{E}_{q^{\texttt{mrc}}}[\|\hat{\boldsymbol{x}}^{\texttt{mrc}}\|^2] + \|\boldsymbol{x}\|_2^2 - 2\langle \mathbb{E}_{q^{\texttt{mrc}}}[\hat{\boldsymbol{x}}^{\texttt{mrc}}], \boldsymbol{x}\rangle \\
&\overset{(a)}{\leq} \mathbb{E}_{q^{\texttt{mrc}}}[\|\hat{\boldsymbol{x}}^{\texttt{mrc}}\|^2] + \|\boldsymbol{x}\|_2^2 + 2\sqrt{\mathbb{E}_{q^{\texttt{mrc}}}[\|\hat{\boldsymbol{x}}^{\texttt{mrc}}\|^2] \cdot \|\boldsymbol{x}\|^2} \\
&\overset{(b)}{\leq} \left(\frac{1}{m_{\texttt{mrc}}}\right)^2 + 1 + \frac{2}{m_{\texttt{mrc}}} \\
&\overset{(c)}{\leq} \left(\frac{1+\lambda}{m_{\texttt{pu}}}\right)^2 + 1 + \frac{2(1+\lambda)}{m_{\texttt{pu}}} \\
&\overset{(d)}{\leq} (1+\lambda)^2\,\mathbb{E}_{q^{\texttt{pu}}}\big[\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\|^2\big] + 2(1+\lambda)(2+\lambda)\sqrt{\mathbb{E}_{q^{\texttt{pu}}}\big[\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\|^2\big]} + (2+\lambda)^2
\end{aligned}$$

where (a) follows from Cauchy–Schwarz inequality, (b) follows because $\|\hat{\boldsymbol{x}}^{\texttt{mrc}}\| = 1/m_{\texttt{mrc}}$ and $\|\boldsymbol{x}\| \leq 1$, (c) follows from Lemma E.2 (which shows $m_{\texttt{pu}} - m_{\texttt{mrc}} \leq \lambda \cdot m_{\texttt{mrc}}$), and (d) follows using (48) and some simple manipulations. □

In the following Lemma, we show that with on the order of ε-bits of communication, the mean squared error associated with $\texttt{MRC}$ simulating $\texttt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\texttt{mrc}}}\big[\,\|\hat{\boldsymbol{x}}^{\texttt{mrc}} - \boldsymbol{x}\|_2^2\,\big]$) is close to the mean squared error associated with $\texttt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\texttt{pu}}}\big[\,\|\hat{\boldsymbol{x}}^{\texttt{pu}} - \boldsymbol{x}\|_2^2\,\big]$).

Lemma E.3. *Let $q^{pu}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP $\texttt{PrivUnit}_2$ mechanism with parameters p_0 and γ and estimator $\hat{\boldsymbol{x}}^{pu}$. Let $q^{mrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the $\texttt{MRC}$ privatization mechanism simulating $\texttt{PrivUnit}_2$ with N candidates and estimator $\hat{\boldsymbol{x}}^{mrc}$. Consider any $\lambda > 0$. Then,*

$$\mathbb{E}_{q^{mrc}}\big[\,\|\hat{\boldsymbol{x}}^{mrc} - \boldsymbol{x}\|_2^2\,\big] \leq (1+\lambda)^2\,\mathbb{E}_{q^{pu}}\big[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\big] + 2(1+\lambda)(2+\lambda)\sqrt{\mathbb{E}_{q^{pu}}\big[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\big]} + (2+\lambda)^2$$

as long as

$$N \geq 2e^{2\varepsilon}\left(\frac{2(1+\lambda)}{\lambda\,(p_0 - 1/2)}\right)^2 \ln\left(\frac{4(1+\lambda)}{\lambda\,(p_0 - 1/2)}\right).$$

Proof. The proof follows from Proposition E.1 and Lemma E.2. □

E.2.3 Simulating $\texttt{PrivUnit}_2$ using Minimal Random Coding

The following Theorem shows that, for mean estimation, $\texttt{MRC}$ can simulate $\texttt{PrivUnit}_2$ in a near-lossless manner (when n is large and λ is small) while only using on the order of ε bits of communication.

Theorem E.1. *Let $r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right)$ and $r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{mrc}, q^{mrc}\right)$ be the empirical mean estimation error for $\texttt{PrivUnit}_2$ with parameter p_0 and $\texttt{MRC}$ simulating $\texttt{PrivUnit}_2$ with N candidates respectively. Consider any $\lambda > 0$. Then,*

$$r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{mrc}, q^{mrc}\right) \leq (1+\lambda)^2\, r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right) + 2(1+\lambda)(2+\lambda)\sqrt{\frac{r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right)}{n}} + \frac{(2+\lambda)^2}{n}.$$

as long as

$$N \geq 2e^{2\varepsilon}\left(\frac{2(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right)^2 \ln\left(\frac{4(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right).$$

Proof. The proof follows directly from Lemma E.3 since for all $i \in [n]$, $\hat{\boldsymbol{x}}_i^{\texttt{mrc}}$ are independent of each other as well as unbiased. □

E.3 Empirical Comparisons

In this section, we compare $\texttt{MRC}$ simulating $\texttt{PrivUnit}_2$ (using its approximate DP guarantee) against $\texttt{PrivUnit}_2$ and SQKR for mean estimation with $d = 500$ and $n = 5000$. We use the same data generation scheme described in Section 4.3 and set $\delta = 10^{-6}$. As before, SQKR uses #-bits $= \varepsilon$ because it leads to a poor performance if #-bits $> \varepsilon$. We show the privacy-accuracy tradeoffs for these three methods in Figure 3. We see that $\texttt{MRC}$ simulating $\texttt{PrivUnit}_2$ can attain the accuracy of the uncompressed $\texttt{PrivUnit}_2$ for the range of ε's typically considered by LDP mechanisms while only using $(3\varepsilon/\ln 2) + 6$ bits. In comparison with the results from Section 4.3, the results in this section come with an approximate guarantee ($\delta = 10^{-6}$) and with a higher number of bits of communication. In other words, along with the obvious gains of pure privacy instead of approximate privacy, $\texttt{MMRC}$ results in a lower communication cost (and therefore a lower computation cost) compared to $\texttt{MRC}$.

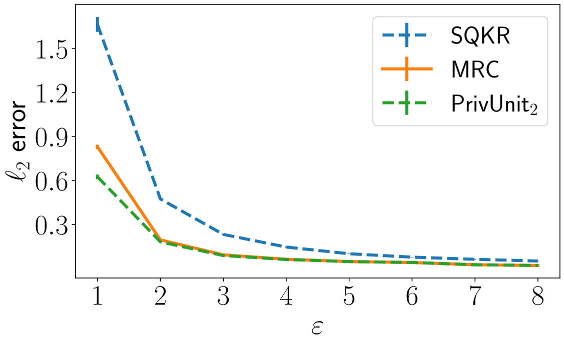

Figure 3: Comparing $\texttt{PrivUnit}_2$, $\texttt{MRC}$ simulating $\texttt{PrivUnit}_2$ and SQKR for mean estimation in terms of ℓ_2 error vs ε with $d = 500$, $n = 5000$, and #bits $= (3\varepsilon/\ln 2) + 6$.

F MODIFIED MINIMAL RANDOM CODING SIMULATING $\texttt{PrivUnit}_2$

In this section, we prove Lemma 4.1 (in Appendix F.1) and Theorem 4.1 (in Appendix F.2.3). To prove Theorem 4.1, first, in Appendix F.2.1, we show that when the number of candidates N is exponential in ε, the scaling factor $m_{\texttt{mmrc}}$ is close to the scaling parameter associated with $\texttt{PrivUnit}_2$ (i.e., $m_{\texttt{pu}}$). Next, in Appendix F.2.2, we provide the relationship between the mean squared error associated with $\texttt{MMRC}$ simulating $\texttt{PrivUnit}_2$ and the mean squared error associated with $\texttt{PrivUnit}_2$. Finally, in Appendix F.3, we provide some empirical comparisons in addition to the ones in Section 4.3 between $\texttt{MMRC}$ simulating $\texttt{PrivUnit}_2$ and $\texttt{PrivUnit}_2$.

F.1 Unbiased Modified Minimal Random Coding simulating $\texttt{PrivUnit}_2$

Consider the $\texttt{PrivUnit}_2$ ε-LDP mechanism q^{pu} described in Section 2 with parameters p_0 and γ. $\texttt{PrivUnit}_2$ is a cap-based mechanism with $\mathsf{Cap}_{\boldsymbol{x}} = \{\boldsymbol{z} \in \mathbb{S}^{d-1} \mid \langle \boldsymbol{z}, \boldsymbol{x} \rangle \geq \gamma\}$ as discussed in Appendix D. Let $\pi^{\texttt{mmrc}}$ be the distribution and $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_N$ be the candidates obtained from Algorithm 2 when the reference distribution is $\text{Unif}(\mathbb{S}^{d-1})$. Let $K \sim \pi^{\texttt{mmrc}}(\cdot)$. Define $p_{\texttt{mmrc}} := \mathbb{P}(\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}})$ to be the probability with which the sampled candidate $\boldsymbol{z}_K$ belongs to the spherical cap associated with $\texttt{PrivUnit}_2$. Define $m_{\texttt{mmrc}}$ as the scaling factor in (2) when p_0 in (2) is replaced by $p_{\texttt{mmrc}}$. Define $\hat{\boldsymbol{x}}^{\texttt{mmrc}} := \boldsymbol{z}_K / m_{\texttt{mmrc}}$ as the estimator of the $\texttt{MMRC}$ mechanism simulating $\texttt{PrivUnit}_2$.

Lemma 4.1. *Let $\hat{\boldsymbol{x}}^{mmrc}$ be the estimator of the* $\texttt{MMRC}$ *mechanism simulating* $\texttt{PrivUnit}_2$ *as defined above. Then, $\mathbb{E}_{q^{mmrc}}[\hat{\boldsymbol{x}}^{mmrc}] = \boldsymbol{x}$.*

Proof. The proof is similar to the proof of Lemma E.1. □

F.2 Utility of Modified Minimal Random Coding simulating $\texttt{PrivUnit}_2$

F.2.1 The scaling factors of $\texttt{PrivUnit}_2$ and $\texttt{MMRC}$ are close when N is of the right order

In the following Lemma, we show that when the number of candidates N is exponential in ε, then the scaling parameters associated with $\texttt{PrivUnit}_2$ and the $\texttt{MMRC}$ scheme simulating $\texttt{PrivUnit}_2$ are close.

Lemma F.1. *Let N denote the number of candidates used in the* $\texttt{MMRC}$ *scheme. Let $K \sim \pi^{mmrc}$ where π^{mmrc} is the distribution over the indices $[N]$ associated the* $\texttt{MMRC}$ *scheme simulating* $\texttt{PrivUnit}_2(\boldsymbol{x}, \gamma, p_0)$*. Consider any $\lambda > 0$. Then, the scaling factor m_{pu} associated with* $\texttt{PrivUnit}_2$ *and the scaling factor m_{mmrc} associated with the* $\texttt{MMRC}$ *scheme simulating* $\texttt{PrivUnit}_2$ *are such that*

$$m_{pu} - m_{mmrc} \leq \lambda \cdot m_{mmrc}$$

as long as

$$N \geq \frac{e^{2\varepsilon}}{2} \left(\frac{2(1+\lambda)}{\lambda\,(p_0 - 1/2)} \right)^2 \ln \left(\frac{4(1+\lambda)}{\lambda\,(p_0 - 1/2)} \right).$$

Proof. The proof follows a structure similar to the proof of Lemma E.2. As in the proof of Lemma E.2, we have

$$\frac{1}{m_{\texttt{mmrc}}} - \frac{1}{m_{\texttt{pu}}} \leq \frac{1}{m_{\texttt{pu}}} \left(\frac{1}{\dfrac{p_0 - 1/2}{p_0 - p_{\texttt{mmrc}}} - 1} \right)$$

We will now upper bound $\dfrac{p_0 - p_{\texttt{mmrc}}}{p_0 - 1/2}$. We start by obtaining expressions for $p_{\texttt{mmrc}}$ and p_0.

To compute $p_{\texttt{mmrc}} := \mathbb{P}\{\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}}\}$, recall that θ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$. Let $c_1(\varepsilon, d)$ and $c_2(\varepsilon, d)$ be as defined in (35). Let $\bar{c}_1(\varepsilon, d) = c_1(\varepsilon, d) \times A(1, d)$ and $\bar{c}_2(\varepsilon, d) = c_2(\varepsilon, d) \times A(1, d)$. It is easy to see from Algorithm 3 and (35) that $\mathbb{P}(\boldsymbol{z}_k \in \mathsf{Cap}_{\boldsymbol{x}}) = \bar{c}_1(\varepsilon, d)/p_0$. Further, since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\theta \sim \frac{1}{N} \mathsf{Binom}\left(N, \frac{\bar{c}_1(\varepsilon, d)}{p_0} \right),$$

so we have

$$\begin{aligned} p_{\texttt{mmrc}} = \mathbb{P}\{\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}}\} &= \mathbb{E}\left[\mathbb{P}\{\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}} | \theta\}\right] \\ &\overset{(a)}{=} \mathbb{E}\Bigg[\frac{\theta \bar{c}_1(\varepsilon, d)}{\bar{c}_2(\varepsilon, d) + \mathbb{E}[\theta]\left(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)\right)} \times \mathbb{1}\left(\theta \leq \mathbb{E}[\theta]\right) \\ &\qquad + \frac{\mathbb{E}[\theta]\,\bar{c}_1(\varepsilon, d) + (\theta - \mathbb{E}[\theta])\bar{c}_2(\varepsilon, d)}{\bar{c}_2(\varepsilon, d) + \mathbb{E}[\theta]\left(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)\right)} \times \mathbb{1}\left(\theta > \mathbb{E}[\theta]\right) \Bigg] \end{aligned} \tag{49}$$

where (a) follows from Algorithm 2.

Similarly, with some simple manipulations on the definition of p_0, we have

$$p_0 = \frac{\mathbb{E}\left[\theta\right] \bar{c}_1(\varepsilon, d)}{\bar{c}_2(\varepsilon, d) + \mathbb{E}\left[\theta\right] \left(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)\right)} \tag{50}$$

From (49) and (50), we have

$$\begin{aligned} p_0 - p_{\mathtt{mmrc}} &= \frac{\mathbb{E}\left[\bar{c}_1(\varepsilon, d)(\mathbb{E}[\theta] - \theta) \times \mathbb{1}\left(\theta \leq \mathbb{E}\left[\theta\right]\right) + \bar{c}_2(\varepsilon, d)(\mathbb{E}[\theta] - \theta) \times \mathbb{1}\left(\theta > \mathbb{E}\left[\theta\right]\right)\right]}{\bar{c}_2(\varepsilon, d) + \mathbb{E}\left[\theta\right] \left(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)\right)} \\ &\overset{(a)}{\leq} \frac{\mathbb{E}\left[\bar{c}_1(\varepsilon, d)(\mathbb{E}[\theta] - \theta) \times \mathbb{1}\left(\theta \leq \mathbb{E}\left[\theta\right]\right)\right]}{\bar{c}_2(\varepsilon, d) + \mathbb{E}\left[\theta\right] \left(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)\right)} \end{aligned}$$

where (a) follows because $(\mathbb{E}[\theta] - \theta) \times \mathbb{1}\left(\theta > \mathbb{E}\left[\theta\right]\right) \leq 0$. Now, using the Hoeffding's inequality, we have $\mathbb{P}\left\{|\theta - \mathbb{E}[\theta]| \geq \sqrt{\frac{\ln(2/\beta)}{2N}}\right\} \leq \beta$. Conditioned on the event $\left\{|\theta - \mathbb{E}[\theta]| \leq \sqrt{\frac{\ln(2/\beta)}{2N}}\right\}$ and using the fact that $|p_0 - p_{\mathtt{mmrc}}| \leq 1$, we have

$$\begin{aligned} p_0 - p_{\mathtt{mmrc}} &\leq \frac{\bar{c}_1(\varepsilon, d)\sqrt{\frac{\ln(2/\beta)}{2N}}}{\bar{c}_2(\varepsilon, d) + \mathbb{E}\left[\theta\right] \left(\bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)\right)} + \beta \\ &\overset{(a)}{\leq} \frac{\bar{c}_1(\varepsilon, d)}{\bar{c}_2(\varepsilon, d)}\sqrt{\frac{\ln(2/\beta)}{2N}} + \beta \overset{(b)}{\leq} \exp(\varepsilon)\sqrt{\frac{\ln(2/\beta)}{2N}} + \beta \overset{(c)}{\leq} \frac{\lambda(p_0 - 1/2)}{1 + \lambda}. \end{aligned}$$

where (a) follows because $\mathbb{E}\left[\theta\right] \geq 0$, (b) follows because $\bar{c}_1(\varepsilon, d)/\bar{c}_2(\varepsilon, d) \leq e^{\varepsilon}$, and (c) follows if we pick

$$\beta \leq \frac{\lambda(p_0 - 1/2)}{2(1 + \lambda)} \qquad \text{and} \qquad N \geq \frac{\exp(2\varepsilon)\ln(2/\beta)}{2\left(\frac{\lambda(p_0 - 1/2)}{1+\lambda} - \beta\right)^2} = \frac{\exp(2\varepsilon)}{2}\left(\frac{2(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right)^2 \ln\left(\frac{4(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right).$$

The rest of the proof is similar to the proof of Lemma E.2. □

F.2.2 Relationship between the mean squared errors associated with $\mathtt{PrivUnit}_2$ and $\mathtt{MMRC}$ simulating $\mathtt{PrivUnit}_2$

In the following Proposition, we show that if the scaling factor $m_{\mathtt{mmrc}}$ is close to the scaling parameter $m_{\mathtt{pu}}$, then the mean squared error associated with $\mathtt{MMRC}$ simulating $\mathtt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\mathtt{mmrc}}}\left[\, \|\hat{\boldsymbol{x}}^{\mathtt{mmrc}} - \boldsymbol{x}\|_2^2\, \right]$) is close to the mean squared error associated with $\mathtt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\mathtt{pu}}}\left[\, \|\hat{\boldsymbol{x}}^{\mathtt{pu}} - \boldsymbol{x}\|_2^2\, \right]$).

Proposition F.1. *Let $q^{pu}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP $\mathtt{PrivUnit}_2$ mechanism with parameters p_0 and γ and estimator $\hat{\boldsymbol{x}}^{pu}$. Let $q^{mmrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the $\mathtt{MMRC}$ privatization mechanism simulating $\mathtt{PrivUnit}_2$ with N candidates and estimator $\hat{\boldsymbol{x}}^{mmrc}$. Let m_{pu} denote the scaling factor associated with $\mathtt{PrivUnit}_2$ and m_{mmrc} denote the scaling factor associated with the $\mathtt{MMRC}$ scheme simulating $\mathtt{PrivUnit}_2$. Consider any $\lambda > 0$. If $m_{pu} - m_{mmrc} \leq \lambda \cdot m_{mmrc}$, then*

$$\mathbb{E}_{q^{mmrc}}\left[\, \|\hat{\boldsymbol{x}}^{mmrc} - \boldsymbol{x}\|_2^2\, \right] \leq (1+\lambda)^2\, \mathbb{E}_{q^{pu}}\left[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\right] + 2(1+\lambda)(2+\lambda)\sqrt{\mathbb{E}_{q^{pu}}\left[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\right]} + (2+\lambda)^2.$$

Proof. The proof is similar to the proof of Proposition E.1. □

In the following Lemma, we show that with on the order of ε-bits of communication, the mean squared error associated with $\mathtt{MMRC}$ simulating $\mathtt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\mathtt{mmrc}}}\left[\, \|\hat{\boldsymbol{x}}^{\mathtt{mmrc}} - \boldsymbol{x}\|_2^2\, \right]$) is close to the mean squared error associated with $\mathtt{PrivUnit}_2$ (i.e., $\mathbb{E}_{q^{\mathtt{pu}}}\left[\, \|\hat{\boldsymbol{x}}^{\mathtt{pu}} - \boldsymbol{x}\|_2^2\, \right]$).

Lemma F.2. *Let $q^{pu}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP $\mathtt{PrivUnit}_2$ mechanism with parameters p_0 and γ and estimator $\hat{\boldsymbol{x}}^{pu}$. Let $q^{mmrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the $\mathtt{MMRC}$ privatization mechanism simulating $\mathtt{PrivUnit}_2$ with N candidates and estimator $\hat{\boldsymbol{x}}^{mmrc}$ as defined above. Consider any $\lambda > 0$. Then,*

$$\mathbb{E}_{q^{mmrc}}\left[\, \|\hat{\boldsymbol{x}}^{mmrc} - \boldsymbol{x}\|_2^2\, \right] \leq (1+\lambda)^2\, \mathbb{E}_{q^{pu}}\left[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\right] + 2(1+\lambda)(2+\lambda)\sqrt{\mathbb{E}_{q^{pu}}\left[\|\hat{\boldsymbol{x}}^{pu} - \boldsymbol{x}\|^2\right]} + (2+\lambda)^2$$

as long as

$$N \geq \frac{e^{2\varepsilon}}{2}\left(\frac{2(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right)^2 \ln\left(\frac{4(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right).$$

Proof. The proof follows from Proposition F.1 and Lemma F.1. □

F.2.3 Simulating $\mathtt{PrivUnit}_2$ using Modified Minimal Random Coding

Now, we provide a proof of Theorem 4.1.

Theorem 4.1. *Let $r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right)$ and $r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{mmrc}, q^{mmrc}\right)$ be the empirical mean estimation error for $\mathtt{PrivUnit}_2$ with parameter p_0 and $\mathtt{MMRC}$ simulating $\mathtt{PrivUnit}_2$ with N candidates respectively. Consider any $\lambda > 0$. Then,*

$$r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{mmrc}, q^{mmrc}\right) \leq (1+\lambda)^2\, r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right) + 2(1+\lambda)(2+\lambda)\sqrt{\frac{r_{\mathsf{ME}}\left(\hat{\boldsymbol{\mu}}^{pu}, q^{pu}\right)}{n}} + \frac{(2+\lambda)^2}{n}.$$

as long as

$$N \geq \frac{e^{2\varepsilon}}{2}\left(\frac{2(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right)^2 \ln\left(\frac{4(1+\lambda)}{\lambda\left(p_0 - 1/2\right)}\right). \tag{11}$$

Proof. The proof follows directly from Lemma F.2 since for all $i \in [n]$, $\hat{\boldsymbol{x}}_i^{\mathtt{mmrc}}$ are independent of each other as well as unbiased. □

F.3 Additional Empirical Comparisons

In Section 4.3, we empirically demonstrated the privacy-accuracy-communication tradeoffs of $\mathtt{MMRC}$ simulating $\mathtt{PrivUnit}_2$ against $\mathtt{PrivUnit}_2$ and SQKR in terms of ℓ_2 error vs #bits and ℓ_2 error vs ε (see Figure 1). In this section, we provide comparisons between these methods in terms of ℓ_2 error vs d (see Figure 4 (left)) and ℓ_2 error vs n (see Figure 4 (right)) for a fixed ε (=6) and a fixed #bits (=11). As before, SQKR uses #bits $= \varepsilon$ for both because it leads to a poor performance if #bits $> \varepsilon$.

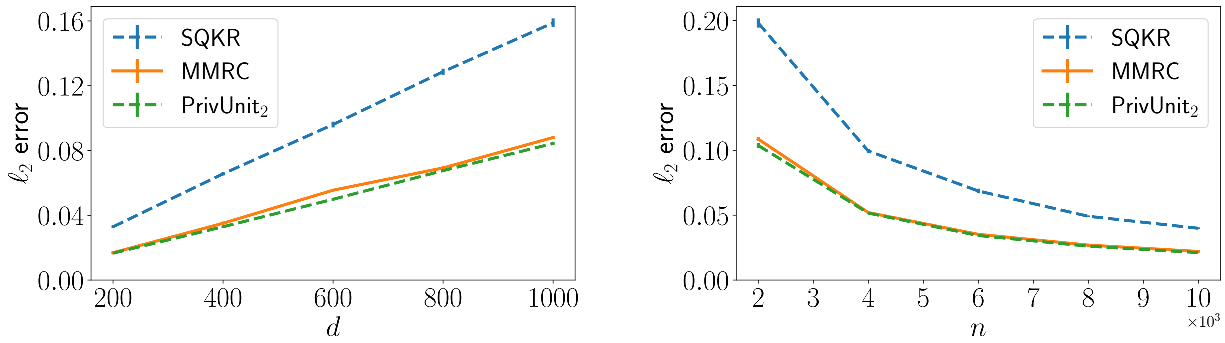

Figure 4: Comparing $\mathtt{PrivUnit}_2$, $\mathtt{MMRC}$ simulating $\mathtt{PrivUnit}_2$ and SQKR for mean estimation with $\varepsilon = 6$ and #bits $= 11$. **Left:** ℓ_2 error vs d for $n = 5000$. **Right:** ℓ_2 error vs n for $d = 500$.

G PRELIMINARY ON $\mathtt{Subset\ Selection}$

In this section, we briefly recap the $\mathtt{Subset\ Selection}$ (SS) mechanism proposed in Ye and Barg (2018). Let $\boldsymbol{x} = (x_1, x_2, ..., x_d) \in \{0,1\}^d$ be the one-hot representation of an input symbol in $\mathcal{X} = [d] = \{1, \cdots, d\}$[11]. Let $q^{\mathsf{ss}}(\boldsymbol{z}|\boldsymbol{x})$ be the $\mathtt{Subset\ Selection}$ mechanism defined in Ye and Barg (2018) where the output alphabet is the set of all $d-$bit binary strings with Hamming weight $s \in [d]$, i.e.,

$$\mathcal{Z} = \left\{ \boldsymbol{z} = (z_1, z_2, ..., z_d) \in \{0,1\}^d : \sum_{i=1}^{d} z_i = s \right\}. \tag{51}$$

[11] With a slight abuse of notation, when context is clear, we sometime use $\boldsymbol{x} = i$ for some $i \in [d]$ to indicate the one-hot representation of symbol i

Given $\boldsymbol{x} \in \mathcal{X}$, `Subset Selection` maps it to $\boldsymbol{z} \in \mathcal{Z}$ with the following conditional probability:

$$q^{\mathsf{ss}}(\boldsymbol{z}|\boldsymbol{x}=i) \coloneqq \begin{cases} \dfrac{e^{\varepsilon}}{\binom{d-1}{s-1}e^{\varepsilon}+\binom{d-1}{s}} & \text{if } \boldsymbol{z} \in \mathcal{Z}_i \\ \dfrac{1}{\binom{d-1}{s-1}e^{\varepsilon}+\binom{d-1}{s}} & \text{if } \boldsymbol{z} \in \mathcal{Z} \setminus \mathcal{Z}_i \end{cases} \tag{52}$$

where $\mathcal{Z}_i = \left\{\boldsymbol{z} = (z^{(1)}, \cdots, z^{(d)}) \in \mathcal{Z} : z^{(i)} = 1\right\}$ is the set of elements in $\mathcal{Z}$ with 1 in the i^{th} location.

Ye and Barg (2018) show that the marginal distribution of $\boldsymbol{z}$ is a linear function of that of $\boldsymbol{x}$. In particular, if we define $p_i \coloneqq \mathbb{P}\{x_i = 1\}$ for all $i \in [d]$ and let $\boldsymbol{z} \sim q^{\mathsf{ss}}(\cdot|\boldsymbol{x})$, then (53) is due to (5) in Ye and Barg (2018),

$$\begin{aligned} q_i^{\mathsf{ss}} \coloneqq \mathbb{P}\{z_i = 1\} &= \frac{\binom{d-1}{s-1}e^{\varepsilon}p_i + \left(\binom{d-2}{s-2}e^{\varepsilon}+\binom{d-2}{s-1}\right)(1-p_i)}{\binom{d-1}{s-1}e^{\varepsilon}+\binom{d-1}{s}} && (53)\\ &= \frac{s(d-s)(e^{\varepsilon}-1)}{(d-1)(s(e^{\varepsilon}-1)+d)}p_i + \frac{s((s-1)e^{\varepsilon}+(d-s))}{(d-1)(s(e^{\varepsilon}-1)+d)} \\ &= m_{\mathsf{ss}} \cdot p_i + b_{\mathsf{ss}}, && (54) \end{aligned}$$

where

$$m_{\mathsf{ss}} \coloneqq \frac{s(d-s)(e^{\varepsilon}-1)}{(d-1)(s(e^{\varepsilon}-1)+d)}, \qquad b_{\mathsf{ss}} \coloneqq \frac{s((s-1)e^{\varepsilon}+(d-s))}{(d-1)(s(e^{\varepsilon}-1)+d)}. \tag{55}$$

The final estimator of $\boldsymbol{x}$ is denoted by $\hat{\boldsymbol{x}}^{\mathsf{ss}}$ and is defined as $\frac{1}{m_{\mathsf{ss}}} \cdot (\boldsymbol{z} - b_{\mathsf{ss}} \cdot \mathbf{1}_d)$, where $\mathbf{1}_d \triangleq [1, \cdots, 1]^{\intercal} \in \mathbb{R}^d$. In other words, m_{ss} and b_{ss} are used de-bias the outcome $\boldsymbol{z}$. The scheme is summarized in Algorithm 4.

Algorithm 4: Subset Selection

```
Require: $\boldsymbol{x} \in [d]$, $s \in [d]$.
  Draw a $s$-hot random vector $\boldsymbol{z}$ according to the distribution $q^{\mathsf{ss}}(\boldsymbol{z}|\boldsymbol{x})$ in (52).
      return $\hat{\boldsymbol{x}}^{\mathsf{ss}} = \frac{1}{m_{\mathsf{ss}}} \cdot (\boldsymbol{z} - b_{\mathsf{ss}} \cdot \mathbf{1}_d)$
```

G.1 `Subset Selection` is unbiased and order-optimal

The following proposition borrowed from Ye and Barg (2018) shows that the output of the `Subset Selection` mechanism (a) is unbiased and (b) has order-optimal utility.

Proposition G.1. *Let $\hat{\boldsymbol{x}}^{ss}$ = `Subset Selection`$(\boldsymbol{x}, s)$ for some $\boldsymbol{x} \in \mathcal{X}$ and $s \in [d]$. Then, $\mathbb{E}[\hat{\boldsymbol{x}}^{ss}] = \boldsymbol{x}$. Further, the ℓ_2 estimation error is*

$$\mathbb{E}\left[\|\hat{\boldsymbol{x}}^{ss} - \boldsymbol{x}\|_2^2\right] = \left(\frac{(s(d-2)+1)\,e^{2\varepsilon}}{(d-s)\,(e^{\varepsilon}-1)^2} + \frac{2(d-2)}{(e^{\varepsilon}-1)^2} + \frac{(d-2)(d-s)+1}{s\,(e^{\varepsilon}-1)^2} - \sum_i p_i^2\right).$$

Moreover, if we pick $s \coloneqq \lceil \frac{d}{1+e^{\varepsilon}} \rceil$, then

$$\mathbb{E}\left[\|\hat{\boldsymbol{x}}^{ss} - \boldsymbol{x}\|_2^2\right] = \frac{d}{\min\left(e^{\varepsilon}, (e^{\varepsilon}-1)^2, d\right)},$$

which is order-optimal.

G.2 `Subset Selection` is a cap-based mechanism

As discussed in Section 3, q^{ss} defined in (3) is a cap-based mechanism with $\mathsf{Cap}_{\boldsymbol{x}} = \mathcal{Z}_{\boldsymbol{x}}$, $c_1(\varepsilon, d) = \dfrac{e^{\varepsilon}}{\binom{d-1}{s-1}e^{\varepsilon}+\binom{d-1}{s}}$, and $c_2(\varepsilon, d) = \dfrac{1}{\binom{d-1}{s-1}e^{\varepsilon}+\binom{d-1}{s}}$.

Further, $\mathbb{P}_{\boldsymbol{z} \sim \mathrm{Unif}(\mathcal{Z})}\left(\boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{x}}\right) = \frac{\binom{d-1}{s-1}}{\binom{d}{s}} = \frac{s}{d}$. Therefore,

$$\frac{c_1(\varepsilon, d)}{c_2(\varepsilon, d)} \times \mathbb{P}_{\boldsymbol{z} \sim \mathrm{Unif}(\mathcal{Z})}\left(\boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{x}}\right) = e^{\varepsilon} \times \frac{s}{d} \overset{(a)}{=} \frac{e^{\varepsilon}}{d} \times \lceil \frac{d}{1+e^{\varepsilon}} \rceil \geq \frac{e^{\varepsilon}}{d} \times \frac{d}{1+e^{\varepsilon}} \overset{(b)}{\geq} \frac{1}{2}$$

where (a) follows by plugging in $s = \lceil \frac{d}{1+e^{\varepsilon}} \rceil$ and (b) follows because $\varepsilon \geq 0$.

H SIMULATING `Subset Selection` USING MINIMAL RANDOM CODING

In this section, we simulate `Subset Selection` using `MRC` analogous to how we simulate `Subset Selection` using `MMRC` in Section 5. First, in Appendix H.1, we provide an unbiased estimator for `MRC` simulating `Subset Selection`. Next, in Appendix H.2 we provide the utility guarantee associated with `MRC` simulating `Subset Selection`. To do that, first, in Appendix H.2.1, we show that when the number of candidates N is exponential in ε, the scaling factor $m_{\texttt{mrc}}$ is close to the scaling parameter associated with `Subset Selection` (i.e., $m_{\texttt{ss}}$). Next, in Appendix H.2.2, we provide the relationship between the mean squared error associated with `MRC` simulating `Subset Selection` and the mean squared error associated with `Subset Selection`. In Appendix H.2.3, we combine everything and show that, for frequency estimation, `MRC` can simulate `Subset Selection` in a near-lossless manner while only using on the order of ε-bits of communication. Finally, in Appendix H.3, we provide some empirical comparisons.

H.1 Unbiased Minimal Random Coding simulating `Subset Selection`

Consider the `Subset Selection` ε-LDP mechanism $q^{\texttt{ss}}$ with parameter s as described in Section 2 and Appendix G. Let $\pi^{\texttt{mrc}}$ be the distribution and $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_N$ be the candidates obtained from Algorithm 1 when the reference distribution is $\mathrm{Unif}(\mathcal{Z})$ where $\mathcal{Z}$ is as defined in (51). Let θ denote the fraction of candidates inside $\mathsf{Cap}_{\boldsymbol{x}} = \mathcal{Z}_{\boldsymbol{x}}$ where $\mathcal{Z}_{\boldsymbol{x}}$ is the set of elements in $\mathcal{Z}$ with 1 in the same location as $\boldsymbol{x}$. It is easy to see that $\theta \sim \frac{1}{N}\mathsf{Binom}\left(N, \frac{s}{d}\right)$. Let $q_i^{\texttt{mrc}} = \mathbb{P}(z_i = 1)$ where $\boldsymbol{z} \sim q^{\texttt{mrc}}(\cdot|\boldsymbol{x})$ i.e., $q_i^{\texttt{mrc}} = \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1\right\}$ where $K \sim \pi^{\texttt{mrc}}(\cdot)$.

The following lemma shows that the marginal distribution of $q_i^{\texttt{mrc}}$ can be written as a linear function of p_i similar to $q_i^{\texttt{ss}}$ in (54). This allows us to provide an unbiased estimator for `MRC` simulating `Subset Selection`.

Lemma H.1. *Let $K \sim \pi^{\texttt{mrc}}(\cdot)$ and $q_i^{\texttt{mrc}} = \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1\right\}$ for $i \in [d]$. Then,*

$$q_i^{\texttt{mrc}} = p_i m_{\texttt{mrc}} + b_{\texttt{mrc}}$$

where

$$m_{\texttt{mrc}} := \mathbb{E}\left[\frac{\theta e^{\varepsilon}}{e^{\varepsilon}\theta + (1-\theta)}\right] - \frac{1}{d-1}\mathbb{E}\left[s - \frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right],$$

$$b_{\texttt{mrc}} := \frac{1}{d-1}\mathbb{E}\left[s - \frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right].$$

Further, $\hat{\boldsymbol{x}}_{\texttt{mrc}} := (\boldsymbol{z}_K - b_{\texttt{mrc}} \cdot \mathbf{1}_d)/m_{\texttt{mrc}}$ is an unbiased estimator of $\boldsymbol{x}$, i.e., $\mathbb{E}[\hat{\boldsymbol{x}}_{\texttt{mrc}}] = \boldsymbol{x}$.

Proof. We have

$$\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1\right\} = \sum_j p_j \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\} \overset{(a)}{=} p_i \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i\right\} + (1-p_i)\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\}. \quad (56)$$

where (a) follows by symmetry. Next, we compute $\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i\right\}$ and $\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\}$ separately.

To compute $\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i\right\}$, recall that θ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$. From Appendix G.2, recall that $c_1(\varepsilon, d) := \frac{e^{\varepsilon}}{\binom{d-1}{s-1}e^{\varepsilon} + \binom{d-1}{s}}$, $c_2(\varepsilon, d) := \frac{1}{\binom{d-1}{s-1}e^{\varepsilon} + \binom{d-1}{s}}$. Further, since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\theta \sim \frac{1}{N}\mathsf{Binom}\left(N, \frac{\binom{d-1}{s-1}}{\binom{d}{s}}\right) = \frac{1}{N}\mathsf{Binom}\left(N, \frac{s}{d}\right),$$

so we have

$$\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i\right\} = \mathbb{P}\left\{\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}} | \boldsymbol{x} = i\right\} \overset{(a)}{=} \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}} | \boldsymbol{x} = i, \theta\right\}\right] = \mathbb{E}\left[\frac{c_1(\varepsilon,d)\theta}{c_1(\varepsilon,d)\theta + (1-\theta)c_2(\varepsilon,d)}\right] \overset{(b)}{=} \mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right], \tag{57}$$

where (a) follows by the law of total probability and (b) is due to $c_1(\varepsilon,d)/c_2(\varepsilon,d) = e^{\varepsilon}$.

To compute $\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\}$, we decompose it into

$$\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\} = \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 | \boldsymbol{x} = j\right\} + \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j\right\}, \tag{58}$$

for any $j \neq i$ and calculate each of the terms separately.

As before, let θ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$. Further, let $\bar{\theta}$ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$ as well as have 1 in the j^{th} location. Since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\bar{\theta} \sim \frac{1}{N}\mathsf{Binom}\left(N\theta, \frac{\binom{d-2}{s-2}}{\binom{d-1}{s-1}}\right) = \frac{1}{N}\mathsf{Binom}\left(N\theta, \frac{s-1}{d-1}\right),$$

so we have

$$\begin{aligned}\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 | \boldsymbol{x} = j\right\} &\overset{(a)}{=} \mathbb{E}_{\theta}\left[\mathbb{E}_{\bar{\theta}}\left[\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 \middle| \boldsymbol{x} = j, \bar{\theta}, \theta\right\}\right]\right] \\ &= \mathbb{E}_{\theta}\left[\mathbb{E}_{\bar{\theta}}\left[\frac{c_1(\varepsilon,d)\bar{\theta}}{c_1(\varepsilon,d)\theta + (1-\theta)c_2(\varepsilon,d)}\right]\right] \\ &\overset{(b)}{=} \frac{s-1}{d-1}\mathbb{E}_{\theta}\left[\frac{c_1(\varepsilon,d)\theta}{c_1(\varepsilon,d)\theta + (1-\theta)c_2(\varepsilon,d)}\right] \overset{(c)}{=} \frac{s-1}{d-1}\mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right]\end{aligned} \tag{59}$$

where (a) follows by the law of total probability, (b) follows because $\mathbb{E}[\bar{\theta}] = \frac{s-1}{d-1} \times \theta$, and (c) is due to $c_1(\varepsilon,d)/c_2(\varepsilon,d) = e^{\varepsilon}$.

Similarly, to compute the term $\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j\right\}$, let $\bar{\theta}$ denote the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$ as well as have 0 in the j^{th} location. Since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\bar{\theta} \sim \frac{1}{N}\mathsf{Binom}\left(N(1-\theta), \frac{\binom{d-2}{s-1}}{\binom{d-1}{s}}\right) = \frac{1}{N}\mathsf{Binom}\left(N(1-\theta), \frac{s}{d-1}\right),$$

so we have

$$\begin{aligned}\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j\right\} &\overset{(a)}{=} \mathbb{E}_{\theta}\left[\mathbb{E}_{\bar{\theta}}\left[\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 \middle| \boldsymbol{x} = j, \bar{\theta}, \theta\right\}\right]\right] \\ &= \mathbb{E}_{\theta}\left[\mathbb{E}_{\bar{\theta}}\left[\frac{c_2(\varepsilon,d)\bar{\theta}}{c_1(\varepsilon,d)\theta + (1-\theta)c_2(\varepsilon,d)}\right]\right] \\ &\overset{(b)}{=} \frac{s}{d-1}\mathbb{E}_{\theta}\left[\frac{c_2(\varepsilon,d)(1-\theta)}{c_1(\varepsilon,d)\theta + (1-\theta)c_2(\varepsilon,d)}\right] \\ &\overset{(c)}{=} \frac{s}{d-1}\mathbb{E}\left[\frac{(1-\theta)}{e^{\varepsilon}\theta + (1-\theta)}\right],\end{aligned} \tag{60}$$

where (a) follows by the law of total probability, (b) follows because $\mathbb{E}[\bar{\theta}] = \frac{s}{d-1} \times (1-\theta)$, and (c) is due to $c_1(\varepsilon,d)/c_2(\varepsilon,d) = e^{\varepsilon}$. Using (59) and (60) in (58), we have

$$\begin{aligned}\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\} &= \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 | \boldsymbol{x} = j\right\} + \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j\right\} \\ &= \frac{s-1}{d-1}\mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right] + \frac{s}{d-1}\mathbb{E}\left[\frac{(1-\theta)}{e^{\varepsilon}\theta + (1-\theta)}\right]\end{aligned}$$

$$= \frac{1}{d-1}\left(s - \mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right]\right) \tag{61}$$

Combining everything, we have

$$\begin{aligned}
q_i^{\mathtt{mrc}} &= \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1\right\} \\
&\overset{(a)}{=} p_i \times \left[\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i\right\} - \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\}\right] + \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\}. \\
&\overset{(b)}{=} p_i \times \left[\mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right] - \frac{1}{d-1}\left(s - \mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right]\right)\right] + \frac{1}{d-1}\left(s - \mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\theta + (1-\theta)}\right]\right) \\
&\overset{(c)}{=} p_i m_{\mathtt{mrc}} + b_{\mathtt{mrc}}
\end{aligned}$$

where (a) follows from (56), (b) follows from (57) and (61), and (c) follows from the definitions of $m_{\mathtt{mrc}}$ and $b_{\mathtt{mrc}}$. Note that the above conclusion holds for all prior distribution $\boldsymbol{p} = (p_1, ..., p_d)$ such that $\boldsymbol{x} \sim \boldsymbol{p}$. Thus by setting $\boldsymbol{p} = \boldsymbol{x}$ (here $\boldsymbol{x}$ is viewed as a one-hot vector), i.e., letting $\boldsymbol{p}$ be the point mass distribution at $\boldsymbol{x}$, we have

$$\mathbb{E}[\hat{\boldsymbol{x}}_{\mathtt{mrc}}] = (\mathbb{E}[\boldsymbol{z}_K] - b_{\mathtt{mrc}} \cdot \mathbf{1}_d)/m_{\mathtt{mrc}} = (q^{\mathtt{mrc}} - b_{\mathtt{mrc}} \cdot \mathbf{1}_d)/m_{\mathtt{mrc}} = \left((m_{\mathtt{mrc}} \cdot \boldsymbol{p} + b_{\mathtt{mrc}} \cdot \mathbf{1}_d) - b_{\mathtt{mrc}} \cdot \mathbf{1}_d\right)/m_{\mathtt{mrc}} = \boldsymbol{p} \overset{(a)}{=} \boldsymbol{x},$$

where (a) is due to our construction of $\boldsymbol{p}$. □

H.2 Utility of Minimal Random Coding simulating `Subset Selection`

H.2.1 The scaling factors of `Subset Selection` and `MRC` are close when N is of the right order

In the following Lemma, we show that when the number of candidates N is exponential in ε, then the scaling parameters associated with `Subset Selection` and the `MRC` scheme simulating `Subset Selection` are close.

Lemma H.2. *Let N denote the number of candidates used in the* `MRC` *scheme. Let $K \sim \pi^{\mathtt{mrc}}$ where $\pi^{\mathtt{mrc}}$ is the distribution over the indices $[N]$ associated the* `MRC` *scheme simulating* `Subset Selection`. *Consider any $\lambda > 0$. Then, the scaling factors m_{ss} and b_{ss} associated with* `Subset Selection` *and the scaling factors m_{mrc} and b_{mrc} associated with the* `MRC` *scheme simulating* `Subset Selection` *are such that*

$$m_{ss} - m_{mrc} \leq \lambda \cdot m_{mrc}$$

and $b_{ss} \leq b_{mrc}$ as long as

$$N \geq \frac{2(e^{\varepsilon}+3)^2(1+\lambda)^2}{0.24^2\lambda^2} \ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right).$$

Proof. First, we will obtain convenient expressions for $m_{\mathtt{ss}}$ and $b_{\mathtt{ss}}$ defined in (55). We can write

$$m_{\mathtt{ss}} := \left(\frac{\mathbb{E}[\theta]e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])}\right) - \frac{1}{d-1}\left(s - \frac{e^{\varepsilon}\mathbb{E}[\theta]}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])}\right) \tag{62}$$

$$b_{\mathtt{ss}} := \frac{1}{d-1}\left(s - \frac{e^{\varepsilon}\mathbb{E}[\theta]}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])}\right). \tag{63}$$

To verify these, we simply plug $\mathbb{E}[\theta] = \frac{s}{d}$ into (62) resulting in:

$$m_{\mathtt{ss}} = \frac{d}{d-1}\frac{se^{\varepsilon}}{se^{\varepsilon} + (d-s)} - \frac{s}{d-1} = \frac{dse^{\varepsilon} - s^2e^{\varepsilon} - s(d-t)}{(d-1)\left(se^{\varepsilon} + d - s\right)} = \frac{s(d-s)(e^{\varepsilon}-1)}{(d-1)\left(se^{\varepsilon} + d - s\right)}.$$

and into (63) resulting in:

$$b_{\mathtt{ss}} = \frac{1}{d-1}\left(s - \frac{se^{\varepsilon}}{se^{\varepsilon} + d - s}\right) = \frac{1}{d-1}\left(\frac{s^2e^{\varepsilon} + s(d-s) - se^{\varepsilon}}{se^{\varepsilon} + d - s}\right) = \frac{1}{d-1}\left(\frac{s(s-1)e^{\varepsilon} + s(d-s)}{se^{\varepsilon} + d - s}\right).$$

Recall the definitions of $b_{\mathtt{ss}}$ and $m_{\mathtt{ss}}$ from Lemma H.1. Applying Jensen's inequality on the concave function $x \mapsto \frac{x}{x+c}$ for some $c > 0$ yields $m_{\mathtt{mrc}} \le m_{\mathtt{ss}}$ and $b_{\mathtt{mrc}} \ge b_{\mathtt{ss}}$.

Now, we will bound $|m_{\mathtt{mrc}} - m_{\mathtt{ss}}|$:

$$\begin{aligned}
|m_{\mathtt{mrc}} - m_{\mathtt{ss}}| &= \left(\frac{d}{d-1}\right)\left(\frac{\mathbb{E}[\theta]e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} - \mathbb{E}\left[\frac{\theta e^{\varepsilon}}{e^{\varepsilon}\theta + (1-\theta)}\right]\right) \\
&\overset{(a)}{\le} 2\left(\frac{\mathbb{E}[\theta]e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} - \mathbb{E}\left[\frac{\theta e^{\varepsilon}}{e^{\varepsilon}\theta + (1-\theta)}\right]\right) \\
&= 2\left(\mathbb{E}\left[\frac{\left(\mathbb{E}[\theta]-\theta\right)e^{\varepsilon}}{\left((e^{\varepsilon}-1)\mathbb{E}[\theta]+1\right)\left((e^{\varepsilon}-1)\theta+1\right)}\right]\right),
\end{aligned} \tag{64}$$

where (a) holds since $d \ge 2$. Next, we condition on the event $\mathcal{E} := \left\{|\mathbb{E}[\theta] - \theta| \le \sqrt{\frac{\ln(2/\beta)}{2N}}\right\}$, which has probability $\mathbb{P}_{\theta}\left\{\mathcal{E}\right\} \ge 1-\beta$ by Hoeffding's inequality. We continue to upper bound (64):

$$\begin{aligned}
|m_{\mathtt{mrc}} - m_{\mathtt{ss}}| &= 2\left(\mathbb{P}\left\{\mathcal{E}\right\}\mathbb{E}\left[\frac{\left(\mathbb{E}[\theta]-\theta\right)e^{\varepsilon}}{\left((e^{\varepsilon}-1)\mathbb{E}[\theta]+1\right)\left((e^{\varepsilon}-1)\theta+1\right)}\middle|\mathcal{E}\right] + \mathbb{P}\left\{\mathcal{E}^c\right\}\mathbb{E}\left[\frac{\left(\mathbb{E}[\theta]-\theta\right)e^{\varepsilon}}{\left((e^{\varepsilon}-1)\mathbb{E}[\theta]+1\right)\left((e^{\varepsilon}-1)\theta+1\right)}\middle|\mathcal{E}^c\right]\right) \\
&\overset{(a)}{\le} 2\left(\mathbb{E}\left[\frac{\left(\mathbb{E}[\theta]-\theta\right)e^{\varepsilon}}{\left((e^{\varepsilon}-1)\mathbb{E}[\theta]+1\right)\left((e^{\varepsilon}-1)\theta+1\right)}\middle|\mathcal{E}\right] + \beta\right) \\
&\overset{(b)}{\le} 2\left(\mathbb{E}\left[\frac{\left(\mathbb{E}[\theta]-\theta\right)e^{\varepsilon}}{\left((e^{\varepsilon}-1)\mathbb{E}[\theta]+1\right)\left((e^{\varepsilon}-1)E[\theta]/2+1\right)}\middle|\mathcal{E}\right] + \beta\right) \\
&\le 4\mathbb{E}\left[\frac{\left(\mathbb{E}[\theta]-\theta\right)e^{\varepsilon}}{\left((e^{\varepsilon}-1)\mathbb{E}[\theta]+1\right)^2}\middle|\mathcal{E}\right] + 2\beta \overset{(c)}{\le} 4\sqrt{\frac{\ln(2/\beta)}{2N}}\frac{e^{\varepsilon}(1+e^{\varepsilon})^2}{4e^{2\varepsilon}} + 2\beta \\
&= \sqrt{\frac{\ln(2/\beta)}{2N}}\left(e^{\varepsilon} + 2 + \frac{1}{e^{\varepsilon}}\right) + 2\beta \le \sqrt{\frac{\ln(2/\beta)}{2N}}(e^{\varepsilon}+3) + 2\beta,
\end{aligned} \tag{65}$$

where (a) holds since

$$\frac{\left(\mathbb{E}[\theta]-\theta\right)e^{\varepsilon}}{\left((e^{\varepsilon}-1)\mathbb{E}[\theta]+1\right)\left((e^{\varepsilon}-1)\theta+1\right)} = \frac{\mathbb{E}[\theta]e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} - \frac{\theta e^{\varepsilon}}{e^{\varepsilon}\theta + (1-\theta)} \le 1,$$

(b) holds if we pick N large enough so that $|\theta - \mathbb{E}[\theta]| \le \frac{\mathbb{E}[\theta]}{2}$ for which a sufficient condition is $\sqrt{\frac{\ln(2/\beta)}{2N}} \le \frac{\mathbb{E}[\theta]}{2}$ i.e., $N \ge \frac{2\ln(2/\beta)}{\mathbb{E}[\theta]^2} = 2(d/s)^2\ln(2/\beta)$, and (c) holds since $\mathbb{E}[\theta] = s/d \ge 1/(1+e^{\varepsilon})$. Notice that the constraint $N \ge 2(d/s)^2\ln(2/\beta)$ in inequality (b) can be further satisfied as long as $N \ge 2\ln(2/\beta)(1+e^{\varepsilon})^2$ since $s/d \ge 1/(1+e^{\varepsilon})$.

Next, we lower bound $m_{\mathtt{ss}}$ in (62):

$$\begin{aligned}
m_{\mathtt{ss}} &= \left(\frac{d}{d-1}\right)\left(\frac{\mathbb{E}[\theta]e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} - \frac{s}{d}\right) \ge \frac{\mathbb{E}[\theta]e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} - \frac{s}{d} \\
&\overset{(a)}{=} \frac{s}{d}\left[\frac{(e^{\varepsilon}-1)(d-s)}{(e^{\varepsilon}-1)\cdot s + d}\right] = \frac{s}{d}\left[\frac{(e^{\varepsilon}-1)(1-s/d)}{(e^{\varepsilon}-1)\cdot s/d + 1}\right] \\
&\overset{(b)}{\ge} \frac{1}{1+e^{\varepsilon}}\left[\frac{(e^{\varepsilon}-1)\left(\frac{e^{\varepsilon}}{1+e^{\varepsilon}} - \frac{1}{d}\right)}{(e^{\varepsilon}-1)\left(\frac{1}{1+e^{\varepsilon}} + \frac{1}{d}\right) + 1}\right] \\
&\overset{(c)}{\ge} \frac{1}{1+e^{\varepsilon}}\left[\frac{(e^{\varepsilon}-1)\frac{e^{\varepsilon}-1}{1+e^{\varepsilon}}}{(e^{\varepsilon}-1)\left(\frac{2}{1+e^{\varepsilon}}\right) + 1}\right] = \frac{\left(e^{\varepsilon}-1\right)^2}{(3e^{\varepsilon}-1)(e^{\varepsilon}+1)} \\
&\overset{(d)}{\ge} \frac{(e-1)^2}{(3e-1)(e+1)} \ge 0.24,
\end{aligned} \tag{66}$$

where (a) holds by plugging in $\mathbb{E}[\theta] = s/d$, (b) holds since $s = \lceil d/(1+e^{\varepsilon})\rceil$ (so $\frac{1}{1+e^{\varepsilon}} \le \frac{s}{d} \le \frac{1}{1+e^{\varepsilon}} + \frac{1}{d}$), (c) holds since we only focus on the regime where $\varepsilon \le d-1$ (so $\frac{1}{d} \le \frac{1}{1+\varepsilon}$), and (d) holds by observing that $f(x) := \frac{(x-1)^2}{(3x-1)(x+1)}$

is an increasing function for $x \geq 1$ and we have $\varepsilon \geq 1$. Putting things together, we obtain

$$\frac{m_{\mathtt{ss}} - m_{\mathtt{mrc}}}{m_{\mathtt{mrc}}} = \frac{m_{\mathtt{ss}} - m_{\mathtt{mrc}}}{m_{\mathtt{ss}} - (m_{\mathtt{ss}} - m_{\mathtt{mrc}})} \overset{(a)}{\leq} \frac{\sqrt{\frac{\ln(2/\beta)}{2N}}(e^{\varepsilon}+3) + 2\beta}{0.24 - \left(\sqrt{\frac{\ln(2/\beta)}{2N}}(e^{\varepsilon}+3) + 2\beta\right)} \overset{(b)}{\leq} \lambda,$$

where (a) follows from (65) and (66) and (b) follows as long as

$$\sqrt{\frac{\ln(2/\beta)}{2N}}(e^{\varepsilon}+3) + 2\beta \leq \frac{0.24\lambda}{1+\lambda}. \tag{67}$$

To ensure (67), we let

$$\beta \leq \frac{0.24\lambda}{4(1+\lambda)} \qquad \text{and} \qquad N \geq \frac{1}{2}\left(\frac{(e^{\varepsilon}+3)}{\frac{0.24\lambda}{(1+\lambda)} - 2\beta}\right)^2 \ln(2/\beta) = \frac{2(e^{\varepsilon}+3)^2(1+\lambda)^2}{0.24^2\lambda^2} \ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right).$$

It is easy to verify that this choice of N satisfies $N \geq 2\ln(2/\beta)(1+e^{\varepsilon})^2$. □

H.2.2 Relationship between mean squared errors associated with Subset Selection and MRC simulating Subset Selection

In the following Proposition, we show that if $m_{\mathtt{mrc}}$ is close to $m_{\mathtt{ss}}$ and $b_{\mathtt{mrc}} \geq b_{\mathtt{ss}}$, then the mean squared error associated with MRC simulating Subset Selection (i.e., $\mathbb{E}_{q^{\mathtt{mrc}}}\left[\|\hat{\boldsymbol{x}}^{\mathtt{mrc}} - \boldsymbol{x}\|_2^2\right]$) is close to the mean squared error associated with Subset Selection (i.e., $\mathbb{E}_{q^{\mathtt{ss}}}\left[\|\hat{\boldsymbol{x}}^{\mathtt{ss}} - \boldsymbol{x}\|_2^2\right]$).

Proposition H.1. *Let $q^{ss}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP* Subset Selection *mechanism with estimator $\hat{\boldsymbol{x}}^{ss}$. Let $q^{mrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the* MRC *privatization mechanism simulating* Subset Selection *with N candidates and estimator $\hat{\boldsymbol{x}}^{mrc}$. Let m_{ss} and b_{ss} denote the scaling factors associated with* Subset Selection *and m_{mrc} and b_{mrc} denote the scaling factors associated with the* MRC *scheme simulating* Subset Selection. *Consider any $\lambda > 0$. If $m_{pu} - m_{mrc} \leq \lambda \cdot m_{mrc}$ and $b_{mrc} \geq b_{ss}$, then*

$$\mathbb{E}_{q^{mrc}}\left[\|\hat{\boldsymbol{x}}^{mrc} - \boldsymbol{x}\|_2^2\right] \leq \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right)\mathbb{E}_{q^{ss}}\left[\|\hat{\boldsymbol{x}}^{ss} - \boldsymbol{x}\|^2\right]$$

Proof. We have

$$\mathbb{E}_{q^{\mathtt{mrc}}}\left[\|\hat{\boldsymbol{x}}^{\mathtt{mrc}} - \boldsymbol{x}\|_2^2\right] \overset{(a)}{=} \sum_{i=1}^{d} \mathsf{Var}\left(\hat{\boldsymbol{x}}_i^{\mathtt{mrc}}\right) \overset{(b)}{=} \left(\frac{1}{m_{\mathtt{mrc}}}\right)^2 \sum_i \mathsf{Var}\left((\boldsymbol{z}_K)_i\right) \overset{(c)}{=} \left(\frac{1}{m_{\mathtt{mrc}}}\right)^2 \sum_i q_i^{\mathtt{mrc}}(1 - q_i^{\mathtt{mrc}}).$$

where (a) follows because $\boldsymbol{x}$ is a constant, (b) follows because $\hat{\boldsymbol{x}}_{\mathtt{mrc}} = (\boldsymbol{z}_K - b_{\mathtt{mrc}})/m_{\mathtt{mrc}}$, and (c) follows because $(\boldsymbol{z}_K)_i \sim \mathsf{Ber}(q_i^{\mathtt{mrc}})$. Similarly, we have We have

$$\mathbb{E}_{q^{\mathtt{ss}}}\left[\|\hat{\boldsymbol{x}}^{\mathtt{ss}} - \boldsymbol{x}\|_2^2\right] \overset{(a)}{=} \sum_{i=1}^{d} \mathsf{Var}\left(\hat{\boldsymbol{x}}_i^{\mathtt{ss}}\right) \overset{(b)}{=} \left(\frac{1}{m_{\mathtt{ss}}}\right)^2 \sum_i \mathsf{Var}\left(z_i\right) \overset{(c)}{=} \left(\frac{1}{m_{\mathtt{ss}}}\right)^2 \sum_i q_i^{\mathtt{ss}}(1 - q_i^{\mathtt{ss}}).$$

where (a) follows because $\boldsymbol{x}$ is a constant, (b) follows because $\hat{\boldsymbol{x}}^{\mathtt{ss}} = (\boldsymbol{z} - b_{\mathtt{ss}})/m_{\mathtt{ss}}$, and (c) follows because $z_i \sim \mathsf{Ber}(q_i^{\mathtt{ss}})$.

Now, let us look at the difference i.e.,

$$\begin{aligned}
&\mathbb{E}_{q^{\mathtt{mrc}}}\left[\|\hat{\boldsymbol{x}}^{\mathtt{mrc}} - \boldsymbol{x}\|_2^2\right] - \mathbb{E}_{q^{\mathtt{ss}}}\left[\|\hat{\boldsymbol{x}}^{\mathtt{ss}} - \boldsymbol{x}\|_2^2\right] \\
&= \left(\frac{1}{m_{\mathtt{mrc}}}\right)^2 \sum_i q_i^{\mathtt{mrc}}(1 - q_i^{\mathtt{mrc}}) - \left(\frac{1}{m_{\mathtt{ss}}}\right)^2 \sum_i q_i^{\mathtt{ss}}(1 - q_i^{\mathtt{ss}}) \\
&\leq \left(\frac{1}{m_{\mathtt{mrc}}}\right)^2 \sum_i \left(q_i^{\mathtt{mrc}}(1 - q_i^{\mathtt{mrc}}) - q_i^{\mathtt{ss}}(1 - q_i^{\mathtt{ss}})\right) + \left[\frac{1}{m_{\mathtt{mrc}}^2} - \frac{1}{m_{\mathtt{ss}}^2}\right]\left(\sum_i q_i^{\mathtt{ss}}(1 - q_i^{\mathtt{ss}})\right).
\end{aligned}$$

Now, first, we will bound $\left(\frac{1}{m_{\texttt{mrc}}}\right)^2 \sum_i (q_i^{\texttt{mrc}}(1-q_i^{\texttt{mrc}}) - q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}}))$. To that end, observe that $m_{\texttt{pu}} - m_{\texttt{mrc}} \leq \lambda \cdot m_{\texttt{mrc}}$ implies

$$\frac{1}{m_{\texttt{mrc}}} \leq (1+\lambda)\frac{1}{m_{\texttt{ss}}}. \tag{68}$$

Further, we have

$$\begin{aligned} q_i^{\texttt{mrc}} &\overset{(a)}{=} m_{\texttt{mrc}} p_i + b_{\texttt{mrc}} \overset{(b)}{=} q_i^{\texttt{ss}} + (m_{\texttt{mrc}} - m_{\texttt{ss}})p_i + (b_{\texttt{mrc}} - b_{\texttt{ss}}) \overset{(c)}{\geq} q_i^{\texttt{ss}} - \lambda \cdot m_{\texttt{mrc}} \cdot p_i + (b_{\texttt{mrc}} - b_{\texttt{ss}}) \\ &\overset{(d)}{\geq} q_i^{\texttt{ss}} - \lambda \cdot m_{\texttt{mrc}} \cdot p_i \overset{(e)}{\geq} q_i^{\texttt{ss}} - \lambda \cdot m_{\texttt{ss}} \cdot p_i \overset{(f)}{\geq} (1-\lambda) q_i^{\texttt{ss}}, \end{aligned} \tag{69}$$

where (a) follows from Lemma H.1, (b) follows from (54), (c) follows because $m_{\texttt{pu}} - m_{\texttt{mrc}} \leq \lambda \cdot m_{\texttt{mrc}}$, (d) follows because $b_{\texttt{mrc}} \geq b_{\texttt{ss}}$, (e) follows because $m_{\texttt{ss}} \geq m_{\texttt{mrc}}$ as seen in Lemma H.2, and (f) follows because $b_{\texttt{ss}} \geq 0$. Next, we have

$$\frac{q_i^{\texttt{mrc}}(1-q_i^{\texttt{mrc}}) - q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}})}{q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}})} = \frac{(q_i^{\texttt{ss}} - q_i^{\texttt{mrc}})(q_i^{\texttt{ss}} + q_i^{\texttt{mrc}} - 1)}{q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}})} \overset{(a)}{\leq} \frac{\lambda q_i^{\texttt{ss}}(q_i^{\texttt{ss}} + q_i^{\texttt{mrc}} - 1)}{q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}})} \overset{(b)}{\leq} \frac{\lambda}{1-q_i^{\texttt{ss}}} \tag{70}$$

where (a) follows from (69) and (b) follows since $q_i^{\texttt{ss}} \leq 1$ and $q_i^{\texttt{mrc}} \leq 1$.

Let us now upper bound $q_i^{\texttt{ss}}$. We have

$$q_i^{\texttt{ss}} = m_{\texttt{ss}} \cdot p_i + b_{\texttt{ss}} \overset{(a)}{\leq} m_{\texttt{ss}} + b_{\texttt{ss}} \overset{(b)}{=} \left(\frac{\mathbb{E}[\theta]e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])}\right) \overset{(c)}{\leq} \frac{1}{2} \tag{71}$$

where (a) follows because $p_i \leq 1$, (b) follows from (62) and (63), and (c) follows because $\mathbb{E}[\theta] = \frac{s}{d} \geq \frac{1}{e^{\varepsilon}+1}$. Combining (70) and (71), and then re-arranging results in

$$\sum_i q_i^{\texttt{mrc}}(1-q_i^{\texttt{mrc}}) - \sum_i q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}}) \leq 2\lambda \sum_i q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}}).$$

Together with (68), we obtain

$$\left(\frac{1}{m_{\texttt{mrc}}}\right)^2 \sum_i (q_i^{\texttt{mrc}}(1-q_i^{\texttt{mrc}}) - q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}})) \leq \frac{2\lambda(1+\lambda)^2}{m_{\texttt{ss}}^2} \sum_i q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}}).$$

To bound $\left[\frac{1}{m_{\texttt{mrc}}^2} - \frac{1}{m_{\texttt{ss}}^2}\right] (\sum_i q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}}))$, simply note that (68) implies $\frac{1}{m_{\texttt{mrc}}^2} \leq (1+\lambda)^2 \frac{1}{m_{\texttt{ss}}^2}$ resulting in

$$\left[\frac{1}{m_{\texttt{mrc}}^2} - \frac{1}{m_{\texttt{ss}}^2}\right] \left(\sum_i q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}})\right) \leq \frac{2\lambda + \lambda^2}{m_{\texttt{ss}}^2} \left(\sum_i q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}})\right).$$

Combining everything, we have

$$\mathbb{E}_{q^{\texttt{mrc}}}\left[\|\hat{\boldsymbol{x}}^{\texttt{mrc}} - \boldsymbol{x}\|_2^2\right] \leq \left(1 + 2\lambda(1+\lambda)^2 + 2\lambda + \lambda^2\right) \frac{1}{m_{\texttt{mrc}}^2} \sum_i q_i^{\texttt{ss}}(1-q_i^{\texttt{ss}}) = \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right) \mathbb{E}_{q^{\texttt{ss}}}\left[\|\hat{\boldsymbol{x}}^{\texttt{ss}} - \boldsymbol{x}\|^2\right]$$

□

In the following Lemma, we show that with on the order of ε-bits of communication, the mean squared error associated with `MRC` simulating `Subset Selection` (i.e., $\mathbb{E}_{q^{\texttt{mrc}}}\left[\|\hat{\boldsymbol{x}}^{\texttt{mrc}} - \boldsymbol{x}\|_2^2\right]$) is close to the mean squared error associated with `Subset Selection` (i.e., $\mathbb{E}_{q^{\texttt{ss}}}\left[\|\hat{\boldsymbol{x}}^{\texttt{ss}} - \boldsymbol{x}\|_2^2\right]$).

Lemma H.3. *Let $q^{ss}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP* `Subset Selection` *mechanism with estimator $\hat{\boldsymbol{x}}^{ss}$. Let $q^{mrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the* `MRC` *privatization mechanism simulating* `Subset Selection` *with N candidates and estimator $\hat{\boldsymbol{x}}^{mrc}$. Consider any $\lambda > 0$. Then,*

$$\mathbb{E}_{q^{mrc}}\left[\|\hat{\boldsymbol{x}}^{mrc} - \boldsymbol{x}\|_2^2\right] \leq \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right) \mathbb{E}_{q^{ss}}\left[\|\hat{\boldsymbol{x}}^{ss} - \boldsymbol{x}\|^2\right]$$

as long as

$$N \geq \frac{2(e^{\varepsilon}+3)^2(1+\lambda)^2}{0.24^2\lambda^2} \ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right).$$

Proof. The proof follows from Proposition H.1 and Lemma H.2. □

H.2.3 Simulating `Subset Selection` using Minimal Random Coding

The following Theorem shows that, for frequency estimation, `MRC` can simulate `Subset Selection` in a near-lossless manner (when λ is small) while only using on the order of ε bits of communication.

Theorem H.1. *Let* $r_{\mathsf{FE}}\left(\hat{\Pi}^{ss}, q^{ss}\right)$ *and* $r_{\mathsf{FE}}\left(\hat{\Pi}^{mrc}, q^{mrc}\right)$ *be the empirical frequency estimation error for* `Subset Selection` *and* `MRC` *simulating* `Subset Selection` *with* N *candidates respectively. Consider any* $\lambda > 0$. *Then*

$$r_{\mathsf{FE}}\left(\hat{\Pi}^{mrc}, q^{mrc}\right) \leq \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right) r_{\mathsf{FE}}\left(\hat{\Pi}^{ss}, q^{ss}\right),$$

as long as

$$N \geq \frac{2(e^{\varepsilon}+3)^2(1+\lambda)^2}{0.24^2\lambda^2} \ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right).$$

Proof. The proof follows directly from Lemma H.3 since for all $i \in [n]$, $\hat{\boldsymbol{x}}_i^{\mathtt{mrc}}$ are independent of each other as well as unbiased. □

H.3 Empirical Comparisons

In this section, we compare `MRC` simulating `Subset Selection` (using its approximate DP guarantee) against `Subset Selection` and RHR for frequency estimation with $d = 500$ and $n = 5000$. We use the same data generation scheme described in Section 5.3 and set $\delta = 10^{-6}$. As before, RHR uses #-bits $= \varepsilon$ because it leads to a poor performance if #-bits $> \varepsilon$. We show the privacy-accuracy tradeoffs for these three methods in Figure 5. We see that `MRC` simulating `Subset Selection` can attain the accuracy of the uncompressed `Subset Selection` for the range of ε's typically considered by LDP mechanisms while only using $(3\varepsilon/\ln 2) + 6$ bits. In comparison with the results from Section 5.3, the results in this section come with an approximate guarantee ($\delta = 10^{-6}$) and with a higher number of bits of communication. In other words, along with the obvious gains of pure privacy instead of approximate privacy, `MMRC` results in a lower communication cost (and therefore a lower computation cost) compared to `MRC`.

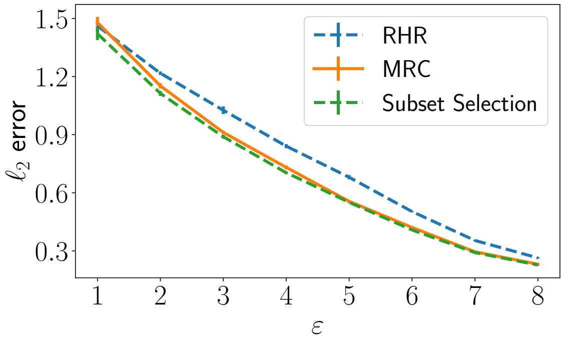

Figure 5: Comparing `Subset Selection`, `MRC` simulating `Subset Selection` and SQKR for frequency estimation in terms of ℓ_2 error vs ε with $d = 500$, $n = 5000$, and #bits $= (3\varepsilon/\ln 2) + 6$.

I MODIFIED MINIMAL RANDOM CODING SIMULATING `Subset Selection`

In this section, we prove Lemma 5.1 (in Appendix I.1) and Theorem 5.1 (in Appendix I.2.3). To prove Theorem 5.1, first, in Appendix I.2.1, we show that when the number of candidates N is exponential in ε, the scaling factor $m_{\mathtt{mmrc}}$ is close to the scaling parameter associated with `Subset Selection` (i.e., $m_{\mathtt{ss}}$). Next, in Appendix I.2.2, we provide the relationship between the mean squared error associated with `MMRC` simulating `Subset Selection` and the mean squared error associated with `Subset Selection`. Finally, in Appendix I.3, we provide some empirical comparisons in addition to the ones in Section 5.3 between `MMRC` simulating `Subset Selection` and `Subset Selection`.

I.1 Unbiased Modified Minimal Random Coding simulating `Subset Selection`

Consider the `Subset Selection` ε-LDP mechanism q^{ss} described in Section 2 with $s := \lceil \frac{d}{1+e^{\varepsilon}} \rceil$. `Subset Selection` is cap-based mechanism as discussed in Section 3 and Appendix G with $\mathsf{Cap}_{\boldsymbol{x}} = \mathcal{Z}_{\boldsymbol{x}}$ and $\mathbb{P}_{\boldsymbol{z} \sim \text{Unif}(\mathcal{Z})} \left(\boldsymbol{z} \in \mathsf{Cap}_{\boldsymbol{x}} \right) = s/d$. Let $\pi^{\texttt{mmrc}}$ be the distribution and $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_N$ be the candidates obtained from Algorithm 2 when the reference distribution is $\text{Unif}(\mathcal{Z})$ where $\mathcal{Z}$ is as defined in (51). Let θ denote the fraction of candidates inside $\mathsf{Cap}_{\boldsymbol{x}} = \mathcal{Z}_{\boldsymbol{x}}$ where $\mathcal{Z}_{\boldsymbol{x}}$ is the set of elements in $\mathcal{Z}$ with 1 in the same location as $\boldsymbol{x}$. It is easy to see that $\theta \sim \frac{1}{N} \mathsf{Binom} \left(N, \frac{s}{d} \right)$. Let $q_i^{\texttt{mmrc}} = \mathbb{P}(z_i = 1)$ where $\boldsymbol{z} \sim q^{\texttt{mmrc}}(\cdot | \boldsymbol{x})$ i.e., $q_i^{\texttt{mmrc}} = \mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 \right\}$ where $K \sim \pi^{\texttt{mmrc}}(\cdot)$.

Lemma I.1. *Let $K \sim \pi^{\mathit{mmrc}}(\cdot)$ and $q_i^{\mathit{mmrc}} = \mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 \right\}$ for $i \in [d]$. Then,*

$$q_i^{\mathit{mmrc}} = p_i m_{\mathit{mmrc}} + b_{\mathit{mmrc}}$$

where

$$m_{\mathit{mmrc}} := \frac{d}{d-1} \mathbb{E} \left[\frac{e^{\varepsilon} \theta}{e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1} \left(\theta \leq \mathbb{E}[\theta] \right) + \frac{e^{\varepsilon} \mathbb{E}[\theta] + \theta - \mathbb{E}[\theta]}{e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1} \left(\theta > \mathbb{E}[\theta] \right) \right] - \frac{s}{d-1} \tag{72}$$

$$b_{\mathit{mmrc}} := \frac{1}{d-1} \left(s - \mathbb{E} \left[\frac{e^{\varepsilon} \theta}{e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1} \left(\theta \leq \mathbb{E}[\theta] \right) + \frac{e^{\varepsilon} \mathbb{E}[\theta] + \theta - \mathbb{E}[\theta]}{e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1} \left(\theta > \mathbb{E}[\theta] \right) \right] \right). \tag{73}$$

Proof. Following the proof of Lemma H.1, we compute $\mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i \right\}$ and $\mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j \right\}$ separately.

To compute $\mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i \right\}$, recall that θ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$. From Appendix G.2, recall that $c_1(\varepsilon, d) := \frac{e^{\varepsilon}}{\binom{d-1}{s-1} e^{\varepsilon} + \binom{d-1}{s}}$, $c_2(\varepsilon, d) := \frac{1}{\binom{d-1}{s-1} e^{\varepsilon} + \binom{d-1}{s}}$. Further, since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\theta \sim \frac{1}{N} \mathsf{Binom} \left(N, \frac{\binom{d-1}{s-1}}{\binom{d}{s}} \right) = \frac{1}{N} \mathsf{Binom} \left(N, \frac{s}{d} \right),$$

so we have

$$\begin{aligned} \mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i \right\} &= \mathbb{P} \left\{ \boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}} | \boldsymbol{x} = i \right\} \overset{(a)}{=} \mathbb{E} \left[\mathbb{P} \left\{ \boldsymbol{z}_K \in \mathsf{Cap}_{\boldsymbol{x}} | \boldsymbol{x} = i, \theta \right\} \right] \\ &\overset{(b)}{=} \mathbb{E} \left[\frac{e^{\varepsilon} \theta}{e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1} \left(\theta \leq \mathbb{E}[\theta] \right) + \frac{e^{\varepsilon} \mathbb{E}[\theta] + \theta - \mathbb{E}[\theta]}{e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1} \left(\theta > \mathbb{E}[\theta] \right) \right] \end{aligned} \tag{74}$$

where (a) follows by the law of total probability and (b) is due to Algorithm 2 and $c_1(\varepsilon, d) / c_2(\varepsilon, d) = e^{\varepsilon}$.

To compute $\mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j \right\}$, we decompose it into

$$\mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j \right\} = \mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 | \boldsymbol{x} = j \right\} + \mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j \right\}, \tag{75}$$

for any $j \neq i$ and calculate each of the terms separately.

As before, let θ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$. Further, let $\bar{\theta}$ denotes the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$ as well as have 1 in the j^{th} location. Since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\bar{\theta} \sim \frac{1}{N} \mathsf{Binom} \left(N\theta, \frac{\binom{d-2}{s-2}}{\binom{d-1}{s-1}} \right) = \frac{1}{N} \mathsf{Binom} \left(N\theta, \frac{s-1}{d-1} \right),$$

so we have

$$\begin{aligned} &\mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 | \boldsymbol{x} = j \right\} \overset{(a)}{=} \mathbb{E}_{\theta} \left[\mathbb{E}_{\bar{\theta}} \left[\mathbb{P} \left\{ (\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 \middle| \boldsymbol{x} = j, \bar{\theta}, \theta \right\} \right] \right] \\ &\overset{(b)}{=} \mathbb{E}_{\theta} \left[\mathbb{E}_{\bar{\theta}} \left[\frac{e^{\varepsilon}}{e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \times \bar{\theta} \right] \mathbb{1} \left(\theta \leq \mathbb{E}[\theta] \right) + \mathbb{E}_{\bar{\theta}} \left[\frac{e^{\varepsilon} \mathbb{E}[\theta] + \theta - \mathbb{E}[\theta]}{\theta \left(e^{\varepsilon} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta]) \right)} \times \bar{\theta} \right] \mathbb{1} \left(\theta > \mathbb{E}[\theta] \right) \right] \end{aligned}$$

$$\overset{(c)}{=} \frac{s-1}{d-1}\mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right)}\cdot\mathbb{1}\left(\theta\leq\mathbb{E}\left[\theta\right]\right)+\frac{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\theta-\mathbb{E}\left[\theta\right]}{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right)}\cdot\mathbb{1}\left(\theta>\mathbb{E}\left[\theta\right]\right)\right]. \tag{76}$$

where (a) follows by the law of total probability, (b) follows from Algorithm 2, and (c) is because $\mathbb{E}[\bar{\theta}] = \frac{s-1}{d-1} \times \theta$.

Similarly, to compute the term $\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j\right\}$, let $\bar{\theta}$ denote the fraction of candidates that belong inside the $\mathsf{Cap}_{\boldsymbol{x}}$ i.e., have 1 in the same location as $\boldsymbol{x}$ as well as have 0 in the j^{th} location. Since $\boldsymbol{z}_k$ are generated uniformly at random,

$$\bar{\theta} \sim \frac{1}{N}\mathsf{Binom}\left(N(1-\theta), \frac{\binom{d-2}{s-1}}{\binom{d-1}{s}}\right) = \frac{1}{N}\mathsf{Binom}\left(N(1-\theta), \frac{s}{d-1}\right),$$

so we have

$$\begin{aligned}
\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j\right\} &\overset{(a)}{=} \mathbb{E}_{\theta}\left[\mathbb{E}_{\bar{\theta}}\left[\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 \middle| \boldsymbol{x} = j, \bar{\theta}, \theta\right\}\right]\right] \\
&\overset{(b)}{=} \mathbb{E}_{\theta}\left[\mathbb{E}_{\bar{\theta}}\left[\frac{\bar{\theta}}{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right)}\cdot\mathbb{1}\left(\theta>\mathbb{E}\left[\theta\right]\right)+\frac{\left(1-\mathbb{E}\left[\theta\right]\right)+\left(\mathbb{E}\left[\theta\right]-\theta\right)e^{\varepsilon}}{(1-\theta)(e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right))}\bar{\theta}\cdot\mathbb{1}\left(\theta\leq\mathbb{E}\left[\theta\right]\right)\right]\right] \\
&\overset{(c)}{=} \frac{s}{d-1}\mathbb{E}\left[\frac{(1-\theta)}{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right)}\cdot\mathbb{1}\left(\theta>\mathbb{E}\left[\theta\right]\right)+\frac{\left(1-\mathbb{E}\left[\theta\right]\right)+\left(\mathbb{E}\left[\theta\right]-\theta\right)e^{\varepsilon}}{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right)}\cdot\mathbb{1}\left(\theta\leq\mathbb{E}\left[\theta\right]\right)\right].
\end{aligned} \tag{77}$$

where (a) follows by the law of total probability, (b) follows from Algorithm 2, and (c) is because $\mathbb{E}[\bar{\theta}] = \frac{s}{d-1} \times \theta$. Using (76) and (77) in (75), we have

$$\begin{aligned}
\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\} &= \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 1 | \boldsymbol{x} = j\right\} + \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1, (\boldsymbol{z}_K)_j = 0 | \boldsymbol{x} = j\right\} \\
&= \frac{1}{d-1}\left(s - \mathbb{E}\left[\frac{e^{\varepsilon}\theta}{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right)}\cdot\mathbb{1}\left(\theta\leq\mathbb{E}\left[\theta\right]\right)+\frac{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\theta-\mathbb{E}\left[\theta\right]}{e^{\varepsilon}\mathbb{E}\left[\theta\right]+\left(1-\mathbb{E}\left[\theta\right]\right)}\cdot\mathbb{1}\left(\theta>\mathbb{E}\left[\theta\right]\right)\right]\right).
\end{aligned} \tag{78}$$

Combining everything, we have

$$q_i^{\text{mmrc}} = \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1\right\} = p_i \times \left[\mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = i\right\} - \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\}\right] + \mathbb{P}\left\{(\boldsymbol{z}_K)_i = 1 | \boldsymbol{x} = j\right\} \overset{(a)}{=} p_i m_{\text{mmrc}} + b_{\text{mmrc}}$$

where (a) follows from (74) and (78), and the definitions of m_{mmrc} and b_{mmrc}. □

Lemma 5.1. *Let $\hat{\boldsymbol{x}}^{mmrc}$ be the estimator of the* MMRC *mechanism simulating* `Subset Selection` *as defined above. Then, $\mathbb{E}[\hat{\boldsymbol{x}}^{mmrc}] = \boldsymbol{x}$.*

Proof. Given Lemma I.1, the proof follows from the proof of Lemma H.1. □

I.2 Utility of Modified Minimal Random Coding simulating `Subset Selection`

I.2.1 The scaling factors of `Subset Selection` and `MMRC` are close when N is of the right order

In the following Lemma, we show that when the number of candidates N is exponential in ε, then the scaling parameters associated with `Subset Selection` and the `MMRC` scheme simulating `Subset Selection` are close.

Lemma I.2. *Let N denote the number of candidates used in the* `MMRC` *scheme. Let $K \sim \pi^{mmrc}$ where π^{mmrc} is the distribution over the indices $[N]$ associated the* `MMRC` *scheme simulating* `Subset Selection`*. Consider any $\lambda > 0$. Then, the scaling factors m_{ss} and b_{ss} associated with* `Subset Selection` *and the scaling factors m_{mmrc} and b_{mmrc} associated with the* `MMRC` *scheme simulating* `Subset Selection` *are such that*

$$m_{ss} - m_{mmrc} \leq \lambda \cdot m_{mmrc}$$

and $b_{ss} \leq b_{mmrc}$ as long as

$$N \geq \frac{2(e^{\varepsilon}+1)^2(1+\lambda)^2}{0.24^2\lambda^2}\ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right).$$

Proof. The proof is similar to the proof of Lemma H.2. We only show the key steps here.

From (73) and (63), we have

$$\begin{aligned}
b_{\texttt{mmrc}} - b_{\texttt{ss}} &= \frac{1}{d-1} \cdot \frac{1}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} \cdot \mathbb{E}\left[e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right) + (\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta > \mathbb{E}\left[\theta\right]\right)\right] \\
&\overset{(a)}{\ge} \frac{1}{d-1} \cdot \frac{1}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} \cdot \mathbb{E}\left[(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right) + (\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta > \mathbb{E}\left[\theta\right]\right)\right] \\
&= \frac{1}{d-1} \cdot \frac{1}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} \cdot \mathbb{E}\left[(\mathbb{E}[\theta]-\theta)\right] = 0.
\end{aligned}$$

where (a) follows because $e^{\varepsilon} \ge 1$. From (72) and (62), we have

$$\begin{aligned}
m_{\texttt{ss}} - m_{\texttt{mmrc}} &= \frac{d}{d-1} \cdot \frac{1}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} \cdot \mathbb{E}\left[e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right) + (\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta > \mathbb{E}\left[\theta\right]\right)\right] \\
&\le \frac{d}{d-1} \cdot \frac{1}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} \cdot \mathbb{E}\left[e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)\right] \\
&\overset{(a)}{\le} \frac{2}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} \cdot \mathbb{E}\left[e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)\right] \qquad (79)
\end{aligned}$$

where (a) holds since $d \ge 2$. Next, we condition on the event $\mathcal{E} \coloneqq \left\{ |\mathbb{E}[\theta] - \theta| \le \sqrt{\frac{\ln(2/\beta)}{2N}} \right\}$, which has probability $\mathbb{P}_{\theta}\left\{\mathcal{E}\right\} \ge 1 - \beta$ by Hoeffding's inequality. We continue to upper bound (79):

$$\begin{aligned}
m_{\texttt{ss}} - m_{\texttt{mmrc}} &= 2\left(\mathbb{P}\left\{\mathcal{E}\right\}\mathbb{E}\left[\frac{e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])}\middle|\mathcal{E}\right] + \mathbb{P}\left\{\mathcal{E}^c\right\}\mathbb{E}\left[\frac{e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])}\middle|\mathcal{E}^c\right]\right) \\
&\overset{(a)}{\le} 2\left(\mathbb{E}\left[\frac{e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])}\middle|\mathcal{E}\right] + \beta\right) \\
&\overset{(b)}{\le} (1+e^{\varepsilon})\sqrt{\frac{\ln(2/\beta)}{2N}} + 2\beta
\end{aligned}$$

where (a) holds since

$$\frac{e^{\varepsilon}(\mathbb{E}[\theta]-\theta)\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} = \frac{e^{\varepsilon}\mathbb{E}[\theta]\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} - \frac{e^{\varepsilon}\theta\cdot \mathbb{1}\left(\theta \le \mathbb{E}\left[\theta\right]\right)}{e^{\varepsilon}\mathbb{E}[\theta] + (1-\mathbb{E}[\theta])} \le 1,$$

and (b) holds since $\mathbb{E}[\theta] = s/d \ge 1/(1+e^{\varepsilon})$.

The rest of the proof is similar to the proof of Lemma H.2. □

I.2.2 Relationship between the mean squared errors associated with `Subset Selection` and `MMRC` simulating `Subset Selection`

In the following Proposition, we show that if $m_{\texttt{mmrc}}$ is close to $m_{\texttt{ss}}$ and $b_{\texttt{mmrc}} \ge b_{\texttt{ss}}$, then the mean squared error associated with `MMRC` simulating `Subset Selection` (i.e., $\mathbb{E}_{q^{\texttt{mmrc}}}\left[\,\|\hat{\boldsymbol{x}}^{\texttt{mmrc}} - \boldsymbol{x}\|_2^2\,\right]$) is close to the mean squared error associated with `Subset Selection` (i.e., $\mathbb{E}_{q^{\texttt{ss}}}\left[\,\|\hat{\boldsymbol{x}}^{\texttt{ss}} - \boldsymbol{x}\|_2^2\,\right]$).

Proposition I.1. *Let $q^{ss}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP* `Subset Selection` *mechanism with estimator $\hat{\boldsymbol{x}}^{ss}$. Let $q^{mmrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the* `MMRC` *privatization mechanism simulating* `Subset Selection` *with N candidates and estimator $\hat{\boldsymbol{x}}^{mmrc}$. Let m_{ss} and b_{ss} denote the scaling factors associated with* `Subset Selection` *and m_{mmrc} and b_{mmrc} denote the scaling factors associated with the* `MMRC` *scheme simulating* `Subset Selection`*. Consider any $\lambda > 0$. If $m_{pu} - m_{mmrc} \le \lambda \cdot m_{mmrc}$ and $b_{mmrc} \ge b_{ss}$, then*

$$\mathbb{E}_{q^{mmrc}}\left[\,\|\hat{\boldsymbol{x}}^{mmrc} - \boldsymbol{x}\|_2^2\,\right] \le \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right)\mathbb{E}_{q^{ss}}\left[\|\hat{\boldsymbol{x}}^{ss} - \boldsymbol{x}\|^2\right]$$

Proof. The proof is similar to the proof of Proposition H.1. □

In the following Lemma, we show that with on the order of ε-bits of communication, the mean squared error associated with `MMRC` simulating `Subset Selection` (i.e., $\mathbb{E}_{q^{\texttt{mmrc}}}\big[\,\|\hat{\boldsymbol{x}}^{\texttt{mmrc}} - \boldsymbol{x}\|_2^2\,\big]$) is close to the mean squared error associated with `Subset Selection` (i.e., $\mathbb{E}_{q^{\texttt{ss}}}\big[\,\|\hat{\boldsymbol{x}}^{\texttt{ss}} - \boldsymbol{x}\|_2^2\,\big]$).

Lemma I.3. *Let $q^{ss}(\boldsymbol{z}|\boldsymbol{x})$ be the ε-LDP* `Subset Selection` *mechanism with parameters d and $s = \lceil \frac{d}{1+e^{\varepsilon}} \rceil$ and estimator $\hat{\boldsymbol{x}}^{ss}$. Let $q^{mmrc}(\boldsymbol{z}|\boldsymbol{x})$ denote the* `MMRC` *privatization mechanism simulating* `Subset Selection` *with N candidates and estimator $\hat{\boldsymbol{x}}^{mmrc}$ as defined above. Consider any $\lambda > 0$. Then,*

$$\mathbb{E}_{q^{mmrc}}\Big[\|\hat{\boldsymbol{x}}^{mmrc}-\boldsymbol{x}\|_2^2\Big]\leq(1+4\lambda+5\lambda^2+2\lambda^3)\mathbb{E}_{q^{ss}}\Big[\|\hat{\boldsymbol{x}}^{ss}-\boldsymbol{x}\|_2^2\Big],$$

as long as

$$N \geq \frac{2(e^{\varepsilon}+1)^2(1+\lambda)^2}{0.24^2\lambda^2} \ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right).$$

Proof. The proof follows from Proposition I.1 and Lemma I.2. □

I.2.3 Simulating `Subset Selection` using Modified Minimal Random Coding

Now, we provide a proof of Theorem 5.1.

Theorem 5.1. *Let $r_{\mathsf{FE}}\left(\hat{\Pi}^{ss}, q^{ss}\right)$ and $r_{\mathsf{FE}}\left(\hat{\Pi}^{mmrc}, q^{mmrc}\right)$ be the empirical mean estimation error for* `Subset Selection` *and* `MMRC` *simulating* `Subset Selection` *with N candidates respectively. Consider any $\lambda > 0$. Then*

$$r_{\mathsf{FE}}\left(\hat{\Pi}^{mmrc}, q^{mmrc}\right) \leq \left(1+4\lambda+5\lambda^2+2\lambda^3\right) r_{\mathsf{FE}}\left(\hat{\Pi}^{ss}, q^{ss}\right),$$

as long as

$$N \geq \frac{2(e^{\varepsilon}+1)^2(1+\lambda)^2}{0.24^2\lambda^2} \ln\left(\frac{8(1+\lambda)}{0.24\lambda}\right). \tag{13}$$

Proof. The proof follows directly from Lemma I.3 since for all $i \in [n]$, $\hat{\boldsymbol{x}}_i^{\texttt{mmrc}}$ are independent of each other as well as unbiased. □

I.3 Additional Empirical Comparisons

In Section 5.3, we empirically demonstrated the privacy-accuracy-communication tradeoffs of `MMRC` simulating `Subset Selection` against `Subset Selection` and RHR in terms of ℓ_2 error vs #bits and ℓ_2 error vs ε (see Figure 2). In this section, we provide comparisons between these methods in terms of ℓ_2 error vs d (see Figure 6 (left)) and ℓ_2 error vs n (see Figure 6 (right)) for a fixed ε (=6) and a fixed #bits (=14). As before, RHR uses #bits = ε for both because it leads to a poor performance if #bits > ε.

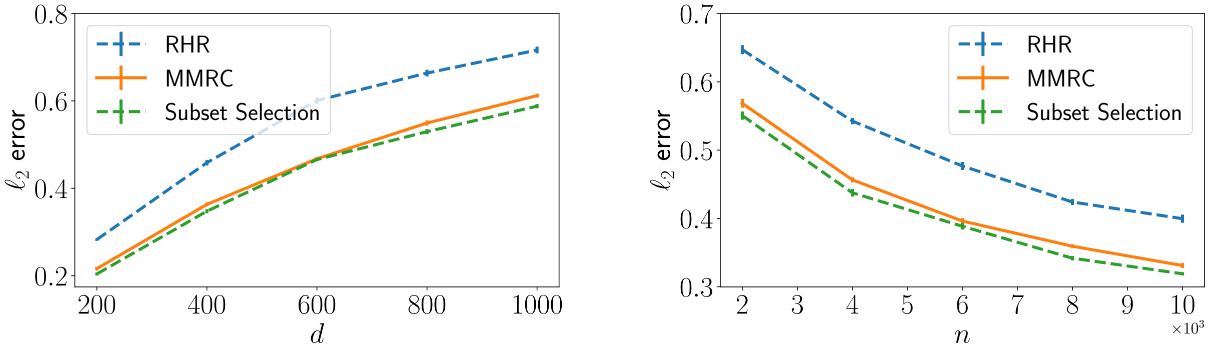

Figure 6: Comparing `Subset Selection`, `MMRC` simulating `Subset Selection` and RHR for frequency estimation with $\varepsilon = 6$ and #bits = 14. **Left:** ℓ_2 error vs d for $n = 5000$. **Right:** ℓ_2 error vs n for $d = 500$.